

The Plant Sciences

*Series Editors:* Mark Tester · Richard Jorgensen

Russell K. Monson *Editor*

# Ecology and the Environment

---

# The Plant Sciences

## **Series Editors**

Mark Tester  
King Abdullah University of Science & Technology  
Thuwal, Saudi Arabia

Richard Jorgensen  
School of Plant Sciences  
University of Arizona  
Tucson, AZ, USA

The volumes in this series form the world's most comprehensive reference on the plant sciences. Composed of ten volumes, *The Plant Sciences* provides both background and essential information in plant biology, exploring such topics as genetics and genomics, molecular biology, biochemistry, growth and development, and ecology and the environment. Available through both print and online mediums, the online text will be continuously updated to enable the reference to remain a useful authoritative resource for decades to come.

With broad contributions from internationally well-respected scientists in the field, *The Plant Sciences* is an invaluable reference for upper-division undergraduates, graduate students, and practitioners looking for an entry into a particular topic.

### **Series Titles**

1. Genetics and Genomics
2. Molecular Biology
3. Biochemistry
4. Cell Biology
5. Growth and Development
6. Physiology and Function
7. Biotic Interactions
8. Ecology and the Environment
9. Evolution, Systematics and Biodiversity
10. Applications

More information about this series at: <http://www.springer.com/series/11785>

---

Russell K. Monson  
Editor

# Ecology and the Environment

With 196 Figures and 21 Tables

 Springer Reference



*Editor*

Russell K. Monson  
School of Natural Resources and the Environment and  
Laboratory for Tree Ring Research  
University of Arizona  
Tucson, USA  
and  
Professor Emeritus, Ecology and Evolutionary Biology  
University of Colorado  
Boulder, CO  
USA

ISBN 978-1-4614-7500-2                      ISBN 978-1-4614-7501-9 (eBook)  
ISBN 978-1-4614-7502-6 (print and electronic bundle)  
DOI 10.1007/978-1-4614-7501-9  
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014948194

© Springer Science+Business Media New York 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

---

## Preface

The content of this volume is intended to place plants within the context of their surrounding environment, including both biotic and abiotic interactions. Interactions between plants and their environment occur across multiple scales in space and time, and as the Editor of the volume, I strived to invite and assemble a series of chapters that cover interactive scales from the organism to the ecosystem and that are driven by processes spanning seconds to decades. Understanding the fact that plant–environment interactions span multiple spatiotemporal scales and that the processes that control these interactions change with scale is a useful point of departure for deeper investigations within the field of ecology. This understanding lies at the foundation of advanced topics such as plant–environment feedbacks, nonlinear responses of plants to climate change, extinction dynamics of plants in fragmented landscapes, and earth system modeling. Starting from this point of understanding, we can develop strategies for effective management and conservation of natural resources in the face of the daunting environmental challenges that we face as a global society. The continuity of topics from fundamental ecology to sustainable protection of ecosystems is crucial as a theme and pedagogic framework in the academic courses offered to undergraduate students in the plant sciences. Nearly all topics involving plant ecology can be developed within the conceptual framework of spatiotemporal scaling. This book has been prepared with this conceptual framework in mind. In all chapters, we have tried to make connections from smaller to larger scales of ecological organization. We tried to communicate the fundamental nature of these connections in as simple and clear a manner as was possible as a means to reach mid-program to advanced-program undergraduate students, the primary intended audiences for this book.

The book is divided informally into three sections. In the first eight chapters, fundamental principles of plant–environment interactions are discussed.

In the first chapter, Reichstein et al. provide an overview of the scales and types of interactions that determine how plants respond to their environments. Topics in this chapter extend from global productivity to organismic phenology. A common theme is control over organism and ecosystem dynamics by climate, and an emphasis is placed on integrating observations with computer modeling as a means of understanding ecological processes across multiple scales.

The chapter by Bierzychudek takes up the topic of plant populations and the factors that control their persistence. Important factors discussed in this chapter

include the importance of population size to the maintenance of genetic diversity and controls over population resilience in the face of environmental change. A link is established between the reproductive success of individual plants and population dynamics – once again bridging scales by which we consider ecological interactions.

Kraft and Ackerly consider the ecological “rules” by which plant communities are formed. These rules can be traced to the nature of traits and interactions of species, especially those that determine interspecific competition and facilitation.

The chapter by Linhart delves into the processes by which plants are pollinated and seeds are dispersed. These processes largely determine patterns of species migration and are important determinants for the rate of species evolution.

Pham and McConnaughay discuss the potential for adaptive “plasticity” in the expression of plant traits given environmental variation. This critical link between a plant’s genotype and phenotype explains much about the limits to stress tolerance in plant populations, their capacity to adjust to short- and long-term changes in climate, and their ability to expand into new environments and community niches.

Trowbridge provides an evolutionary context for plant–insect chemical interactions, emphasizing the two-way nature of a chemical “arms race” in which the chemical defenses in plants must change over time to stay one step ahead of insects that are on their own evolutionary trajectories to resist plant defenses.

The chapter by Lipson and Kelley focuses on the belowground ecology of plants, particularly those interactions between roots and microorganisms. Belowground plant and microbe ecology provides the foundations for understanding the recycling of nutrients through decomposition and the processes that ultimately determine the sustainable nature of soil and its associated biogeochemical cycles. In the ecological research community, considerable effort has recently been devoted to understanding the links between biogeochemical cycles defined at biome-to-global scales and the specific microbial “species” that control soil processes.

Finally, Knapp et al. provide a chapter on abiotic and biotic controls over primary production. Primary production ultimately sustains all global food webs and determines the balance of carbon that is exchanged between ecosystems and the atmosphere – a relationship with important implications for global climate change.

The next nine chapters focus on specific types of ecosystems and cover the unique abiotic and biotic factors that control ecosystem integrity and determine key vulnerabilities that threaten sustainable persistence.

Gallery provides a chapter on tropical forests, emphasizing the wealth of biodiversity contained in these ecosystems and its importance for the stability of ecosystem processes. Maintenance of high levels of biodiversity in the face of increased human exploitation of tropical forests and the emergence of abiotic stresses associated with climate warming and drying in equatorial regions has produced grand challenges for those interested in the conservation and sustainable management of these ecosystems – which are among the earth’s most magnificent.

The chapter by Monson covers the ecology of mid-latitude, northern hemisphere forests – often called “temperate forests.” After discussing fundamental processes of primary production and nutrient cycling, he takes up the issue of recent changes

in climate that threaten forest sustenance through increased frequencies of large-scale insect attacks, increased numbers and sizes of wildfires, and exploitation of wood and water resources.

Sandquist takes up the topic of plants in desert ecosystems. He develops the concept that plants have evolved highly unique adaptive strategies to deal with the extremes of heat and drought in desert climates. The novel nature of desert plant adaptations has fueled the curiosity of plant ecologists for the last two millennia and provides clear examples of how form and function must be considered together as the “adaptive clay” that is sculpted by natural selection.

The chapter by Germino takes us to another extreme of environmental tolerance – that of the short growing seasons and cold temperatures in alpine ecosystems. Plants in these ecosystems have evolved unique morphological forms that allow them to persist in the warmer surface boundary layer next to the ground and thus become uncoupled from the cold temperatures that occur higher up. In both deserts and alpine ecosystems, seedling establishment is difficult and infrequent, and so disturbance due to biotic and abiotic stresses have the potential to exert long-term impacts on community composition and ecosystem processes.

The chapter by Peterson takes us to another example of abiotic extremes in discussing the ecology of arctic ecosystems. In these high-latitude regions, cold temperatures slow the rate of decomposition and create extremely low levels of soil fertility. Animals take on novel facilitative roles that redistribute and recycle nutrients, and unique plant adaptations have evolved to provide access to nutrient sources that are not commonly used in temperate ecosystems.

Blair et al. discuss the nature of grasslands. Grassland communities have high root-to-shoot ratios and are maintained by climate, fire, and frequent disturbance due to grazing. Together, these processes provide natural impediments to the invasion of woody species. However, when these natural mechanisms break down due to overgrazing or landscape fragmentation from human land use, community dynamics can shift, allowing invasion of both woody and nonwoody exotic species. This chapter on grassland ecology provides a nice case study on the challenges we face due to species invasions into novel niches.

Moving to the boundary between terrestrial and ocean ecosystems, Armitage considers the nature of coastal wetlands and in particular salt marshes and mangrove swamps. As in the case for desert, alpine, and arctic ecosystems, the saline extremes of these coastal wetlands has produced a type of vegetation with unique adaptations – in this case, adaptations to avoid or tolerate salt uptake. These ecosystems are extremely vulnerable to the deposition of pollution from human industries. The direct and indirect effects of this pollution create imbalances in the availability of oxygen and nutrients, which in turn reduce plant productivity and threaten food webs.

Kirkman discusses the nature of immersed seagrass ecosystems, moving our perspective even further offshore. Seagrass communities are among the most valuable on earth for providing goods and services valued by humans – they represent the natural hatcheries for our most valued seafood fishes. Though the term “seagrass” would suggest ecosystems based on a monotypic life form, here we

find some of the most biologically diverse communities on earth. The same pollution that threatens near coastal wetlands and swamps, however, has caused an unraveling of natural species interactions in seagrass ecosystems and has destabilized the hidden mechanisms that sustain diversity and community structure.

Finally, Geider et al. take us to the deeper ocean biomes, where phytoplankton ecology emerges as the primary topic of plant–environment interactions. In a rather comprehensive treatment, these authors provide details of how marine algae tolerate the near-surface ocean environment characterized by high solar radiation and low nutrient availability, how oceanographers study these interactions, and how excess nutrient burdens, climate change, and increases in acidity are capable of changing ocean productivity and altering the global carbon cycle.

In the final four chapters of the book, we consider some of the issues associated with plants and their role in environmental sustainability.

Leakey tackles the issue of recent increases in the mean global atmospheric CO<sub>2</sub> concentration and its influence on plant photosynthesis and the efficiency by which water is used. He discusses this topic from the foundations of photosynthetic biochemistry and stomatal function and describes how environmental changes in the atmospheric CO<sub>2</sub> concentration interact with these processes to influence crop yield and food security.

Wiedinmyer et al. provide a chapter on plant volatile organic compound emissions and their influences on air quality. In particular, they consider recent increases in the production of tropospheric ozone and atmospheric aerosols, both of which affect global climate. It has been known for several decades that the emission of volatile organic compounds from forests can affect a vast number of atmospheric chemical reactions. However, the final products of these reactions, such as ozone and aerosols, have been difficult to quantify primarily because the chemistry has been studied in theoretical terms. We are just now beginning to accumulate the results from field campaigns and studies of forests such that accurate quantitative predictions are becoming possible. This issue is also relevant to our expanded reliance on global agriforests for wood, pulp, and energy production. Most agriforest tree species emit relatively high amounts of reactive volatile organic compounds and are thus capable of affecting regional and global air quality.

O’Keefe et al. discuss the development of cellulosic biofuels as an alternative to our reliance on fossil fuels. Consideration of biofuels within the context of environmental impacts must be generated from knowledge of total resource use and the potential for hidden resource costs. These authors take on the complexities of this issue and consider the costs of biofuel production in comprehensive terms – including the costs of water, nutrients, and overall energy.

In the final chapter of the book, Hamilton provides a new framework for sustainability science. He focuses specifically on the need for integration of knowledge on natural systems such as that provided in the preceding chapters into the social, economic, and political discussions that ultimately determine how we manage our natural resources. His chapter brings us to the conclusion that “human well-being” is intricately tied to the relations between societies and natural

ecosystems and that this nexus, with human well-being as a central concern, should be the focus of strategies for action that improve natural resource management.

As a member of the “baby-boom” generation, I have observed immense changes in the earth system over the past five decades. The population of the earth has nearly doubled since the year of my birth. From hindsight, it is clear that as the population of the earth has expanded, the margin for error in how we manage our natural and agricultural ecosystems has contracted. As future generations take on the responsibility for managing our natural resources, one of the most effective things we can contribute is our accumulated knowledge – organized in a way that educates them and allows them to avoid some of the catastrophic mistakes that prior generations have made. This book hopefully provides some movement in that direction. Although a tendency often exists to attack a problem at the scale of its impact, knowledge of the processes and interactions that lie beneath the scale of impact will often lead to better-informed solutions – from the bottom-up. Hopefully, the emphasis on processes and interactions that cross all scales of plant–environment interaction, which we have tried to produce in this book, will contribute to future solutions.

Tucson, AZ, USA  
June 2014

Russell K. Monson



---

## Series Preface

Plant sciences is in a particularly exciting phase, with the tools of genomics, in particular, turbo-charging advances in an unprecedented way. Furthermore, with heightened attention being paid to the need for increased production of crops for food, feed, fuel, and other needs and for this to be done both sustainably and in the face of accelerating environmental change, plant science is arguably more important and receiving more attention than ever in history. As such, the field of plant sciences is rapidly changing, and this requires new approaches for the teaching of this field and the dissemination of knowledge, particularly for students. Fortunately, there are also new technologies to facilitate this need.

In this 10-volume series, *The Plant Sciences*, we aim to develop a comprehensive online and printed reference work. This is a new type of publishing venture exploiting Wiki-like capabilities, thus creating a dynamic, exciting, cutting-edge, and living entity.

The aim of this large publishing project is to produce a comprehensive reference in plant sciences. *The Plant Sciences* will be published both in print and online; the online text can be updated to enable the reference to remain a useful authoritative resource for decades to come. The broader aim is to provide a sustainable super-structure on which can be built further volumes or even series as plant science evolves. The first edition will contain 10 volumes.

*The Plant Sciences* is part of SpringerReference, which contains all Springer reference works. Check out the link at <http://www.springerreference.com/docs/index.html#Biomedical+and+Life+Sciences-lib1>, from where you can see the volumes in this series that are already coming online.

The target audience for the initial 10 volumes is upper-division undergraduates as well as graduate students and practitioners looking for an entry on a particular topic. The aim is for *The Plant Sciences* to provide both background and essential information in plant biology. The longer-term aim is for future volumes to be built (and hyperlinked) from the initial set of volumes, particularly targeting the research frontier in specific areas.

*The Plant Sciences* has the important extra dynamic dimension of being continually updated. *The Plant Sciences* has a constrained Wiki-like capability, with all original authors (or their delegates) being able to modify the content.

Having satisfied an approval process, new contributors will also be registered to propose modifications to the content.



It is expected that new editions of the printed version will be published every 3–5 years. The project is proceeding volume by volume, with volumes appearing as they are completed. This also helps to keep the text fresher and the project more dynamic.

We would like to thank our host institutions, colleagues, students, and funding agencies, who have all helped us in various ways and thus facilitated the development of this series. We hope this volume is used widely and look forward to seeing it develop further in the coming years.

King Abdullah University of Science & Technology,  
Thuwal, Saudi Arabia

Mark Tester

School of Plant Sciences, University of Arizona,  
Tucson, AZ, USA  
22 July 2014

Richard Jorgensen

---

## Editor Biography



**Russell K. Monson** is Louise Foucar Marshall Professor at the University of Arizona, Tucson, and Professor Emeritus at the University of Colorado, Boulder. In recognition of his past research and writings, he has been awarded several fellowships, including the John Simon Guggenheim Fellowship, the Fulbright Senior Fellowship, and the Alexander von Humboldt Fellowship. He is an elected fellow of the American Geophysical Union. Professor Monson's research is focused on forest carbon cycling, photosynthetic metabolism, and the production of biogenic volatile organic compounds from forest ecosystems. Professor Monson is Editor-in-Chief of the journal *Oecologia*, an international journal on ecology, and he has authored or coauthored over 200 peer-reviewed publications.

---

## Series Editors Biography



**Mark Tester** is Professor of Bioscience in the Center for Desert Agriculture and the Division of Biological and Environmental Sciences and Engineering, King Abdullah University for Science and Technology (KAUST), Saudi Arabia. He was previously in Adelaide, where he was a Research Professor in the Australian Centre for Plant Functional Genomics and Director of the Australian Plant Phenomics Facility. Mark led the establishment of this Facility, a \$55 m organisation that develops and delivers state-of-the-art phenotyping facilities, including The Plant Accelerator, an innovative plant growth and analysis facility. In Australia, he led a research group in which forward and reverse genetic approaches were used to understand salinity tolerance and how to improve this in crops such as wheat and barley. He moved to KAUST in February 2013, where this work is continuing, expanding also into work on the salinity tolerance of tomatoes.

Mark Tester has established a research program with the aim of elucidating the molecular mechanisms that enable certain plants to thrive in sub-optimal soil conditions, in particular in soils with high salinity. The ultimate applied aim is to modify crop plants in order to increase productivity on such soils, with consequent improvement of yield in both developed and developing countries. The ultimate intellectual aim is to understand the control and co-ordination of whole plant

function through processes occurring at the level of single cells, particularly through processes of long-distance communication within plants.

A particular strength of Professor Tester's research programme is the integration of genetics and genomics with a breadth of physiological approaches to enable novel gene discovery. The development and use of tools for the study and manipulation of specific cell types adds a useful dimension to the research. Professor Tester received training in cell biology and physiology, specialising in work on ion transport, particularly of cations across the plasma membrane of plant cells. His more recent focus on salinity tolerance is driven by his desire to apply his training in fundamental plant processes to a problem of practical significance.

Professor Tester was awarded a Junior Research Fellowship from Churchill College, Cambridge in 1988, a BBSRC (UK) Research Development Fellowship in 2001, and an Australian Research Council Federation Fellowship in 2004. Professor Tester obtained his Bachelor's degree in botany from the University of Adelaide in 1984, and his PhD in biophysics from the University of Cambridge in 1988.



**Dr. Richard Jorgensen**, Professor Emeritus, School of Plant Sciences, University of Arizona, Tucson, AZ, USA

Dr. Jorgensen is a recognized international leader in the fields of epigenetics, functional genomics, and computational biology. His research accomplishments include the discovery in plants of a gene-silencing phenomenon called cosuppression, which led to the discovery in animals of RNA interference, a gene-silencing tool that has major potential implications for medicine including the treatment of diseases such as cancer, hepatitis, and AIDS. In 2007, he was awarded the Martin Gibbs Medal for this groundbreaking work in cosuppression and RNAi by the American Society of Plant Biologists (ASPB). He was elected a

Fellow of the American Association for the Advancement of Sciences (AAAS) in 2005 and an Inaugural Fellow of the ASPB in 2007.

Dr. Jorgensen was the founding Director of the iPlant Collaborative, a 5 years, \$50 M NSF project to develop cyber-infrastructure for plant sciences. Dr. Jorgensen also served as the Editor in Chief of *The Plant Cell*, the leading research journal in plant biology, from 2003 to 2007. He is currently Editor in Chief of *Frontiers in Plant Science*, a cutting-edge, open-access journal allied with Nature Publishing Group. He is also Series Editor for the book series *Plant Genetics and Genomics: Crops and Models* for Springer Publishing. Dr. Jorgensen has published numerous scientific articles and is regularly invited to present his research findings at universities, research institutions, and scientific conferences nationally and internationally.





---

# Contents

<b>1 Plant–Environment Interactions Across Multiple Scales</b> . . . . .	1
Markus Reichstein, Andrew D. Richardson, Mirco Migliavacca, and Nuno Carvalhais	
<b>2 Plant Biodiversity and Population Dynamics</b> . . . . .	29
Paulette Bierzychudek	
<b>3 Assembly of Plant Communities</b> . . . . .	67
Nathan J. B. Kraft and David D. Ackerly	
<b>4 Plant Pollination and Dispersal</b> . . . . .	89
Yan Linhart	
<b>5 Plant Phenotypic Expression in Variable Environments</b> . . . . .	119
Brittany Pham and Kelly McConnaughay	
<b>6 Evolutionary Ecology of Chemically Mediated Plant-Insect Interactions</b> . . . . .	143
Amy M. Trowbridge	
<b>7 Plant-Microbe Interactions</b> . . . . .	177
David A. Lipson and Scott T. Kelley	
<b>8 Patterns and Controls of Terrestrial Primary Production in a Changing World</b> . . . . .	205
Alan K. Knapp, Charles J. W. Carroll, and Timothy J. Fahey	
<b>9 Ecology of Tropical Rain Forests</b> . . . . .	247
Rachel E. Gallery	
<b>10 Ecology of Temperate Forests</b> . . . . .	273
Russell K. Monson	
<b>11 Plants in Deserts</b> . . . . .	297
Darren R. Sandquist	
<b>12 Plants in Alpine Environments</b> . . . . .	327
Matthew J. Germino	



---

<b>13</b>	<b>Plants in Arctic Environments</b> .....	363
	Kim M. Peterson	
<b>14</b>	<b>Grassland Ecology</b> .....	389
	John Blair, Jesse Nippert, and John Briggs	
<b>15</b>	<b>Coastal Wetland Ecology and Challenges for Environmental Management</b> .....	425
	Anna R. Armitage	
<b>16</b>	<b>Near-Coastal Seagrass Ecosystems</b> .....	457
	Hugh Kirkman	
<b>17</b>	<b>Ecology of Marine Phytoplankton</b> .....	483
	Richard J. Geider, C. Mark Moore, and David J. Suggett	
<b>18</b>	<b>Plants in Changing Environmental Conditions of the Anthropocene</b> .....	533
	Andrew D. B. Leakey	
<b>19</b>	<b>Plant Influences on Atmospheric Chemistry</b> .....	573
	Christine Wiedinmyer, Allison Steiner, and Kirsti Ashworth	
<b>20</b>	<b>Biofuel Development from Cellulosic Sources</b> .....	601
	Kimberly O’Keefe, Clint J. Springer, Jonathan Grennell, and Sarah C. Davis	
<b>21</b>	<b>Plant Ecology and Sustainability Science</b> .....	631
	Jason G. Hamilton	
	<b>Index</b> .....	655

---

## Contributors

**David D. Ackerly** Department of Integrative Biology, University of California, Berkeley, CA, USA

**Anna R. Armitage** Department of Marine Biology, Texas A&M University at Galveston, Galveston, TX, USA

**Kirsti Ashworth** Ecosystems–Atmosphere Interactions Group, Karlsruhe Institute of Technology, Garmisch–Partenkirchen, Germany

Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA

**Paulette Bierzychudek** Department of Biology, MSC 53, Lewis & Clark College, Portland, OR, USA

**John Blair** Division of Biology, Kansas State University, Manhattan, KS, USA

**John Briggs** Division of Biology, Kansas State University, Manhattan, KS, USA

**Charles J. W. Carroll** Graduate Degree Program in Ecology, Department of Biology, Colorado State University, Fort Collins, CO, USA

**Nuno Carvalhais** Department of Biogeochemical Integration, Max–Planck–Institute for Biogeochemistry, Jena, Germany

Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, Caparica, Portugal

**Sarah C. Davis** Voinovich School of Leadership and Public Affairs, Ohio University, Athens, OH, USA

**Timothy J. Fahey** Department of Natural Resources, Cornell University, Ithaca, NY, USA

**Rachel E. Gallery** School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

**Richard J. Geider** School of Biological Sciences, University of Essex, Colchester, Essex, UK

**Matthew J. Germino** Forest and Rangeland Ecosystem Science Center, US Geological Survey, Boise, ID, USA

**Jonathan Grennell** Voinovich School of Leadership and Public Affairs, Ohio University, Athens, OH, USA

**Jason G. Hamilton** Department of Environmental Studies and Sciences, Ithaca College, Ithaca, NY, USA

**Scott T. Kelley** Department of Biology, San Diego State University, San Diego, CA, USA

**Hugh Kirkman** Australian Marine Ecology Pty Ltd, Kensington, Australia

**Alan K. Knapp** Graduate Degree Program in Ecology, Department of Biology, Colorado State University, Fort Collins, CO, USA

**Nathan J. B. Kraft** Department of Biology, University of Maryland, College Park MD, USA

**Andrew D. B. Leakey** Department of Plant Biology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

**Yan Linhart** Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

**David A. Lipson** Department of Biology, San Diego State University, San Diego, CA, USA

**Kelly McConnaughay** Department of Biology, Bradley University, Peoria, IL, USA

**Mirco Migliavacca** Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany

Department of Earth and Environmental Science, University of Milano-Bicocca, Milan, Italy

**Russell K. Monson** School of Natural Resources and the Laboratory for Tree Ring Research, University of Arizona, Tucson, AZ, USA

**C. Mark Moore** Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK

**Jesse Nippert** Division of Biology, Kansas State University, Manhattan, KS, USA

**Kimberly O'Keefe** Division of Biology, Kansas State University, Manhattan, KS, USA

**Kim M. Peterson** Department of Biological Sciences, University of Alaska Anchorage, Anchorage, AK, USA

**Brittany Pham** Department of Biology, Bradley University, Peoria, IL, USA

---

**Markus Reichstein** Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany

**Andrew D. Richardson** Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

**Darren R. Sandquist** Department of Biological Science, California State University, Fullerton, CA, USA

**Clint J. Springer** Department of Biology, Saint Joseph's University, Philadelphia, PA, USA

**Allison Steiner** Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA

**David J. Suggett** Functional Plant Biology & Climate Change Cluster, University of Technology, Sydney, NSW, Australia

**Amy M. Trowbridge** Department of Biology, Indiana University, Bloomington, IN, USA

**Christine Wiedinmyer** Atmospheric Chemistry Division, NCAR Earth System Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

---

# Plant–Environment Interactions Across Multiple Scales

# 1

Markus Reichstein, Andrew D. Richardson, Mirco Migliavacca,  
and Nuno Carvalhais

## Contents

Environmental Controls on Vegetation: Introduction .....	3
Environmental Controls: Climate .....	5
Environmental Controls: CO <sub>2</sub> , O <sub>3</sub> , Pollutants, and Nitrogen Deposition .....	7
Ozone and Air Pollutants .....	9
Soil Properties .....	9
Animals Including Humans .....	10
Plant Responses to the Environment .....	11
Influences of Vegetation on Environment .....	12
Microclimate .....	12
Transpiration .....	13
Surface Energy Budget .....	14

---

M. Reichstein (✉)

Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany

e-mail: [mreichstein@bgc-jena.mpg.de](mailto:mreichstein@bgc-jena.mpg.de)

A.D. Richardson

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

e-mail: [arichardson@oeb.harvard.edu](mailto:arichardson@oeb.harvard.edu)

M. Migliavacca

Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany

Department of Earth and Environmental Science, University of Milano-Bicocca, Milan, Italy

e-mail: [mmiglia@bgc-jena.mpg.de](mailto:mmiglia@bgc-jena.mpg.de)

N. Carvalhais

Department of Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany

Departamento de Ciências e Engenharia do Ambiente, DCEA, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, Caparica, Portugal

e-mail: [ncarval@bgc-jena.mpg.de](mailto:ncarval@bgc-jena.mpg.de)

Biogeochemical Cycling, Including Carbon .....	15
Emissions of Biogenic Volatile Organic Compounds (VOCs) .....	17
Observation Strategies .....	17
Classical Observations: Surveys, Biometry, and Tree Rings .....	18
Flux Measurements .....	19
Remote Sensing .....	19
Atmospheric Observation of Trace Gases .....	21
Modeling Strategies .....	21
From Leaf Level to Community Dynamics .....	22
Bringing Models and Observations Together .....	23
Representing Ecosystem Functioning from Local to Regional Scales .....	24
Future Directions .....	25
References .....	25

## Abstract

- It has been known for a long time that the environment shapes the appearance of vegetation (vegetation structure). The systematic description of these effects has led to classifications of life forms at the organismic scale and biomes at the global scale by Alexander von Humboldt, Christen C. Raunkiær, Wladimir Köppen, and other early plant geographers and plant ecologists.
- Consequently, plant traits and processes carried out by plants (vegetation function) are influenced by climate and other environmental conditions. However, given the previous limitations of both observations and theory, systematic and comparative studies of plant ecology and physiological ecology only began in the twentieth century.
- Through their adaptive and genetic constitutions, plants can react to environmental changes by different mechanisms involving various time scales. These mechanisms include acclimation, plasticity, and evolution.
- Plant reactions, in turn, can feed back to influence the environment at different scales by exchanges of matter and energy. For example, plants humidify the air, change turbulence and wind field, and hence influence cloud formation; they absorb carbon dioxide, produce oxygen and reactive volatile organic compounds, and modify, protect, and stabilize soils.
- There are a large variety of techniques available to researchers for the observation of vegetation–environment interactions at different time scales. No single technique can answer all questions; they have to be used synergistically, and often times these “suites” of observations have to be deployed across broad geographic areas and in multiple types of biomes.
- Due to the complexity of interactions and feedbacks between vegetation and the environment, numerical modeling has become a pivotal tool in conjunction with model–data fusion techniques. This new emphasis on fusing observations and theory has provided scientists with unprecedented insight into the mechanisms governing plant–atmosphere interactions, permitted the scaling of mechanisms across broad spans of space and time, and provided an integrated picture of global ecological processes.

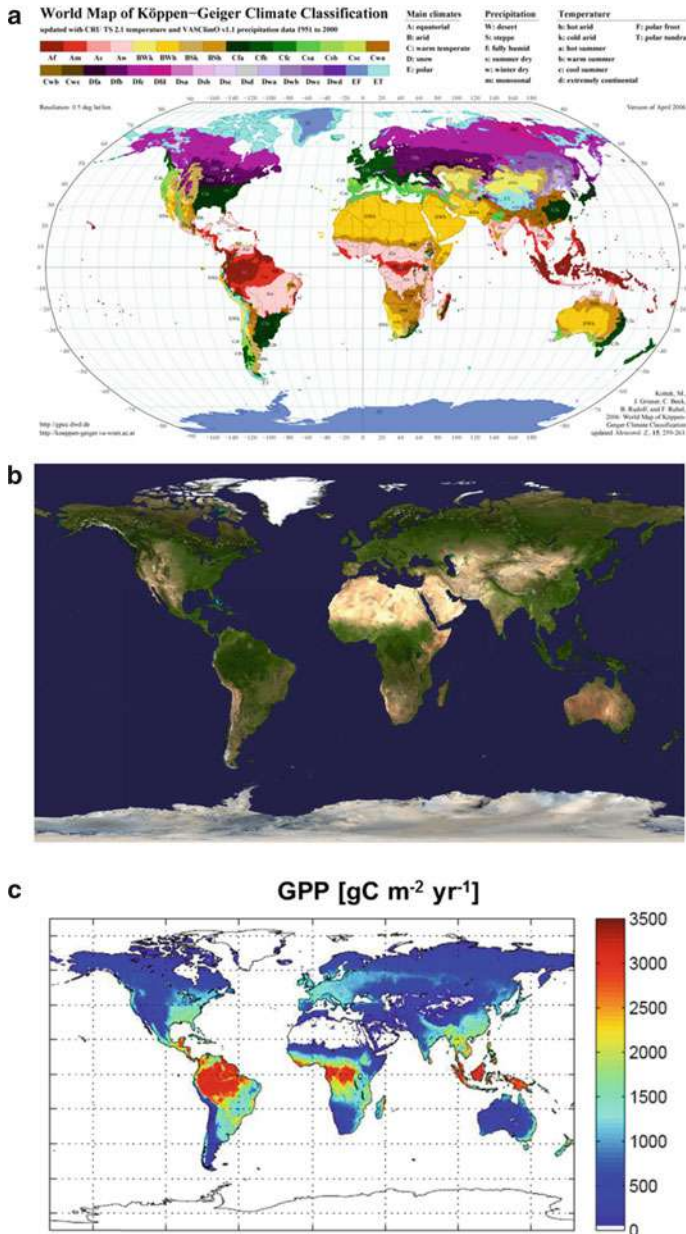
*Leaves, when present, exert a paramount influence on the interchanges of moisture and heat. They absorb the sunshine and screen the soil beneath. Being freely exposed to the air they very rapidly communicate the absorbed energy to the air, either by raising its temperature or by evaporating water into it.*

*Lewis Fry Richardson (1922), Weather Prediction by Numerical Process*

---

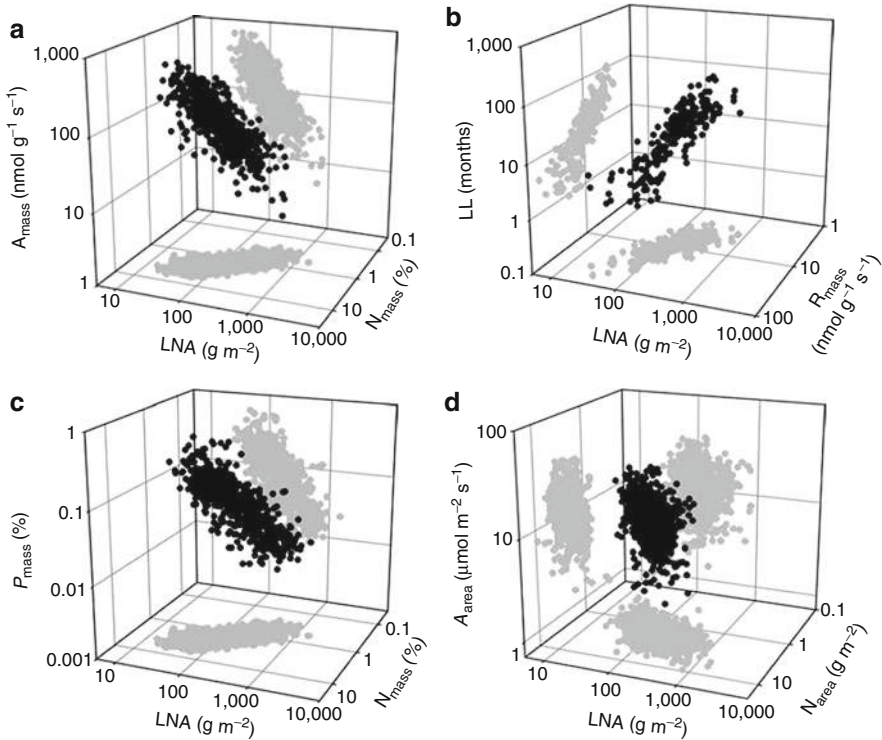
## Environmental Controls on Vegetation: Introduction

The effect of the environment, and in particular climate, on vegetation has been recognized since Aristotle and Theophrastus in ancient Greece (Greene 1909). Comprehensive and systematic descriptions of how the distributions of plants relate to environmental factors were pioneered by Alexander von Humboldt in the early nineteenth century and largely based on physiognomical (structural) observations. Raunkiaer classified plant life forms according to the position of their buds during the unfavorable season of the year (too cold, too dry) and identified diverse strategies to respond to recurrent adverse conditions (Raunkiaer 1934). The refinement of these life forms (e.g., based on leaf habit and longevity) and consideration of them in the context of vegetation formations and landscapes in relation to climate led to global climate and biome life-zone classifications (e.g., Köppen 1923; Holdridge 1947). These classification systems are still widely used today and updated with current climatological measurements (e.g., Kottek et al. 2006; Fig. 1a). Today, satellite remote sensing observation systems allow for an objective, repeated, spatially complete, and contiguous study of vegetation structure because the interactions with electromagnetic waves (in particular those interactions that lead to surface reflectance) depend on vegetation density and arrangement. A global composite of average vegetation greenness is strikingly similar to the Köppen climate map and underlines the continued value of eco-climatological classifications, even though they are only based on physiognomy and not on processes or functions, which are ultimately feeding back to influence the environment. Nevertheless, structure and function are related at an organismic level as has been noted for many decades by plant physiologists and are also emerging as a central organizing principle at the global level; this is seen, for example, when Fig. 1c, an estimate of photosynthesis, is compared to Fig. 1b, an estimate of vegetation density and cover. A similar argument about the correlation between structure and function of vegetation has been made at the leaf level with the so-called leaf economics spectrum, where traits such as leaf mass per area, nitrogen content, and maximum photosynthesis covary across global biomes (Fig. 2). The general principles of structure and function in plant ecology are described in textbooks by Barbour et al. (1999) and Schulze et al. (2005). How environmental factors act on plants and how plant processes feedback to the environment at different levels of integration are described more in detail here.



**Fig. 1** Different global views on similar spatial patterns of climate and vegetation. **(a)** Climate classification by Köppen (1923), update by Kottek et al. (2006). **(b)** Remote sensing view from the NASA MODIS sensor (From <http://svs.gsfc.nasa.gov/vis/a000000/a003100/a003191/frames/2048x1024/background-blumarmble.png>). **(c)** Annual carbon dioxide uptake by photosynthesis of vegetation (GPP) inferred from a statistical model, derived from ground observations and remote sensing





**Fig. 2** Three-way trait relationships among the six leaf traits with reference to LMA, one of the key traits in the leaf economics spectrum. The direction of the data cloud in three-dimensional space can be ascertained from the shadows projected on the floor and walls of the three-dimensional space. *LMA* leaf mass per area, *P* phosphorus, *N* nitrogen, *A<sub>max</sub>* light-saturated photosynthesis, *R* respiration, *LL* leaf life span (From Wright et al. 2004)

## Environmental Controls: Climate

The state of the atmosphere affects the rate at which plants and other living organisms produce and consume trace gases such as carbon dioxide ( $\text{CO}_2$ ), methane, and water vapor. The main fundamental processes of the biosphere (evaporation, photosynthesis, transpiration, respiration, and decomposition) are controlled by five climatic factors: radiation, temperature, precipitation, relative humidity, and wind speed.

Solar radiation (SR) is the primary source of energy for autotrophic organisms. Light energy directly drives many fundamental plant and biophysical processes, such as photosynthesis and evapotranspiration, by influencing stomatal conductance, transpiration, and leaf temperature. A portion of the incoming SR, the photosynthetically active radiation (PAR) spectral region between 400 and 700 nm, is absorbed by pigments and photosynthetic organs of vegetation and serves as one of the major biophysical variables directly related to photosynthesis and  $\text{CO}_2$  assimilation by vegetation. The amount of absorbed PAR primarily

depends on the leaf area index (LAI) of the ecosystem (defined as the amount of one-sided green leaf area per unit ground surface area,  $\text{m}^2/\text{m}^2$ ) and on the architecture of the canopy, and it is converted into chemical energy in sugars and secondary metabolites. Photosynthetic processes are affected not only by the amount of PAR but also by its quality. Recent studies showed higher ecosystem  $\text{CO}_2$  assimilation efficiency under “skylight” conditions that foster a high fraction of diffuse radiation (Mercado et al. 2009). A more uniform distribution of irradiance causes an increase in the proportion of light penetration through the canopy and irradiance per unit of LAI, once again illustrating the interaction between a driving environmental variable, vegetation (or, in this case, canopy) structure, and a physiological variable, such as  $\text{CO}_2$  assimilation rate. Moreover, at the canopy level the redistribution of the solar radiation load from photosynthetically light-saturated leaves to non-saturated (or shaded) leaves results in a greater increase in leaf photosynthesis rate. This is due to the fact that shaded leaves conduct most of their photosynthetic  $\text{CO}_2$  assimilation in the interactive domain located in the linear part of the light curve response (approximating a first-order relationship with absorbed radiant energy), while the saturated, sunlit leaves operate in the interactive domain located in the plateau of the light response curve (approximating a zero-order relationship with absorbed radiant energy). SR directly/indirectly influences many secondary plant processes such as seedling regeneration, leaf morphology, and the vertical structure of stands. The seasonal variation of photoperiod is also an important factor controlling both leaf flush and leaf senescence and therefore, together with temperature and water availability, controls plant phenology and the growing season length.

From the molecular to ecosystem scales, temperature influences biological processes by controlling the kinetics of enzyme-catalyzed chemical reactions and thus controlling the rates of plant growth, the patterns of seasonal phenology in ecosystems, the distribution of species and diversity of communities, and the decomposition and mineralization of soil organic matter. Generally, the control by temperature causes process kinetics to exhibit an optimum at intermediate temperatures. The response of processes to temperature variations can be flexible, leading to time-dependent acclimation responses that allow for maintaining the performance of processes across a range of temperature conditions (Atkin et al. 2005).

Aside from direct impacts on ecosystems, increasing temperatures can trigger indirect effects on plants in the ecosystem; many of which interact with one another to produce subtle synergies. On the one hand, warmer temperatures may enhance decomposition, releasing nutrients through mineralization; on the other hand, enhanced evaporation may decrease soil water content, reducing decomposition rates and its consequent release of nutrients and decreasing the mobility of nutrients from the soil into plants. As another example, on the one hand, warmer springs, as a result of climate change, can induce plants in temperate-latitude biomes to initiate their seasonal growth earlier and thus increase their potential to assimilate  $\text{CO}_2$  from the atmosphere; but warmer autumns can also potentially interfere with cold-temperature hardening, placing plants at increased risk of physiological damage during a critical phase of seasonality when frosts are interspersed with favorable weather.

Precipitation is another of the crucial environmental drivers of ecosystem functioning at different spatial and temporal scales. At short time scales, precipitation and soil water content control stomatal conductance, and because stomatal conductance is coupled with photosynthesis, soil water thus influences the rate of  $\text{CO}_2$  assimilation by vegetation. At longer time scales, the depletion of soil water content due to scarce precipitation may lead to prolonged water stress with a consequent modification of vegetation structure, such as leaf area index, rooting depth, and chlorophyll content. Since higher plants do not directly rely on precipitation but rather on water stored in the soil, the timing of precipitation in relation to the evaporative demand of the atmosphere, and thus mean air temperature, is of high importance.

Relative humidity (rH) is defined as the ratio of actual water vapor content to the saturated water vapor content at a given temperature and pressure. rH determines the vapor pressure deficit (VPD) between the soil and atmosphere and between the plant and atmosphere, and thus, climate and the spatial distribution of humidity in the atmosphere control potential evaporation rates and surface energy budgets at the global scale. The VPD directly influences plant water relations and indirectly affects hydraulic connectivity between leaves and the soil, leaf growth, photosynthesis, and evapotranspiration processes through stomatal control and leaf water potential.

Wind speed is another key factor controlling vegetation processes. Different regimes of wind speed and direction may influence physiological and mechanical aspects of vegetation. The main physiological effects are related to an enhancement of evapotranspiration. Wind removes the more humid air around the leaf by replacing it with drier air and, thus, increases the rate of transpiration. Finally, wind speed influences photosynthesis rates. Turbulence increases with wind speed in the atmosphere, which mixes  $\text{CO}_2$  from higher levels in the atmosphere downward toward the canopy, and thus increases the availability of  $\text{CO}_2$  for photosynthesis. Turbulence also mixes heat energy between the canopy surface and areas higher in the atmosphere, affecting the potential for vegetated surfaces to exchange sensible heat (through convection) with the atmosphere and thus contribute to the surface energy balance. Wind may also have mechanical impacts on vegetation by damaging shoots, controlling the allocation of carbon to stem thickening, and controlling the timing and patterns of leaf, flower, and fruit shedding. Crops and trees with shallow roots may be uprooted, leading to other secondary effects such as soil erosion, nutrient deposition, and recruitment opportunities for seedlings requiring a gap in the vegetated canopy. At the landscape scale, high wind speeds, associated with conditions of low rH and moisture of vegetation, may also contribute to vegetation drying and thus enhancement of the ignition potential of wildfires and, once ignited, the spread and intensity of fires.

## **Environmental Controls: $\text{CO}_2$ , $\text{O}_3$ , Pollutants, and Nitrogen Deposition**

$\text{CO}_2$  is one of the essential drivers of photosynthesis. Leaf photosynthesis increases nonlinearly with the leaf internal  $\text{CO}_2$  concentration, reaching a saturation plateau.

Since the  $\text{CO}_2$  concentration in the intercellular air spaces of the leaf is about 70 % of atmospheric  $\text{CO}_2$ , leaf photosynthesis is expected to respond positively to the atmospheric increase of  $\text{CO}_2$  observed since the preindustrial era, which is related to the increase of anthropogenic emissions from fossil fuel combustion and land-use change. Empirical evidence from  $\text{CO}_2$  fumigation experiments (FACE, Free-Air  $\text{CO}_2$  Enrichment studies) has shown that the expected increase of  $\text{CO}_2$  concentration in the atmosphere of the future enhances plant growth, the so-called “ $\text{CO}_2$  fertilization” effect (Norby and Zak 2011). These studies have also revealed a response of leaf photosynthesis to elevated  $\text{CO}_2$  that is dependent on the conditions at which the plant was grown. In essence, plants grown at elevated  $\text{CO}_2$  accumulate sugars at a greater rate than those grown at lower atmospheric  $\text{CO}_2$  concentrations. The accumulated sugars trigger changes in the expression of the genes for Rubisco, the primary  $\text{CO}_2$ -fixing enzyme of photosynthesis, such that fewer enzyme molecules are produced. Rubisco is the most abundant protein on Earth, and its production by plants utilizes approximately 30 % of the nitrogen resource available to plants. At elevated  $\text{CO}_2$ , a reduction in the allocation of nitrogen to the production of Rubisco per unit of leaf area means that more nitrogen can be allocated to the production of new leaf area. Thus, the high- $\text{CO}_2$  feedback enhances the nitrogen-use efficiency of plants and enhances the potential growth rate of plants in an elevated  $\text{CO}_2$  (future) atmosphere.

Besides the increase of  $\text{CO}_2$ , anthropogenic activities cause an increase in atmospheric nitrogen (N) deposition, particularly of nitrogen oxide compounds ( $\text{NO}_x$ ), and N input to the biosphere caused by the use of fertilizers. The combustion of fossil fuels and the burning of biomass associated with forest clearing and agricultural development tend to create a high-temperature process, called the Zeldovich reaction, which “scrambles” the N released from plant tissues with the  $\text{O}_2$  consumed from the atmosphere and creates  $\text{NO}_x$  compounds that are deposited back to ecosystems. Once deposited to the soil, microorganisms can convert the deposited  $\text{NO}_x$  to nitrate and ammonium ions, capable of plant uptake. Due to their tendency to be leached from soils, nitrate and ammonium are scarce in natural, unperturbed ecosystems and play a critical role in the biosphere by determining the potential rates of primary productivity. N availability especially limits gross primary productivity (GPP) and terrestrial carbon (C) sequestration in the boreal and temperate zone. Human activities associated with the burning of fossil fuels and the production of agricultural fertilizers have doubled the input of N since 1860. These anthropogenic changes have had consequences for the turnover of N and storage of C. In particular, an enhancement of forest growth associated with N fertilization and a reduction of soil respiration (Janssens et al. 2010) have been observed. The terrestrial C and N cycles are tightly related. At low N availability, a doubled  $\text{CO}_2$  concentration shows a small effect on biomass and photosynthetic rates, with a negative feedback due to the sequestration of N into the increment of biomass: the  $\text{CO}_2$  fertilization increases the terrestrial C storage, as well as the terrestrial N stock, with a consequent reduction of N availability in the soil (Zaehle 2013).

## Ozone and Air Pollutants

Ozone ( $O_3$ ) is produced by photochemical reactions between  $NO_x$ , which is produced by natural soil processes as well as anthropogenic fossil fuel/biomass burning, and volatile organic carbon compounds (VOCs), which are principally emitted from forests but can be produced from anthropogenic sources as well.  $O_3$  is phytotoxic and causes deleterious effects on plants that span from the cellular to community scales (Ainsworth et al. 2012). Effects at the community-scale include reduced primary productivity and shifts in species composition. At the scale of individual organisms, ozone uptake causes reduced rates of biomass and leaf area production, reduced reproductive output, and shifts in the phenological sequences associated with seasonality, such as the timing of leaf senescence. At the leaf scale, ozone uptake causes reductions of photosynthesis, discoloration and production of necrotic lesions on the leaf surface, increased respiration rates due to energetic demands of tissue repair, and cuticular wax accumulation. Finally, at the cellular level, ozone uptake causes reduced Rubisco activity and content, increased rates of flavonoid biosynthesis, and increased rates of protein turnover. Elevated  $CO_2$  causes partial stomatal closure, so the combined effect of high  $CO_2$  and ozone is less than the negative effect of ozone alone. These processes emphasize, first, that ozone exposure, determined on the basis of atmospheric ozone concentrations (traditionally used to calculate the damage), needs to be substituted by the cumulative uptake (or dosage) of ozone to a plant and, second, that for a full evaluation of the impact of  $O_3$  on plant function within the context of global change (e.g., including increasing N deposition and atmospheric  $CO_2$  concentration), the feedbacks and interactions among all three components need to be addressed in observation networks and Earth system modeling.

## Soil Properties

Soils have a fundamental influence on vegetation by providing the most important reservoir of nutrients and water needed for the biological activities of plants, as well as serving as a medium for structural anchorage. Soils are more than the inorganic products of crushed and weathered rocks; rather soils are living systems, a dynamic component of the Earth system because of the organisms they hold (Bahn et al. 2010). Carbon is exuded by roots and root-associated fungi, and these exudates supply carbon to heterotrophic bacteria and other microorganisms that in turn mineralize soil organic matter, freeing nutrients to be reabsorbed by plants. In fact, plants must be considered as part of the soil (through their roots). There are physical, chemical, and biological soil factors that exert profound influences on vegetation. The main physical characteristics are soil texture, structure, and depth. Soil texture is determined by the content of silt, clay, and sand, as well as larger solid matter such as gravel and rocks. Soil texture determines the water holding capacity of soils, hydraulic conductivity through soils to roots, and the cation exchange capacity of a soil. Soil depth is determined by the position of the bedrock

or of the water table and by site characteristics such as slope and topography. Soil depth and its association with soil organic matter content also determine the portion of soil usable by plant roots and, therefore, the total water and nutrient holding capacities.

Chemical characteristics of soils include fertility and acidity (pH), which influence the capacity of soil to sustain growth and maintenance of metabolism in plants. Soil pH affects the availability of macro- and micronutrients by controlling the chemical forms of the nutrients. The optimum soil pH range for most plants is between 5.5 and 7.0, although many plants have adapted to pH values outside this range. The concentration of available N is less sensitive to pH than the concentration of available phosphorus (P). In order for P to be available for plants, soil pH needs to be in the range 6.0–7.5. If pH is lower than 6, P starts forming insoluble compounds with iron and aluminum, and if pH is higher than 7.5, P starts forming insoluble compounds with calcium.

## **Animals Including Humans**

Direct animal–plant interactions include mutualistic relationships such as pollination and antagonistic relationships such as herbivory. In addition, there are several indirect effects of animals (especially soil invertebrates and protozoans) on plants because they change the environment, particularly the soil, through reworking it (e.g., earthworms) and by feeding on dead plant material and other animals, which enhances nutrient cycling. Interactions occur between climate and animal–plant relations. For example, widespread forest insect outbreaks have been shown to be muted or amplified by climate, which controls life cycle frequencies and the potential for winter mortality in the insects, as well as stress intensity in trees, both of which in turn affect the rates of insect damage. During warmer and drier climate extremes, insect damage to forests is generally increased, causing increased rates of leaf and root litter deposition to the soil and increased rates of tree mortality. Animal–plant interactions are described in more detail in Malmstrom (2010).

Humans influence virtually all environmental factors discussed above and, hence, directly and indirectly influence vegetation in important ways. The direct effects of humans include the CO<sub>2</sub> and N deposition that occurs to ecosystems as a result of fossil fuel and biomass burning. Examples of indirect influences include the climate change associated with increasing atmospheric CO<sub>2</sub> levels and increases in the oxidative capacity of the atmosphere due to photochemically reactive air pollution. Humans have imposed rates of land-use change and impacts to natural communities and populations of plants that are unprecedented in relation to natural animal impacts on the landscape. In fact, the magnitude of human impact has been so great that many scientists now refer to the current time as the Anthropocene. Virtually all natural animal–plant interactions have been affected by human activities. This interaction between humans and the Earth system, while relatively well characterized within the realm of climate change, has been virtually unstudied

within the realm of how nonhuman animals influence ecosystems, communities, and populations. An emphasis on the level of interaction is required before we can properly understand how climate change, biological extinction, and human enterprise are mutually connected. Concrete examples of impacts on vegetation include those by animals through grazing and by humans through land fertilization, forest and land management, and disturbances such as deforestation, fires, and, more generally, land-use change. Deforestation and fires are the main disturbances at global scale. Vast areas covered by forests have been converted to agriculture. On one hand, humans influence fire patterns by intentionally or accidentally igniting fires; on the other hand, humans actively suppress both anthropogenic and natural fires (Bowman et al. 2009).

## Plant Responses to the Environment

Unlike animals, which are often mobile and can relocate in response to environmental change, plants are at the mercy of the environment, at least at the time scale of the current generation. However, most plants have the capacity to respond to environmental change in the short term (within a generation) through ecophysiological responses and via phenotypic plasticity (the expression of different phenotypic traits depending on growth environment), and all plants have the capacity to environmental change in the long term (multiple generations) through evolution. Phenotypic plasticity involves changes that can be reversible over the life span of an organism. Consistent with the theme of processes occurring across multiple scales, there is concern that the current rate of climate change is faster than that experienced by species in the past history of the Earth system. While phenotypic plasticity can accommodate some level of change in the short term, it is unlikely that species can evolve fast enough to sustain their populations in the face of continued change. Acceleration in the rate of species extinctions is likely to occur. This is particularly relevant to tropical species, which have evolved within relatively narrow limits of climate variability. Tropical species are likely to be in greater danger of extinction in the face of future climate change, compared to temperate species, which often have greater capacities for phenotypic plasticity and greater genetic variance within populations.

As an example of the differences between adaptation and phenotypic plasticity, we can consider the case of plant responses to drought. The adaptation of plants to drought has involved many different types of evolutionary change, including the leaf sclerophylly (thickened, hardened foliage) and succulence; the former tends to *resist* drought by producing leaves that are protected against herbivory and mechanical damage from the wind so that the cost of replacing foliage in a resource-limited environment is reduced, whereas the latter tends to *avoid* drought by producing internal supplies of stored water that can be drawn down slowly. Metabolic pathways such as C4 and CAM photosynthesis are examples of the entire metabolic pathways that have evolved to facilitate high rates of carbon assimilation with limited loss of water through transpiration to a dry atmosphere. Phenotypic plasticity in response to drought includes the seasonal

drought deciduous loss of leaves, which is reversible once moisture becomes available once again, and the accumulation of physiological regulator compounds, such as abscisic acid (ABA), which can accumulate in leaves during drought and cause stomata to not open as much during daylight periods; when moisture becomes available again, the ABA can be metabolized and stomatal opening can once again be increased. Once again, these responses occur across vastly different time scales. Adaptation occurs across generations, whereas phenotypic plasticity occurs within individuals of a single generation.

---

## Influences of Vegetation on Environment

Just as the environment influences the growth, form, and reproductive success of individual plants and the structure and composition of plant communities, there are, in turn, profound influences of vegetation on the environment (Pielke et al. 1998). The effects of vegetation on the environment occur at a range of scales (McPherson 2007), from microclimate to local weather to global climate. For example, a large tree not only influences microclimate by providing shade on a warm day, but it is also responsible for the transport of water from the soil to the atmosphere, thereby affecting regional cloud and rainfall patterns. Through photosynthesis, the same tree also removes CO<sub>2</sub> – an important greenhouse gas – from the atmosphere, thus affecting long-term global climate trends.

While these effects have long been recognized, as illustrated at the beginning of this chapter by the quotation from Richardson (1922), our understanding of the associated processes and how they vary among ecosystem types has advanced greatly in recent decades. This section will provide an overview of the various ways in which vegetation can influence the environment at different spatial and temporal scales. The focus is mostly on how vegetation can affect the atmosphere and climate system, but microclimatic effects both above- and belowground are also considered. The nature and magnitude of these effects vary among the world's biomes, according to the amount and type of vegetation present, soil and climate conditions, and seasonality (Richardson et al. 2013).

### Microclimate

Plants influence microclimate in numerous ways. Forest trees provide perhaps the best example, because their vertical trunks and elevated foliage create unique three-dimensional gradients of environmental conditions and resource availability from the top of the canopy to the forest floor. The evolution of tall, woody plants was therefore a critical event for life on our planet because it resulted in remarkable habitat diversity through vertical stratification. It created a novel niche within which organisms could evolve and adapt – the vertical niche. Today, this diversity is best exemplified by the tropical rainforests, in which highly specialized communities of plants and animals are adapted to different canopy strata, each of which has its own microclimate.



In forests, the dominant environmental gradient is related to light availability. While leaves at the top of the canopy are regularly exposed to full sun, understory plants commonly grow in less than 5 % (and sometimes even less than 1 %!) of full sunlight. Over the course of the day, the understory light environment is heterogeneous, as periodic sunflecks, or brief periods when the direct solar beam penetrates to the forest floor, may account for a disproportionate share of the total flux of solar radiation. Furthermore, there is also a vertical gradient in the quality, or spectral distribution, of light. This occurs because individual leaves typically absorb roughly 90 % of solar radiation in PAR wavelengths but absorb less than 50 % of solar radiation in near-infrared wavelengths (700–1,000 nm). Thus, relatively less visible radiation, and relatively more near-infrared radiation, penetrates through the canopy to lower layers. Note that in seasonally deciduous forests, these gradients also vary over the course of the year, according to variation in LAI and leaf angle distribution.

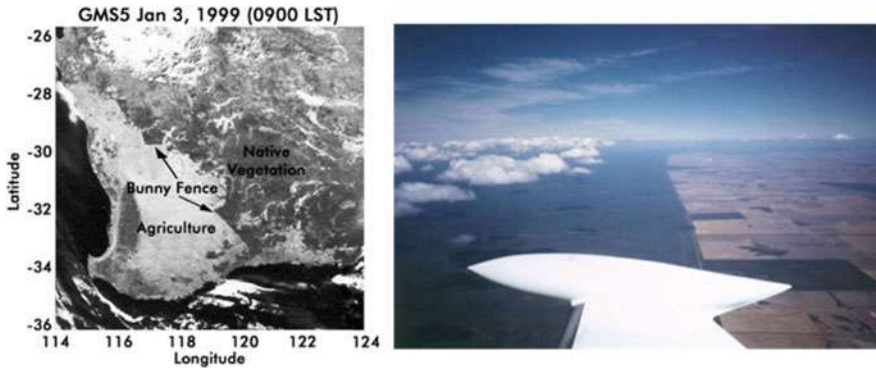
In the shaded understory, environmental conditions are more mesic than at the top of the canopy. Temperature extremes are reduced, resulting in a narrower diurnal temperature range. At the same time, relative humidity is generally increased, and the evaporative demand of the local atmosphere is reduced, in the understory compared to the top of the canopy. Leaves and trunks also exert drag, thereby reducing wind speeds within and below the canopy relative to above the canopy.

Vegetation also affects the soil microclimate. By providing shade, overstory vegetation reduces soil temperature extremes. A substantial amount of precipitation is intercepted by canopy foliage, thereby reducing throughfall, the process by which rainwater drips through the canopy. Although some of the intercepted precipitation is redirected to flow down branches and stems to the forest floor, the net effect is to increase spatial heterogeneity in soil moisture. Leaf litter on the forest floor acts as mulch, reducing evaporation from the soil surface, and in some cases it acts as a hydrophobic barrier, intercepting throughfall water and holding it at the surface until it evaporates, thus reducing penetration into the soil. These effects on soil microclimate are ecologically important because they will influence decomposition processes and nutrient cycling.

There are countless other examples of the ways in which vegetation influences microclimate. For example, in mountain areas, trees and shrubs affect surface roughness and hence drifting and spatial patterns of snow accumulation. In boreal ecosystems, moss and other surface vegetations insulate the underlying permafrost and maintain cold root-zone temperatures. Arctic and alpine cushion plants create their own microclimate, using a prostrate growth form and densely packed leaves to increase the thickness of the boundary layer near the ground. Compared to the surrounding air, these plants grow in a warmer, more humid, and less windy environment that is more favorable to photosynthesis and growth.

## Transpiration

Gas exchange for photosynthesis occurs through stomata on the leaf surface. Stomata open during the day, allowing CO<sub>2</sub> to diffuse into the leaf. At the same



**Fig. 3** The potential effect of vegetation on local climate. See text for more information (From Lyons 2002)

time, however, diffusion of water vapor from inside the leaf to the free atmosphere is driven by the non-saturated moisture state of the atmosphere and its evaporative demand for water. Through this process, called transpiration, plants are responsible for the movement, each day, of massive quantities of water from the soil column to the atmosphere. Transpiration has a significant cooling effect on surface climate, removing heat through the process of latent heat exchange. Additionally, at regional scales, transpiration by plants results in the increased abundance of clouds, which both moderate surface temperature and enhance precipitation. A nice example of such effects is the so-called “bunny fence” experiment in Australia (Fig. 3). The fence has led to a sharp boundary of vegetation types across a relatively homogeneous terrain, with a visible influence on cloud cover.

## Surface Energy Budget

Vegetation influences climate through biogeophysical effects related to the surface energy budget and the partitioning of net radiation to latent and sensible heat fluxes (Bonan 2008a). Albedo, the proportion of incident solar radiation that is reflected by the land surface, determines net shortwave radiation; net longwave radiation is driven by surface and sky temperatures. Darker surfaces (low albedo) absorb more shortwave radiation than bright surfaces (high albedo) and hence have a warming effect on local climate. During the growing season, there are large differences in albedo among different vegetation types, with grasslands and crops having higher albedo than broadleaf forests, which in turn have higher albedo than conifer forests. However, during the winter months, the difference in albedo between deciduous and conifer forests at high latitudes is even greater, because of the high albedo of snow on the ground that is visible through the leafless deciduous canopy.

The climate effects of differences in albedo may be offset by the cooling effects of evaporation, i.e., latent heat flux. For example, the conversion of tropical forest in the Amazon to agriculture has a net warming effect on surface climate, with a

modest increase in albedo (cooling effect) more than offset by a large decrease in transpiration (warming effect). For a given amount of net radiation, lower latent heat flux must be offset by higher sensible heat flux. Thus, boreal conifer forests, which have lower rates of evapotranspiration than boreal deciduous forests, have higher rates of sensible heat flux, which returns energy to the atmosphere and promotes the development of a deeper atmospheric boundary layer. When available soil moisture is reduced during drought, driving reductions in evapotranspiration, there is similarly a corresponding increase in sensible heat flux, ultimately affecting mesoscale circulation and atmospheric transport.

## **Biogeochemical Cycling, Including Carbon**

On geologic time scales, photosynthesis has had a profound influence on the Earth's atmosphere (Beerling 2007). One important event was the evolution, approximately 3 billion years ago, of the cyanobacteria. Although not considered plants, the cyanobacteria were the first organisms to conduct photosynthesis in a manner similar to the way that plants do. Through endosymbiosis, cyanobacteria evolved possession of both Photosystems 1 and 2, which allowed them to make use of water as an electron donor for photosynthesis and produce oxygen as a by-product. Photosynthesis by the cyanobacteria thus resulted in the oxygenation of the atmosphere, which ultimately enabled the evolution of large, multicellular life forms. A second important event was the evolution of woody plants during the Paleozoic era. Between 400 and 300 million years ago, CO<sub>2</sub> concentrations in the atmosphere dropped from roughly 4,000 ppm to less than 500 ppm. This occurred because early vascular plants, growing in steamy swamps, used the carbohydrates resulting from photosynthesis to build lignified tissues that could not be broken down by existing decomposer organisms. As a result, massive amounts of C, in the form of dead plant biomass, came to be sequestered in deep peat deposits rather than respired back to the atmosphere. Over time, this peat was converted to the coal that powered the Industrial Revolution.

Today, terrestrial vegetation continues to play a critical role in the biogeochemical cycling of carbon. Carbon is the building block of life, and on a dry-matter basis, plants are about 45 % carbon. There is almost as much (600 Pg) carbon stored in living plant matter as there is in the atmosphere (750 Pg), while the reservoir of dead plant matter in the soils of terrestrial ecosystems is even larger (1,600 Pg). At the same time, CO<sub>2</sub> in the atmosphere is the substrate for photosynthesis and thus a prerequisite for the process by which plants convert solar energy into stored chemical energy. However, CO<sub>2</sub> is also a potent greenhouse gas and one of the main factors driving climate change. The levels of atmospheric CO<sub>2</sub> have been rising since the start of the Industrial Revolution, from 280 ppm in the early 1800s to over 400 ppm by 2013. This rise has been driven by the combustion of coal and other fossil fuels formed, over millions of years, from dead organic matter. Over the next 100 years, the future climate of our planet largely depends on the trajectory of atmospheric CO<sub>2</sub>. In the last century, global temperatures have risen by approximately 1 °C, but future increases of less than 1° or more than 3 °C are forecasted by 2100,

depending on what level of atmospheric CO<sub>2</sub> concentrations is reached by the end of the century. In this context, an important point is that the stores of carbon in biomass and soils are large relative to the atmospheric reservoir. This suggests that disturbance, extreme climate events, or other exogenous factors affecting vegetation C reserves could have a direct impact on atmospheric CO<sub>2</sub> concentrations and thus either enhance or reduce future climate change.

Each year, about one-quarter of CO<sub>2</sub> in the atmosphere is turned over through photosynthetic processes. 60 % of this photosynthesis occurs on land, while the remainder is in the oceans. Current estimates put total global gross primary productivity of terrestrial vegetation at 122 Pg C year<sup>-1</sup>. The flux of carbon from terrestrial ecosystems back to the atmosphere, driven by decomposition and cellular respiration processes, is almost as large. Although the net balance between these fluxes is small (and varies in magnitude from year to year, depending on variations in weather and disturbance factors), individually these fluxes dwarf the rates of anthropogenic emissions of CO<sub>2</sub>, which total roughly 8 Pg C year<sup>-1</sup>. Thus, the ability of terrestrial ecosystems to remove CO<sub>2</sub> from the atmosphere and sequester it in long-lived C pools is an important consideration in the context of mitigation of future climate change.

The rates of gross primary productivity vary widely among biomes. With annual gross primary productivity of 41 Pg C year<sup>-1</sup>, tropical forests account for about one-third of the world's terrestrial primary productivity. On a unit-area basis, gross primary productivity of tropical forests (2,300 g C m<sup>-2</sup> year<sup>-1</sup>, on average) is also higher than in any other ecosystem. For example, it is 8–10 times higher than the rates of gross primary productivity in tundra and desert ecosystems. By comparison, the gross primary productivity of temperate (950 g C m<sup>-2</sup> year<sup>-1</sup>) and boreal (600 g C m<sup>-2</sup> year<sup>-1</sup>) ecosystems is intermediate between these two extremes.

A key factor driving spatial patterns in annual gross primary productivity is growing season length, which varies from a month (or less) in high-latitude and high-altitude tundra ecosystems to a full 12 months in tropical and subtropical ecosystems where neither water nor temperature is seasonally limiting. Interannual variability in weather (principally temperature and precipitation) drives year-to-year variation in seasonality (i.e., phenology or the annual rhythms of vegetation development and senescence), which can directly increase or decrease annual carbon uptake.

Even at short (<1 year) time scales, vegetation has a measurable impact on atmospheric CO<sub>2</sub>. This is demonstrated by the strong annual cycle in atmospheric CO<sub>2</sub> as measured, for example, at monitoring stations such as Mauna Loa, Hawaii. There, atmospheric CO<sub>2</sub> concentrations drop during the summer, when vegetation in the northern hemisphere is photosynthetically active and photosynthesis greatly exceeds respiration, and rise during the winter, when the reverse is true. The seasonal amplitude of atmospheric CO<sub>2</sub>, about 10 ppm, is about five times greater than the annual increase in atmospheric CO<sub>2</sub>, thereby illustrating the importance of vegetation in the annual global carbon budget.

Plants also play key roles in the cycling of other elements besides carbon. For example, symbiotic relationships between some plant species (including alder, as

well as soybeans and other legumes) and bacteria such as *Rhizobium* and *Frankia* are important because the bacteria can fix  $N_2$  gas (which cannot be otherwise used by plants) from the atmosphere to  $NH_3$ , or ammonium ( $NH_4^+$ ), which is available for plant uptake. Additionally, plant root exudates can enhance the chemical weathering of soil minerals to forms that are phyto-available.

## Emissions of Biogenic Volatile Organic Compounds (VOCs)

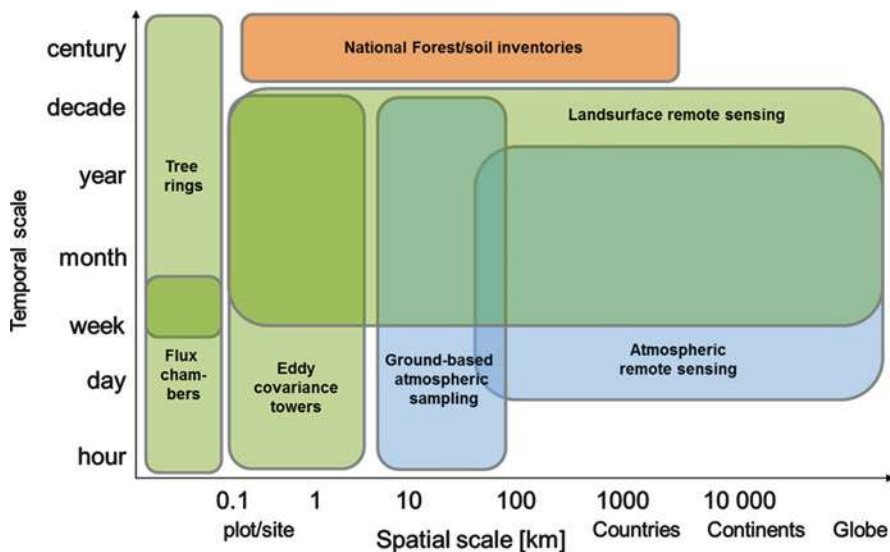
Biogenic VOCs, which are a class of reactive hydrocarbons that includes isoprene, monoterpenes, and sesquiterpenes, are emitted by most of the world's plants. VOCs are physiologically important to the plant because they play a role in protection against thermal and oxidative stress and ecologically important because they are involved in allelopathy, defense against pathogens and herbivores, and signaling to pollinators and seed dispersers. VOC emissions vary seasonally in relation to temperature and phenology. Some plant species, such as eucalypts and oaks, emit much larger quantities of VOCs than other species. For example, Australia's Blue Mountains are so named because of the large amounts of terpenoids emitted by eucalyptus trees. These VOCs are oxidized to secondary aerosols, which then scatter light at the violet end of the visible spectrum, causing a characteristic blue haze over the forest. Similarly, VOC emissions by oaks and conifers are the cause of the "smoke" after which the Smoky Mountains of the southeastern United States are named. This haze increases the flux of diffuse solar radiation, which enhances canopy-level photosynthesis.

VOCs are important to the climate system for a number of reasons (Peñuelas and Staudt 2010). Secondary aerosols formed from VOCs are a major source of cloud condensation nuclei and affect cloud abundance and thickness, as well as precipitation. Clouds, in turn, affect the radiation balance of the Earth in complex ways, trapping heat in the lower atmosphere but also enhancing the planetary albedo, resulting in an increase in the fraction of incident solar radiation that is reflected back into space. VOCs can also have an impact on atmospheric concentrations of other important greenhouse gases. By reducing the atmospheric oxidation potential, VOCs indirectly increase the expected lifetime of atmospheric methane and react with  $O_3$ . However, the total impact of these processes on global climate is difficult to quantify, in part because the estimates of total global VOC emissions have high uncertainties.

---

## Observation Strategies

As discussed above, plant–environment relations are manifold and operate at different scales. Accordingly, observation strategies, which span across several spatial and temporal scales, have been developed and are needed for a complete exploration of plant–environment interactions (Fig. 4).



**Fig. 4** Observational systems related to vegetation and ecosystem function across temporal and spatial scales with special emphasis on carbon balance and trace gas exchange with the atmosphere

### Classical Observations: Surveys, Biometry, and Tree Rings

Traditional forestry measurements have been used widely in forest inventories to assess the aboveground biomass of forest stands. Aboveground woody carbon biomass is usually inferred from measurements of volume increase and wood density. All existing biometric studies rely on measurements of stem diameter variations that can be derived from dendrometer data and repeated inventories of tree stand densities and sizes and tree rings. More sophisticated methods also account for tree height to avoid a priori relationships between diameter and stem volume. Tree rings allow for the investigation of not only current but also past variations of radial growth in trees with a pronounced seasonal cycle of cambial activity. Hence, reconstructions of tree growth in relation to the climate variability spanning several centuries are possible. If it is known which climate factor has been limiting growth, reconstructions of this climate factor from tree ring chronologies is possible. This research field of dendrochronology has a long tradition. One interesting emerging technique is the use of microcore sampling which allows for monitoring of the seasonal variability of tree stem growth.

Leaf area index, or LAI (cf. paragraph Environment: climate), is an important biometric measurement that is needed to characterize plant canopies (e.g., light and precipitation interception) for the validation of satellite products and for model parameterization. LAI is typically measured with destructive methods (leaf harvesting) or with indirect methods such as hemispherical photograph or optical measurements. Recently, important biometric measurements (specific leaf area,

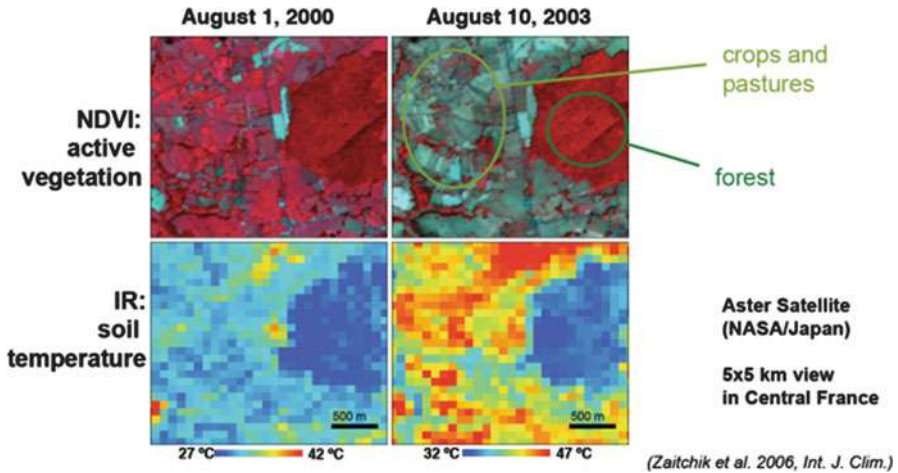
LAI, aboveground biomass) have been collected in a harmonized way in the context of a research initiative called “TRY,” which focuses on the collection of data and knowledge on plant traits at the global scale (Kattge et al. 2011).

## Flux Measurements

Flux measurements, i.e., the measurement of gas exchange between plants and the surrounding air, can be done at organ, whole-plant, or ecosystem level. They are classically performed with enclosures (cuvettes or chambers) that surround leaves, branches, or the whole plant and where air is blown through, and the concentration differences between the inlet and outlet are measured. These measurements are very precise, but the presence of the chambers can change the microenvironment around the object to be measured, e.g., by changing the radiation balance, thus altering the respective gas exchange. This problem is overcome by eddy covariance (EC), a micrometeorological technique that relies on the combination of high-frequency measurement (10–20 Hz), temperature, wind speed, and gas concentration (e.g., CO<sub>2</sub>, water vapor, methane, etc.) (Baldocchi 2008). In the last three decades, this technique has been widely used for monitoring carbon, water, and energy fluxes and, more recently, fluxes for methane and other greenhouse gases, in more than 500 research sites, scattered across a variety of biomes and climatic regions. The long-term measurements of CO<sub>2</sub> and greenhouse gas fluxes obtained using the eddy covariance technique make it a useful tool for elucidating the carbon balance of terrestrial ecosystems and the causes of its interannual variability and for improving the understanding of the interaction between carbon, water, energy fluxes, and climate. Measuring the abundance and fluxes of stable isotopes has become possible with high temporal resolution and yields complementary information on plant ecophysiology (Griffis 2013).

## Remote Sensing

Remote sensing (RS) observations can provide spatial and temporal variability of ecosystem properties driving carbon, water, and energy fluxes, as well as important information about vegetation and ecosystem structure (e.g., aboveground biomass, leaf area index). RS data provides spatial (global, regional, and local) and temporal (decadal, seasonal, and interannual) information about the important properties of the ecosystem. Moreover, by using multitemporal classification methods, RS can be used to gather information about land-use change and disturbance (in particular fires and deforestation). However, RS data can be hampered by the contamination of the signal by aerosols and clouds and by the fact that the parameters are estimated by using empirical relationships or radiative transfer models and not by direct measurement; whenever models must be inserted into a diagnostic or prognostic process, gaps in knowledge produce uncertainty in calculation. Nevertheless, interactions between climate and vegetation type can often be clearly inferred,



**Fig. 5** Remotely sensed images of vegetation cover (*top*) and land surface temperature (*bottom*) before (*left*) and during (*right*) the 2003 European heat wave. Denser vegetation cover with forest yields less surface heating (From Zaitchik et al. 2006)

as in Fig. 5, where strong gradients of land surface temperature are found depending on vegetation type and density as a consequence of an extreme heat wave.

The typology of measurements and parameters retrieved via RS depends on the characteristics of the sensors. With the development of hyperspectral imaging or reflectance sensors, it is possible to look at objects (target) using a vast portion of the electromagnetic spectrum. Targets such as leaves or tree canopies have unique “fingerprints” (spectral signatures) across the electromagnetic spectrum. By exploiting this information, it is possible to derive important properties such as chlorophyll/pigments, leaf nitrogen, extractable water content, etc. RS data can be collected at different spatial scales by using satellite products and airborne platforms with hyperspectral sensors, as well as in the proximity of the surface (proximal sensing). Proximal sensing is increasingly growing because it is one way to better understand the relationships between RS data and ecosystem processes at high temporal resolution, if associated with EC measurements. An emergent branch of RS is the direct inference of physiological processes, in particular photosynthesis. Among these, the measurement of sun-induced chlorophyll fluorescence (SIF) by passive (i.e., without artificial excitation sources) RS systems at field scale has been proven to be a valuable method for the assessment of plant photosynthesis (Meroni et al. 2009).

Another important technique, which yields three-dimensional structural information (e.g., leaf and branch distribution), is terrestrial LiDAR scanners (Levick and Rogers 2008). Terrestrial LiDAR measurements are generally collected using an instrument placed on a survey tripod above the ground in the experimental site. Their usage for estimating leaf area stems from a very high spatial resolution and a relatively small laser footprint size with respect to the typical dimensions of leaves



and other tree organs. New LiDAR missions will be used in the future to precisely describe the canopy structure, in particular the vertical distribution of elements in a canopy, the tree height, and the tree cover.

## Atmospheric Observation of Trace Gases

Any spatial divergence in flux into or out of the atmosphere will lead to a change in the concentration of the respective gas, e.g., CO<sub>2</sub>. Spatial divergence of gas flux (i.e., change in the magnitude of the flux as a function of space) and its influence on the time-dependent accumulation or depletion of gas concentration is determined by the principle of “continuity,” which in turn is required to adhere to the principle of mass conservation. Hence, the atmosphere works as a natural integrator of gas fluxes and concentrations over large scales (Fig. 4). However, due to atmospheric circulation, the coupling of spatial divergence in flux to time-dependent divergence in concentration is often “smeared,” and respective signals in concentration are transported away from the causal sink or source represented in the fluxes, both vertically and horizontally. Therefore, inferring fluxes from observations of atmospheric concentration is a challenge as spatial and temporal scales are increased. Connection between observed concentrations and inferred fluxes relies on the inverse modeling of atmospheric transport (Heimann and Kaminski 1999); the transport model must be used to go back in time and figure out from where on the landscape the flux divergence that gave rise to the concentration originated. For instance, with this approach, a net CO<sub>2</sub> uptake by northern hemisphere ecosystems has been inferred. In addition, oscillations of the climate system (El Niño Southern Oscillation) have been synchronized with respective oscillations of vegetation activity through this same atmospheric inverse modeling approach (Heimann and Reichstein 2008). As at the ecosystem level, the measurement of gas concentrations is fruitfully complemented by observations of stable isotopes, which help infer ecophysiological properties at larger scales. For instance, drought effects on photosynthesis have recently been detected at large by atmospheric <sup>13</sup>C observations.

---

## Modeling Strategies

From a theoretical point of view, models can be defined as representations of systems or processes underlying a wide variety of observed properties and functions. Ecosystem models are simplifications of the complex organizational structures and interactions observed in nature but ultimately synthesize our knowledge and theory. In this regard, very different modeling approaches have been developed, depending on the characteristics or dynamics to be represented or hypotheses to be explored. These range from simple empirical univariate approaches describing processes of decomposition or primary productivity to process-based approaches with a more mechanistic representation of physical chemical reactions in living organisms to describing dynamics of vegetation changes. But independent of the

approaches, testing the concepts and hypotheses embedded in models is an essential step in model construction, for which observations are crucial. Recent technological advances in observational strategies, from in situ to satellite remote sensing retrievals of biophysical properties of vegetation, representing unique sources of information for sophisticated computer models, which simulate the entire Earth system, have been used to conduct global-scale experiments and probe the effects of changes in surface vegetation on the climate system (Bonan 2008b).

## From Leaf Level to Community Dynamics

A comprehensive representation of ecosystem dynamics integrates processes that are relevant and observed at different temporal and spatial scales, from leaf to globe. Comprehensive ecosystem models simulate carbon assimilation processes at the leaf level, where carbon uptake is mediated by photosynthesis and stomatal conductance controls. Simple empirical models for primary productivity have followed the radiation use efficiency paradigm set by Monteith (1972), where primary productivity results from the efficiency at which plants convert absorbed radiant energy into carbon, while more mechanistic approaches have been following biochemical descriptions of photosynthesis based on enzyme kinetics and coupled to stomatal controls over CO<sub>2</sub> diffusion. Additionally, photosynthesis is mediated by nitrogen-rich enzymes, which results in a dependence of primary productivity on the environmental nitrogen availability and ability to mobilize it to leaves, and the new generations of models attempt for explicit coupling between the C and N cycles in order to describe these interactions. Upon assimilation, carbon is allocated to maintenance processes and structural development in different plant organs. Plant respiration results from metabolic activities associated with plant maintenance and growth, which can be modeled empirically based on response functions to climate or more mechanistically, linking environmental conditions to rates of enzymatic activity in the processes of cellular maintenance (see Amthor 2000). Plant growth depends on how the assimilated carbon is allocated to different plant organs. The distribution of assimilated carbon throughout the different plant organs is still one of the most unknown aspects of plant functioning, and its description in models can range from simple fixed fractions based on allometric relationships to schemes that prescribe it according to environmental limiting factors or evolutionary survival strategies (Franklin et al. 2012). At seasonal scales, the allocation of carbon is strongly controlled by day length, temperature, and precipitation patterns, which motivates the simulations of seasonality in leaf development to be frequently described by empirical phenology models (Richardson et al. 2013). However, at longer time scales, climate regimes, nutrient availability, and water storage capacity in soils control the long-term carbon investments between above- and belowground pools. These are not only controlled by abiotic factors, such as climate or soil properties, but also driven by between-plant competition for the same resources. The life cycle of plants depends strongly on these strategies and the ability of different species to cope with extreme

environmental conditions and disturbances. The emergence of dynamic vegetation models attempts to describe individual development as well as spatial and temporal changes in vegetation communities by embodying the principles of population dynamics (growth, mortality, reproduction, dispersal, and competition for resources) and succession rules (Prentice et al. 2007). The spatial distribution of vegetation is dominated by multiple mechanisms occurring at different temporal scales that are not explained exclusively by environment–vegetation relationships. Overall, these conceptually different modeling strategies stem from the different perspectives of various scientific disciplines, with particular interest in simulating the terrestrial biosphere including ecology, forestry, biogeochemistry, and climate-related sciences.

## Bringing Models and Observations Together

With the current increase in observational methods and data streams, today's main challenges relate to the comprehensive integration between the theory embedded in models and observational data to corroborate hypotheses of ecosystem functioning at different temporal and spatial scales.

To this end, relevant observations include measurements of vegetation and ecosystem pools and fluxes as well as of variables that influence or translate variations in ecosystem states. Measurements of CO<sub>2</sub> exchange at the leaf level are a primary source of information for building and parameterizing photosynthesis models. However, from a reductionist point of view, appropriately scaling up these processes to the whole-tree or canopy level would imperatively entail the description of biochemical states of leaves, tree hydraulic properties, and radiation regimes throughout the vertical profile. In this regard, observations of whole-ecosystem exchange of carbon and water with the atmosphere represent a top-down estimate of the whole total net ecosystem fluxes, including respiratory fluxes from heterotrophic decomposition. The partitioning of the different flux sources is possible through the measurement of component fluxes, such as transpiration or soil respiration. On the other side, biometric observations of above- and belowground biomass pools represent the temporal integral of assimilation, respiration, allocation, and litterfall processes occurring since establishment, hence ranging from instantaneous to decadal time scales. At longer scales and from regional to global extents, satellite remote sensing retrievals of vegetation properties like LAI and tree density and height, as well as spatial distribution of vegetation types, are important benchmarks to evaluate the representation of integrated processes of vegetation dynamics.

Comparisons between simulations and observations usually reveal deficiencies in modeling approaches, although they are limited in diagnosing the actual sources of errors, especially in complex models that incorporate multiple processes from leaf to biome level. Model–data fusion (MDF) approaches aim at transferring the information content of observations to modeling structures through parameter optimization (calibration) or adjustment of simulated states based on the minimization of cost functions that translate the mismatch between modeled and observed quantities. MDF is based on the principle that comparing patterns in responses or

states between observations and model simulations allows inferring the likelihood of the underlying mechanisms and hypothesis about ecosystem functioning (Reichstein and Beer 2008). MDF approaches enable the explicit treatment of the main sources of uncertainty arising from model structure, parameters, initial conditions, and observational data used in driving or constraining the model (Liu and Gupta 2007). Model parameters control the sensitivities of ecosystem responses to environmental conditions but also regulate internal dynamics related, for instance, to the maximum photosynthetic capacity, optimum temperature for photosynthesis, allocation of carbon to plant organs, surface to leaf area, etc. Although some of these parameters can be, and have been, measured, there are uncertainties related to observational methods as well as to its spatial and temporal representativeness, many times translated in the high variability of observations. The model structure is tested by exploring the likelihood of the model given the observations within the feasible distributions of parameters. The observational uncertainty can also be formally integrated in MDF approaches by weighing higher (lower) the observational records with lower (higher) uncertainties in the cost function. But the evaluation of the model is very dependent on the construction of the cost function, and modeling exercises have emphasized the challenges in the comprehensive representation of ecosystems. Given a multivariate comparison of model outputs with observations that translate different components of an ecosystem, the construction of an unbiased and comprehensive estimator of likelihood becomes a challenge per se. If integrating the multivariate observations of carbon and water fluxes and pools in ecosystem modeling provides a comprehensive test to model structures, it may also bias parameterizations when the datasets' dimensions can vary orders of magnitude, which would tend to favor model behavior for the most observed variable(s). Another aspect relates to inconsistencies between datasets, which could lead to parameterization biases and erroneous identification of poor model structures. The advantage of MDF lies in its ability to formally account for all these sources of uncertainties in bringing the theory embedded in models and observations together (Williams et al. 2009).

Overall, exploring model and data integration approaches reflects the possibility to test theories and hypotheses about ecosystem functioning corroborated by observations. Given the complexity of ecosystems, a comprehensive analysis values the overall coherence of our understanding of ecosystem functioning, but that does not detract from using simpler approaches that target exploring conceptual hypotheses. Ultimately, the association between ecosystem properties and functional behavior reflects the potential to extrapolate and scale the representation of ecosystem functioning.

## **Representing Ecosystem Functioning from Local to Regional Scales**

To generalize the representation of ecosystem functioning in space and time has been and still is a significant challenge. Up to what extent can the functional responses and internal dynamics of observed ecosystems be generalized to

unobserved regions? The link between plant structural types and the seasonality of phenology has motivated the classification of vegetation according to plant functional types (PFT). This classification assumes similar behavior in responses to environmental conditions, effects on ecosystem structure, and inherent processes. But the existing diversity holds a multiplicity of structural and functional characteristics that is well beyond the extent of a classification scheme. The possibility to move beyond classification schemes relies on the ability to link functional responses of plants and ecosystems to ubiquitous observations of relevant biotic and abiotic properties or states (Kattge et al. 2011).

---

## Future Directions

It is evident that plants react to the environment and influence the environment at different scales, from local to global scale. Direct responses to normal variation are relatively well understood, but in the future the regional feedbacks between plants and weather, i.e., the regional coupling between vegetation and the atmosphere, need to be understood better. These feedbacks are largely mediated through the water and energy cycles. For example, forest and grasslands were shown to exhibit very different energy fluxes to the atmosphere during heat waves and drought (Teuling et al. 2010). This way, they contribute differently to the development and stabilization of heat waves (Seneviratne et al. 2010). Moreover, direct and indirect responses of vegetation to extreme conditions need further study, with the main question, under which conditions irreversible processes like mortality are triggered? In this context it has recently been argued that vegetation responses to climate extremes can cause a positive global climate feedback by reducing the photosynthetic uptake (Reichstein et al. 2013). Last but not least, the fate of vegetation under a rapidly changing climate, as is being experienced now, will depend on its ability and velocity to adapt to those changing conditions. This is currently completely ignored in climate models (Stocker et al. 2013). Thus, joint studies in genetics, developmental biology, biogeochemistry, and biosphere modeling (e.g., Scheiter et al. 2013) need to be integrated in future research efforts.

---

## References

- Ainsworth EA, Yendrek CR, Sitch S, Collins WJ, Emberson LD. The effects of tropospheric ozone on net primary production and implications for climate change. *Annu Rev Plant Biol.* 2012;63:637–61.
- Amthor JS. The McCree–de Wit–Penning de Vries–Thornley respiration paradigms: 30 years later. *Ann Bot.* 2000;86(1):1–20.
- Atkin OK, Bruhn D, Hurry VM, Tjoelker MG. Evans Review No. 2: The hot and the cold: unravelling the variable response of plant respiration to temperature. *Funct Plant Biol.* 2005;32(2):87–105.
- Bahn M, Janssens IA, Reichstein M, Smith P, Trumbore SE. Soil respiration across scales: towards an integration of patterns and processes. *New Phytol.* 2010;186(2):292–6.

- Baldocchi D. Turner review No. 15. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Aust J Bot.* 2008;56(1):1–26.
- Barbour MG, Burk JH, Pitts WD, Gilliam FS, Schwartz MS. *Terrestrial plant ecology*. 3rd ed. Menlo Park: Addison Wesley Longman; 1999.
- Beerling DJ. *The emerald planet*. New York: Oxford University Press; 2007.
- Bonan GB. *Ecological climatology – concepts and application*. 2nd ed. Cambridge, UK: Cambridge University Press; 2008a.
- Bonan GB. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science*. 2008b;320(5882):1444–9.
- Bowman DMJS, et al. Fire in the earth system. *Science*. 2009;324(5926):481–4.
- Franklin O, Johansson J, Dewar RC, Dieckmann U, McMurtrie RE, Brännström Å, Dyzinski R. Modeling carbon allocation in trees: a search for principles. *Tree Physiol.* 2012;32(6):648–66.
- Greene EL. *Landmarks of botanical history: a study of certain epochs in the development of the science of botany: part 1, Prior to 1562 A.D.* Washington, DC: Smithsonian Institution; 1909.
- Griffis TJ. Tracing the flow of carbon dioxide and water vapor between the biosphere and atmosphere: a review of optical isotope techniques and their application. *Agr Forest Meteorol.* 2013;174–175(0):85–109.
- Heimann M, Kaminski T. Inverse modeling approaches to infer surface trace gas fluxes from observed atmospheric mixing ratios. In: Bouwman A-F, editor. *Approaches to scaling of trace gas fluxes in ecosystems*. Amsterdam: Elsevier; 1999. p. 275–95.
- Heimann M, Reichstein M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*. 2008;451(7176):4.
- Holdridge LR. Determination of world plant formations from simple climatic data. *Science*. 1947;105(2727):367–8.
- Janssens IA, et al. Reduction of forest soil respiration in response to nitrogen deposition. *Nat Geosci.* 2010;3(5):315–22.
- Kattge J, et al. TRY – a global database of plant traits. *Glob Chang Biol.* 2011;17(9):2905–35.
- Köppen W. *Die Klimate der Erde*. Berlin: Walter de Gruyt; 1923.
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World Map of the Köppen-Geiger climate classification updated. *Meteorol Z.* 2006;15(3):259–63.
- Levick SR, Rogers KH. Structural biodiversity monitoring in savanna ecosystems: integrating LiDAR and high resolution imagery through object-based image analysis. In: Blaschke T, Lang S, Hay G, editors. *Object-based image analysis*. Berlin/Heidelberg: Springer; 2008. p. 477–91.
- Liu Y, Gupta HV. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour Res.* 2007;43:W07401.
- Lyons T. Clouds prefer native vegetation. *Meteorol Atmos Phys.* 2002;80(1–4):131–40.
- Malmstrom C. Ecologists study the interactions of organisms and their environment. *Nat Educ Knowl.* 2010;3(10):88.
- McPherson RA. A review of vegetation – atmosphere interactions and their influences on mesoscale phenomena. *Prog Phys Geogr.* 2007;31(3):261–85.
- Mercado LM, Bellouin N, Sitch S, Boucher O, Huntingford C, Wild M, Cox PM. Impact of changes in diffuse radiation on the global land carbon sink. *Nature*. 2009;458(7241):1014–7.
- Meroni M, Rossini M, Guanter L, Alonso L, Rascher U, Colombo R, Moreno J. Remote sensing of solar-induced chlorophyll fluorescence: review of methods and applications. *Remote Sens Environ.* 2009;113(10):2037–51.
- Monteith JL. Solar-radiation and productivity in tropical ecosystems. *J Appl Ecol.* 1972;9(3):747–66.
- Norby RJ, Zak DR. Ecological lessons from free-air CO<sub>2</sub> enrichment (FACE) experiments. *Annu Rev Ecol Evol Syst.* 2011;42(1):181–203.

- Peñuelas J, Staudt M. BVOCs and global change. *Trends Plant Sci.* 2010;15(3):133–44.
- Pielke Sr RA, Avissar R, Raupach M, Dolman AJ, Zeng X, Denning AS. Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. *Glob Chang Biol.* 1998;4(5):461–75.
- Prentice IC, Bondeau A, Cramer W, Harrison SP, Hickler T, Lucht W, Sitch S, Smith B, Sykes MT. Dynamic global vegetation modelling: quantifying terrestrial ecosystem responses to large-scale environmental change. In: Canadell JG, editor. *Terrestrial ecosystems in a changing world.* Berlin/New York: Springer; 2007.
- Raunkjær C. The life forms of plants and statistical plant geography, being the collected papers of C. Raunkjær. Oxford: Oxford University Press; 1934.
- Reichstein M, Beer C. Soil respiration across scales: the importance of a model–data integration framework for data interpretation. *J Plant Nutr Soil Sci.* 2008;171(3):344–54.
- Reichstein M, Bahn M, Ciais P, Frank D, Mahecha MD, Seneviratne SI, Zscheischler J, Beer C, Buchmann N, Frank DC. Climate extremes and the carbon cycle. *Nature.* 2013;500(7462):287–95.
- Richardson LF. *Weather prediction by numerical process.* Cambridge University Press; 1922. 236 pp.
- Richardson AD, Keenan TF, Migliavacca M, Ryu Y, Sonnentag O, Toomey M. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agr Forest Meteorol.* 2013;169:156–73.
- Scheiter S, Langan L, Higgins SI. Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist.* 2013;198(3):957–69.
- Schulze E-D, Beck E, Muller-Hoenstein K. *Plant ecology.* New York: Springer; 2005.
- Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth-Science Rev.* 2010;99(3):125–61.
- Stocker TF, Dahe Q, Plattner G-K. Climate change 2013: the physical science basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers. IPCC; 2013.
- Teuling AJ, Seneviratne SI, Stöckli R, Reichstein M, Moors E, Ciais P, Luysaert S, van den Hurk B, Ammann C, Bernhofer C, Dellwik E, Gianelle D, Gielen B, Grünwald T, Klumpp K, Montagnani L, Moureaux C, Sottocornola M, Wohlfahrt G. Contrasting response of European forest and grassland energy exchange to heatwaves. *Nat Geosci.* 2010;3:722–7.
- Williams M, et al. Improving land surface models with FLUXNET data. *Biogeosciences.* 2009;6(7):1341–59.
- Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JH, Diemer M. The worldwide leaf economics spectrum. *Nature.* 2004;428(6985):821–7.
- Zaehle S. Terrestrial nitrogen–carbon cycle interactions at the global scale. *Philos Trans R Soc B Biol Sci.* 2013;368(1621).
- Zaitchik BF, Macalady AK, Bonneau LR, Smith RB. Europe’s 2003 heat wave: a satellite view of impacts and land–atmosphere feedbacks. *Int J Climatol.* 2006;26(6):743–69.

## Further Reading

- Larcher W. *Physiological plant ecology: ecophysiology and stress physiology of functional groups.* Springer; 2003.
- Purkis SJ, Klemas VV. *Remote sensing and global environmental change.* Chichester: Wiley; 2010.

Paulette Bierzychudek

## Contents

Introduction .....	30
Structure of Plant Populations .....	32
Temporal Patterns of Population Dynamics .....	37
Causes of Different Temporal Patterns of Plant Population Dynamics .....	38
What Forces Regulate the Sizes of Plant Populations? .....	42
The Role of Stochastic Influences, Especially in Small Populations .....	46
Incorporating Population Structure into Models and Analyses .....	49
Spatial Patterns of Population Dynamics .....	59
A Brief Guide to Methodological Approaches Used in Field Studies of Plant Population Dynamics .....	61
Defining the Boundaries of a Population .....	61
Censusing Populations .....	61
Future Directions .....	63
References .....	63

---

## Abstract

- Population dynamics is the study of how and why population sizes change over time.
- Repeated censuses of individuals within populations are the core data collected by plant ecologists studying population dynamics.
- Plant populations are characterized by their size (or density) and their structure (the numbers of individuals of different ages and sizes).
- Plant population ecologists use observations, experiments, and mathematical models to document and understand patterns of population dynamics.
- Most plant populations appear to be regulated by density-dependent forces; resource competition and natural enemies are the most likely forces responsible for regulation.

---

P. Bierzychudek (✉)  
Department of Biology, MSC 53, Lewis & Clark College, Portland, OR, USA  
e-mail: [bierzych@lclark.edu](mailto:bierzych@lclark.edu)



- Stochastic forces have particularly strong effects on small populations.
- Population viability analyses assess how stochastic forces affect a population's probability of extinction and can be used to identify effective management options.
- Demographic differences among individuals affect their potential contributions to population dynamics.
- Transition matrix models are the most important model used to study plant populations and guide the management of harvested populations and species of conservation concern.
- Regional dynamics of assemblages of plant subpopulations, such as metapopulations, have not been well studied in plants and are an active area of research.

---

## Introduction

The Haleakala silversword, *Argyroxiphium sandwicense* subsp. *macrocephalum*, is an unusual plant for many reasons, not the least of which is its striking appearance, like the offspring of a marriage between a footstool and a pincushion (Fig. 1).

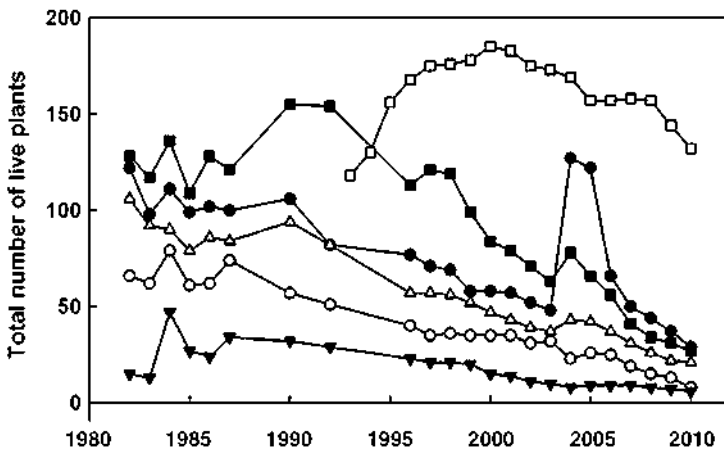
Found only on Mt. Haleakala, a dormant volcanic cinder cone on the Hawaiian island of Maui, this remarkable plant lives on mostly barren, rocky, unstable slopes at elevations of 2,100–3,000 m. Individuals live for up to 50 years before sending up a flowering stalk that bears as many as 600 flower heads. After this one reproductive episode, the plant dies.

The Haleakala silversword population has survived the cattle and goats that once grazed the mountain and persists despite the fact that tourists impressed by their bizarre appearance once routinely “bowled” these plants down the mountainside or uprooted them for souvenirs. Protection from these threats in the 1930s greatly increased the silversword's numbers over the next 60 years. By the late 1990s, the silversword population was estimated to be 16 times larger than it had been in 1935, and this iconic plant came to be considered one of the Hawaiian Islands' conservation success stories. However, since the mid-1990s, the silversword population is once again in decline (Fig. 2; Krushelnycky et al. 2013).

These trends would not have been apparent except for observers who chose to census the number of silversword individuals in the Mt. Haleakala population, starting with park ranger S.H. Lamb in 1935 (U.S. Fish and Wildlife Service 1997). Census data are key to understanding the dynamics of plant populations, i.e., how numbers of individuals change over time, and to determining the causes of those changes. This chapter will examine the history, key concepts, main methodologies, and important unanswered questions in the field of plant population dynamics.

A population is a group of individuals belonging to the same species, living in the same area. The study of plant population dynamics, i.e., how and why plant populations change in numbers over time, is a relatively recent chapter in plant ecology. While a few earlier workers had carried out repeated censuses of plant populations, British ecologist John L. Harper (1925–2009) revolutionized

**Fig. 1** A flowering Haleakala silversword (Photo by Forest and Kim Starr)



**Fig. 2** Numbers of Haleakala silversword individuals at a high-elevation canyon rim site (*open squares*) and at five sampling areas on the crater floor (*other symbols*) (Figure from Krushelnycky et al. 2013)

how ecologists thought about plants with his 1967 paper, “A Darwinian Approach to Plant Ecology,” and his 1977 book, *Population Biology of Plants*. Before Harper, it was mostly zoologists, not botanists, who studied the biology of populations. Harper, his students, and many other ecologists he influenced developed the quantitative, process-oriented, and often experimental approach to the study of plant population dynamics that characterizes the field today. In fact, John Harper argued that plants were more suitable than most animals for the study of population dynamics because “plants stand still to be counted and do not have to be trapped, shot, chased, or estimated” (Harper 1977, p. v).

Plant population ecologists are interested in knowing what trends characterize plant populations over time – do they increase? Decrease? Remain constant? Are these patterns predictable or stochastic? What forces are responsible for the different patterns? These questions are of interest not only for their own sake, but also because their answers can lead to effective problem solving in the fields of agriculture, forestry, range management, natural area management, and species conservation.

This chapter will begin by describing the structure of plant populations and by considering some aspects of plant biology that affect how plant populations are studied, such as the relationship between size and age, and how “individuals” are defined. This will be followed by a description of some of the spatial and temporal patterns displayed by different populations and a consideration of the possible causes of these different patterns. The chapter will briefly review some of the primary methodological approaches used to study plant populations in the field. Throughout, it will illustrate some of the ways these approaches have been applied to address particular practical problems, especially in the area of biodiversity conservation.

---

## Structure of Plant Populations

A consideration of the structure of plant populations starts with the question “what is an individual?” Many herbaceous and woody plant species, including some tree species, are capable of spreading horizontally by means of rhizomes and runners. For such species, an “individual” is a nebulous concept and not necessarily a meaningful distinction. It is easy to recognize a newly germinated seedling as a single individual, but that individual can grow into a patch of grass many meters in diameter or an aspen clone that covers an entire hillside. These differences between individuals are a consequence of the modular growth form typical of most plant species. Deciding how to quantify the number of individuals in a population is often the first challenge that must be confronted when studying a plant population’s dynamics.

Plant ecologists have found it useful to distinguish between two kinds of individuals. Individuals that arise from different propagules and are thus genetically distinct from one another are known as *genets*. However, because an individual that has spread horizontally may break up into physically independent units, not all

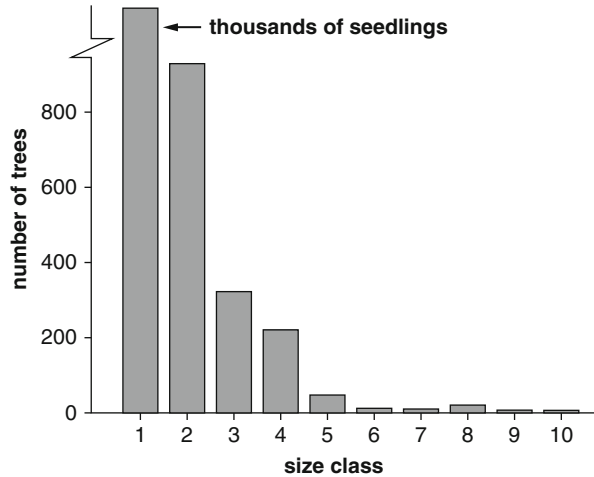
independent units are distinct genets. Individuals that are physiologically independent of one another are considered separate *ramets*, regardless of their genetic similarity. The number of genets in a plant population can be much lower than the number of ramets. Ramets are often easier to recognize than genets, so this definition of an individual is more frequently used. Because the identification of individuals can be so challenging in many species, studies of plant populations have historically been biased toward those species in which individuals are relatively easy to define; we know much less about species with strong propensities toward vegetative spread than about species that tend to restrict their growth to the vertical dimension.

Once the issue of how to define an individual has been addressed, there are two ways to express the size of a population. Sometimes a population's size is described as the number of individuals it contains; other times it is the population's density that is reported, i.e., the mean number of individuals per unit of area. It is important to keep in mind that density is an average measure for the entire population and that individuals can be distributed in space in three different ways. Individuals of a species are sometimes spaced regularly, such that the mean density of individuals in a series of sampling plots is greater than the variance in density among plots. Alder shrubs in the Alaskan tundra are regularly spaced; Chapin et al. (1989) suggested that regular spacing is most likely to be found in habitats with low species diversity and intense competition for resources, like desert or tundra. Rarely, individuals are randomly distributed in space (Hutchings 1997); in this case the mean density of individuals among plots is similar to the variance. Finally, individuals are most often found in a clumped distribution (Hutchings 1997), with the variance in the density of individuals among sample plots being greater than the mean. A clumped distribution pattern can occur if the underlying physical environment is heterogeneous, with individuals clustered within the suitable patches and absent from the unsuitable ones. It can also arise from the fact that many plant species have rather localized seed dispersal, so that seedlings are often found in close proximity to their parents.

In addition to variation in their spatial distribution, individuals within a population can vary in such characteristics as their size, their age, or their sex. These so-called demographic parameters often have important effects on how each individual contributes to a population's dynamics. Because most plant species have perfect flowers, there is only one sex in most plant populations; all individuals are hermaphrodites. In such species, sex is not a particularly important demographic characteristic. Sex is a more important demographic parameter in many animal populations and in those plant species with separate sexes. In such species, the ratio of male to female individuals can strongly affect a population's potential for increase.

In animals with determinate growth, age is a very important demographic parameter. Individual animals often must reach a certain age before achieving sexual maturity, and an individual's probabilities of dying and of giving birth (probabilities often referred to as *vital rates*) are well correlated with its age. By contrast, consider a seedling *Eucalyptus*, a 5 m tall *Eucalyptus* in the forest understory, and a mature 100 m tall *Eucalyptus* tree, each of which has very different probabilities of dying and of reproducing. While it is certain that the mature tree is older than the seedling, the age of the understory individual is more

**Fig. 3** Numbers of different-sized *Araucaria cunninghamii* individuals per hectare in Papua, New Guinea (Data from Enright and Ogden 1979)



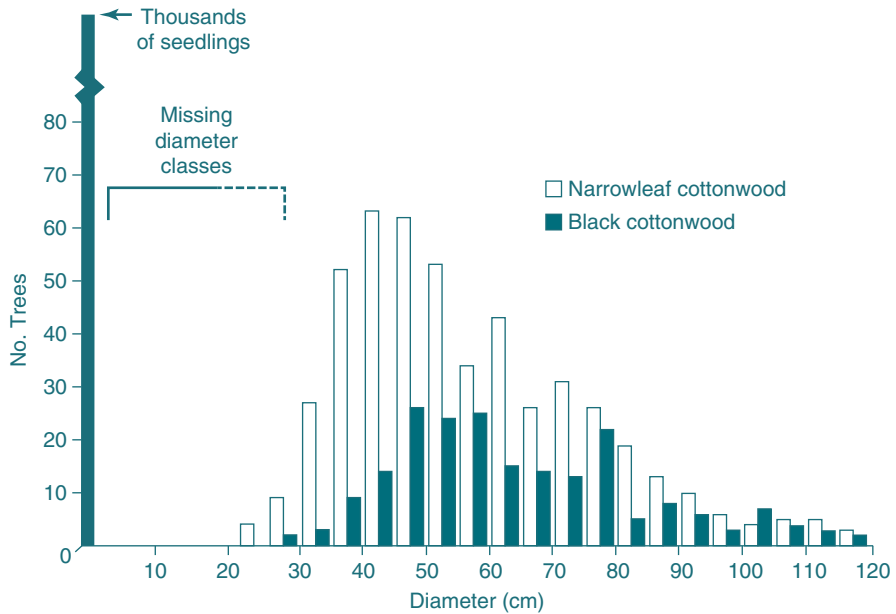
difficult to predict. Such an individual might be quite young, if it germinated in a light gap and there were no other individuals growing nearby to compete for water or nutrients. Alternatively, such an individual might be considerably older if its growth has been suppressed by competition with larger neighbors for many years. But in an important sense, its age doesn't matter; this individual won't flower or set seed until or unless it reaches the canopy. The indeterminate growth form of this and many other plants means that an individual plant's probability of dying or reproducing tends to be more closely related to its size, or to its growth stage, than to its age (Gurevitch et al. 2002). Individuals of different sizes or stages have very different potentials to influence the population's future size.

Therefore, many studies of plant populations record information on the *size* or *growth stage* of each individual in the population. This information can be displayed in the form of a histogram. Many plant populations in nature display a size structure like that shown by the tropical tree *Araucaria cunninghamii* in Fig. 3.

This pattern has three primary causes: first, many plants tend to produce large numbers of small propagules. Second, individuals experience mortality as they grow. And third, small individuals are generally more vulnerable to mortality than larger ones are, which is why numbers of individuals in the larger size classes diminish much more gradually than those in the smaller size classes do.

The observation of deviations from this pattern can generate interesting questions about a population's history. For example, in Yellowstone National Park, USA, in the floodplain of the Lamar River, there are mature cottonwood trees and large numbers of seedling cottonwoods, but almost no individuals intermediate in size between these two classes (Fig. 4).

According to Beschta (2003), this gap suggests that little or no recruitment of this riparian species occurred between 1920, when wolves were hunted to extinction in the Park, and 1995, when they were reintroduced. While wolves were absent from the Park, Beschta hypothesized, elk boldly grazed in these open river valleys, eating

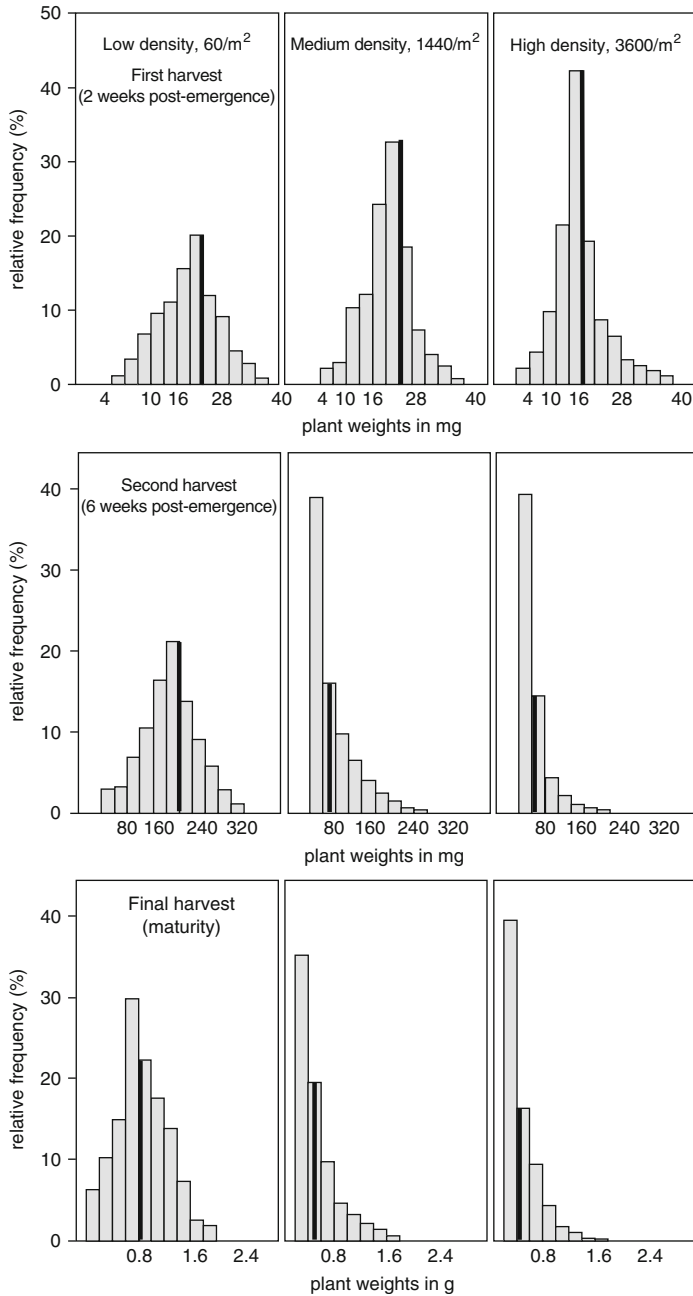


**Fig. 4** Numbers of cottonwood trees of different trunk diameter size classes in the Lamar Valley of Yellowstone National Park, USA (Reprinted from Beschta 2003)

young seedlings and saplings and preventing the establishment of mature trees. With the recent return of wolves to the valley, elk have become more wary, rarely venturing out of the forest into the open floodplain habitat (Beschta 2003), allowing seedling cottonwoods to survive unbrowsed. While this hypothesis for the cottonwood stage structure in Yellowstone remains controversial (Winnie 2012), it is clear that the unexpected size structure of this cottonwood population demands an explanation.

In even-aged populations of agricultural or greenhouse plants, other patterns of size structure are observed, and it becomes possible to examine how these patterns develop and change over time. Frequency distributions of seedling weights are typically approximately normal (Fig. 5, top row).

Variation in seedling size exists because seed sizes are rarely uniform, and the size of a seed has a strong influence on the size of the seedling that emerges from it (Hutchings 1997). Over time, as seedlings grow, their weight distributions tend to become increasingly skewed (Fig. 5, middle, bottom rows), especially at higher densities, for several reasons (Hutchings 1997). First, there is genetic variation for growth rate among a group of individuals. Second, the timing of a seedling's emergence relative to that of its closest neighbors can give certain seedlings an initial growth advantage or disadvantage. Third, the spacing of a growing plant's immediate neighbors determines the amount of resources available to it. For all these reasons, many individuals may remain small, spindly, and fail to flower or produce seeds. This effect is most extreme and rapid in high-density populations (Fig. 5, right-hand column).



**Fig. 5** Frequency distributions of plant weights for flax, *Linum usitatissimum*, sown at three densities. Y-axis = percent of the population in each weight class. Heavy black bar represents the mean plant weight. Top row: seedlings harvested 2 weeks after emergence; weights in mg. Middle row: 6-week-old plants; weights in mg. Bottom row: mature plants; weights in g (Figure from Harper (1977))

Over time, the death of some of the small individuals in a dense population can allow other individuals to achieve larger size, and it is common to observe the size structure of such populations shifting over time as shown in Fig. 5. Mortality resulting from competition simultaneously alters the population density. Thus density and individual plant weight change in concert. In many populations this process of “self-thinning” has been shown to follow a temporal pattern represented by the relationship

$$w = cN^{-k} \quad (1)$$

where  $w$  represents mean individual plant weight,  $N$  is density of surviving plants, and  $c$  is a constant that varies among species. The value of the parameter  $k$  is approximately  $3/2$  for a wide range of plant species (Harper 1977). Differences in plant size caused by intraspecific competition ultimately lead to differences in performance. These differences among individuals within a population can have important effects on the potential of a population to change in numbers in the future.

---

## Temporal Patterns of Population Dynamics

The size of any population changes over time because individuals are born and die and/or migrate into or out of the population. In other words,

$$N_{t+1} = N_t + B - D + I - E \quad (2)$$

where  $N_{t+1}$  = a population’s size or density at time  $t + 1$ ,  $N_t$  = its size/density one unit of time (usually a year) earlier,  $B$  = the number of births,  $D$  = the number of deaths,  $I$  = the number of immigrants into the population, and  $E$  = the number of emigrants from the population during the period between  $t$  and  $t + 1$ . Because plants are sessile, changes in the size of a plant population are typically much more influenced by births and deaths than by immigration/emigration (though the influence of dispersal will be addressed in section “[Spatial Patterns of Population Dynamics](#)”).

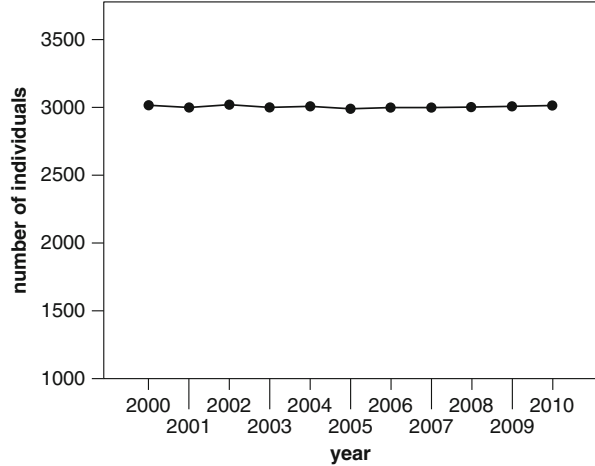
It is easy to imagine that most plant populations must be in a state of equilibrium ( $N_{t+1} = N_t$ ), with births balancing deaths (Fig. 6).

Dramatic changes in the abundance of plant species are rarely observed. But long-term monitoring of plant populations reveals that few populations are static, at least not for long, and that even those that appear static are actually undergoing considerable turnover (Silvertown and Charlesworth 2001). Static populations tend to be restricted to species where individuals are long-lived, like trees, to habitats that rarely experience disturbances and to locations where environmental conditions are predictable from year to year. Few species or environments fit this description.

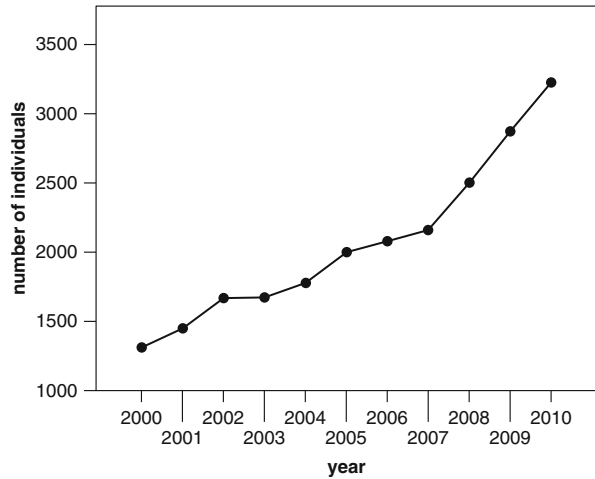
Instead, the sizes of plant populations typically fluctuate over time, either deterministically, stochastically, or both. Some populations appear to be increasing in numbers (Fig. 7); others appear to be decreasing (Fig. 8).



**Fig. 6** A hypothetical population with little or no change in numbers with time



**Fig. 7** A hypothetical population in which numbers of individuals are increasing with time

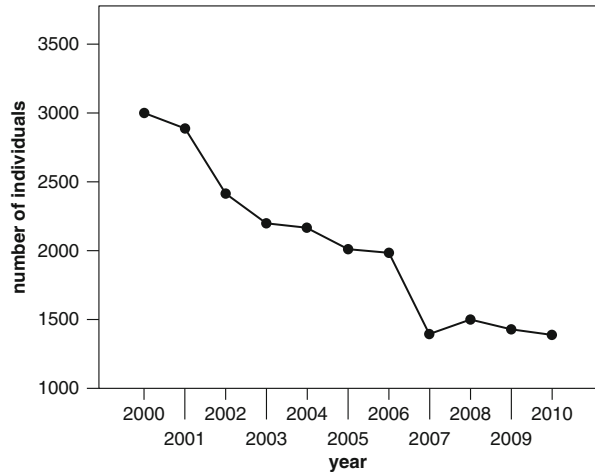


Over longer time periods, the same population can display both patterns. Often superimposed on these trends, and also evident in populations with little overall change, is an unpredictable “wobble” in numbers of individuals (Fig. 9).

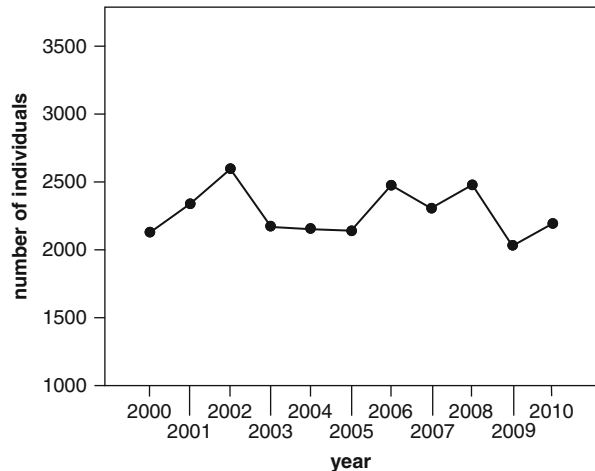
## Causes of Different Temporal Patterns of Plant Population Dynamics

What causes these different patterns? One important approach to understanding patterns of population dynamics is to build mathematical models that vary in the assumptions they make about the forces that might influence a population's

**Fig. 8** A hypothetical population in which numbers of individuals are decreasing with time



**Fig. 9** A hypothetical population with unpredictable fluctuations but no overall trend in numbers with time

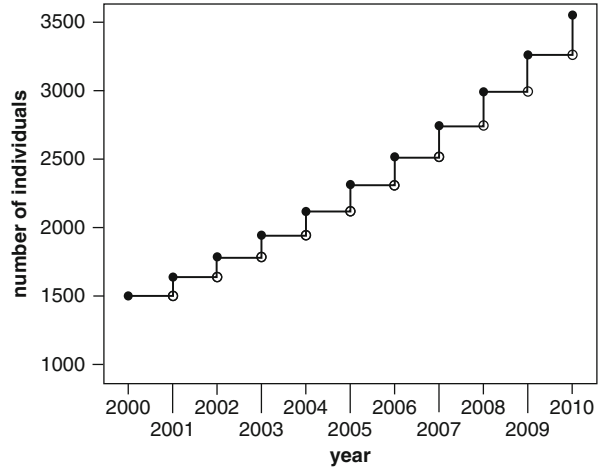


dynamics and then to compare the dynamics of model populations to information about natural populations obtained by regular censuses.

A model is simply a mathematical representation of a hypothesis; assumptions about possible forces at work are represented as elements of that mathematical expression. The goal of model building is to develop a model: (a) that is as simple as possible, (b) that captures the essential forces responsible for a population's dynamic behavior, and (c) that omits details that do not provide additional explanatory power. Such a model will concisely explain the reasons for a particular pattern of population dynamics.

This section will consider a series of such models/hypotheses, starting with simple ones and moving on to models of increasing complexity and realism. The simpler models, so-called unstructured models, treat all individuals as equal,

**Fig. 10** A hypothetical population with an initial size,  $N_0$ , of 1,500 individuals and an annual growth rate,  $\lambda$ , of 1.09



ignoring demographic characteristics such as size differences among individuals. While such models may be unrealistic, they provide an important foundation upon which to build more realistic versions. The more complex versions, so-called structured models, incorporate demographic variation among individuals.

The simplest representation of population growth is the *geometric model*:

$$N_{t+1} = N_t \cdot \lambda \quad (3)$$

where  $\lambda$  = the population's net reproductive rate, i.e., the ratio of  $N_{t+1}$  to  $N_t$ . In Eq. 3,  $\lambda$  is a constant; in other words, this model contains the implicit assumption that the population's net reproductive rate does not change as a function of the population's size, and is not influenced by changing environmental conditions. This model can be generalized to longer time periods:

$$N_t = N_0 \cdot \lambda^t \quad (4)$$

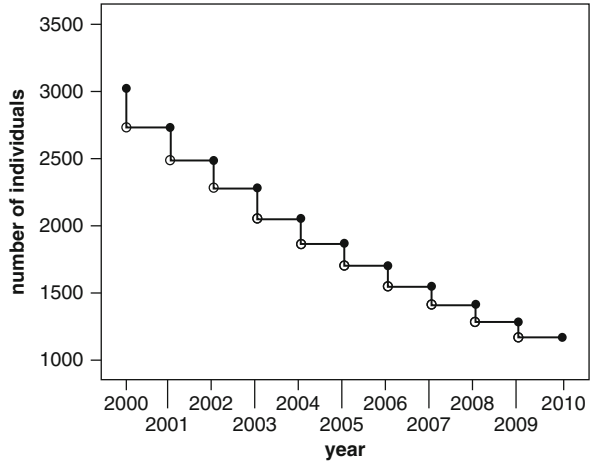
A population with  $\lambda > 1$  is increasing geometrically (see Fig. 10), one with  $\lambda < 1$  is decreasing geometrically (see Fig. 11), and one with  $\lambda = 1$  is not changing in size. Because this model is in the form of a difference equation, it is a particularly apt way to describe a population whose size grows (or shrinks) in "spurts" that occur once a year. This is the case, for example, for annual species in which individuals live for one growing season, produce seeds, and die at the end of that season, their seeds germinating at the beginning of the next growing season.

It is also possible to express the hypothesis that the population growth rate is a constant in continuous time, a form that some readers may find more familiar:

$$\frac{dN}{dt} = rN \quad (5)$$

In this continuous-time model of exponential (i.e., geometric) population growth,  $r$  is a parameter known as the intrinsic or instantaneous rate of increase

**Fig. 11** A hypothetical population with an initial size,  $N_0$ , of 3,000 individuals and an annual growth rate,  $\lambda$ , of 0.91



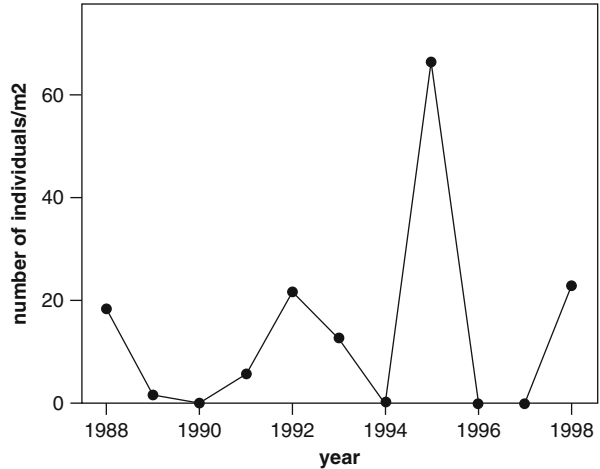
and is defined as the difference between the per capita birth and death rates. A growing population has an  $r > 0$ , while a declining one has an  $r < 0$ . This model produces the same results as those shown in Figs. 10 and 11 except that the change in population size is continuous rather than stepwise. For more about the correspondence between the difference-equation and continuous-time forms of the geometric/exponential growth model, see Begon et al. (1996).

When a population's dynamics fit the pattern of change in numbers over time shown in Fig. 10, it suggests that necessary resources are superabundant relative to the resource requirements of individuals in the population. This pattern can be observed in plant populations that have recently colonized an environment where competitors and predators are rare and where resources are temporarily superabundant, such as species occupying a recently abandoned agricultural field, a newly logged forest, or the site of a recent fire, flood, or other catastrophic disturbance. Many species are specifically adapted to these habitats and are rarely seen in other circumstances, surviving from disturbance to disturbance by means of long-lived seed banks.

However, few populations exhibit a pattern of geometric growth for more than a short time; no population is capable of increasing forever without limit. One obvious cause of population decline is a directional change in the suitability of the environment resulting from successional change, e.g., as a meadow is colonized by shrubs and trees, herb and grass species decrease in abundance. It is more challenging to understand changes in numbers that occur in environments that are not undergoing such obvious environmental change.

Populations that experience a positive growth phase at first are often limited (eventually) by abiotic or biotic factors. Some of these factors act with an intensity that is independent of the size of the population subject to them; these are often referred to as density-independent limiting forces. For example, a severe drought might cause the death of all of the seedlings whose roots failed to reach a particular soil depth, no matter whether the density of seedlings was relatively high or

**Fig. 12** Changes in plant density of *Linanthus parryae* over 10 years (Data from Schemske and Bierzychudek 2001)



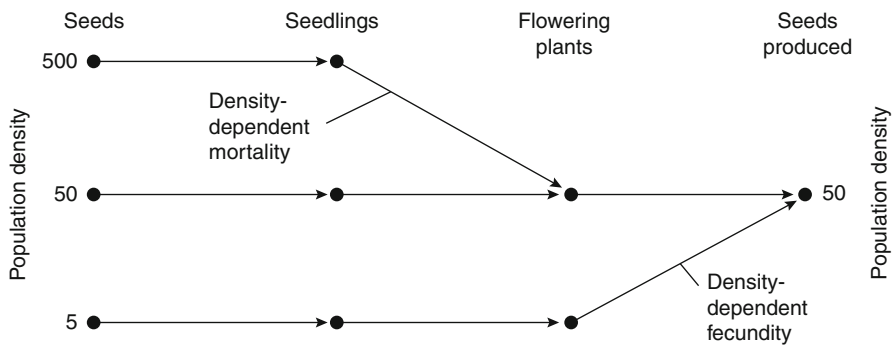
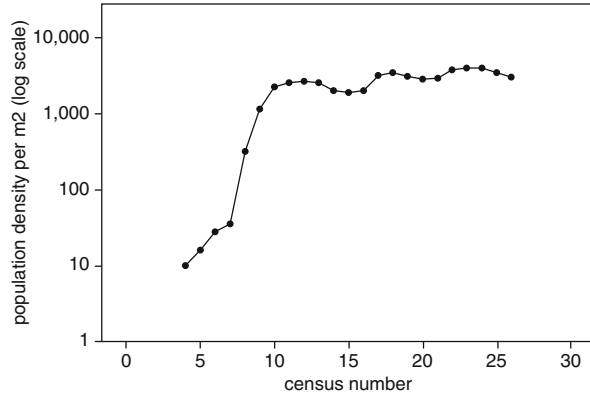
relatively low. A late frost might cause the abortion of all the developing seeds in a population. Density-independent mortality can periodically reduce the size of a population; Fig. 12 shows population trends for *Linanthus parryae*, a desert annual, in the Mojave Desert of southern California, USA. The years when no adults were recorded had extremely low rainfall; the population persisted during these periods by means of dormant seeds. It is hard to imagine a population in which density-independent forces have no effect on population density or dynamics. However, while density-independent mortality sources can limit the size of a population, they cannot regulate it (Watkinson 1997).

## What Forces Regulate the Sizes of Plant Populations?

Many populations appear to be regulated, i.e., to behave as though there were upper and lower bounds on their size, in that the population tends to return to its previous size or density following a perturbation. The population of the fast-growing annual *Poa annua* shown in Fig. 13, for example, has reached a more or less stable density in a relatively short time. Density-independent mortality sources cannot explain the existence of these bounded patterns. To understand regulated patterns of population dynamics, it is necessary to look to forces whose effects are proportionally more severe when the density of a population is high than when it is low, i.e., forces whose effects are density dependent.

For example, a plant seed might not germinate successfully unless it falls in a *safe site*, a microsite that has the appropriate physical and biological conditions that will permit a seedling to emerge safely from a seed (Harper 1977). Because any environment contains a limited number of safe sites, the mortality rate from failure to land in a safe site will be greater when large numbers of seeds are produced than when few seeds are produced (Fig. 14). Or, consider a fungal pathogen that infects and kills individual hemlock trees that are too weak to mount a defense. An individual tree is

**Fig. 13** Changes in density (on a log scale) of *Poa annua* colonizing abandoned land (Data from Law (1981), figure redrawn from Watkinson (1997))

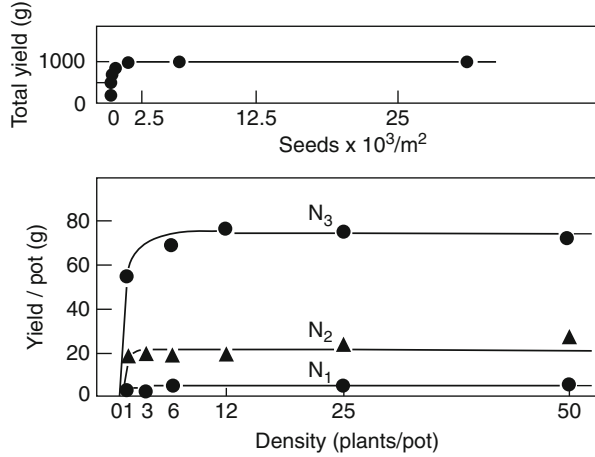


**Fig. 14** Schematic diagram illustrating the role of density-dependent forces in regulating population density (Figure reprinted from Silvertown and Charlesworth 2001)

less vulnerable to infection in a population where individuals are widely spaced than in one where trees are crowded and sunlight or nutrients are in short supply. Thus mortality due to fungal attack may be density dependent. Finally, when the number of adult plants is small, each individual will grow larger and produce more seeds than when individuals are denser (Fig. 14). All these forces tend to dampen variations in population density and thus to regulate population numbers.

Because so many plant populations appear to be regulated in some way, the existence of density dependence has been investigated in a wide range of species. Both observational and experimental approaches have been used. Two kinds of observational studies have provided evidence for density-dependent population regulation. First, ecologists have looked for positive correlations between plant size and interplant distance, considering such patterns to be evidence that plant size is controlled, to some degree, by the intensity of competition with neighbors. Other kinds of observational studies have taken advantage of natural variation in population density, either in time or in space, to determine whether and how a population's birth and death rates vary with density. However, tightly regulated populations are

**Fig. 15** Relationships between original density of seeds or plants and the final yield biomass (in g) for clover, *Trifolium subterraneum* (top), and for a grass, *Bromus unioloides* (bottom) (Figure from Harper 1977)



expected to exhibit little natural variation in density; thus the stronger the regulation, the harder it is to detect (Silvertown and Charlesworth 2001). Another shortcoming of both kinds of observational studies is that spatial variation in environmental factors could complicate the interpretation of observed trends (Antonovics and Levin 1980). An alternative approach has been to alter density experimentally, either in the field or in the greenhouse, and to measure how survival and fecundity rates vary with density.

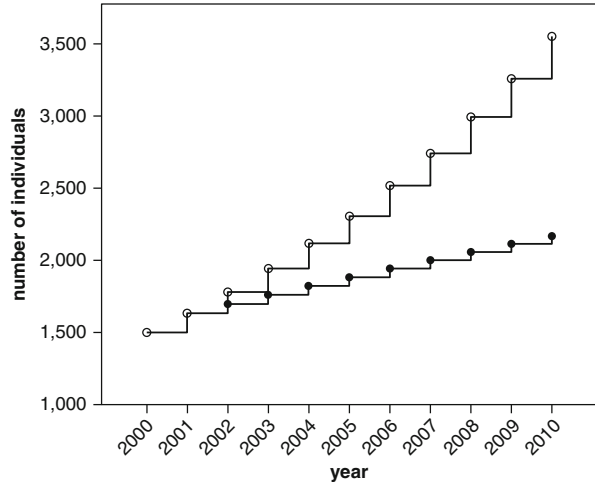
A large number of such studies have repeatedly demonstrated that variation in population density can have dramatic effects on individual growth rates, fecundity rates, and mortality rates (Harper 1977; Antonovics and Levin 1980). At relatively low densities, individual plants tend to exhibit few reductions in performance. However, at medium densities, reductions are often seen in growth rate and reproductive output. Finally, at relatively high densities, mortality rates can dramatically increase. For example, studies of how final biomass depends on the density of seeds originally sown have repeatedly confirmed the “law of constant final yield” (Fig. 15). Similarly, the relationship between plant weight and plant density represented by the “ $-3/2$  self-thinning law” (Eq. 1) illustrates the powerful influence of density. Because these reductions are observed even in controlled environments where herbivores and parasites are absent, it is clear that these reductions are very often a consequence of resource competition among conspecific neighbors.

The potential effect of intraspecific competition can be incorporated into the previous model of population growth, shown here in the form of a difference equation (contrast this with Eq. 3):

$$N_{t+1} = \frac{N_t \lambda}{1 + aN_t} \quad (6)$$

In this so-called logistic model,  $a$  equals  $(\lambda - 1)/K$ , where  $K$  is the carrying capacity of the environment for the species (in units of numbers of individuals).

**Fig. 16** These two hypothetical populations show the contrast between geometric and logistic growth. The population represented by open circles is growing geometrically at an annual rate,  $\lambda$ , of 1.09. The population represented by the closed circles is growing logistically at the same annual rate,  $\lambda$ , of 1.09, but with  $K = 3,000$ , i.e.,  $\alpha = 0.0003$



This model differs from the geometric model only in its modification of the assumption that  $\lambda$  is a constant. The logistic model assumes that the growth rate is equal to  $\lambda$  when  $N_t$  is near 0 and that it decreases linearly toward 1 as  $N_t$  approaches  $K$ . The logistic model generates the population dynamics shown by the closed circles in Fig. 16. A derivation of this model can be found in Begon et al. (1996).

Some readers may be more familiar with the continuous-time form of the logistic model,

$$\frac{dN}{dt} = rN \left( \frac{K - N}{K} \right) \quad (7)$$

Equation 7 contains the same assumption about the linear dependence of the population growth rate on  $N$  as does Eq. 6. Both models predict that a population's numbers should grow until they reach an equilibrium size ( $K$ ), at which point deaths balance births. The observation in nature of a trajectory like that in Fig. 13 implies that a population's dynamics are largely governed by intraspecific competition for one or more limited resources. Many populations that initially display a pattern of geometric growth eventually reach a more or less stable size like that predicted by the logistic model.

Resource competition is not the only biotic interaction potentially capable of regulating plant populations; interactions with enemies like herbivores, seed predators, and plant parasites such as fungi and bacteria also have the potential to act as regulatory forces. An influential paper by Hairston et al. (1960) argued that the fact that plants generally appear "abundant and largely intact" implied that it was unlikely that plant populations could be regulated by their enemies. However, this argument has been challenged for many reasons (see review by Crawley 1989). Indeed, it is often assumed that natural enemies can regulate plant populations; for example, efforts to use biological control to reduce weed populations are grounded in this assumption (Halpern and Underwood 2006). In addition, a popular hypothesis



to explain why plant species that have been transported from their native location to a new geographical region often become invasive is the “enemy release hypothesis.” This hypothesis proposes that movement to a new location releases nonnative species from the regulatory effects of the enemies that held them in check where they were native.

The relative importance of natural enemies in regulating plant populations remains controversial, however, because they have been less well investigated experimentally. Much of the evidence supporting the role of natural enemies comes from large-scale releases of herbivores for purposes of weed control; such releases are neither randomized nor replicated. Better evidence comes from controlled experiments in which plants in plots protected from herbivore activity by caging or insecticide application are compared to plants in unprotected control plots. The results of such studies have been mixed, with vertebrate herbivores typically exerting stronger regulatory effects than insects and some studies showing no evidence for herbivore regulation (Crawley 1989). Because these methods of herbivore exclusion have been shown to have unintentional treatment effects, even those studies implicating herbivores as important do not necessarily provide compelling evidence for the role of natural enemies in regulating plant population dynamics (Crawley 1989). Additionally, such studies are often limited to measuring the impact of enemies on individual plant performance, and their results cannot easily be “scaled up” to provide insights about the regulation of entire populations. For example, a herbivore that reduces an individual’s seed production might not affect the population’s dynamics if the availability of safe sites limits the numbers of seeds that can germinate successfully (Crawley 1989; Halpern and Underwood 2006). Finally, studies investigating the effect of natural enemies on plant performance rarely investigate whether such effects are density dependent, as they must be if they are to be able to regulate plant population dynamics (Halpern and Underwood 2006). The role of natural enemies in regulating plant populations is an important area in need of additional investigation, especially because the findings of these efforts have important implications for the control of pests and the management of plant invasions (Halpern and Underwood 2006).

## **The Role of Stochastic Influences, Especially in Small Populations**

In addition to seeking to understand the forces that regulate sizes or densities of plant populations, plant ecologists are also interested in understanding the role of stochastic influences on population dynamics. Such influences are especially important in small, at-risk plant populations. Ecologists recognize two kinds of stochastic influences. Environmental stochasticity refers to erratic, unpredictable variation among years in abiotic and biotic parameters such as rainfall, temperature, winter snow depth, dates of first and last frost, or population sizes of predators, parasites, or interspecific competitors. These forces can be thought of as external to the population, and they affect all individuals in similar ways. Environmental stochasticity, on short or long time scales, leads survival and recruitment rates to

vary from 1 year to the next, producing temporal patterns like that in Fig. 9. All natural populations, regardless of their size, are influenced to some degree by environmental stochasticity.

In contrast to environmental stochasticity, demographic stochasticity refers to variation in vital rates arising from chance differences in the fates of different individuals; this kind of variation arises from within the population itself rather than from external forces. For example, an average plant in a population might be expected to produce 100 seeds, but not every plant conforms to this average. Some might make more than 100 seeds, some fewer. Demographic stochasticity is primarily a concern for small populations, because in large populations, there are abundant opportunities for these random deviations from the mean to cancel one another out. For this reason, large populations are much more likely to follow the law of averages. In a small population, however, it is likely that these random interindividual differences will lead to deviations in the numbers of deaths or births in different years and thus to a population size that varies randomly from 1 year to the next. Since small populations also experience environmental stochasticity, they can fluctuate in size to a considerable degree between years. This fluctuation is important because it greatly increases their vulnerability to extinction.

The way environmental stochasticity affects population dynamics, and thus a population's extinction risk, is important but somewhat counterintuitive. Temporal fluctuations in vital rates do more than cause a population's dynamics to be more variable over time; they can actually cause a population to grow more slowly than it would in the absence of variability. Morris and Doak (2002) illustrate this effect using the following example. Imagine a population of 100 individuals with an annual growth rate,  $\lambda$ , that can take one of two values, 0.86 and 1.16, each value occurring with a 50 % probability. The average of these two values is 1.01; thus, we might reasonably expect that this population would have 14,477 individuals 500 years in the future:

$$100 \cdot (1.01)^{500} = 14,477 \quad (8)$$

However, the population will not grow at a rate of 1.01 every one of these 500 years. Each year, it will grow either at a rate of 1.16 or 0.86. If  $\lambda = 1.16$  in exactly 250 years, and 0.86 in the other 250, which is quite probable, the population would in fact have only 54 individuals 500 years from now, a huge difference from the calculation in Eq. 8:

$$100 \cdot (1.16)^{250} \cdot (0.86)^{250} = 54.8 \quad (9)$$

Of course other outcomes are possible in this probabilistic scenario, but this one is the most likely. It is no accident that the computation that accounted for variation in  $\lambda$  predicted a smaller population than the computation using the mean; incorporating stochasticity into models of population growth makes it likely that populations will do worse than they would in a deterministic model (Morris and Doak 2002).

In the preceding example, the simple average of the two values of  $\lambda$ , 1.01, generated a very poor (and wildly overoptimistic) prediction of the population's future dynamics. This simple average (the sum of  $n$  values divided by  $n$ ) is also known as the arithmetic mean. A less-familiar mean is the geometric mean (the  $n$ th root of the product of  $n$  values). The geometric mean of 1.16 and 0.86 is 0.9988, and using it instead of the arithmetic mean generates a more accurate prediction of the population's growth rate in the face of environmental stochasticity:

$$100 \cdot (0.9988)^{500} = 54.8 \quad (10)$$

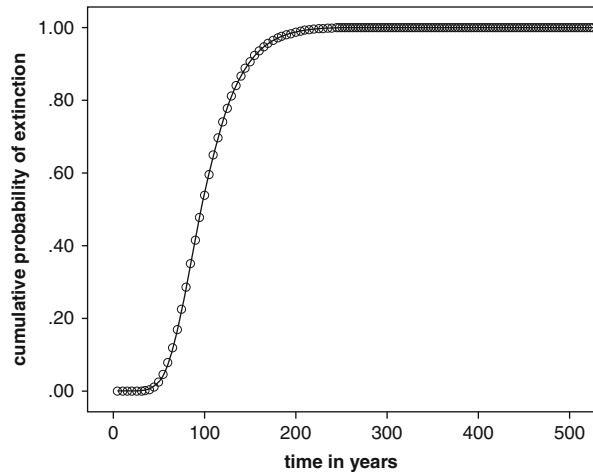
The geometric mean of a series of numbers is always less than or equal to the arithmetic mean. That the geometric mean  $\lambda$  yields a more accurate population prediction should make sense, given that population growth is a multiplicative process.

As this example illustrates, a population experiencing temporal variability in vital rates might decline over time, even if in some years its growth rate,  $\lambda$ , is well above 1.0. This fact has important implications for the persistence of species of conservation concern. Using information about the amount of temporal variability a population experiences, a prediction can be made about the likelihood that a population will persist or go extinct within some specified time frame. Such information can also be used to identify effective management options. These investigations use a variety of modeling approaches collectively known as population viability analysis (PVA).

Over the last several decades, the development of models to assess the extinction risk of threatened or endangered populations has been one of the most active areas of research in plant (and animal) population dynamics. Morris et al. (1999) is an excellent introduction to some of the most commonly used PVA approaches, and Morris and Doak (2002) provide further elaboration; Brigham and Thomson (2003) provide a good, brief overview. PVA models allow  $\lambda$  to vary over a range of values from year to year, with that range representing the degree of environmental variation a population experiences. Such models cannot forecast the future size of the population with certainty; instead, they aim to forecast the probability that a population will achieve a particular size (or become effectively extinct) by some specified future time. The greater the interannual variability in population growth rates, the greater the uncertainty associated with these forecasts.

To illustrate this approach, some of the data in Fig. 2 for the Hawaiian silversword are analyzed here using the simple PVA for "count data" (i.e., unstructured data) presented in Morris et al. (1999). The data come from 11 permanent plots that were established on Mt. Haleakala in 1982 to permit long-term monitoring of the silversword population. All individuals in the plots were censused in 23 of the years between 1982 and 2010 (Krushelnicky et al. 2013). The population in Fig. 2 shown by the closed squares has fluctuated in numbers over the census period and since 2000 has appeared to be declining. What are the survival prospects for this population if current trends continue? The first step in performing a count-based PVA is to estimate values of  $\mu$ , which is a stochastic version of the log of the population growth rate (see

**Fig. 17** The cumulative distribution function of extinction probabilities for the Hawaiian silversword population represented by the closed squares in Fig. 2. The population appears likely to be extinct within 200 years if current trends continue



Morris and Doak 2002 for details), and of  $\sigma^2$ , a measure of the stochastic variance in  $\mu$ . Morris et al. (1999) and Morris and Doak (2002) provide formulas for computing these parameters. Following their procedure yields a value for  $\mu$  of  $-.001$ . The fact that  $\mu$  is negative means that the population will certainly go extinct; this is a reasonable expectation given the population trend evident in Fig. 2. But how much time will elapse before extinction occurs? To determine the likely time frame for this event, the cumulative distribution function (CDF) of extinction probabilities can be estimated (code for this computation is available in the R package “popbio”). To estimate a CDF, it is important to define an extinction “threshold,” i.e., a number of individuals below which the population becomes effectively extinct. In this example, that threshold has been set to four individuals. The resulting CDF, shown in Fig. 17, illustrates that without active management of some kind, this population of Hawaiian silverswords is likely to be extinct within 200 years.

## Incorporating Population Structure into Models and Analyses

Even these more complex models incorporating stochastic variation described in the previous section are relatively simple in that they are unstructured. They track total population numbers, treating all individuals as making the same contribution to population growth, ignoring the fact that individuals can vary with respect to the demographic parameters introduced in section “[Structure of Plant Populations.](#)” Structured models of population dynamics take a different approach, tracking the vital rates of different age, stage, or size classes separately and making predictions not only about how the size of an entire population might change under different assumptions, but also about how the abundances of each class are expected to change. A great deal of research in plant population dynamics over the last several decades has made use of these models.

**Table 1** A life table for the grass *Poa annua*, data from Law (1975), table adapted from Begon et al. (1996)

x, age (in 3-month periods in this example)	$a_x$ , number of individuals that live to age x	$l_x$ , proportion surviving to age x	$m_x$ , mean number of seeds produced by an individual while age x
0	843	1.0	0
1	722	0.856	300
2	527	0.625	620
3	316	0.375	430
4	144	0.171	210
5	54	0.064	60
6	15	0.018	30
7	3	0.004	10
8	0	0	–

Structured models are based on the notion of a *life table*, a convenient way to summarize demographic information for age-structured populations. First developed for human populations, life tables contain information on how probabilities of survival and reproduction vary with an individual's age. A life table summarizes data collected during repeated regular censuses of a cohort, which is a group of individuals all born at the same point in time. This information can then be used to calculate the cohort's (and, by extension, the population's) rate of increase.

Each age is represented as a separate row in a life table (see Table 1), and information on the survival and fecundity for each age is organized as a series of columns. The first column of a life table contains the ages ( $x$ ) of individuals in the cohort, with  $x = 0$  representing the age of a newborn individual. (Because seeds are so hard to observe, "birth" in plant life tables is often defined as the appearance of a seedling.) While censuses are often conducted annually for organisms in seasonal environments, census intervals may be chosen to be shorter (as in Table 1) or longer than a year, depending on the life history of the organism. The life table here is for an annual grass, *Poa annua*, and censuses were carried out every three months.

At each census, the numbers of survivors of the cohort are counted. These data are presented in the second column ( $a_x$ ). The original number of individuals in the cohort, 843 in this example, is  $a_0$ . These values can be used to compute each age class's age-specific survivorship,  $l_x$  ( $a_x/a_0$ ), which is the proportion of the original cohort that lives at least until age  $x$ . Age-specific fecundity,  $m_x$ , is typically quantified as the mean number of seeds (or seedlings) produced per individual while it is age  $x$ .

The symbols used to represent these different vital rates are unfortunately not standardized; some authors use  $N_x$  in place of  $a_x$  or  $B_x$  in place of  $m_x$ . Likewise, survivorship ( $l_x$ ) is sometimes represented as the proportion of a cohort still alive, as is the case here, and other times as a standardized number of survivors from a hypothetical original cohort of 1,000. It is also worth noting that for organisms

with separate sexes, life tables are based on the number of *female* offspring produced by a typical *female*, since the population growth rate in such species is typically determined by the rate at which females reproduce. Since most plant species are hermaphroditic, life tables for most plants need not make this distinction.

The data in a life table can be used to compute the cohort's net reproductive rate,  $R_0$ , the average number of offspring that a typical individual produces over its lifetime, i.e., per generation. The formula for  $R_0$  is

$$R_0 = \sum_{x=0}^k l_x m_x \quad (11)$$

where  $k$  is the final age used in the life table. Note that  $R_0$  differs from a simple sum of the numbers of offspring produced at each age; it weights each reproductive episode by the likelihood that an individual will live to that age. The units of  $R_0$  are the expected numbers of offspring produced per newborn individual per generation. In order to convert  $R_0$  to  $\lambda$  or to  $r$ , the generation time,  $G$ , must be computed, as follows:

$$G = \frac{\sum_{x=0}^k l_x m_x x}{\sum_{x=0}^k l_x m_x} \quad (12)$$

The relationship between  $R_0$  and  $\lambda$  is then

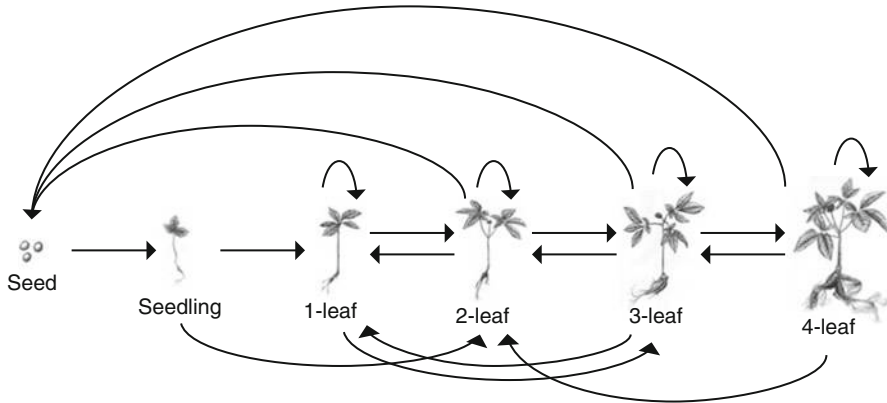
$$\lambda = (R_0)^{\frac{1}{G}} \quad (13)$$

while the relationship between  $R_0$  and  $r$  is

$$r = \frac{\ln(R_0)}{G} \quad (14)$$

It is important to note that life tables, like the simplest unstructured models presented in section “[Causes of Different Temporal Patterns of Plant Population Dynamics](#),” assume that an individual's fecundity depends only on its age and is not affected by population density. Thus a life table is implicitly a geometric growth model. In that sense, it can accurately compute a population's current reproductive rate, but it might do a poor job of forecasting future reproduction. Secondly, because in many plants the correlation between age and size is not very strong (Gurevitch et al. 2002), life tables are not appropriate tools for the study of many plant populations; they are probably most appropriately applied to annual species, as in Table 1. However, they provide a useful introduction to other kinds of structured models.

Structured models of most plant species tend to use size classes rather than age classes. The use of size classes introduces some complications into the modeling process. In a life table, in which individuals are classified by their age, an individual can have only two possible fates between successive censuses: it may



**Fig. 18** Life-cycle diagram for *Panax quinquefolius*, American ginseng. In this scheme, individuals are divided into six possible size classes. Information from annual censuses allows researchers to estimate the probability of each of the “transitions” represented by the arrows (Reprinted from Farrington et al. 2008)

move into the next age class, or it may die. When individuals are classified by size rather than age, there are more possibilities. Between censuses, an individual may (a) move from a smaller size class to one or more larger size classes (“growth”), (b) move from a larger class to one or more smaller classes (“regression”), (c) remain in the same class (“stasis”), or (d) die. These complex possibilities are often displayed in the form of a *life-cycle diagram*. Figure 18 shows a life-cycle diagram for American ginseng, *Panax quinquefolius*, an herbaceous perennial. The arrows represent the possible changes that individual plants can undergo between successive censuses, as well as the fact that plants having at least two leaves can also produce seeds. Individuals that die between censuses are not shown in the diagram.

To accommodate these complications, plant ecologists generally model a structured population’s dynamics with size-structured transition matrix models, also known as Leslie matrix models, Lefkovitch models, or simply matrix models. A “transition” is a period of time between successive population censuses, during which individuals in the population may undergo changes in their status, like those in Fig. 18. These models represent the population’s status changes during each of these transitions as a matrix of vital rates (Fig. 19). Each vital rate is estimated from annual censuses of individually marked plants. A transition matrix is square (i.e., it has equal numbers of rows and columns). There are as many rows and columns as there are size classes. Each entry in the matrix has two subscripts: the first ( $i$ ) representing its row (i.e., the class it has transitioned to) and the second ( $j$ ) representing its column (i.e., the class it has transitioned from). Each entry in the matrix,  $a_{ij}$ , represents the proportion of individuals originally present in class  $j$  that transitioned to class  $i$  between the first and second census.

**Fig. 19** A generic size-classified transition matrix model for a species with four size classes, not yet parameterized with data

	From class (at time t):			
	1	2	3	4
1	$a_{11}$	$a_{12}$	$a_{13}$	$a_{14}$
2	$a_{21}$	$a_{22}$	$a_{23}$	$a_{24}$
3	$a_{31}$	$a_{32}$	$a_{33}$	$a_{34}$
4	$a_{41}$	$a_{42}$	$a_{43}$	$a_{44}$

$$\begin{matrix}
 \mathbf{M} & \times & \mathbf{n}_t & = & & \mathbf{n}_{t+1} \\
 \begin{bmatrix} 0 & 0 & .125 \\ .601 & .091 & 0 \\ .011 & .633 & .82 \end{bmatrix} & \times & \begin{bmatrix} 15 \\ 30 \\ 100 \end{bmatrix} & = & \begin{bmatrix} (0 \times 15) & (0 \times 30) & (.125 \times 100) \\ (.601 \times 15) & (.091 \times 30) & (0 \times 100) \\ (.011 \times 15) & (.633 \times 30) & (.82 \times 100) \end{bmatrix} & = & \begin{bmatrix} 12.5 \\ 11.745 \\ 101.155 \end{bmatrix}
 \end{matrix}$$

**Fig. 20** This example represents a population divided into three size classes. At time t there were 15 class-1 individuals, 30 class-2 individuals, and 100 class-3 individuals. Multiplication of the matrix **M** by the vector **n<sub>t</sub>** as shown produces a new vector, **n<sub>t+1</sub>**, of 12.5 class-1 individuals, 11.745 class-2 individuals, and 101.155 class-3 individuals (since a fractional individual cannot exist, these are often rounded to the nearest whole number)

Though a transition matrix does not explicitly include survival/mortality rates for each size class, the proportion of individuals in class j experiencing mortality between the two censuses can be calculated as

$$1 - \sum_{i=1}^{i=k} a_{ij} \tag{15}$$

Conventionally, the first class in a transition matrix represents newborn individuals (i.e., individuals present at the second census that were not present at the first), so the entries in the top row of the matrix are zero until reproduction has been incorporated. The reproductive contribution of class j is defined as the mean number of class-1 individuals present at time t + 1 that were produced between the first and second censuses by individuals in class j at time t. Morris et al. (1999) and Morris and Doak (2002) provide clear accounts of how to construct a transition matrix from census data.

Figure 20 shows an example of a matrix (**M**) for a hypothetical plant population in which individuals can belong to any of three size classes. In this example, these transitions are possible: class-1 individuals can grow to class 2 or to class 3 or die; class-2 individuals can stay in class 2, grow to class 3, or die; and class-3 individuals can stay in class 3 or die. Only class-3 individuals can reproduce. Figure 20 also shows two vectors (columns of numbers). These vectors represent the population’s *size structure*, i.e., the numbers of individuals present in each size class at some particular census period. The sum of these numbers equals n<sub>t</sub>, the total number of individuals in the population at time t.

Matrix models place vital rate data into a matrix format so that the operations of matrix algebra can be used to project the population’s size structure into the future,



given particular assumptions. When a transition matrix is multiplied by a vector that represents a population's current size structure, the resulting vector gives the population's size structure 1 year in the future. (Figure 20 shows how matrix multiplication is carried out.) Repeated multiplication of the matrix by the resulting vector (using mathematical software such as MATLAB or Mathematica) can project the population any number of years into the future. Iterative multiplication eventually yields a population size structure that is stable, in the sense that the proportion of the population in each size class does not change, even as the total population size continues to grow (or shrink). The dominant eigenvalue of the matrix, which can be easily computed with mathematical software, is equivalent to  $\lambda$ , the population's rate of increase, the rate at which the population size will change once it has achieved its stable size structure. This one parameter,  $\lambda$ , integrates multiple vital rates into a single metric.

Because  $\lambda$  indicates whether a population is stable, increasing, or declining, it provides important basic information about a population's status. Matrix models also allow researchers to determine other important information about a species. Through approaches known as sensitivity and elasticity analyses and life table response experiments (Caswell 2001), the contribution of individual vital rates or of particular matrix entries to the overall population growth rate can be assessed. These analyses allow researchers to explore the specific mechanisms underlying observed variation in  $\lambda$  over time or between different populations. More complex versions of these models can be created to incorporate the production of vegetative propagules, seed dormancy, and other life history variations.

But the growth in the use of matrix models since their introduction in the early 1970s is due particularly to their usefulness for guiding management (Crone et al. 2011). For the last several decades, conservation biologists have studied the population dynamics of plant species of conservation concern to better document the status of sensitive species of plants, to quantify extinction risk, to understand the causes of population declines, to explore possible ways to reverse those declines, and to assess the effects of possible changes in management or environmental conditions. For those charged with managing these species, managing invasive species, or setting guidelines for sustainable harvesting,  $\lambda$  provides important information about population status.

Furthermore, matrix models can allow a researcher to model the potential long-term effects of events that a natural or managed population might experience, such as herbivory, harvesting, controlled burning, etc. This can be done in a variety of ways. A sensitivity analysis allows ecologists to evaluate the effectiveness of management alternatives that are expected to alter particular elements in a matrix. Alternatively, potential management approaches can be simulated by repeatedly multiplying alternative matrices, representing different environmental states, in different orders (see the example of *Hudsonia montana* described below). Such information can help managers decide whether a particular harvesting rate is sustainable or how frequently to mow or burn a meadow or grassland they are managing for a sensitive species. For example, American ginseng is a plant that is harvested as a medicinal herb; its market value makes it a tempting target for illegal

overharvesting. Farrington et al. (2008) modified vital rates in a matrix model to investigate how different levels of harvesting, in association with browsing by deer, influenced ginseng's population growth rate.

For all of the reasons described above, matrix models have become the primary analytical tool for studying plant population dynamics; by 2009, well over 300 such studies had been published (Crone et al. 2011), and their numbers continue to grow. However, some caveats about the use of matrix models are in order. One of the assumptions of the basic transition matrix model is that the population's vital rates as represented in the matrix will remain constant over the time frame over which  $\lambda$  is being projected. However, vital rates are not fixed; they vary from 1 year to the next, as a consequence of stochastic environmental variation. Two censuses – one transition – cannot capture the full range of environmental variation that a population experiences. Ecologists have invested considerable effort in developing ways to incorporate this year-to-year variation in vital rates into matrix models.

There are two general approaches for incorporating environmental variability into matrix models; both require census data from multiple years. The first approach is to construct a series of transition matrices, one for each pair of censuses. Then,  $\lambda$  is computed by computer simulation, by drawing individual matrices at random (with replacement) from the pool of those available. The second approach is to represent each vital rate in the matrix as a random variable capable of taking on a range of possible values (determined using census data from multiple years) and then to use computer simulation to create a unique matrix from these ranges of allowable values for each time step in the simulation. In both approaches, because the sequence of matrices used will affect the value of  $\lambda$ , researchers compute the mean and variance of  $\lambda$  from a large number of simulations (1,000 or more). These approaches thus also provide researchers with important information about the uncertainty associated with their estimates of  $\lambda$ . Both approaches to incorporating temporal variability in vital rates have strengths and weaknesses and many variations (see Morris and Doak 2002).

A good example of the utility of the matrix model approach for the management of threatened species is provided by a study of mountain golden heather, *Hudsonia montana*, a threatened shrub from North Carolina, USA (Gross et al. 1998). Once thought to be extinct, *H. montana* was rediscovered in 1979. The reasons for its low numbers were hypothesized to be either competition from other plants as a result of fire suppression and/or trampling by hikers and campers. Gross et al. (1998) used matrix modeling to address these questions about *H. montana*: How can recovery be achieved? Would protection from trampling be sufficient to permit recovery? Can the implementation of controlled burns achieve recovery? Must both strategies be implemented? If controlled burns are important, given their high cost, what is the least frequent burn interval that can achieve a positive population growth rate? Gross et al.'s (1998) study used census data on *H. montana* collected over 5 years from an unmanipulated population as well as from one subjected to a controlled burn. Observations of the reasons for each observed mortality event allowed the quantification of trampling-caused mortality. Multiple censuses provided Gross

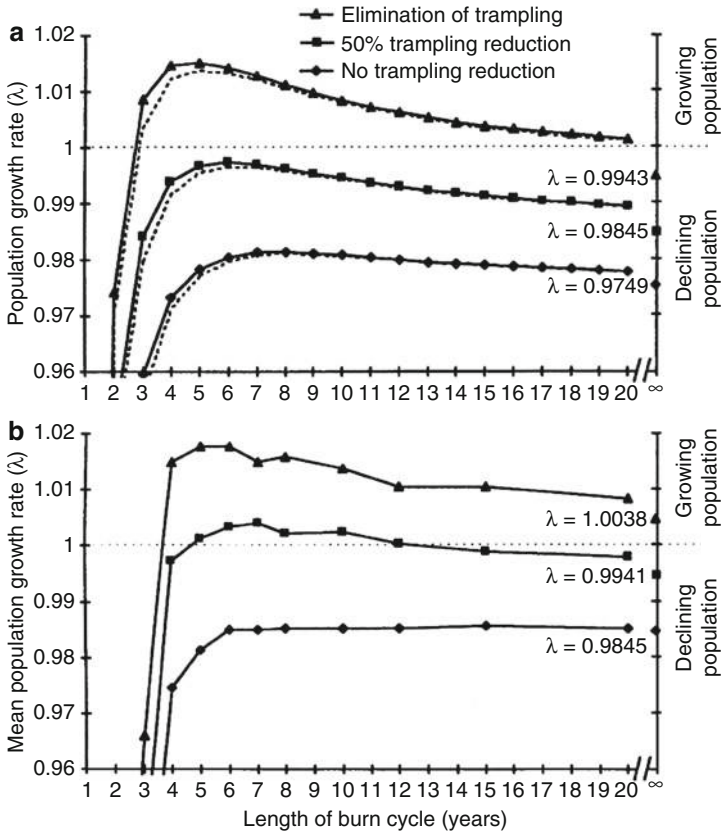
et al. (1998) with data on vital rates in the burned population during the year of the burn as well as 1 and 2 years afterward.

Gross et al. (1998) performed both a deterministic analysis as well as one that incorporated stochastic variability by treating each vital rate as a random variable. In the deterministic analysis, they created different matrices that represented populations subject to one of three levels of trampling (no reduction from current levels and 50 % and 100 % reductions of trampling mortality) in non-burn, burn, and postburn years. By multiplying different matrices together, they created product matrices that simulated a variety of burn scenarios (e.g., burning every other year, every 5th year, every 10th year) in combination with any of the three trampling scenarios and computed  $\lambda$  for each one. In their stochastic analysis, they explored 39 different management strategies, consisting of the three different trampling levels combined with 13 different burn scenarios, ranging from no burning to control burns carried out at intervals of between 1 and 20 years.

The study's results demonstrated that neither management strategy by itself was sufficient to reverse the decline of *H. montana* (Fig. 21). However, they found that population growth ( $\lambda > 1$ ) was possible if burning was combined with the elimination of some or all of the trampling. While one burn every 6–8 years was predicted to maximize *H. montana*'s growth rate, Gross et al. (1998) found that decreasing the burn frequency to as much as once every 12–16 years would still allow the numbers of this threatened plant to increase. The stochastic analysis produced a somewhat more optimistic outlook (compare Fig. 21) than the deterministic one. This finding runs counter to the idea described in section “[The Role of Stochastic Influences, Especially in Small Populations](#)” that incorporating environmental variability often leads to forecasts of slower population growth. This result could be due to the nature of the variability in this particular example or to negative correlations in the variability of different vital rates (Doak et al. 2005).

Gross et al. (1998) asked what strategies would be effective in reversing *H. montana*'s observed decline. The same data can be used to carry out a PVA. The goal of such an analysis is to forecast the probability of extinction if no management were implemented. Morris and Doak (2002) reanalyzed Gross et al.'s (1998) data to produce such a forecast. Incorporating environmental variability by using a matrix-selection approach, Morris and Doak (2002) computed the cumulative probability of extinction (which they defined as the population's falling below 500 individuals, since most of the “individuals” are dormant seeds in the soil) as a function of time. They found that, in the absence of any management action, the population has nearly a 50 % probability of extinction within 50 years (Fig. 22). Methods for these and other analyses using matrix models can be found in Caswell (2001) and in Morris and Doak (2002), and code for carrying them out is available in the R package “popbio.”

The incorporation of environmental variability is not the only important concern when using matrix models. Another assumption of matrix models is that the population has attained a stable size distribution. Until this occurs, the actual population growth rate can be quite different and either larger than or smaller than  $\lambda$ . A population in a highly variable environment may not have the opportunity

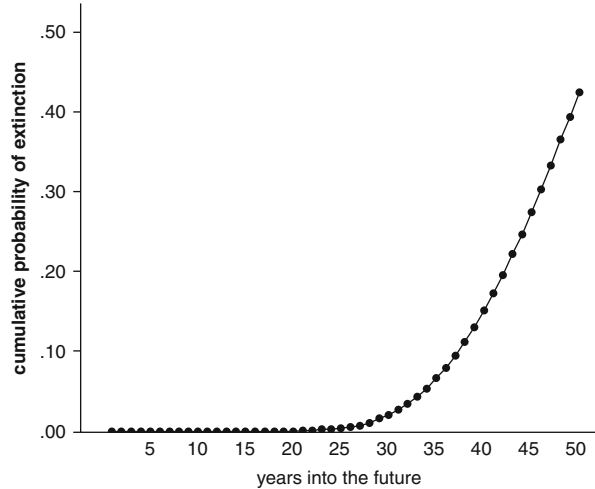


**Fig. 21** The annual population growth rate,  $\lambda$ , of *Hudsonia montana* as a function of simulated burn cycle length and level of trampling reduction for deterministic (a) and stochastic (b) transition matrix models. Dashed lines in a represent the population growth expected while the population achieves a stable size distribution

to achieve a stable size distribution, in which case the  $\lambda$  generated by a matrix model may provide a poor forecast of population behavior. However, Williams et al. (2011) surveyed data from 46 plant species and found that most were near their stable size distributions. For populations that are not, methods of transient analyses (references in Williams et al. 2011) can be used to arrive at forecasts of population growth rates.

Thirdly, it is important to recognize that while these models are structured, they are variations of the simple geometric growth model first presented in section “Causes of Different Temporal Patterns of Plant Population Dynamics,” in which  $\lambda$  is assumed to be independent of population density. In that section, it was acknowledged that the geometric model is quite unrealistic. However, measured values of  $\lambda$  for plant populations tend to center around 1.0 (Crone et al. 2013), which implies that most of the populations analyzed using these models are not

**Fig. 22** The cumulative probability of extinction for Gross et al.'s (1998) population of *Hudsonia montana*, in the absence of any management, as analyzed by Morris and Doak (2002) (Figure redrawn from Morris and Doak 2002)



changing in size very rapidly; therefore, the geometric model may often be an appropriate one. For species of conservation concern, whose population sizes are by definition well below  $K$ , the assumption of a lack of density dependence is certainly appropriate, justifying the widespread use of these models for this purpose. Nevertheless, it is clear that there are some kinds of plant populations for which this density-independent approach is unsuitable. For this reason, density-dependent versions of matrix models have been developed (Caswell 2001; Morris and Doak 2002).

The widespread use of matrix models, coupled with an appreciation of their limitations/assumptions, has raised questions about their value and applicability. Crone et al. (2013) used long-term data from 20 plant species to compare the forecasts of matrix models for these species with their observed population dynamics. They concluded that matrix models provided a good integration of a population's vital rates during the time period during which those vital rates had been estimated and that  $\lambda$  was indeed a suitable way to assess a population's status and to evaluate management options. However, they found that in many instances, matrix models failed to accurately forecast future population sizes. In evaluating the possible causes of this failure, Crone et al. (2013) ruled out density dependence and shortcomings in the number of sampled plants or census years, two often-cited concerns about matrix models.

Instead, they concluded that the most plausible explanation for why matrix models sometimes fail to accurately forecast future population behavior is that the assumption of environmental constancy (even allowing for stochastic variation about some mean) is not met (Crone et al. 2013). Especially in the face of the environmental changes in temperature and precipitation currently occurring as a result of anthropogenically increased levels of atmospheric  $\text{CO}_2$ , it is clearly desirable to develop ways to incorporate the likely effects of directionally changing environmental parameters into models of population dynamics.

Krushelnycky et al. (2013) took such an approach to try to understand the reasons why, after such a successful population recovery, the Hawaiian silversword population has once again begun to decline. Since climate data indicated that conditions on the volcano had become drier and warmer over time, they investigated this possible cause for the declining population growth rate by modeling the dependence of annual values of  $\lambda$  on various measures of rainfall and temperature. They found that  $\lambda$  was positively correlated with the number of wet season days having  $>10$  mm of rainfall and negatively correlated with the number of rainless days during the dry season. However,  $\lambda$  was also negatively correlated with the number of rainy season days where rainfall exceeded 15 mm. These associations explained 64 % of the observed variation in  $\lambda$ . Population growth rate did not depend significantly on temperature. These results suggested that changes in rainfall patterns are affecting the persistence of the silversword population, though not in a straightforward way. The authors concluded that the view of the Haleakala silversword as “secure” is no longer justified, now that global climate change has begun to significantly affect rainfall patterns on the volcano. Despite successful efforts to address earlier threats of vandalism and grazing, it now seems that the silversword has a bigger problem, one not so easily solved by building fences or educating visitors; climate change appears to be causing most of these high-altitude populations to decline (Krushelnycky et al. 2013).

Increasingly, plant ecologists are looking for ways to incorporate the role of changing environmental factors into their analyses of past population dynamics, as in the above example, as well as into forecasts of future dynamics. For example, Salguero-Gomez et al. (2012) used structured demographic models, coupled with high-resolution climatic models projecting future global changes in temperature and precipitation, to assess how these climatic changes would be likely to affect two species of desert plants, one from Utah in southwestern North America and one from Israel’s Negev Desert. Their surprising result was that projected changes in precipitation in these regions (increases in Utah, decreases in Israel) were expected to lead to increased population growth for both plant species (Salguero-Gomez et al. 2012).

---

## Spatial Patterns of Population Dynamics

Up until now, the emphasis in this chapter has been on how births and deaths contribute to changes in the size or density of plant populations. Immigration and emigration, though included in Eq. 2, have been ignored. But just as population densities vary in time, they can also vary spatially. Ecologists are discovering that this spatial variation is fluid rather than static. They are asking questions about what determines these patterns and developing tools to study them.

The study of spatial patterns of population dynamics is driven in large part by the recognition that suitable habitat for many species is fragmented rather than contiguous. The fragmentary nature of suitable habitat is caused not only by natural physical phenomena (e.g., variation in parent material of soil, in elevation and hydrology, and the ephemeral nature of many habitats) but also, very importantly,

by human activities like urban and agricultural development, forest harvesting, etc. Such anthropogenic habitat fragmentation has been recognized as one of the greatest threats to species diversity. Regardless of the cause of patchiness, many plant species are distributed within discrete patches of suitable habitat embedded in an unsuitable habitat matrix; these patches can be connected by the dispersal of seeds and/or pollen. Understanding the persistence of species in fragmented habitats often requires adopting a spatial perspective that includes more than a single local population.

Regional assemblages of populations of the same species can take many forms. The best-studied type of regional population assemblage is the *metapopulation*. A metapopulation is a network of relatively small, local subpopulations connected by migration. Because of their small size, individual subpopulations within the larger metapopulation are prone to local extinction. Metapopulation theory has led to the conclusion that in order for a metapopulation to persist over the long term, there must be asynchronous, reciprocal dispersal between existing subpopulations and from existing subpopulations to unoccupied patches of suitable habitat and that the density of suitable habitat patches must exceed some threshold (Freckleton and Watkinson 2002). The dynamics of the entire metapopulation are determined by these processes of extinction, dispersal, and recolonization and thus are not a simple function of the collective dynamics of local populations (Freckleton and Watkinson 2002). Likewise, the dynamics of local populations that are part of a metapopulation cannot be completely understood without adopting a metapopulation perspective.

While metapopulation theory has had a strong influence on how animal populations are studied, there are limited numbers of studies of plant populations that take a metapopulation perspective, in part because the existence of seed dormancy in many plant species complicates the quantification of extinction rates (Husband and Barrett, 1996) and also because it is difficult to recognize what constitutes a suitable habitat patch when it is unoccupied (Freckleton and Watkinson 2002). Another way in which regional assemblages of plant populations may differ from those of animals is that plants and their propagules are often very long-lived, and their dispersal abilities are more limited than those of animals; thus processes such as extinction and colonization may take place on much longer time scales. Consequently, few studies have attempted to measure colonization, extinction, and recolonization rates and the density of suitable habitat patches for regional assemblages of plant species (Freckleton and Watkinson 2002; Ouborg and Eriksson 2004). In fact, the very applicability of the metapopulation concept to plant species continues to be the topic of vigorous debate (Husband and Barrett, 1996; Freckleton and Watkinson 2002; Ouborg and Eriksson 2004).

Determining whether a particular plant species has a true metapopulation structure is more than an academic concern; it has important implications for how species conservation should be approached. For species that exist as metapopulations, it is inevitable that local populations will go extinct, so conservation efforts must not only protect existing subpopulations; they must also protect unoccupied but suitable habitat and conserve dispersal opportunities (e.g., through the creation of corridors). This is not necessary when local processes dominate spatial dynamics. In addition to metapopulations, ecologists recognize other kinds



of regional population assemblages. For example, some species occupy networks of habitat patches in which dispersal is primarily one way. Such networks (termed “source-sink” or “mainland-island” models) can persist if there are one or more source populations (where reproduction rates typically exceed mortality rates) that periodically provide emigrants to sink populations (where mortality rates typically exceed reproductive rates). In other species, different subpopulations may be so isolated from one another that the subpopulations are more or less unconnected and regional-scale spatial dynamics are governed almost completely by the dynamics of local populations. Finally, there are species that do not occupy distinct habitat patches, but exist as spatially distinct subpopulations within an essentially continuous habitat; spatial dynamics in this case are also governed largely by local processes (Freckleton and Watkinson 2002; Ouborg and Eriksson 2004). Given the importance of understanding spatial population dynamics to ecology, evolution, and conservation, the study of metapopulations in particular and of spatial dynamics in general is and will continue to be an active area of ongoing research in plant ecology (Ouborg and Eriksson 2004).

---

## **A Brief Guide to Methodological Approaches Used in Field Studies of Plant Population Dynamics**

### **Defining the Boundaries of a Population**

A population is a group of individuals belonging to the same species. How do ecologists determine the boundaries of a population? Sometimes boundaries are obvious, e.g., when a plant species lives on an island or in a natural area surrounded by developed land. But other times, a population’s boundaries are not so obvious; in these situations, ecologists define the boundaries of a population somewhat arbitrarily. Knowledge of the typical dispersal distance of seeds, or of the flight distance of pollinators, can be helpful in defining boundaries. In practice, ecologists usually define boundaries as regions where a population’s density falls off. Unless the population of interest is assumed to be closed to immigration/emigration, such a loose definition does not usually present a problem. The concept of a population is, after all, a human construct.

Anyone who studies population dynamics must make choices about how many and which populations to include in their study. These might be a random sample of known populations, or populations might be chosen because of some factor of interest that is being investigated. Issues that arise in sampling from a set of possible populations are addressed by Morris and Doak (2002).

### **Censusing Populations**

In the beginning of this chapter, repeated censuses were described as being at the heart of studies of plant population dynamics. Of course, annual censuses must be made at approximately the same time each year. Some studies of population



dynamics (known as count-based studies) only require information about how the numbers of individuals in the population change over time. For these studies, it is not necessary to know how each individual plant's status changes temporally and thus marking plants individually is unnecessary. It is not even necessary to count the numbers of seeds in the soil, because such censuses are useful as long as they represent counts of a constant fraction of the population each year (Morris and Doak 2002). If a population is at or near a stable size distribution (see section “[Incorporating Population Structure into Models and Analyses](#)”), this assumption is likely to be met and seeds can be ignored. However, careful records do need to be kept about the location of population boundaries, so repeated counts can be made in the same area. Count data are the easiest data to acquire and are the kinds of data most often collected by land managers responsible for monitoring sensitive species. Analysis of these data is done by means of unstructured models (see section “[The Role of Stochastic Influences, Especially in Small Populations](#)”).

However, it is relatively easy to track changes in the status of individual plants over time and thus to go beyond count-based studies to incorporate information on a population's age or size structure and how it changes over time. [A video by plant ecologist James McGraw demonstrates some of these techniques using wild ginseng, *Panax quinquefolium*. <http://www.youtube.com/watch?v=u3CxPUR6cy4>.] These data can then be used to parameterize structured models (see section “[Incorporating Population Structure into Models and Analyses](#)”). Gathering such information typically requires marking each individual in the population (or a randomly chosen subset of individuals) with a unique number, usually by attaching numbered metal tags to the plants or inserting them into the ground nearby. A metal detector can be a useful tool for relocating buried tags. Alternatively, for very small plants, the corners of small sampling plots can be marked with nails and a pantograph, photograph, or other method used to locate and relocate particular individuals within the plot. However, rhizomatous plants and those whose position may be altered by burrowing animals or by frost heaving can move a surprising amount from 1 year to the next, making reliable re-identification difficult.

For structured population studies, decisions must be made about how to demarcate size classes or stages. This decision is partly based on convenience and feasibility, but it is also important to find a reasonable compromise between creating too few and too many classes. The more individuals in each class, the more accurately their vital rates can be estimated. But the wider the boundaries of the class, the more likely it is that the class will pool individuals of widely varying sizes, with divergent demographic fates. See Caswell (2001) and Morris and Doak (2002) for detailed advice about defining size class boundaries.

While most size classes are relatively easy to recognize, others are more problematic. Some perennial plants have underground corms or other perennating organs that, though alive, may remain dormant for one or more growing seasons. Distinguishing dormancy from mortality requires multiple census years. Accurately estimating individual fecundity can be difficult without repeated visits to a population at the time of seed production, and many species have seeds that remain dormant in the seed bank for anywhere from a few months to many years.

Sometimes experiments involving buried seeds are necessary to quantify seed dormancy and survival rates. Other species form vegetative propagules (e.g., cormlets) that can be dispersed and must be accounted for. Each species requires its own set of methodological decisions.

---

## Future Directions

Transition matrix models will continue to be an important way to study the dynamics of plant populations and to guide management decisions. Every year these models grow increasingly sophisticated (Salguero-Gomez and de Kroon 2010). Some of the newest developments include ways to represent networks of populations connected by dispersal, investigate the importance of ecological drivers of population dynamics, explore the transient dynamics of populations responding to changing conditions, and make better population forecasts in the face of temporal and spatial stochasticity.

Understanding the effects of climate change on plant population dynamics, in particular, is an area of high priority. Climate change is a long-term, uncontrolled experiment whose effects on population dynamics are of great scientific and practical importance. The large numbers of published studies making use of matrix models facilitate the asking of questions such as: can we make robust predictions about whether species in particular habitats or with particular life histories are more or less vulnerable to the effects of stochasticity or climate change than others?

In the study of population dynamics in general, advances in molecular technologies are making it possible to identify and quantify soil microorganisms, permitting researchers to begin to explore how interactions with soil biota determine plant population dynamics (Bever et al. 1997). And there are growing links between the study of population dynamics and other biological subdisciplines, such as community ecology, ecosystem ecology, and ecophysiology, with the goal of providing a greater mechanistic understanding of the processes underlying population dynamics and a better understanding of large-scale ecological processes.

---

## References

- Antonovics J, Levin DA. The ecological and genetic consequences of density-dependent regulation in plants. *Annu Rev Ecol Systemat.* 1980;11:411–52.
- Begon M, Mortimer M, Thompson DJ. *Population ecology, a unified study of animals and plants.* 3rd ed. Oxford: Blackwell; 1996.
- Beschta RL. Cottonwoods, elk, and wolves in the Lamar Valley of Yellowstone National Park. *Ecol Appl.* 2003;13(5):1295–309.
- Bever JD, Westover KM, Antonovics J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J Ecol.* 1997;85:561–73.
- Brigham CA, Thomson DM. Approaches to modeling population viability in plants: an overview. In: Brigham CA, Schwartz MW, editors. *Population viability in plants.* Berlin: Springer; 2003. p. 145–71.

- Caswell H. Matrix population models: construction, analysis and interpretation. 2nd ed. Sunderland: Sinauer; 2001.
- Chapin FS, McGraw JB, Shaver GR. Competition causes regular spacing of alder in Alaskan shrub tundra. *Oecologia*. 1989;79:412–6.
- Crawley MJ. Insect herbivores and plant population dynamics. *Annu Rev Entomol*. 1989;34:531–64.
- Crone EE, Menges ES, Ellis MM, Bell T, Bierzychudek P, Ehrlen J, Kaye TN, Knight TM, Lesica P, Morris WF, Oostermeijer G, Quintana-Ascencio PF, Stanley A, Ticktin T, Valverde T, Williams JL. How do plant ecologists use matrix population models? *Ecol Lett*. 2011;14:1–8.
- Crone EE, Ellis MM, Morris WF, Stanley A, Bell T, Bierzychudek P, Ehrlen J, Kaye TN, Knight TM, Lesica P, Oostermeijer G, Quintana-Ascencio PF, Ticktin T, Valverde T, Williams JL, Doak DF, Ganesan R, McEachern K, Thorpe AS, Menges ES. Ability of matrix models to explain the past and predict the future of plant populations. *Conserv Biol*. 2013;27(5):968–78.
- Doak DF, Morris WF, Pfister C, Kendall BE, Bruna EM. Correctly estimating how environmental stochasticity influences fitness and population growth. *Am Nat*. 2005;166(1):E14–21.
- Enright N, Ogden J. Applications of transition matrix models in forest dynamics: *Araucaria* in Papua New Guinea and *Nothofagus* in New Zealand. *Aust J Ecol*. 1979;4:3–23.
- Farrington SJ, Muzika R-M, Drees D, Knight TM. Interactive effects of harvest and deer herbivory on the population dynamics of American ginseng. *Conserv Biol*. 2008;23(3):719–28.
- Freckleton RP, Watkinson AR. Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *J Ecol*. 2002;90(3):419–34.
- Gross K, Lockwood Jr JR, Frost CC, Morris WF. Modeling controlled burning and trampling reduction for conservation of *Hudsonia montana*. *Conserv Biol*. 1998;12(6):1291–301.
- Gurevitch J, Scheiner SM, Fox GA. The ecology of plants. Sunderland: Sinauer; 2002.
- Hairston NG, Smith FE, Slobodkin LB. Community structure, population control and competition. *Am Nat*. 1960;94:421–5.
- Halpern SL, Underwood N. Approaches for testing herbivore effects on plant population dynamics. *J Appl Ecol*. 2006;43:922–9.
- Harper JL. A Darwinian approach to plant ecology. *J Ecol*. 1967;55:247–70.
- Harper JL. Population biology of plants. London: Academic; 1977.
- Husband BC, Barrett S. A metapopulation perspective in plant population biology. *J Ecol*. 1996;84:461–9.
- Hutchings MJ. The structure of plant populations. In: Crawley MJ, editor. *Plant ecology*. 2nd ed. Oxford: Blackwell; 1997. p. 325–58.
- Krushelnicky PD, Loope LL, Giambelluca TW, Starr F, Starr K, Drake DR, Taylor AD, Robichaux RH. Climate-associated population declines reverse recovery and threaten future of an iconic high-elevation plant. *Glob Chang Biol*. 2013;19:911–22.
- Morris WF, Doak DF. Quantitative conservation biology, theory and practice of population viability analysis. Sunderland: Sinauer; 2002.
- Morris WF, Doak D, Groom M, Kareiva P, Fieberg J, Gerber L, Murphy P, Thomson D. A practical handbook for population viability analysis. The Nature Conservancy; Washington, DC, 1999.
- Ouborg NJ, Eriksson O. Toward a metapopulation concept for plants. In: Hanski I, Gaggiotti OE, editors. *Ecology, genetics, and evolution of metapopulations*. Burlington: Elsevier Academic Press; 2004. p. 447–69.
- Salguero-Gomez R, de Kroon H. Matrix projection models meet variation in the real world. *J Ecol*. 2010;98:250–4.
- Salguero-Gomez R, Siewert W, Casper BB, Tielborger K. A demographic approach to study effects of climate change in desert plants. *Philos Trans R Soc B Biol Sci*. 2012;367:3100–14.
- Schemske DW, Bierzychudek P. Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. *Evolution*. 2001;55(7):1269–82.

- Silvertown J, Charlesworth D. Introduction to plant population biology. 4th ed. Oxford: Blackwell; 2001.
- U.S. Fish and Wildlife Service. Recovery plan for the Maui plant cluster. Portland: U.S. Fish and Wildlife Service; 1997.
- Watkinson AR. Plant population dynamics. In: Crawley MJ, editor. Plant ecology. 2nd ed. Oxford: Blackwell; 1997. p. 359–400.
- Williams JL, Ellis MM, Bricker MC, Brodie JF, Parsons EW. Distance to stable stage distribution in plant populations and implications for near-term population projections. *J Ecol.* 2011;99:1171–8.
- Winnie Jr JA. Predation risk, elk, and aspen: tests of a behaviorally mediated trophic cascade in the greater Yellowstone ecosystem. *Ecology.* 2012;93(12):2600–14.

## Further Reading

- Brigham CA, Schwartz MW, editors. Population viability in plants: conservation, management, and modeling of rare plants. Berlin: Springer; 2003.
- Caswell H. Matrix population models. 2nd ed. Sunderland: Sinauer; 2001.
- Gibson DJ. Methods in comparative plant population ecology. Oxford: Oxford University Press; 2002.
- Gotelli NJ. A primer of ecology. Sunderland: Sinauer; 2008.

Nathan J. B. Kraft and David D. Ackerly

## Contents

Introduction .....	68
A Brief History of the Development of Community Assembly Concepts .....	68
Dispersal .....	71
Abiotic Filtering .....	73
Biotic Interactions .....	75
Relationship Between Community Assembly and Coexistence Theory .....	76
Phylogenetic Patterns .....	77
Biogeography and the Build Up of Species Pools .....	80
Scale Dependence .....	83
Future Directions .....	84
Pattern-to-Process Mapping .....	84
Coexistence Theory and Community Assembly .....	85
Methods for Multitrophic Interactions .....	85
References .....	85

---

## Abstract

- Communities are located within a larger species pool of potential colonists. The study of community assembly considers the mechanisms by which local communities are formed from the species pool.
- Dispersal from the species pool, abiotic tolerance of colonists, and biotic interactions can all influence membership in local communities.
- Phenotypic similarities and differences of co-occurring species can be used (within limits) to make inferences about the role of alternative processes contributing to community assembly.

---

N.J.B. Kraft (✉)

Department of Biology, University of Maryland, College Park MD, USA

e-mail: [nkraft@umd.edu](mailto:nkraft@umd.edu)

D.D. Ackerly

Department of Integrative Biology, University of California, Berkeley, CA, USA

e-mail: [dackerly@berkeley.edu](mailto:dackerly@berkeley.edu)

- In many plant groups, close relatives tend to share similar phenotypic traits. Therefore, patterns of phylogenetic relatedness within a community can also be used to make inferences about community assembly mechanisms.
- As the community and the species pool can be defined at a number of different spatial and temporal scales, community assembly patterns often show strong scale dependence. In some cases, a single process can produce contrasting phenotypic patterns at different scales of analysis, while in other cases different processes may have stronger influences on community assembly at different scales.
- Species pools are shaped by dispersal of lineages among biogeographic regions, in situ speciation within regions, and extinction. The characteristics of the species pool often persist in local community patterns.
- Community assembly studies are often limited in the extent to which specific mechanisms can be inferred from community pattern. Future work should focus on improved models of competition and coexistence dynamics in community assembly as well as methods for considering multitrophic interactions.

---

## Introduction

Community assembly is the study of the processes that shape the identity and abundance of species within ecological communities. Central to most studies of community assembly is the concept of a species pool that is larger in geographic scope than the local community under study. The species pool contains potential colonists of the community, and many studies in this area focus on developing an understanding of the role of dispersal, responses to abiotic conditions, and biotic interactions in shaping local assemblages. Thus, community assembly considers both the ecological interactions that shape local communities and the evolutionary and biogeographic processes that lead to variation in the diversity and composition of species pools across the globe.

---

## A Brief History of the Development of Community Assembly Concepts

There are two persistent and central concepts in the study of community assembly. The first is the “species pool,” defined as the suite of possible colonists for a local site under study, and the second is the metaphor of a “filter” or a “sieve” that represents abiotic or biotic barriers to successful establishment at a local site. The two concepts can be traced back to two distinct sources: the study of species assemblages on oceanic islands and the study of succession following disturbance.

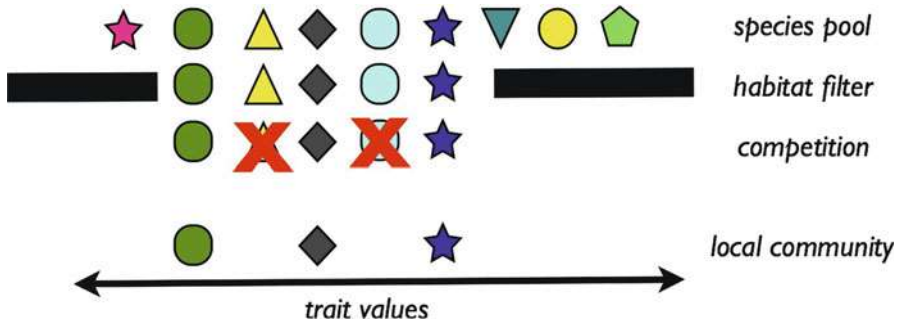
Perhaps the best-known precursor to community assembly theory is MacArthur and Wilson’s seminal theory of island biogeography, which describes the fate of an

oceanic island biota that is envisioned as receiving a supply of immigrants from a larger mainland species pool (MacArthur and Wilson 1967). The distance of the island from the mainland is predicted to influence the frequency with which new colonists arrive, and the size of the island influences the rate at which species go extinct on the island. Together these two properties predict the equilibrium number of species that the island will support at any point in time. Biotic interactions between species are implicit in island biogeography theory, as local extinction rates increase with species richness, though the primary focus of the theory is on the dynamics of dispersal to a community from a larger mainland species pool.

Following MacArthur and Wilson, the next step in the evolution of community assembly theory was Jared Diamond's study of bird communities on islands near New Guinea (Diamond 1975). Diamond was the first to use the concept of "assembly" in this context. In contrast to island biogeography, Diamond primarily focused on the role of biotic interactions in shaping local communities, and in particular he proposed "assembly rules" that captured the competitive exclusion of species that were too ecologically similar to co-occur. Diamond was criticized for lacking a proper null hypothesis for species differences when testing his assembly rules, as a null hypothesis is needed to permit the falsification of the hypothesis that competition shapes community assembly. If the process of competition is the only mechanism of community assembly that is considered, then there is no opportunity to allow for the role of other processes.

Shortly following the publication of Diamond's work, and at least in part in response to it, null models were developed that offer a solution to this issue (Pielou and Routledge 1976; Connor and Simberloff 1979; Strong et al. 1979; Colwell and Winkler 1984). Central to the null model concept is the idea of a species pool that is used to create a null hypothesis for the assembly of communities within a region. The null model captures much of the ecology of the system, but removes the key process of interest, such as competitive interactions, from the model. Thus, a null model for assembly rules might simply contain the dispersal of species from the species pools into local communities, without any consideration of competitive interactions. Random samples from the species pool can then be used to generate a distribution under the null hypothesis describing what species assemblages should look like in the absence of competition. Since the original application of null models to island data, the approach has been developed extensively (Gotelli and Graves 1996) and still remains central to many studies of community assembly.

In addition to the concept of a species pool, another central idea in community assembly theory is the concept of a "filter" that allows some species to pass through while serving as a barrier to unsuitable species as they arrive and attempt to establish at a site. This concept is first seen in the study of succession following disturbance, when Nobel and Slatyer (1977) describe an "environmental sieve" during succession. This concept was used extensively throughout the development of plant community assembly studies, often in terms of a "filter" that only permitted particular phenotypes to establish and persist (van der Valk 1981; Woodward and Diamond 1991).



**Fig. 1** Basic conceptual model of community assembly in terms of species functional traits (phenotypes). Empirically, an ecologist can consider the local community in relation to the species pool of potential colonists. Habitat filtering is often hypothesized to limit the range of traits that can successfully survive and establish at a site, as well as sometimes shifting the mean value relative to the species pool. Competition, in its earliest forms in community assembly theory, was predicted to favor the coexistence of species that differed in resource use or requirements, reflected here in their functional traits. In the example here, competitive exclusion leads to a local community of species with trait values that are more dissimilar than species in the original species pool. Additional community assembly processes are not shown. (After Woodward and Diament (1991))

While some early community assembly studies focused on forbidden combinations of species, much of the later research quickly transitioned to a focus on patterns of phenotypic traits of community members rather than on species identity *per se*. In plant community ecology, the focus has often been on functional traits, which are defined as aspects of the plant phenotype that are indicative of variation in ecological strategies of resource use, growth, and distribution in relation to environmental conditions (Westoby and Wright 2006). Plant functional traits that are relevant to community assembly can be anatomical or morphological traits, such as specific leaf area ( $\text{m}^2 \text{g}^{-1}$  biomass), root depth, or seed size, or they can be ecophysiological measures that reflect the integrated activities of several related plant processes, such as maximum photosynthesis rate or photosynthetic water use efficiency ( $\text{mol CO}_2$  assimilated  $\text{mol}^{-1}$  water transpired). As these functional traits can be measured on most if not all plants within and across communities, they offer a phenotypic common currency that can be used to draw generalizations across species and to make inferences about the mechanisms that shape community patterns. For example, in studies of succession, environmental sieves or filters are often hypothesized to drive convergence or clustering in phenotypic traits (relative to a null model), whereas competition patterns of the sort originally proposed by Diamond are typically predicted to produce phenotypic overdispersion, where co-occurring species are more dissimilar in traits than expected (Fig. 1; Weiher and Keddy 1999). More recent community assembly studies have refined these predictions in a number of ways, as discussed in subsequent sections.

A number of methodological considerations arise when sampling and analyzing functional traits in a community assembly context. Functional traits may vary



across modules (e.g., branches, leaves) within an individual plant, individuals within species, and across species. Some of this variation is environmentally driven plasticity, which is often correlated across species. For example, leaves produced in full sun environments often have lower specific leaf area (thicker or more dense tissue) than leaves from the same plant grown in shaded environments. Functional traits may also change throughout the development of an individual organ (e.g., a leaf or a stem) and through the ontogeny of whole plant. In sampling plant functional traits, the convention has often been attempt to minimize the role of ontogenetic and environmentally driven plastic variation among individuals within a species by standardizing trait sampling to particular environmental conditions and ontogenetic stages (Cornelissen et al. 2003). For example, many leaf traits are typically sampled on fully expanded and hardened leaves growing in the outer canopy of adult trees. Sampling in this way then emphasizes the role of genetic differences among individuals in driving any intraspecific variation in traits.

In many plant community assembly analyses, trait values among individuals within species are averaged, and analyses conducted on species trait means. This is justifiable if intraspecific variation is modest relative to interspecific variation in traits. However, in some communities, particularly those with low species richness, intraspecific variation can be substantial relative to interspecific variation, and as such there is growing interest in incorporating intraspecific trait variation into community assembly analyses (Violle et al. 2012). One of the important considerations in these analyses is whether the trait values measured on each individual area direct measure of that individual's growth and function, mediating interactions with the environment and with other individuals, or whether the traits represent proxies for underlying life history strategies of the species. In the latter case, it may be more appropriate to focus on species means rather than the particular manifestation of a trait in one individual or one local environment. When the data is available, it is also possible to determine from a quantitative standpoint the importance of intraspecific variation. For example, Cornwell and Ackerly (2009) evaluated the shift in community level mean trait values across a gradient of soil water availability and found that incorporation of intraspecific variability led to a steeper shift across the gradient, but the difference was fairly modest due to the greater role of interspecific turnover. In addition to averaging individuals within species, some community assembly studies take the additional step of grouping species with similar functional traits into functional groups or functional types, such as "C<sub>4</sub> grasses" or "broadleaf evergreen trees," which can further simplify analyses.

---

## Dispersal

Dispersal refers to the movement of an individual organism during its lifetime, from its place of birth to the location where it produces offspring. As plants are sessile organisms, in most species, dispersal only occurs once during the life cycle at the seed stage. Once a plant germinates, it occupies a single location for the rest of its life. In addition to this mechanism, a small number of species are able to disperse via

vegetative fragmentation, where disarticulated modules of a plant are able to initiate new roots after being transported to a new location. Dispersal is a key component of the community assembly process, as a plant must arrive at a location first before it can become a member of the community. The various mechanisms of propagule dispersal, together with the sessile habit, have many important consequences for plant population and community ecology. Four critical aspects of the dispersal process are considered here: dispersal mechanisms, dispersal distances, seed dormancy, and the role of dispersal limitation in shaping community assembly.

Seed dispersal is accomplished by a wide variety of different mechanisms, including gravity (i.e., large seeds that fall directly below the adult), ballistic mechanisms that eject a seed a short distance, floating on water, movement by wind, and dispersal by animals (either attached to the outside on fur or feathers or ingested and carried internally until deposited after a short time). The importance of these mechanisms for community ecology is that they are often undirected relative to the sites where a plant may be best suited to grow (unlike many animals, which can search for appropriate habitats). For example, many early successional plants are wind dispersed and produce numerous small seeds; this increases the likelihood that at least a few seeds will land in recently disturbed sites by chance, but wind dispersal will not generally be targeted at disturbed sites. Animal-mediated dispersal may be more directed, as when many seeds are deposited below perch trees where birds rest after eating. However, seeds are often dispersed more or less randomly with respect to the distribution of environments or communities where a particular species is most likely to germinate and successfully establish.

Most seeds travel only a short distance. In wind-dispersed species, tree height, seed size, and dispersal structures (wings, hairs, etc.) all influence dispersal distances, but even for tall trees with small seeds, most seeds travel less than 1 km. Thus, on short timescales, community assembly may be dispersal limited, in the sense that new species would arrive from external seed sources only infrequently and in small numbers. However, while most seeds travel short distances, plants have a remarkable ability to achieve rare, long-distance dispersal events. The best evidence for these events comes from remote oceanic islands, such as Hawaii, where the entire native flora is descended from hundreds of independent colonization events, in which seeds traveled thousands of miles across open water at some point over the past 5–10 million years. The recovery of Northern Hemisphere vegetation following widespread glaciation also demonstrates the importance of long-distance dispersal. Following past glacial epochs, species moved north in Europe and North America far faster than would be predicted based on the more common, short-distance seed dispersal from adult plants (Clark et al. 1998).

At the landscape scale, seed dormancy can be thought of as an important component of dispersal. Opportunities for germination and establishment may only occur infrequently, especially for species that colonize after disturbances such as wildfire or treefalls or in highly variable environments such as deserts where periods of sufficient rainfall for germination and establishment are sporadic and unpredictable. For these species, dormancy represents dispersal in time, allowing a seed to persist in a particular spot until suitable conditions occur.

Thus, while dispersal is undirected in space, the combination of dormancy and specific germination cues (discussed in the next section) allows some species to disperse in time so seedlings can occupy suitable environments for establishment and growth. Many examples have been documented of viable seeds germinating after hundreds or even thousands of years of dormancy. However, it is likely that most seeds in natural populations germinate from the seed bank within a few years, before they are lost to burial, predation or fungal attack.

What are the consequences of dispersal for community assembly? On the one hand, over short timescales most plants move short distances, and arrival of new species in a community may be infrequent. On the other hand, over longer timescales (e.g., thousands of years), many plants have a remarkable capacity for long-distance dispersal, and the history of vegetation response to climate change demonstrates that the composition of plant communities is highly dynamic. As a practical matter, many studies of community assembly assume that the plant species in a regional species pool (on the scale of tens to hundreds of kilometers) have the capacity to disperse anywhere within that region, given a reasonable amount of time. To the extent that is true, then community assembly patterns reflect local abiotic and biotic interactions that determine the composition and abundance of co-occurring species. However, it is difficult to establish the exact temporal and spatial scales at which the assumption of unlimited dispersal is a reasonable approximation; at local scales, over short durations, and at biogeographic scales over longer time periods, dispersal may be a critical process that explains patterns of species distributions and community composition.

---

## Abiotic Filtering

One of the central metaphors in community assembly is that of a habitat filter, where the abiotic environment “filters out” species by limiting establishment or survival at particular sites. As plant dispersal is often relatively undirected, seeds may often arrive at locations where conditions are not favorable for germination or long-term survival. These filters can impact plants at any life stage and can involve any of a number of abiotic factors singly or in combination.

Many plant species have specific abiotic requirements for successful germination, and thus the germination stage represents the first point at which habitat filtering can occur. Germination cues can include moisture, temperature, light, photoperiod, and even fire or smoke in some species adapted to fire-prone environments. Many species require specific combinations of abiotic cues, such as a period of cold temperature followed by a photoperiod indicative of long days. Reliance on these cues can help to ensure that a species will not germinate and die in unfavorable conditions. Some species are able to persist in a dormant state as a seed for long periods of time waiting for the proper cues to trigger germination, but the length of time that seeds remain viable varies widely among species. The ability of some species to persist for extended periods in the seed bank can complicate the task of quantifying community membership at a particular site, and an examination of

the seed bank (and testing for seed viability) may be required to definitively conclude that a species is absent from a site. This is most relevant in communities that exhibit substantial variation in abiotic conditions over time, as different species may use the same habitat at different times of the year or in different years, depending on year-to-year variation in weather, remaining dormant in the seed bank at other times.

Abiotic factors can also cause mortality or prevent successful reproduction at any time during the life cycle from germination through reproductive maturity. Species vary in requirements for light, nutrients, and water as well as in tolerance to drought and temperature, and any of these factors can cause mortality at any stage. An important consideration is that brief, extreme climatic events can have strong impacts on species survival. For example, the average climatic conditions at a site may be ideal for the growth and reproduction of a species, but a brief period of extreme cold or heat or a short but severe drought that occurs infrequently can cause significant mortality and effectively remove particular species from a site. For example, a severe drought associated with an El Niño event in the 1980s is thought to have had persistent and long-lasting impacts on the species composition, and associated functional traits, of a tropical forest on Barro Colorado Island, Panama (Feeley et al. 2011). Therefore, in considering the role of abiotic conditions in filtering species from a site, it may be just as important to consider the variance or the extremes of abiotic conditions as it is to consider the average values.

Practically speaking, it can be challenging to distinguish between habitat filtering and dispersal limitation when a species is completely absent from a site. Simple experiments can be helpful in testing for habitat filtering. On the most basic level, these experiments involve transplanting individuals either as adults or as seeds to the site and monitoring germination and/or survival. In situations where these experiments are impractical, seed traps or detailed examination of the seed bank can be useful in ruling out dispersal limitation as the cause of a species absence.

An important consequence of abiotic filtering is that species composition typically changes along environmental gradients. For example, there is widespread evidence that plant communities change in predictable ways along gradients of light, water availability, soil fertility, elevation and latitude, among other factors. These changes in species identity are also often reflected in changes in the functional traits of species, such that average trait values across species in the community can shift along a gradient. For example, woody plant leaf functional traits change consistently across a gradient of soil water availability in coastal California and across microtopographic gradients in the Ecuadorian Amazon (Kraft et al. 2008; Cornwell and Ackerly 2009). Another frequently documented pattern is that the breadth or variance of strategies seen at any point along the gradient is often smaller than is seen across the gradient as a whole. The significance of these observations – i.e., shifts in the mean of trait values and reduction in the range or variance in trait values at points along a gradient – is typically documented using a null model approach, comparing observed communities to hypothetical communities assembled at random from the regional species pool.

## Biotic Interactions

Just as abiotic factors can serve as filters to prevent establishment of species, interactions between plants and other organisms can have important consequences for community assembly. Competition and natural enemies (herbivores, parasites, and pathogens) can reduce growth and survival of plants at a particular site, and positive interactions can allow species to establish and persist at sites where they would otherwise be unable to survive. In many conceptual models of community assembly, biotic interactions are often considered to impact community assembly after abiotic filtering has occurred. While this may be true if the primary habitat filter occurs at the germination stage, in reality biotic and abiotic factors are likely important throughout the lifecycle of most plants. Persistence in a community requires tolerance of stresses in the germination, establishment, and adult reproductive phases, to ensure reproduction of the next generation.

As stated earlier, competition has long been considered to be a central biotic factor in community assembly, dating back to Jared Diamond's initial study of bird communities on islands (and before that back to Darwin, writing in the *Origin of Species*). Competition is hypothesized to impact community assembly by the failure of species to establish or persist at a location in the face of competitive interactions. Early community assembly theory focused on the competitive exclusion principle (Hardin 1960), which hypothesizes that "complete competitors cannot coexist," meaning that species are more likely to be able to coexist if they have niche differences. Early work in this area focused on the concept of limiting similarity, which hypothesized that there was a finite limit to how similar two coexisting species could be. While theoretical work has since suggested that there is not likely to be an absolute limit to similarity, the general idea that differences between species promote coexistence by reducing competition has persisted as a central theme in many community assembly studies. To date, many plant community assembly studies have approached competition by documenting differences in the niches or phenotypes of co-occurring species and testing whether those differences are greater than what might be expected by chance. For example, co-occurring plants in sand dune plant communities in New Zealand and forests in the Ecuadorian Amazon are often more phenotypically distinct from each other than predicted by null models (Stubbs and Wilson 2004; Kraft et al. 2008). In many ways, this approach has direct links to Jared Diamond's initial approach of documenting "forbidden combinations" of species on islands. While phenotypic patterns that are consistent with competition are regularly detected in plant communities, they are far less common than evidence for habitat filtering.

Herbivores, parasites, and pathogens, collectively referred to as natural enemies, can also have important and wide-reaching consequences for community assembly. One challenge in this area is that community assembly studies typically focus just on members of one guild or functional type (e.g., trees or herbaceous plants) and often have not considered other trophic levels. In some cases, the impact of natural enemies can be studied primarily through plant distributions. For example, if species suffer primarily from natural enemies that are species specific, seedlings growing near adult

trees of the same species should suffer more negative effects than seedlings growing far from adults, as natural enemies can become concentrated near adult trees (reviewed in Wright 2002). In this case, the study of plant distribution patterns within communities can offer some insight into the role of natural enemies. However, in other cases, we likely need improved conceptual models and approaches to effectively incorporate natural enemies into community assembly studies.

Positive interactions between species can also have profound impacts on community assembly, allowing species to establish or persist at sites where they would otherwise be unable to survive. Many of these associations are between plants and other organisms. For example, associations with mycorrhizae and nitrogen-fixing bacteria allow many plant species to gain access to essential nutrients more effectively, and many species rely on insect or animal pollinators for reproduction. The absence of these mutualist partners can effectively exclude plants from particular sites. Our understanding of these relationships in a community assembly context is hampered by the same limitations as our understanding of natural enemies – many studies typically focus just on plants, not on other groups within a community. While it is possible to study some consequences of plant-pollinator interactions primarily through the plant community (Sargent and Ackerly 2008), new approaches will be needed to fully incorporate positive interactions that extend beyond a single trophic or functional group into community assembly studies.

It is also well understood that plants can have positive effects on each other. These impacts most commonly involve an amelioration of environmental stress or a reduction in herbivore pressure via associational defenses. For example, in hot and dry environments, some species are known to function as “nurse plants” by modifying the nearby microclimate enough to allow other species to be able to establish. However, many positive interactions between plants are known to be highly context dependent. For example, in one globally replicated experimental study, plants growing at lower elevations on mountains were often found to compete with one another, while species growing at higher elevations on the same mountains (which is presumably a more stressful environment) were found to have positive effects on one another (Callaway et al. 2002). These findings highlight that most positive interactions (and perhaps many species interactions in general) typically include both a positive and a negative component and that the relative importance of these components for community assembly can shift as abiotic conditions change. This also highlights a general but understudied challenge within the topic of community assembly – disentangling the interactions among abiotic and biotic filters.

---

## **Relationship Between Community Assembly and Coexistence Theory**

An important ongoing area of development in community assembly theory is in improving the models of competition to incorporate insights from coexistence theory that have occurred since the development of the community assembly

approach (HilleRisLambers et al. 2012). Community assembly analyses, as discussed above, have typically considered competition to be a process that favors coexistence between species that are phenotypically distinct, assuming that differences in functional traits or functional types between species are related to niche differences. This assumption is grounded in early theories of competition and differentiation in resource use but is incomplete with respect to continuing developments in the theory of species coexistence.

In particular, Chesson (2000) has advocated for a consideration of two distinct phenomena in competitive interactions. First, coexistence between a pair of species is made more likely by the presence of stabilizing niche differences, defined as differences in resource use that give both species an advantage when rare. These advantages allow each species to recover from low abundance, buffering each species against competitive exclusion. Thus, coexistence is enhanced by niche differences, exactly as modeled in much of community assembly theory. However, Chesson goes on to consider another component of the interaction, termed average fitness differences. Average fitness differences reflect differences in the average competitive ability of species, and these differences can lead to the exclusion of the less fit species even if there are niche differences between the two species. Therefore, a pair of species is most likely to be able to coexist when they have large niche differences and minimal fitness differences.

This viewpoint leads to multiple potential phenotypic outcomes of competitive exclusion. If the functional traits of organisms under consideration primarily reflect niche differences, then competition should result in phenotypic disparity between co-occurring organisms. However, if the traits under study correlate instead with average fitness differences, then phenotypically similar species may be more likely to coexist. If traits correlate with both niche and fitness differences, then the outcome may be a combination of patterns or something that appears essentially random. One of the major limitations in making progress in this area is a lack of understanding of the extent to which commonly measured plant functional traits correlate with niche differences, average fitness differences, or some combination of the two. Detailed manipulations that measure functional traits in communities as well as quantifying niche and fitness differences among species will be needed to make progress in this area.

One major difference between community assembly theory and many coexistence approaches is that coexistence theory primarily focuses on species that are able to survive and persist at a site, whereas community assembly considers a broader suite of species and the process of abiotic habitat filtering. This gap highlights the progress that could be made from a better unification of these two areas. Some of the issues and challenges in unifying these approaches are discussed in the last section of this chapter.

---

## Phylogenetic Patterns

In a famous quote from the *Origin of Species* (1859), Darwin noted that “As the species of the same genus usually have...much similarity in habits and constitution, . . . the struggle will generally be more severe between them, if

they come into competition with each other, than between the species of distinct genera.” His observation reflected the general knowledge of any experienced systematist or field naturalist that related species tend to be ecologically similar; e.g., one would expect two grass-eating rabbit species to compete directly for the same food sources, whereas a grain-eating mouse and a carnivorous fox are utilizing quite different resources. In the first half of the twentieth century, experimental studies of competition by Gause and the development of Lotka-Volterra competition theory led to the development of the competitive exclusion principle (Hardin 1960), discussed earlier, which posits that species competing for the same resources could not coexist in a community. Putting these ideas together, ecologists in the mid-twentieth century suggested that species of the same genus would not live together in local communities, at least not as often as one might expect if communities were assembled randomly from the available species in a regional species pool. This prediction was supported in studies of animal communities on islands, compared to the fauna of adjacent mainland regions. These studies provide some of the earliest examples of null models in ecology, discussed above.

Starting in the 1960s, the study of phylogenetics was revolutionized by conceptual, computational, and empirical advances, most notably the breakthroughs in molecular biology leading to the modern era of DNA sequencing. With high-resolution, well-supported phylogenies available, new methods have been developed to reexamine classical questions in ecology and evolutionary biology. The study of plant communities presented particular challenges, as the deeper structure of the angiosperm phylogeny had never been well understood and molecular data brought a number of surprises. The first breakthroughs came in the 1990s, quickly leading to a broad community effort under the Angiosperm Phylogeny Group and a rapidly growing consensus about major patterns in flowering plant phylogeny. Plant ecologists moved quickly to utilize the newly available phylogenies to tackle large-scale problems in adaptive evolution, diversity, and community assembly (Webb 2000). Molecular data also provide branch lengths that quantify the degree of relatedness among species, and fossil calibrations can be applied to estimate branch lengths in millions of years since species diverged from their most recent common ancestor. The phylogenetic distance between two species is defined as the distance from one species down the phylogeny to the common ancestor and back up to the other species (in other words, two times the age of their most recent common ancestor).

The phylogenetic structure of a community can be described in a number of ways, using quantitative metrics based on the phylogenetic relationships for the community, “pruned” from the larger phylogeny of all plants. As in the examples discussed above, statistical analyses of phylogenetic community structure consider a local community relative to a null model of communities assembled from a regional species pool. Two simple measures of phylogenetic community structure are the mean phylogenetic distance, defined as the average of the phylogenetic distances between all pairs of taxa in a community, and the mean nearest neighbor distance, defined as the average distance from each species to its closest relative. Using these measures, the net relatedness index (NRI) and nearest taxon index (NTI)



can be calculated comparing observed communities to hypothetical communities constructed under a null model. These metrics allow us to describe communities along a continuum, from those in which co-occurring species are more closely related ( $NRI > 0$ ) or more distantly related ( $NRI < 0$ ) than expected by chance under the null model. Similarly, NTI measures whether each species' closest relative is more closely ( $NTI > 0$ ) or more distantly ( $NTI < 0$ ) related than expected by chance. NTI is loosely analogous to the study of species/genus ratios, asking whether species tend to co-occur with very close relatives (Webb 2000).

Advances in phylogenetics, together with the assembly of large trait databases, have also allowed broad tests of the extent to which closely related species tend to be ecologically similar. This pattern is referred to as phylogenetic signal, where a high degree of signal indicates that close relatives exhibit similar trait values. A null model can be used to evaluate the significance of these patterns, by randomizing trait values across the tips of the phylogeny to determine the extent of phylogenetic signal that would occur by chance. In broad-scale studies, especially those spanning large global databases, most ecological traits exhibit moderate to strong patterns of phylogenetic signal. However, this pattern may not be observed in smaller, local communities where the set of co-occurring taxa represents a very few representatives across numerous major clades; in these situations, each species closest relative in a community may not be close at all, on a global scale, so the signal of trait evolution is diluted.

As described above, the concept of habitat filtering suggests that species living together in a community will be more ecologically similar than expected, relative to a broader regional species pool. If traits exhibit significant phylogenetic signal, then ecologically similar species will also tend to be closely related. In this situation, local communities may be composed of closely related species (relative to the null model), with  $NRI$  values  $> 0$ . Thus, studies that detect positive  $NRI$  values may be used to infer that habitat filtering processes are significant in the assembly of a local community. On the other hand, there are two scenarios that could lead to  $NRI < 0$ , with communities composed of distantly related species. First, habitat filtering may occur, but the traits that influence species habitat distributions may exhibit low signal, possibly due to rapid divergence among close relatives. Thus, ecologically similar species that co-occur would be distant relatives. Alternatively, co-occurring species may be ecologically distinct from each other, reflecting the outcome of biotic interactions or one of the coexistence mechanisms discussed above. If the associated traits exhibit high phylogenetic signal, then again the species co-occurring in communities will be widely dispersed across the phylogeny and more distantly related than expected by chance.

A study of the community assembly of oaks (*Quercus*) in northern Florida illustrates these latter patterns. The communities are strongly structured along soil moisture gradients, from seasonally flooded bottomlands to dry, sandy uplands. As expected, species that live together share traits related to drought tolerance, a case of habitat filtering with respect to these traits. However, these traits exhibit very low phylogenetic signal, with convergent evolution of low, medium, and high soil moisture tolerance across several sub-clades within *Quercus*. As a result, local

communities are composed of distantly related species, and other differences between the sub-clades suggest that they may exhibit trait differences that enhance coexistence within these communities (Fig. 2).

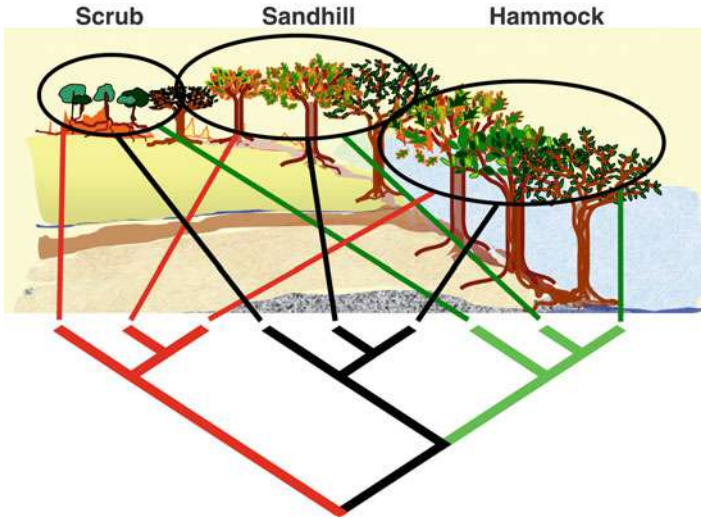
---

## **Biogeography and the Build Up of Species Pools**

As we have discussed above, local communities are assembled from a regional pool of available species. Thus, an understanding of the processes that shape the diversity and the functional and phylogenetic composition of regional biota is valuable for a deeper understanding of local communities. How big is the regional species pool? This is a difficult, perhaps impossible, question to answer precisely as the appropriate temporal and spatial scale defining the regional pool will depend on the size of the communities and the dispersal biology and longevity of the organisms under consideration. In many cases, the delineation of the regional pool for the purposes of empirical studies is constrained practically by availability of data, though the growth and refinement of regional and global biodiversity databases (e.g., GBIF, the Global Biodiversity Information Facility) may eventually overcome this obstacle.

The diversity of a regional flora and fauna reflects the interaction of three fundamental processes: arrival of new lineages by dispersal from other regions, speciation within the region, and local extinction of lineages. Dispersal can occur across substantial distances. As described earlier, the colonization of oceanic islands by seeds dispersed by wind or water currents or by animal agents provides direct evidence of the potential for long-distance dispersal. Similar types of long-distance dispersal events occur across and among continents as well, though they are harder to detect. As climates shift and the combination of continental drift and fluctuating sea levels have altered the connections between landmasses, dispersal can also occur as a stepping-stone process with populations migrating along corridors that provide favorable environments for at least short-term establishment and subsequent reproduction. For example, in the Northern Hemisphere, there has been extensive migration between North America and Europe across a North Atlantic land bridge, during the Eocene and possibly into the Oligocene, while more recently Asia and North America have been connected by the Bering land bridge during periods of low sea level.

On evolutionary timescales, speciation is a key process increasing diversity within biogeographic regions. During the speciation process, evolutionary shifts in habitat affiliation and the climatic tolerances of a lineage tend to change slowly. As a result, individual clades will tend to diversify and spread across major climatic zones, and these biotic similarities then come to define distinctive biogeographic regions around the world. Based on phylogenetic and fossil evidence, it is believed that flowering plants originated in tropical regions and diversified extensively during the Cretaceous. Around 55 million years ago during the Eocene, the world was much warmer overall, and tropical forests extended to midlatitudes, far beyond their current distribution. Cooling and drying trends since then have led to the



**Fig. 2** Phylogenetic community structure of oak-dominated communities in Florida, demonstrating phylogenetic overdispersion within each of the three habitat types. Oaks within each of the three major phylogenetic lineages occur in each community, and null model analysis reveals that this pattern is not expected by chance (Redrawn with permission from Cavender-Bares et al. (2004b))

emergence of temperate climates, and many of these ancestral, tropical lineages gave rise to temperate plant groups that spread and diversified as the cooler climate spread at mid- and high latitudes. Many well-known clades in the temperate flora, such as maples and oaks, first appear in the fossil record during this time and then spread around the Northern Hemisphere in the temperate regions of Asia, Europe, and North America. Drying trends that began in the Oligocene led to the emergence of the semiarid and arid floras, including the world's modern deserts and Mediterranean-type climate zones. Diversification and adaptation to arid climates in these areas tends to be very recent, resulting in many distinctive and locally endemic groups. For example, close to half of the native flora of California is endemic to the Mediterranean-climate region west of the Sierra Nevada.

The Miocene and Pliocene also witnessed a profound ecological transition that continues to shape our modern ecosystems and regional floras around the world. During this time, woodland ecosystems gradually transitioned to open grasslands, likely due to drying and then to an increase in the frequency of wildfire. The hot, open conditions, combined with relatively low atmospheric  $\text{CO}_2$ , promoted the evolution and diversification of  $\text{C}_4$  grasses (grasses that utilize a specialized photosynthetic pathway to concentrate  $\text{CO}_2$  at the biochemical site of carbon fixation). Subsequently,  $\text{C}_4$  grasses spread and became the dominant species in subtropical and warm temperate grasslands, though the particular characteristics of species that become ecosystem dominants are not well understood (Edwards et al. 2010). As illustrated by this example, past environmental conditions have a direct

influence on the evolutionary history and adaptive evolution of regional floras around the world. The functional diversity available in the regional species pool reflects the cumulative results of diversification and adaptive evolution, and the footprint of the past is evidence (though sometimes only on close inspection) in the structure and function of modern ecosystems.

The third important process in the development of regional species pools is extinction. Extinction is also the most difficult to document, as evidence must be sought primarily in the fossil record or by indirect inference from biogeographic distributions and phylogenetic relationships. The balance of speciation and extinction can generate very different levels of diversity, even in climatically similar regions. The Northern Hemisphere flora provides important examples, as the forests of East Asia have much higher diversity of tree species than North America, and North America is in turn more diverse than Europe. This pattern is thought to be due in part to higher rates of extinction in Europe and North America during the Ice Ages (the last 500,000 years), as glaciers extended further south in these areas and plants were pushed up against oceanic barriers (e.g., the Mediterranean).

In recent decades, there has been extensive debate over the role of regional versus local processes as influences on diversity and ecology of local communities. The influence of the regional biota is evident when diversity of local communities is correlated with regional diversity, even under similar climatic and environmental conditions. This is observed in north temperate forests, as local diversity (in small areas, e.g., one hectare) is highest in East Asia, intermediate in eastern North America, and lowest in Europe. While the abiotic and biotic filtering processes discussed above may be operating in all of these forests, the resulting diversity of local communities is still higher when there are more species in the regional pool contributing to the assembly process. This suggests that communities may be structured by factors such as niche differentiation, but may not be ecologically saturated in the sense of reaching a maximum limit on diversity that is set by local ecological factors.

Patterns of ecological and functional diversity in local communities also bear the footprint of evolutionary history. The adaptation of lineages to climatic conditions experienced during their evolutionary history is an important example of “niche conservatism”, a general term for the observation that related species often maintain ecological similarities over very long periods of time. Many examples are now known where diversity and distributions at local and landscape scales reflect niche conservatism of the constituent lineages. For example, in California plants derived from northern lineages are most diverse in cooler and moister parts of the state and are also more common on cool, north-facing slopes or riparian zones of a local landscape. Plants derived from semiarid and subtropical lineages, in contrast, have primarily diversified in the drier, Mediterranean-climate zones of California and are the primary contributors to the drought-adapted chaparral (i.e., evergreen shrubland) vegetation. However, like many patterns in ecology and evolution, there is no one rule that covers all situations. The oak example discussed above illustrates how several clades within a genus can exhibit convergent evolution in habitat

tolerances, so close relatives spread out across the landscape and occupy different habitats. Improved knowledge of phylogenetic history, the fossil record, and the climatic history of different regions of the world will continue to shed light on these fundamental questions in the evolution of regional floras and their influence on the assembly of local communities.

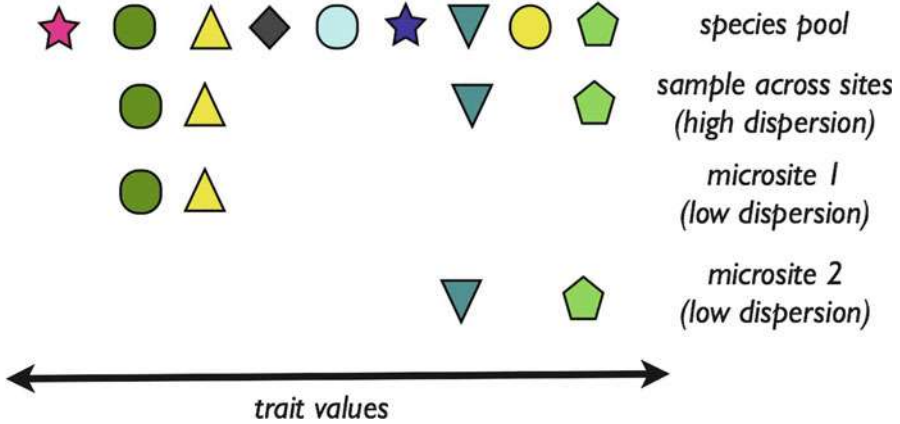
---

## Scale Dependence

Community assembly studies typically focus on comparing the members of a focal community to a regional pool of potential colonists. While this step might seem clear in theory, in practice the definition of an appropriate boundary for a community and a species pool is fraught with uncertainty. It is often best to simply acknowledge that there will be several possible ways in which to delineate the community and the species pool and that each combination may reflect the action of different assembly processes. For example, a species pool could be defined as any species in the vicinity of the focal community that might be able, based on known dispersal distances, to disperse a seed into the community within 1 year or one generation of the focal species. A broader-scale analysis might consider the species pool to be any species in the region, even if it would likely take longer than one generation for some species in the pool to disperse into the local community. It is essential to include an understanding then of how the pool and community were defined when drawing conclusions based on community assembly analyses – an analysis based on a narrowly defined species pool might only be appropriate for making inferences about short-term ecological processes, whereas an analysis based on more broadly defined pool could reflect the action of processes acting over multiple generations.

With this scale dependence in mind, there are a number of cases where a single ecological process is predicted to produce contrasting patterns depending on the scale of analysis. For example, a narrowly defined community sample at a small spatial scale that only contains a single habitat type might readily demonstrate phenotypic clustering or other patterns consistent with habitat filtering when compared to a broader species pool that contains multiple habitat types. But if the community sample is broadened to include two or more habitat types in the same sample, it is conceivable that new, larger-scale analysis will reveal overdispersion, reflecting the aggregation of two or more distinct phenotypic clusters of species that are different from each other (Fig. 3). In this case, a single ecological phenomenon – environmental filtering – will produce different phenotypic dispersion patterns depending on the scale of analysis.

In summary, it is essential for researchers to be cognizant of the criteria that are used to delineate a community and a species pool in a community assembly analysis and also to recognize that any inferences from the analysis will be conditioned on those criteria, as patterns will likely shift as the scope of the pool and community is altered. When possible, explicitly varying the scope of the pool and the sample can be used to detect the action of processes operating at different spatial scales.



**Fig. 3** A single community assembly process (habitat filtering into specific microsites) can produce contrasting phenotypic patterns depending on the scale of analysis. If a microsite is compared to the species pool, either microsite will show low dispersion (phenotypic clustering). However, if a large spatial scale is used to define the community that includes both microsites, this new larger community sample across sites may reveal higher phenotypic dispersion than expected when compared to the species pool

## Future Directions

The study of community assembly theory has seen considerable development since it was pioneered in the middle of the twentieth century. However, considerable challenges remain, and here we highlight three essential areas where additional work is needed, including better methods to distinguish between multiple processes in producing community patterns, a more complete incorporation of coexistence theory into community assembly studies, and better consideration of multitrophic interactions.

## Pattern-to-Process Mapping

Early work in community assembly focused on phenotypic convergence driven by abiotic filters and phenotypic disparity driven by competitive exclusion. However, it has become increasingly apparent that many community assembly processes can produce similar patterns. For example, high phenotypic similarity of species within a community can be produced by habitat filtering, in situ speciation, or pollinator facilitation (Emerson and Gillespie 2008; Sargent and Ackerly 2008), among other processes (Cavender-Bares et al. 2009). In some cases, additional information such as the timescale of the analyses (which is implicit in the spatial and temporal criteria used to define the species pool and the community) or the pollination syndromes of the species in the community can be used to distinguish between potential mechanisms, but in other cases, definitive links between process and pattern can be

challenging. In some cases, approaches that go beyond simple comparisons of a community list to the species pool will be needed. Analyses that focus on variation in performance of individuals over time appear to be particularly promising in this area (Uriarte et al. 2010).

## Coexistence Theory and Community Assembly

Despite a long history of considering the role competition in community assembly, many of the predictions commonly tested in community assembly fail to fully reflect recent developments in coexistence theory. In particular, the recognition that phenotypic similarity can both increase the chances of competitive exclusion (if traits reflect niche differences) and decrease the chances of competitive exclusion (if traits reflect average fitness differences) has only recently been recognized in the context of community assembly (Mayfield and Levine 2010; HilleRisLambers et al. 2012). Progress in this area will depend first and foremost on a better understanding of the extent to which niche and relative fitness differences are correlated with components of the plant phenotype, as the assumption has long been that functional traits primarily capture niche differences. The data needed to properly quantify niche and fitness differences typically require more detailed measurements than most community assembly studies to date (Levine and HilleRisLambers 2009; Adler et al. 2010), and therefore new approaches will be needed to bring a consideration of these phenomena into community assembly analyses.

## Methods for Multitrophic Interactions

Most community assembly analyses focus within a trophic level, and most plant-focused studies do not explicitly consider other trophic levels except as those interactions play out implicitly among plants (e.g., Sargent and Ackerly 2008). However, given the ubiquity of trophic interactions in shaping community patterns, community assembly will not be able fully consider the multitude of ecological interactions shaping local communities until it is able to explicitly incorporate trophic interactions into analyses. Given the rapid pace of development of network theory in recent years, it may be that a robust solution to this issue will emerge from an integration of these two approaches.

---

## References

- Adler PB, Ellner SP, Levine JM. Coexistence of perennial plants: an embarrassment of niches. *Ecol Lett.* 2010;13:1019–29.
- Callaway RM, Brooker RW, Choler P, Kikvidze Z, Lortie CJ, Michalet R, Paolini L, et al. Positive interactions among alpine plants increase with stress. *Nature.* 2002;417:844–8.
- Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. The merging of community ecology and phylogenetic biology. *Ecol Lett.* 2009;12:693–715.

- Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst.* 2000;31:343–66.
- Clark JS, Fastie C, Hurtt G, Jackson ST, Johnson C, King GA, Lewis M, et al. Reid's paradox of rapid plant migration. *Bioscience.* 1998;48:13–24.
- Colwell RK, Winkler DW. A null model for null models in biogeography. In: Strong DR, Simberloff DS, Abele LG, Thistle AB, editors. *Ecological communities: conceptual issues and the evidence.* Princeton: Princeton University Press; 1984. p. 344–59.
- Connor EF, Simberloff D. The assembly of species communities – chance or competition. *Ecology.* 1979;60:1132–40.
- Cornelissen JHC, Lavorel S, Garnier E, Diaz S, Buchmann N, Gurvich DE, Reich PB, et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust J Bot.* 2003;51:335–80.
- Cornwell WK, Ackerly D. Community assembly and shifts in the distribution of functional trait values across an environmental gradient in coastal California. *Ecol Monogr.* 2009;79:109–26.
- Darwin C. On the origin of species by means of natural selection. London: John Murray; 1859.
- Diamond JM. Assembly of species communities. In: Cody ML, Diamond JM, editors. *Ecology and evolution of communities.* Cambridge: Harvard University Press; 1975. p. 342–444.
- Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, C4\_Grasses\_Consortium. The origins of C4 Grasslands: integrating evolutionary and ecosystem science. *Science.* 2010;328:587–91.
- Emerson BC, Gillespie RG. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol Evol.* 2008;23:619–30.
- Feeley KJ, Davies SJ, Perez R, Hubbell SP, Foster RB. Directional changes in the species composition of a tropical forest. *Ecology.* 2011;92:871–82.
- Gotelli NJ, Graves GR. *Null models in ecology.* Washington: Smithsonian Institution Press; 1996.
- Hardin G. Competitive exclusion principle. *Science.* 1960;131:1292–7.
- HilleRisLambers J, Adler PB, Harpole WS, Levine JM, Mayfield MM. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology Evolution and Systematics.* 2012, 43:227–48.
- Kraft NJB, Valencia R, Ackerly D. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science.* 2008;322:580–2.
- Levine JM, HilleRisLambers J. The importance of niches for the maintenance of species diversity. *Nature.* 2009;461:254–7.
- MacArthur RH, Wilson EO. *The theory of island biogeography: monographs in population biology.* Princeton: Princeton University Press; 1967.
- Mayfield MM, Levine JM. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol Lett.* 2010;13:1085–93.
- Nobel IR, Slatyer RO. Post-fire succession of plants in Mediterranean ecosystems. In: Mooney HA, Conrad CE, editors. *Proceedings of the symposium on the environmental consequences of fire and fuel management in Mediterranean ecosystems.* California: Palo Alto; 1977. p. 27–36.
- Pielou EC, Roush RD. Salt-marsh vegetation – Latitudinal gradients in zonation patterns. *Oecologia.* 1976;24:311–21.
- Sargent RD, Ackerly DD. Plant-pollinator interactions and the assembly of plant communities. *Trends Ecol Evol.* 2008;23:123–30.
- Strong DR, Szyska LA, Simberloff DS. Tests of community-wide character displacement against null hypotheses. *Evolution.* 1979;33:897–913.
- Stubbs WJ, Wilson JB. Evidence for limiting similarity in a sand dune community. *J Ecol.* 2004;92:557–67.
- Uriarte M, Swenson NG, Chazdon RL, Comita LS, John Kress W, Erickson D, Forero-Montaña J, et al. Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: implications for community assembly. *Ecol Lett.* 2010;13:1503–14.
- van der Valk AG. Succession in wetlands – a Gleasonian approach. *Ecology.* 1981;62:688–96.
- Violle C, Enquist BJ, McGill BJ, Jiang L, Albert CH, Hulshof C, Jung V, et al. The return of the variance: intraspecific variability in community ecology. *Trends Ecol Evol.* 2012;27:244–52.



- Webb CO. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am Nat.* 2000;156:145–55.
- Weiher E, Keddy PA. *Ecological assembly rules: perspectives, advances, retreats.* Cambridge: Cambridge University Press; 1999.
- Westoby M, Wright IJ. Land-plant ecology on the basis of functional traits. *Trends Ecol Evol.* 2006;21:261–8.
- Woodward FI, Diament AD. Functional approaches to predicting the ecological effects of global change. *Funct Ecol.* 1991;5:202–12.
- Wright SJ. Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia.* 2002;130:1–14.

## Further Reading

- Abrams P. The theory of limiting similarity. *Annu Rev Ecol Syst.* 1983;14:359–76.
- Ackerly DD. Adaptation, niche conservatism, and convergence: comparative studies of leaf evolution in the California chaparral. *Am Nat.* 2004;163:654–71.
- Ackerly DD, Cornwell WK. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecol Lett.* 2007;10:135–45.
- Bertness MD, Callaway R. Positive interactions in communities. *Trends Ecol Evol.* 1994;9:191–3.
- Bruno JF, Stachowicz JJ, Bertness MD. Inclusion of facilitation into ecological theory. *Trends Ecol Evol.* 2003;18:119–25.
- Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA. Phylogenetic overdispersion in Floridian oak communities. *Am Nat.* 2004;163:823–43.
- Cavender-Bares J, Kitajima K, Bazzaz FA. Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. *Ecol Monogr.* 2004;74:635–62.
- Cornelissen JHC, Lavorel S, Garnier E, Diaz S, Buchmann N, Gurvich DE, Reich PB, et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust J Bot.* 2003;51:335–80.
- Cornwell WK, Schwiik DW, Ackerly DD. A trait-based test for habitat filtering: convex hull volume. *Ecology.* 2006;87:1465–71.
- Diaz S, Noy-Meir I, Cabido N. Can grazing response of herbaceous plants be predicted from simple vegetative traits? *J Appl Ecol.* 2001;38:497–508.
- Fine PVA, Mesones I, Coley PD. Herbivores promote habitat specialization by trees in Amazonian forests. *Science.* 2004;305:663–5.
- Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen-host range. *Proc Natl Acad Sci USA.* 2007;104:4979–83.
- Kraft NJB, Ackerly DD. Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian forest. *Ecol Monogr.* 2010;80:401–22.
- McGill BJ, Enquist BJ, Weiher E, Westoby M. Rebuilding community ecology from functional traits. *Trends Ecol Evol.* 2006;21:178–85.
- Paine CET, Baraloto C, Chave J, Hérault B. Functional traits of individual trees reveal ecological constraints on community assembly in tropical rain forests. *Oikos.* 2011;120:720–7.
- Ricklefs RE. Community diversity – relative roles of local and regional processes. *Science.* 1987;235:167–71.
- Ricklefs RE, Schluter D. *Species diversity in ecological communities: historical and geographical perspectives.* Chicago: University of Chicago Press; 1993.
- Ricklefs RE, Travis J. A morphological approach to the study of avian community organization. *Auk.* 1980;97:321–38.
- Shipley B. *From plant traits to vegetation structure: chance and selection in the assembly of ecological communities.* Cambridge: Cambridge University Press; 2009.

- Swenson NG, Enquist BJ. Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. *Ecology*. 2009;90:2161–70.
- Swenson NG, Enquist BJ, Pither J, Thompson J, Zimmerman JK. The problem and promise of scale dependency in community phylogenetics. *Ecology*. 2006;87:2418–24.
- Valiente-Banuet A, Verdu A. Facilitation can increase the phylogenetic diversity of plant communities. *Ecol Lett*. 2007;10:1029–36.
- Vamosi SM, Heard SB, Vamosi JC, Webb CO. Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol Ecol*. 2009;18:572–92.
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ. Phylogenies and community ecology. *Annu Rev Ecol Syst*. 2002;33:475–505.
- Weiher E, Keddy PA. Assembly rules, null models, and trait dispersion – new questions front old patterns. *Oikos*. 1995;74:159–64.

Yan Linhart

## Contents

Introduction .....	90
Pollination .....	91
How Do Plants Benefit? .....	91
How Do Animals Benefit? .....	91
Dispersal Agents: Who Is Involved? .....	92
Precision of Delivery Systems .....	92
How Plants Manipulate Fertilization .....	96
Genetic and Evolutionary Consequences of Pollination Patterns .....	98
Evolutionary Dynamics: Evolution in Action .....	100
Pollination Is Just Part of the Story .....	101
Dispersal .....	102
How Do Plants Benefit? .....	102
How Do Animals Benefit? .....	102
Seed Packaging .....	103
Dispersal Agents: Who Is Involved? .....	103
Fruit Characteristics .....	105
Patterns of Dispersal .....	106
Evolutionary Dynamics: Evolution in action .....	109
Seed Dispersal Is Just Part of the Story .....	109
Synthesis and Conclusions .....	110
The Systematics of Associations .....	110
Effects of Dispersal Patterns and Ecological Heterogeneity on Genetic Organization of Populations .....	110
Pollination, Dispersal, and Human Activities .....	112
Future Directions .....	114
Summary .....	115
References .....	115

---

Y. Linhart (✉)

Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

e-mail: [yan.linhart@colorado.edu](mailto:yan.linhart@colorado.edu)

---

**Abstract**

- Plants depend on a wide diversity of animals for pollination and seed dispersal.
- The devices used by plants to attract animals span the whole range of animal senses.
- The diversity of plant sexual systems is very broad.
- Plants can manipulate their own fertilization.
- Pollen delivery systems must be precise.
- Changes in floral attraction signals can have evolutionary consequences.
- Seeds must be dispersed away from parent plants and each other.
- Pollination and seed dispersal influence the genetic structure of populations and their evolution.
- Human activities have significant and often negative impacts on pollinators and seed dispersers. Problems are predictable.
- The complex nature of these plant-animal interactions means that many questions await future studies.

---

**Introduction**

Pollination events can make headline news. Seriously. Consider the headlines below, and note the sources: Huffington Post, New York Magazine, Washington Post. . .not bad for press coverage of a botanical topic.

“Corpse Flower,” World’s Stinkiest Plant, Blooms In Washington At U.S. Botanic Garden *Huffington Post* 7/21/2013

Washington’s Stinkiest Flower to Reach Peak Smell on Monday  
*New York Magazine* 7/21/13

The corpse flower is in bloom *Washington Post* 7/22/13

Not to be outdone, fruits can also get press coverage. The ginkgo tree produces fruit that smells so foul that stories in the New York Times (November 5, 2010), Washington Post (October 10, 2009), and other newspapers (e.g., Chicago Tribune, Orange County Register, Toronto Sun) have commented on the unhappiness of residents of neighborhoods where the fruit-producing trees have been planted, and Washington DC horticulturists even tried to sterilize female ginkgos to avoid the olfactory assault of ripening fruit. . .to no effect.

Plants are immobile and therefore depend on the actions of curious and hungry strangers who will disperse their pollen and seeds. In some cases, the evolutionary drive to attract pollen or seed dispersers gets pretty extreme as in the case of the corpse flower and ginkgo, about which more later. But the excitement conveyed by the news coverage – at least for the corpse flower – illustrates nicely why the topics of pollination and dispersal provide entertainment and fascination. These topics are also sources of scientific curiosity provided by the esthetics of colorful and odoriferous flowers and fruit and the impressive gymnastics that animals often go through to get at rewards of pollen, nectar, fruit, and other goodies.

In plants, sexual reproduction involves the fertilization of an ovule by a traveling pollen grain. Once that has occurred, the embryo develops inside a seed, and at maturity, this seed must be dispersed away from the parent. In both scenarios, the movement must be provided by an external force.

The primary focus of this discussion will be on pollination and dispersal by animals and wind, followed by descriptions of the influences of these interactions upon gene flow and the implications of pollination and seed dispersal in ecological, evolutionary, and practical contexts. Most of these interactions are complex and have multiple ramifications. Consequently, some are poorly understood and in need of further work. The topics that need such work will be pointed out.

Globally, animals are by far the most common pollen vectors. It is estimated that they pollinate about 78 % of temperate species and 94 % of tropical ones (Ollerton et al 2011). Animals are also responsible for the bulk of seed dispersal in tropical communities and many temperate ones. The types of vectors that plants take advantage often depend on the ecology of the areas they grow in. For example, wind dispersal of pollen and seeds tends to occur in certain relatively harsh climates and also in certain taxonomic groups. Seed dispersal by fish has developed in the seasonally flooded regions of the Brazilian *Pantanal*. Nonflying marsupial mammals are common in Australia and serve as important pollinators to a number of plants there. Given the importance and diversity of animals in these interactions, as well as in herbivory (see Chap. 6, ► [“Evolutionary Ecology of Chemically Mediated Plant-Insect Interaction”](#)), it is no surprise that the plant ecologist John Harper once quipped:

The plant kingdom is very largely what the animal kingdom made it.

---

## Pollination

### How Do Plants Benefit?

Sex makes the world go ‘round even for sessile plants, so that is what pollen dispersal is about: improving the likelihood that individuals produce seeds that will pass on their genes. It is true that under some specific circumstances, certain plants are able to set seed without the intervention of animals or wind. These plants are said to *beautogamous* or self-pollinating. However, given the great preponderance of cross-pollination in the plant kingdom, clearly there are significant advantages to mating with other individuals.

### How Do Animals Benefit?

Plants obviously benefit from these interactions, but what do animals obtain in return? They have access to various rewards, including nectar or pollen or a combination of both, or as in the case of some orchids, a resin that males of certain

orchid bees use to mark territories. Some plants also produce various oils (used as food) or resins (as nest-building materials) within their flowers. Some plants exploit the passionate libido of male bees, wasps, or flies by mimicking the appearance and sometimes the odor of females. These female look-alikes get mounted by males whose exertions then either pick up pollen or deposit it (Gaskett 2011; Ellis and Johnson 2010). Finally, other plants like the corpse flower take advantage of the propensity of certain insects to be attracted to smelly corpses: they emit odors sure to keep us away but wonderfully evocative to these carrion-feeding beetles and flies, which are fooled by the smells and pollinate the flowers.

The various rewards offered by plants to attract pollinators are important drivers of ecosystem function. For example, virtually all of the 25,000 or so bee species in the world depend on pollen and/or nectar for their food. At least 650 species of birds are obligate nectar feeders (about half of them are hummingbirds) and many more such as orioles, warblers, and finches will partake periodically. About 10 % of bats are pollinators for over 500 species of plants belonging to about 70 families.

## Dispersal Agents: Who Is Involved?

Within angiosperms, on an evolutionary timescale, animal pollination is basal, meaning that the earliest ancestors of flowering plants depended on animals from their origins onwards. Indeed, there is evidence that even among long-extinct seed plants and their surviving descendents such as the living fossil *Welwitschia* as well as Gnetales and cycads, animal pollination was and continues to be the *modus operandi* (Barrett 2002; Hu et al. 2008).

The animal groups involved include vertebrates, primarily birds and mammals, but also lizards, while the invertebrates include primarily bees, bumblebees and wasps (Hymenoptera), butterflies and moths (Lepidoptera), flies (Diptera), beetles (Coleoptera), and thrips (Thysanoptera). The agents that provide the pollen transfer are used to give names to the flowers using those agents. Thus, flowers pollinated by wind are said to exhibit *anemophily*, those visited by birds exhibit *ornithophily*, bat pollination is *chiropterophily*, bee pollination is *melittophily*, and so on (Proctor et al. 2012; Olesen and Valedo 2003; Waser and Ollerton 2006).

Abiotic pollination is mostly by wind, which is the primary pollen transporter for various groups such as pines, oaks, walnuts, grasses –including corn, wheat, and rice – and sedges. Water can also provide pollen transport and does so in aquatic species such as *Ceratophyllum*, *Potamogeton*, and *Zostera*.

## Precision of Delivery Systems

If all animals were attracted to all flowers, the outcomes would leave much to be desired. Imagine if all the mail and text messages sent out by individuals to specific addresses within a single city arrived haphazardly in batches to various recipients

who had never heard of the senders. Some precision is needed to ensure delivery to appropriate locations. This is where plants can take advantage of the diversity of animals that visit flowers and the diversity of sensory systems in these animals.

Plants have diversified greatly in terms of their mechanisms of attraction. Figure 1 provides a small sample of the diversity of floral shapes, colors, and architectures used to attract pollen dispersers, including the corpse flower.

As the animals approach the flowers, they perceive one or more of the following features: (1) fragrance, (2) color, (3) morphology, and finally (4) rewards provided. Also, the timing (day or night, and flowering season) of the rewards can be manipulated to take advantage of the times of activity of their pollinators. Finally, depending on the plant and pollinator involved, it seems that all sensory abilities of pollinators can be exploited: sight, smell, taste, touch, and even hearing.

A combination of these features can attract specific pollinators and ensure reasonably accurate pollen delivery. For example, birds such as hummingbirds in the Americas and sunbirds, sugarbirds, and honeyeaters on other continents are attracted to red or orange flowers with little or no fragrance, which secrete large amounts of dilute nectar, and are often long and tube shaped. Conversely, bees tend to be attracted to flowers with strong fragrances, which are often yellow or blue and produce smaller amounts of more concentrated nectar and lots of pollen (Proctor et al. 2012).

Some features of flowers are only visible under ultraviolet light: these patterns can manifest themselves as lines or spots on petals and often serve the important function of guiding pollinators to nectar rewards. As such these patterns are referred to as nectar guides and are very poorly studied (Primack 1982).

At night, moths tend to gravitate towards flowers with delicate sweet smells like jasmines that are white in color, while bats tend to visit greenish to purplish flowers with strong smells of fermentation. These sets of floral characteristics are usually called *pollination syndromes* (Proctor et al. 2012).

These descriptions do not imply that pollinators are fixed in their preferences and will not visit flowers that deviate from these characteristics. For example, there are multiple reports of hummingbirds visiting thistles, white-flowering jasmines, lavender with intense blue flowers, and pink apple blossoms. Conversely, bees can visit red or white flowers, while hawk moths will go to yellow or pink flowers. Such behaviors indicate that pollinators have to be expedient in their choices: when their preferred menus are not available, they make do. These patterns of pollinator flexibility have led some students of pollination to doubt the accuracy of pollination syndromes. More recent work has addressed this issue in a comprehensive manner and has concluded that these patterns of preferences are valid in many situations, but it is important to remember that many species have flowers whose signals are understood by many animals and offer rewards accessible to many species (Fenster et al. 2004).

Given that the signals broadcast by flowers often generate predictable behaviors by specific visitors, it follows that when floral signals change, new visitors can be attracted. Such shifts in visitor identity can change the patterns of pollen dispersal from this plant. This matter is discussed in detail later in the section dealing with evolution in action.



**Fig. 1** Sample of diversity of flower colors and architectures used by plants to attract pollinators. *Top row*: paperwhite, *Narcissus papyraceus*; cardoon, *Cynara cardunculus*. *Second row*: corpse flower, *Amorphophallus titanicus*; dahlia, *Dahlia* sp.; saffron, *Crocus sativus*. *Third row*: sage, *Salvia cinnabarina*; flamingo flower, *Anthurium andraeanum*; pincushion tree, *Leucospermum* sp. *Bottom row*: tomato, *Lycopersicon esculentum*; violet, *Viola* sp. (Corpse flower photo from Wikia; all other photos by YBL)

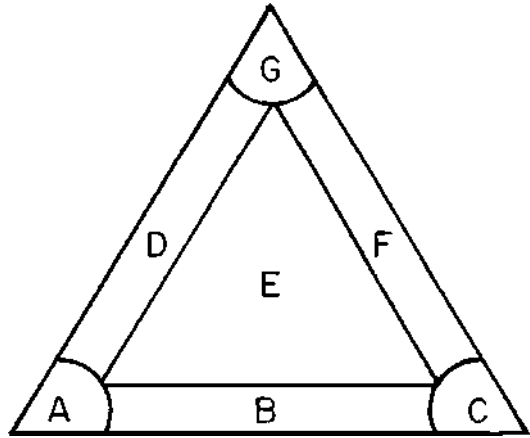


Touch and hearing have also been documented as factors relevant to some specific pollinators. Certain bumblebees are especially fond of flowers such as snapdragons, whose petals have rough surfaces, which enable the insects to grip the flowers firmly and extract the nectar rewards more easily (Whitney et al. 2009). Bats depend on sonar and hearing as they navigate their world, and at least two plants are now known to help bats locate flowers and nectar sources by focusing their hearing. In the tropical vine *Marcgravia*, there is a dish-shaped leaf positioned so as to provide characteristic echo signatures that serve as a beacon towards the open flowers (Simon et al. 2011). In the vine *Mucuna*, the flowers contain a small concave “mirror” produced by two petals that also sends out signatures that enable bats to distinguish between flowers with abundant nectar and those without (von Helversen and von Helversen 2003).

The need for precision of pollen delivery, coupled with the ability of plants to exploit the whole panoply of animal senses, leads to one logical question: how specialized can these interactions get? For animals, as noted above, they must be flexible and willing to exploit any resources that are available. As for plants, they very seldom rely on one, two, or three species. However, there are a couple of remarkably tight associations between specific groups of plants and pollinators. *Ficus* have been an evolutionarily active genus, with hundreds of species in tropical and subtropical regions of several continents, and they rely on small wasps for their pollination. Often, this reliance is so tight that one or a few species of wasps pollinate a single fig species. The other genus that depends on very specific pollinators is *Yucca*, which is pollinated by small moths. *Yuccas* and figs are often cited as unusual examples of close specialization. However, detailed studies of both associations indicate that, under specific circumstances, the plants can evolve novel solutions to their needs for pollinators (Patel et al. 1993; Dodd and Linhart 1994).

There are many unresolved questions about the influences of floral signals upon pollinators; they await the next generation of students of pollination. One of the more contentious questions has focused on why bird-pollinated flowers tend to be red. It seems that at least in hummingbirds, there is no innate attraction to the color red, and it has been suggested that that bird flowers are red because they are not as easily detectable by various bees, which means there is less competition for nectar. However, it is not that simple. For starters, not all reds are created equal: reds vary in their wave lengths, and then some reflect ultraviolet (UV) light, others absorb it. The UV reflectors may attract bees, while the UV absorbers do not. In addition, different groups of pollinating birds have somewhat different visual systems. Recent work provides more details. For example, in a detailed analysis of 206 plant species in Australia, Shrestha et al. (2013) demonstrated that bird-pollinated and insect-pollinated flowers differ significantly in the chromatic cues of their flowers and that, although there is a good deal of variation among the species, the wavelengths involved are concentrated near the optima useful for discrimination by the two groups.

**Fig. 2** Diversity of mating systems in plants. The peaks of the triangle indicate: *A* = cross-fertilized species (including dioecious, self-incompatible, dichogamous ones); *C* = primarily selfing species; and *G* = apomicts. *B* = partially self-pollinated species; *D, E, F* = various sorts of mixed mating systems, including apomixis. Their location implies that they fall closer to *A, B*, or *C*, respectively, when sexual (Figure modified from Kearns and Inouye 1993)



## How Plants Manipulate Fertilization

Overall, plant mating systems show a range of possibilities far more diverse than anything that animals have been able to come up with (Fig. 2). In addition, for any one species, its mating system is dynamic and can vary in space and time. A recent synthesis of the evolutionary dynamics of these mating systems is provided in a special issue of the *Annals of Botany* (Karron et al. 2012).

Just as in many animals, there are some plant species that have separate male and female plants. This condition is known as *dioecy*. Examples of dioecious plants include the ginkgo and many yews (*Taxus*) and junipers (*Juniperus*) among the conifers, as well as hollies, hops, date palms, poplars, and willows. But this is a relatively uncommon condition in plants, probably because of the reproductive challenges of being sedentary. One solution is to have separate male and female flowers on the same plant, a condition called *monoecy*. Examples of monoecious species include many conifers such as pines, along with oaks, corn, and squashes.

The majority of flowering plants have so-called *perfect* or hermaphroditic flowers which means that both the female and male structures are within the same flower. However, many species have also evolved variants on that theme, and the primary evolutionary driver of such alternative morphologies is the promotion of *allogamy*, or *outcrossing*. Other variants include *gynodioecy* whereby some individuals have only female flowers while others have perfect flowers. Certain saxifrages, thyme, and other species use that system. *Androdioecy* involves some individuals being males and others hermaphrodites. Examples include some relatives of potatoes and also asparagus. *Subdioecy* involves three types of individuals: some are male, some female, and some are hermaphrodites.

Plants with hermaphroditic flowers also have ways to ensure or at least increase the likelihood of outcrossing. These include morphological variation in flower shape such as variation in length of the style, the column that supports the stigma or female part of the flower where the pollen must land to initiate pollination.

These styles can be either long or short, a condition known as *distyly*. Given individuals produce only flowers of one type or the other. In order to effect pollination only pollen from the opposite style length will be acceptable to a given individual. The most famous distylous species are primroses. Darwin became intrigued by this phenomenon, and his studies of primroses contributed significantly to his ideas about evolution. As the horticulturist Henry Mitchell once pointed out, who knew that from those modest primroses one could develop such revolutionary concepts. Tristylous species are rare but operate on a similar principle. Another form of separation of floral parts involves both the location of male and female parts within the flower and the timing of floral development. For example, in many species with tubular flowers, as the flower opens, the first organs to be exposed are the anthers that carry pollen. After some time, typically 1–3 days, the anthers dry up and the style elongates, so that the stigma protrudes furthest out of the corolla and is most likely to receive pollen from another flower. In some species, including *mimulus* and some members of the family Bignoniaceae, the stigma has two lobes that can close and prevent pollen deposition in response to touch. This has been posited to be an adaptation to prevent self-pollination, but the evidence is modest.

Physiological mechanisms also prevent self-pollination. These are called self-incompatibility and basically involve the ability of a plant to differentiate between self-produced pollen and pollen from another plant. The former either cannot germinate or the pollen tube grows more slowly down the style. Such self-incompatibility is very common throughout the plant kingdom.

When pollinators are unreliable, plants can also manipulate these systems in other ways. For example, at the periphery of species distributions or on islands, populations evolve away from these mechanisms. Thus, on islands, species characterized by distyly, including the primroses noted above have evolved to become homostylous. *Yuccas* and other species that are characteristically self-incompatible in the center of their ranges evolve towards being self-compatible on the periphery (Dodd and Linhart 1994).

Once pollen has fertilized an ovule, some plants can still have some control over the genetic quality of their offspring. For example, if they have two or more embryos within a single seed (a condition known as polyembryony), there can be competition among embryos, and the slower-growing ones fall by the way side, and only one emerges at germination and feeds on the seed resources. This condition is known in several grasses including corn, rye, and wheat and also in many conifers such as pines and *Araucaria*.

About those self-pollinators, it is known from basic genetics that inbreeding is deleterious, so what about those species that self-pollinate their flowers? These species have several features that mitigate the consequences of such inbreeding. First, natural selection has reduced the frequency of deleterious alleles to such an extent that the probabilities of homozygous combinations with lethal effects are very low. When they do occur, the seeds are simply aborted, and this is no great loss, as such species typically produce many hundreds to thousands of seeds per reproductive episode. Second, many of these species are polyploid, which provides a reservoir of genetic variability. Third, there is periodic outcrossing in many of these species, which replenishes the reservoir of variability needed.

Some species have yet another reproductive mechanism that can be useful: apomixis which involves the ability to reproduce without fertilization. This is especially useful when pollinators are unreliable but seeds must be produced at all costs. It is an especially useful attribute in many weeds. However, even in species that are apomictic, some opportunity for pollination and the associated recombination is often maintained. For example, the dandelion *Taraxacum officinale* is a well-known apomict and notorious weed. Yet it still produces pollen and nectar (not needed if seeds are simply produced via mitosis) and is visited by many insects. Careful analyses show that indeed at least some seeds in some plants are produced sexually and help maintain genetic variability (Richards 2003).

## Genetic and Evolutionary Consequences of Pollination Patterns

### Mating Patterns

Given the central role of animals in pollination, any factors that influence the behavior of pollinators can have important repercussions on pollen dispersal and, therefore, mating patterns in plant populations. For example, in the Americas, many plants are pollinated by hummingbirds. Some hummingbirds tend to set up territories around dense clusters of flowering plants. As a result, pollen dispersal is limited (Fig. 3). A similar pattern occurs among bees: highly social bees such as honeybees tend to forage in groups on concentrations of flowers.

These behaviors in turn lead to a high frequency of cross-pollination among near neighbors. Turner et al. (1982) simulated the gradual change in genetic structure of a plant population where such long-term near-neighbor mating is maintained: the population changes from being a complete and random mix of genotypes in Generation 1 to genetic patchiness by Generation 100 and beyond (Fig. 4).

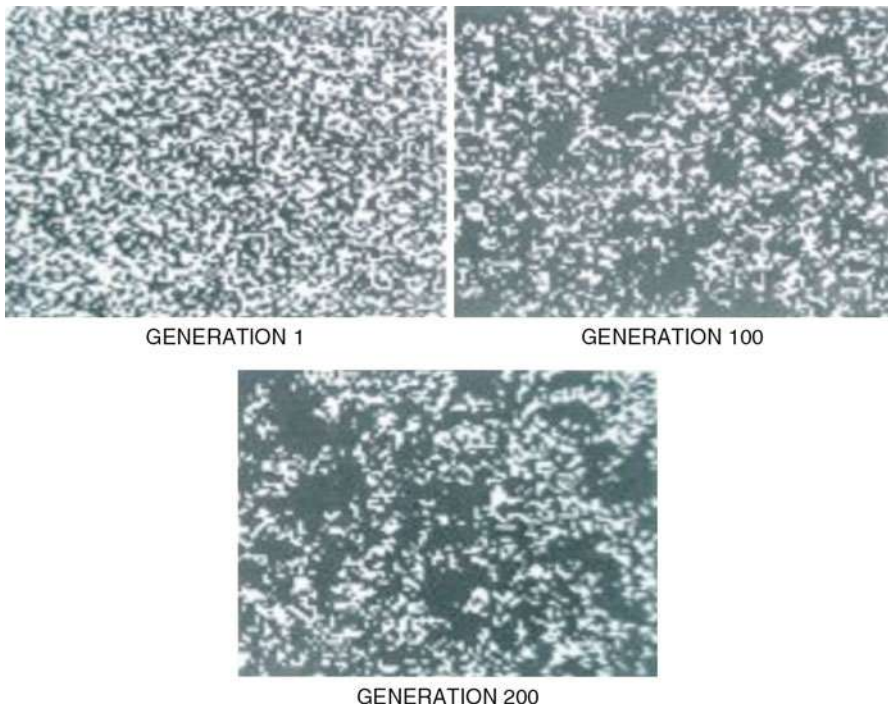
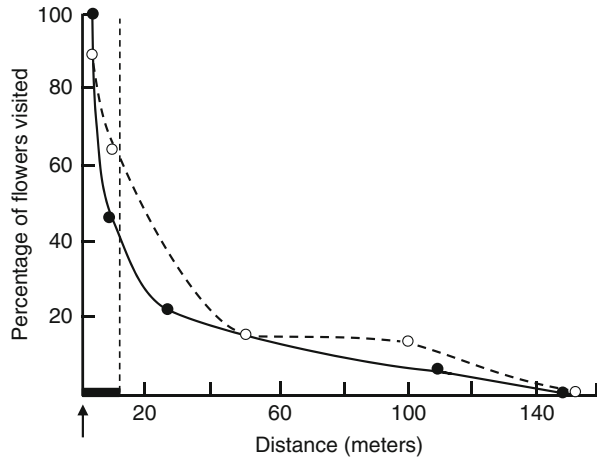
Mating can also occur between unrelated individuals living far from each other. For example, some hummingbirds, called hermits, prefer plants with few large flowers with high nectar rewards and travel longer distances to feed at specific locations. The result is that such plant species exhibit higher outcrossing. This behavior is called *trapline* foraging, and in addition to hummingbirds, some bees and bats are also long-distance pollen dispersers. These trapliners all play essential functions especially in tropical ecosystems where many plants are present in populations of very low densities, often in fragmented habitats, and therefore need reliable long-distance delivery systems.

Wind pollination, being passive, generates a very different pattern of pollen deposition. The general shape is often that of a curve with a leptokurtic distribution, which means that, compared to a normal distribution, a higher than expected amount of pollen is deposited near the source and at long distances and lower than expected proportion is deposited at intermediate distances.

### Population Structuring

One of the important features of populations is their so-called effective size, usually symbolized as  $N_e$ , which is defined as the number of individuals among whom

**Fig. 3** Dispersal of pollen of *Heliconia latispatha* by the territorial hummingbird *Amazilia saucerrottei*. The pollen was labeled with dye at distance 0. The center of the territory is denoted by the arrow and the edge by the dotted vertical line. The two curves represent data from two separate days (From Linhart 1973)



**Fig. 4** Distribution of genotypes in a simulated population at generations 0, 100, and 200. As a result of nearest-neighbor pollination, note the shift of pattern from random distribution of black and white genotypes to patches of white and black denoting groups of homozygous genotypes developing gradually (Modified from Turner et al 1982)

mating is random. This effective population size is strongly influenced by pollination. As might be expected,  $N_e$  tends to be smaller in plants that are self-compatible and that are pollinated by small insects for which energetic constraints and optimal foraging limit long-distance movement. In contrast, larger pollinators, including the trapliners discussed above, as well as wind pollination usually lead to the development of larger population sizes. Whenever plants occupy environments that are ecologically heterogeneous, such as mosaics of soil conditions, strong elevational gradients, variable moisture, or light, these living conditions impose natural selection, and the evolutionary response to this selection will depend on the extent of gene flow. In other words, gene flow is a homogenizing force across landscapes unless it is limited. This means that small  $N_e$  will promote genetic differentiation across small distances.

## Evolutionary Dynamics: Evolution in Action

Interactions between plants and pollinators provide wonderful opportunities to flesh out the comment by Harper noted above and elucidate just exactly how it is that plants are what animals made them (Patin [2011](#)). For example, Sapir and Armbruster ([2010](#)) and their collaborators have addressed this issue recently and illustrate how such interactions can influence evolutionary changes starting with the genetic basis for variation in floral features and how such variation then influences pollinator behavior, leading to subtle population divergence, then speciation, and the detection of these patterns with the help of phylogenetic analyses.

One kind of analysis shows the logic involved in some of this research. Given the tendency for pollinators to pay attention to specific features of flowers, what happens when these features change? The identity of pollinators can change as well. The best examples of such changes involve shifts in color or scent. Both can be under the control of one or a few genes, so any mutations that change gene function and lead to the production of flowers of different colors or odors can have significant effects on pollinator visitation and, therefore, gene flow within and among populations. Several studies have documented the association between differences in color or odor and differences in pollinator suites (Adler and Irwin [2012](#); Sheehan et al. [2013](#)). With respect to color, much of the evidence supporting this pattern of change come from studies of closely related species which show that differences in flower color are associated with differences in pollinator preferences. For example, in the genus *Penstemon*, one study analyzed the shape and color of 49 species and documented the fact that these species are strongly differentiated with respect to pollination by different groups. Flower color distinguished hummingbird from bee-pollinated species, and among the latter, flowers visited by larger bees have relatively open and short floral tubes, while those pollinated by smaller bees tend to have long narrow floral tubes. In *Petunia*, *P. integrifolia* has purple flowers and is pollinated by solitary bees, while *P. axillaris* has white flowers pollinated by hawk moths, and *P. exserta* has red, hummingbird-pollinated flowers. In the case of fragrance, studies with *Polemonium viscosum* have documented the existence of so-called *scent*

morphs which are either sweet or skunky to the human nose. Sweet-smelling flowers are pollinated by large bees and have wider corolla lobes, have longer corolla tubes, and are generally more flared out to accommodate these large bees. Skunky flowers are pollinated by small flies, whose body mass is about 1.4 % that of the big bees. They are smaller but produce just as much nectar per flower. Populations of *Polemonium* often have both morphs, but bees provide about 75 % of the visits in treeless tundra but only about 10 % in the lower zone of scattered trees called krummholz. Flies show the reverse pattern (Galen 1989).

Such intra-specific variation in plants can indeed be an agent of genetic differentiation, but how often it leads to speciation is still open to debate and in need of further studies with a broad array of species. These issues and their complexities are discussed in detail by Kay and Sargent (2009) who conclude that floral differences, and the associated differences in pollinator identity and behavior, are rarely if ever sufficient to lead to speciation by themselves, while Schiestl (2011) suggests that they may.

## Pollination Is Just Part of the Story

Evolutionary interactions between plants and pollinators are never simple quid pro quo affairs. There are always complications that are pertinent to the outcomes. The most common challenge for plants is the issue of attractiveness: plants have to deal with the quintessential quandary – to flash or not to flash...to smell or not to smell...those are the choices, because the signals they emit can make them attractive to pollinators and to herbivores! For example, in the wild radish, *Raphanus sativus*, plants can produce flowers of variable colors, which range from white and yellow to pink and bronze. Pollinators prefer plants with white and yellow flowers. The problem is that so do many herbivores. The resulting diversity of selection pressures helps maintain a flower color polymorphism in the species (Irwin et al. 2003). In *Petunia*, floral odors can attract both pollinators and florivores. The solution is for the plant to emit a complex blend of odors: some components attract the former, while others are demonstrably repellent to the latter (Kessler et al. 2013). These sorts of conflicting selection pressures have repercussions on many features of plant anatomy and flowering patterns (Strauss and Irwin 2004). But it gets more interesting still: the colorful attractiveness of flowers can be exploited by other members of the community. For example, there is a whole suite of spiders called crab spiders that take advantage of these situations. Some crab spiders hide inside flowers and catch unsuspecting visitors. In Australia, certain crab spiders mimic flower colors, and even UV reflectance, to lure bees into their grip (Llandres et al. 2011). In other communities, small mites reside inside flowers visited by hummingbirds, feed on pollen and nectar, and hop onboard for quick transport to other flowers as needed. Finally, one may well ask why do flowers vary at all. It is tempting to think that the attraction of pollinators is the primary selective agent responsible for this variation, but in fact, in addition to floral herbivores and nectar thieves, other constraints are always at work; they include limited resources



that must be partitioned in some optimal way, the need to complete one's life cycle before harsh conditions set in, and various demands associated with genetic variability (Galen 1999).

---

## **Dispersal**

### **How Do Plants Benefit?**

There are multiple reasons why it is advantageous for seeds to be transported away from their seed parent and from each other. All of them can exert strong selection pressures favoring dispersal.

#### **Escape**

Seeds need to move away so as to reduce the likelihood that they will be damaged by herbivores, parasites, or disease organisms that befall their seed parent. In addition, when seeds and the seedlings they produce are in high densities, such settings increase the likelihood of density-dependent attacks by seed or seedling consumers.

#### **Improved Growing Conditions**

Soils near adult plants may be depleted of nutrients and/or have less water, more shade, and perhaps an accumulation of toxic secondary compounds such as terpenes leached into the soil from mother plants.

#### **Colonization of New Habitats**

Whenever a plant can establish in a habitat where it was absent before, it may benefit for a variety of reasons, including increases of population size and escape from herbivores and other consumers and diseases.

#### **Genetic Recombination**

Variability is a basic requirement for survival and adaptive evolution. Given that once established plants will most likely exchange genes with near neighbors, if such neighbors are genetically related, inbreeding ensues, and the next generation will suffer the consequences. Conversely, if neighbors are somewhat different, the next generation can benefit from being more variable.

### **How Do Animals Benefit?**

Seeds and fruit are important sources of food for all animal dispersers. Just as in the case of pollen and nectar, this food is dependable enough that many species of diverse animals, described below, have evolved to become seed and fruit consumers and, in some cases, are highly specialized on these diets.



## Seed Packaging

At their simplest, seeds are covered by a hard envelope that protects them from fluctuations in temperature and moisture. If they have no structural modifications, they tend to be round or ovoid. Such seeds will simply fall to the ground when the structure within which they develop matures and crack as it dries up. Obviously if they fall, they have not traveled far from the mother plant or from each other, and, as noted above, this is often problematic. There are situations when such limited dispersal is useful, and they often involve plants adapted to live in very specific environments. For example, plants living in temporary pools surrounded by dry habitats restrict their dispersal to those pools. Plants that live on islands have often evolved reduced dispersal abilities because it does not pay to get dispersed into salt water. However, in general, there has been strong selection favoring devices that help the seeds travel away from mother plants. Solutions to the challenge of dispersal come in many shapes. These include having some way to exploit wind or water currents.

To get dispersed by animals, seeds must either attach themselves to a disperser or offer a reward. Those that attach themselves tend to do so with hooks, bristles, and barbs or have adhesive surfaces. A look at one's socks after a walk through a dense grassland or a weed patch illustrates the effectiveness of such mechanisms. Dwarf mistletoes of the genus *Arceuthobium* employ a different method. Their seeds are inside fruit that at maturity are very sensitive to touch. When touched, they explode and send sticky seeds out at a speed approaching 100 km/h. These seeds then travel along, attached to the visiting bird or mammal until they are rubbed off.

Rewards come in two major categories. The most common ones are in the form of fleshy fruits, which often have bright, visually attractive coloration. In some plants, the seeds alone are large enough to be attractive to animals that collect them, transport them to specific locations and cache them for future consumption. Examples of such large seeds include the oaks, chestnuts, walnuts, hazelnuts, pistachios, pinon pines, and their relatives (Fenner and Thomson 2005).

## Dispersal Agents: Who Is Involved?

Beyond the broad categories below, there are no tidy groupings as there are in pollination, because most seeds and fruit are routinely dispersed by multiple animal species. As a result, spatial patterns are complex (Levine and Murrell 2003).

### Wind

Wings attached to the seeds offer one solution and are common in groups as diverse as grasses, conifers, ashes, and maples. Other devices that exploit wind are parachutes such as the ones found in dandelions or plumes of various shapes and many other daisy relatives. Seeds can also be so small that they behave like dust particles and are dispersed by the slightest breeze. Orchids use that solution.

## Water

Coconuts provide a perfect example of the benefits of water for long-distance dispersal: they are buoyant and protect their seeds from seawater well enough to stay viable for many months. As a result they can colonize shores far distant from their place of origin. They are pantropical in distribution, and it has been suggested that they achieved this long-distance travel entirely on their own. Of course they are also eminently edible, so human-aided movement on boats may have also been important. It is sure that no one can agree on their geographical origin.

Water-mediated dispersal is uncommon and poorly studied. Species that rely on it are few, with the notable exception of trees and shrubs called mangroves, which belong to some 20 different families and have all adapted to life in coastal wetlands. As such they are very important, for they provide the structural frameworks for coastal ecosystems in the tropics and subtropics. In most mangroves, the seeds germinate while still attached to the plants, so that the units that are dispersed are actually seedlings (Kathiresan and Bingham 2001).

## Ants

They are the only insects involved on a regular basis as seed dispersers, and they play important roles in some specific settings. In temperate habitats of the Northern hemisphere, some 300 plant species depend primarily on ants. They tend to be herbs of the forest floor such as anemones, cyclamens, trilliums, and violets. In contrast, in Australia and South Africa, the plants tend to be shrubs inhabiting dry sclerophyllous fire-prone woodland. To attract the ants, the plants produce food bodies called elaiosomes that are attached to the seeds. Ants transport these items to their nests, where they eat the elaiosomes and discard the seeds. These seeds fall on refuse piles, which provide more nutrients than surrounding soils. The germinating seedlings thus get an extra boost in their early life (Gomez and Espadaler 2013).

## Vertebrates

The species involved in seed dispersal are a remarkably diverse array of vertebrate groups. Birds, rodents, and bats are the most frequent contributors. At this time, it is impossible to ascertain the exact proportion of species in these various groups that are involved as seed dispersers. However, it is estimated that over 1/3 of terrestrial bird species eat fruit and about 1/5 of terrestrial mammals do so. Some important seed dispersers are unexpected as they include various carnivores such as maned wolves, coyotes, foxes, jackals, and even tigers. Figure 5 illustrates the diversity of fruit and seed shapes and sizes seen at just one location in a tropical forest in Ecuador. These fruits and seeds will be dispersed by many species of birds and mammals.

In addition, in riparian habitats especially in the tropics, but in other regions as well, fishes are important fruit dispersers and have been so for a very long time, perhaps since the Paleozoic. Indeed, it may be that they were the first vertebrates to act as seed dispersers. At least 275 species are frugivorous. They belong to various groups including piranhas, catfishes, carps, and minnows. They are especially important in Neotropical forests that are periodically flooded such as the Brazilian



**Fig. 5** Diversity of seed and fruit types in a forest in Ecuador. The plant genera represented include *Spondius* sp. (Anacardiaceae, mango and sumac family), *Schefflera* (Araliaceae, ivy and ginseng family), *Raffia* (Arecaceae, palms), and *Guarea* (Meliaceae, mahogany family). Other families include the Annonaceae (sweetsop family), Araceae (anthurium – Fig. 1 – family), and Lauraceae (laurel family) (Photo and information courtesy of K.M. Holbrook)

Pantanal where they help disperse a large number of shrubs and trees and can be the principal dispersers for many trees (Galetti and Gouling 2011). Other inhabitants of riparian regions can also be useful to plants in that regard. For example, among the Crocodylia, at least 13 species are documented as seed dispersers, and they consume seeds or fruits in at least 46 genera belonging to 34 families (Platt et al. 2013). Lizards and turtles are also known to get involved in seed dispersal, especially on islands (Olesen and Valedo 2003), and one species of frog has been reported as a frugivore so far (da Silva et al. 1989).

## Fruit Characteristics

### Visual

Ripe fruits dispersed by birds are often brightly colored, and the predominant colors are red and black, and often reflect ultraviolet light. In contrast, mammal-dispersed fruit are not as visually striking and tend to have more subdued colors. The fact that colors can promote specificity of dispersers is illustrated by the bright red color of ripe chili peppers. The red color is produced by capsaicins which attract the birds

and have no effects on their palates. Conversely, they produce a memorably spicy burning sensation in mammals. This helps keep mammalian frugivores away from the fruit (Schulze and Spiteller 2009).

### **Olfactory**

Mammals have more highly developed olfactory abilities than birds, so detectable smells are often important features of mammal-dispersed fruits. Bat-dispersed fruit including various figs are best known in that context. This area of plant-disperser interactions is poorly understood and a wide-open field for study. It is sure that certain fruits such as durians which live in the forests of Malaysia and Indonesia are very good advertisers in that context. Their taste is heavenly, but the smell they emit has been described as fermenting dirty socks with elements of onions and leaking gas and a background hint of long-dead corpse. Given the wonderful complexity of this bouquet, it is not surprising that they attract diverse local denizens including tigers, elephants, and monkeys that disperse them with gusto. See, for example, a video of a tiger checking out a durian (*Sumatran tiger inspects durian fruit on forest floor* [www.arkive.org](http://www.arkive.org) > *Species* > *Mammals* > *Tiger*). As for ginkgo, there is no idea what animals might have dispersed its stinky fruit. Unfortunately, it is now a living fossil that grows only in urban habitats. It has disappeared from the wilds of Asia but has left enough fossils around so it is well known that it was a forest dweller about 200 million years ago. Whether it was dispersed by dinosaurs or small mammals, or both, will never be known.

### **Size and Nutritional Value**

There is a large range of sizes in fruit, and there are very general patterns of variation between geographic regions. Thus, in multiple families, fruits are on average larger in the tropics of Asia and Africa than in the Americas, presumably because of the absence of larger seed dispersers in the latter regions. As for fruit composition, there is also a large range of variation of fruit composition: some (e.g., think cherries, apples, or citrus) are mostly carbohydrates and offer little reward to the dispersers beyond a quick energy boost, while others are rich in proteins (e.g., avocado, guava, dates) and lipids (e.g., olives, as well as magnolia, dogwoods (*Cornus* spp.), and Virginia creeper (*Parthenocissus*)) (Johnson et al. 1985).

### **Patterns of Dispersal**

The diversity of seed dispersal mechanisms is clearly the outcome of strong evolutionary pressures generated by the advantages listed above. However, the behavior of frugivores following ingestion is highly variable and often poorly known: as a result, the shapes of dispersal distributions away from sources are erratic and usually not quantified. In addition, the diversity of dispersers in natural landscapes, especially in the tropics, is very high, so studies that focus on just a few species cannot provide an accurate picture of dispersal patterns (Table 1).

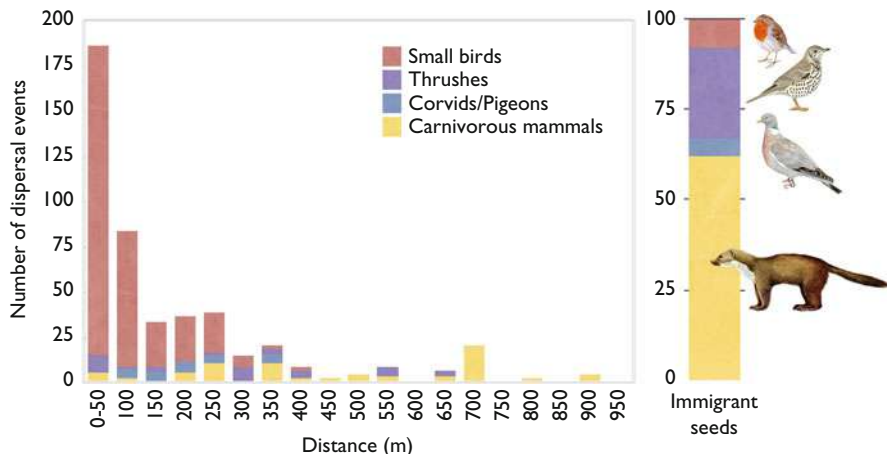
**Table 1** Routine maximum seed dispersal distances achievable by various combinations of plants and dispersal agents

Distance	Vector (propagule type)
0–10 m	Mechanical
	Ants
10–100 m	Wind (large winged fruits)
	Rodents
	Macaques (large seeds, not swallowed)
100 m–1 km	Small- and medium-sized forest passerines
	Fruit bats (large seeds)
	Most primates (seeds swallowed)
1–10 km	Large canopy birds
	Open-country passerines
	Small fruit bats (tiny seeds)
	Orangutans
	Carnivores, including civets, martens, and bears
	Most terrestrial herbivores
>10 km	Wind (tiny seeds), water
	Fruit pigeons
	Large fruit bats (tiny seeds)
	Elephants, rhinoceroses
	People

From Corlett 2009

For this reason, until recently, the only species for which reasonable data on dispersal patterns was available were for wind-dispersed seeds collected from containers at varying distances from sources. For such species, the patterns are straightforward: the distributions tend to be leptokurtic, and small numbers of seeds can travel several km. For animal-dispersed species, distributions definitely show attenuation with distance, but over those distances, they are often very patchy (Cain et al. 2000). However, the use of DNA microsatellite analysis is documenting the complexities of such dispersal very nicely. For example, Jordano et al. (2007) have found that both birds and mammals disperse the fruit of *Prunus mahaleb*; small birds tend to move seeds shorter distances and into covered habitats, while mammals move them longer distances and into open areas and also account for about two-thirds of introduction of immigrant seeds into populations (Fig. 6). Other studies document long-distance patterns and show that the tails of distributions can be much longer – up to several km – than previously thought (Ashley 2010).

One important question is the extent to which the dispersers actually deliver the seeds to locations where the seeds can get established. These are often known as *safe sites*, and such dispersal to useful locations is usually called *directed dispersal* (Wenny et al. 2011). Such dispersal is becoming recognized as an important alternative to the notion that all seeds are dispersed as clouds over the landscape. The best understood examples in order of the number of plant species involved include (1) ant-dispersed plants with seeds attached to elaiosomes. The ants transport the seeds to their nest areas. At least 3000 species in 60 plant families have



**Fig. 6** Frequency distribution of seeds of the cherry *Prunus mahaleb* dispersed by small birds including warblers *Sylvia* spp., and robins *Erithacus rubecula*; thrushes (*Turdus* spp.); large birds, including pigeons *Columba*, and corvids such as carrion crows *Corvus corone*; and carnivorous mammals including red fox *Vulpes vulpes*, marten *Martes foina*, and badger *Meles meles* (Adapted from Jordano et al. 2007)

evolved this strategy, so obviously it works. (2) Mistletoes of several families. These mistletoes are parasites and must establish themselves on the branches of their host woody plants. The seeds are very sticky, so that whether they are dispersed externally by attaching to plumage or fur or ingested, many will tend to stick to the rough bark of their hosts and will not land on unsuitable sites such as soil or leaves. (3) Several pines that produce large seeds that attract corvids such as nutcrackers and jays. The birds, often called scatter hoarders, collect seeds and bury them in areas away from the trees where they collected them but in habitats suitable for the next generation of trees (Tomback and Linhart 1990).

Other scenarios that fit the pattern of directed dispersal include activities within gaps in closed-canopy forests. These gaps admit more light to the forest floor; therefore, plants can germinate more readily, grow faster, and produce flower and fruit more often. Hence, frugivorous birds and mammals tend to visit gaps frequently since their foods can be found in such habitats more predictably. After feeding they can move off to other gaps and drop off the seeds in their feces, thus promoting dispersal to suitable habitats. Even wind-dispersed species can end up preferentially in gaps because of turbulence associated with those openings in canopies (Wenny (2001), but see Puerta-Pinero et al. (2013) for a different perspective).

In arid areas, soil surfaces tend to be inhospitably hot, dry, and/or windy. In contrast, conditions around established plants are shadier and often moister. That is why such plants are referred to as nurse plants, and they often serve as places where birds or small mammals come to rest and can deposit seeds which find more suitable sites for germination and survival than in open habitats.

Secondary dispersers can also promote directed dispersal of seeds to preferred habitats. Thus, even when plants such as pines are wind dispersed, small mammals, such as chipmunks and ground squirrels, then pick up seeds and bury them, often in somewhat sheltered places, and these buried seeds have a much higher probability of germinating and producing live progeny than do seeds that land on the forest floor. In addition certain mammals such as monkeys and tapirs have latrine sites, and smaller mammals and dung beetles can pick up seeds at those sites and bury them in places where they are more likely to thrive.

From these descriptions, it seems that the overall pattern of distribution after dispersal from a specific source for many species is leptokurtic, with many seeds remaining relatively near the origin and with a long tail that has bumps wherever there are local aggregations of seeds. These distributions will influence the genetic constitution of the populations produced. For example, whenever there is aggregation of seeds with some genetic relatedness, there will be patches of such individuals in the adult populations. Such patches have been detected in both wind-dispersed taxa, such as pines and eucalyptus, and animal-dispersed taxa, such as figs and *Cecropia* (Hamrick and Trapnell 2011).

## Evolutionary Dynamics: Evolution in action

Given the central importance of seed dispersal for the survival of plants, it is no surprise that mechanisms that alter dispersal patterns are evolutionarily flexible. There is evidence for rapid evolution of altered dispersal in settings where such dispersal is counterproductive. For example, in urban environments, the weed *Crepis sancta* grows in small patches of soil surrounded by inhospitable concrete. This species produces two kinds of seeds; some are dispersers, while others are non-dispersers. In these urban environments, dispersing seeds have a much lower probability of reaching suitable habitats, so the pressure is strong to produce non-dispersers who stay close to home. Cheptou et al. (2008) found that in about 5–12 generations, urban populations were evolving towards reduced dispersal. Populations of several invasive species that reached islands in the Pacific Northwest have also evolved towards reduced seed dispersal in about five generations (Cody and Overton 1996).

## Seed Dispersal Is Just Part of the Story

Just as in pollination scenarios, plants face conflicting selection pressures in the context of seed production and dispersal. Seeds represent concentrated packages of carbohydrates, proteins, and lipids designed to nourish developing seedlings. No wonder much of our basic nutrition is based on seeds and grains. And no wonder that thousands of species of all manner of seed parasites and consumers, from fungi to insects to birds to mammals, have caught on to that fact and have evolved to focus on seeds. So, once again, plants must adapt to these hordes. At the same time, they must put out large enough numbers of fruits and seeds to attract dispersers.



One solution that plants have evolved is to protect these valuable packages with various toxic compounds (see Chap. 6, ► “[Evolutionary Ecology of Chemically Mediated Plant-Insect Interaction](#)”). Another is to put out very large numbers of seeds simultaneously so as to overwhelm the seed consumers, but do so on an irregular basis so that the consumers cannot track those bonanzas. This phenomenon is known as *masting*, and for any one species in any one location, masting episodes occur every few years. Many species of trees and other perennials follow this pattern, which also has the advantage of synchronizing flowering thus improving the probability of pollination and outcrossing for all members of the participating population (Kelly and Sork 2002).

---

## Synthesis and Conclusions

### The Systematics of Associations

Adaptation to a specific pollination or dispersal mode does not occur at the family level. For example, even within small families such as the Brazil nut family (Lecythidaceae) with about 300 species, various species are pollinated by birds, bats, and/or insects. As for dispersal, their woody fruits are adapted for dispersal by primates, birds, fish, and even wind and water.

In general terms, while there are a few situations where a whole family, e.g., Pinaceae or Poaceae, are wind pollinated, this is uncommon, and especially in the context of animal pollination, specialization typically occurs at the genus level at most: for example, the genus *Ficus* has a close association with wasps that belong to several families and *Heliconia* depends primarily on hummingbirds and *Yucca* on moths. That is, there is no plant family where the whole family is narrowly adapted to a small group of related pollinators. Instead, what one typically sees within a family is a diversity of pollination syndromes, with some species attracting bees, others butterflies, and others yet moths or flies or other groups. Seed dispersal follows the same pattern. A few families such as Fagaceae produce big, rewarding seeds that get collected by various animals and buried, but this seems unusual, as even among Pinaceae where the bulk of species are wind dispersed, some species such as Pinon pines and their relatives are bird dispersed; among grasses (Poaceae) and daisies (Compositae), you get both wind and animal dispersal.

### Effects of Dispersal Patterns and Ecological Heterogeneity on Genetic Organization of Populations

When considering the genetic structure of plant populations, it is most useful to visualize mosaics. Analyses of genetic patterns show that alleles and genotypes are usually not homogeneous in their distributions within or among populations. This indicates that dispersal of pollen and/or seeds can be limited and generates clusters of genotypes within populations. For example, if at least some of the seed dispersal



is limited, then populations can consist of family groups of related genotypes. Then, if pollen exchange is also among neighbors, this will produce strong genetic heterogeneity. Such outcomes have been observed in several species. The scale of these mosaics depends on the biology of individual species: in clonally reproducing species, individual genotypes have multiple stems and span several meters or more in diameter. In annual plants and many forest trees, individual genotypes have single stems but they usually also show genetic patchiness because genetically related seeds are often clustered (e.g., Figs. 3 and 4 and also Hamrick and Trapnell 2011).

In nature, these populations occupy heterogeneous habitats that vary in physical conditions, including moisture and nutrients, and biotic conditions such as competition, pollination, and herbivory. One analysis that illustrates how a mosaic pattern can be generated by interactions between these multiple selection pressures is provided by Gomez et al. (2009). They describe what they call a geographic selection mosaic in the species *Erysimum mediohispanicum* (Brassicaceae). They studied eight populations and quantified patterns of selection imposed upon different populations by pollinators and herbivores. The mean interpopulation distance was about 800 m, but some populations were about 200 m. apart. They found that different populations were pollinated by insects with different characteristics and behaviors. These included flies, large bees, small bees, and beetles. The primary herbivores were wild Spanish ibex (*Capra pyrenaica*) and domestic sheep; the intensity of damage they inflicted varied among populations. As a result of this variation in pollination and herbivory, populations were exposed to variable types and intensity of selection, which produced significant interpopulation differentiation in several traits. Some traits including flower features such as tube length and corolla diameter were under selection pressures in some populations and not others. Other features of corolla shape including tube width and overall shape showed evidence of diversifying selection. These variable patterns mean that some populations are under intense selection, which they call “hotspots,” while they refer to others where selection is less intense as “cold spots.” This sort of study shows that even when the organisms involved in pollination and herbivory are generalists, they can produce intense selection, and the selection mosaics can operate at small scales.

While populations often consist of genetic mosaics, there is still connection among those patches. The use of genetic analyses, and especially DNA microsatellites, is providing important insights into such connections by analyzing movement of plant genes across landscapes via pollen and seeds. It is now being learned that both can be dispersed across hundreds of meters or more in natural ecosystems so that genetic neighborhoods can be much larger than earlier estimates based on movement alone. Some impressive examples include several reports of wind pollination in *Populus* and *Pinus* across several km and up to 80 or more km in *Ficus* pollinated by wasps which are themselves carried by prevailing winds. This means that gene flow can be a strong homogenizing force, even in landscapes where populations are relatively isolated or fragmented, or where solitary individuals are very distant from conspecifics (Ashley 2010).

## Pollination, Dispersal, and Human Activities

### Protection Needed by Pollinators and Dispersers

Consider this: without the evolutionary driving force generated by interactions between animals and flowering plants, many of the seeds and fruits that make up nearly 80 % of the human diet would not exist. Since much of our food comes from plants, and is therefore dependent on pollination and seed dispersal, we need to be decent stewards of our ecosystems in order to feed ourselves. So far, our record is not too good. Fully 50 years ago, Rachel Carson warned us in *Silent Spring* that our pollinators and fruit dispersers were imperiled. Seventeen years ago, Stephen Buchmann and Gary Nabhan (1996) reminded us in *The Forgotten Pollinators* that a great diversity of animals work for us in those roles. The problems continue and are getting worse. In a recent review, Potts et al. (2010) describe the current situation of pollinator declines. Honeybees have been introduced all over the world because of their efficiency of their service: fully 96 % of crops that are pollinated by this species show increased yields when serviced by honeybees compared to other insects. As a result, we have become hugely dependent on their good services. Their numbers have been declining as a result of parasites and poor management. That is why the difficulties faced by honeybees and the crops they pollinate are worthy of serious concern. Their recent declines are spectacular (e.g., 59 % loss of colonies in the USA between 1947 and 2005 and 25 % loss of colonies in central Europe between 1985 and 2005 according to their figures). Potts et al. also stress that while much remains to be learned about honeybee declines, even less is known about the status of wild pollinators. One example they provide comes from work with bumblebees (*Bombus*) in the UK, where 6 of 16 nonparasitic bumblebees have declined significantly in the past decades (including *B. subterraneus* which has become extinct) and another 4 may be in trouble. They go on to argue that coordinated and standardized monitoring programs are urgently needed. The same goes for fruit dispersers who are also declining. For example, Sekercioglu et al. (2004) warned that globally, a quarter or more of fruit-dispersing birds were extinction-prone. Many natural ecosystems are also vulnerable to human-induced changes such as climate change and are already showing signs of stress (Corlett 2009).

### Evolutionary Dynamics in Agronomic Ecosystems

The activities of pollinators and seed dispersers also influence evolution in ways that make our lives difficult. Two examples will be noted: the evolution of herbicide resistance in weeds and the unwanted spread of genes involved in GMO crops.

At this time, over 200 species – and growing – are reported to be resistant to herbicides (<http://www.weedscience.org>) and the numbers keep increasing. There is resistance to all known types of herbicides (Delye et al. 2013). One reason why this has become such a serious problem is because many crop species have weeds as close relatives. For example, a quick survey of the list of resistant weeds includes several species of *Raphanus* (relatives of beets, radishes, and cabbages), *Solanum* (relatives of potatoes, tomatoes, and eggplants), and *Avena* (relatives of oats), as well as weedy versions of rice, sorghum, sunflowers, carrots, and the list goes on.

The issue is that these weeds can and do exchange genes with their cultivated relatives. It must be borne in mind that in today's agriculture, many crops have been bred to be herbicide-resistant themselves. The rationale is that if crop plants are herbicide resistant, then herbicides can be used in crop fields with impunity to control the weeds, which is much cheaper than other means of weeding. Beautiful logic, until biology intervenes. The problem was predicted by the work of the botanist Jack Harlan who was the first to draw attention to the fact that crop plants often grew in close proximity to weeds that were close relatives. For example, he observed that in Mexico and Central America, maize grew in the company of its ancestor and competitor, teosinte. In Africa, he saw cultivated and weedy sorghum in close association, in Asia cultivated and weedy rice grew side by side, and so on. These observations coupled with the recognition that these weeds and crops could interbreed led him to formulate the concept of the "compilospecies" which posited that whenever groups of species were closely related, they could exchange genes and thereby compile useful information. In retrospect, it is no wonder that herbicide resistance has evolved so quickly in so many species. We have helped the process along: we have introduced genes for herbicide resistance into crops, whose pollen and seeds move about, sometimes great distances (as per Ashley 2010), and help pass on those genes to weedy relatives.

As for the escape of transgenes, this possibility was brought up at least two decades ago. So far, it seems that relatively few transgenes have ended up in wild populations, but still, thanks to unexpected dispersal of pollen and/or seeds, they are found in settings where they were not intended to be (Ellstrand 2012). One issue is that escape into wild populations is not the only problem. Escape of multiple transgenes into populations free of such genes is another. This is happening in Mexico. This is very problematic given the dependence upon maize as a food crop in humans worldwide and because Mexico is the original home of maize and the center of diversity of this species; at least 60 distinct land races adapted to very different ecological conditions, and several wild relatives of maize are unique to the country. This genetic diversity represents a very important reservoir for future breeding of maize. The majority of maize fields in Mexico are small, family enterprises, and seeds are usually replanted within the area where they were produced. This method contributes to the maintenance of these land races. The accidental introduction of foreign transgenes into such varieties can disrupt the integrated nature of their genomes. If one imagines that the genome of a variety is like a blueprint that guides its construction, the sudden introduction of new components into the design can alter the appearance and/or function of the finished product. In addition, the blueprints are no longer useful for future work. For these reasons, there was concern about the introduction of genetically engineered corn in Mexico, and a moratorium on such introduction was put in place in 1998. Despite this moratorium, transgenes have been detected in native populations, and the consequences of these careless introductions are being assessed (Pineyro-Nelson et al. 2009).

As for transgenes for herbicide resistance, they are also spreading in our landscapes and creating problems as illustrated in this case study. Creeping bent grass (*Agrostis stolonifera*) is commonly used in golf courses. In 2002, a version of this

species carrying genes for resistance to the herbicide glyphosate (aka Roundup<sup>®</sup>) was planted by the Scotts Company on 162 ha in Oregon. Wind-dispersed pollen carrying the resistance genes moved from that population and fertilized ovules of two local species (*A. stolonifera* and *A. gigantea*), and the hybridizations occurred on sentinel plants as far as 21 km away. In addition, winds helped move transgenic seeds into nearby areas. Recently the situation has become more complicated because of the detection of an intergeneric hybrid which carries the transgenes and consists of a combination of the bent grass with rabbit-foot grass (*Polypogon monspeliensis*) (Snow 2012).

Overall, given the warning provided by Ashley (2010) about the fact that long-distance gene flow via pollen and seed is much more prevalent than we thought, the message is clear. . . there are problems afoot.

## Future Directions

There are over 250,000 species of flowering plants in the world (80–90 % are pollinated by animals) and another 1,000 or so species of non-flowering plants that disperse by seed. There are probably well over 30,000 species of animals involved in the tasks of pollination or seed dispersal. It is no wonder that we still have much to learn about the interactions. The issues that are especially poorly known are noted in the text and summarized below:

- The variability of the color spectra and UV nectar guides produced by plants to attract animals and the ability of various animals to detect those signals. In more general terms, the intricacies of visual and chemical communication in the contexts of pollination, seed dispersal, and herbivory deserve greater attention (Schaefer and Ruxton 2011).
- The extent to which shifts in signals, especially olfactory and visual ones, can produce shifts in pollinator visitation patterns and resulting gene flow and population differentiation is also open to question.
- Animal-mediated seed dispersal outside of the temperate zones of North America and Europe is a very open field, both in the tropics and in the Southern Hemisphere. Even within temperate areas, we still have a very limited understanding of the ecosystem services that birds provide (Wenny et al. 2011). Large vertebrate dispersers are major contributors to seed dispersal networks, especially in the tropics (Table 1), but much of our information is anecdotal. Their real contributions are poorly known because they are difficult to study and can be very rare. In addition, some are already missing from some ecosystems (Corlett 2009; Vidal et al. 2013).
- This introduction to pollination and dispersal should be used in combination with the discussion of herbivory (Chap. 6, ► “[Evolutionary Ecology of Chemically Mediated Plant-Insect Interaction](#)”) and biodiversity and population dynamics (Chap. 2, ► “[Plant Biodiversity and Population Dynamics](#)”) to get a synthetic understanding of linkages among populations, metapopulation dynamics in space and time, and long-term dynamics.

## Summary

Plants are stationary and depend on external agencies to help them reproduce and disperse their seeds. Most plant species utilize animal pollinators and seed dispersers, although in specific ecosystems some plants can use wind or water for such transport.

To attract these animal vectors, plants use various food rewards including pollen, nectar, seeds, and fruits.

In terms of species numbers, the majority of pollinators are insects, and the majority of seed dispersers are vertebrates.

The genetic structure of plant populations is strongly influenced by their pollen and seed vectors. When wind is the dispersing agent, pollen and seed movement are relatively straightforward and can be described by leptokurtic distributions, with most of the pollen grains or seeds transported short distances and tails extending long distances away from the source. When the dispersal is by animals, the behaviors of individuals and species are so variable as to render generalizations difficult.

The ecology and evolution of plants is not just about plants: animals are important actors in those plays. Consequently, the effects of humans upon pollinators and seed dispersers should influence management decisions in natural and agronomic ecosystems.

---

## References

- Adler LS, Irwin RE. What you smell is more important than what you see? Natural selection on floral scent. *New Phytol.* 2012;195:510–1.
- Ashley MV. Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. *Crit Rev Plant Sci.* 2010;29:148–61.
- Barrett SCH. The evolution of plant sexual diversity. *Nat Rev Genet.* 2002;3:274–84.
- Buchmann SL, Nabhan GP. *The forgotten pollinators.* Washington D.C.: Island Press; 1996.
- Cain ML, Milligan BG, Strand AE. Long-distance seed dispersal in plant populations. *Am J Bot.* 2000;87:1217–27.
- Cheptou PO, Carrue O, Rouifed S, Cantarel A. Rapid evolution of seed dispersal in an urban environment in the weed *Crepis sancta*. *Proc Natl Acad Sci.* 2008;105:3796–9.
- Cody ML, Overton JM. Short-term evolution of reduced dispersal in island plant populations. *J Ecol.* 1996;84:53–61.
- Corlett RT. Seed dispersal distances and plant migration potential in tropical East Asia. *Biotropica.* 2009;41:592–8.
- da Silva HR, de Britto-Pereira MC, Caramaschi U. Frugivory and seed dispersal by *Hyla truncata*, a neotropical treefrog. *Copeia.* 1989;781–3.
- Delye C, Jasieniuk M, Le Corre V. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* 2013. In press.
- Dodd RJ, Linhart YB. Reproductive consequences of interactions between *Yucca glauca* (Agavaceae) and *Tegeticula yuccasella* (Lepidoptera) in Colorado. *Am J Bot.* 1994;81:815–25.
- Ellis AG, Johnson SD. Floral mimicry enhances pollen export: the evolution of pollination by sexual deceit outside of the Orchidaceae. *Am Nat.* 2010;176:E143–51.
- Ellstrand NC. Over a decade of crop transgenes out-of-place. In: *Regulation of agricultural biotechnology: the United States and Canada.* The Netherlands: Springer; 2012. p. 123–35.

- Fenner M, Thomson K. The ecology of seeds. New York: Cambridge University press; 2005.
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. Pollination syndromes and floral specialization. *Annu Rev Ecol Evol Syst.* 2004;35:375–403.
- Galen C. Measuring pollinator-mediated selection on morphometric floral traits: bumblebees and the alpine sky pilot, *Polemonium viscosum*. *Evolution.* 1989;43:882–90.
- Galen C. Why do flowers vary? *BioScience.* 1999;49:631–40.
- Galetti M, Goulding M. Seed dispersal by fishes in tropical and temperate fresh waters: the growing evidence. *Acta Oecol.* 2011;37:561–77.
- Gaskett AC. Orchid pollination by sexual deception: pollinator perspectives. *Biol Rev.* 2011;86:33–75.
- Gómez C, Espadaler X. An update of the world survey of myrmecochorous dispersal distances. *Ecography.* 2013;36:1193–1201.
- Gomez JM, Perfectti F, Bosch J, Camacho JPM. A geographic selection mosaic in a generalized plant–pollinator–herbivore system. *Ecol Monogr.* 2009;79:245–63.
- Hamrick JL, Trapnell DW. Using population genetic analyses to understand seed dispersal patterns. *Acta Oecol.* 2011;37:641–9.
- Hu S, Dilcher DL, Jarzen DM, Winship D. Early steps of angiosperm–pollinator coevolution. *Proc Natl Acad Sci.* 2008;105:240–5.
- Irwin RE, Strauss SY, Storz S, Emerson A, Guibert G. The role of herbivores in the maintenance of a flower color polymorphism in wild radish. *Ecology.* 2003;84:1733–43.
- Johnson RA, Willson M, Thomson JN, Bertin RI. Nutritional values of wild fruit and consumption by migrant frugivorous birds. *Ecology.* 1985;66:819–27.
- Jordano P, Garcia C, Godoy JA, García-Castaño JL. Differential contribution of frugivores to complex seed dispersal patterns. *Proc Natl Acad Sci.* 2007;104:3278–82.
- Karron JD, Ivey CT, Mitchell RJ, Whitehead MR, Peakall R, Case AL. Viewpoint: part of a special issue on plant mating systems. *Ann Bot.* 2012;109:493–503.
- Kathiresan K, Bingham BL. Biology of mangroves and mangrove ecosystems. *Adv Mar Biol.* 2001;40:81–251.
- Kay KM, Sargent RD. The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. *Annu Rev Ecol Evol Syst.* 2009;40:637–56.
- Kearns CA, Inouye DW. Techniques for pollination biologists. Niwot: University Press of Colorado; 1993.
- Kelly D, Sork VL. Mast seeding in perennial plants: why, how, where? *Annu Rev Ecol Syst.* 2002;33:427–47.
- Kessler D, Diezel C, Clark DG, Colquhoun TA, Baldwin IT. *Petunia* flowers solve the defence/apparency dilemma of pollinator attraction by deploying complex floral blends. *Ecol Lett.* 2013;16:299–306.
- Levine JM, Murrell DJ. The community-level consequences of seed dispersal patterns. *Annu Rev Eco Syst.* 2003;34:549–74.
- Linhart YB. Ecological and behavioral determinants of pollen dispersal in hummingbird-pollinated *Heliconia*. *Am Nat.* 1973;107:511–23.
- Llandres AL, Gawryszewski FM, Heiling AM, Herberstein ME. The effect of colour variation in predators on the behaviour of pollinators: Australian crab spiders and native bees. *Ecol Entomol.* 2011;36:72–81.
- Olesen JM, Valedo A. Lizards as pollinators and seed dispersers: an island phenomenon. *Trends Ecol Evol.* 2003;18:177–81.
- Ollerton J, Winfree R, Tarrant S. How many flowering plants are pollinated by animals? *Oikos.* 2011;120:321–6.
- Patel A, Hossaert-Mckey M, Mckey D. Ficus-pollinator research in India: past, present and future. *Curr Sci.* 1993;65:243–53.
- Patiny S. Evolution of plant-pollinator relationships. Cambridge: Cambridge University Press; 2011.

- Pineyro-Nelson A, Van Heerwaarden J, Perales HR, Serratos-Hernandez JA, Rangel A, Hufford MB, Álvarez-Buylla ER. Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. *Mol Ecol*. 2009;18(4):750–61.
- Platt SG, Elsey RM, Liu H, Rainwater TR, Nifong JC, Rosenblatt AE, Mazzotti FJ. Frugivory and seed dispersal by crocodilians: an overlooked form of saurochory? *J Zool*. 2013;291:87–99.
- Potts SG, Jacobus C, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE. Global pollinator declines: trends, impacts and drivers. *Trends Ecol Evol*. 2010;25:345–53.
- Primack RB. Ultraviolet patterns in flowers, or flowers viewed by insects. *Arnoldia*. 1982;42:146–59.
- Proctor M, Yeo P, Lack A. The natural history of pollination. London: Collins New Naturalist Library; 2012.
- Puerta-Piñero C, Muller-Landau HC, Calderón O, Wright SJ. Seed arrival in tropical forest tree fall gaps. *Ecology*. 2013;94:1552–62.
- Richards AJ. Apomixis in flowering plants: an overview. *Philos Trans R Soc Lond B Biol Sci*. 2003;358:1085–93.
- Sapir Y, Armbruster SC. Pollinator-mediated selection and floral evolution: from pollination ecology to macroevolution. *New Phytol*. 2010;188:303–6.
- Schaefer HM, Ruxton GD. Plant-animal communication. Oxford: Oxford University Press; 2011.
- Schiestl FP. Animal pollination and speciation in plants: general mechanisms and examples from the orchids evolution of plant-pollinator relationships; 2011. [books.google.com](http://books.google.com)
- Schulze B, Spiteller D. Capsaicin: tailored chemical defence against unwanted “frugivores”. *ChemBioChem*. 2009;10:428–9.
- Şekercioglu ÇH, Daily GC, Ehrlich PR. Ecosystem consequences of bird declines. *Proc Natl Acad Sci*. 2004;101:18042–7.
- Sheehan H, Hermann K, Kuhlemeier C. Color and scent: how single genes influence pollinator attraction. *Cold Spring Harb Symp Quant Biol*. 2013;77:117–133.
- Shrestha M, Dyer AG, Boyd-Gerny S, Wong BBM, Burd M. Shades of red: bird-pollinated flowers target the specific colour discrimination abilities of avian vision. *New Phytol*. 2013;198:301–10.
- Simon R, Holderied MW, Corinna U, Koch CU, von Helversen O. Floral acoustics: conspicuous echoes of a dish-shaped leaf attract bat pollinators. *Science*. 2011;333:631–3.
- Snow AA. Illegal gene flow from transgenic creeping bentgrass: the saga continues. *Mol Ecol*. 2012;21:4663–4.
- Strauss SY, Irwin RE. Ecological and evolutionary consequences of multispecies plant-animal interactions. *Annu Rev Ecol Evol Syst*. 2004;35:435–66.
- Tomback DF, Linhart YB. The evolution of bird-dispersed pines. *Evol Ecol*. 1990;4:185–219.
- Turner ME, Stephens JC, Anderson WW. Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. *Proc Natl Acad Sci*. 1982;79:203–7.
- Vidal MM, Pires MM, Guimarães Jr PR. Large vertebrates as the missing components of seed-dispersal networks. *Biol Conserv*. 2013;163:42–8.
- von Helversen D, von Helversen O. Object recognition by echolocation: a nectar-feeding bat exploiting the flowers of a rain forest vine. *J Comp Physiol A*. 2003;189:327–36.
- Waser NM, Ollerton J, editors. Plant-pollinator interactions: from specialization to generalization. Chicago: University of Chicago Press; 2006.
- Wenny DG. Advantages of seed dispersal: a re-evaluation of directed dispersal. *Evol Ecol Res*. 2001;3:51–74.
- Wenny DG, Devault TL, Johnson MD, Kelly D, Sekercioglu CH, Tomback DF, Whelan CJ. The need to quantify ecosystem services provided by birds. *Auk*. 2011;128:1–14.
- Whitney H, Chittka L, Bruce T, Glover BJ. Conical epidermal cells allow bees to grip flowers and increase foraging efficiency. *Curr Biol*. 2009;19:1–6.

---

# Plant Phenotypic Expression in Variable Environments

# 5

Brittany Pham and Kelly McConnaughay

## Contents

Introduction .....	120
Phenotypic Plasticity Is a Particular Form of Variable Phenotypic Expression .....	121
Definition of Phenotypic Plasticity .....	121
Phenotypic Plasticity Is Not the Only Mechanism that Generates Variable Phenotypic Expression .....	121
Challenges in Defining Phenotypic Plasticity .....	122
The Particular Importance of Phenotypic Plasticity in Plants .....	123
Phenotypic Plasticity Is Often, but Not Always, Interpreted as an Adaptive Response to Variable Environments .....	124
Phenotypic Plasticity as Adaptive .....	124
Phenotypic Plasticity: A Highly Selected Trait or a Consequence of Selection for Multiple Phenotypes? .....	124
Theoretical Limits to Selection for Phenotypic Plasticity .....	125
Phenotypic Plasticity as Nonadaptive or Maladaptive .....	126
The Role of Phenotypic Plasticity in Evolution .....	126
Phenotypic Plasticity Versus Developmentally Programmed Changes in Phenotypic Expression .....	127
Techniques for Evaluating Phenotypic Plasticity .....	127
Norms of Reaction Characterize Phenotypic Expression for One or More Genotypes Across a Range of Environments .....	127
Use of Developmentally Sensitive (Common Size or Developmental Stage) Comparisons Versus Common Time or Age Comparisons .....	128
Growth Analysis and Allometric Approaches .....	133
Modular Growth as a Platform for Evaluating Phenotypic Plasticity in Plants .....	134
Selecting Methodological Approaches .....	136
Future Directions .....	138
References .....	138

---

B. Pham • K. McConnaughay (✉)

Department of Biology, Bradley University, Peoria, IL, USA

e-mail: [bpham@fsmail.bradley.edu](mailto:bpham@fsmail.bradley.edu); [kdm@fsmail.bradley.edu](mailto:kdm@fsmail.bradley.edu)



---

**Abstract**

- Phenotypic expression is the result of a complex interplay between an organism's genes and its environment.
- During growth and development, organisms undergo a programmed series of phenotypic changes. Phenotypic expression thus varies throughout growth and development, even when the environment is homogenous and static. This has been termed "ontogenetic drift."
- Phenotypic expression may also vary with environmental conditions. The ability to vary phenotypic expression in response to environmental conditions is known as "phenotypic plasticity."
- The ability of an organism to express variable phenotypes in heterogeneous environments has been thought to confer adaptive benefits that increase fitness. Plants, as immobile organisms, cannot relocate to more favorable environments; plant phenotypic plasticity could be under strong selective pressure in predictably variable environments.
- Plant growth rates and developmental trajectories are generally plastic; i.e., they frequently vary with local environmental conditions.
- Whenever environmentally induced plasticity in growth and development occurs, interpretations of phenotypic plasticity are confounded with changes in phenotypic expression associated with ontogenetic drift.
- Plant phenotypic plasticity should be evaluated in a developmentally explicit context. Phenotypic expression should be characterized in light of developmental trajectories of phenotypic change whenever possible.
- Comparing plant phenotypes at a common age versus a common developmental stage may result in incorrect conclusions regarding the nature of the observed phenotypic variation.
- Selection of methodological approaches to evaluate plant phenotypic expression should align with the hypothesis under investigation.

---

**Introduction**

Biologists have developed a small handful of unifying themes to explain the astonishing diversity of form and function exhibited by organisms. Phenotypic plasticity is one of those themes that continues to fascinate biologists from diverse backgrounds from ecologists and geneticists to developmental and evolutionary biologists. It is often a subject that students have difficulty grasping, for phenotypic plasticity is the result of the interplay between two distinct but interacting identities – the genetics of an organism and its environment – but is responsible for much of the intraspecific variation observed in ecological contexts. In this chapter, we will describe phenotypic plasticity, offer examples of how it can confer putative adaptive advantages for species in predictably variable environments, explore how phenotypic expression is facilitated and constrained by predetermined patterns of phenotypic expression throughout growth and development, and discuss methodological approaches to assessing phenotypic plasticity.

## **Phenotypic Plasticity Is a Particular Form of Variable Phenotypic Expression**

### **Definition of Phenotypic Plasticity**

Phenotypic plasticity is defined as the ability of an organism – with its singular genotype – to express a range of phenotypes depending on its environmental conditions (for an exhaustive list of definitions, see Whitman and Agrawal 2009).

Pigliucci (2001) begins dissecting phenotypic plasticity with a discussion of the relationship between genotype and phenotype, the basis of the concept of phenotypic plasticity. Students' initial exposure to the concepts of genotype and phenotype invariably begins with Gregor Mendel and how one gene produces one phenotype. However, in reality, genes do not operate independently from one another, and a single gene rarely codes for one and only one phenotypic trait; epistasis and pleiotropy are more ubiquitous than the “one-gene-one-phenotype” model suggests. Phenotypic plasticity adds yet another layer of complexity in that gene networks can act together to produce distinctly different phenotypes in different environments. Ultimately, Pigliucci concludes that the environment cannot be discarded as “noise.” Understanding how an individual responds in different environments is as integral a part of describing its characteristics, as is its color or age (Bradshaw 1965; Pigliucci 2001). That is not to say in heterogeneous environments a trait may not remain the same, indicating that the trait has no plasticity and is not under environmental influence (Bradshaw 1965).

The role of the environment in inducing the observed variation in phenotypic expression is critical to an assessment of phenotypic plasticity; variable phenotypic expression that is solely a consequence of genetics is not considered plasticity. Plasticity can be in response to a particular environment (e.g., low resource patch vs. high resource patch) or in response to a change in that environment over time (e.g., a pulse of resources made available in an otherwise low resource patch). Phenotypic traits for which plastic expression has been documented in plants include morphological (e.g., leaf shape, branching patterns), allocational (e.g., root to shoot mass ratios, leaf area ratios, reproductive effort), anatomical (e.g., cuticle thickness, palisade mesophyll depth, stomatal density), physiological (e.g., light saturated photosynthetic rates, basal metabolic rates), and biochemical (e.g., defensive chemical production, Rubisco contents) traits.

### **Phenotypic Plasticity Is Not the Only Mechanism that Generates Variable Phenotypic Expression**

With the advent of increasingly sophisticated molecular techniques, developmental biologists have exploded the myth that gene expression is a simple and predictable linear sequence of events that result in one and only one phenotypic variant for any unique combination of genetics and environment. Rather, a variety of biochemical processes at the cellular and molecular level include elements of

stochasticity, such that gene and protein expression, and thus trait development, can vary at least in part due to small accumulations of chance events (e.g., Yampolsky and Scheiner 1994).

Phenotypic variation that does not correlate with a specific genotype or specific environmental cue, but is the result of stochasticity in the biochemical processes involved in gene and protein expression and other cellular noise that occurs throughout development, is referred to as “developmental noise” (Bradshaw 1965). If developmental noise generates sufficient variation in phenotypic expression, genetically identical individuals grown in the same environment will exhibit different phenotypes. The resulting phenotype could be adaptive, maladaptive, or neutral depending on the environmental conditions (DeWitt and Scheiner 2004). For example, in times of stabilizing selection where a mean phenotype is more desirable, developmental noise might reduce fitness; alternatively, variable phenotypic expression may increase the probability that at least some members of a population are able to survive and reproduce in stressful or rapidly changing environments.

## Challenges in Defining Phenotypic Plasticity

Many have argued that the basic definition of phenotypic plasticity is too broad to be of utility. DeWitt and Scheiner (2004) note that, at the most basic level, all traits are in some way influenced by the environment, causing everything to fall under the realm of phenotypic plasticity. Plastic morphological responses are themselves the result of physiological changes; thus plasticity at one level is likely correlated – causal or not – with plasticity at another level (Bradshaw 1965; Whitman and Agrawal 2009). Much of the early work on plant phenotypic plasticity focused on observable changes in plant morphology – leaf size or shape, plant size, root size, etc. (e.g., Vogel 1968). These morphological changes often have functional significance that aid in important activities like light capture or nutrient acquisition that directly affect fitness (Sultan 1987). Increasingly, however, the term phenotypic plasticity has been used more broadly to include changes in biochemistry, physiology, and life history as every morphological change in response to environmental conditions has resulted from a change in physiology.

Additionally, one must consider the context of “environment.” To the ecologist, the environment is comprised of the abiotic and biotic surroundings external to the organism, while a physiologist examining phenotypic plasticity may define environment in terms of the surrounding cells, hormones, enzymes, etc. However, external environmental changes are thought to lead to more extensive effects on phenotypic plasticity than internal changes (Garland and Kelly 2006). One must note that the conditions to which plants are exposed should be strongly rooted in an ecological context – exposing a plant that lives in Death Valley to lethally cold temperatures probably is not relevant, while studying a plant exposed to elevated CO<sub>2</sub> levels has real-life significance.

Finally, Weiner (2004) argues that the definition of “trait,” the aspect of phenotype under evaluation as plastic or not, is itself too broadly defined and suggests that the fact that a given trait is measurable does not guarantee that the trait is relevant to the organism’s ecological persistence or evolutionary success.

In light of the difficulties in developing a clear and consistent definition of phenotypic plasticity, one should take care to understand the context in which researchers frame their individual questions. DeWitt and Scheiner (2004, p. 2) eloquently state:

Such breadth of scope reinforces the idea that a particular trait value as observed in a given environment always is a special case of a potentially more complex relationship. That is, specific phenotype-environment observations are a fraction of a multidimensional space. This view promotes in our thinking the constant and useful caveat that given phenotype distributions may only apply for the environment in which observation is conducted. Extrapolation beyond given conditions must be justified rather than assumed.

## The Particular Importance of Phenotypic Plasticity in Plants

Plants are often viewed as passive organisms, subject to the local light conditions and nutrients necessary for normal growth and development and in some instances subject to the whims of pollinators and animals for reproductive success. But when one considers the incredible degree of phenotypic plasticity that plants exhibit, they are clearly superior to animals in regard to fine-tuning their phenotype beyond expressing a narrowly fixed set of traits dictated by genetic and developmental constraints. Animals have a fixed body plan that follows strict developmental trajectories and allometries (Wu et al. 2003) – gametes are either male or female, cell types become differentiated for highly specialized, irreversible functions, etc. In contrast, plant development is continuous, organogenesis in particular occurs throughout growth and development, and plant cell fates are less determinate (Walbot 1996). Consequently, plant body plans are highly variable, much of that variability is determined through environmentally induced changes in gene expression.

The study of phenotypic plasticity *is* especially important in plants because they *are* generally immobile organisms; therefore, they must tolerate, acclimate, or adapt to their immediate abiotic and biotic environment or die (Bradshaw 1965; Schlichting 1986). Although tolerating environmental stresses is important, the benefit of plasticity is that it allows an organism or a population to utilize a greater ecological niche. Since there are fewer trophic niches plants can occupy compared to animals, competition for resources is more prevalent making it beneficial for plants to have a mechanism to alter size parameters, biological phenomena (e.g., flowering), life history patterns, etc. Variable environments are ubiquitous in both time and space. Because resources are patchily distributed, being able to exploit greater amounts or types of resources and/or environments could enable a plant to survive and to have higher reproductive output than other individuals or other species (Fitter and Hay 2002).

## **Phenotypic Plasticity Is Often, but Not Always, Interpreted as an Adaptive Response to Variable Environments**

### **Phenotypic Plasticity as Adaptive**

Phenotypic plasticity is often thought to confer an adaptive advantage to the organism; i.e., phenotypic expression is fine-tuned to the organism's environmental conditions, allowing for optimal resource acquisition or maximizing fitness in some other way. For example, many plant species produce leaves that vary in morphology, physiology, or a myriad of other phenotypic parameters depending on whether the plant (or the leaf in some cases) is in a sunny or shady environment. "Shade leaves" are larger and thinner, maximizing surface area per unit tissue, and have greater photosynthetic efficiencies (i.e., more CO<sub>2</sub> fixed per unit light absorbed), while "sun leaves" have lower surface area per volume tissue, higher stomatal densities, and greater capacity to dissipate heat (e.g., Vogel 1968). These differences are thought to maximize light capture and minimize heat and photooxidative stress in low- and high-light environments, respectively, thus conferring a selective advantage to those genotypes capable of developing leaves of these phenotypes in the appropriate environments. Plastic sunshade responses have been noted for other phenotypic traits, including leaf area index and biomass allocation to roots and the production of carbon-based chemical defenses (Fitter and Hay 2002).

Assessing plant response to changing environmental conditions in both time and space is essential in testing ecological and evolutionary models concerning whether phenotypic plasticity is adaptive (Wright and McConnaughay 2002). For example, optimal partitioning models predict that plants will respond to changes in resource availability by changing biomass partitioning to above- versus belowground structures, such that the acquisition of the most limiting resource(s), and thus overall growth, is maximized (e.g., Bloom et al. 1985). These models are based on assumed trade-offs in allocation of biomass, or some other unit of investment, to "competing" structures or functions. Individuals have limited resources to invest – in organs, tissues, or metabolic pathways – in growth, defense, reproduction, or other functions, and a unit of biomass (or nitrogen) invested in a leaf (or defensive chemical) cannot be simultaneously invested in a root (or heat-shock protein). An optimal investment strategy would partition biomass or other internal resources such that all plant requirements for growth, reproduction, and defense are balanced. For example, if belowground resources are limited and light and CO<sub>2</sub> are relatively abundant (aboveground resources), then plants would be predicted to allocate more resources to building root structures to aid in nutrient capture (Bloom et al. 1985).

### **Phenotypic Plasticity: A Highly Selected Trait or a Consequence of Selection for Multiple Phenotypes?**

If environmentally induced variation in phenotype is correlated with enhanced fitness across a range of environments, plasticity confers a selective advantage (Sultan 1987).

But if phenotypic plasticity is adaptive, how does selection work to increase variable phenotypic expression? Considerable debate exists regarding whether phenotypic plasticity itself is a trait that can undergo selection versus the consequence of selection for different phenotypes in different environments (e.g., Schlichting 1986; Via 1993). At the heart of the debate is whether plasticity – the ability to produce a range of phenotypic responses depending on environmental conditions – is a trait under independent genetic control and can be selected for apart from its relation to phenotypic development (Via 1993; DeWitt and Scheiner 2004). Part of the support for this logic is provided by specific examples of genetically related individuals that differ greatly in their plastic responsiveness to changes in the environment (Bradshaw 1965; Via 1993). However, Via (1993) suggests that the array of phenotypes observed among related species could be a result of the ubiquitous nature of environmental heterogeneity and that current models support the idea that phenotypic plasticity can evolve as a byproduct of an environment favoring a certain phenotype.

### Theoretical Limits to Selection for Phenotypic Plasticity

If phenotypic plasticity is adaptive, irrespective of the mechanism for selection, key questions remain: why is phenotypic plasticity not more prevalent and why does it not always result in an optimal phenotype? Newman (1992) discusses limits to plasticity, including deficient sensory capabilities, inability to respond, and lack of genetic variation. The events leading to a change in phenotype start with some environmental cue that must be detected by the organism. The organism must be able to respond to the sensory input in a timely manner. If it can, altered gene expression may occur, which can yield an observable or measurable phenotype (Garland and Kelly 2006). Without the cue and the appropriate genetic variation, a strong plastic response cannot be initiated. The accuracy or adaptive value of a plastic response depends on the degree to which the cue predicts future environmental conditions (DeWitt and Scheiner 2004; Garland and Kelly 2006). DeWitt et al. (1998) outline other limits to phenotypic plasticity. Possible costs of phenotypic plasticity include maintenance costs, production costs, information acquisition costs, developmental instability, and genetic costs. Limits to the benefit of phenotypic plasticity in the achievement of optimality were explained including information reliability limit, lag-time limit, developmental range limit, and the epiphenotype problem. Ecological and evolutionary models that have incorporated these types of costs and limits predict that the selective pressure required to increase plasticity under these circumstances would be low, but without empirical data a ranking of the most important factors cannot be determined (DeWitt and Scheiner 2004). More work on this subject matter is needed and will greatly improve the predictions from these models, but they will require thoughtful experimental designs that are more realistic in environmental conditions. Models must also consider that species are never isolated from one another in real life (e.g., DeWitt et al. 1998; Pigliucci 2001).

## Phenotypic Plasticity as Nonadaptive or Maladaptive

Theoretical evaluations of the selective advantage of variation in phenotypic traits historically relied on identifying the underlying genetic structure of the traits under consideration; it is only in the last 30 years or so that the effect of the environment in modulating those traits has received due attention. In fact, environmentally induced variation in traits was previously viewed as “noise” in an otherwise highly evolved system. A review by M. J. West-Eberhard (1989, p. 249) recounts how Sir Vincent Wigglesworth, a British entomologist, “described some geneticists as being ‘apologetic’ about environmentally cued polymorphisms, which they considered examples of unfortunate defects in the delicate genetic apparatus.” She further reported that A. D. Bradshaw “noted that botanists were carefully avoiding any mention of plasticity; environmental effects in experiments were considered ‘only an embarrassment’”!

Indeed, early studies looking for broad patterns to explain variable phenotypic expression focused on the constraints that development imposes upon organisms and consequent restriction on the kinds of phenotypes that can be expressed (e.g., West-Eberhard 1989). The most extreme form of developmentally constrained phenotypic expression is canalization, the expression of a single phenotype regardless of environmental pressure. Canalization is traditionally argued to be a result of stabilizing selection, which reduces variation by favoring individuals with intermediate values for a trait (Waddington 1942), but has more recently been interpreted as an evolutionary result for an inherent need for developmental stability (Siegal and Bergman 2002). The assumption that maintaining stability in a changing environment is adaptive carries with it an implicit assumption that lack of stability (i.e., plasticity) equates to lack of adaptation or is maladaptive (Bradshaw 1965).

Phenotypic variation can be detrimental to organisms if a single phenotype is best in all conditions, further complicating interpretations of variation in phenotypic expression (DeWitt and Scheiner 2004).

Although there tends to be a focus on the adaptive responses of plants to changing environments, some argue that plasticity does not solely function as an adaptive response to increase plant fitness; phenotypic plasticity can result to nonadaptive or even maladaptive responses, particularly in novel environments (Ghalambor et al. 2007).

## The Role of Phenotypic Plasticity in Evolution

There is much debate whether phenotypic plasticity accelerates or slows evolution. Some have argued that phenotypic plasticity at the individual level allows organisms, especially plants, to respond to changes in the environment in a way that does not alter the underlying genetic sequence (regulation by gene regulatory process such as epigenetics), meaning the trait cannot be selected for or against (Schlichting 1986). Others have asserted that phenotypic plasticity speeds evolutionary change as it can accelerate selection because it produces real-time variation that matches

current environmental conditions (Sultan 1987; West-Eberhard 1989). Still others have asserted that it depends on the context. If a trait exists in a wide range of phenotypes, directional selection cannot act upon or favor a particular phenotype, especially if that phenotype does not occur repeatedly. However, when the environment produces cues in a recurrent fashion so that a phenotype is produced repeatedly giving sufficient time for evolution to occur, selection can act upon it and one would conclude that phenotypic plasticity accelerates evolution (West-Eberhard 2003).

---

## **Phenotypic Plasticity Versus Developmentally Programmed Changes in Phenotypic Expression**

Pigliucci (2001) and others use the concept of developmental noise to refer to that aspect of phenotypic expression which is under neither genotypic nor environmental control – the random changes in phenotype due to stochastic events in gene expression. However, phenotypic traits often change in very specific and highly organized ways as organisms grow and develop. G.C. Evans (1972) noted that most phenotypic traits change dramatically over the course of plant growth or development, often in a highly predictable, fixed manner. He termed this phenomenon “ontogenetic drift,” with ontogeny defined as the sequence of events occurring from the single-cell through maturity. For example, different plant organs (leaves, stems, roots, etc.) will increase in biomass throughout growth and development, but the proportion of biomass allocated to each of these organs is rarely constant over time. Changes in allocation during plant growth and development (i.e., ontogeny) reflect the plants’ changing allocation priorities as growth proceeds (Weiner 2004). For herbaceous annuals, it has been found that the proportion of biomass allocated to roots is initially high during the establishment in soil, but decreases dramatically after a few weeks of growth (e.g., Evans 1972; Coleman et al. 1994). Additionally, patterns of biomass allocation may differ as plant life strategy differs. For example, allocation to roots in perennials increases over the course of development (e.g., Niinemets 2004). Coleman et al. (1994) observed that traits relating to resource acquisition, allocation, and partitioning rarely have the same rate of change throughout development, making the study of how ontogeny can place developmental limitations on plasticity necessary.

---

## **Techniques for Evaluating Phenotypic Plasticity**

### **Norms of Reaction Characterize Phenotypic Expression for One or More Genotypes Across a Range of Environments**

The range of phenotypic variation that results from systematic, repeatable responses to degrees of an environmental cue is termed the norm of reaction (Via 1993). The concept, originally described by Woltereck in the early 1900s, was largely ignored, but has become an important research tool in the study of



phenotypic plasticity (Pigliucci 2001). This approach attempts to integrate two components, the phenotype and the environment, into a singular descriptor (e.g., Scheiner 1999) and is used to visualize how a *single* phenotype (y-axis) may vary in different environmental conditions (x-axis) (Whitman and Agrawal 2009). The degree of plasticity is related to the slope of the line depicting the reaction norm and the comparison between or among slopes depicts whether there is variation for plasticity (Pigliucci 2001). For example, two parallel lines with zero slope would have no plasticity and no variation in plasticity, while two parallel lines with nonzero slopes would have plasticity but not variation in plasticity (same slope value), and two nonparallel lines offset from one another would have both plasticity in trait and variation in plasticity (different slopes).

Norms of reaction allow for rapid comparison of phenotypic expression by any number of genetic entities (genotypes, species) across a range of environments. They do not account for ontogenetically induced variation in phenotypic expression, however, which may confound interpretations of phenotypic plasticity.

### **Use of Developmentally Sensitive (Common Size or Developmental Stage) Comparisons Versus Common Time or Age Comparisons**

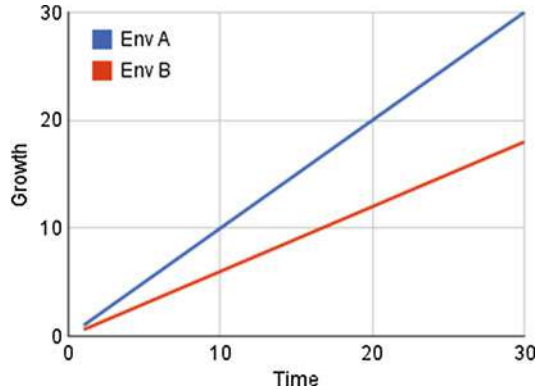
The easiest and probably most common method used to assess phenotypic change in response to environmental conditions is to grow or identify plants across the range of environments of interest and to assess phenotypic variation at a common time (reviewed in Coleman et al. 1994). While this method allows for the researcher to assess phenotypic variation at key time points (e.g., 1 week after germination, when key pollinators are present, following a significant rain event, etc.), it does not allow one to evaluate phenotypic plasticity per se.

Plant growth and development rates vary widely – and often independently – as a function of environment, and developmentally programmed changes in phenotypic expression (i.e., ontogenetic drift) are common; the result is that most traits exhibit variable phenotypic expression over time even when the environment is held constant, confounding the interpretation of phenotypic plasticity (e.g., Evans 1972; Coleman et al. 1994; Wright and McConnaughay 2002). McConnaughay and colleagues have demonstrated how interpretations of phenotypic plasticity can change if plants are compared at a common time (obscuring the effects of developmentally programmed changes in phenotypic expression) versus at a common size or developmental stage (Coleman et al. 1994; Coleman and McConnaughay 1995; Gedroc et al. 1996; McConnaughay and Coleman 1999; Wright and McConnaughay 2002).

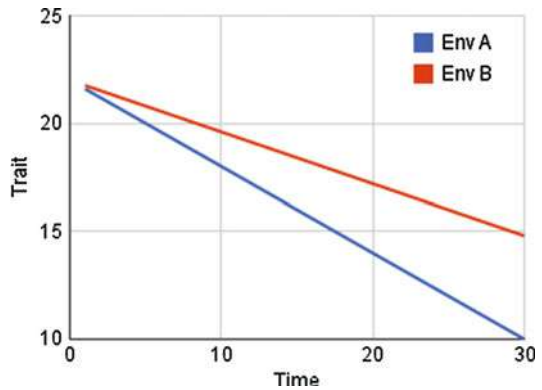
### **“Apparent Plasticity”: Variable Phenotypic Expression that Arises Solely from Plastic Growth and Developmental Rates Coupled with Ontogenetic Drift**

Figures 1, 2, and 3 demonstrate how interpretation of experimental results may differ when relationships among developmental patterns and growth rates are considered

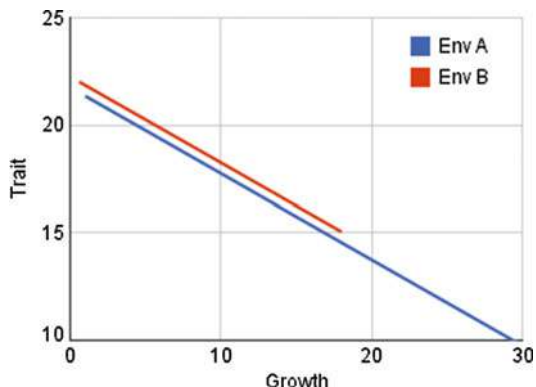
**Fig. 1** Idealized depiction of plasticity in plant growth and development. Genetically identical plants may exhibit different growth and developmental rates in different environments. In this example, *environment A* supports higher rates of growth and development



**Fig. 2** Idealized depiction of phenotypic trait that exhibits ontogenetic drift. In this example, the phenotypic trait decreases in value as ontogeny proceeds. The trait appears to also exhibit plasticity, as phenotypic expression differs across environments



(Coleman et al. 1994). Figures 1 and 2 depict two common patterns exhibited by plants in variable environments: plastic growth responses (Fig. 1) and developmentally programmed changes in phenotypic expression or ontogenetic drift (Fig. 2). If one were to compare phenotypes from these two environments at a common time, they would correctly conclude that phenotypic expression differs in the two environments, which may be interpreted as support for phenotypic plasticity. However, in Fig. 3, the pattern of change in phenotypic expression throughout growth and development is fixed (i.e., the slope is constant for each line). The variable phenotypic expression is not apparent when plants are compared at similar stages in growth and development. In other words, environmental heterogeneity causes plasticity in growth rates, and ontogenetic drift occurs, but the ontogenetic program of phenotypic change throughout development is constant (i.e., not plastic). Thus, when compared at a common point in growth and development, there are no phenotypic differences among treatments, although phenotypes will differ when compared at a common time (Coleman et al. 1994). This trait is said to exhibit “apparent plasticity,” which is defined as variation in a trait because of environmentally induced variation in growth or development coupled with



**Fig. 3** The phenotypic trait values from Fig. 2 are replotted against the growth values from Fig. 1, using the time point data for each x,y pair. In this example, the ontogenetic trajectory for the phenotypic trait does not vary with environment, that is, to say it is not plastic. Thus, variations in phenotypic expression are consequences of plasticity in growth and developmental rates coupled with ontogenetic drift and not plasticity in the ontogenetic program for phenotypic expression

ontogenetic drift in the trait of interest (McConnaughay and Coleman 1999; Wright and McConnaughay 2002).

Apparent plasticity is usually the result of environmental conditions that the plant does not have appreciable control over (toxins, soil nutrients, oxygen levels, or temperature) as opposed to a condition the plant can respond to by actively changing development in such a way as to maintain constant growth rates (Scheiner 1999). These environmental conditions alter biochemical processes within the plant, which in turn affect development, typically resulting in smaller size plants compared to their non-limited cohorts (Whitman and Agrawal 2009).

Those studying optimal partitioning theory (OPT) models will find this scenario particularly relevant. For example, plants grown in shade conditions invest a greater proportion of assimilated resources to growing leaves and will have a greater leaf area ratio (LAR) than the same species grown in more abundant light conditions. However, shade-grown plants (or plants in any unfavorable condition) typically grow and develop more slowly, and plants typically invest more biomass in structural support relative to leaf area as they grow. Is the reduced LAR in shade-grown plants the result of structural and functional adjustments of resource allocation as predicted by OPT or a consequence of slow growth rates and delayed development along a fixed ontogenetic trajectory? If examined at a common age, the conclusion may be in favor of OPT in which the plant is apparently optimizing function, but if plants are compared at the same size, differences in allocation may disappear, diminishing the discussion on the effects of differing light treatment on LAR (Coleman et al. 1994).

Mooney et al. (1988) studied the effects of sulfur dioxide ( $\text{SO}_2$ ) on the growth and resource acquisition of cultivated radish, *Raphanus sativus*, by measuring changes in photosynthetic activity, biomass accumulation, and root to leaf allocation relative to controls. Excess atmospheric sulfur dioxide is caused by human

industrial activities and can cause direct inhibition of photosynthesis by increasing stomatal opening, leading to excessive water loss (Varshney et al. 1979). They measured reduced photosynthetic performance in SO<sub>2</sub>-exposed plants and attributed this to a reduction in carboxylating capacity (Mooney et al. 1988). Recall, carboxylation is the first step of the Calvin-Benson cycle in which the enzyme Rubisco adds a CO<sub>2</sub> to ribulose 1,5-bisphosphate (RuBP) resulting in two molecules of PGA. They explained that the lower growth rate observed in SO<sub>2</sub>-exposed plants was mitigated by increased allocation to new leaf material, an observation consistent with optimal partitioning theory (Mooney et al. 1988). However, when Coleman and McConnaughay (1995) reexamined these data, plants were compared at a common size or via an allometric approach. Differences in root to shoot ratio and leaf area ratio that is reported to be statistically different in Mooney et al. (1988) were actually similar when compared at a common size and support the conclusion that these reductions were a result of ontogenetic drift and not changes in functional allocation (Coleman and McConnaughay 1995).

### **Additional Examples of Misinterpretation When Developmental Context Is Ignored**

When growth rates, phenotypic expression throughout ontogeny, and the pattern of change throughout ontogeny are all variable, three different scenarios can result in which the presence, magnitude, or direction of phenotypic plasticity will differ when plants are compared at a common age versus a common size or developmental stage (Coleman et al. 1994; Wright and McConnaughay 2002).

In the first scenario, phenotypes look similar when plants are compared at a common age or time, but there are clear phenotypic differences when plants are compared at a common stage in development (i.e., the depiction looks opposite of the scenario discussed in regard to apparent plasticity). Rice and Bazzaz (1989) examined the effect of varying light conditions on different plant traits and quantified plasticity in those traits in an annual plant (*Abutilon theophrasti*) at both a common plant age and a common plant size. They found that treatment-induced differences in leaf number and height became apparent when plants were compared at a common size, although not at a common age.

A second scenario is that the comparison at a common plant age or plant size may result in quantitatively different results, but the direction is the same, so the conclusions drawn without explicitly incorporating the developmental context will be similar to the more developmentally explicit test, but the estimation of the magnitude of the plastic response will differ (Coleman et al. 1994). Poorter et al. (1994) examined the differences in morphology, carbon economy, and chemical composition of fast- and slow-growing seedlings over a period of 1–2 months when exposed to low nitrogen conditions to see if these conditions would yield similar results of previous research done at non-limiting resource levels. Measures between these two species, such as those for specific leaf area and growth rate, were more pronounced when plants were compared at a common age versus a common size, likely because the slow-growing plants were at an earlier stage of ontogeny for the common age comparisons (Coleman et al. 1994).

A third scenario is predicted in which the direction of the results is reversed, with one treatment having a greater value, when compared at plant age, and a smaller value compared to the other treatment, when compared at a common plant size or vice versa (Coleman et al. 1994). Evans (1972) in his work with *Impatiens parviflora* demonstrated this pattern in regard to leaf weight as a function varying light intensity.

### **“Complex Plasticity”: Variable Phenotypic Expression Arising from the Interplay of Plasticity in Growth and Development and in the Ontogenetic Trajectory of the Trait of Interest**

As we have noted earlier, whenever there is plasticity in both the growth rate (assuming that ontogenetic drift in phenotypic expression occurs) and in the ontogenetic program itself (i.e., the pattern of ontogenetic drift changes with environment), “complex plasticity” is observed, which is defined as plasticity in both the growth rate and the ontogenetic program/trajectory (Wright and McConnaughay 2002). Yet still the situation can be further complicated and is more realistic, as plasticity in the ontogenetic program may occur at specific windows of time or may be expressed differently at different points of ontogeny depending on the species and the specific trait examined (Wright and McConnaughay 2002). For example, plasticity in root to shoot ratios was examined for two species of annuals to determine if changes were consistent with optimal partitioning theory (OPT) or exhibited ontogenetic drift. Substantial ontogenetic drift was found in partitioning and the period at which plasticity was expressed, as root to shoot ratios decreased through ontogeny and plasticity in partitioning only occurred early in the experiment, respectively. It was concluded that root to shoot partitioning was partially consistent with OPT, but that it was ontogenetically constrained, and that the ontogenetic program could exhibit plastic responses early, but not later, in development (Gedroc et al. 1996).

Other studies have evaluated phenotypic plasticity using a developmental context. Geng et al. (2007) went a step farther and evaluated root allocation under different resource levels in the perennial *Alternanthera philoxeroides* to test predictions based on this developmentally explicit model of phenotypic plasticity. In annual plants, root allocation is initially high and declines with growth and development (Hunt 1990); plants that are growing in environments where belowground resources are limiting – assuming that growth is impaired – already have favorable root allocation patterns. Conversely, when aboveground resources limit growth, the normal developmental pattern of high root (low shoot) allocation provides a distinct disadvantage for resource acquisition. The developmentally explicit model thus predicts that plants should exhibit true plasticity in response to aboveground resources (e.g., light and CO<sub>2</sub>) but not necessarily in response to belowground resources (McConnaughay and Coleman 1999). Since root allocation in perennials is different than annuals in that root allocation increases over time, Geng et al. (2007) predicted opposite responses for *Alternanthera philoxeroides* (i.e., true plasticity in belowground resources and apparent plasticity in aboveground resources). They exposed genetically similar clonal stem fragments to varying

levels of light, nutrients, and water. This approach can be likened to the split-brood design explained by Via (1993), in which a family member or clone is split among different environments. If one tests random samples from a population, only mean plasticity can be estimated and values for genetic variation in plasticity cannot be obtained, this approach is probably more realistic as not all plants reproduce asexually, but small differences in treatment may be confounded due to lack of genetic similarity. In this experiment, when one resource was kept low, the other two were maintained at moderate or high levels. Growth parameters were measured over a period of 81 days with frequent harvests over time. Root allocation was examined over time and as a function of total plant biomass (i.e., same ontogenetic stage). When examined in both manners, root allocation did increase over the 81-day growth period in a direction that was opposite of that predicted of annual herbs. When compared across time, all three treatments resulted in significant differences between high and low resource levels, with an increase in allocation to roots for low nutrient and low water treatments and an increase in allocation to shoots in the light-limited treatments. When compared across size, water and nutrient treatments remained significant; however, there was no longer a difference between low- and high-light conditions. Therefore, root allocation in response to light limitation resulted in slowed growth rates along a fixed ontogenetic trajectory – a response in agreement with apparent plasticity– while root allocation patterns exhibited in response to belowground resources (water and nutrients) were consistent with complex plasticity. These results agreed with predictions of the developmentally explicit model (McConnaughay and Coleman 1999). As this study demonstrates, evaluating phenotypic plasticity in the context of ontogeny is important as it allows a more complete understanding of an organism’s ability to respond to a heterogeneous environment.

### **When Is Developmental Context Not Important?**

If the trait of interest does not exhibit ontogenetic drift, environmental effects’ rates of growth and development will obviously not alter phenotypic expression. Similarly, if the trait does exhibit ontogenetic drift, but there is no plasticity in growth or development, all observed phenotypic variation across environments must be attributed to plasticity in the ontogenetic trajectory. In either case, phenotypic plasticity may be inferred at a common time without misinterpretation as a consequence of ignoring developmentally induced variation in phenotypic expression.

### **Growth Analysis and Allometric Approaches**

The study by Gedroc et al. (1996) highlights two important aspects of the ontogenetic pattern of phenotypic expression. First, a trait’s ontogenetic trajectory is not necessarily a simple linear function of plant growth or development. Phenotypic expression of a trait may not change in a simple linear or even monotonic fashion as growth and development proceeds. It may be harder to “capture” the developmental

context of phenotypic expression during periods of rapid developmentally induced phenotypic change. Second, the developmental trajectory of a trait is not always either plastic or invariant, but can depend on the stage of development within which an environmental cue is perceived. Only by following phenotypic expression throughout growth and development, i.e., explicitly evaluating the developmental trajectory of the trait itself, is it possible to evaluate the degree to which phenotypic expression was altered due to environmentally induced phenotypic plasticity in the trait or due to environmentally induced plasticity in growth and development and developmentally coordinated nonplastic changes in phenotypic expression.

Traditional growth analysis techniques (e.g., Hunt 1990) allow one to explicitly evaluate phenotypic expression across growth and development by allocating experimental units (replicates) across time and using regression techniques to evaluate the developmental trajectory directly (e.g., Coleman et al. 1994; Coleman and McConnaughay 1995; Gedroc et al. 1996; McConnaughay and Coleman 1999).

In some instances, the researcher is more interested in the relationship between two functionally related traits. For example, root allocation is almost always considered in the context of allocation to other organs (e.g., leaves) or more simply allocation to all other functions. It may be useful in such cases to evaluate the allometric relationship between root mass production and leaf (or shoot) mass production directly.

## **Modular Growth as a Platform for Evaluating Phenotypic Plasticity in Plants**

Plant growth is distinctly modular in nature. The plant body exists as an assemblage of repeated units, also known as modules or metamers, which can be arrayed in varying number, size, and pattern (White 1979). In addition, the production of these modules is often indeterminate, meaning that the production and placement of modules is not predetermined but varies, often in response to environmental conditions, and the modules themselves often exhibit plasticity, in size, shape, or even biochemical or physiological features. The result is a highly flexible body plan that can respond to spatial and temporal variation in environmental conditions.

### **Meristem Fates Determine Plant Architecture and Body Plan**

Module development in plants occurs in highly localized regions. Immature, undifferentiated cells that are capable of dividing, called meristematic cells, are found in distinct places in the plant such as the tips of roots or shoots (apical meristems) and along the stem to increase plant diameter. How much and where meristematic tissue is localized helps determine how a plant grows and overall plant architecture (Silvertown and Charlesworth 2001). There are two fates of meristematic tissue – reproductive and vegetative. The shoot apical meristem (SAM) is a regenerative cell population responsible for aboveground biomass. Cells of this type are located at the tip of the shoot and are responsible for lateral organs such as leaves and flowers. SAMs may remain vegetative, helping establish the location of



nodes and branches, leaves, and distance between internodes. Vegetative tissues exhibit indeterminate growth patterns and during ontogenetic development could produce node after node forming repeating units called modules. Meristematic tissue also allows for horizontal growth of plants which is termed clonal growth. Each individual that arises is genetically identical to the original plant and is referred to as a ramet of the larger genet; physical connections to exchange nutrients may be temporary or long lasting (Silvertown and Charlesworth 2001). However, SAMs can also exhibit determinate growth patterns if production of inflorescences or flowers occurs.

### Phenotypic Plasticity in a Modular Body Form

Research on how meristems “determine” the fate of cells is extensive and involves complex chemical/hormonal and genetic regulatory pathways that are beyond the scope of this chapter. But in relation to phenotypic plasticity, the modular lifestyle of plants deserves consideration. The plant body can be thought of as a series of repeated units (also known as metamers or phytomers). A single leaf blade, the portion of the stem immediately subtending that leaf, and the branch meristem located in the axil of the leaf taken together comprise a module which can be repeated to create the aboveground plant body. Typically studies on modularity have been done with clonal plants, as plasticity across clonal units has been long noted, but modularity has been found to be equally as prevalent across metamers of nonclonal species (e.g., de Kroon et al. 2005).

As unitary organisms, humans have comparatively set mechanisms and biomechanics of growth and development (Wu et al. 2003), so it is often difficult for us to think non-anthropomorphically about plant development. As noted earlier, plant development is comparatively indeterminate, so unlike unitary organisms in which growth ceases at a particular age or size, plants have the potential to keep growing, within the constraints of biomechanics, even upon reaching reproductive maturity (Fitter and Hay 2002; Weiner 2004). Therefore, each module can be exposed to a slightly different environment than the modules that came before it, making the determination of a plant’s success in a given environment more difficult to assess especially if it is not studied over a lifetime (de Kroon et al. 2005).

It is important to understand the role of development when considering how plants respond to their environment, so instead of phenotypic plasticity being traditionally viewed in terms of the whole plant summation of all modular responses, some have argued that phenotypic plasticity should be viewed in terms of the module (de Kroon et al. 2005). Pamela K. Diggle (1994) explored this long-ignored phenomenon in her investigation of andromonoecy (plants bearing both hermaphroditic and staminate flowers) in *Solanum hirtum*. Andromonoecy is a trait that is said to exhibit phenotypic plasticity because resource availability to reproduction affects the relative abundance of each flower type. She found that floral primordia located at the base of the plant stem invariably became hermaphroditic flowers, while those located at the distal ends could develop into either hermaphroditic or staminate flowers (Diggle 1994). Based on these observations, she developed the term ontogenetic contingency, stating “the developmental fate of a



primordium depends upon where and when it is produced within the architecture of an organism and what events [environmental conditions] have preceded it during ontogeny” (Diggle 1994, p. 1354).

The benefit of modular integration is that it can help a plant overcome spatial and temporal environmental variation that decreases their success. Traditional methods of studying plasticity, such as developmental reactions norms, do not explicitly recognize that the whole plant phenotype is the integrated sum of many modules that may develop and exist in different environmental states and that plasticity may be expressed at the level of modules (de Kroon et al. 2005).

---

## Selecting Methodological Approaches

The aforementioned models demonstrate the importance of developing developmentally explicit models to distinguish between the various plasticities (passive, true (ontogenetic and developmental), and complex). In order to do this, methodological approaches may need to consider ontogenetic effects depending on the hypothesis being tested. Many of the methodological considerations have already been highlighted in the examples discussed thus far.

Since we recognize the importance of examining phenotypic expression over growth and development, our sampling must also reflect this idea. Samples should be collected over the entire period of ontogeny and through time, with harvests occurring more frequently as opposed to only at a few time points or at the end of the season. Given practical considerations, this will almost certainly mean that the number of replicate samples evaluated at any time point will be fewer. Lower replication at any given time point often raises concerns about statistical power; however, when the response of interest is the developmental trajectory itself, and not the response at a given point along the developmental trajectory (i.e., the linear or curvilinear change in phenotype over time as measured by a line or curve vs. the value of the phenotypic trait at any specific time as measured by a point on the line or curve), power is best conserved by spreading samples across the range of the line or curve one is attempting to characterize (Hunt 1990; Coleman et al. 1994).

Many plant processes are best understood in terms of size rather than age, but the standard of comparison will depend on the specific research goal (Coleman et al. 1994; Wright and McConnaughay 2002). When making single phenotypic observations that are related to characteristics that vary during growth and development, an ontogenetic standard should be implemented that utilizes comparisons at a common point in ontogeny. Examples of phenotypic processes that may benefit from this approach include proportional biomass allocation, functional adjustment in biomass partitioning, leaf area production, branching frequency, whole-plant nitrogen uptake, and reproductive output in relation to life history stage (Coleman et al. 1994; Wright and McConnaughay 2002).

Not all traits require examination at common ontogenetic stages and comparisons at a common age can be more appropriate for studying real-time processes like

herbivory, intra- and interspecific competition, plant-plant or plant-animal interactions, or analysis of plant-environment interactions, such as those related to changes in season. Examples of phenotypic processes that may benefit from this approach include leaf nitrogen levels in relation to herbivory, flowering loads and pollinator activity, or total fitness in annual plants affected by end of season frosts (Coleman et al. 1994; Wright and McConnaughay 2002). Agrawal et al. (2012) examined plant resistance mechanisms in *Oenothera biennis*. By suppressing insect herbivores in the field, they found that after only 5 years, protected plants diverged from control plants, producing less defensive chemicals in their fruits while increasing competitive ability likely due to greater energy available for biomass allocation. The approach of chronological age was appropriate for this real-time study that demonstrated rapid, ecological, and evolutionary change.

The examination of two or more functionally related phenotypic traits is highly dependent upon the degree of ontogenetic drift that each trait exhibits. An allometric approach is best used to assess each of these traits individually and in relation to each other. Allometric growth is an unequal change in the size of one body part relative to the change in size of another body part, or sometimes the entire body, whereas isometric growth is the condition of directly proportional change among body parts arising from identical growth rates of the individual parts (West-Eberhard 2003; Wu et al. 2003). Most traits exhibit allometry with respect to one another, as opposed to isometry (West-Eberhard 2003). One example of isometry in plants is that for every leaf there will be exactly one petiole attaching the leaf to the stem (though the sizes of the leaf blade and petiole in question may be allometric). Isometry between two traits would be characterized by a simple linear relationship with a slope of 1.0. Any deviation from a slope of 1.0 represents an allometric relationship between the traits such that change in one trait during growth and development is greater or lesser than the change in the other trait. If the relationship between two traits exhibits curvilinearity during development, the traits are said to exhibit complex rather than simple allometry (Coleman et al. 1994). When structures compete with each other for resources (like roots and shoots), a change in either allometric ratio of either trait would be expected to influence the other (West-Eberhard 2003). Other examples of functionally related phenotypic traits include root to shoot biomass accumulation, comparisons of reproductive versus vegetative biomass, relationships between height and diameter, tissue carbon to nitrogen ratios, or leaf nitrogen composition and photosynthetic relationships (Coleman et al. 1994; Wright and McConnaughay 2002).

Ontogenetic drift can only be ignored when relationships between biomass variables are isometric and linear (i.e., simple allometry). When biomass allocation patterns are allometric, patterns will differ throughout growth and development regardless of environmental conditions, meaning they exhibit ontogenetic drift (McConnaughay and Coleman 1999). Many allocation patterns follow allometric trajectories, which are intrinsically a function of plant size (Weiner 2004), so any factor that influences size will change allocation. It is well known that plant allocation patterns are size dependent, but methods traditionally used to assess biomass allocation, such as optimal partitioning theory, make the assumption that

plant allocation is size independent. Taking an allometric approach to these studies incorporates the dynamics of size changes in response to environmental conditions. It also helps allow one to differentiate between plasticity in growth rate (apparent plasticity) and plasticity as a result of environmental heterogeneity (true plasticity; see Coleman et al. 1994; Gedroc et al. 1996; Weiner 2004; Geng et al. 2007).

---

## Future Directions

Studies of plant phenotypic plasticity aim to increase our understanding of how plants cope with variable environments. At this time, no simple unified theory exists that predicts when, how, or to what extent plants can respond to changes in the environment with changes in phenotype or under what circumstances any such phenotypic changes will increase fitness. Past work has obscured our evaluation of phenotypic plasticity by confounding environmentally induced variation in phenotypic expression with environmentally induced variation in growth and development and developmentally fixed patterns of phenotypic expression. A more developmentally explicit approach to evaluating the mechanisms of variable phenotypic expression could lead to a greater understanding of the limits of phenotypic plasticity and its potential significance in evolutionary and ecological contexts.

---

## References

- Agrawal AA, Hastings AP, Johnson MTJ, Maron JL, Salminen J. Insect herbivores drive real-time ecological and evolutionary change in plant populations. *Science*. 2012;338:113–6.
- Bloom AJ, Chapin III FS, Mooney HA. Resource limitation in plants – an economic analogy. *Annu Rev Ecol Evol Syst*. 1985;16:363–92.
- Bradshaw AD. Evolutionary significance of phenotypic plasticity in plants. *Adv Genet*. 1965;13:115–55.
- Coleman JS, McConnaughay KDM. A non-functional interpretation of a classical optimal-partitioning example. *Funct Ecol*. 1995;9:951–954.
- Coleman JS, McConnaughay KDM, Ackerly DD. Interpreting phenotypic variation in plants. *Trends Ecol Evol*. 1994;9:187–91.
- de Kroon H, Heidrun H, Stuefer JF, van Groenendael JM. A modular concept of phenotypic plasticity in plants. *New Phytol*. 2005;166:73–82.
- DeWitt TJ, Scheiner SM. *Phenotypic plasticity: functional and conceptual approaches*. New York: Oxford University Press; 2004.
- DeWitt TJ, Sih A, Wilson DS. Costs and limits of phenotypic plasticity. *Trends Ecol Evol*. 1998;13:77–81.
- Diggle PK. The expression of andromonoecy in *Solanum hirtum* (Solanaceae): phenotypic plasticity and ontogenetic contingency. *Am J Bot*. 1994;81:1354–65.
- Evans GC. *The quantitative analysis of plant growth*. Oxford: Blackwell Scientific; 1972.
- Fitter A, Hay R. *Environmental physiology of plants*. 3rd ed. London: Academic; 2002.
- Garland T, Kelly SA. Phenotypic plasticity and experimental evolution. *J Exp Biol*. 2006;209:2344–61.
- Gedroc JJ, McConnaughay KDM, Coleman JS. Plasticity in root shoot partitioning: optimal, ontogenetic, or both? *Funct Ecol*. 1996;10:44–50.

- Geng Y, Pan X, Xu WZ, Li B, Chen J. Plasticity and ontogenetic drift of biomass allocation in response to above- and belowground resource availabilities in perennial herbs: a case study of *Alternanthera philoxeroides*. *Ecol Res.* 2007;22:255–60.
- Ghalambor CK, McKay JK, Carroll SP, Reznick DN. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol.* 2007;21:394–407.
- Hunt R. Basic growth analysis. London: Unwin Hyman Press; 1990.
- McConnaughay KDM, Coleman JS. Biomass allocation in plants: ontogeny or optimality? A test along three resource gradients. *Ecology.* 1999;80:2581–93.
- Mooney HA, Küppers M, Koch G, Gorham J, Chu C, Winner WE. Compensating effects to growth of carbon partitioning changes in response to SO<sub>2</sub>-induced photosynthetic reduction in radish. *Oecologia.* 1988;75:502–6.
- Newman RA. Adaptive plasticity in amphibian metamorphosis. *BioScience.* 1992;42:671–8.
- Niinemets U. Adaptive adjustments to light in foliage and whole-plant characteristics depend on relative age in the perennial herb *Leontodon hispidus*. *New Phytol.* 2004;162:683–96.
- Pigliucci M. Phenotypic plasticity: beyond nature and nurture. Baltimore: The John Hopkins University Press; 2001.
- Poorter H, Claudius ADM, van de Vijver CADM, Boot RGA. Growth and carbon economy of a fast-growing and a slow-growing grass species as a dependent on nitrate supply. *Plant Soil.* 1994;171:217–27.
- Rice SA, Bazzaz FA. Quantification of plasticity of plant traits in response to light intensity: comparing phenotypes at a common weight. *Oecologia.* 1989;78:502–7.
- Scheiner SM. Towards a more synthetic view of evolution. *Am J Bot.* 1999;86:145–8.
- Schlichting CD. The evolution of phenotypic plasticity in plants. *Annu Rev Ecol Evol Syst.* 1986;17:667–93.
- Siegal ML, Bergman A. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci U S A.* 2002;99:10528–32.
- Silvertown J, Charlesworth D. Introduction to plant population biology. 4th ed. Oxford: Blackwell Science; 2001.
- Sultan SE. Evolutionary implications of phenotypic plasticity in plants. *Evol Biol.* 1987;21:127–78.
- Varshney CK, Garg JK, Lauenroth WK, Heitschmidt RK. Plant responses to sulfur dioxide pollution. *Crit Rev Environ Control.* 1979;9:27–50.
- Via S. Adaptive phenotypic plasticity: target or by-product of selection in a variable environment? *Am Nat.* 1993;142:352–65.
- Vogel S. “Sun leaves” and “shade leaves”: differences in convective heat dissipation. *Ecology.* 1968;49:1203–4.
- Waddington CH. Canalization of development and the inheritance of acquired characters. *Nature.* 1942;15:563–5.
- Walbot V. Sources and consequences of phenotypic and genotypic plasticity in flowering plants. *Trends Plant Sci.* 1996;1:27–32.
- Weiner J. Allocation, plasticity and allometry in plants. *Perspect Plant Ecol Evol Syst.* 2004;6:207–15.
- West-Eberhard MJ. Phenotypic plasticity and the origins of diversity. *Annu Rev Ecol Syst.* 1989;20:249–78.
- West-Eberhard MJ. Developmental plasticity and evolution. New York: Oxford University Press; 2003.
- White J. The plant as a metapopulation. *Annu Rev Ecol Syst.* 1979;10:109–45.
- Whitman DW, Agrawal AA. What is phenotypic plasticity and why is it important? In: Whitman DW, Ananthakrishnan TN, editors. Phenotypic plasticity of insects: mechanisms and consequences. Enfield: Science Publishers; 2009. p. 1–63.
- Wright SD, McConnaughay KDM. Interpreting phenotypic plasticity: the importance of ontogeny. *Plant Species Biol.* 2002;17:119–31.

- Wu R, Ma C, Lou X, Casella G. Molecular dissection of allometry, ontogeny, and plasticity: a genomic view of developmental biology. *BioScience*. 2003;53:1041–7.
- Yampolsky LY, Scheiner SR. Developmental noise, phenotypic plasticity, and allozyme heterozygosity in *Daphnia*. *Evolution*. 1994;5:1715–22.

## Further Reading

- Bernacchi CJ, Coleman JS, Bazzaz FA, McConnaughay KDM. Biomass allocation in old-field annual species grown in elevated CO<sub>2</sub> environments: no evidence for optimal partitioning. *Glob Change Biol*. 2000;8:855–63.
- Bernacchi CJ, Thompson JN, Coleman JS, McConnaughay KDM. Allometric analysis reveals relatively little variation in nitrogen versus biomass accrual in four plant species exposed to varying light, nutrients, water and CO<sub>2</sub>. *Plant Cell Environ*. 2007;30:1216–22.
- Caldwell MM, Pearcy RW. Exploitation of environmental heterogeneity by plants: ecophysiological processes above- and belowground. London: Academic; 1994.
- Chevin L, Lande R, Mace GM. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol*. 2010;8:e1000357. doi:10.1371/journal.pbio.1000357.
- Coen E. The art of genes: how organisms make themselves. New York: Oxford University Press; 1999.
- Diggle PK. A developmental morphologist's perspective on plasticity. *Evol Ecol*. 2002;16:267–83.
- Gianoli E, Valladares F. Studying phenotypic plasticity: the advantage of a broach approach. *Biol J Linn Soc*. 2012;105:1–7.
- Hallgrímsson B, Hall BK. Variation: a central concept in biology. San Diego: Academic; 2005.
- Hodge A. The plastic plant: root responses to heterogeneous supplies of nutrients. *New Phytol*. 2004;162:9–24.
- Huber H, Lukács S, Watson MA. Spatial structure of stoloniferous herbs: an interplay between structural and blue-print, ontogeny and phenotypic plasticity. *Plant Ecol*. 1999;141:107–15.
- Leyser O, Day S. Mechanisms of plant development. Oxford: Blackwell; 2003.
- Matesanz S, Gianoli E, Valladares F. Global change and the evolution of phenotypic plasticity in plants. *Ann NY Acad Sci*. 2010;1206:35–55.
- McCarthy MC, Enquist BJ. Consistency between an allometric approach and optimal partitioning theory in global patterns of plant biomass allocation. *Funct Ecol*. 2007;21:713–20.
- McConnaughay KDM, Coleman JS. Can plants track changes in nutrient availability via changes in biomass partitioning? *Plant and Soil*. 2008;202:201–9.
- Miner BG, Sultan SE, Morgan SG, Padilla DK, Relyea RA. Ecological consequences of phenotypic plasticity. *Trends Ecol Evol*. 2005;20:685–92.
- Mooney HA, Winner WE, Pell EJ. Response of plant to multiple stresses. London: Academic; 1991.
- Moriuchi KS, Winn AA. Relationships among growth, development and plastic response to environmental quality in a perennial plant. *New Phytol*. 2005;166:149–58.
- Müller I, Schmid B, Weiner J. The effect of nutrient availability on biomass allocation patterns in 27 species of herbaceous plants. *Perspect Plant Ecol Evol Syst*. 2000;3:115–27.
- Novoplansky A. Developmental plasticity in plants: implications of noncognitive behavior. *Evol Ecol*. 2002;16:177–88.
- Pigliucci M. Evolution of phenotypic plasticity: where are we now? *Trends Ecol Evol*. 2005;20:481–5.
- Pigliucci M, Hayden K. Phenotypic plasticity is the major determinant of changes in phenotypic integration in *Arabidopsis*. *New Phytol*. 2001;152:419–30.
- Pigliucci M, Murren CJ, Schlichting CD. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol*. 2006;209:2362–7.

- Porter JR. A modular approach to plant growth analysis. I. Theory and principles. *New Phytol.* 1983;94:183–90.
- Porter JR, Lawlor DW. Plant growth interactions with nutrition and environment. Cambridge: Cambridge University Press; 1991.
- Reekie EG, Bazzaz FA. Reproductive allocation in plants. London: Academic; 2005.
- Smith JM, Burian R, Kaufman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. Developmental constraints and evolution. *Q Rev Biol.* 1985;60:265–87.
- Stanton ML, Roy BA, Thiede DA. Evolution in stressful environments I: phenotypic variability, phenotypic selection, and response to selection in five distinct environmental stress. *Evolution.* 2000;54:93–111.
- Steinger T, Roy BA, Stanton ML. Evolution in stressful environments II: adaptive value and costs of plasticity in response to low light in *Sinapis arvensis*. *J Evol Biol.* 2003;16:313–23.
- Sultan SE. Phenotypic plasticity for plant development, function and life history. *Trends Plant Sci.* 2000;5:537–42.
- Sultan SE, Bazzaz FA. Phenotypic plasticity in *Polygonum persicaria*. III. The evolution of ecological breadth for nutrient environment. *Evolution.* 1993;47:1050–71.
- Valladares F, Sanchez-Gomez D, Zavala MA. Quantitative estimation of phenotypic plasticity: bridging the gap between evolutionary concept and its ecological applications. *J Ecol.* 2006;94:1103–16.
- Valladares F, Gianoli E, Gómez JM. Ecological limits to plant phenotypic plasticity. *New Phytol.* 2007;176:749–63.
- Via S, Lande R. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution.* 1985;39:505–22.
- Via S, Gomulkiewicz R, De Jong G, Scheiner SM, Schlichting CD, Van Tienderen PH. Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol Evol.* 1995;10:212–7.

---

# Evolutionary Ecology of Chemically Mediated Plant-Insect Interactions

# 6

Amy M. Trowbridge

## Contents

Introduction .....	144
A Primer on Plant-Herbivore Evolution .....	146
Techniques and Analyses for Testing Classic Macroevo­lutionary Hypotheses .....	146
A Framework for Explaining the Diversity and Function of Secondary Metabolites .....	147
Ecological Costs, Trade-Offs, and the Emergence of Plant Defense Theories .....	148
This Ain't a Scene, It's an Arms Race .....	152
Fitting Plants with Weapons in the Form of Chemicals .....	153
Advantages of Chemical Mixtures .....	157
Spatiotemporal Patterns of Plant Secondary Chemistry Alter Herbivore Performance ...	158
Running to Stay in the Same Place .....	161
Basis of Plant Selection and Evolution of Feeding Deterrents .....	162
Insect Detoxification Systems .....	163
Sequestration .....	164
Induced Responses .....	165
Regulation of Costly Defenses .....	165
Plant Perception and Signal Transduction .....	167
Direct Inducible Defenses .....	169
Indirect Inducible Defenses .....	170
Future Directions .....	171
References .....	172

---

## Abstract

- Approaches for testing macroevolutionary theories.
- Coevolution: understanding the diversity and role of plant secondary compounds.
- Costs associated with synthesis: trade-offs and defense theory.
- Plants are armed with an arsenal of chemical defenses against insects.

---

A.M. Trowbridge (✉)  
Department of Biology, Indiana University, Bloomington, IN, USA  
e-mail: [amtrowbr@indiana.edu](mailto:amtrowbr@indiana.edu)

- Spatiotemporal patterns of plant defense chemistry: timing and location of synthesis can impact herbivory.
- Insect host preference and response to secondary compounds.
- Regulation of defense: plant perception of herbivory, signal transduction, and induced responses.
- Plant volatile-mediated defenses against herbivores.

*It takes all the running you can do to keep in the same place.  
The Red Queen to Alice  
Through the Looking Glass (1960)*

---

## Introduction

Plants and their associated insect herbivores account for more than half of all described species and interactions between these organisms are among the most dominant relationships in nature. Terrestrial plants serve as the primary food source for more than one million insect species scattered across diverse taxa, and these insect herbivores ingest >20 % of annual net primary productivity (Schoonhoven et al. 2005). Insects have developed diverse feeding strategies to obtain nutrients from their host plants, yet plants have not remained passive in the face of these attacks. Rather, plants have developed constitutive and dynamic forms of both physical and chemical resistance over evolutionary time to mitigate herbivory. These plant traits consequently influence the evolutionary trajectories of herbivores, thus resulting in the reciprocal evolution of herbivore countermeasures to thwart defenses. Ehrlich and Raven (1964) famously coined this phenomenon as “coevolution,” a concept that serves as a framework for the discussion within this chapter.

Beginning in the Early Devonian, and followed by a more extensive pulse at the Mississippian-Pennsylvanian boundary, plants have been exposed to herbivory, resulting in a vast array of trophic connections and evolutionary radiations. The chemical interactions that exist between plants and herbivores have long been documented, but the field of chemical ecology has experienced substantial developments over the past 40 years for a number of reasons including (1) increasingly successful identifications of organic molecules, (2) a merging of state-of-the-art chemical techniques with a desire to understand complex biological systems, and (3) the awareness that secondary metabolites play a significant role in complex multitrophic interactions (Harborne 1997). These advancements, biological phenomena to the development and merging of technologies for studying chemically mediated biological phenomena, have led to exciting opportunities in the field, creating an area ripe for integrative research and important ecological discoveries.

It was the seminal work of Fraenkel (1959) that highlighted the fact that secondary metabolites (compounds not directly involved in growth or reproduction) were not simply to be considered waste products of a plant’s primary metabolism



but that there may be ecological and evolutionary reasons for the existence of the overwhelming chemical diversity of these compounds, namely, defense against pathogens and herbivores. The composition of secondary metabolites in plants varies not only across different plant taxa but can fluctuate substantially among different populations, between individuals, among different organs, across developmental stages, and under varying environmental conditions. The dynamic nature of secondary metabolism has led to the development of a number of theories describing the various factors and selection pressures responsible for the qualitative and quantitative patterns of defense compounds observed today. Thus, it remains essential to explain the patterns of diversity and the distribution of various classes of secondary metabolites to mechanistically understand *how* they function in plant defense against herbivores.

Much research has demonstrated the role of plant secondary metabolites as defense mechanisms against insect feeding via direct toxicity, reducing digestibility, deterring feeding, and attracting the natural enemies of the herbivore. Similar to functional analyses of other plant traits, understanding the evolution of chemical defenses is evaluated in the context of plant fitness; thus, the term “defense” is usually reserved for traits that increase *plant* fitness while “resistance” refers to a trait that reduces *herbivore* preference, survival, and/or fecundity (Karban and Baldwin 1997). The distinction between these terms is important, and while few studies have truly demonstrated the defensive role of secondary compounds via observed increases in plant fitness, this chapter uses this term interchangeably with resistance, assuming the traits they describe result in a reduction in the impact of herbivores. While the potential selective pressures responsible for shaping defensive chemistry are discussed, the use of these terms (i.e., defense and resistance) is not meant to make any assumptions as to the driving forces behind the evolution of plant chemical traits.

Suites of plant secondary compounds that confer resistance to insects can be controlled genetically and expressed *constitutively*, providing a relatively constant barrier to attack, or they can be *induced*, produced in response to tissue damage. Plants possess the ability to assess stimuli in their environment and respond accordingly via several different potential modes of signaling (e.g., chemical, electrical, etc.), resulting in both general and herbivore-specific chemical responses. Furthermore, plants can respond to herbivory not only by altering the concentrations of defense compounds within their tissues but also by actively releasing volatile organic compounds (VOCs) into the environment. These volatiles serve as an indirect defense by recruiting the natural enemies of a plant’s associated herbivore(s). As such, recent research is moving beyond chemically mediated interactions between pairs of species and beginning to focus on large-scale multitrophic effects of plant secondary metabolites at multiple scales in time and space in an effort to understand their significance at the community and ecosystem level.

Although many plant secondary compounds have been shown to have antagonistic effects on herbivores, insects have also evolved counteradaptations to avoid, tolerate, and detoxify defense chemicals as well as use them to their benefit in response to predation, disease, and environmental variability. Many studies have stressed the importance of plant chemistry in driving the evolution of herbivore

preference and thus herbivore host ranges. Furthermore, the distributions of secondary compounds among plant taxa have been used as evidence of their defensive roles and their involvement in what has been termed a biochemical coevolutionary “arms race.” However, plant secondary metabolites not only respond to insect feeding but are also influenced by environmental variability. Changes in secondary chemistry associated with fluctuating environmental conditions and observed global changes to atmospheric composition can lead to altered trophic interactions and may have important feedbacks on ecosystem function, a topic beyond the scope of this chapter. To summarize, plant chemical responses are controlled by complex and interacting abiotic and biotic factors with cascading effects on food webs, communities, ecosystems, and atmospheric composition. Thus, a comprehensive understanding of the role of secondary compounds on ecosystems, including under-investigated multitrophic interactions, requires a multidisciplinary experimental approach that combines inference from evolution, chemistry, organismal science, and ecology.

---

## **A Primer on Plant-Herbivore Evolution**

### **Techniques and Analyses for Testing Classic Macroevolutionary Hypotheses**

The incredible diversity of plant-insect interactions, how they change over time, and the mechanisms driving them continue to spark the research of evolutionary biologists and ecologists. Particular interest has been given to the diversity of secondary metabolites and their role in shaping specialized plant-herbivore associations through their function as defensive compounds. Macroevolutionary hypotheses have suggested that reciprocal evolution of adaptations and subsequent speciation events have given rise to the existence and maintenance of this observed chemical diversity and thus the specificity that exists between plants and their associated herbivores. While testing these theories has remained challenging, the relatively recent advent of molecular analysis, phylogenetics, and other advanced genetic tools has now allowed scientists to begin to approximate the timing, patterns, distribution, and evolution of plant traits.

Phylogenetic approaches and historical biogeography offer complementary strategies for understanding the origin and evolution of plant-herbivore associations. To develop ecological hypotheses regarding the evolutionary histories of particular insect herbivores and their host plants, many researchers have opted to use phylogenetic data based on molecular analyses. The degree to which an association exists between monophytic clades of host plants and those of insect herbivores can be evaluated statistically and categorized as parallel cladogenesis (where associations established over long periods of time result in the cladograms of host plants and their insects exhibiting relatively high concordance) or diffuse arrangements. The best estimates of divergence within clades (groups containing a common ancestor and all its descendants) are not solely based on the timing of

molecular evolution (often termed the molecular clock) but are also calibrated to fossil records. In other words, not all insights on past plant-insect associations can be gleaned from phylogenetic analyses of modern taxa alone. The fossil record provides invaluable insight into the role of plants and insects in past ecosystems by offering data on the intensity of herbivore attack, recognizable damage types, levels of host specialization, and temporal trends of herbivore pressure on plant lineages. As will be discussed in more detail below, combining phylogenetic history, fossil record patterns, and manipulative field experiments will lead us into a new era of understanding the evolution and ecological significance of plant defenses.

## **A Framework for Explaining the Diversity and Function of Secondary Metabolites**

One central theme surrounding the evolution of plant-herbivore interactions is the synthesis and diversity of plant secondary metabolites and their role as defense mechanisms against insect herbivory. Due to the diverse defensive properties of these compounds (e.g., toxicity, deterrents, altered nutritional status, etc.) and the narrow host range of most phytophagous insects, it has been suggested that herbivores act as selective agents on the chemical makeup of plants. Plant secondary chemistry has, in turn, shaped the specialized associations between certain plant species and insects. In other words, many evolutionary biologists postulate that plants and their herbivores engage in *coevolution* or population-level processes of reciprocal adaptation of pairwise (two species) or diffuse (multispecies) interactions (see Futuyma and Keese 1992).

Over the past 60 years, a number of theories have emerged describing the diversity and evolution of plant-herbivore interactions beginning with the idea that insect host shifts were determined by shared host-plant chemistry (Dethier 1954). This hypothesis was then expanded upon by Fraenkel (1959), who suggested that plant secondary compounds serve as adaptive defenses against herbivory. The integration of these ideas culminated in the concept of coevolution set forth by Ehrlich and Raven (1964) where they provided a conceptual framework for the “escape and radiate” hypothesis, suggesting that a plant species may evolve a novel chemical defense that enables them to escape from most or all of its associated herbivores and radiate into new species. Over time, one or more insect species colonize plants within this new clade and adapt to its chemical makeup and undergo adaptive radiation themselves. These stepwise adaptive radiations, otherwise known as the “evolutionary arms race” (also known as the biochemical arms race hypothesis) between plants and herbivores may explain much of the biological diversity that exists between and within plant taxa. In addition, plants possess a range of biologically active compounds to effectively defend against a constant onslaught of attackers, which may also contribute to the maintenance of chemical diversity across taxa (Jones and Firn 1991). While high diversity of defense function is advantageous, it likely incurs trade-offs with other plant processes (e.g., growth, reproduction) with consequences for selection, a topic that will be

explored later in this chapter. Despite these strong theoretical foundations and years of empirical work, understanding the ecological and evolutionary processes that lead to variation in plant defenses among species remains a significant challenge. To demonstrate reciprocal coevolutionary adaptations requires information on *function* and selection; however, these data can be difficult to obtain and few studies have rigorously tested this hypothesis.

Beyond explaining the *diversity* of plant chemical compounds, the *role* these compounds play in mediating plant-herbivore interactions and the development of host specificity among phytophagous insects remains a highly debated topic in chemical ecology. Since Fraenkel's (1959) article, the past 50 years have seen the development, testing, and modeling of the ecological roles of secondary compounds, and the major functional roles of these compounds have been broadly acknowledged. Secondary metabolites are heritable and produced as a result of the selective advantages that they confer, which is consistent with their sophisticated structures, complex mechanisms of action, and multiple observable functions in nature. While the primary function of secondary compounds appears to be that of defense, in many instances defensive activities coincide with the same compounds also serving as herbivore attractants as well as transport, storage, and/or signaling molecules. However, the possibility that the evolution of function can occur through insect adaptation allows for much faster and more complicated evolutionary changes, a topic that will be discussed in more detail later in this chapter. In light of the diversity of secondary compounds, their myriad ecological and physiological roles, and the speed with which those roles change, a better understanding of the regulation of secondary metabolite production (i.e., gene expression) will offer welcomed insight into controls over the *quantities* of secondary compounds and changes in their function through time.

## **Ecological Costs, Trade-Offs, and the Emergence of Plant Defense Theories**

While the diversity of compounds found within plants results from interactions between plants and their major pests, a more challenging issue has been predicting *quantities* of secondary metabolites. A number of models have been put forth in an attempt to take into account the costs and benefits associated with the evolution of and investment in secondary compounds. There are costs to producing any trait, but these may be difficult to detect, particularly given the fact that plants contain hundreds of secondary compounds that potentially contribute to resistance against herbivory and trade-offs likely occur between suites of traits or syndromes (e.g., tolerance and resistance). In addition, the genotypes and/or the resource environment can mask the cost or trade-off between two traits (i.e., general vigor issues; Agrawal 2011). Furthermore, energy, resources, and other types of costs associated with strategies or syndromes introduce a level of subjectivity when making cost/benefit comparisons. Nonetheless, costs associated with producing secondary compounds should be measured in terms of fitness impacts on the plant.

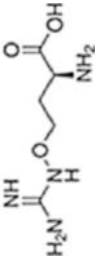
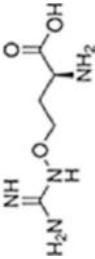
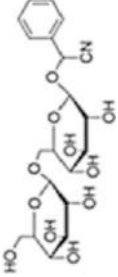
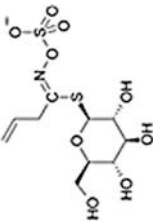
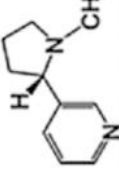
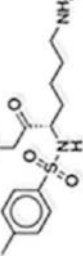
Thus, natural selection is expected to favor plants that possess a composition and concentration of defense compounds that not only maximize diversity but also minimize costs (Jones and Finn 1991).

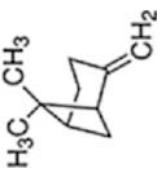
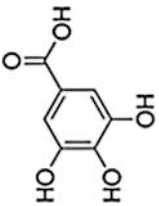
The idea that there must be a cost associated with the production of secondary compounds, or defense, was first put within an optimality framework by McKey (1974), who suggested that an increase in reproduction resulting from an increased allocation of resources to defense is likely due to a plant maintaining a certain level of biomass (e.g., forgoing attack by pests or pathogens). Applying this concept to risk assessment theory, Feeny (1976) developed the apparency model, where plants that are noticeable to herbivores are more likely to invest in defense in contrast to unapparent or ephemeral plants. However, assessing the apparency of a plant can be subjective; rather than focusing on how apparent plants are to particular herbivores, Janzen (1974) suggested that slow-growing plants, particularly those inhabiting resource-poor environments, should invest heavily in chemical defenses due to the “value” of each leaf to the plant. Plants growing in resource-rich soils would thus be less likely to invest in defense as they would be able to grow faster and better tolerate herbivory. Similarly, the Resource Availability Hypothesis set forth by Coley et al. (1985) postulated that abiotic resources (e.g., high-resource light gaps) are the driving factor behind plant evolutionary strategies or syndromes, with “escape” or pioneer species having few chemical defenses but rapid leaf expansion and low nutritional quality and “defense” species (understory) having high levels of chemical defenses. Around the same time, the Carbon/Nutrient Balance Hypothesis was developed (Bryant et al. 1983) and described how the supply of carbon and nutrients in the environment influences the production of plant defenses. Namely, if the C:N ratio acquired by a plant controls allocation of resources to plant functions, carbon-based defenses will be produced under nitrogen-poor conditions and more nitrogen-based defenses synthesized when carbon is limited. In the 1990s, Herms and Mattson (1992) offered a synthesis and expansion of the Growth-Differentiation Balance Hypothesis (Loomis 1932), stating that plant defenses are a result of a trade-off between growth and differentiation (i.e., processes that enhance the structure or function of existing cells) and a plant will only produce chemical defenses when sufficient energy is available from photosynthesis.

While each of the above hypotheses have been cited at one time or another as the theoretical basis for published studies on plant defense, considerable confusion remains, namely, due to the fact that (1) there is a large diversity of secondary metabolite structure and function (Table 1), (2) the hypotheses are not mutually exclusive and are difficult to test, and (3) contradictory results have led to the perception that there is no tangible theory of plant defense. For a more detailed account of plant defense hypotheses, see Stamp (2003). While each hypothesis and model study system has contributed to our current understanding of plant defenses, they also demonstrate how unrelated plant species have converged evolutionarily on suites of similar defense strategies.

Much evidence suggests that sets of traits have evolved independently to maximize fitness under given environmental and ecological conditions. For example,

**Table 1** Common classes of plant defense compounds

Chemical class	Common plant families	Modes of defense	Example(s)	Structure
<i>Non-protein amino acids</i>	Fabaceae	Mimick protein amino acids	L-canavanine <sup>a</sup>	
	Poaceae	Interfere with enzymes and neurotransmitters	L-DOPA	
<i>Cyanogenic glycosides</i>	Fabaceae	Toxicity of HCN	Amygdalin <sup>a</sup>	
	Rosaceae	Feeding deterrent	Linamarin	
	Linaceae		Dhurrin	
	Compositae			
<i>Glucosinolates</i>	Brassicaceae	Inactivation of proteins and nucleic acids	Sinigrin <sup>a</sup>	
	Tropaeolaceae	Growth inhibition		
	Capparidaceae	Feeding deterrent		
<i>Alkaloids</i>	Solanaceae	Alter enzyme activity	Caffeine	
	Papaveraceae	Inhibit DNA synthesis and repair	Atropine	
	Apocynaceae	Interfere with the nervous system	Nicotine <sup>a</sup>	
	Ranunculaceae		Strychnine	
			Coniine	
<i>Proteinase inhibitors</i>	Ubiquitous	Prevent degradation and turnover of proteins	Serine	
			Cysteine <sup>a</sup>	
			Aspartic	

<b>Terpenoids</b>	Ubiquitous	Interfere with insect molting  Disrupt cell membranes Inhibit ATP-synthase Interfere with nervous system Feeding deterrent	$\beta$ -pinene <sup>a</sup>				
			(E)- $\beta$ -farnesene				
			Avenacoside-B				
			Digitoxin				
			Squalene				
			Retinol				
			<b>Phenolics</b>	Ubiquitous	Protein and lipid peroxidation Protein inactivation DNA disruption and cell death	Capsaicin	
						Salicylic acid	
						Quercetin	
						Gallic acid <sup>a</sup>	
Flavone							
Psoralen							
Angelicin							

<sup>a</sup>Represents the example structure given for each class

different plant lineages have evolved the ability to make the same specialized metabolites present in other lineages or make different compounds that fulfill the same functional role. There are a number of reasons that multiple resistance traits may evolve together and repeatedly across species. First, most plants are subject to multiple attackers, with specific traits negatively impacting particular pests. Thus, diversifying resistance strategies will likely increase defense against a large number of potential herbivores. In addition, multiple resistance traits may be adaptive considering that some traits, while conferring defense under some circumstances, may fail to provide resistance under another set of ecological conditions. Finally, defensive synergism may provide higher levels of resistance than any single defense strategy alone, although evidence of this phenomenon remains scant.

That herbivory imposes natural selection on plants, particularly in terms of the defensive function of plant secondary metabolites, is well documented. However, not all secondary compounds are necessarily used for plant defense, and phylogenetic comparisons can elucidate the convergent evolution of suites of plant features or “defense syndromes” in response to particular herbivores and/or the environment (Agrawal and Fishbein 2006). The observed parallelism begs the question as to whether variation among plant taxa is mostly the result of shared biosynthetic pathways and minor genetic changes (i.e., similar phenotypic origins) or the result of differential histories and selection pressures. Unfortunately, the study of convergent evolution in plant defense chemistry is limited by (1) an incomplete knowledge of the secondary metabolites within each plant species (where the 200,000 identified to date is likely a gross underestimate) and (2) a lack of knowledge regarding the genes and biosynthetic pathways responsible for the production of these compounds. Studying the modes of action and ecological roles of different classes of secondary compounds will offer insights into the evolution of plant-insect interactions.

---

## **This Ain't a Scene, It's an Arms Race**

Over the past 350 million years, plants have developed a number of strategies for tolerating and defending themselves against the plethora of herbivores that rely on them for energy. One of these strategies involves the synthesis of over 200,000 secondary compounds that belong to various chemical classes, including nonprotein amino acids (5.1.2), cyanogenic glycosides (5.1.3), glucosinolates (5.1.4), alkaloids (5.1.5), proteinase inhibitors (5.1.6), terpenoids (5.1.7), and phenolics (5.1.8) (Table 1). All of these compounds differ not only in the biosynthetic pathways responsible for their production but also in their molecular structures and their physiological consequences for herbivores. These diverse compounds can exhibit a wide range of variation in terms of costs (ecological, resource, and energetic), where they are produced within an individual plant (leaves vs. roots) and across plant lineages (found across a wide range of plant taxa or restricted to specific genera), the timing of their accumulation/toxicity (constitutive vs. induced), their general palatability, and consequences for insects (deterrents vs. attractants).



The role of secondary compounds has been extensively explored within a coevolutionary framework, particularly the idea of plants possessing a “chemical armory” to avoid being overeaten and selective pressures in the form of insect feeding requirements. To complement the ideas and theories presented in the previous section, this chapter continues to discuss plant-insect interactions as mediated by secondary compounds in the context of biochemical coevolutionary theory, but with a more specialized focus on the types of secondary compounds produced, their associated costs, the specificity of their production, and the impacts they have on herbivores. The following section offers evidence and insight into the theories presented on the continuing coevolutionary arms race between plants and insects for mutual survival and the patterns of host utilization and diversity of plant secondary chemistry observed today.

### **Fitting Plants with Weapons in the Form of Chemicals**

A large number of secondary compounds play an important ecological role in plant defense against herbivores, but how many of these metabolites disrupt insect physiological processes and metabolism remains to be elucidated. Many secondary compounds appear to disrupt insect membrane function, namely, by inhibiting nutrient and ion transport as well as deterring signal transduction processes, metabolism, and hormone-controlled physiological processes. Furthermore, some classes of compounds are structurally similar to neurotransmitters and have been shown to interfere with insect neuroreceptors. Regardless of their molecular mode of action, all of these compounds are considered “toxic” to the herbivores that ingest them. However, the toxicity of a chemical is always relative, dependent on the dosage over time; the age, size, and health of the insect; the mechanism of absorption; and the mode of excretion. Furthermore, to minimize the risk of self-toxicification, many defense compounds are stored in specialized compartments, such as a vacuole, the apoplast, and resin ducts among other plant structures. The importance of the costs associated with these storage strategies and the timing of release will be discussed in more detail below. Next, a brief account is provided of some of the major classes of plant defense compounds, common plant families that produce them, and examples of their ecological role in mediating plant-insect interactions (Table 1).

#### **Nonprotein Amino Acids**

Nonprotein amino acids are widely distributed across plant taxa but are notably characteristic of legumes (Fabaceae) and grasses (Poaceae). In addition to serving as intermediates in the biosynthesis of primary metabolites and acting as nitrogen storage compounds (e.g., L-canavanine in legume seeds), nonprotein amino acids are some of the simplest N-containing secondary compounds with known toxic effects on herbivores. Most of the 300 known nonprotein amino acids act as antimetabolites by mimicking one of the 20 protein amino acids and, thus, being mistakenly incorporated into a nonfunctional protein resulting in unnatural function and death, such as the mis-incorporation of L-canavanine in place of L-arginine.

Other nonprotein amino acids, such as L-DOPA, have been shown to harm insects by interfering with essential enzymes, such as those responsible for the hardening and darkening of the insect cuticle. Yet other nonprotein amino acids mimic neurotransmitters (e.g., dopamine, norepinephrine), resulting in abnormal growth and development. While these compounds are fairly effective in defending plants against herbivores, there is a risk in deploying these nitrogen-rich compounds as defense agents. Some species have developed the ability to detoxify these compounds, converting them to usable forms of nitrogen, which can be an extremely limiting nutrient in many terrestrial ecosystems and particularly for insects.

### **Cyanogenic Glycosides**

While cyanogenic glycoside are not themselves toxic, when enzymatically broken down, they release hydrogen cyanide (HCN), which affects the terminal cytochrome oxidase system in the mitochondrial respiratory pathway, resulting in oxygen starvation and death. To prevent autotoxicity, the plant must take precautions during biosynthesis, forming multienzyme complexes which prevent the release of harmful intermediates. Following their production, plants then store these N-containing substances as inactive glycosides (a molecule in which a sugar is bound to a noncarbohydrate structure) in the vacuole separate from the cytoplasmic hydrolases ( $\beta$ -glucosidases and  $\alpha$ -hydroxynitrilelyases). Upon herbivore feeding, the cell structures are ruptured, including the vacuole, allowing the two substances to interact, resulting in the cleaving of the aglycone moiety and the conversion to HCN. Approximately 60 variations of cyanogenic glycosides have been identified and are characteristically found in more than 2,600 plant species, including ferns, gymnosperms, and angiosperms. Cyanogenic glycosides have been extracted from almonds and the fruits of the Rosaceae family (e.g., cherries, apples, plums, peaches, raspberries) and in several important crops, such as cassava (*Manihot esculenta*), sorghum (*Sorghum bicolor*), and barley (*Hordeum vulgare*). Despite the effective toxicity of HCN, its effect on herbivores is dosage dependent, as are most defense compounds, and some specialists are capable of tolerating relatively high levels of HCN. Furthermore, recent studies have suggested that the primary defensive role of cyanogenic glycosides does not appear to be its toxicity, but rather its ability to serve as an effective feeding deterrent due to its bitter taste.

### **Glucosinolates**

Close to 150 different glucosinolates, or mustard oil glycosides, have been identified within the Brassicaceae, Capparidaceae, and Tropaeolaceae plant families. Glucosinolates are biosynthetically related to cyanogenic glycosides as both are spatially separated from their hydrolyzing enzyme, in this case a thioglucosidase myrosinase. Similar to the production of HCN from cyanogenic glycosides, the enzyme and glucosinolate substrate come into contact upon tissue damage from herbivory, and the unstable aglycones are released resulting in various active compounds including nitriles and isothiocyanates. The latter hydrolysis product affects herbivores by reacting spontaneously with compounds containing unshared

pairs of electrons, mainly proteins and nucleic acids, making them inactive. However, the role of glucosinolates as defensive compounds is complicated by the extreme variation in their composition and concentration within species, between plant tissues, and across ontogenetic stages, providing both a challenge and opportunity for specialist and generalist insects. For example, while significant negative correlations have been shown between glucosinolate content and insect fecundity, several specialist species preferentially feed on Brassicaceae, using antennal receptors to locate their preferred hosts by the presence of glucosinolates.

### **Alkaloids**

Alkaloids are one of the most structurally diverse groups of N-containing secondary compounds and are present in ~20 % of higher plant families, including Solanaceae, Papaveraceae, Apocynaceae, and Ranunculaceae. More than 12,000 alkaloids have been identified to date, and these can be subdivided into more than 20 different classes including pyrrolidines, tropanes, piperidines, and pyridines. With individual alkaloids having the ability to carry out multiple functions, it is not surprising that these compounds can exhibit a variety of deleterious effects on metabolic function and physiology by affecting enzymes, inhibiting DNA synthesis and repair, and affecting the nervous system. In addition to the famous use of alkaloids extracted from hemlock to put the philosopher Socrates to death, other typical alkaloids include caffeine, atropine, and nicotine. Besides their well-known effects on vertebrates, including humans, alkaloids act as natural defense compounds by paralyzing and having toxic effects on herbivores, such as targeting insect postsynaptic receptors, as in the case of nicotine.

### **Proteinase Inhibitors**

Proteinase inhibitors do just that, inhibit different types of proteinases that occur in the herbivore gut, ultimately serving as anti-digestive proteins. Proteinase inhibitors interact with the active site of target proteases, attenuating protein processing and turnover by causing enzymes to become inactive, thus preventing the degradation of anti-nutritional or toxic proteins and interfering with digestion in the gut to prevent effective nutrient utilization. Plants contain a variety of proteinase inhibitors (e.g., serine, cysteine, aspartic), and the various classes are identified by the structure of their polypeptide backbone. Some proteinase inhibitors are found constitutively in seeds and tubers, likely because the integrity of these organs is essential for survival. Herbivore attack can also induce proteinase inhibitor gene expression, both locally and systemically. While the effects of proteinase inhibitors on herbivore mortality or performance are relatively minor, even small effects on development or fecundity may be ecologically relevant.

### **Terpenoids**

Terpenoids are ubiquitous across plant families, with over 22,000 of these lipophilic compounds having been described, and play multiple roles in plant defense. Terpenoids share a common biosynthetic origin and are synthesized from

five-carbon isoprene units creating monoterpenes ( $C_{10}$ ), sesquiterpenes ( $C_{15}$ ), diterpenes ( $C_{20}$ ), and triterpenes ( $C_{30}$ ). Mono- and sesquiterpenes are primary components of essential oils (e.g., those found in conifer resin), while diterpenes and triterpenes tend to have similar molecular structures as sterols and steroid hormones. In fact, some triterpenes, known as phytoecdysones, are mimics of insect-molting hormones and can disrupt larval development. Another group of triterpenoids, cardiac glycosides (e.g., digitoxin found in foxglove *Digitalis* spp.), are known to cause heart attacks in bird and mammalian herbivores if ingested in high quantities. However, some herbivores, such as the monarch butterfly, can overcome the dangerous effects of these compounds and store them safely within their bodies to avoid predators (see section “[Sequestration](#)” below). Saponins are yet another group of glycosylated triterpenoids. These compounds have detergent-like properties are found in the cell membranes of many plants species, and have been shown to disrupt cell membranes of herbivores as well as fungal pathogens. The literature is full of examples demonstrating the effects of plant terpenoids on herbivores, with a particular focus on toxicity and feeding/oviposition deterrence. While the mechanisms by which terpenoids directly act on insect pests remain to be elucidated for many systems, a number of studies have suggested that terpenes inhibit ATP-synthases, interfere with insect molting, and/or disturb the nervous system. In addition to directly defending host plants from their associated herbivores, many terpenes have been known to serve as plant indirect defenses, in which volatile compounds are exploited by the natural enemies of herbivores as host location cues (see section “[Indirect Inducible Defenses](#)” below).

## Phenolics

Similar to terpenoids, phenolics are a large, ubiquitous group of carbon-based secondary compounds, of which over 9,000 have identified, and include a wide variety of subclasses such as flavonoids, tannins, lignin, and furanocoumarins. Phenolics consist of a hydroxyl group ( $-OH$ ) bonded directly to an aromatic hydrocarbon group and are classified based on the number of phenol units in the molecule. Plants store phenolics in the vacuole. Some common and naturally occurring phenolic compounds include cannabinoids found in *Cannabis* spp., capsaicin in chili peppers (*Capsicum* spp.), and salicylic acid from *Salix* spp., which is used to produce aspirin. Phenolics can have negative effects on non-adapted insects, likely due to oxidative mechanisms in the midgut that result in the formation of superoxide radicals and other reactive oxygen species that can lead to protein and lipid peroxidation. Flavonoids play a variety of biological activities (e.g., phytoalexins, detoxifying agents, UV filters, allelochemicals, etc.), protecting plants from different biotic and abiotic stresses. However, these multiple roles make the interpretation of experimental results regarding flavonoids rather difficult when trying to elucidate their primary role in plant resistance. Tannins have also been shown to be toxic to insects due to their ability to bind to salivary proteins and digestive enzymes (e.g., trypsin and chymotrypsin), resulting in protein inactivation, the inability to gain weight, and death. Another group of phenolic compounds, furanocoumarins, is found primarily in species of the

Apiaceae and Rutaceae families and is also produced in response to pathogen or herbivore attack (see section “[Induced Responses](#)” below). Furanocoumarins are activated by UV light and are highly toxic to herbivores as a result of their integration into DNA (although they can also interact with protein and lipids), thus resulting in cell death. The interaction between wild parsnips (*Pastinaca sativa*) and the parsnip webworm (*Depressaria pastinacella*) is mediated by furanocoumarins, and the work of May Berenbaum has elegantly demonstrated the coevolution of the efficient detoxification systems possessed by webworms in the form of cytochrome P450s to cope with the presence of furanocoumarins and the response in the chemical evolution in wild parsnips (see Berenbaum citations in “[Further Reading](#)”).

### Advantages of Chemical Mixtures

Plants contain a complex array of secondary compounds from varying chemical classes, the quality and quantities of which are subject to change under varying environmental and biotic conditions. Understanding the ecological and evolutionary roles of secondary compounds in the context of their variation across plant taxa, populations, individuals, and even organs on the same plant remains a major focus across scientific disciplines. Some studies have suggested that it is to the plant’s advantage to employ such diverse mixtures of secondary compounds so to be protected against a wide range of current (and potential) herbivore attackers. Another benefit of chemical mixtures is the idea of synergism: that two or more defense components together provide greater toxicity or deterrence to herbivores than the equivalent amount of a single defensive compound alone. Synergistic effects of mixtures have been observed among compounds in the same chemical class and among compounds within different classes, suggesting that the effects of individual compounds on herbivores should be assessed alone as well as within the context of their naturally occurring background chemicals. A number of factors have been postulated as contributing to the effectiveness of synergisms, including the idea that one component may facilitate the transport of another to the target sites, for example, by altering the permeability of cell membranes. Studies on conifer resin have suggested that lower molecular weight monoterpenes may serve as a diluting factor, making the resin more soluble for diterpenes and the solution more fluid overall. Another potential way in which mixtures are more effective is that compounds can affect each other’s metabolism, for example, classes of compounds inhibiting detoxification enzymes can lead to greater overall toxicity. In addition, volatile secondary compounds have been shown to attract the natural enemies of herbivores (see section “[Induced Responses](#)” below), and complex mixtures of these airborne defense compounds may be critical for the level of sophisticated communication needed for parasitoids to effectively find suitable hosts. Regardless of the mechanism by which chemical mixtures are effective forms of plant defense, the benefits accrued from deploying a range of chemical defense compounds must outweigh the costs of producing them.

So how do such chemically complex mixtures arise in different plant species without significant metabolic costs? Recent molecular studies have shown that secondary metabolism is characterized by highly branched biosynthetic pathways that result in a variety of target molecules from only a few different precursors supplied by primary metabolic processes. In many cases the same basic building blocks are repeatedly added to make a range of compound intermediates of different sizes, which can undergo further diversification via the activity of enzyme families (e.g., terpene synthases). In addition to the presence of many different synthases, each synthase enzyme is capable of forming multiple products from a single substrate, lowering the metabolic cost of producing chemical mixtures while maintaining a relatively high level of structural diversity. Thus, each pathway has the ability to produce mixtures that maximize fitness and survival while decreasing metabolic and ecological costs. For a detailed account of the metabolic origins and ecological benefits of plant chemical mixtures, see Gershenzon et al. (2012).

## **Spatiotemporal Patterns of Plant Secondary Chemistry Alter Herbivore Performance**

### **Spatial Heterogeneity**

The interactions between plants and their associated herbivores are contingent upon the condition of both species as they coexist across space and time. Yet the secondary chemistry of plants is quite variable not only between taxa, but within species, between individuals, and even among plant organs, resulting in a highly heterogeneous foraging environment. The fact that most plant defenses are not evenly distributed within an individual is largely due to a combination of both environmental and genetic factors that are exacerbated by selection via biotic agents. This “patchiness” occurs concurrently at multiple spatial scales, and its effects on herbivory are dependent upon the herbivore’s power of perception as well as their foraging mobility. This phenological variation is central to the ecology and evolution of plants, affecting both intra- and interspecific interactions, which have important implications for plant fitness and the development of theories seeking to describe plant function.

The Optimal Defense Theory states that organisms evolved to allocate their defenses in such a way as to maximize fitness (McKey 1974), resulting in allocation toward plant parts that confer the greatest fitness value (i.e., reproductive organs, developing leaves). Substantial variation exists in levels of secondary compounds found within leaves, flowers, roots, and stems, which may reflect movement of compounds within the plant or the fact that some plant parts are relatively more expendable than others. To most effectively serve as plant defenses against leaf-chewing insects, secondary metabolites tend to be located where they are most likely to have the greatest effect on herbivore attackers, namely, at the plant surface. Studies have identified a range of secondary compounds to be present in such

structures as glandular hairs or trichomes, leaf waxes, leaf resins, latex, and in the vacuoles of epidermal cells. While the Optimal Defense Theory predicts roots to have lower levels of defense compounds compared to shoots due to a lower probability of attack, recent work has shown roots to contain a wide range of secondary compounds at relatively high concentrations. In fact, root herbivores have been shown to do as much or even more damage to wild plants as aboveground feeding insects. These findings have, among others, resulted in a recent body of literature focused on how plant secondary compounds mediate interactions between root and shoot herbivores and vice versa. While aboveground herbivory can alter belowground interactions via changes in root chemistry, leaf damage has also been shown to induce the production of secondary compounds in petals, nectar, and pollen, which may defend the plant against florivores but also deter pollinator visitation and potentially plant fitness (see section “[Induced Responses](#)” below). Understanding these differential pressures and allocation strategies will ultimately aid in our understanding of how organisms adapt, evolve, and express particular traits.

In addition to variation within plants, there is also considerable variability in leaf secondary chemistry among different individuals within a population. While trees that exhibit higher levels of defense compounds tend to experience lower levels of herbivore damage, this can come with a significant cost to the tree in terms of biomass (fewer leaves) resulting in cascading negative effects on nutrient acquisition and fitness. In some cases, patchiness can benefit the herbivore if it results in particular plant species expressing low levels of defense compounds clumped together in space (Moore and DeGabriel 2012). However, in other instances, an herbivore’s search for a suitable host can be similar to trying to find a needle in a haystack, with the herbivore being forced to spend more time moving within the canopy and thus increasing its chance of being predated upon by its natural enemies. Furthermore, the heterogeneous chemical environment of the canopy can result in patches of plants defended by different secondary compounds, which can decrease the ability of herbivores that rely on mixed diets to ameliorate the detrimental effects of some secondary compounds. Canopy and landscape heterogeneity in levels of defense compounds can also lead to associational resistance, where plant susceptibility to insect pests is influenced by the quality and proximity of neighboring plants (Barbosa et al. 2009). Associational resistance can occur if herbivores select hosts at the *patch* scale and plants gain additional resistance if neighboring plants are unpalatable. However, having unpalatable neighbors can also result in associational susceptibility if herbivores forage at an *individual* plant scale, where the contrast in nutrient value and levels of defense become more apparent. However, whether or not variability in defense compounds among individuals results in associational resistance or susceptibility depends on the herbivore’s specific foraging movements and host location strategies, a topic that requires coupling observations in natural landscapes with foraging theory (see Moore and DeGabriel 2012).

## Temporal Heterogeneity

Secondary metabolites can change in response to a plant's developmental trajectory (ontogeny) and to seasonal conditions (e.g., water availability, temperature, photoperiod), with subsequent effects on a plant's physiology and metabolism. Meta-analyses performed by Barton and Koricheva (2010) revealed general patterns for defense compounds, particularly that concentrations of secondary metabolites tend to increase during seedling growth but decrease during leaf development. Furthermore, they found that most metabolites remain relatively stable in mature leaves through the season, with some changes in composition for specific compounds (e.g., tannins and lignin). Temporal shifts in patterns of secondary compounds may result from a number of different mechanisms, including a potential dilution effect as other metabolites accumulate in greater concentrations over time, translocation of compounds from one plant tissue or organ to another, time lags associated with the differentiation of specialized storage structures (e.g., resin ducts, trichomes), altered foliar concentrations due to volatilization (particularly with changes in temperature), and/or catabolism (Koricheva and Barton 2012). But what are the evolutionary causes responsible for temporal changes in plant defense compounds? A number of theories (e.g., the Growth-Differentiation Balance Hypothesis) suggest resource and metabolic constraints, substrate-level competition, and trade-offs with other plant processes, although little support for these mechanistic hypotheses has emerged. Another potential explanation for the evolution of temporal changes in plant defense compounds is that they are adaptive responses to selection pressures imposed by herbivores. For example, a number of studies support the Plant Apparency Hypothesis (Feeny 1976) with higher levels of secondary compounds observed in young developing leaves and a general increase in plant allocation to chemical defenses through ontogeny. The Optimal Defense Theory (Rhoades 1979) is also supported with many studies showing the majority of chemical defense present in high concentrations in young leaves, thus allocating resources to defenses proportional to the risk from herbivores and the value of the tissue to fitness. While there is a significant amount of support for the idea that herbivores drive temporal patterns of plant defense, other selective agents (e.g., pollinators, pathogens, large herbivores, plants, and the environment) may also play a role when particular plants defense compounds are expressed.

Regardless of how temporal patterns of plant defenses evolved, it is well known that these changes have profound effects on herbivores and the amount of damage sustained by plants. Variation in herbivore preference and performance on leaves of varying ages across the season may result from temporal changes in the concentration and composition of defense compounds. Likewise, changes in plant chemistry may also affect when herbivores can feed. For instance, long-lived leaves of tropical species are most susceptible to herbivory during the short period of leaf expansion, at which time these leaves usually exhibit the highest concentrations of secondary defenses (Coley et al. 1985). However, it is difficult to attribute changes in herbivore feeding with seasonal changes in secondary compounds alone considering the many other physiological, and thus nutritional, changes simultaneously occurring.



Further complicating the interpretation of how temporal changes in defense compounds alter herbivore performance is the lack of consideration for the third trophic level. Volatile secondary metabolites indirectly serve as plant defense compounds by attracting the natural enemies of herbivores (see section “[Indirect Inducible Defenses](#)” below), and their composition and concentrations are also impacted by leaf age and the overall developmental stage of the plant. Being more or less attractive to parasitoids during different phases of the herbivore’s life cycle can have profound effects on the effectiveness of parasitoids and result in important consequences for current herbivore population densities and the potential for future outbreak events.

Not only do plants exhibit an astounding level of variability in the levels of defense compounds present in their tissues over relatively short temporal scales (e.g., a growing season), but plant chemical defense has also been shown to respond to temporal changes on a large environmental scale, including shifts in soil conditions and competitive interactions among plants during succession. Herbivore communities may also vary temporally with succession, particularly in response to altered plant composition. Thus, it is not surprising that certain types and concentrations of plant chemical defenses may be more prominent and effective during specific successional stages. While plant defense theories (e.g., the Resource Availability Hypothesis) suggest mechanisms to explain the composition and concentration of plant secondary compounds during succession, many alternative scenarios have been observed. For example, greater environmental heterogeneity in late successional communities may tend to increase intraspecific variation in the quantity or quality of defense compounds, independent of nitrogen or carbon availability relative to demand. Unfortunately, the level of intraspecific variation in secondary chemistry among individual plants and the mechanisms responsible for altering concentrations during succession remains to be elucidated. A greater understanding of the specific processes structuring these observed patterns will likely be gained by quantifying the distribution of defense compounds within natural populations.

---

## Running to Stay in the Same Place

Current patterns of herbivory are dynamic and are shaped by both past evolutionary forces and continual environmental and phenotypic variability. Most herbivores demonstrate specific feeding habits and are associated with only one or a few plant genera or only a single plant family. Even herbivores with broad feeding patterns have specific adaptations to feed on such a range of hosts. Regardless of the range of host plants used by phytophagous insects, many studies have stressed the importance of plant chemistry in driving the evolution of herbivore preference. The distributions of secondary compounds among plant taxa have been used as evidence of their defensive roles and their involvement in the biochemical coevolutionary arms race. Natural selection imposed by herbivores causes the evolution of a new plant deterrent/resistance trait, or compound, which reduces susceptibility to herbivory. In response to the chemical defenses of plants, much evidence has been put forth

indicating reciprocal evolution in their associated herbivores and the tendency of insects to overcome these barriers to herbivory by adapting their feeding habits, preferences, and detoxification mechanisms to cope with the myriad secondary compounds they may encounter. However, it is important to note that chemical constraints do exist for many herbivores, limiting their host-use ability, with only 10 % of all species feeding on more than three plant families (Bernays and Graham 1988).

Due to the multifunctionality of most plant defense compounds (antifungal, antibacterial, frost tolerance, nutrient storage, signaling, etc.), the idea that selection for particular chemical profiles within plants is primarily imposed by herbivores remains controversial. Given the vast number of herbivores that attack a single plant species across its entire life and the array of compounds that a plant expresses, is it possible that all associated herbivores act as selective agents? And are the secondary compounds present the result of biotic and/or abiotic pressures? In light of the early reliance on circumstantial evidence supporting coevolutionary theory (e.g., Ehrlich and Raven 1964), other possibilities were put forth to explain host-plant use by herbivores including host-finding mechanisms, mate location, and natural enemies. However, more recent studies using quantitative genetics have detected genetic-based variability in chemistry, heritability, correlations to herbivore preference, and the type and direction of herbivore selection. These findings offer strong evidence that herbivores are the primary agents of natural selection on some specific secondary compounds. While new molecular approaches are providing insight into the generality of these coevolutionary processes, determining the defensive function of each compound specific for each herbivore attacker separate from its other ecological functions remains a difficult task. The following section discusses the evolution of feeding deterrents as well as the reciprocal counteradaptations of herbivores to cope with these defense compounds.

## **Basis of Plant Selection and Evolution of Feeding Deterrents**

It is well-accepted that most plants are unpalatable to most insects due to their secondary chemistry (Bernays and Chapman 1987). It is assumed that these avoidance responses are due to particular compounds acting as feeding deterrents or being toxic to the herbivore, thus playing a major role in insect preference and host selection. However, it is important to note that the toxicity of a compound may be subtle, incomplete, difficult to measure, or nonexistent (i.e., the compound may serve as a deterrent for other reasons than toxicity) and the relative effects on herbivore fitness can be equally challenging to assess. Furthermore, even chemicals referred to as feeding stimulants/attractants can play a defensive role in so much as the insect's dependency can become a hazard if the substance is not present and results in the inability to feed (Harborne 1997). Thus, any chemical affecting insect feeding plays a role in plant defense.

Given the range of plant secondary chemicals that are present within and among plant families and their effects on herbivores, insects can exhibit a variety of taxonomic relationships with plants. In light of the chemical constraints imposed

upon insect host use, it has been suggested that the development of novel insect detoxification systems may have allowed insects to use other hosts in different plant taxa, resulting in the radiation of local populations into new species. Thus, insects can be polyphagous (eating any plant, such as leaf cutting ants and locusts), oligophagous (feeding on relatively few related species belonging to one or only a few plant genera or families, such as danaid butterflies), or monophagous (feeding on a single plant species, such as the silkworm on mulberry leaves). Most insects are mono- or oligophagous and have been shown to feed on obviously toxic plants, and some even have the ability to exploit these defense compounds for their own use (see section “[Sequestration](#)” below). In accordance with coevolutionary theory, an evolving pattern of feeding deterrents has arisen across plant taxa with a trend toward chemical complexity in structures. While the production of such a diversity of chemical structures serves as an effective defensive function for plants, some insects have evolved defense mechanisms to cope with the negative effects of these compounds. In addition to the adaptive processes of insects, biochemical information has also shed light on the plasticity of plants and their energetic, metabolic, and chemical responses to producing costly chemical defenses following herbivory (see section “[Induced Responses](#)” below). Taken together, these interactions (insect detoxification mechanisms and the relative costs and benefits of plant chemical defenses) offer insight into the diversity of both plant and insect species and the theory of reciprocal evolutionary interactions mediated by plant chemistry.

## **Insect Detoxification Systems**

Phytophagous insects have evolved a range of mechanisms to avoid the plant chemical defenses present within their host plants. These strategies include avoidance, tolerance, impermeable guts, accumulation/sequestration, and/or detoxification. The insect enzymatic detoxification system requires both the degradation and neutralization of plant toxins and several detoxification systems having been described (Després et al. 2007). One of the best-documented mechanisms of detoxification occurs via cytochrome P450 monooxygenases (P450s), which are capable of oxidizing a variety of lipophilic compounds and converting them into polar molecules prior to their absorption by gut tissues. Given the variety of defense compounds to which insects are exposed, P450 transcriptional responses can be quite complex. Yet despite their importance in the ability of insects to cope with the presence of plant toxins in their diet, the activity of P450s has been studied in only a few insect species. Glutathione S-transferases have also been shown to effectively detoxify allelochemicals by conjugating reduced glutathione into electron acceptors, resulting in less toxic metabolites. Esterases are another group of detoxification enzymes that possess the ability to hydrolyze esters and amides, converting them into more polar substances for easier absorption. In addition, some species of Coleopteran and Lepidopteran herbivores are capable of adapting to dietary proteinase inhibitors by selectively upregulating the production of proteinases insensitive to inhibition. Regardless of the mechanism employed by the insect, the

production or upregulation of these detoxification enzymes is induced by exposure to toxic plant secondary compounds or wound signaling molecules present in the host plant (e.g., jasmonate and salicylate). The modified compounds can then either be stored or excreted, reducing their toxicity to the herbivore. However, detoxification can incur significant metabolic and energetic costs, resulting in trade-offs between the ability to metabolize plant secondary compounds, growth, reproduction, immunity, and overall insect fitness.

## Sequestration

In addition to having the ability to biochemically digest/assimilate plant defense compounds present in their hosts, herbivores can also make use of secondary compounds obtained from plants by storing them in their own body tissues and integument. An impressive variety of plant defense compounds can be sequestered by insects including alkaloids, cyanogenic glycosides, glucosinolates, and isoprenoids. More than 250 herbivorous insect species in six orders have demonstrated the ability to sequester these metabolites from at least 40 plant families (Opitz and Müller 2009). The amounts of sequestered compounds found in insects can vary dramatically, in part due to the variability of secondary metabolites present in host plants (Nishida 2002). The amount of sequestered compounds can also differ between insect sexes depending on their use for reproductive purposes (e.g., serving as precursors for pheromone production, nuptial gifts or spermatophores, or offspring protection). In many Lepidoptera and Hymenoptera, sequestered compounds can also be transferred from the leaf-chewing larval stage to the adult stage, a strategy that can be used to the insect's advantage by making adults unpalatable to natural enemies. One of the most famous examples of defense chemicals benefiting insects through sequestration is that of the specialist monarch butterfly (*Danaus plexippus* Nymphalidae), its milkweed (*Asclepias syriaca*) host plant, and its natural predator, the blue jay (*Cyanocitta cristata*). The milkweed provides monarchs with cardiac glycosides, which they sequester, rendering them poisonous to most vertebrates. Blue jays that attempt to eat monarch butterflies tend to have a strong visceral reaction to the toxic compounds and learn to associate the markings of the butterfly with this response, thereby learning to avoid them in the future. However, it is important to note that the benefits gained by monarchs from cardiac glycosides do come with a cost: monarchs are negatively affected when feeding on milkweed plants with low nitrogen levels (having to consume more plant tissue per day and making them more vulnerable to predation) and can also be negatively affected by high levels of cardiac glycosides and/or latex. Thus, the heterogeneous nature of individual plant quality within a population plays an important role in host choice via the ability to cope with a plant's defense mechanisms.

In order to successfully sequester and exploit plant defense compounds without inflicting harm on themselves, insects have evolved a number of interesting physical and biochemical strategies. To avoid autotoxicity, insects are forced to

construct specialized storage structures (e.g., glands) to compartmentalize the toxins away from their hemolymph. However, demonstrating the metabolic and ecological costs associated with this adaptation has proven difficult. In addition to storing these defense compounds, insects must also possess the ability to sequester certain compounds out of the diverse assemblage of chemicals most plants express. Thus, insects rely on selective transporters to carry specific compounds from the gut to the hemolymph and from the hemolymph into specialized compartments for storage. While potential transport mechanisms for polar and nonpolar defense compounds have been proposed, more work is needed to elucidate the details of these biochemical pathways. In addition to utilizing transport mechanisms to store plant toxins, insects can also modify sequestered compounds prior to storage (e.g., via epimerization and re-esterification) or alter them into more polar compounds to more easily facilitate excretion. Sequestration has also been observed by the third trophic level (parasitoids) and, in a few cases, fourth trophic level (hyperparasitoids), but the physiological adaptations of these organisms to toxic metabolites remain unknown. Despite the benefits of sequestration, the costs to the herbivore can have important consequences for herbivore immunity and fitness; however, this topic is beyond the scope of this chapter.

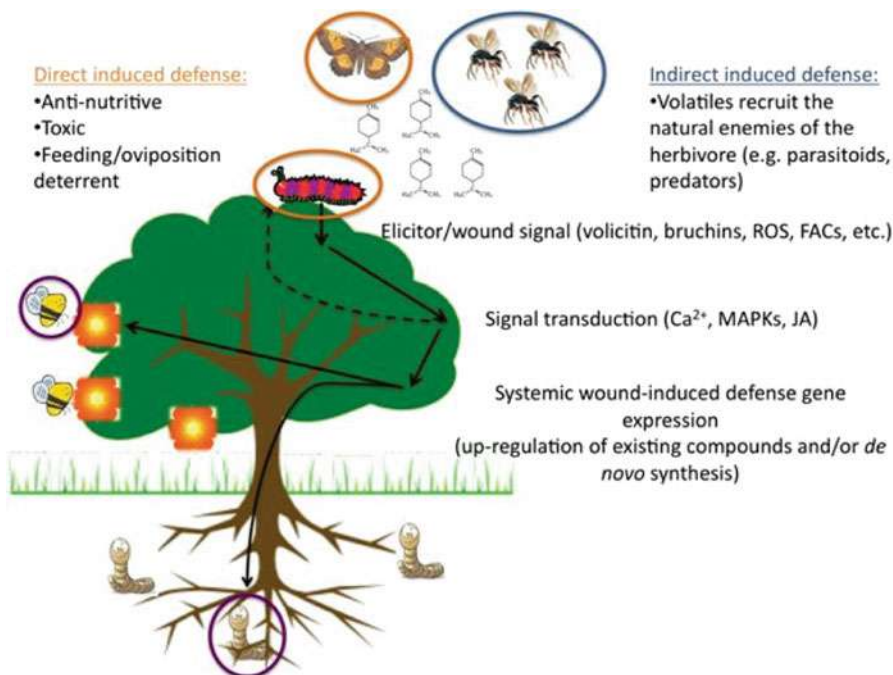
---

## Induced Responses

An element of plant-insect interactions that eludes a simple evolutionary description is that of induced responses. Despite the amount of attention induced resistance has received over the past few decades, there remains a significant gap in our knowledge pertaining to the evolution of herbivore-induced specificity of defense strategies and perception. In addition, the heritability of herbivore-induced responses is yet to be determined, particularly the genetic basis for hormonal signaling, the interactions between pathways, and the selective forces that act on these traits. In terms of the latter, the influence of induced responses across trophic levels within communities is complex, eliciting various responses in different arthropod species resulting in varying degrees of selection. While more work is needed, it is clear that a community perspective is critical to understanding costs of chemical defense syndromes and the evolution of specificity in plant induced responses (Agrawal and Fishbein 2006)

## Regulation of Costly Defenses

While the previous sections in this chapter focus mostly on constitutive defenses (i.e., defenses preformed before insect attack), another mechanism of plant chemical defense known as inducible defenses will now be discussed (Fig. 1). For a plant response or defense to be considered “induced,” the plant must first perceive damage, which initiates a downstream molecular signal (i.e., signal transduction), resulting in the synthesis of novel secondary metabolites or an increased production



**Fig. 1** Simplified scheme of the signal cascade involved in plant perception of herbivory and the synthesis of secondary defense compounds. *Solid lines* represent systemic upregulation of genes responsible for secondary metabolite production (*leaves, flowers, and roots*) and the *dashed line* indicates a local response at the area of wounding (*leaf level*). Resulting direct and indirect herbivore-induced defenses mediated by changes in plant chemistry are indicated across trophic levels with *colored circles*. *Orange circles* represent direct defenses in the form of reducing nutritive value, toxicity, and/or volatile feeding and oviposition deterrents for conspecifics, and *blue circles* represent indirect defenses or the volatile attraction of the natural enemies of herbivores, while *purple circles* indicate variability in the attractive or deterrent properties of altered chemistry in distal plant organs

of existing secondary metabolites (e.g., Kessler and Baldwin 2002). Herbivore-induced changes in plant chemistry can have either *direct* effects on the susceptibility of host plants to insects or can serve as attractants to the natural enemies of the herbivore, serving as an *indirect* defense (see section “[Indirect Inducible Defenses](#)” below). As with constitutive defenses, inducible defenses can only truly be selected as defense systems if there is heritable variation and if the plants experience a higher fitness by exhibiting the induced chemical response than not. However, because a true measure of fitness is difficult to obtain, proxies must be used and the results can be inconsistent and variable. As noted in the beginning of the chapter, the use of the term “defense” makes no assumptions about selection by herbivores, only that the trait defends the plant. Whether it evolved specifically to do so is another issue and it is difficult to determine the specific selective factors that shape a trait. Fortunately, phylogenetic reconstructions have begun to offer

insight into the validity of the coevolutionary theory and the history of induced responses. By placing theories describing the variability of induced defenses within the framework of costs and constraints, one can hope to begin to understand the evolutionary development of induced chemical traits.

Investing in compounds that defend plants can be costly and plants must allocate their finite resources among defense, growth, and reproduction as needed. Thus, investing in the synthesis of defense compounds unnecessarily can be directly costly in terms of fitness. However, if the compounds conferring defense also provide other benefits, such as dissipating heat, providing structure, etc., then the ecological cost may be relatively little in terms of plant fitness, even in the absence of herbivores. Plants may also pay a cost in terms of preventing autotoxicity, by synthesizing structures to safely store toxic compounds if the defense is constitutively maintained. Plants may also incur resistance costs, which can result from higher-level ecological interactions, such as in the case of compounds serving as deterrents against generalist herbivores as well as pollinators (Kessler and Baldwin 2002). The inconsistent phenotype expressed by plants with induced defenses may benefit the plant by resulting in a lag in insect counteradaptations, as herbivores are less likely to adapt to defenses that are intermittently expressed as compared to those they encounter on a more regular basis. Overall, in the absence of herbivores, it appears to be in the best interest of the plant to avoid the aforementioned metabolic and ecological fitness costs, favoring the evolution of inducible defenses. Despite the vast number of induced secondary compounds described and their potential roles in plant protection, definitive proof of the particular defense function of each compound remains to be determined in most cases.

## Plant Perception and Signal Transduction

Plants possess the ability to recognize and respond in a relatively sophisticated way to insect attack, as opposed to casual mechanical wounding, resulting in a variety of herbivore-induced chemical responses. Each cell can perceive specific “danger” signals and transmit this information systemically to prevent future attacks and defend itself either directly or indirectly. In comparison to plant-pathogen interactions, relatively little is known regarding molecular recognition and active response to insect herbivores. However, it is to the plant’s benefit to be sensitive to the multitude of life histories and feeding behaviors employed by its enemies. Thus, plants possess a recognition system that involves the perception of molecules or elicitors found in the saliva or secretions of insects that enter the plant following injury or feeding (see Kessler and Baldwin 2002; Howe and Jander 2008; Wu and Baldwin 2010 for reviews of this topic). One of the most widely studied groups of elicitors is fatty acid-amino acid conjugates (FACs, e.g., volicitin in *Spodoptera exigua* oral secretions). While the mechanism by which plants perceive FACs remains to be elucidated, the involvement of an FAC-specific receptor dependent on jasmonic acid signaling has been proposed. Another identified group of elicitors



is the inceptins, which are produced when a plant ATP-synthase subunit is cleaved in the midgut of the insect. Small molecular elicitors known as caeliferins can also induce chemical responses as well as lytic enzymes, such as  $\beta$ -glucosidase (isolated, e.g., from *Pieris brassicae*), glucose oxidases (*Helicoverpa zea*), alkaline phosphatase (piercing whitefly *Bemisia tabaci*), and watery digestive enzymes from aphid saliva. Furthermore, the feeding behaviors of insect larvae (e.g., speed, mode, frequency) can also be differentially recognized by the plant and play an important role in the specificity of the induced response. In addition to herbivore feeding, plants can also perceive insects' oviposition activities and express induced direct or indirect defenses in response. Bruchins (isolated, e.g., from the oviposition fluid of pea weevils) have been shown to result in neoplasma growth while benzyl cyanide (isolated from the large cabbage white butterfly *Pieris brassicae*) can induce the arrest of the parasitoid *Trichogramma brassicae* on Brussels sprout (*Brassica oleracea* var. *gemmifera*). While relatively little is known about *how* plants perceive herbivores, many small molecules have been identified in the complex signaling networks responsible for deploying the appropriate downstream defenses.

The use of model plants, artificially generated mutants, sequencing technologies, microarrays, and transcriptional profiling tools has greatly enhanced our understanding of the genetic basis of plant signaling and stress response. A number of regulatory networks have been suggested to mediate herbivore-induced responses in plants including  $\text{Ca}^{2+}$  ion fluxes, mitogen-activated protein kinases (MAPKs), jasmonic acid (JA), salicylic acid (SA), ethylene (ET), and reactive oxygen species (ROS). Molecular and genomic tools are being used to uncover the complexity of the induced defense signaling networks that have evolved during the arms race between plants and their attackers. While descriptions of each of these signaling pathways is beyond the scope of this chapter, many studies have shown JA, SA, and ET to be key players in the regulation of the signaling cascades responsible for induced defenses. Following insect attack, plants produce varying amounts of SA, JA, and ET, which contribute to the specific induced defenses that are synthesized. Plants may be attacked by a number of different insect pests, requiring regulatory mechanisms that can adapt with the various challenges they encounter. Thus, cross talk between induced defense signaling pathways not only provides flexibility in a plant's response due to the antagonistic or synergistic interactions between the hormones produced, but allows for a level of specificity while minimizing energy costs. While cross talk between pathways generally aids the plant in determining which defense strategy to deploy, some insects have evolved to manipulate plants for their own benefit by suppressing or adjusting the production of induced defenses. For example, some herbivores may activate the SA-signaling pathway, which antagonistically interacts with JA-dependent defenses, thus resulting in enhanced insect performance. Despite the progress in identifying molecular mechanisms responsible for interactions between defense signaling pathways, explaining the evolution and maintenance of variation in induced responses and its effect on the fitness of plants within complex communities remains a challenge.



Regardless of the signaling responsible for the induced responses observed in herbivore-challenged plants, the altered plant chemistry may affect the insect pest or other herbivores that attempt to use the plant in the future. To this end, the induced response may be considered rapid or delayed (Haukioja 1991), affecting herbivores during the season when damage occurs *or* in subsequent seasons with consequences for herbivore population dynamics. Whether rapid or delayed, induced defense has been observed in many diverse plant families (mostly long-lived perennials) within a wide range of habitats and across multiple spatiotemporal scales, from single leaves to entire trees, from hours to years. Despite the plant families that exhibit herbivore-induced defense via changes in plant chemistry, this response is certainly not considered ubiquitous, and in fact, some plants experiencing herbivore damage have been described as better hosts for insects, with subsequent increases in herbivore performance and survival (i.e., induced susceptibility). Furthermore, when and where plant defenses will be found are dependent on inherent plant growth rate, plant ontogeny, the associated selective environment, the type and extent of herbivore damage, and the evolutionary history of the plant-insect interaction (Karban and Baldwin 1997). Thus, the type and level of induction can vary significantly in different systems, making it difficult to apply generalizations as to how plants are perceived by insects, the effects on individual herbivores and higher trophic levels, and the consequences for herbivore population dynamics at the community level.

## Direct Inducible Defenses

Direct defenses affect the susceptibility to and/or the performance of attacking insects, resulting in an increase in plant fitness (Kessler and Baldwin 2002). Chemical defenses are typically categorized by their mode of action, namely, either anti-nutritive or toxic, with the former affecting either pre- (e.g., limiting food supply) or post-ingestion (e.g., reducing nutrient value) processes with the latter causing growth and development disruptions to the herbivore (Chen 2008). For example, proteinase inhibitors affect post-digestive processes, serving an anti-nutritive role, whereas alkaloids, terpenoids, and phenolics have all been shown to be toxic to a number of generalist herbivores, resulting in a trade-off between detoxification and growth/development. The release of volatiles in response to herbivore feeding can also provide a direct defensive benefit by deterring further conspecific feeding and oviposition. Deception is another way in which plants use herbivore-induced volatiles to their advantage, such as in the case of the sesquiterpene (*E*)- $\beta$ -farnesene, which is also an aphid alarm pheromone that signals aphids to stop feeding and disperse. In addition to directly resisting the attacking herbivore, induced volatiles can also influence herbivores on neighboring plants by priming non-infested plants to chemically respond faster to future insect attacks. Furthermore, herbivore-induced volatiles also offer an indirect benefit to the plant by attracting the natural enemies of herbivores, which will be discussed in more detail below.

## Indirect Inducible Defenses

Plants release a wide array of volatile compounds following damage, some general to mechanical damage (typically mixtures of  $C_6$  alcohols, aldehydes, and esters via the oxidation of membrane-derived fatty acids, also known as green leaf volatiles) and others specific to the herbivore species and its instar as well as the intensity and frequency of feeding (e.g., terpenoids, fatty acid derivatives, phenylpropanoids, and benzenoids). These herbivore-induced changes in a plant's volatile chemistry can influence the predators, pathogens, and parasitoids of herbivores via volatile compounds that increase the host location efficiency of the herbivore's natural enemies. It is worth noting that because the fitness benefits of herbivore-induced volatiles have not been clearly demonstrated for many plant systems, their generalized function as indirect "defenses" remains debatable (Dicke and Baldwin 2010). Furthermore, their reliability as sophisticated indicators of herbivory has come into question due to the variation observed in production among individuals both constitutively and following herbivore attack, the complex background chemical landscapes in which they are perceived, and their function in nearly every aspect of plants' biotic and abiotic interactions.

Regardless of the debate over the fitness benefits of herbivore-induced volatiles, the past 40 years has seen an explosion of research describing how the vast array of herbivore-induced plant volatiles effectively recruit insects of the third trophic level that prey upon or parasitize larval herbivores, as well as eggs. By doing so, these volatiles reduce the preference and/or performance of herbivore, thus being considered an indirect defense and an important mediator of tritrophic interactions (Karban and Baldwin 1997). The unique suite of compounds released following herbivore damage is quite sophisticated, differing in total abundance and composition following attack by different herbivores. The species-specific plumes present within the local environment, which are dependent upon the existing abiotic conditions as well, contain critical host location information for parasitoids, which have developed the ability to learn chemical cues associated with the presence and quality of their specific host. For instance, some parasitoids are capable of differentiating between parasitized and unparasitized larval hosts in flight due to the different odor blends induced by each caterpillar. While herbivore-induced volatile blends can be quite complex, a number of individual volatiles involved in attraction of parasitoids have been identified. However, it is highly unlikely that a parasitoid will be exposed to only one volatile compound in nature, and the context within which a volatile blend is perceived may be important. Thus, while individual herbivore-induced volatiles may be involved in parasitoid host location, it is often critical that they are perceived in the context of other volatiles so as to distinguish variation in quality and quantity.

Herbivore-induced volatiles impact evolutionary pressures on herbivores and parasitoids through their role in determining fitness. As previously mentioned, plant volatiles are involved in a range of ecological functions beyond indirect plant defense, including altering the apparency of plants to mutualists, being involved in plant-plant communication, varying the palatability of plant tissue, reducing

microbial colonization, and alleviating abiotic stress such as drought, UV, and heat. As such, their role in plant evolution is dynamic. A number of adaptive explanations have been offered to address the diversity of volatiles found among and within plant families, and it has also been suggested that natural selection exploits the volatility of the compounds themselves and the context in which they are perceived by herbivores and their natural enemies. Similar to foliar compounds, the precise ecological function and evolutionary consequence of every plant volatile is not yet known so their full contribution to plant-insect evolution has yet to be characterized. However, the importance of herbivore-induced volatiles to plant, herbivore, and parasitoid signaling and fitness highlights their potentially important role in the coevolution among taxonomic groups.

---

## Future Directions

The ability of plants to chemically defend themselves against the constant onslaught of herbivores that rely on them for food and energy has fascinated scientists for years. Since, Fraenkel (1959), many studies have sought to describe the variation in plant chemical defense strategies that exists among and within plant families, primarily in the context of coevolutionary theories. Coupling phylogenetic and molecular tools with historical biogeography, studies have shown patterns in plant chemical defense and insect host use, including convergent evolution, and researchers must continue to enhance the molecular and chemical toolbox and design experiments in the context of broader ecological scales to understand larger macroevolutionary patterns. It is also necessary to understand the trade-offs that exist between costs of chemical defense and the benefits obtained from them to appreciate how these strategies are selected upon and evolve. However, given the numerous secondary compounds plants produce and the range of herbivores, pathogens, and abiotic stresses that may select for these each chemical trait over time, determining true defensive functions of mixtures, classes of compounds, or even individual chemicals can be daunting. Furthermore, bioassays aimed at determining the effects of these compounds on potential pests are required and should be coupled with other genetic methods (e.g., transcriptional profiling, mutants, genetic knockouts, etc.) to elucidate not only the effects on herbivore performance but also the molecular mechanisms responsible for them. Along this vein, it is critical to identify the genetic basis for hormonal signaling and interactions between pathways in order to link plant perception of herbivores, signaling cascades, and the production of defensive compounds with the ecological repercussions at the community level.

Plant chemical defenses cannot be considered solely on a pairwise level with a single herbivore but must be framed within a large community perspective considering the multitude of herbivores that plants must defend against and the myriad higher-trophic-level interactions and environmental factors that also influence plant traits (Fig. 1). Thus, future work should focus not only on the defensive properties of secondary compounds in terms of affecting herbivore performance, but also on

the responses of other insects (e.g., pollinators, parasitoids, etc.) to gain a more comprehensive understanding of the cascading effects of plant defenses on community structure. Furthermore, the extent of the specificity of plant chemical defenses should be taken into account, particularly induced defenses, to untangle the primary drivers of community interactions and their role in shaping plant-insect relationships and evolutionary trajectories. In regard to specificity, it is important to expand some of the more conventional targeted chemical analyses (i.e., only focusing on one group of compounds) and to integrate metabolomics into plant-insect research. Current studies may be missing other important secondary compounds that might be contributing to a plant's defense against herbivores and a more mechanistic understanding of defense allocation in plants would be gained by linking primary and secondary metabolic processes through metabolomics. Techniques from metabolomics may be able to detect subtle changes in plant responses over time offering a better idea of the temporal scales over which responses might be most effective against insect pests. While the production of plant secondary compounds can vary significantly over time and space, it is also influenced by a suite of abiotic factors including changes in atmospheric CO<sub>2</sub>, O<sub>3</sub>, temperature, precipitation, nutrient availability, etc., thus having important consequences for plant defense and a number of ecological interactions. To identify general patterns of plant defense strategies under natural conditions, future research must focus on the interactive effects of herbivory and climate on plant secondary production and the consequences for insect population dynamics. Thus, the impact of climate and herbivory on more classes of compounds must be assessed in a wider range of species (i.e., outside the boreal and temperate zone bias) and couched within a whole ecosystem context. Such a multifactor approach is critical to understand the impacts of predicted climate change, insect population dynamics, and their interactions in the future.

---

## References

- Agrawal AA. Current trends in the evolutionary ecology of plant defence. *Funct Ecol.* 2011;25:420–32.
- Agrawal AA, Fishbein M. Plant defense syndromes. *Ecology.* 2006;87:S132–49.
- Barbosa P, Hines J, Kaplan I, et al. Associational resistance and associational susceptibility: having right or wrong neighbors. *Annu Rev Ecol Evol Syst.* 2009;40:1–20.
- Barton KE, Koricheva J. The ontogeny of plant defense and herbivory: characterizing general patterns using meta-analysis. *Am Nat.* 2010;175:481–93.
- Bernays E, Chapman R. The evolution of deterrent responses in plant-feeding insects. In: Chapman RF et al., editors. *Perspectives in chemoreception and behavior.* New York: Springer; 1987. p. 159–73.
- Bernays E, Graham M. On the evolution of host specificity in phytophagous arthropods. *Ecology.* 1988;69:886–92.
- Bryant JP, Chapin III FS, Klein DR. Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos.* 1983;40:357–68.
- Chen M-S. Inducible direct plant defense against insect herbivores: a review. *Insect Sci.* 2008;15:101–14.

- Coley PD, Bryant JP, Chapin III FS. Resource availability and plant antiherbivore defense. *Science*. 1985;230:895–9.
- Després L, David J-P, Gallet C. The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol*. 2007;22:298–307.
- Dethier VG. Evolution of feeding preferences in phytophagous insects. *Evolution*. 1954;8:33–54.
- Dicke M, Baldwin IT. The evolutionary context for herbivore-induced plant volatiles: beyond the “cry for help”. *Trends in Plant Science*. 2010;15:167–75.
- Ehrlich PR, Raven PH. Butterflies and plants: a study in coevolution. *Evolution*. 1964;18:586–608.
- Feeny P. Plant apparency and chemical defense. In: Wallace JW, Mansell RL, editors. *Biochemical interactions between plants and insects*. New York: Springer; 1976. p. 1–40.
- Fraenkel GS. The raison d’être of secondary plant substances. *Science*. 1959;129:1466–70.
- Futuyma DJ, Keese MC. Evolution and coevolution of plants and phytophagous arthropods. In: Rosenthal GA, Berenbaum MR, editors. *Herbivores: their interactions with secondary plant metabolites vol II: ecological and evolutionary processes*. San Diego: Academic Press; 1992. p. 439–475.
- Gershenzon J, Fontana A, Burow M, et al. Mixtures of plant secondary metabolites: metabolic origins and ecological benefits. In: Iason GR, Dicke M, Hartley SE, editors. *The ecology of plant secondary metabolites: from genes to global processes*. New York: Cambridge University Press; 2012. p. 56–77.
- Harborne JB. *Introduction to ecological biochemistry*. 4th ed. San Diego: Academic; 1997.
- Haukioja E. Induction of defenses in trees. *Annu Rev Entomol*. 1991;36:25–42.
- Hermes DA, Mattson WJ. The dilemma of plants: to grow or defend. *Q Rev Biol*. 1992;67:283–335.
- Howe GA, Jander G. Plant immunity to insect herbivores. *Annu Rev Plant Biol*. 2008;59:41–66.
- Janzen DH. Tropical blackwater rivers, animals, and mast fruiting by the Dipterocarpaceae. *Biotropica*. 1974;6:69–103.
- Jones CG, Firn RD. On the evolution of plant secondary chemical diversity. *Philos Trans Biol Sci*. 1991;333:273–80.
- Karban R, Baldwin IT. *Induced responses to herbivory*. Chicago: Chicago University Press; 1997.
- Kessler A, Baldwin IT. Plant responses to insect herbivory: the emerging molecular analysis. *Annu Rev Plant Biol*. 2002;53:299–328.
- Koricheva J, Barton KE. Temporal changes in plant secondary metabolite production: patterns, causes, and consequences. In: Iason GR, Dicke M, Hartley SE, editors. *The ecology of plant secondary metabolites: from genes to global processes*. New York: Cambridge University Press; 2012. p. 34–55.
- Loomis WE. Growth-differentiation balance vs. carbohydrate-nitrogen ratio. *Proc Am Soc Hortic Sci*. 1932;29:240–5.
- McKey D. Adaptive patterns in alkaloid physiology. *Am Nat*. 1974;108:305–20.
- Moore B, DeGabriel JL. Integrating the effects of PSMs on vertebrate herbivores across spatial and temporal scales. In: Iason GR, Dicke M, Hartley SE, editors. *The ecology of plant secondary metabolites: from genes to global processes*. New York: Cambridge University Press; 2012. p. 226–46.
- Nishida R. Sequestration of defensive substances from plants by lepidoptera. *Annu Rev Entomol*. 2002;47:57–92.
- Opitz SEW, Müller C. Plant chemistry and insect sequestration. *Chemoecology*. 2009;19:117–54.
- Rhoades DF. Evolution of plant chemical defense against herbivores. In: Rosenthal GA, Janzen DH, editors. *Herbivores: their interaction with secondary plant metabolites*. New York: Academic; 1979. p. 3–54.
- Schoonhoven LM, van Loon JJA, Dicke M. *Insect-plant biology*. Oxford: Oxford University Press; 2005.
- Stamp N. Out of the quagmire of plant defense hypotheses. *Q Rev Biol*. 2003;78:23–55.
- Wu J, Baldwin IT. New insights into plant responses to the attack from insect herbivores. *Annu Rev Genet*. 2010;44:1–24.

## Further Reading

- Agrawal AA. Natural selection on common milkweed (*Asclepias syriaca*) by a community of specialized insect herbivores. *Evolut Ecol Res.* 2005;7:651–67.
- Agrawal AA, Lau JA, Hambäck PA. Community heterogeneity and the evolution of interactions between plants and insect herbivores. *Q Rev Biol.* 2006;81:349–76.
- Agrawal AA, Conner JK, Rasmann S. Tradeoffs and negative correlations in evolutionary ecology. In: Bell M, Eanes W, Futuyma D, Levinton J, editors. *Evolution after Darwin: the first 150 years.* Sunderland: Sinauer Associates; 2010. p. 243–68.
- Arnason JT, Bernards M. Impact of constitutive plant natural products on herbivores and pathogens. *Can J Zool.* 2010;88:615–27.
- Ayres MP, Clausen TP, MacLean SEJ, et al. Diversity of structure and antiherbivore activity in condensed tannins. *Ecology.* 1997;78:1696–712.
- Bailey JK, Schweitzer JA, Rehill BJ, et al. Rapid shifts in the chemical composition of aspen forests: an introduced herbivore as an agent of natural selection. *Biol Invasions.* 2007;9:715–22.
- Berenbaum M. Toxicity of a furanocoumarin to armyworms: a case of biosynthetic escape from insect herbivores. *Science.* 1978;201:532–4.
- Berenbaum M. Patterns of furanocoumarin distribution and insect herbivory in the Umbelliferae: plant chemistry and community structure. *Ecology.* 1981;62:1254–66.
- Berenbaum M. Coumarins and caterpillars: a case for coevolution. *Evolution.* 1983;37:163–79.
- Berenbaum MC. The expected effect of a combination of agents: the general solution. *J Theor Biol.* 1985;114:413–31.
- Berenbaum MR, Zangerl AR. Furanocoumarin metabolism in *Papilio polyxenes*: biochemistry, genetic variability, and ecological significance. *Oecologia.* 1993;95:370–5.
- Berenbaum MR, Nitao JK, Zangerl AR. Adaptive significance of furanocoumarin diversity in *Pastinaca sativa* (Apiaceae). *J Chem Ecol.* 1991;17:207–15.
- Berenbaum MR, Favret C, Schuler MA. On defining “key innovations” in an adaptive radiation: cytochrome P450s and papilionidae. *Am Nat.* 1996;148:S139–55.
- Bergvall UA, Rautio P, Kesti K, et al. Associational effects of plant defences in relation to within- and between-patch food choice by a mammalian herbivore: neighbour contrast susceptibility and defence. *Oecologia.* 2006;147:253–60.
- Bernasconi ML, Turlings TCJ, Ambrosetti L, et al. Herbivore-induced emissions of maize volatiles repel the corn leaf aphid, shape *Rhopalosiphum maidis*. *Entomol Exp Appl.* 1998;87:133–42.
- Bowers MD. The evolution of unpalatability and the cost of chemical defense in insects. In: Roitberg BD, Isman MG, editors. *Insect chemical ecology: an evolutionary approach.* New York: Chapman and Hall; 1992. p. 216–44.
- Castañeda LE, Figueroa CC, Fuentes-Contreras E, et al. Energetic costs of detoxification systems in herbivores feeding on chemically defended host plants: a correlational study in the grain aphid, *Sitobion avenae*. *J Exp Biol.* 2009;212:1185–90.
- Close DC, McArthur C. Rethinking the role of many plant phenolics—protection from photodamage not herbivores? *Oikos.* 2002;99:166–72.
- Coley PD. Herbivory and defensive characteristics of tree species in a lowland tropical forest. *Ecol Monogr.* 1983;53:209–34.
- De Moraes CM, Lewis WJ, Pare PW, et al. Herbivore-infested plants selectively attract parasitoids. *Lett Nat.* 1998;393:570–3.
- De Moraes CM, Mescher MC, Tumlinson JH. Caterpillar-induced nocturnal plant volatiles repel conspecific females. *Nature.* 2001;410:577–80.
- Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry.* 2009;70:1621–37.
- Dicke M. Behavioural and community ecology of plants that cry for help. *Plant Cell Environ.* 2009;32:654–65.

- Dyer LA, Dodson CD, Stireman JO, et al. Synergistic effects of three Piper amides on generalist and specialist herbivores. *J Chem Ecol.* 2003;29:2499–514.
- Fatouros NE, van Loon JJA, Hordijk KA, et al. Herbivore-induced plant volatiles mediate in-flight host discrimination by parasitoids. *J Chem Ecol.* 2005;31:2033–47.
- Fine PVA, Mesones I, Coley PD. Herbivores promote habitat specialization by trees in Amazonian forests. *Science.* 2004;305:663–5.
- Fine PVA, Miller ZJ, Mesones I, et al. The growth-defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology.* 2006;87:S150–62.
- Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. *Proc Natl Acad Sci.* 2009;106:18054–61.
- Futuyma DJ, Mitter C. Insect-plant interactions: the evolution of component communities. *Philos Trans R Soc Lond B Biol Sci.* 1996;351:1361–6.
- Gerber E, Hinz HL, Blossey B. Interaction of specialist root and shoot herbivores of *Alliaria petiolata* and their impact on plant performance and reproduction. *Ecol Entomol.* 2007;32:357–65.
- Gershenson J, Dudareva N. The function of terpene natural products in the natural world. *Nat Chem Biol.* 2007;3:408–14.
- Gouinguéné SP, Turlings TCJ. The effects of abiotic factors on induced volatile emissions in corn plants. *Plant Physiol.* 2002;129:1296–307.
- Hakes AS, Cronin JT. Environmental heterogeneity and spatiotemporal variability in plant defense traits. *Oikos.* 2011;120:452–62.
- Halitschke R, Stenberg JA, Kessler D, et al. Shared signals – ‘alarm calls’ from plants increase apparency to herbivores and their enemies in nature. *Ecol Lett.* 2008;11:24–34.
- Hopkins RJ, van Dam NM, van Loon JJA. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu Rev Entomol.* 2009;54:57–83.
- Huang T, Jander G, de Vos M. Non-protein amino acids in plant defense against insect herbivores: representative cases and opportunities for further functional analysis. *Phytochemistry.* 2011;72:1531–7.
- Ibrahim MA, Nissinen A, Holopainen JK. Response of *Plutella xylostella* and its parasitoid *Cotesia plutellae* to volatile compounds. *J Chem Ecol.* 2005;31:1969–84.
- Irwin RE, Adler LS. Correlations among traits associated with herbivore resistance and pollination: implications for pollination and nectar robbing in a distylous plant. *Am J Bot.* 2006;93:64–72.
- Janz N, Nylin S. The oscillation hypothesis of host-plant range and speciation. In: Tilmon KJ, editor. *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects.* Berkeley: University of California Press; 2008. p. 203–15.
- Johnson MTJ, Agrawal AA, Maron JL, Salminen J. Heritability, covariation and natural selection on 24 traits of common evening primrose (*Oenothera biennis*) from a field experiment. *J Evol Biol.* 2009;22:1295–307.
- Kaplan I, Halitschke R, Kessler A, et al. Physiological integration of roots and shoots in plant defense strategies links above-and belowground herbivory. *Ecol Lett.* 2008;11:841–51.
- Kessler A, Baldwin IT. Defensive function of herbivore-induced plant volatile emissions in nature. *Science.* 2001;291:2141–4.
- Koornneef A, Pieterse CMJ. Cross talk in defense signaling. *Plant Physiol.* 2008;146:839–44.
- Koricheva J. Interpreting phenotypic variation in plant allelochemistry: problems with the use of concentrations. *Oecologia.* 1999;119:467–73.
- Kostenko O, Bezemer TM. Intraspecific variation in plant size, secondary plant compounds, herbivory and parasitoid assemblages during secondary succession. *Basic Appl Ecol.* 2013;14:337–46.
- Kursar TA, Coley PD. Convergence in defense syndromes of young leaves in tropical rainforests. *Biochem Syst Ecol.* 2003;31:929–49.
- Kursar TA, Dexter KG, Lokvam J, et al. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proc Natl Acad Sci.* 2009;106:18073–8.

- Lerdau M, Gray D. Ecology and evolution of light-dependent and light-independent phytogetic volatile organic carbon. *New Phytol.* 2003;157:199–211.
- Lindroth R. Impacts of elevated atmospheric CO<sub>2</sub> and O<sub>3</sub> on forests: phytochemistry, trophic interactions, and ecosystem dynamics. *J Chem Ecol.* 2010;36:2–21.
- Milchunas DG, Noy-Meir I. Grazing refuges, external avoidance of herbivory and plant diversity. *Oikos.* 2002;99:113–30.
- Pass GJ, Foley WJ. Plant secondary metabolites as mammalian feeding deterrents: separating the effects of the taste of salicin from its post-ingestive consequences in the common brushtail possum (*Trichosurus vulpecula*). *J Comp Physiol B.* 2000;170:185–92.
- Peñuelas J, Llusà J. Plant VOC emissions: making use of the unavoidable. *Trends Ecol Evol.* 2004;19:402–4.
- Pichersky E, Lewinsohn E. Convergent evolution in plant specialized metabolism. *Annu Rev Plant Biol.* 2011;62:549–66.
- Pieterse CMJ, Dicke M. Plant interactions with microbes and insects: from molecular mechanisms to ecology. *Trends Plant Sci.* 2007;12:564–9.
- Rausher MD. Co-evolution and plant resistance to natural enemies. *Nature.* 2001;411:857–64.
- Schuler MA. P450s in plant–insect interactions. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* 2011;1814:36–45.
- Smilanich AM, Vargas J, Dyer LA, Bowers MD. Effects of ingested secondary metabolites on the immune response of a polyphagous caterpillar *Grammia incorrupta*. *J Chem Ecol.* 2011;37(3):239–45.
- Stinchcombe JR, Rausher MD. Diffuse selection on resistance to deer herbivory in the ivyleaf morning glory, *Ipomoea hederacea*. *Am Nat.* 2001;158:376–88.
- Theis N, Lerdau M. The evolution of function in plant secondary metabolites. *Int J Plant Sci.* 2003;164:S93–102.
- Tholl D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr Opin Plant Biol.* 2006;9:297–304.
- Thompson JN. Specific hypotheses on the geographic mosaic of coevolution. *Am Nat.* 1999;153: S1–14.
- Tuomi J, Niemelä P, Chapin III FS, et al. Defensive responses of trees in relation to their carbon/nutrient balance. In: Mattson WJ, Levieux J, Bernard-Dagan C, editors. *Mechanisms of woody plant defenses against insects*. New York: Springer; 1988. p. 57–72.
- Van Dam NM, Tytgat TOG, Kirkegaard JA. Root and shoot glucosinolates: a comparison of their diversity, function and interactions in natural and managed ecosystems. *Phytochem Rev.* 2009;8:171–86.
- Venditti C, Meade A, Pagel M. Multiple routes to mammalian diversity. *Nature.* 2011;479:393–6.
- Wiggins NL, McArthur C, Davies NW, McLean S. Spatial scale of the patchiness of plant poisons: a critical influence on foraging efficiency. *Ecology.* 2006;87:2236–43.
- Wink M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry.* 2003;64:3–19.
- Yuan JS, Himanen SJ, Holopainen JK, et al. Smelling global climate change: mitigation of function for plant volatile organic compounds. *Trends Ecol Evol.* 2009;24:323–31.
- Zagobelny M, Bak S, Rasmussen AV, et al. Cyanogenic glucosides and plant–insect interactions. *Phytochemistry.* 2004;65:293–306.
- Zangerl AR, Rutledge CE. The probability of attack and patterns of constitutive and induced defense: a test of optimal defense theory. *Am Nat.* 1996;147:599–608.
- Zarate SI, Kempema LA, Walling LL. Silverleaf whitefly induces salicylic acid defenses and suppresses effectual jasmonic acid defenses. *Plant Physiol.* 2007;143:866–75.



David A. Lipson and Scott T. Kelley

## Contents

Introduction .....	178
Mutualisms .....	179
N-Fixing Mutualisms .....	180
Mycorrhizae .....	185
Leaf Endophytes .....	188
Host Controls Over Mutualisms .....	189
Plant Growth-Promoting Rhizobacteria .....	190
Plant-Microbe Signaling in the Rhizosphere .....	191
Nutrient Relations in the Rhizosphere .....	191
The Rhizosphere Effect .....	191
Plant-Microbe Competition for N .....	192
Impact of Plants on Soil Microbial Processes .....	193
PMI Effects on the Soil Matrix .....	195
Impacts of Plants on Microbial Diversity .....	195
Culture-Independent Characterization of Microbial Diversity .....	195
Microbial Diversity of the Phyllosphere .....	196
Microbial Diversity of the Rhizosphere .....	198
Impacts of Microbes on Plant Diversity .....	199
The Role of Plant-Microbe Interactions in Global Change .....	200
Future Directions .....	202
References .....	203

---

## Abstract

- Globally, the majority of nitrogen and phosphorus uptake by plants is mediated by mutualistic root microbes, which form intricate and complex biochemical and genetic interactions with plants.
- Plant leaves host a variety of beneficial bacteria and fungi that contribute to plant nutrition and/or defense against pathogens.

---

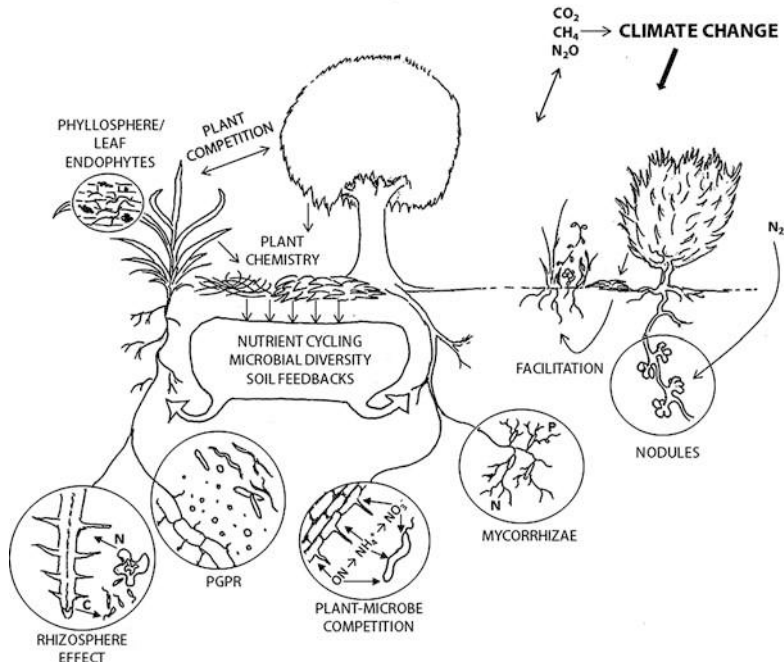
D.A. Lipson (✉) • S.T. Kelley  
Department of Biology, San Diego State University, San Diego, CA, USA  
e-mail: [dlipson@mail.sdsu.edu](mailto:dlipson@mail.sdsu.edu); [skelley@mail.sdsu.edu](mailto:skelley@mail.sdsu.edu)

- In addition to mutualistic bacteria intimately associated with roots, there exist plant growth-promoting rhizobacteria more loosely associated with roots that contribute to plant nutrition, protection from pathogens, and environmental stress reduction.
- The region surrounding roots, the rhizosphere, is a dynamic environment, rich in chemical communication among plants and microbes, where nutrient cycling is altered by root exudation and heightened microbial activity.
- Plants profoundly impact the biogeochemical cycling activities of soil microbes through their effects on microclimate and soil chemistry.
- Plants and microbes collaborate to produce soil organic matter such as humic substances, which determine important soil properties such as water and nutrient holding capacity and the stability of soil carbon.
- The species composition of plant-associated microbial communities is extremely diverse and variable, but is strongly influenced by plant species.
- Soil microbial communities can mediate changes in plant diversity during invasions or succession through positive and negative soil feedbacks.
- Plant-microbe interactions are involved in several feedback mechanisms in which the biosphere reacts to and influences climate change.
- There are currently gaps in the understanding of plant-microbe interactions, particularly in terms of genetics of certain plant-microbe mutualisms, the diversity of plant-associated microbial communities, and the role of plant-microbe interactions in producing feedbacks to climate change; however, new technologies are emerging that should help fill existing gaps.

---

## Introduction

Plant-microbe interactions (PMI) are central to the functioning of terrestrial ecosystems. No model of plant biology is complete without taking into account their associated microbes, just as soil microbes cannot be understood without considering the plants that shape their habitat. In fact, given the origin of chloroplasts and mitochondria from endosymbiotic bacteria, it could be argued that PMI are inherent to the very biology of plant cells. PMI form a continuum, ranging from highly coevolved, species-specific mutualisms tightly associated with plant tissues, to the more variable and general communities of microbes in the soil, which produce strong feedbacks that drive plant growth and, in turn, are largely controlled by plant chemistry and microclimate. Plant-microbe mutualisms show an extraordinarily intricate signaling/gene expression network between host and symbiont, but even some of the more general PMI are mediated by surprisingly complex and intimate interactions. PMI can shape both the plant and microbial communities and provide strong feedbacks in important global processes, such as biological invasions and climate change. Important gaps remain in the current understanding of PMI, but methodologies and research are advancing rapidly that may address some of these gaps.



**Fig. 1** A summary of the PMI and their roles in the environment discussed in this chapter

This chapter first describes the nature of the PMI at the individual plant-microbe level, working from specific to more general associations. It then considers larger-scale implications of these interactions for plant and microbial communities, ecosystems, and global change. The chapter concludes with an assessment of the current gaps in knowledge and how newly developed tools could help fill those gaps. Highlights of the topics discussed here are depicted in Fig. 1. We do not deal explicitly with plant pathogens. However, mutualistic associations such as mycorrhizae can span the mutualism-parasitism continuum, and so parasitism is considered briefly within this context. There is also some mention of pathogenic microbes in the discussion of microbial impacts on plant diversity. Similarly, while this chapter deals mainly with positive associations, plant-microbe competition is also considered, as it is an inherent factor in the functioning of the rhizosphere.

## Mutualisms

Mutualisms are differentiated from other positive associations in that they are generally essential (at least in practical terms) for the survival of one or both partners, species-specific and show an especially high degree of coevolution between the partners. The most widely studied plant-microbe mutualisms are those between leguminous plants and nitrogen (N)-fixing bacteria (collectively

referred to as rhizobia) and mycorrhizal associations between roots and fungi. However, many other important plant-microbe mutualisms exist, examples of which are included in this section.

## N-Fixing Mutualisms

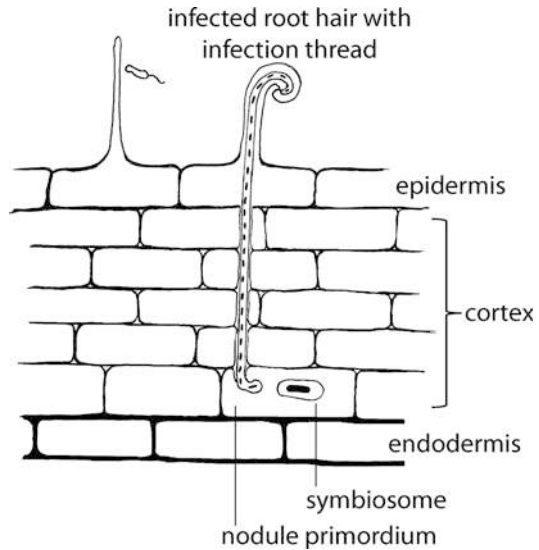
Dinitrogen gas ( $N_2$ ) makes up 80 % of the earth's atmosphere, yet N is the most commonly limiting nutrient for plant growth in most terrestrial ecosystems (with the exception of the tropics, where phosphorus (P) is more commonly limiting). One of the main reasons for this apparent paradox is that because of the highly Stable  $N\equiv N$  triple bond, enzymatic fixation of  $N_2$  into a biologically useable form is an energetically expensive process, requiring about 16 mol ATP per mole N fixed. While only prokaryotes (Bacteria and Archaea) are known to carry out this process, a great variety of plants maintain mutualisms with N-fixing bacteria.

### Rhizobia-Legume Mutualism

Most globally important is the legume-rhizobia mutualism. Leguminous plants (in the “bean” family, *Fabaceae*, formerly *Leguminosae*) form mutualisms with nodule-forming, N-fixing bacteria, known collectively as rhizobia. These bacteria are generally *Alphaproteobacteria* from closely related genera such as *Rhizobium*, *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, and *Sinorhizobium*. More recently, members of the *Betaproteobacteria* (such as *Burkholderia*) have been found to form similar associations with *Mimosa*, these sometimes referred to as beta-rhizobia. There are also reports of *Gammaproteobacteria* capable of producing nodules and fixing N in legumes (Shiraishi et al. 2010). Rhizobia typically form nodules on roots, though may also form these structures on stems, as is the case for the mutualism between the tropical tree, *Sesbania rostrata*, and its partner, *Azorhizobium* spp. The *nod* genes required for nodulation and the *nif* genes required for N fixation are often found on a *Sym* plasmid or other mobile genetic element. In particular, the *nod* genes appear to have been horizontally transferred among the various nodulating bacteria of the  $\alpha$ -,  $\beta$ -, and even  $\gamma$ -*Proteobacteria* (Masson-Boivin et al. 2009; Shiraishi et al. 2010).

The infection and nodulation process involves a complex interplay between bacteria and host. The signaling and genetics have been reviewed in great detail, especially for the *Sinorhizobium-Medicago* and *Mesorhizobium-Lotus* systems (Oldroyd et al. 2011). The general picture is described here, though these relationships are impressively diverse and likely include many exceptions (Masson-Boivin et al. 2009). Roots produce flavonoids that attract Rhizobia from the surrounding soil (where they can survive independently of plants, but will not generally fix N). Rhizobia bind specifically to lectins (proteins that bind carbohydrates) on the surface of root hairs via sugar residues on the bacterial surface. Bacterial cell surface polysaccharides also appear to play key roles in avoiding the host's defense response. The bacteria invade the root hair, producing polysaccharide-degrading enzymes such as polygalacturonase or cellulose to soften the root hair cell wall.

**Fig. 2** Root hair infection of legumes by rhizobia



Bacteria produce nodulation factors, typically lipochitooligosaccharides (several N-acetyl glucosamine units with an acyl chain on the end), causing the root hair to curl, effectively trapping the bacteria (Fig. 2). An infection thread is formed by an invagination of the plant cell wall and plasma membrane, together with polysaccharide production by the bacteria. The bacterial colony in the infection thread is separated from the plant cytoplasm by the plant membrane. As the bacteria divide, the infection thread travels down the root hair, through the epidermis, and into the cortex, passing through cortical cells along the way via cytoplasmic bridges, until finally the bacteria are released from the thread and transported into the plant cortical cell that will give rise to the nodule (“nodule primordium”). During infection thread growth, this cell has already altered its gene expression in various ways (e.g., becoming polyploid), in response to bacterial nodulation factors that affect plant hormones, cytokinin, and auxin. Further plant and bacterial growth lead to nodule formation.

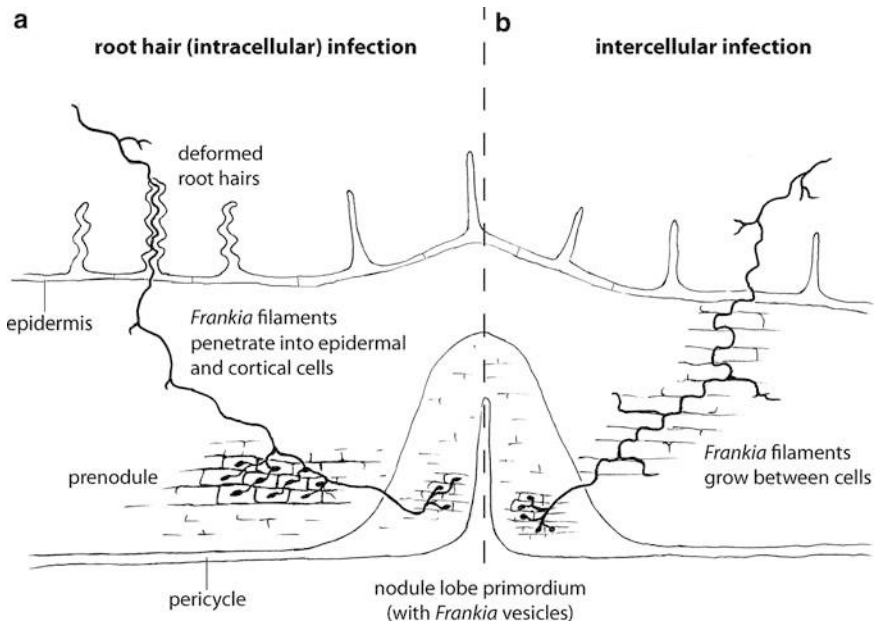
Alternatively, some rhizobia, particularly those adapted to aquatic/semiaquatic tropical hosts, infect via crack entry. *Azorhizobium caulinodans* is the prime example of this strategy. They form pockets between epidermal cells, loosened by the emergence of lateral roots. In some cases, infection threads may then form. Eventually, bacteria are taken into root cortical cells and form nodules. As this type of infection is independent of the nodulation factors mentioned above for root hair entry (and which are absent in some *Bradyrhizobia* that infect via cracks), it is proposed that cytokinins produced by the bacteria induce development of the nodule primordium. In either mode of entry, bacteria are eventually transported into the cell interior of the nodule primordium, surrounded by a peribacteroid membrane. As the nodule develops, the bacteria differentiate into bacteroids. Their gene expression is altered, shutting down many “housekeeping genes” required for plant-independent growth while upregulating N fixation.

Even after nodules are established, there is further complex interplay between the bacteroid and host. A bacteroid together with its surrounding membrane is called a symbiosome. Symbiosomes resemble organelles, such as mitochondria or chloroplasts, in that they are highly dependent on their hosts. Amino acid synthesis is shut down in bacteroids of some rhizobia species, who instead rely on their hosts for amino acids (Oldroyd et al. 2011). Similarly, rhizobia require homocitrate from the host to produce the iron-molybdenum (FeMo) cofactor of nitrogenase, the primary enzyme of N fixation. The plant host is also primarily responsible for regulating O<sub>2</sub>, which is inhibitory to nitrogenase. Plants control the overall O<sub>2</sub> permeability of the nodule through a barrier in the nodule cortex and produce the O<sub>2</sub>-binding protein, leghemoglobin, reducing free O<sub>2</sub> concentrations to the nanomolar range while facilitating O<sub>2</sub> diffusion to the rapidly respiring bacteroids. The plant supplies energy to the bacteroids in the form of organic acids such as malate, succinate, and fumarate. The primary product of N fixation, NH<sub>4</sub><sup>+</sup>, is transported across the peribacteroid membrane, assimilated into glutamate and glutamine in plant cytosol and exported from the nodule as other N-rich amino acids such as arginine. An extraordinary amount of detailed information is known about the few model legume-rhizobia systems mentioned above. However, nature is full of surprising variation on these themes. Striking examples include the discovery of methylophony in the legume-nodulating  $\alpha$ -Proteobacterium, *Methylobacterium nodulans*, and photosynthesis in a group of *Bradyrhizobium* spp. that nodulate legumes of the *Aeschynomene* genus. In both cases, the unexpected energy-generating metabolism by the bacterial endosymbionts (recycling of plant-produced methanol and photosynthesis in the nodule, respectively) appear to contribute to the efficiency of the mutualisms (Masson-Boivin et al. 2009).

Another variation is the association of rhizobia with *Parasponia*, a non-leguminous genus of tropical tree. These associations appear to be less efficient and sophisticated than those in legumes (Santi et al. 2013). In terms of host plant phylogeny and nodule morphology, these mutualisms are similar to actinorhizal associations, which form with a very different group of bacteria.

### Actinorhizal Associations

Actinorhizal N-fixing associations have not been as thoroughly studied as those of legumes and rhizobia, but they are globally important, contributing possibly about 25 % of terrestrial nitrogen fixation worldwide (Pawlowski and Newton 2008). Actinorhizal associations form between various plants (generally woody trees or shrubs) in the Fagales, Cucurbitales, and Rosales and the filamentous Actinobacterium, *Frankia*. Actinorhizal nodules have a coral-like morphology, with multiple lobes. The actinorhizal plants appear to form a single N-fixing clade of angiosperms with the legumes. Actinorhizal nodules are morphologically similar to those of legumes, and deeper parallels may also exist. For example, the receptor-like kinase SymRK is required for nodulation in both legumes in the actinorhizal tree, *Casuarina glauca* (Masson-Boivin et al. 2009). Actinorhizal plants fill similar niches as legumes, for example, in colonizing N-poor soils in early succession. In cooler regions, actinorhizal plants are often the dominant N



**Fig. 3** Two pathways of *Frankia* infection in actinorhizal mutualisms

fixers among trees and shrubs, whereas leguminous trees generally fill these niches in the tropics. However, the *Myrica faya* invasion of Hawaii (discussed below) is clearly an important exception to this. The specificity between host and bacterial endosymbiont varies. Some species of *Myrica*, such as *M. pensylvanica* and *M. californica*, can host a broad diversity of *Frankia*, whereas as *M. gale* has greater specificity. *Myrica*, *Alnus*, *Dryas*, and *Elaeagnus* species often have the ability to form nodules outside their native ranges.

There are a number of parallels between actinorhizal and legume-rhizobia mutualisms. For example, as in legumes, there are two infection strategies in actinorhizal symbioses: root hair infection and intercellular colonization (Fig. 3). The mechanisms of plant-microbe communication in the actinorhizal nodulation process are known in far less detail than for the legume-rhizobia system, but there are likely several parallels. *Frankia* produces root hair deformation factor (*Had*), possibly similar to rhizobial nodulation factors, except it does not cause cell division in the root cortex. Flavonoids may be involved, both in stimulatory and inhibitory roles, but these are not clearly characterized. There is also some evidence that lectins might be involved in bacterial binding at the plant surface.

In contrast to the root hair infection process in legumes (in which rhizobia stay in an extracellular infection thread until reaching the root cortex), one pathway of infection by *Frankia* is to enter the cytoplasm of a deformed root hair cell (Fig. 3a). An infection thread-like structure is formed by growth of host plasma membrane and cell wall material around the invading *Frankia* filament. In response to root hair

infection by *Frankia*, root cortical cells below the infected root hair start to divide, expand, and eventually become infected by *Frankia*, forming the prenodule. As the prenodule matures, *Frankia* differentiate into vesicles (analogous to bacteroids in rhizobia) and *nif* genes are expressed. Meanwhile, in the root pericycle, a nodule primordium is initiated, which expands, becomes infected intracellularly by *Frankia* from the prenodule, and develops into a mature nodule lobe. Alternatively, *Frankia* can invade by growing between epidermal and cortical cells, in an expanded intercellular zone created by the thickening of host cell walls (Fig. 3b). *Frankia* filaments then penetrate the cytoplasm of cortical cells of the nodule-lobe primordium, which develops into a nodule lobe. In both cases, actinorhizal nodules are essentially modified lateral roots, unlike the case in legumes.

In actinorhizal mutualisms, the host plant generally plays a smaller role in nodule O<sub>2</sub> regulation than in legumes. *Frankia* can fix N under aerobic conditions, in part due to the thick walls and rapid respiration rates of its vesicles. Actinorhizal nodules generally lack the dense cell layers that restrict O<sub>2</sub> in legume nodules, but some host plants form nodules with thick, lignified cell walls and high levels of hemoglobin (of either plant or bacterial origin), and in which the *Frankia* do not form vesicles. “Nodule roots” represent an interesting variation in nodule O<sub>2</sub> relations: these are produced by some actinorhizal plants (such as *Myrica* spp.) in wet soils, growing upwards above the water table to conduct O<sub>2</sub> to submerged nodules through porous (aerenchymous) tissues.

In contrast to rhizobia bacteroids in legume nodules, *Frankia* assimilates the NH<sub>4</sub><sup>+</sup> produced in N fixation and instead exports N to host cells in the form of amino acids such as arginine (Berry et al. 2011). Actinorhizal mutualisms are not nearly as genetically well characterized as in legumes. While this knowledge base is growing and a number of symbiotic genes have been identified, the precise roles for these genes in symbiosis are still being elucidated.

### Plant-Cyanobacterial Mutualisms

Heterocystous cyanobacteria are filamentous photosynthetic bacteria with specialized cells (heterocysts) where N fixation takes place. Heterocystous cyanobacteria form N-fixing mutualisms with bryophytes, the water-fern *Azolla* (Pteridophyta), cycads (Gymnosperms), and the flowering plant *Gunnera* (an angiosperm). Cycads are among the more ancient lineages of extant vascular plants and arose much earlier than the nodule-forming angiosperms. Their N-fixing mutualisms with cyanobacteria appear to be far less sophisticated than those in nodules. Cycads, when infected by *Nostoc* spp., produce coralloid roots in which the cyanobacteria are housed in a mucilaginous extracellular space. In terms of this mutualism, more is known about gene expression in the cyanobacterial partner, which has slower cell division, increased cell volume, altered intracellular structures, and increased frequency of heterocysts compared to the free-living state. However, the cyanobacteria are far more independent of their hosts and less physiologically altered than in the previously mentioned mutualisms. The exact C source provided by the plant is currently not certain but may be simple sugars. Fixed N is assimilated into amino acids (glutamine or citrulline) within the heterocysts and transferred to



the plant. The cyanobacteria are solely responsible for protecting their nitrogenase enzymes from O<sub>2</sub>. This is done by concentrating all N-fixing activity into heterocysts with thick walls and rapid respiration rates (Santi et al. 2013).

The heterocystous cyanobacterium, *Nostoc azollae* (formerly *Anabaena azollae*), grows in cavities on the underside of leaves in the aquatic fern, *Azolla*. This may represent the simplest of all plant-bacterial N-fixing associations, yet some degree of coevolution has occurred. For example, the cyanobiont of *Azolla* is transmitted from generation to generation via megasporocarps (structures *Azolla* uses for dispersal of its spores), rather than relying on a fresh supply of cyanobacteria from the environment for each new generation of plant (Santi et al. 2013).

## Mycorrhizae

Mycorrhizal associations form between a wide variety of plant roots and fungi (Smith and Read 2008). The majority of plant species have some form of mycorrhizae, notable exceptions being the *Brassicaceae* and *Chenopodaceae* families. These two non-mycorrhizal plant families are generally ruderal (weedy) and so grow best in high-nutrient conditions where mycorrhizae would be of less benefit (see section “[Host Controls over Mutualisms](#)”). Three broad classes of mycorrhizae are differentiated by the arrangement of fungal hyphae in or around plant cells: endomycorrhizae penetrate into the plant cytoplasm, ectomycorrhizae (EM) form a dense “mantle” around stunted lateral roots and grow between root cells without penetrating the cell membrane, and ectendomycorrhizae both penetrate into the interior of the host cells while also forming a mantle. Despite each category being quite diverse, these morphologies are fairly well correlated with their ecological roles. These associations mainly provide benefit by effectively extending the root surface for nutrient uptake; but they also may offer the host plant some protection from pathogens or other stresses such as heavy metal toxicity. The most widespread and well-studied type is the arbuscular mycorrhizae (AM), a form of endomycorrhizae.

### Arbuscular Mycorrhizae

AM are the most common mycorrhizae, forming endomycorrhizal associations with about two thirds of all plant species and about 80 % of angiosperms. However, AM are also found among gymnosperms, bryophytes, and ferns. They are particularly common among herbaceous species, and so from an ecosystem perspective, AM is the dominant type of mycorrhizal relationship in grasslands. The fungal partners are now placed in the *Glomeromycota* phylum. These fungi are reliant on their hosts and are therefore considered obligate biotrophs. As such they do not live independently as saprotrophs (decomposers of dead organic matter) in soil, but exist in a dormant form until they encounter a compatible host root. As a result, no AM mycobiont exists as a pure culture, though co-cultures with plant root tissue have been maintained.

Because these fungi are not effective saprotrophs, they are less able to access N that is covalently bound to complex soil organic matter. P is bound to organic matter through ester bonds that require a narrower class of enzymes to cleave, and so the primary nutritional role of AM is to acquire phosphorus (P) for their hosts. AM are named for the highly branched, treelike structures (arbuscules) they form within plant cortical cells. These structures are the primary site of nutrient exchange between the fungus and plant (the highly branched geometry provides high surface area for exchange). Vesicles are also found within the roots in some AM, and in older literature, the term vesicular-arbuscular mycorrhizae (VAM) is frequently found. These structures appear to play a storage/dormancy role in the fungus and are capable of infecting new roots. To survive in the soil between hosts, AM fungi produce spores, such as the large “gigaspores” of *Gigaspora* spp., which can reach about 0.5 mm in diameter.

The AM infection process has been worked out in detail for *Medicago truncatula* (Bonfante and Genre 2010). There are a multitude of signals between plant and fungus, in which each senses and responds to the other. The fungus senses the presence of roots through root exudates and CO<sub>2</sub> from root respiration. These signals stimulate spore germination. In fact, spores can be germinated in the lab under elevated CO<sub>2</sub> but will abort without the presence of a compatible host. Fungal hyphae are stimulated to become highly branched when in close proximity to a host root. The exact signal for this response is unknown at this time, but the response is produced most strongly in P-starved plants. The plant senses the approaching fungus even before physical contact with the roots, through an unresolved soluble signaling molecule. The fungus must avoid triggering the defense response of the host plant. This may be done by altering chitin in the fungal cell walls, degrading the plant-produced defense signals, or producing defense-suppressing compounds. The plant undergoes systemic changes (found in the entire plant rather than just the local infected area) in response to AM infection. These include expression of P-starvation genes and lateral root formation, both serving to increase the efficiency of infection. Additionally, infection induces cell-specific gene expression in roots, such as cellulase, chitinase, and P uptake. The cellulase enzymes presumably act to soften the plant cell wall to allow intracellular penetration, whereas the chitinase could be part of the plant’s general defense response.

### **Ecto-, Ectendo-, and Arbutoid Mycorrhizae**

These three mycorrhizal types share morphological features and are sometimes grouped together. EM relationships are found on the majority of tree species, and so EM are the dominant mycorrhizal type in forested ecosystems. However, they are found on a variety of non-woody plants, such as the alpine sedge, *Kobresia myosuroides*. The fungal partners are Basidiomycetes and Ascomycetes. In contrast to the AM fungi, EM can live freely in the soils as saprotrophs, degrading complex organic matter. Because of this, they can access a broader range of soil nutrients than AM and so transfer N, P, and other nutrients to the host plant. Most trees may be considered obligately ectomycorrhizal in the sense that they would not be likely to compete for nutrients in natural conditions without their EM. EM form dense

mats of hyphae (mantles) around stunted lateral roots, called club roots, due to their club-like appearance. The fungal hyphae penetrate between plant cortical cells, forming a network referred to as the Hartig net. EM greatly extend the length and surface area of the rooting system. Because fungal hyphae have a smaller diameter than plant cells, it is much cheaper for a plant to allocate C to its EM fungus than to produce the equivalent amount of root length or surface area. EM relationships range greatly in the specificity between host and mycobiont. A single host plant may be infected simultaneously by a high diversity of EM fungi, while on the other extreme, some plant species have very specific requirements for infection and their range may be restricted by the presence of compatible EM fungi in the soil.

Ectendomycorrhizae form a mantle and Hartig net like those of EM but also penetrate into plant epidermal and cortical cells. These are formed by Ascomycetes on species of *Pinus* and *Larix*. Interestingly, the same fungal species can form ecto-, ectendo-, or ericoid mycorrhizae, depending on the host plant, illustrating how the plant controls the morphology of these structures. Arbutoid mycorrhizae, formed in *Arbutus* species of the *Ericales*, also form a mantle, Hartig net, and intracellular structures, but are distinguished from ectendomycorrhizae in that they only infect epidermal cells.

### **Ericoid Mycorrhizae**

Ericoid associations are found among the *Ericaceae* plant family, including many wetland species, and so these are the predominant mycorrhizal type in wetlands. The fungal partners are Ascomycetes and Deuteromycetes. As complex organic matter accumulates in wetland soils, ericoid fungi appear to be adapted to access N from highly complex organic molecules and so can allow their hosts access to forms of organic N not generally available to other plants. Ericoid mycorrhizae form intracellular coils, which function analogously to arbuscules in plant-fungus nutrient exchange.

### **Orchid and Monotropoid Mycorrhizae**

These two mycorrhizal types are grouped together here because both include non-photosynthetic plants that use fungi to access organic carbon from other plants or decaying organic matter. Plants from the Orchidaceae (orchids) form obligate mycorrhizal relationships with Basidiomycete fungi (and a few Ascomycetes). Orchids produce very small seeds without major storage reserves. As a result, they rely on mycorrhizae for seed germination and early establishment of seedlings. Some orchids are non-photosynthetic and mycoheterotrophic, meaning they rely on these associations throughout their life, while others are mixotrophic, gaining C both from photosynthesis and mycorrhizal fungi. This relationship is unique among mycorrhizae in that the plant is reliant on organic C from the fungus, whereas typically the plant provides C to the mycobiont in exchange for nutrients. These relationships gain C by parasitizing ectomycorrhizal networks of other plant species or by the saprotrophic activity of the fungal partner. One noteworthy example of the latter form of relationship is that between some mycoheterotrophic orchids (such as *Galeola* and *Gastrodia*) and *Armillaria mellea*, a wood-degrading root pathogen.

The orchid produces an antifungal protein, gastrodianin, which may play a role in preventing root degradation by the fungus (Baumgartner et al. 2011). *Armillaria* is bioluminescent, generating light using the enzyme, luciferase. It is possible that the bioluminescence attracts nocturnal animals for fungal spore dispersal.

Many orchids form associations with the so-called Rhizoctonia complex, actually comprising three groups within the Agaricomycetes: the Sebaciniales, Ceratobasidiaceae, and Tulasnellaceae. Rhizoctonia-type associations are endomycorrhizal: fungi form coils (pelotons) between the cell wall and membrane of root cortical cells. These pelotons are eventually digested by the plant. It is still uncertain to what extent the fungi in these associations benefit from the relationships, and the orchids have often been viewed as parasitizing the fungi. However, this view may be changing as evidence for the mutualism of these relationships emerges (Dearnaley et al. 2012).

Monotropoid mycorrhizae are also formed by non-photosynthetic, parasitic plants. Like some of the orchid mycorrhizae, Monotropoideae species (Ericales) tap into EM networks to access sugars and nutrients. These appear to be exploitative mycorrhizae rather than mutualisms, in that there is no evidence that the fungi benefit. As in classic ectendomycorrhizal structures, a mantle, Hartig net, and intracellular hyphae are formed. The exchange of nutrients presumably occurs in the “fungal pegs” that penetrate into epidermal root cells.

### Root Endophytic Fungi

In addition to the mycorrhizae described above, there are a variety of plant-fungal interactions described generically as “root endophytes.” These interactions range from mutualistic to parasitic. A frequently observed morphology of root endophyte is the “dark septate fungi,” named for their dark pigmentation and the presence of cross walls between hyphal cells (these are absent in AM). *Phialocephala fortinii* is a common example of this type of fungus; many are Ascomycetes belonging to the order Helotiales. These represent a variety of fungal species found in a variety of plants, and both positive and negative growth effects on the host have been reported. A meta-analysis concluded that dark septate endophytes tend to have a net positive effect on plant growth and nutrient uptake (Newsham 2011). In contrast, another meta-analysis that included studies on all varieties of root fungal endophytes (not just dark septate) found a net negative or neutral effect on plants (Mayerhofer et al. 2013). Clearly the relative benefit of these relationships depends on the species of host and fungus, but it appears that the dark septate variety is generally more beneficial to plants than other root endophytic fungi.

### Leaf Endophytes

Mutualisms between microbes and plant roots have received the most attention, but there are a number of important and fascinating mutualistic associations between microbes and leaves. For example, endophytic associations between grass and fungi can protect the host from herbivory, disease, and drought stress and can stimulate

root growth (Saikkonen et al. 2013). *Epichlöe* species (Ascomycetes) are common leaf endophytes in grasses. Like mycorrhizal relationships, leaf endophytic relationships can range from mutualistic to parasitic. In the more mutualistic instances, the fungus is transmitted to new generations of host plants through seeds. In these symbioses, the fungus does not penetrate cell walls, often colonizing vascular tissues. The growth of fungal hyphae and plant tissues are well coordinated, with hyphal growth ceasing once leaf elongation is complete. The fungi protect plants from insect herbivores through production of alkaloids, such as peramine and loline. Some strains produce indoleterpenes and ergot alkaloids, which are also effective against vertebrate herbivores (the latter, including lysergic acid amine, is often responsible for poisoning livestock). Endophytic fungi can provide protection against root-feeding nematodes, despite their absence in roots. This might be caused by translocation of toxins synthesized by the fungus, induction of plant defenses, or morphological changes in the roots of the host plant. Stimulation of root growth by endophytes may also be responsible for increased stress resistance in the host.

Leaf surfaces also support thriving communities of bacteria, including numerous beneficial species (see “[Microbial Diversity of the Phyllosphere](#)” section below). Some, such as *Sphingobacterium* spp., provide protection from leaf pathogens (Vorholt 2012). In the nonvascular realm, Sphagnum spp. living in methanogenic wetland ecosystems host CH<sub>4</sub>-oxidizing bacteria that convert CH<sub>4</sub> to CO<sub>2</sub>, which the host plant uses for photosynthesis (Raghoebarsing et al. 2005).

## Host Controls Over Mutualisms

All of the mutualisms described above have the potential to become parasitic under certain conditions when the cost to the plant of sustaining the partner outweighs the benefit. This can occur under high-nutrient conditions when plants roots can easily absorb growth-limiting nutrients without the aid of root symbionts, or under low light conditions when allocation to leaves is a better investment for the plant than allocation to roots and root mutualisms. Also, there is considerable variation in the effectiveness of potential microbial mutualists in the environment, and so the pool of bacterial or fungal strains that can infect plant roots may fall along a mutualism-parasitism continuum. The plant therefore has to have mechanisms for controlling the growth of microbial symbionts on its roots depending on growth conditions and the effectiveness of the infecting microbial strain. In the most general sense, plants have the ability to allocate resources to either above- or belowground structures to maximize growth, and this will determine how much C is provided to root mutualists. More specific control mechanisms also exist. Nodulating plants have the ability to limit the infection process in response to nutrient and light availability (Pawlowski and Newton 2008). Legumes can also exert control on the symbiosis in mature nodules by regulating O<sub>2</sub> supply. In this way, the host plant can impose “sanctions” on ineffective rhizobia strains that “cheat” by taking up resources while not fixing N, by restricting O<sub>2</sub> supply across the diffusive barrier in the nodule or

even across the peribacteroid membrane. Similarly, plants regulate AM infection in response to the plant's P status through P-starvation genes as described above and can abort infection in cases where transport of P to the plant is ineffective. In *Medicago truncatula*, silencing the function of the fungal phosphate transporter, MtPT4, leads to premature death of the arbuscules, indicating that the plant uses the influx of phosphate from the arbuscule to signal the presence of an effective mutualist fungal strain. However, when these plant mutants were grown under N-limiting conditions, they allowed arbuscules to form normally, showing that the host plant also relies on N transfer from its fungal partner (Javot et al. 2011).

---

## Plant Growth-Promoting Rhizobacteria

The mutualisms described above generally involve species-specific interactions with a high degree of coevolution between plant host and microbe. However, plants also benefit from mutually positive interactions with microbes of a less species-specific nature. The microbial community surrounding plant roots (the rhizosphere) contains a variety of bacteria that contribute to plant growth and survival. These are commonly known as plant growth-promoting rhizobacteria (PGPR). PGPR benefit plants by improving nutrient availability, stimulation of root growth, bioremediation of contaminants, reduction of plant stress, and protection from pathogens (Santi et al. 2013).

PGPR include associative N fixers, such as *Azospirillum spp.*, that are fueled by energy from root exudates and produce N that contributes to plant growth. Other frequently encountered associative N fixers include *Acetobacter diazotrophicus*, *Herbaspirillum seropedicae*, *Azoarcus spp.*, and *Azotobacter spp.* While these so-called associative N-fixing relationships are generally distinguished from the mutualistic N-fixing ones described earlier, the boundary between mutualistic and associative N fixation is somewhat arbitrary, as these bacteria form intimate relationships with roots and are similar to mutualistic N fixers in that they colonize surfaces or the interior of roots, synthesize plant hormones (auxins, cytokinins, and gibberellins, but mostly the auxin, IAA), and alter their own gene expression in the colonization process. Like rhizobia, *Azospirillum* species have large plasmids that contain genes for interacting with plants (e.g., chemotaxis and motility genes that allow them to sense and move towards roots). *Azospirillum* colonizes root surfaces, but some other associative N fixers infect root cells (*A. diazotrophicus* infects through cracks at lateral root junctions and enters the host's xylem).

PGPR may also solubilize P from the mineral form, apatite (calcium phosphate), and produce siderophores that solubilize and transport Fe or other metals. PGPR can protect plants from a variety of stresses. Production of extracellular polymeric substances (EPS) can trap water in the rhizosphere and reduce water/desiccation stress. Production of the enzyme, 1-aminocyclopropane-1-carboxylate deaminase (ACCd), lowers the concentration of ethylene that is overproduced by plants in response to stressful conditions. ACCd-producing bacteria can help plants recover from stress due to salinity, drought, and heavy metals and may help promote

nodulation in legumes. Some PGPR produce plant hormones that can have a positive impact on plant growth. Finally, the presence of benign bacteria on the root surface can have a “probiotic” effect, protecting the root from opportunistic pathogens by keeping this niche occupied (Santi et al. 2013).

---

## Plant-Microbe Signaling in the Rhizosphere

There is a complex chemical conversation among plant roots and members of the microbial community in the rhizosphere (Badri et al. 2009). Microbes produce plant hormones, and conversely, plants produce signals that alter microbial gene expression and growth. An important adaptation for rhizosphere bacteria is the ability to colonize roots through biofilm formation. Biofilms are surface-associated microbial colonies that include cells in a matrix of EPS. Once formed, biofilms can be resistant to environmental stresses and predation. Biofilm formation involves quorum sensing, a signaling mechanism among bacteria in which individuals of a population produce a signaling molecule that increases in concentration as the population grows until a threshold is reached, signaling a sufficient population size to initiate gene expression for cooperative activities that require some minimum population to be effective. A strategy among competing rhizosphere bacteria is to disrupt biofilm formation in the competing population by degrading quorum-sensing molecules. In fact an enzyme from a *Bacillus* species has been cloned into tobacco to protect the plant from infection by pathogenic *Erwinia* biofilms.

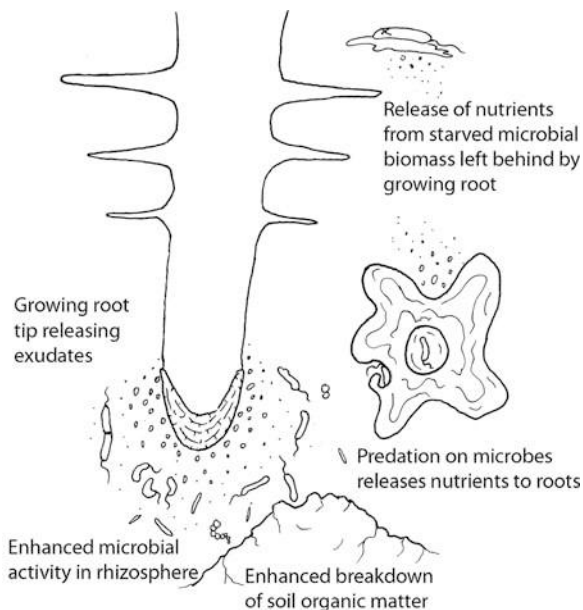
---

## Nutrient Relations in the Rhizosphere

### The Rhizosphere Effect

Arguably the most important function of soil microorganisms is the recycling of mineral nutrients from organic matter. This process occurs throughout the soil but is particularly accelerated in the rhizosphere. The so-called rhizosphere effect arises from the exudation of labile compounds (organic acids, sugars) by roots that stimulate microbial activity, leading to enhanced nutrient cycling (Kuzyakov and Xu 2013). It has been estimated that up to 40 % of C photosynthesized by plants is secreted into the rhizosphere. In the short term, addition of labile C causes microbes to grow and immobilize inorganic N, making it less available to plants. However, this can lead to increased N availability by several mechanisms. Root exudates can “prime the pump” by increasing the activity and populations of rhizosphere microbes which then increase rates of organic matter decomposition and N mineralization. N must be released from the microbial biomass for the plants to benefit from this effect. This can occur as a result of trophic dynamics in the rhizosphere, in which predation by protozoa or other predators causes the rapid turnover of microbial biomass and release of mineral N to plants. The same effect can also

**Fig. 4** An illustration of the “rhizosphere effect,” in which root exudation stimulates net N mineralization by enhancing microbial activity in the rhizosphere (see text for details)



result from the dynamics of root growth in the soil. Exudates are produced maximally near the growing root tip, and so as the root moves through the soil, it creates a dynamic “boom-bust” pattern in its wake, causing previously stimulated microbes to release their N upon starvation. These processes are illustrated in Fig. 4.

## Plant-Microbe Competition for N

When considered on a short time scale, plants and microbes compete for nutrients (Kuzyakov and Xu 2013). Bacteria and fungi have a great advantage over plant roots in terms of absorbing nutrients from the soil solution, mainly because of their high surface to volume ratios (due to their smaller dimensions). Because of this, plants were classically considered to only have access to inorganic forms of N that exist in excess of microbial growth needs. Therefore, plant N availability is often estimated by measuring rates of net N mineralization, the balance between gross N mineralization (inorganic N released from decaying organic matter) and immobilization of inorganic N into microbial biomass. Theoretically, net mineralization occurs when the C:N ratio of the decaying organic matter drops below some threshold relative to the C:N ratio of the microbial biomass, at which point N exists in excess for the growing microbes. However, plant roots can absorb N even when net mineralization is not occurring by directly competing with microbes for uptake. Among inorganic forms, plants generally compete better for nitrate ( $\text{NO}_3^-$ ) than for ammonium, probably because nitrate is much more mobile in soils. This allows plants to absorb nitrate through mass flow of soil solution through the



roots (driven by transpiration at the leaf surface). Also the higher mobility of nitrate allows plant roots to create a larger depletion zone around the roots, leading to a stronger concentration gradient and more rapid diffusion to the root surface.

Because microbes are supposed to outcompete plant roots during nutrient uptake and because soil microbial growth should generally be limited by either C or N, it was initially surprising to learn that some plants acquire significant amounts of their N budget from the uptake of organic forms of N. In particular, amino acid uptake has proven to be widespread among plants, even when in a non-mycorrhizal state. In at least one ecosystem, plants and microbes appear to have different preferences for amino acids: in a study of the alpine sedge, *Kobresia myosuroides*, and alpine soil microbes, the smaller, more rapidly diffusing amino acid, glycine, was preferable to plants, whereas the energy-rich and metabolically central amino acid, glutamate, was taken up more rapidly by microbes (Lipson et al. 1999). These complementary preferences are consistent with the fact that plants are autotrophic and therefore limited by mineral nutrients such as N rather than by C, whereas most soil microbes are heterotrophic and therefore are generally limited by organic C.

When considering short-term direct competition between plants and microbes, the best predictor of the outcome is probably root and mycorrhizal surface area. However, it is important to keep in mind the fact that while roots and soil microbes exist in close proximity, they do not share the same niche, and so plant roots do not need to outcompete microbes in general for some limiting nutrient such as N. Plants tend to grow and senesce on a timescale of months to years, whereas soil microbial biomass turns over many times per year. Also, plants are generally N limited, while soil microbes are more commonly energy limited. Therefore, in the short-term, microbes will outcompete plant roots for available N, but in the longer term, plants gain access to N from the turnover of microbial biomass. In turn, plant tissues senesce and are decayed by C-hungry microbes. This relationship between plants and microbes can help retain N in ecosystems: microbes immobilize nutrients when plants are inactive, preventing gaseous and hydrological losses, and then act as a N source for plants during the growing season.

---

## Impact of Plants on Soil Microbial Processes

Just as plants are reliant on soil microbes for recycling of nutrients trapped in organic matter, as discussed in the previous section, most soil microbes (i.e., the heterotrophs) depend on plants for all of their energy requirements. Therefore, the quantity and quality of plant litter will determine the nature of the microbial community (see section “[Impacts of Plants on Microbial Diversity](#)”). Additionally, plants have effects on soil chemistry and microclimate, which in turn control microbially driven processes such as the biogeochemical cycling of N and C (Eviner and Chapin 2003).

Two commonly used indices for predicting rates of decomposition of plant litter are the ratios of C:N and lignin:N. However, other aspects of plant litter chemistry can also alter microbial processes. For example, these indices fail to predict the

slow decomposition rate of mosses: for some mosses, it seems to be phenolic compounds that retard decomposition, but *Sphagnum* mosses produce highly resistant polysaccharides (Hajek et al. 2011). Tannins and polyphenols are also sometimes good predictors of litter decomposition and N mineralization rates. In some ecosystems, polyphenolic compounds from leaf litter can slow N mineralization rates, leading to higher organic:inorganic N ratios. These compounds form complexes with proteins (and possibly other molecules), making soil organic matter harder to degrade while also inhibiting the extracellular enzymes that breakdown these molecules. Polyphenolics, themselves, are also hard to degrade and have toxic effects on some microbes (Cesco et al. 2012). However, in some cases, it appears that phenolics from plants are not generally inhibitory to soil microbial activity but rather provide a high C:N substrate for microbial growth that induces immobilization of inorganic N (Eviner and Chapin 2003). Regardless of the mechanism, the net effect of phenolic rich litter will generally be lower inorganic N availability.

The interplay between plant litter chemistry and its decomposition by soil microbes can lead to positive feedbacks that reinforce patterns of soil fertility. For example, plants from stressful environments generally have longer-lived, more protected, nutrient-poor, litter that is harder to decompose, which leads to low N mineralization rates in the soil and continued low-nutrient conditions. Conversely, plants in high-fertility soils generally produce higher quality litter, in turn leading to a more active microbial community with higher rates of mineralization and decomposition. In general, the former low-nutrient condition is associated with higher fungal:bacterial ratios, whereas high-fertility soils are thought to be more bacterially dominated.

The classic model of succession predicts that mature successional stages will have more closed, tight N cycles with lower losses than early stages. Related to this idea is the hypothesis that nitrification is inhibited in mature forested ecosystems, as the end product of this process, nitrate, is easily lost through leaching and denitrification (conversion to gaseous products). Microbes that carry out nitrification are not necessarily inhibited by any direct, specific mechanism, but a slower, tighter N cycle should have the same effect: because of the low energy yield of  $\text{NH}_4^+$  and nitrite ( $\text{NO}_2^-$ ) oxidation that fuels these autotrophs, they require high levels of substrate and will not do well when slow litter decomposition and efficient uptake by plants and heterotrophic microbes lead to low inorganic N concentrations (Eviner and Chapin 2003). However, there are reports of inhibition of nitrifying microbes by polyphenolic compounds (Cesco et al. 2012).

Plants can also control microbial processes through effects on soil pH and redox (Eviner and Chapin 2003). One proposed mechanism for the inhibition of nitrifying bacteria in mature forested ecosystems is the development of acidic soils. However, the discovery of ammonia-oxidizing archaea that are active at low pH helps account for continued nitrification at low pH in some ecosystems. Plant communities may regulate microbial metabolic pathways in Arctic ecosystems through pH and redox effects. Arctic plants with aerenchymous roots, such as various sedges, transport  $\text{O}_2$  to the rhizosphere, potentially inhibiting the strictly anaerobic process of methanogenesis, though also potentially allowing the rapid escape of methane

from saturated soil layers. Mosses, such as *Sphagnum*, tend to create more water-logged, anoxic conditions because of their tremendous water holding capacity. The effect of *Sphagnum* on soil water content is also an example of how plants can alter the soil microclimate. These mosses can also act as efficient insulators of soil, regulating thaw depth in arctic tundra soils. In most ecosystems, either temperature or soil water content will limit microbial activity at some point. The plant community affects both of these variables through shading, sheltering, and transpiration. These effects depend on plant community characteristics such as canopy structure, growth rate, and root:shoot allocation patterns.

---

## PMI Effects on the Soil Matrix

Plants and microbes collaborate to alter the physical and chemical nature of the soil environment. Humic substances are a diverse and complex set of organic compounds that form from plant and microbially produced organic compounds as these are modified by soil microorganisms and animals. The quantity and quality of humic substances determine major soil properties such as water and nutrient holding capacity, soil structure, redox processes, and rates of C sequestration. The nature of organic matter that accumulates in soils depends on interactions among plants, their microbial mutualists, and the saprotrophic microbes that degrade and modify the plant litter. As discussed in the previous section, plant chemistry influences the decomposition rates by microbes, which in turn controls the recycling of nutrients and the accumulation of stable soil organic matter. In addition to saprotrophic microbes, mycorrhizal fungi can also have important effects on soil organic matter. For example, the ectomycorrhizal fungus, *Cenococcum geophilum*, promotes the buildup of recalcitrant organic matter in a thick litter layer, in part by the production of antibiotics which it transfers to its host plant (Ponge 2013). Arbuscular mycorrhizae in the genus, *Glomus*, produce the glycoprotein, glomalalin. This compound stabilizes soil aggregates, leading to a soil structure that is more favorable for root growth, O<sub>2</sub> diffusion, etc. In a more general sense, plant roots and fungal hyphae improve soil structure, mechanically by creating channels as they grow through the soil and chemically by the production of various polysaccharides and other substances that glue together fine mineral particles into larger aggregates.

---

## Impacts of Plants on Microbial Diversity

### Culture-Independent Characterization of Microbial Diversity

Until the late 1980s, most microbiological studies of any environment, including studies of microbes associated with plants, relied on being able to culture microbes in the laboratory. However, laboratory culturing methods recover a small fraction of the true microbial diversity in any given environment. In most cases, this fraction

was less than 1 % of the existing microbial diversity, and often far less. The development of what is now known as culture-independent molecular techniques revolutionized the investigation of environmental microbiology and radically altered our understanding of microbial diversity in countless environments, including those associated with plants. Instead of determining microbial species by growing them in liquid or solid media before chemical or morphological analysis, culture-independent methods directly analyze the genetic information in the microbes, typically the DNA. Also, unlike culturing methods that focus on one species at a time, culture-independent molecular methods can simultaneously investigate all the members of a particular community using the information of the genetic sequences to determine the types of microorganisms present in a sample.

The most common gene targeted for this type of analysis is the small-subunit ribosomal RNA (rRNA) gene, known as the 16S rRNA in Bacteria and Archaea. (In Eukarya it is known as the 18S rRNA because the RNA is significantly larger in eukaryotes). Small-subunit rRNA gene sequences are effective genetic markers for culture-independent microbial studies for a number of reasons. First, this gene sequence is found in all forms of cellular life: Bacterial, Archaeal, and Eukaryal. Second, there exists a large and rapidly growing database of rRNA gene sequences from both cultured and uncultured microbes, allowing ready species identification and phylogenetic analyses. Third, when comparing the sequences of 16S rRNA genes among organisms, it was found to have both highly conserved regions and highly variable regions of sequence. For example, some regions of the sequence were exactly the same between extremely diverse organisms, such as all of the Bacteria or between *E. coli* and humans, while other regions were so variable one can detect sequence differences between different closely related species of microbes. The conserved regions were critical for designing PCR primers that could amplify this gene from, for example, all the bacterial species in a soil sample, while the variable regions were important for telling the species apart. Recently, these methods have been combined with high-throughput sequencing approaches, also called next-generation sequencing (NGS). NGS allows the generation of hundreds of thousands to millions of DNA sequences simultaneously. Furthermore, using PCR primers labeled with unique “barcode” sequences at their 5' end, one can use NGS to describe gene diversity from many environmental samples in a single sequencing reaction. After the sequencing, computational methods are used to determine which sequences came from which samples and what organisms are present in each sample.

## Microbial Diversity of the Phyllosphere

The parts of the plant that live above ground, the stem, branches, and leaves, together comprise what is known as the phyllosphere (Vorholt 2012). Like the rhizosphere, the phyllosphere, and leaves in particular, provides plenteous habitats for microbes. Altogether, the collective global surface area of terrestrial plant leaves is roughly double the total of the land surface area upon which the plants

grow, providing living space for an astonishing  $1 \times 10^{26}$  microbial cells! The morphology of the leaves, including the three-dimensional surface contours and leaf structures (veins, stomata, trichomes, etc.), the chemical composition of the surfaces and the local environmental conditions determine to a large degree the types of microbes that persist and grow on leaves. The top of leaf surfaces is exposed to direct sunlight and UV radiation, and the waxy cuticle prevents plant desiccation and helps retain the plant's own metabolites. This makes for an oligotrophic (nutrient poor) environment, selecting microbes able to survive and grow in these stressful conditions. The undersides of leaves are less exposed to light and, while still covered in a waxy cuticle, tend to retain moisture more readily.

Leaf morphological structures also influence the ability of microbes to colonize the surface of the leaf. Microbial communities tend to form in clumps, called aggregates, in the crevices formed at epidermal cell junctions, along the leaf veins and at the base of trichomes. These aggregate cells can form biofilms by secreting extracellular polymeric substances to protect from desiccation and other stresses. The leaf aggregates tend to also be found in the relatively moist surface depressions. The aggregates contain fungi as well as bacteria, but archaea are rare in the phyllosphere.

While the presence and abundance of bacteria have long been known, the advent of culture-independent molecular methods and NGS, in particular, have allowed for a much deeper appreciation of the true extent of microbial diversity in the phyllosphere. They are also providing the means for the comprehensive analysis of microbial communities across hundreds of thousands of samples. This will be necessary to determine the subtler abiotic and biotic factors affecting phyllosphere diversity, especially given the enormous environmental variability across environments and even within a single plant (e.g., the microbial diversity of leaf surfaces at the top versus bottom branches of a redwood tree (Vorholt 2012)).

So, what have culture-independent methods revealed about the diversity of the phyllosphere? First, the studies done so far have determined that the species richness tends to be very high and increases as one moves from temperate to tropical environments. Given that moisture seems to be a limiting factor, this may be a function of greater rainfall in the tropics and perhaps higher growing temperatures and slower leaf turnover. Second, while species richness is relatively high, the phyllosphere as a whole is less diverse than in typical soil rhizosphere communities. The phyllosphere is typically more nutrient poor and short-lived than the rhizosphere. Third, culture-independent analysis of phyllosphere microbial diversity from very different plant species found it was dominated by bacterial species from a fairly limited range of bacterial phyla. The Proteobacteria, particularly Alphaproteobacteria families such as Methylobacteriaceae and Sphingomonadaceae, were dominant, comprising upwards of 70 % of the bacterial species on leaves. Other common and abundant phyla included the Bacteroidetes and the Actinobacteria. Of the four different plant species investigated in one study, researchers found that between 30 and 40 genera of bacteria were consistently common on leaves, though the proportions of these genera (and certainly the specific strains or species) varied considerable across the various plant species.

In terms of the particular abiotic and biotic factors that determine microbial community diversity in the phyllosphere beyond the general ones mentioned, there are still far more questions than answers. What is known is that environmental factors, such as nitrogen-fertilization, exposure to solar radiation and pollution, as well as biotic factors such as leaf age, do significantly affect the structure of microbial communities. Plant genotype also appears to play an important role in the microbes that persist on leaves. Moreover, overall diversity within a plant species tends to be consistently lower than between species. For instance, a study of pine tree phyllosphere microbial diversity found significantly higher microbial diversity among the phyllosphere of different pine species with overlapping geographic distributions than within the same species.

### **Microbial Diversity of the Rhizosphere**

The microbial communities in the soils associated with plant roots and the immediately surrounding soil (the rhizosphere) represent one of the most diverse and least understood ecologies on the planet (Berendsen et al. 2012). On average, one gram of rhizosphere soil contains on the order of  $10^8$ – $10^9$  microbial cells per gram, which include an estimated 30,000 species. As is the case with the human (mammalian) gut microbiome, the total number of genes in the microbes of the rhizosphere greatly exceeds that of the plant itself, making this a so-called second genome of the plant. In many respects, this microbiome provides similar services to the plant as the mammalian microbiome to its host: nutrient uptake (e.g., phosphorus, nitrogen, minerals, and organic matter), pathogen defense, and host-immunity modulation. In turn, the plant provides the rhizosphere with food in the form of root exudates and sometimes protection.

Culture-independent molecular analyses show that the effects of plants on rhizosphere microbial diversity are significant. Soils are typically carbon poor. Plants secrete as much as 40 % of the products of photosynthesis into the rhizosphere, fostering a high local microbial growth rate compared with the surrounding bulk soils, which are often in dormant state. However, while the cell abundance of the rhizosphere is higher compared with non-plant-associated soils, microbial diversity is lower. Plant root exudates have been shown to enhance the growth of particular strains of heterotrophic bacteria while simultaneously inhibiting others. This results in a higher abundance of cells but a reduced overall diversity. Tomatoes, cucumbers, and sweet peppers have all been shown to secrete various organic acids (e.g., citric acid) that alter local pH around the root and enhance the growth of microbes able to use these carbon sources. Cereal crops, on the other hand, have been shown to inhibit the growth of specific microbe strains by secreting secondary metabolites into the soils. Plants can also affect biofilm formation by interfering with the microbes' ability to produce quorum-sensing molecules.

The effects of plant root exudates on the rhizosphere are also known to be plant-species-specific and can even vary among different genotypes within the same species. Culture-independent microbial diversity studies have found the rhizosphere

microbiome differs among different plant species in the same soil types. Moreover, transplanting a plant species in a different soil can alter that soil's microbial community to resemble the soil from which the plant originated. Different plant genotypes of *Arabidopsis* were shown to alter the plant rhizosphere community, particularly the Alphaproteobacteria and fungal communities. For example, an *Arabidopsis* mutant that produced increased phenolic compounds and decreased sugars had distinct rhizosphere microbiomes compared with wild-type plants. Interestingly, the effects of different plant genotypes on the rhizosphere can also alter the levels of pathogenic microorganisms in the soils. For instance, certain potato cultivars favor higher levels of Pseudomonadales, Streptomycetaceae, and Micromonosporaceae, all of which are known to control plant pathogens to some degree. Finally, it appears that plants can alter their recruitment of beneficial bacteria when under attack by insect herbivores or pathogenic organisms. Several plant species have been shown to release root compounds that increase beneficial organisms (e.g., *Bacillus subtilis*) in the soils when under attack.

---

## Impacts of Microbes on Plant Diversity

Soil microbes have a profound influence on the composition of plant communities (Bever et al. 2012). Microbial communities can differentially affect the success of plant species through impacts on nutrient cycling (as discussed earlier) and by direct positive and negative interactions with plants (such as mutualisms and pathogens, respectively). Many examples that demonstrate these links come from studies of invasion of ecosystems by exotic plant species (van der Putten et al. 2007). For example, it was found that rare plant species tend to be limited by the accumulation of pathogens, whereas invasive plants tended to perform well outside of their native range by forming fewer negative interactions with microbes. One explanation for this could be that invasive plants escape negatively interacting microbes from their native range.

The community of mycorrhizal fungi available to colonize plant roots can also determine plant success, and the diversity of mycorrhizal fungal communities is linked to the diversity and productivity of plant communities. Restoration efforts to reintroduce rare, native plants often require inoculation of mycorrhizal fungi to the soil, particularly for orchid species (Dearnaley et al. 2012). Mycorrhizal diversity can influence the outcome of competition among plant species. Often a variety of mycorrhizal and/or endophytic fungi can colonize a single plant host, but there is variability in how well each fungus benefits the host. The outcome of competition among plant species could be determined by the differential effects of its root mycobionts. Similarly, many invasive plants are non-mycorrhizal, while others are highly general in their mycorrhizal partners, partly explaining their ability to rapidly expand their ranges. Some invasive plants (such as the non-mycorrhizal, *Alliaria petiolata*) produce glucosinolates that reduce the abundance of mycorrhizal fungi that support native plants; other invasive plants (such as *Centaurea maculosa*) exploit native mycorrhizae by tapping into the hyphal network to parasitize the host plants (van der Putten et al. 2007).



Another example of a plant invasion mediated by plant-microbe mutualisms is that by the grass, *Lolium arundinaceum*, and its fungal leaf endophyte, *Neotyphodium coenophialum*. The presence of this leaf endophyte makes the host less palatable to insects while increasing herbivory on its competitors (Saikkonen et al. 2013). This example highlights the fact that plant-plant interactions can be mediated at multiple trophic levels, including microbes and herbivores. N-fixing mutualisms can figure prominently in plant invasions. The actinorhizal shrub, *Myrica faya*, is a noxious invader of forests in Hawaii, leguminous *Acacia* species have invaded South African ecosystems, and a large number of invasive species in North America are also legumes (Ehrenfeld 2003). Invasions by N fixers can lead to increased soil N, facilitating invasion by other plant species. The above examples of plant invasions that are mediated by PMI also have parallels in plant succession. In some cases, early successional plant species are less reliant on mycorrhizae, but the accumulation of pathogens and mycorrhizal fungi in the soil eventually shifts the competitive advantage to later successional species (Bever et al. 2012). Similarly, early N-fixing stages (such as lupines) can facilitate the establishment of other plants.

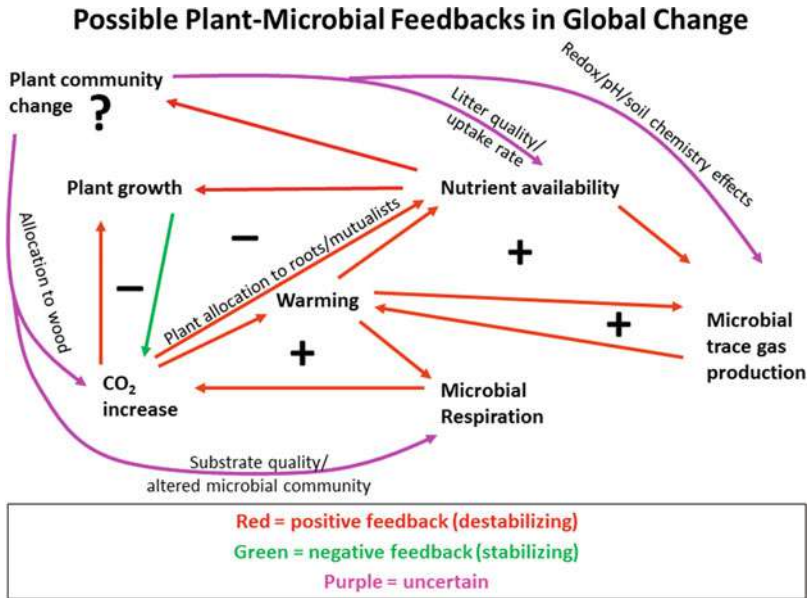
Microbes also mediate interspecific plant competition through their involvement in allelopathic interactions (chemical warfare) among plants (Cipollini et al. 2012). As mentioned above, mycorrhizal fungi of competing species are targeted by some allelopathic plants. There are also reports of allelochemicals that inhibit PGPR and rhizobia. In other cases, soil microbes protect plants against allelopathic competitors, either by degrading potentially inhibitory chemicals produced by plants, by increasing plant resistance to allelochemicals (certain mycorrhizae do this), or by reducing the production of these compounds by plants (leaf pathogens are one example). The opposite effect has also been reported, in which transformation by soil microbes enhances allelopathic effects. For example, gallotannin produced by *Phragmites australis* is metabolized by soil microbes to the more phytotoxic, gallic acid. In the association between *Festuca* species and their endophytic fungus, production of inhibitory compounds appears to be linked to the fungal partner. Allelopathic plants may also utilize mycorrhizal networks among species to deliver phytotoxins.

---

## The Role of Plant-Microbe Interactions in Global Change

The magnitude of biological feedbacks to climate change represent one the largest current uncertainties in climate models (Dieleman et al. 2012). The biosphere engages in both positive and negative feedbacks with the Earth system, exacerbating or stabilizing conditions, respectively. Plants and soil microbes play central roles in these biological feedbacks to changes in atmospheric CO<sub>2</sub> and climate (Fig. 5). Of the roughly eight gigatons of CO<sub>2</sub>-C per year produced through human activities, about half is currently reabsorbed by sinks on land and sea. Plants contribute to the biological sink on land through increased uptake of atmospheric





**Fig. 5** A diagram of potential feedbacks of PMI on global change processes

CO<sub>2</sub> (sometimes referred to as CO<sub>2</sub> fertilization), creating a negative feedback. There are also direct effects of warming on soil respiration (CO<sub>2</sub> production) and greenhouse gas production (N<sub>2</sub>O, CH<sub>4</sub>) by microbes that can produce positive feedbacks, amplifying the warming effect. However, there are numerous additional feedbacks that operate through PMI with less certain effects. Because plants are primarily limited by mineral nutrients, such as N and P, in response to elevated CO<sub>2</sub>, plants often increase allocation to roots, exudates, and mutualists in an attempt to increase nutrient acquisition. Because plant growth responses to elevated CO<sub>2</sub> are generally nutrient limited, this effect can create a negative feedback by helping plants to absorb more CO<sub>2</sub>. However, the differential ability of plant species to effect changes in nutrient acquisition and the resulting changes in plant litter, rhizosphere activity, and nutrient cycling can alter plant communities in uncertain ways. For example, because N fixation is an energetically expensive process, the extra photosynthate afforded by elevated CO<sub>2</sub> might give symbiotic N fixers a competitive advantage. Other, direct climatic effects not shown here can also lead to changes in plant communities. Changes in plant communities produce direct feedbacks to the C cycle, depending especially on their allocation to woody tissues, a highly stable form of C. The remaining feedbacks shown in Fig. 5 are mediated by PMI, as discussed earlier in Sect. [Impacts of Plants on Microbial Diversity](#). Altered plant chemistry (resulting from a change in plant species composition or a CO<sub>2</sub>-induced change in secondary chemistry) could alter rates of litter decomposition, soil sequestration, and nutrient mineralization. Altered soil N

availability will translate into altered N<sub>2</sub>O fluxes. Changes in plant community could feedback to alter trace gas production through impacts on soil redox conditions, pH, and chemistry. For example, loss of mosses in the Arctic (say, because of sensitivity to warmer, drier conditions) could drastically alter soil conditions and the relative production of CO<sub>2</sub> and CH<sub>4</sub>. Changes in plant communities will lead to altered microbial communities with different metabolic properties. For example, changes in fungal:bacterial ratio or overall species composition can affect the biomass-specific respiration rate (or C use efficiency, CUE), leading to different amounts of CO<sub>2</sub> produced per unit microbial biomass per unit time. The tangled web shown in Fig. 5 indicates the great complexity of PMI and their implications for the planet. The magnitudes of these effects are active areas of research.

---

## Future Directions

Numerous gaps still remain in the current understanding of PMI. For example, the roles of PMI in feedbacks to global change, especially multifactor changes such as increased CO<sub>2</sub> and temperature, are not yet included in climate change models (Dieleman et al. 2012). Similarly, the mediation of biological invasions by PMI is an active area of research. There is still much unknown about the genetics of plant-microbe mutualisms. For example, sequencing the genome of *Glomus intraradices* is challenging due to its heterozygosity and lack of a uninucleate stage. And while rhizobia-legume mutualisms have been studied in great detail, current understanding of non-rhizobial N-fixing symbioses lags behind. And while great progress has been made towards understanding the factors that control microbial diversity in soils, a detailed understanding of how plants shape microbial communities in the phyllosphere and rhizosphere has yet to emerge. Finally, it appears that plants may be emerging as reservoirs for bacterial pathogens of humans, but this phenomenon is not yet well understood.

There are numerous new techniques emerging to help answer these lingering questions. The development of high-throughput sequencing technology and other molecular techniques is rapidly changing the face of microbial ecology, making the study of complex microbial communities more tractable. Meanwhile, analytical techniques are making rapid advancements, allowing sensitive detection of processes at unprecedented spatial and temporal resolution. For example, new technology such as laser and cavity ring-down techniques allow the real-time measurement of trace gases and their stable isotopes. Novel visualization techniques, such as reporter genes and synchrotron-based methods, are creating new windows into the rhizosphere (Raab and Lipson 2010). These and other novel methods allow the quantification and identification of C compounds transported from plants into the rhizosphere and to root mutualists. Given the urgent nature of some of the unanswered questions surrounding PMI and the advent of these new techniques, the next decade should produce some very interesting work in these areas.

## References

- Badri DV, Weir TL, Dvd L, Vivanco JM. Rhizosphere chemical dialogues: plant–microbe interactions. *Curr Opin Biotechnol.* 2009;20:642–50.
- Baumgartner K, Coetzee MPA, Hoffmeister D. Secrets of the subterranean pathosystem of *Armillaria*. *Mol Plant Pathol.* 2011;12:515–34.
- Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. *Trends Plant Sci.* 2012;17:478–86.
- Berry AM, Mendoza-Herrera A, Guo Y-Y, Hayashi J, Persson T, Barabote R, et al. New perspectives on nodule nitrogen assimilation in actinorhizal symbioses. *Funct Plant Biol.* 2011;38:645–52.
- Bever JD, Platt TG, Morton ER. Microbial population and community dynamics on plant roots and their feedbacks on plant communities. *Annu Rev Microbiol.* 2012;66:265–83.
- Bonfante P, Genre A. Mechanisms underlying beneficial plant – fungus interactions in mycorrhizal symbiosis. *Nat Commun.* 2010;1:48.
- Cesco S, Mimmo T, Tonon G, Tomasi N, Pinton R, Terzano R, et al. Plant-borne flavonoids released into the rhizosphere: impact on soil bio-activities related to plant nutrition. A review. *Biol Fertil Soils.* 2012;48:123–49.
- Cipollini D, Rigsby CM, Barto EK. Microbes as targets and mediators of allelopathy in plants. *J Chem Ecol.* 2012;38:714–27.
- Dearnaley JDW, Martos F, Selosse M-A. Orchid mycorrhizas: molecular ecology, physiology, evolution and conservation aspects. In: Hock B, editor. *Fungal associations*. 2nd ed. Berlin/Heidelberg: Springer; 2012. p. 207–30.
- Dieleman WIJ, Vicca S, Dijkstra FA, Hagedorn F, Hovenden MJ, Larsen KS, et al. Simple additive effects are rare: a quantitative review of plant biomass and soil process responses to combined manipulations of CO<sub>2</sub> and temperature. *Glob Chang Biol.* 2012;18:2681–93.
- Ehrenfeld JG. Effects of exotic plant invasions on soil nutrient cycling processes. *Ecosystems.* 2003;6:503–23.
- Eviner VT, Chapin FS. Functional matrix: a conceptual framework for predicting multiple plant effects on ecosystem processes. *Annu Rev Ecol Evol Syst.* 2003;34:455–85.
- Hajek T, Ballance S, Limpens J, Zijlstra M, Verhoeven JTA. Cell-wall polysaccharides play an important role in decay resistance of sphagnum and actively depressed decomposition in vitro. *Biogeochemistry.* 2011;103:45–57.
- Javot H, Penmetsa RV, Breuillin F, Bhattarai KK, Noar RD, Gomez SK, et al. *Medicago truncatula* *mtpt4* mutants reveal a role for nitrogen in the regulation of arbuscule degeneration in arbuscular mycorrhizal symbiosis. *Plant J.* 2011;68:954–65.
- Kuzyakov Y, Xu X. Competition between roots and microorganisms for nitrogen: mechanisms and ecological relevance. *New Phytologist.* 2013;198:656–69.
- Lipson DA, Raab TK, Schmidt SK, Monson RK. Variation in competitive abilities of plants and microbes for specific amino acids. *Biol Fertil Soils.* 1999;29:257–61.
- Masson-Boivin C, Giraud E, Perret X, Batut J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* 2009;17:458–66.
- Mayerhofer MS, Kernaghan G, Harper KA. The effects of fungal root endophytes on plant growth: a meta-analysis. *Mycorrhiza.* 2013;23:119–28.
- Newsham KK. A meta-analysis of plant responses to dark septate root endophytes. *New Phytol.* 2011;190:783–93.
- Oldroyd GED, Murray JD, Poole PS, Downie JA. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet.* 2011;45:119–44.
- Pawlowski K, Newton WE, editors. *Nitrogen-fixing actinorhizal symbioses*. Dordrecht: Springer; 2008.
- Ponge J-F. Plant-soil feedbacks mediated by humus forms: a review. *Soil Biol Biochem.* 2013;57:1048–60.

- Raab TK, Lipson DA. The rhizosphere: a synchrotron-based view of nutrient flow in the root zone. In: Grafe M, Singh B, editors. *Advances in understanding soil environments by application of synchrotron-based techniques*. 1st ed. The Netherlands: Elsevier; 2010.
- Raghoebarsing AA, Smolders AJP, Schmid MC, Rijpstra WIC, Wolters-Arts M, Derksen J, et al. Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature*. 2005;436:1153–6.
- Saikkonen K, Gundel PE, Helander M. Chemical ecology mediated by fungal endophytes in grasses. *J Chem Ecol*. 2013;39:962–8.
- Santi C, Bogusz D, Franche C. Biological nitrogen fixation in non-legume plants. *Ann Bot*. 2013;111:743–67.
- Shiraishi A, Matsushita N, Hougetsu T. Nodulation in black locust by the *Gammaproteobacteria Pseudomonas* sp. and the *Betaproteobacteria Burkholderia* sp. *Syst Appl Microbiol*. 2010;33:269–74.
- Smith SE, Read DJ. *Mycorrhizal symbiosis*. 3rd ed. New York: Academic; 2008.
- van der Putten WH, Klironomos JN, Wardle DA. Microbial ecology of biological invasions. *ISME J*. 2007;1:28–37.
- Vorholt JA. Microbial life in the phyllosphere. *Nat Rev Microbiol*. 2012;10:828–40.

## Further Reading

- Crespi M, editor. *Root genomics and soil interactions*. Ames: Wiley-Blackwell; 2013.
- Maheshwari DK, editor. *Bacteria in agrobiolgy: stress management*. Heidelberg: Springer; 2012.
- Pinton R, Varanini Z, Nannipieri P, editors. *The rhizosphere: biochemistry and organic substances at the soil-plant interface*. 2nd ed. Boca Raton: CRC Press; 2007.

---

# Patterns and Controls of Terrestrial Primary Production in a Changing World

# 8

Alan K. Knapp, Charles J. W. Carroll, and Timothy J. Fahey

## Contents

Introduction .....	207
How Is NPP Measured? .....	211
Direct Field Measurement of Aboveground Primary Production .....	212
Field Approaches for Belowground Primary Production .....	213
Remote-Sensing and Modeling Approaches to Terrestrial Primary Production .....	214
Patterns and Controls on Productivity from Global to Local Scales .....	216
Abiotic Controls on NPP .....	218
Temporal Variability in NPP .....	226
Disturbance and NPP .....	229
Biotic Controls on NPP .....	230
Vegetation Structure and NPP .....	231
Biodiversity Effects on Productivity .....	232
Community Change and NPP .....	235
Herbivory and NPP .....	236
Belowground Productivity: Patterns and Controls .....	237
Controls of NPP and the Future .....	239
Future Directions .....	244
References .....	244

---

## Abstract

- Primary production is the process by which solar energy is converted to chemical energy by autotrophic organisms, primarily green plants on land, providing the energy available to power earth's ecosystems. In this process atmospheric CO<sub>2</sub> is incorporated into organic matter, thereby playing a

---

A.K. Knapp (✉) • C.J.W. Carroll  
Graduate Degree Program in Ecology, Department of Biology, Colorado State University, Fort Collins, CO, USA  
e-mail: [aknapp@colostate.edu](mailto:aknapp@colostate.edu); [cjwcarroll@gmail.com](mailto:cjwcarroll@gmail.com)

T.J. Fahey  
Department of Natural Resources, Cornell University, Ithaca, NY, USA  
e-mail: [tjf5@cornell.edu](mailto:tjf5@cornell.edu)

dominant role in the global carbon cycle with crucial implications for global climate change. Net primary production (NPP) is the amount of fixed energy or organic matter left over after the plants have met their own respiratory needs and represents the amount of energy available to the consumers, including humans. Across the earth's terrestrial biomes, a large range of NPP is observed with the highest values in tropical forests and wetlands, intermediate values in temperate forests and grasslands, and lowest in extremely cold or dry deserts.

- Accurate measurement of NPP is challenging despite the simple concept that it represents the amount of new biomass added to the plants in a given time period. This is because a significant and highly variable proportion of NPP is lost from the plants by processes such as herbivory, volatilization, and carbon flux to the soil. Methods of measuring NPP are diverse, being dependent on the structure and dynamics of the vegetation. For example, harvest methods in which the aboveground tissues are periodically clipped from quadrats of known area can be effective for quantifying aboveground NPP in herbaceous vegetation (e.g., grasslands), whereas in woody vegetation, the growth of woody tissues must also be measured. Moreover, measurements of total NPP in terrestrial ecosystems must account for root growth which can be very challenging. As a result, reliable estimates of total NPP are few.
- Plants allocate a large proportion of their fixed energy to their root systems to fuel additional root growth and to meet their respiratory needs. The proportion of total NPP that goes to belowground NPP ranges from about 25 % to over 50 % and is higher in ecosystems where the degree of limitation by soil resources is greater, i.e., dry or nutrient-poor sites. Surprisingly, over 10 % of NPP is contributed by plants to the soil in the form of rhizosphere carbon flux including exudation, rhizodeposition, and allocation to mycorrhizal fungi and other symbionts.
- Variation in NPP results from differences among ecosystems in the amount of photosynthetically active radiation (PAR) reaching the plant canopy, the amount of that PAR absorbed by the foliage (APAR), the biochemical efficiency of the plants under optimal environmental conditions, and the degree to which actual conditions are less than optimal. The APAR depends in part on the amount of foliage surface area per unit ground area (leaf area index – LAI) which ranges from less than 1 in dry or infertile sites to over 10 in some resource-rich forests. Large-scale monitoring of estimated NPP is possible using satellite imagery of reflected solar radiation that can be converted into vegetation LAI and combined with environmental measurements that indicate the degree of stress reduction to photosynthetic activity.
- Four principal abiotic factors usually limit the amount of NPP on land – light, water, temperature, and mineral nutrients – and all these abiotic factors are changing rapidly as a result of human activity, with highly uncertain implications for global and local NPP. Commonly, two or more of these abiotic factors concurrently or sequentially limit NPP, but water deficit is arguably

the most widespread single factor constraining global NPP. The effect of temperature on NPP is most closely related to subfreezing conditions that limit the length of the growing season in temperate and high-latitude environments. Nitrogen is the most important limiting mineral nutrient in most ecosystems, although in highly weathered tropical soils where nitrogen-fixing organisms are abundant, phosphorus may be the most limiting nutrient.

- Biotic factors can play a key role in regulating NPP so that human activities such as vegetation management and introduction of exotic species will exert a major influence on future patterns of NPP. The effects of biodiversity on NPP have proven difficult to establish, but experimental tests suggest that loss of species can reduce NPP particularly if a dominant species is lost or when species numbers become very low, diminishing complementarity in resource use by coexisting species. Dramatic shifts in plant community structure, for example, the ongoing invasion of grassland vegetation by woody plants, can cause changes in NPP that appear to depend in part on climate. Consumption of plant tissues by herbivores often can have a negative effect on NPP, but in many grasslands, compensatory growth responses to herbivory can result in no reduction in NPP or in some cases even stimulation of NPP by herbivory.
- Temporal variation in NPP results from interannual variation in both environmental and biotic factors as well as pulse disturbance events that can reset the successional clock. The response of NPP to interannual variation in rainfall seems to be greatest in semiarid and subhumid environments where average precipitation is sufficient to sustain highly productive communities (vs. true deserts). Following natural or human disturbances, forests exhibit a recurring pattern in which NPP peaks after a few decades of stand development, followed by a decline with age in older stands.
- Global environmental changes – climate, atmospheric CO<sub>2</sub>, nitrogen deposition, exotic species introductions, etc. – are certain to exert a major influence on global NPP in the future, but the outcomes are highly uncertain because of the complex ways in which all these changes interact with one another to influence the vegetation and NPP. For example, CO<sub>2</sub> enrichment experiments indicate that increasing atmospheric CO<sub>2</sub> concentration can significantly stimulate NPP in young forests, but the effect may be transient because of progressively greater stress by mineral nutrients – unless high N deposition overcomes this limitation.

---

## Introduction

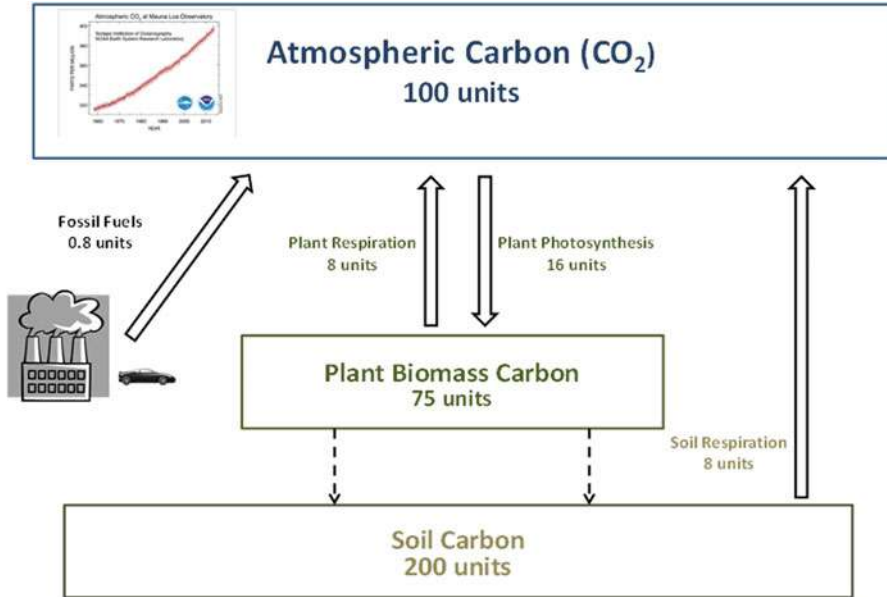
All heterotrophic organisms, from the poles to the tropics, rely on stable forms of chemical energy collectively known as organic (carbon containing) matter derived from biological activity. The energy in virtually all organic matter is ultimately derived from the sun, and the conversion of solar energy to chemical energy is accomplished by autotrophs, primarily green plants in terrestrial ecosystems.

These autotrophs are incredibly diverse in size (<mm to >100 m in height) and life span (a few weeks to thousands of years) and vary widely in their population densities, depending on the resources available. But they all employ a similar photosynthetic process composed of photochemical and biochemical pathways that are highly conserved from an evolutionary perspective. Subtle variations within the photosynthetic process (i.e., C<sub>3</sub>, C<sub>4</sub>, and CAM photosynthetic pathways) can have important implications for determining the amount and global distribution of organic matter produced by plants. But the striking similarities among all autotrophs in the fundamental mechanism by which inorganic CO<sub>2</sub> is converted into organic matter allows us to step back and focus more on the external controls of organic matter production in terrestrial ecosystems and less on physiological variations among autotrophs. This ecosystem perspective is essential for accomplishing the goal of this chapter which is to provide a contemporary and forward looking overview of patterns and determinants of **primary production** (= organic matter production) in terrestrial ecosystems.

Our planet's global carbon cycle, of which atmospheric CO<sub>2</sub> is a key component with direct impacts on climate, depends fundamentally upon terrestrial plant primary production. Why is this so? It is because terrestrial ecosystems account for approximately two-thirds of the global estimate of total primary production, despite covering only a quarter of the earth's surface. The oceans contribute the remainder. Moreover, the annual removal of CO<sub>2</sub> from the atmosphere by the photosynthetic activities of terrestrial plants is about 20 times greater than CO<sub>2</sub> emissions to the atmosphere from fossil fuel burning by humans (Fig. 1). Similarly, CO<sub>2</sub> emitted back to the atmosphere from the respiratory activities of plants is about 10 times that of fossil fuel emissions. Finally, estimates of the amount of carbon stored in terrestrial plants are almost 100 times greater than annual emissions from fossil fuel burning. Indeed, carbon stored in terrestrial plant biomass is equivalent to about 75 % of the carbon found in the atmosphere. The fact that the amount of carbon transferred in and out of the atmosphere by plants is an order of magnitude greater than fossil fuel inputs points to the importance of understanding the dynamics and fate of terrestrial plant primary production. However, the relatively small size of anthropogenic sources of carbon to the atmosphere should not belie their importance. Such emissions have been the dominant cause of the 25 % increase in atmospheric CO<sub>2</sub> levels directly measured in the last 50 years (Fig. 1). Evidence is overwhelming that a consequence of this alteration to the composition of earth's atmosphere will be global warming, an intensification of the global hydrological cycle, and an increase in the number and severity of climatic extremes – and all of these climatic changes will affect plant processes and future levels of primary production. Thus, in order to understand ecological patterns and processes now and in the future, the determinants of primary production across the wide range of earth's terrestrial ecosystems must be understood, from deserts to tropical forests.

Our current understanding of primary production in terrestrial ecosystems is a product of literally thousands of studies conducted during the last 100+ years, but before considering any synthesis of this knowledge, some terms and concepts need to be defined. The total amount of energy fixed (as CO<sub>2</sub> into organic matter) by





From: NOAA (<http://www.esrl.noaa.gov/gmd/ccgg/trends/>) and The Globe Carbon Cycle Project, University of New Hampshire (<http://globecarboncycle.unh.edu/>)

**Fig. 1** Simplified depiction of the global carbon cycle with the central role of processes directly related to production by plants in terrestrial ecosystems highlighted. *Dashed lines* from plant carbon to soil carbon boxes indicate that while plant biomass is the source for most soil carbon, other processes (not shown) determine how much carbon flows from plants to the soil carbon pool. Units are arbitrary and relative to simplify comparisons

plants per unit ground area per unit time is termed **gross primary production (GPP)**. This is the sum of all energy fixed by the autotrophs in the ecosystem. **Net primary production (NPP)** is the amount of energy left over after autotrophs have met their energetic needs through respiration. Thus, NPP is GPP minus respiration by primary producers. NPP represents the amount of energy available to consumers (including humans) in an ecosystem. NPP is typically expressed in units of dry matter ( $\text{grams m}^{-2} \text{ year}^{-1}$ ) rather than units of energy because of the ease of quantifying plant mass and the simplicity of converting mass to energy for plant tissues. As an alternative to units of dry plant matter, grams of carbon also are commonly used to express NPP. Because C content of plant biomass is typically between 45 % and 50 %, converting between plant matter and plant carbon is straightforward. The total mass of plants (per unit area) at any point in time is often referred to as **standing crop** or simply as **biomass**. Many ecologists conceptualize NPP as the amount of *new* biomass added in a given period of time; however, a significant portion of the NPP actually does not appear as new plant tissue but rather is lost from the plant by such pathways as canopy leaching, volatilization, and especially rhizosphere carbon flux, including allocation to mycorrhizal symbionts. Quantifying these components of NPP is very challenging.

**Table 1** Range of NPP and standing biomass (dry matter) for different biomes types (from Huston and Wolverton (2009) (Data are from estimates of above- and belowground components combined and global NPP is based on estimates of the spatial extent of each biome)

Biome	Standing crop biomass (Mg/ha)	Net primary production (g/m <sup>2</sup> /year)	Global net primary production (Pg/year)
<b>Tropical forest</b>	240–388	1,566–2,502	27.4–43.8
<b>Temperate forest</b>	114–268	1,250–1,558	13.0–16.2
<b>Boreal forest</b>	84–128	380–468	5.2–6.4
<b>Tropical savanna and grassland</b>	58	1,080–1,282	29.8–35.4
<b>Temperate grassland and shrubland</b>	14–26	596–786	10.6–14.0
<b>Desert</b>	4–8	102–252	2.8–7.0
<b>Tundra</b>	8–12	178–358	1.0–2.0
<b>Crops</b>	4–6	608–1,008	8.2–13.6
<b>Wetlands</b>	86	2,458	8.3

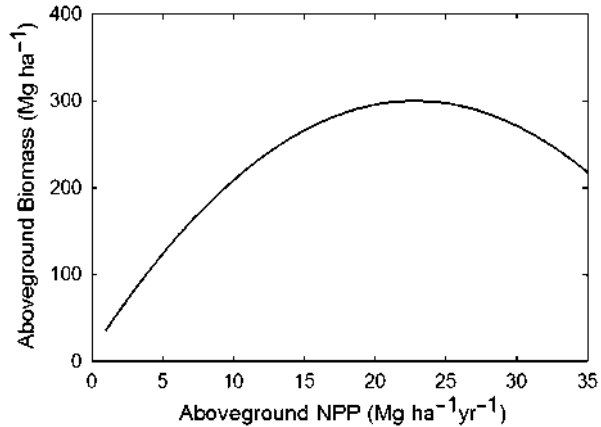
Terrestrial NPP has also been conceptualized by focusing on the ultimate source of energy – the sun. In this approach, think of the vegetation community as a living machine whose growth and metabolism are driven by incoming solar radiation. In this framework, NPP depends upon the efficiency with which the photosynthetically active radiation (PAR) that is absorbed by plant leaves is assimilated into organic matter accumulating in the vegetation. Variation in NPP is the result of differences among ecosystems in the amount of PAR reaching the canopy, the amount of that PAR absorbed by the foliage (APAR), the biochemical conversion efficiency of the plants under optimal environmental conditions, and the degree to which actual conditions are less than optimal. Thus,

$$\text{NPP} = \epsilon * \text{APAR}$$

In this framework the conversion efficiency ( $\epsilon$  – dimensionless) would account for the photochemical efficiency of leaves, energetic costs of growth and maintenance of plant tissues, as well as any environmental stresses, like drought and cold, that reduce photosynthesis below optimal. The APAR term accounts for variation in the amount of PAR reaching the top of the plant canopy, as influenced by day length, cloud cover, etc., as well as the amount of foliage in the plant community (leaf area index or LAI –the leaf surface area per unit ground area) and its architectural arrangement. The LAI depends in part on the availability of soil resources (water, mineral nutrients) and ranges from less than one in deserts to over 10 in some resources-rich forests. Obviously leaves deep in the canopy of such forests receive only enough PAR for minimal photosynthesis, and the energetic costs of growing all the plant tissues – leaves, stem, roots –in these ecosystems set a limit on the maximum NPP attained by terrestrial vegetation (Table 1).

Finally, an alternative to the plant-focused considerations above is a carbon balance perspective on primary production. In this framework, **GPP** is the total

**Fig. 2** Relationship between increasing aboveground NPP in forests and biomass aboveground. Note that at the highest levels of productivity, biomass decreases. This may be because environmental conditions that favor the highest NPP (warm, wet environments) may also favor high turnover of biomass due to death of individuals and rapid decomposition (Modified from Keeling and Philips 2007)



amount of CO<sub>2</sub> (or carbon) that is fixed or taken up by plants in the ecosystem, **ER** (ecosystem respiration) is the amount of CO<sub>2</sub> that is lost or emitted from the ecosystem from the combined metabolic activities of plants and heterotrophs including decomposers (microbes). **Net ecosystem production** or **NEP** is thus **GPP-ER** or the net amount of primary production after losses to respiration by plants, heterotrophs, and decomposers. NEP is a valuable measure for evaluating the balance of CO<sub>2</sub> between ecosystems and the atmosphere. Ecosystems sequester or store carbon when NEP is positive, with the length of time (**residence time**) this carbon remains in the ecosystem determined by its **turnover rate**. The turnover rate is simply the ratio of standing biomass to NPP. Biomass and NPP are mechanistically related to each other, and in general greater NPP will lead to greater standing biomass in terrestrial ecosystems. However, the relationship between NPP and biomass is actually more complex. In forests, for example, aboveground biomass plateaus at intermediate levels of aboveground NPP and may even decline at the highest levels of productivity (Fig. 2). This is because turnover rates may increase in high productivity forests limiting additional biomass accumulation. This relationship is further complicated when comparing the NPP–biomass relationship in different biome types. For example, some forests may have very high standing biomass but low NPP in part due to high respiration rates in large trees; the residence time of C stored in such a system is relatively long and turnover is slow. Conversely, most grasslands have low standing biomass due to consumption by animals or fire, even with relatively high NPP. Indeed, some wetlands have levels of NPP that can match tropical forests, but standing biomass is much lower (Table 1).

## How Is NPP Measured?

Before reviewing what is known about the patterns and controls on NPP (and related processes) in terrestrial ecosystems, it is important to appreciate the wide range of methods used to estimate NPP as well as to understand their limitations.

Such knowledge can be critical for interpreting research and making broader inferences. The accurate measurement of primary production can be very challenging despite the simple concept that it is the amount of new biomass added to the vegetation in a time interval. The principal difficulty is that not all the new biomass that was added is retained at the end of a measurement interval (whether a month, growing season, or year). In most ecosystems a significant proportion of the NPP can be lost to processes such as herbivory. Moreover, direct measurement of changes in the biomass of some tissues, like roots, is very difficult, and a substantial proportion of the belowground production is lost through a variety of rhizosphere carbon flux processes such as exudation, rhizodeposition, and allocation to mycorrhizal fungi.

### **Direct Field Measurement of Aboveground Primary Production**

Field measurement of aboveground NPP (ANPP) is usually conducted on a sample plot basis over an annual time scale. The spatial scale of plot sampling depends on the vegetation structure and its spatial variability. The methods used for ANPP differ categorically between herbaceous and woody dominated vegetation because of the need to quantify woody biomass increment. Some of the field methods for these two categories of terrestrial vegetation are described below.

Estimating ANPP in ecosystems dominated by herbaceous vegetation (e.g., grasslands) is relatively simple compared with those with a substantial woody component (forests, shrublands). Harvest methods are employed in which the aboveground biomass is clipped from a quadrat of a specified size using scissors, separated into components (e.g., species or functional group), dried to constant mass, and weighed. However, harvest methods must account for plant senescence, and the key to success is accurately partitioning the clipped biomass into three pools: green (live) biomass, standing dead produced this year, and any older dead biomass. The frequency of sampling for harvest methods must be adjusted depending on the dynamics of the vegetation. For example, in ecosystems with a short growing season, one clipping at the time of peak standing biomass can provide an accurate estimate of ANPP. In contrast, if the phenology of the dominant species is distinct (e.g., a mix of cool season and warm season floras), then two or more harvests per season may be required and positive differences in green biomass are summed. Also, if production and subsequent senescence and decomposition of plant biomass is substantial during the measurement interval, then the dynamics of all three clipped pools must be measured to account for the turnover of biomass. Thus, in grasslands with long growing seasons, sequential changes in the mass of living and dead pools, as well as losses due to decomposition, must be summed to obtain reliable ANPP estimates.

An additional difficult challenge in many herbaceous communities is accounting for losses due to herbivory. Although it might seem that measuring ANPP in ungrazed exclosures would solve this problem, plants exhibit many compensatory responses to herbivory so that ANPP measurement in the absence of herbivory is

not accurate (see “[Biotic Controls on NPP](#)”). A common solution to this problem is using many temporary, movable enclosures that allow estimation of herbivore consumption and regrowth responses of grazed plants.

Field measurement of ANPP in ecosystems dominated by woody vegetation presents challenges owing to the large and complex dimensions of the plants. The principle underlying field approaches is that

$$\text{ANPP} = \Delta\text{B} + \text{M}$$

where  $\Delta\text{B}$  equals the annual increment in live tree biomass and  $\text{M}$  equals the losses of living tissue to mortality, including litterfall, pruning/herbivory, and tree death. The reason that  $\text{M}$  must be added to  $\Delta\text{B}$  to estimate ANPP is clear for the case where  $\Delta\text{B}$  is zero: if the live biomass is constant from year to year, and losses of live biomass are occurring, then the plants must have replaced this lost biomass in the form of new tissue production. Because of the large size of the plants,  $\Delta\text{B}$  is usually estimated by applying allometric equations that describe the relationship between an easily measured dimension of the plant (e.g., stem diameter or tree height) and plant biomass. Such equations have been developed for most common woody plant species or groups. Next, the annual or multiyear change in diameter can be used to estimate  $\Delta\text{B}$  for each plant in the sample plot. The largest aboveground loss of ephemeral tissue contributing to  $\text{M}$  is fine litterfall which is easily collected using littertraps. Note that in mature woody vegetation, the amount of leaf litterfall is about the same as the new foliage production. For some other litterfall components, especially fruits and woody tissues, the amounts can vary a lot from year to year, and several years of collection will be needed to adequately account for annual variation. If  $\Delta\text{B}$  is estimated on the basis of multiyear changes in live tree biomass, then the value of  $\text{M}$  must also account for trees that died during the measurement interval; this is usually accomplished with tagged tree inventory, but the plots must be large enough to overcome the high spatial variability in tree mortality. Finally, the measurement of ANPP will be incomplete if understory vegetation is ignored and suitable adaptation of herbaceous and woody vegetation measurements may be needed.

## Field Approaches for Belowground Primary Production

A full accounting of terrestrial primary productivity requires estimation of belowground primary production (BNPP). It is helpful to begin by considering the total allocation of fixed carbon or energy to the root system of the plants (total root allocation – TRA) and the components of BNPP including growth of fine and coarse roots and rhizosphere C flux (RCF). A large proportion of TRA is used by the root system for respiratory needs ( $\text{Rr}$ ) and does not contribute to BNPP:

$$\text{BNPP} = \text{TRA} - \text{Rr}$$

Direct measurement of BNPP is not yet possible, but it can be estimated from measurements of its components or from estimates of TRA and  $\text{Rr}$ . The largest

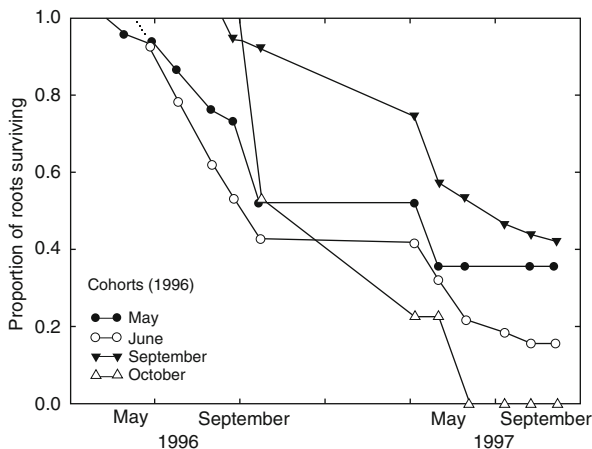
single component of BNPP is usually the growth of ephemeral fine roots, defined as smaller than some arbitrary diameter cutoff (e.g., <1 mm). These fine roots are very important functionally as well for water and nutrient uptake from the soil. A smaller fraction of BNPP goes to the growth of long-lived coarse roots. As noted earlier, a large proportion of BNPP is conveyed to mycorrhizal symbionts or exported from the roots by passive exudation or active rhizodeposition. Finally, a fraction of BNPP is lost to root herbivory, but methodological challenges have limited these measurements to just a handful of studies.

Most estimates of TRA and BNPP in natural vegetation employ a steady-state assumption for either or both soil C content or fine root biomass, meaning that the steady-state parameter is neither increasing nor decreasing substantially. Under this assumption TRA can be estimated as the difference between the annual emission of CO<sub>2</sub>-C from the soil (total soil respiration – TSR) and annual aboveground litterfall, both of which can be measured with high accuracy. Thus, reliable estimates of TRA are available for a variety of global vegetation types. However, to calculate BNPP from TRA requires accurate measurement of R<sub>r</sub> which is challenging because of the complexity of plant root systems, their highly variable metabolic activity, as well as the intimate contact between roots and soil and attendant microbes. Nevertheless, the accurate measurement of TRA has provided useful insights into patterns of BNPP in relation to biotic and environmental factors (see below).

Again, the largest component of BNPP is the growth of short-lived fine roots (FRP). The most reliable way of estimating FRP is combining field measurements of fine root biomass and indices of root turnover, i.e., the proportion of the fine root biomass dying and being replaced annually. Under the steady-state assumption, the fine root turnover coefficient (TC, year<sup>-1</sup>) is the inverse of the average root lifespan, and FRP can be calculated as the product of TC and average fine root biomass. The latter is measured by coring the soil and laboriously sorting the live roots from the soil. Several approaches have been used to estimate fine root TC, including minirhizotrons with which roots can be viewed growing along the surface of a transparent tube inserted into the soil and their survivorship monitored through time (Fig. 3). Measurements of TC based on the decay or dilution of isotopes also have been achieved, but in all cases, a variety of sources of error and bias must be overcome, as summarized by Tierney and Fahey (2007).

## **Remote-Sensing and Modeling Approaches to Terrestrial Primary Production**

Plot-scale field measurements can be very expensive for purposes of routine monitoring, and extrapolation from a few small plots to regional or global scales is challenging. For these purposes, methods have been developed that utilize remotely sensed information from earth-observing satellites combined with computational algorithms that convert satellite data to production estimates. Because the satellites provide complete coverage of the earth's surface at high frequency,



**Fig. 3** An example of fine root survivorship data from a minirhizotron tube (a transparent tube inserted into the soil) beneath a northern hardwood forest in northeastern USA. Periodically, a camera is lowered into the tube and images of roots growing along the surface of the tube are recorded. By identifying and noting the location of a number of roots at one point in time (a cohort), the survival or disappearance of these roots can be reassessed at regular intervals over time. Note in the data above that there were only a few exceptions; fine roots disappear over time with some cohorts (October 1996) experiencing 100 % mortality in less than a year. Root survivorship can be used to estimate the turnover coefficient for calculations of fine root production (adapted from Tierney and Fahey 2001)

these approaches allow both high-resolution and large-scale estimates that are particularly useful for global ecology applications. An overview of these methods also serves to reinforce some of the basic principles of primary production explained earlier.

The basic principle behind remote-sensing approaches is that indices of vegetation structure, especially leaf area index (LAI), are directly related to the photosynthetic capacity of the earth's surface. Passive sensors mounted on satellites, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument, detect solar radiation reflected from the earth's surface, and the ratio of particular wavelength bands that are differentially absorbed by foliage is quite closely related to the LAI, at least below some saturation threshold (about LAI = 4). This remotely sensed fraction of the absorbed PAR (APAR = absorbed photosynthetically active radiation) is then used to estimate maximum GPP, and light-use efficiency (LUE) or production efficiency models adjust for suboptimal environmental conditions and varying respiratory costs. In particular, the maximum conversion of APAR into GPP (i.e., LUE) varies among vegetation types because of differences in the size of plants and consequent total leaf respiration. The LUE will also be reduced by environmental factors that cause stomatal closure, especially subfreezing temperatures, dry soil, and low atmospheric humidity. The production efficiency will further depend on the respiratory costs of growing and maintaining all the other plant tissues.

To evaluate accuracy, point estimates of NPP from the remote-sensing approaches can be compared with plot-based estimates of NPP or with estimates of GPP from eddy flux towers. A comparison of MODIS-based estimates of GPP against flux towers indicated a 20–30 % overestimation by the remote-sensing methods across a range of North American biomes.

The large-scale estimates of NPP based on remote sensing can be applied to a variety of ecological and global questions. The most obvious demand for global productivity estimates is to drive and test coupled general circulation models of climate change and climate feedback mechanisms associated with terrestrial biomes. The large sample sizes available from remote sensing also can help shed light on environmental controls on NPP as spatial data on soils, topography, and climate can be compared against the remote-sensing NPP estimates. Long-term trends in the key global feedback of ecosystem C sequestration also may be identified with global NPP estimates. For example, although global warming might be expected to stimulate higher global terrestrial NPP, remote-sensing estimates for the decade 2000–2009, the warmest on record, suggest a reduction in global productivity owing primarily to large-scale, regional droughts, especially in the southern hemisphere (Zhao and Running 2010). Continued refinement of these approaches could be valuable for informing regional and global environmental policies and investments.

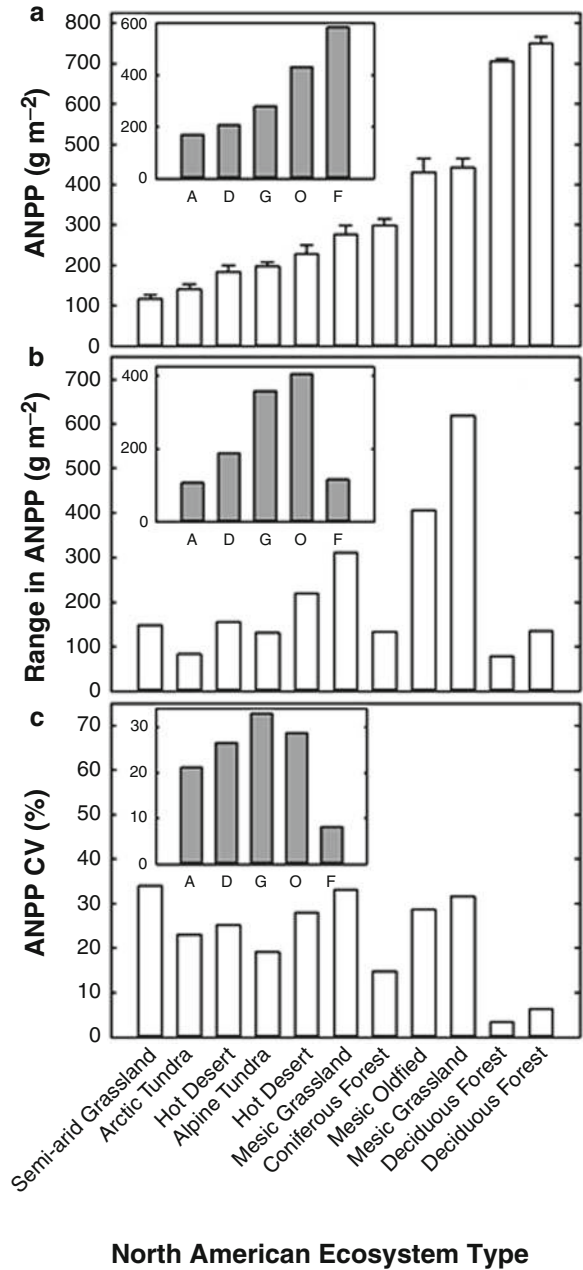
---

## Patterns and Controls on Productivity from Global to Local Scales

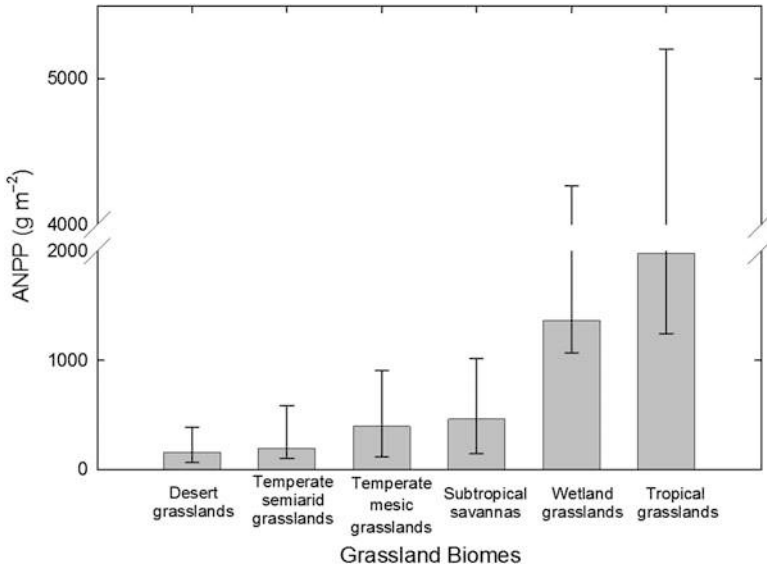
As noted earlier, NPP is the energy input that drives virtually all ecological processes in terrestrial ecosystems. But unlike nutrients which are recycled locally and water which can be recycled regionally and globally, the earth's ecosystems are energetically open, requiring inputs to be continually renewed as energy flows through ecosystems. Ecologists have long recognized that NPP, NEP, and even standing biomass can vary greatly in space and time. This variation occurs among biomes and ecosystem types (Table 1, Fig. 4) and within biomes (Fig. 5). Even at the scale of an individual site, there can be surprising variation in NPP and biomass over short distances. For example, patterns and controls of aboveground NPP have been studied for almost 30 years at the Konza Prairie Long Term Ecological Research (LTER) grassland site in Kansas (Fig. 6). Here, within what appears to be a relatively homogeneous landscape with just a few grass species dominating, aboveground NPP and biomass can vary by a factor of 4 over a distance of less than 100 m. This spatial variation is similar to the fourfold variation observed at a single point over multiple years – driven by climatic variability (wet vs. dry years). Such variation has led to a long-standing interest in understanding the factors that control rates of carbon inputs into ecosystems both across space and time. Historically this interest has been focused on the abiotic factors that best correlate with patterns of variation in NPP, particularly at large spatial scales, or through time with a focus on relationships between interannual variation in NPP and climate. Many field



**Fig. 4** (a) Comparison of average levels of ANPP across 11 sites in 5 biomes (*inset graph*) in North America. Note the sevenfold variation among sites and the fourfold variation among biomes. (b) The range in ANPP measured at each site and (c) the relative variation (coefficient of variation ( $CV$ ) = standard deviation/mean) biome based on an average of 12 years of data from each site. For insets: arctic and alpine tundra, *D* deserts, *G* grasslands, *O* old field, and *F* forests (From Knapp and Smith 2001)



experiments have followed these correlational analyses to more directly evaluate and better quantify the importance of these controls. In the next section, both large-scale correlational approaches and the learning from smaller-scale experiments will be highlighted.

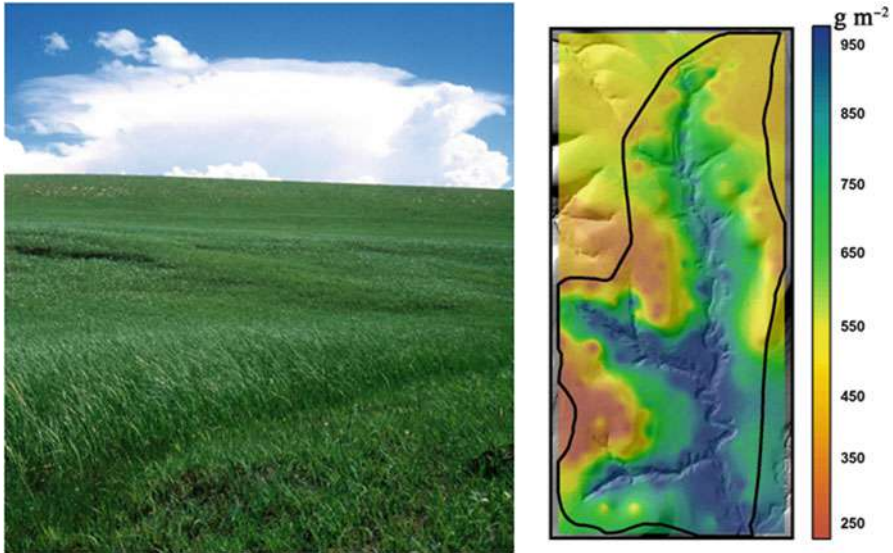


**Fig. 5** Comparison of average ANPP in different types of grasslands and grass-dominated ecosystems from around the world. The error bars depict the range of average values from different individual sites. All of these ecosystems have relatively low standing biomass (compared to forests) but even within systems that are all dominated by a single type of plant (grasses), ANPP can vary from 100 to several thousand  $\text{g}/\text{m}^2$  (Data from Knapp et al. 2007)

## Abiotic Controls on NPP

At global scales, both temperature and precipitation are positively related to patterns of NPP (Fig. 7) with the highest rates of NPP occurring under warm, moist conditions. In general, differences among locations in mean annual precipitation explain more of the global variation in NPP than differences in mean annual temperature. Indeed, recent analyses suggest that for biomes that are not dominated by trees, models with precipitation alone explain the most spatial variation in NPP at global scales. In tree-dominated biomes, precipitation amount is still the best single environmental variable for predicting NPP, but models that also include temperature explain more of the global scale variation in NPP. These two environmental factors can be combined into estimates of **actual evapotranspiration** (AET) which represents the amount of water transpired by plants and evaporated from the plant canopy and land surface. This single variable can be quite difficult to determine precisely, but estimates of spatial patterns of AET correlate quite well with patterns of NPP at a global scale (Fig. 8).

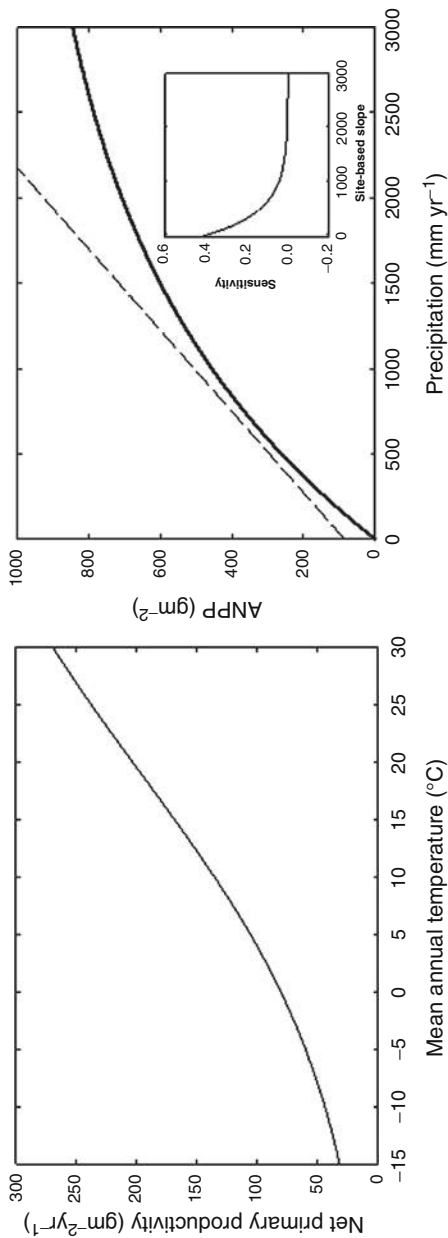
One might conclude from the preceding discussion that globally biomes can be divided into those in which NPP is limited primarily by water vs. those in which water and temperature combine to limit NPP. However, there is abundant evidence from assessments of interannual variability in NPP as well as from manipulative experiments that there are a wide range of abiotic factors that may limit NPP in



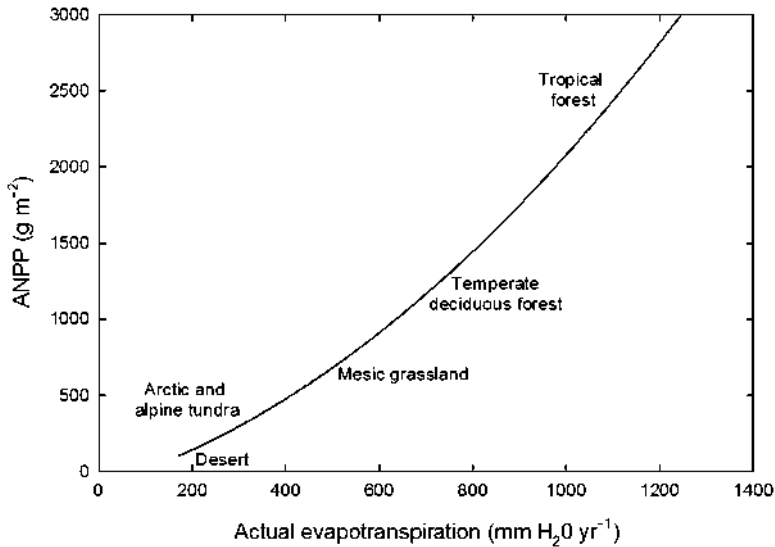
**Fig. 6** *Left:* View of an annually burned tallgrass prairie watershed at the Konza Prairie Biological Station in Kansas (Photo is taken from a lowland topographic position looking to the uplands). *Right:* Spatial variation in aboveground NPP ( $\text{g}/\text{m}^2$ ) in this watershed. Despite similar plant communities and minimal climatic variation at this scale (the watershed is about  $1 \times 0.5$  km in size), aboveground NPP varies fourfold with highest values in the lowlands near ephemeral streams (*dark blue*). Such dramatic variability can be attributed to differences in soil depth, fertility, and microclimate (From Nippert et al. 2011) (Photo credit: Melinda D. Smith)

addition to temperature and precipitation including light availability for some forests and multiple soil nutrients. In fact, there is evidence for both **co-limitation** by multiple factors and **sequential limitation** of NPP. In an experimental setting, co-limitation can be identified when NPP responds to changes in two or more factors individually (N or P for example) and/or to a greater extent when they are combined (N+P). Sequential limitation occurs when a primary limiting factor, for example, water, becomes plentiful, and then the addition of a second, previously non-limiting resource (nitrogen) increases NPP further. Recent analyses of over 600 nutrient addition experiments indicated that co-limitation of NPP occurs about 30 % of the time and sequential limitation 20 % of the time in terrestrial ecosystems (Harpole et al. 2011). If water or other factors had been included in this analysis, the frequency of multiple limiting factors would increase.

Moving down in scale to individual biomes, such as temperate deciduous forests, tropical grasslands, and arctic tundra, environmental controls on NPP can vary substantially. Ecologists often infer climatic controls on NPP at this scale by correlating year-to-year (interannual) variability in climatic parameters with the dynamics of NPP. This temporal approach is necessary in part because spatially within a biome, climatic parameters vary much less than they do at regional to global scales. Indeed, part of the definition of a biome is that it has similar climatic



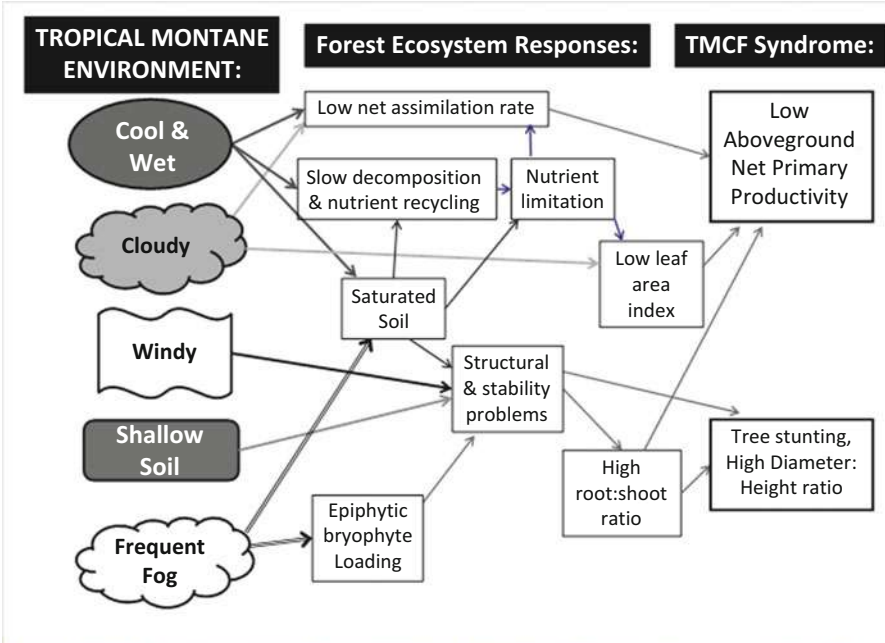
**Fig. 7** Global scale relationships between NPP and mean annual temperature (*left*) and annual precipitation (*right*). The positive NPP–temperature relationship is evident only at very large scales (Modified from Schuur 2003). Within biomes, there is often no relationship or the relationship may be negative since warmer temperatures may lead to greater plant water stress. In contrast, the positive NPP–precipitation relationship is often detected at global, regional, biome, and individual site scales. In this figure, the curvilinear relationship is from Huxman et al. (2004) based on biomes that ranged from deserts to tropical forests. The *dashed line* is the relationship developed by Sala et al. (1988) for grasslands across the central USA. *Inset*: Huxman et al. (2004) hypothesized that the sensitivity of ecosystems to changes in precipitation would depend on mean annual precipitation. Ecosystems that receive mean annual precipitation <1,000 mm are expected to be more sensitive to changes in precipitation amount than ecosystems that have greater mean annual precipitation



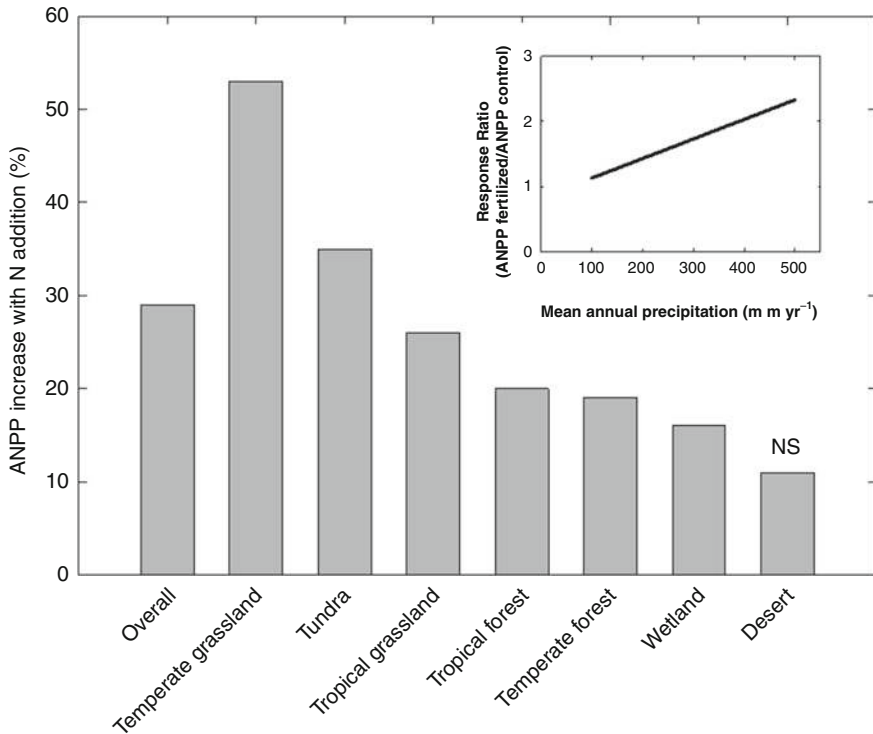
**Fig. 8** Global relationship between actual evapotranspiration (*AET*) and NPP (Modified from Rosenzweig 1968). Because *AET* combines temperature and precipitation, it is often the single best predictor of NPP at global scales

conditions throughout its range – contributing to dominance by a particular vegetation type. With this temporal approach, there is much more evidence for widespread precipitation than temperature limitation to NPP – with the degree of water limitation varying among biomes. For example, when correlations between year-to-year fluctuations in annual precipitation and variations in aboveground NPP were compiled for multiple ecosystems across many biomes, the increase in NPP per unit increase in precipitation in any given year (sensitivity in Fig. 7 inset) was, perhaps not surprisingly, much greater for drier ecosystems than wetter biomes. In the wettest of biomes, too much rainfall can become limiting to NPP through either water-logged soils leading to stressful environments for roots of higher plants (due to oxygen limitations or **anoxia**) or extended periods of cloud cover leading to light limitations to NPP. The most extreme example is in tropical cloud forests where this combination of constraints leads to a syndrome of stunting of tree size and low aboveground productivity (Fig. 9).

When a spatial approach to assessing abiotic controls is taken within a biome, soil fertility and soil depth – often referred to as **edaphic factors** – are most often identified as controlling spatial variation in NPP within many grasslands or forests. Among a multitude of potential edaphic factors, soil nitrogen availability most commonly limits terrestrial production with the degree of limitation inferred from experiments where N is added. Analyses of such experiments in multiple biomes suggest that additional N alone increases NPP in all but desert biomes (Fig. 10), the latter being so severely water limited that additional N has little effect. But even in deserts, there is evidence that in relatively wet years, N addition can increase NPP.



**Fig. 9** *Top.* A conceptual model describing the mechanisms whereby several key environmental factors combined lead to low aboveground productivity and forest stature in tropical montane cloud forests. *TMCF* tropical montane cloud forest. *Bottom:* Photo within a cloud forest in the Dominican Republic. Note the tree trunks covered with epiphytes (plants that grow on other plants but are not parasites). Most of the epiphytes are mosses (bryophytes) (Photo credit: Ruth Sherman)



**Fig. 10** Global analyses of how ANPP responds to N addition experiments conducted in terrestrial ecosystems. The positive response to fertilizer in all but desert ecosystems is interpreted as evidence for widespread N limitation to productivity in terrestrial ecosystems (Data from LeBauer and Treseder (2008)). *Inset*: Relationship between response to N addition and mean annual precipitation (*MAP*) across several desert and semiarid ecosystems. Note that in agreement with the main figure, when *MAP* is lowest, there is little response to added N but that as precipitation increases so do positive ANPP responses (Redrawn from Yahdjian et al. 2011)

Indeed in general, as water becomes more plentiful in arid ecosystems, fertilizer experiments indicate that N becomes more limiting (Fig. 10). This relationship between precipitation and N limitation eventually breaks down in some tropical systems where soil phosphorus may be more limiting than N.

Though one would expect deserts and other arid ecosystems to be water limited and biomes with higher precipitation levels to be more nutrient limited, there are other systems where the primary limitation to NPP can be surprising. In relatively cold biomes such as the arctic tundra and boreal forests, experiments have shown that low soil nutrient availability can be the primary limitation to NPP – not cold temperatures. In these systems, although temperature can exert strong effects via freezing nights and setting the overall length of the growing season, during the growth period, temperature does not directly limit plant physiological processes nearly as much as low N availability to plants – a result of slow decomposition





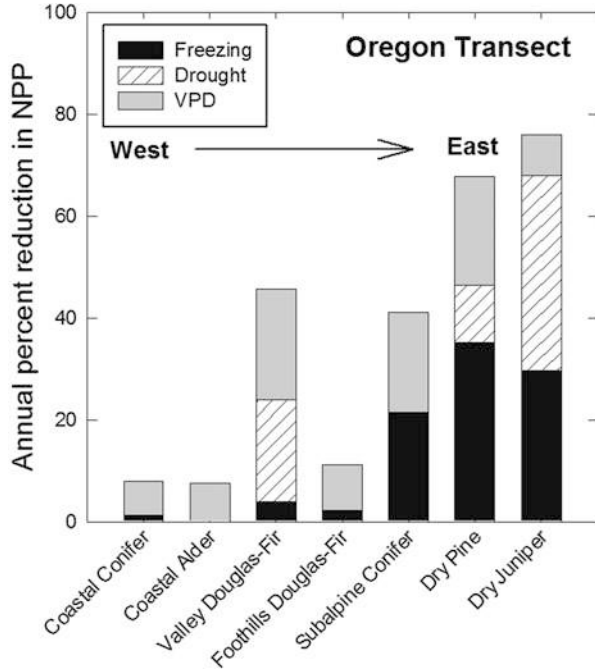
**Fig. 11** (Top) View of arctic tundra at the Toolik Lake Field Station with experimental infrastructure to assess effects of warming (*white structures*) and light availability (*black structures*) on tundra plant communities. Warming air temperatures alone by 4 °C had little effect on ANPP, but warming soil temperatures, which increased soil N or simply adding N fertilizer increased ANPP dramatically. (Bottom left) Example of plant communities outside greenhouse structures. (Bottom right) Increase in biomass and ANPP due to shrub growth inside greenhouse (Photo credit: Alan K. Knapp)

processes and severely reduced microbial activity that maintains nutrients in forms unusable for plants. Experiments have been conducted that have warmed arctic tundra plants or boreal forest trees without warming the soil, and these have demonstrated that there is little impact on plant growth. But if the soil is warmed, or if nutrients are added directly, then much greater NPP responses are measured and in the case of arctic tundra, highly productive shrubs may increase and replace the previously dominant herbaceous vegetation (Fig. 11). Moreover, warmer summer temperatures may *decrease* growth in some boreal forests as a result of increased plant water stress. This interaction between warmer temperatures and water availability is one that will be discussed later when future controls on NPP in a world of climate change are considered.

Finally, another useful approach for identifying the climatic factors that are most likely to limit NPP is to assess limitation of a single vegetation type – such as forests – along transects that capture gradients in several potentially limiting factors. One such study investigated climatic constraints on NPP in coniferous forests along a transect from coastal Oregon over the Cascade Mountains to western Oregon (Runyon et al. 1994). Across this transect an eightfold range of NPP was observed,



**Fig. 12** Annual percent reduction in total NPP of forests along a climatic transect in Oregon associated with three sources of stress: subfreezing temperatures (freezing), high vapor pressure deficit (*VPD*), and soil drought. The transect spans low-elevation humid coastal forests on the west end to high-elevation moist forests to drier forests at the eastern end in central Oregon (Redrawn from Runyon et al. 1994)



with the highest values in mature western hemlock/Douglas-fir stands on moist lower slopes of the west side of the Cascade Mountains and the lowest in juniper stands on the dry eastern slope. Recalling our introductory concept that NPP depends upon absorbed PAR and the overall efficiency of its conversion into biomass, it can be considered how climatic controls and vegetation structure influence NPP across the Oregon transect. Three principal climatic factors differentially reduce the efficiency of APAR conversion along the transect: dry soil (drought), low atmospheric humidity (high VPD, vapor pressure deficit, in Fig. 12), and cold temperatures (subfreezing). Each of these factors can cause plants to close the stomates thereby restricting photosynthetic C gain. In cool, moist coastal sites, these stresses reduced the utilization of APAR by 8–13 %, whereas in cold, dry sites high on the eastern slope, pine and juniper forests experience a 69–77 % reduction (Fig. 12). The distribution of the reduction among these three climatic stress factors also varies markedly across the transect, freezing temperatures being most important in high-elevation forests, dry soil in the juniper stands of central Oregon, and VPD at low-elevation sites in the Willamette Valley.

Climatic constraints also affect the vegetation structure across the Oregon transect; this influences NPP through effects on APAR itself. That is, the LAI of the forest vegetation ranges from less than 1 (juniper) to about 9 (hemlock/Douglasfir), reflecting in part the limited soil moisture available to supply transpiration – in drier sites a high canopy LAI would result in plant death by desiccation. Also, variation in the efficiency of conversion of APAR into NPP depends on the

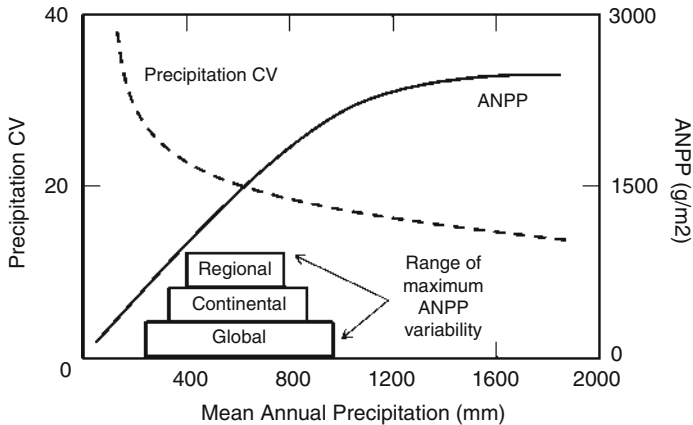
respiratory costs of growing and maintaining all the plant tissues that support the photosynthetic activity in the canopy. In sites with low soil resource availability, a higher proportion of net photosynthesis must be allocated to the roots to compete for and acquire water and mineral nutrients. As a result the relative respiratory costs are much higher on dry and infertile sites. For example, across the Oregon transect, the fraction of NPP allocated belowground ranges from about 20 % to over 60 %; high proportional allocation further constrains NPP by increasing the respiratory costs of the vegetation.

---

## Temporal Variability in NPP

As noted above, NPP can vary significantly from year to year in some types of ecosystems but much less in others (Fig. 4). This type of temporal variability (interannual), and why its magnitude varies among biomes, has interested ecologists for decades. This is because the ecological consequences for organisms adapted to NPP inputs that are more or less stable and predictable from year to year are much different from those subjected to variable and unpredictable input of NPP. By contrast, there also are temporal changes in NPP that are more directional (increasing or decreasing) through time, and this type of temporal variation has also been well studied. Historically, the mechanisms driving these two types of temporal changes in productivity have been considered independent of each other. Below patterns and determinants of interannual variability are discussed first before moving on to directional changes in NPP.

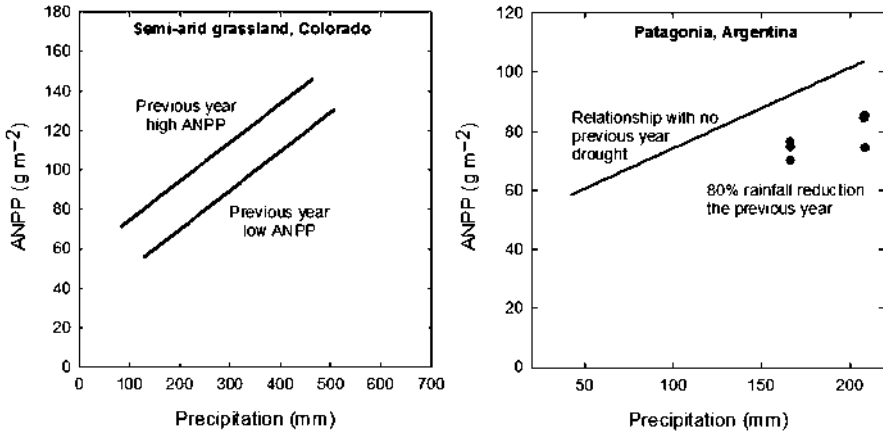
Most research on the patterns and determinants of year-to-year variability in NPP has focused on interannual variability in precipitation as the predominant driver of NPP over much of the world. In general, interannual variability in precipitation is highest where MAP is lowest – at least on a relative basis. Thus, the coefficient of variation (i.e., the relative degree of variation) of annual precipitation in deserts (30–35 %) tends to be much higher than in forests (10–15 %). This occurs because deserts are predominately dry but occasionally experience intense storms that can double the long-term average precipitation amount, while forests typically experience more stable precipitation inputs. Greater interannual variability in precipitation also is, generally, positively correlated with increased variability in ANPP as well as NEP, particularly within certain vegetation types. However, on a global basis, interannual variability in productivity peaks in semiarid to subhumid regions, where grasslands and savannas dominate. This is not where year-to-year variation in precipitation is highest. So why does variability in ANPP and NEP peak here? One explanation is that in regions where precipitation variability is highest (deserts), plants are adapted for coping with a generally water-stressed environment rather than for high potential productivity (Fig. 13). The plant communities in such chronically stressed systems are also typically sparse further limiting production per unit area. Conversely, in forests, plant communities are dense and the growth potential of plants may be higher, but interannual variability in precipitation is relatively low. This also reduces interannual variability in NPP. But in biomes with



**Fig. 13** Relationship between mean annual precipitation (*MAP*) and ANPP (*solid line*) and the coefficient of variation (*CV*) of precipitation and MAP (*dashed line*). The precipitation CV is an indication of how high year-to-year variability is for precipitation. Arid ecosystems generally have the highest CV and ecosystems with high MAP experience much less interannual variability. Although ANPP is strongly related to precipitation amount, interannual variability in ANPP is not highest where the CV of precipitation is highest. Instead, three independent analyses have concluded that the greatest variability in ANPP from year-to-year peaks between 400 and 800 mm MAP (*stacked rectangles*). Global analysis by Jung et al. 2011 (maximum ANPP variability = 250–1,000 mm), continental by Knapp and Smith 2001 (350–850 mm), regional by Paruelo et al. 1999 (450–700 mm)

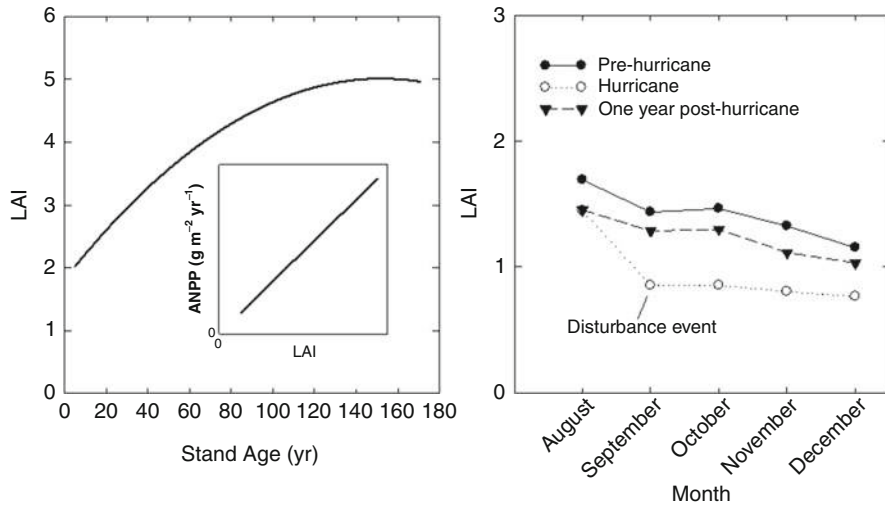
semiarid to subhumid climates, MAP is sufficient to sustain plant communities that are capable of being highly productive *and* precipitation variability is relatively high. This results in substantial variation in NPP between wet and dry years. Independent analyses at regional, continental and global scales indicate that year-to-year variability in ANPP and NEP is highest between 250 and 1000 mm MAP with a peak near 500–600 mm (Fig. 13).

From the above, it is clear that climatic conditions are strong determinants of NPP in most ecosystems, and year-to-year fluctuations can explain much of the temporal variation observed in NPP in most biomes. But there are other determinants of NPP that reflect alterations in resources or ecosystem properties that have occurred in the past but still influence NPP in the present. These are termed **legacy effects** on NPP and in some cases they can be quite substantial. In arid and semiarid ecosystems, legacy effects of past drought or periods of above-average precipitation are well documented. For example, in an analysis of the long-term relationship between ANPP and precipitation, it has been noted that this relationship differs depending on productivity levels the previous year (Fig. 14). In this case, if ANPP was high the previous year – likely due to high levels of precipitation – then ANPP would be higher than expected the next year, even with average amounts of precipitation. Similarly, after years of low ANPP – due to drought – ANPP is lower than expected. These carry-over effects may be related to changes in vegetation structure (plant density, size, and composition), physiological state of the



**Fig. 14** Evidence for climatic legacy effects of the past year's precipitation amounts on the current year's ANPP. *Left panel*: Two relationships between current year's precipitation and ANPP in semiarid shortgrass steppe grasslands of Colorado. These differ depending on the previous year's ANPP. Note that if the previous year had high levels of ANPP – likely due to high precipitation – then the current year has higher ANPP than occurs when the previous year had low ANPP and precipitation. These relationships were derived from long-term data on interannual variability in ANPP and precipitation (Redrawn from Oesterheld et al. 2001). *Right panel*: Results from an experiment in the Patagonia grassland of Argentina where rainfall inputs into plots were reduced by 80 %, and then the next year, these same plots provided average and above-average precipitation amounts and measured ANPP (filled circles). The solid line represents the general relationship between ANPP and precipitation developed for this grassland. Note that ANPP was significantly reduced the year following experimentally imposed drought compared to this general relationship (Modified from Yahdjian and Sala 2006)

plants including their ability to store carbon for growth in future years, as well as increases or decreases in stored soil moisture that may extend beyond the current year. For example, a series of wetter than average years in shrublands of the Chihuahuan Desert of southern New Mexico led to higher ANPP than would be expected based on current year's precipitation. The mechanisms proposed were that a series of wet years allowed for an increase in grass cover and abundance within shrublands, increasing productivity for a given level of rainfall (Peters et al. 2012). A series of dry years has the opposite effect with a decrease in grasses resulting and productivity declining. Results from short-term experiments have also identified the role of **meristem limitation** (Knapp and Smith 2001) in determining the magnitude of legacy effects on ANPP. Yahdjian and Sala (2006) experimentally reduced rainfall inputs by as much as 80 % into a semiarid Patagonian grassland and then measured ANPP in these previously droughted plots and compared these values to adjacent (= control) plots that had not experienced drought (Fig. 14). They found that ANPP was reduced by 20–30 % due to this drought legacy which reduced the density of plants in the ecosystem. Thus, when rainfall was returned to normal or even above-average levels in these previously droughted plots, the density of meristems (growing points in plants) was much lower than in control plots and this limited NPP.



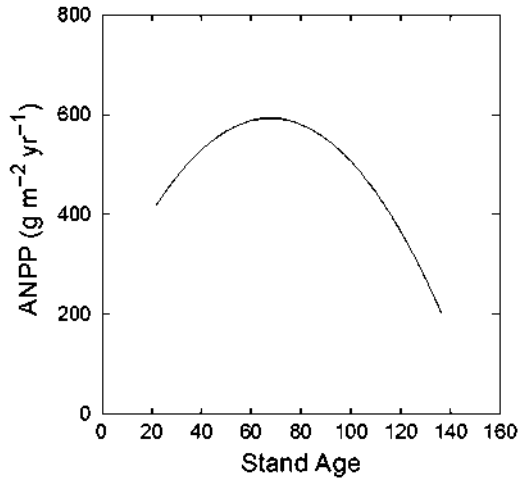
**Fig. 15** *Left.* Relationship between stand age of a temperate forest in Michigan and leaf area index (LAI, m<sup>2</sup> leaf area/m<sup>2</sup> ground area) (Modified from Hardiman et al. 2013). *Inset:* Positive relationship between LAI and ANPP in forests (Modified from Gower et al. (2001)). *Right.* Response of LAI to a hurricane in a scrub-oak forest. Three years are shown, the year before, the year of the hurricane, and the recovery year. The hurricane essentially defoliated the stand, but did not significantly damage twigs, branches, and boles of the trees. Thus, recovery of LAI was relatively rapid (Modified from Li et al. 2007)

## Disturbance and NPP

Disturbances such as fires, insect outbreaks, hurricanes, or direct human activity such as logging in forests can have both immediate- and longer-term legacy effects on NPP in ecosystems. Disturbances, as relatively discrete events that can disrupt ecosystem structure and change resources, substrate, or the physical environment, typically affect forests by disrupting structure. Fires, for example, may result in complete mortality of the overstory trees and reset stand age to zero. A hundred or more years may be required for the leaf area index (LAI) of the forest to recover to prefire levels (Fig. 15). And because LAI and NPP in forests are strongly related (Fig. 15 inset), productivity in forests as they recover is often constrained by low LAI for many years. The importance of the disruption in structure (entire trees killed or windthrown) in forests for recovery is evident when post-disturbance recovery times in which stand age is reset to zero are compared to a disturbance that only removes leaves (high winds from a hurricane) but does not disrupt forest structure (Fig. 15). Here recovery in terms of replacing lost leaf area may take only a year or two.

In many ecosystems that are disturbance dependent, a long period of time without disturbance can also constrain NPP. In mesic, productive grasslands, suppressing fire and removing grazers allows the partially decayed grass biomass from previous years to build up to levels three times greater than what is produced in any single year. This large amount of dead biomass (**detritus**) can shade the

**Fig. 16** Aboveground net primary productivity as a function of stand age in boreal *Picea abies* forests in the vicinity of Karelia, Russia. Note that productivity peaks between 60 and 80 years but by 140 years old, ANPP in this forest stand has decreased by more than 50 % (Modified from Ryan et al. 1997)

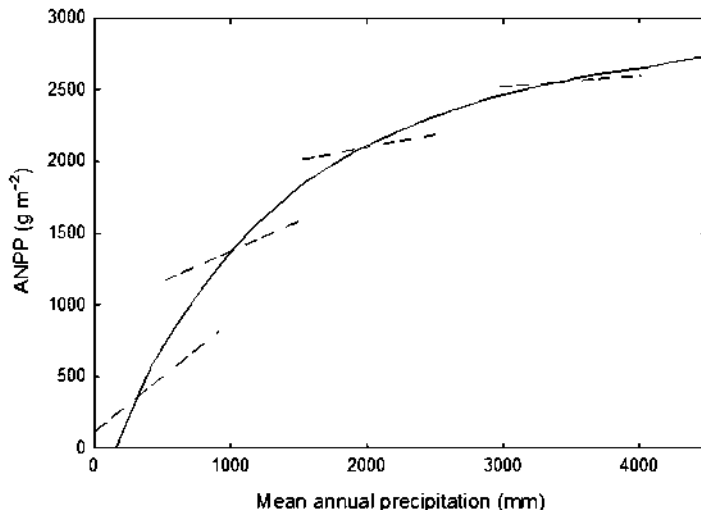


grasses as they emerge from the soil in the spring and reduce soil temperature, slowing rates of nutrient cycling. This large detritus layer can also directly tie up nutrients making them unavailable to the grasses. The net result is that ANPP is often much lower in sites where fire has been suppressed for several years compared to sites in which fire has recently occurred.

Following large-scale disturbances in forest ecosystems, the development of even-aged stands is marked by a recurring pattern of age-related change in aboveground production. For virtually all cases that have been documented, an early peak in forest ANPP is followed by a decline as forests get older (Ryan et al. 1997). Research to explain this phenomenon has been stimulated by its obvious importance to global forest carbon balance. Despite this apparently universal pattern, a parsimonious explanation does not appear to apply – different mechanisms contribute in various forest types. For example, in some cases, a shift in forest composition from faster to slower growing species contributes to the temporal pattern; however, the age-related decline is also observed in forests in the absence of any compositional change (e.g., monospecific plantations; Fig. 16). Declining production in later stages of stand development also has been associated with hydraulic constraints on water transport to the top of tall trees and consequent water stress as well as high respiratory demands in larger trees. Complex effects of canopy architecture and efficient utilization of light resources also have been cited as contributing to the age-related production decline. The implications for optimizing the provision of forest ecosystem services in a changing world continue to be explored.

## Biotic Controls on NPP

Legacy effects on NPP described previously can be explained by a combination of abiotic and biotic changes in ecosystems that influence amounts and controls on



**Fig. 17** Contrasting models of the relationship between ANPP and mean annual precipitation (MAP) based on large-scale spatial data sets (*solid line*) versus multiyear data from single sites (*dashed lines*). Note that in most cases, the slope of the relationship – or sensitivity of changes in ANPP with changes in MAP – is less when relationships are based on data from a single site through time

productivity. Biotic controls on NPP can take many forms besides disturbance or the previous year(s) being wet or dry. Below four examples of how biotic control on NPP has been of interest to ecologists for many years have been highlighted.

### **Vegetation Structure and NPP**

One of the most striking manifestation of how biotic controls – in the form of plant community composition – can influence ANPP is evident when productivity–precipitation relationships developed by combining average ANPP data from multiple ecosystems across a wide range in MAP (spatial relationship, Fig. 17) are contrasted with relationships based on data from a single site with a long-term record of temporal variations in ANPP and precipitation (temporal relationships, Fig. 17). When such temporal relationships are plotted with a spatial relationship, the slope of the temporal relationship is almost always less than the spatial relationship (Fig. 17), particularly at lower levels of precipitation. The interpretation for this pattern is that community composition shifts dramatically along with mean annual precipitation for the spatial relationship (solid line); thus, ANPP values from very different plant communities are combined to create this relationship. However, within any particular site (dashed lines), species composition is relatively constant from year to year, and thus during years with very high rainfall, these sites have lower ANPP than expected from the spatial relationship because

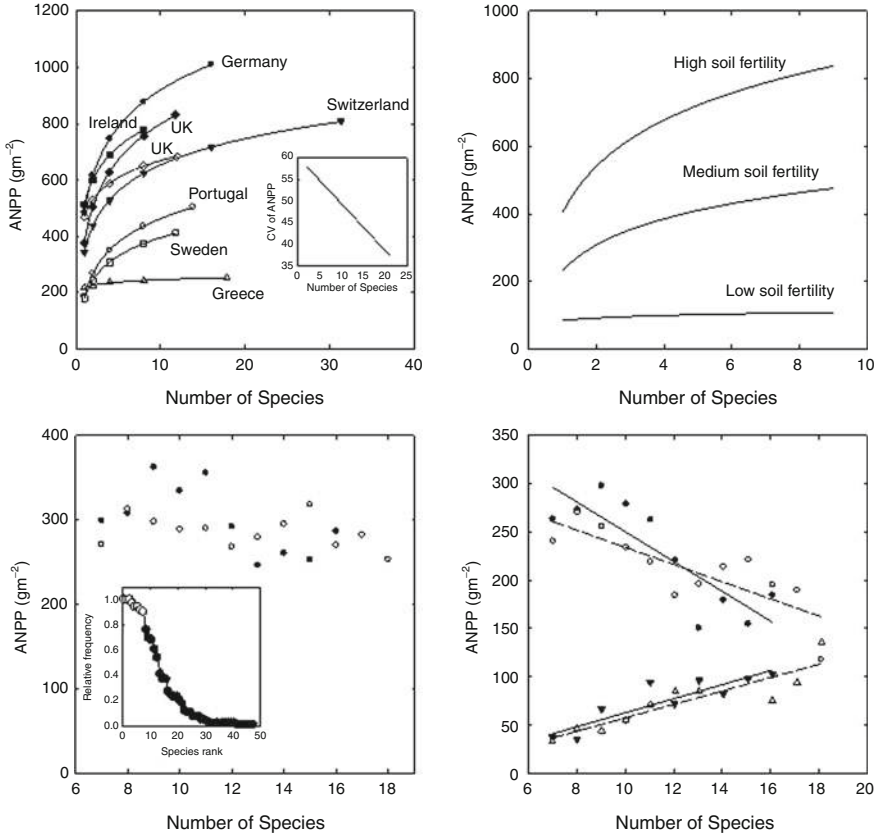
those species with high potential ANPP are not present. For example, in wet years in semiarid grasslands, where the plant community is dominated by a low density of short-statured bunch grasses, ANPP will never be as high as in a more mesic grassland with a high density of very productive grasses – even at the same level of rainfall. This has been termed a **vegetation structure constraint** on productivity (Lauenroth and Sala 1992). Interestingly, the temporal relationship between ANPP and MAP for an individual site predicts higher ANPP in dry years than does the spatial relationship. This also could be related to differences in vegetation structure influencing ANPP as well as carry-over of soil moisture from previous years (see “Legacy Effect” above). This effect would be captured within relationships at the site level but not within spatial relationships.

## Biodiversity Effects on Productivity

Biodiversity – defined conceptually as the number and variety of organisms in a specific area – has long been of interest to ecologists, particularly with regard to understanding if high versus low biodiversity has any impact on ecological pattern and process. During the past few decades, ecologists have focused on the relationship between biodiversity and NPP as concerns heighten over human activities that are leading to species loss locally and extinctions globally (see also ► [Controls of NPP and the Future](#) section below).

Because plant **species richness**, defined operationally as the number of species growing in a plot of specified size, is often highest where productivity is also high, ecologists have suspected that high species richness drives high NPP. The alternative interpretation is that both are simply responding to a third factor such as high resource availability. As a result, a number of experiments have been conducted in which plant species number is varied and productivity – usually ANPP – is measured. When plant communities are assembled randomly from a larger pool of potential species and ANPP is measured in gardens in plots with species number ranging from 1 to 20 or more species, there is a large body of compelling evidence from all over the world that increasing plant species richness increases ANPP. Of course there is evidence for the opposite as well – that reductions in the number of plant species lead to decreased ANPP (Fig. 17 top). In addition to productivity, high species numbers may stabilize ANPP so that it is less variable from year to year (Fig. 18 top). The mechanism invoked to explain these responses with increased species richness is **complementarity** which occurs when a more complete use of available resources is made possible by the co-occurrence of many different species that have varying traits and strategies (**niche partitioning**) for resource use. In the simplest sense, a plot with two plant species, one with shallow roots and one with deep roots, will produce more biomass than either alone, since more of the water in the soil profile is utilized with two species. Similarly, a plot with two species (one drought tolerant and slow growing and one that grows fast with abundant rainfall)



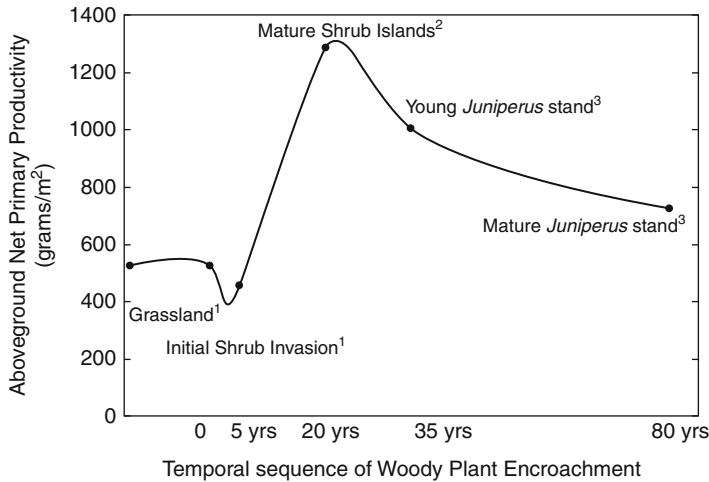


**Fig. 18** *Top: (Left)* Evidence supporting the idea that increasing biodiversity (number of species) of plant communities results in an increase in ANPP and (inset) reduced year-to-year variability in ANPP (Modified from Loreau et al. 2001). *(Right)* How soil fertility influences the biodiversity–ANPP relationship – note that when soil nutrients are very low, increasing species richness has very little effect on ANPP (Modified from Fridley 2002). In both top figures, these relationships were derived from experimental grassland plant communities assembled with different numbers of species from a larger pool of potential species. *Bottom: (Left)* Evidence for no relationship between ANPP and the number of species in plots in a natural grassland community. Open and filled circles denote data from two different years of the experiment. In this study, the number of species was varied by removing species from intact native grassland plots. Species were selected for removal based upon their abundance with the least abundant species removed first. Thus, in the inset depicting species ranked by their abundance (relative frequency of occurrence in plots ~0.8–1.0 for the most abundant or “dominant” species), only those less common to rare species (subordinate species = closed symbols) were removed. *(Right)* Contrasting responses of the ANPP of dominant (circles) versus subordinate species (triangles) to number of species (open and closed symbols denote data from two different years). Note that ANPP of the dominants increased as species number decreased. But production of the subordinate species increased with greater numbers of species (Modified from Smith and Knapp 2003)

will have more stable ANPP levels from year to year as precipitation varies compared to plots with only one of these other species. When multiple plant species are present, some also may **facilitate** the growth of others. Taller species may provide shade and protection from extreme environmental conditions for shorter species in a desert environment, for example. These and other mechanisms have been proposed as responsible for the increase in ANPP with increased plant species richness. Although this richness–ANPP relationship is relatively robust and has been demonstrated in many parts of the world, it is likely to be most important in areas with abundant resources. If resources are low, opportunities for sharing and subdividing resources among many species will also be rare – and this effect of resource availability on the richness–ANPP relationship has been experimentally demonstrated (Fig. 18 top).

The approach taken by ecologists in most experimental tests of the relationship between richness and ANPP is to construct what has been termed “synthetic communities.” These are assembled from random combinations of 1, 2, 4, 8, 16, etc. species and replicated many times in a uniform “garden” environment. This approach has strong statistical rationale, but unfortunately most of the resulting communities are quite dissimilar from natural communities. Natural communities are not random assemblages of the species that can exist in an area. Instead, they are more likely to have a few **dominant species** that make up a large portion of biomass in every plot. Indeed, if one samples hundreds of plots in a natural community, dominant species are typically found in all plots sampled. Thus, synthetic communities, in which each species has an equal chance to be present in each plot, do not reflect the structure of natural communities where some species are very common and many are less common or rare (Fig. 18 bottom). Another experimental approach to assessing the richness–ANPP relationship is to use natural communities and vary richness by removing species. In this case, species are selected nonrandomly with the least abundant species removed preferentially compared to the more abundant dominant species. This has been termed realistic species loss because it has been argued that uncommon species would have the greatest chance of disappearing from communities. In these studies, no relationship between richness and ANPP is detected (Fig. 18 bottom), and productivity of the dominant species, which are present in all plots in these experiments, may actually increase with reduction in overall community richness. Only within the subordinate species (those that are not dominant, Fig. 18 bottom) has a positive richness–ANPP relationship been observed (Fig. 18 bottom). The latter suggests that complementarity may be important within this group of species, but because the dominant species produce so much more biomass than these subordinate species, their response to richness has little impact on overall ANPP.

Thus, research suggests that species loss can reduce ANPP, particularly if a dominant species is lost or if species numbers become very low. Furthermore, mechanisms such as complementarity do operate to increase ANPP as species numbers increase, but the magnitude of this effect may be small.

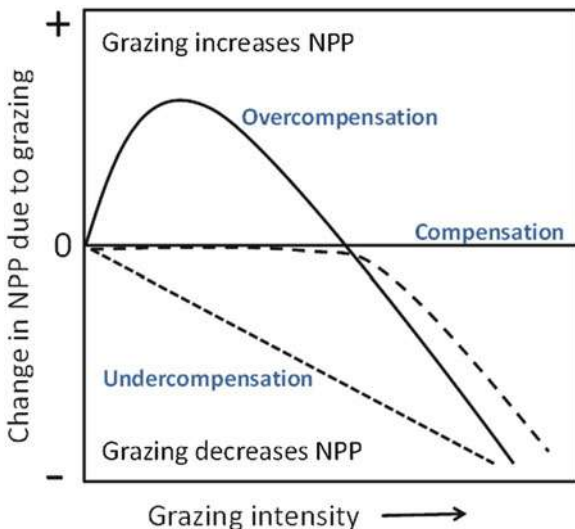


**Fig. 19** How ANPP in a central US grassland changes when shrubs and then trees replace the grassland. In this grassland with relatively high levels of precipitation (>800 mm/year), frequent fire is necessary to maintain grass dominance. If fire is suppressed in this ecosystem (time 0), grassland ANPP (dominated by *Andropogon gerardii*) initially decreases, and shrubs (typically *Cornus drummondii* or dogwood) that are typically present only as isolated and small individuals increase dramatically in abundance and cover. These eventually form dense “shrub islands” that shade the grasses and eliminate them. This shrub island stage can be the very productive (threefold higher than the grassland), but eventually even taller woody plants (*Juniperus virginiana* or eastern red cedar) displace the shrubs and a forest develops with ANPP ~50 % higher than the original grassland (Data are from Heisler et al. (2004), Lett et al. (2004), and Norris et al. (2001))

## Community Change and NPP

Although a plant community that loses or gains species over time is one type of community change that may affect NPP, more dramatic shifts in plant community structure, such as one type of community being replaced by another, might be expected to have much greater impacts on NPP. A striking example of this is occurring globally where woody plants are increasing in abundance in sites that were formerly grasslands, and in some cases, shrubs or forest are completely displacing grassland communities. Interestingly, in drier regions, shrubs displacing former grass-dominated communities may slightly decrease NPP, but in more mesic ecosystems, shrub and forest encroachment into grasslands can dramatically increase productivity – by as much as threefold (Fig. 19).

An important consequence of this increase in NPP with a shift from dominance by herbaceous plants (grasses) to woody plants (shrubs and trees) is that biomass allocation and storage in the ecosystem also shifts from belowground to aboveground. This renders the C in these systems more vulnerable to fire and subsequent release back to the atmosphere as CO<sub>2</sub> (Fig. 1).



**Fig. 20** The relationship between grazing intensity and NPP expressed relative to NPP in the absence of grazing. Shown are four potential relationships. Grazing at any intensity may decrease NPP indicating that the plant community is unable to replace the tissue lost (undercompensation). Communities may be able replace tissue lost at low levels of grazing (compensation) but not at high intensities. Or compensation may occur at all grazing intensities. Finally, there may be low to moderate levels of grazing where NPP is higher than in areas not grazed (overcompensation) (Modified from Hilbert et al. 1981)

## Herbivory and NPP

Most terrestrial plants must cope with some level of biomass loss due to consumption by animals (herbivory) that range in size from elephants and giraffe to mice and voles to insects and nematodes. By removing tissues (leaves and roots), sugars, and nutrient-rich compounds and thus reducing the valuable functions they serve, herbivory often has a negative effect on NPP. Defoliation of plants during insect outbreaks, for example, can result in substantial reductions in biomass produced. However, plants respond to herbivory in a number of ways, and in the case of large grazing animals in herds, the environment may also be changed by the activities of these large herbivores. Combined, these responses may allow plants to grow more rapidly after herbivory, and the resulting increase in productivity may result in complete compensation of biomass lost. Both theory and empirical evidence suggest that in some cases, overcompensation occurs such that NPP is actually *increased* by grazing by animals (Fig. 20).

The evidence for overcompensation of NPP comes mostly from grasslands where the dominant plants (grasses) and the herbivores (large grazers such as wildebeest in Africa or bison in North America) are known to have a long coevolutionary history. Working in the Serengeti in East Africa, McNaughton and colleagues identified a number of mechanisms by which grazing of the

Serengeti grasslands could result in higher NPP. He classified these mechanisms as intrinsic and extrinsic. Intrinsic mechanisms included changes in plants after they were grazed that increased their growth. For example, younger tissues have higher photosynthetic rates than older tissues and since grazing removed most old tissue, plants may grow faster. Removing this old tissue can also increase light available for young leaves. In addition, grazers reduce the transpiring leaf area of plants and overall plant water loss, as well as increase the root to shoot ratio of plants. These responses can decrease the degree of water stress the remaining tissues experience, allowing them to grow faster particularly during periods of limited water availability. Extrinsic factors included increased levels of soil water due to the total leaf area of the grass canopy being reduced and thus whole ecosystem transpiration being reduced. Additionally, nutrients tied up in plant tissues can be slow to be recycled, but consumption by animals allows for more rapid recycling of nutrients and thus grazing may increase nutrient availability – contributing to increased growth rates after grazing. These and other mechanisms, when combined, have the potential to both compensate for biomass lost to grazers and in some cases overcompensate.

---

## Belowground Productivity: Patterns and Controls

Plants transport a portion of net photosynthesis belowground to supply root systems that provide anchorage and acquire soil resources. Of this total root allocation (TRA), a large proportion – about two-thirds – is used for the respiratory needs of growing and maintaining the roots. The remainder is classified as BNPP and includes three principal components: (1) growth of ephemeral fine roots, analogous to foliar ANPP; (2) growth of long-lived coarse roots, including woody roots and belowground storage organs such as tubers; and (3) rhizosphere C flux (RCF) which includes diverse processes like sloughing of root cap cells, active secretion of mucilage, passive exudation, and allocation to mycorrhizal symbionts. The complexity of all these BNPP components, and the difficulty of measuring them, has limited our understanding of patterns and controls of BNPP, but recent advances have provided at least a partial picture.

Two general principles form the basis for understanding BNPP. The first is obvious – the higher the total NPP for an ecosystem, the higher the BNPP. The second principle has been designated the **functional equilibrium hypothesis** and posits that a stable ratio of resource acquisition by shoots and roots is maintained in the face of constraints to resource acquisition, so that one organ does not greatly outgrow the other and overall plant performance is optimized. Thus, the root to shoot production ratio is expected to be higher in dry or infertile soils. Experimental and survey research generally supports the functional equilibrium hypothesis. A summary of observations for several terrestrial biomes indicates that the fraction of NPP comprising BNPP differs considerably among biomes, with the highest BNPP:NPP ratio occurring in grasslands and distinctly lower values in forests (Table 2). Intense competition for limited soil moisture stimulates higher belowground allocation in grasslands. However, it is also notable that the turnover coefficient

**Table 2** Belowground primary production in four terrestrial biomes and its relation with total net primary production. Units are  $\text{g/m}^2/\text{year}$  (Adapted from Tierney and Fahey 2007)

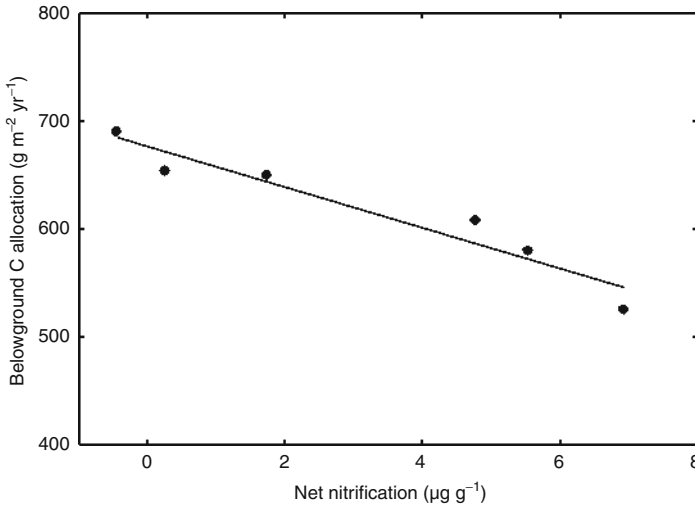
Biome	Mean BNPP	Mean total NPP	BNPP:NPP
<b>Grassland</b>	498	1,032	0.52
<b>Boreal evergreen forest</b>	312	774	0.4
<b>Temperate deciduous forest</b>	380	1,470	0.26
<b>Temperate evergreen forest</b>	426	1,772	0.24

**Table 3** Soil respiration, aboveground litter fall and belowground carbon allocation in three global forest biomes. Units are  $\text{g/m}^2/\text{year}$  (Adapted from Davidson et al. 2002)

Forest biome	Soil respiration	Litterfall	Belowground carbon allocation	Litterfall: BCA
<b>Tropical evergreen</b>	1,603	410	1,193	0.34
<b>Temperate deciduous</b>	840	186	654	0.28
<b>Temperate evergreen</b>	809	188	621	0.3

( $\text{TC}, \text{year}^{-1}$ ) of fine roots in grasslands is higher than in forests so that more rapid replacement of fine roots contributes significantly to the higher BNPP. The causes of the higher TC in grasslands are not known, but TC also seems to increase in warmer and more productive grassland environments. In general, the lifespan of fine roots (inverse of TC) probably decreases as a result of higher metabolic activity, nutrient uptake rates, and herbivory in warmer, more fertile soils, thereby contributing to higher fine root production.

As noted earlier, accurate measurement of BNPP is difficult; however, insights into the process have been provided by the straightforward measurement of TRA as the difference between total soil respiration (TSR) and aboveground litterfall flux of C. On average in the world's forests, annual TSR is about three times greater than aboveground litterfall, and hence, TRA is about twice as large as aboveground litterfall. A synthesis for temperate forests also indicates that TRA is about three times greater than BNPP, implying that about two-thirds of TRA is used in root respiration (Table 3). Measurements of TRA also provide a basis for evaluating BNPP responses to varying soil resource availability. In many temperate zone ecosystems, nitrogen is the most growth-limiting soil nutrient, and the availability of N has increased markedly as a result of anthropogenic activity, with likely consequences for NPP. Increasing soil N availability might be expected to cause a decrease in fine root biomass, but at the same time, it could stimulate higher fine root turnover. The functional equilibrium hypothesis would argue for a reduction in proportional TRA in more fertile soils, and some evidence supports this conjecture (Fig. 21). A further complication is that a large proportion of BNPP goes to RCF. For example, Jones et al. (2009) summarized available evidence to estimate that an average of 27 % of TRA goes to RCF, equivalent to 11 % of net photosynthesis. These observations emphasize the importance of RCF in facilitating soil resource



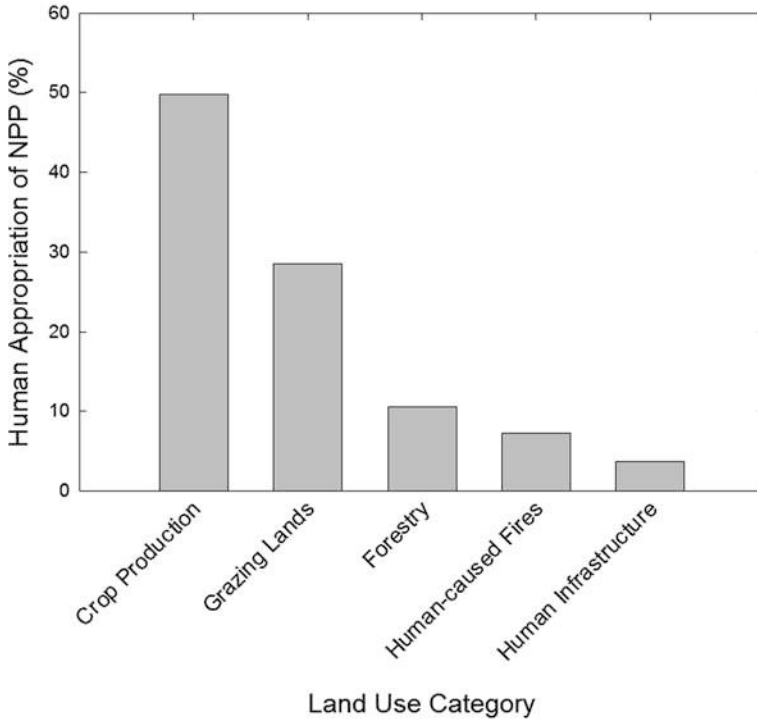
**Fig. 21** Relationship between belowground carbon allocation and soil nitrogen availability (as indicated by net nitrification) in six northern hardwood forests in New Hampshire. Note that as soil fertility increases, estimated carbon allocated belowground decreases (Kikang Bae, unpublished data)

acquisition. The remaining challenge is to understand the proportion of these large fluxes that goes to different RCF pathways and how these may differ depending upon biotic and environmental factors.

## Controls of NPP and the Future

It is estimated that humans are utilizing or otherwise altering almost one-quarter of terrestrial NPP – an amazing amount for a single species (Fig. 22, Haberl et al. 2007). In addition, human population growth is correlated with global changes in atmospheric  $\text{CO}_2$  levels, increased N deposition, and warming temperatures. Because all of these global changes – and many others such as more frequent extreme climatic events and altered disturbance regimes – will impact to some degree the remaining NPP not *directly* influenced by humans, it is safe to say that human activities will be a primary controller of NPP globally in the future. The ways in which increased N deposition and disturbance regimes can influence NPP were discussed above. Here the focus is on the potential impacts of warming temperatures and increased  $\text{CO}_2$ .

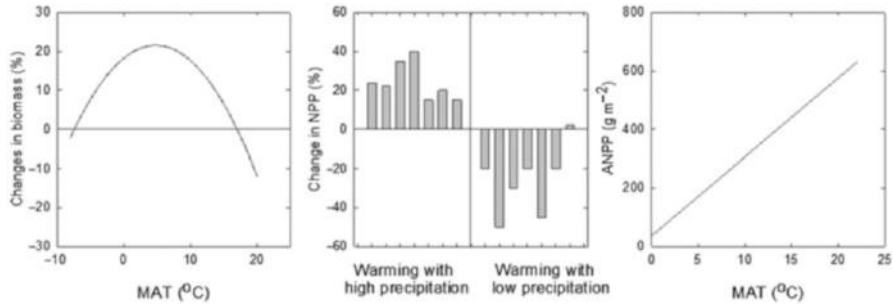
Of all the climatic changes forecast for the future by the Intergovernmental Panel on Climate Change (IPCC), warming of the atmosphere has the highest degree of confidence. Indeed most climate scientists argue that human-caused global warming has already occurred. Thus, a very active area of research today involves studies of warming effects on NPP and related components, with many recent and



**Fig. 22** Human appropriation of NPP in terrestrial ecosystems globally divided by land-use category. Of the almost 25 % of NPP utilized directly or otherwise altered by human activities, almost 80 % is accounted for by food production (crop production and grazing) (Data from Haberl et al. 2007)

ongoing experiments involving heating portions of ecosystems and comparing responses to plots with ambient temperatures. A **meta-analysis** (analysis of numerous independent but similar experiments to assess overall statistical significance) of warming experiments completed in ecosystems ranging from the arctic tundra to the tropics found that warming had little measureable impact on total biomass, belowground biomass, ANPP, or NEE (Wu et al. 2011). However, positive effects of warming were detected on aboveground biomass, NPP, BNPP, and ecosystem respiration (ER). Although the overall effects of warming on NPP and biomass were positive, it is more instructive to assess how responses vary and what determines this variation. In another analysis in which results from numerous warming experiments were combined, responses in ANPP (estimated by changes in biomass) to warming depended strongly on the mean annual temperature (MAT) of the ecosystem studied (Fig. 23). Warming has little impact on very cold ecosystems and negative impacts on very warm ecosystems. But positive warming effects peaked where MAT was between 3 °C and 10 °C indicating that some ecosystem types will be more sensitive than others to warming. The negative impact of warming in areas where MAT > 15 °C may have several explanations.

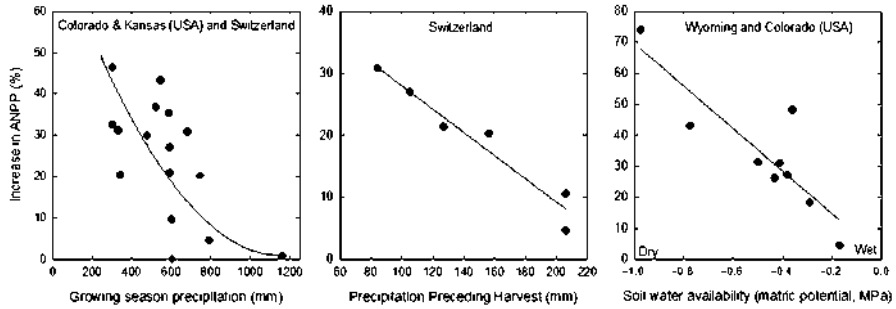




**Fig. 23** (Left) Interaction between mean annual temperature (MAT) and response of ANPP to experimental warming. This was derived by combining results from 127 different warming experiments (warming magnitude ranged from 1 °C to 5 °C) across sites that varied widely in MAT (Modified from Lin et al. 2010). (Center) Response of NPP to warming under conditions of either high or low precipitation modeled for 7 ecosystems ranging from annual grassland to tropical and boreal forests (Modified from Luo et al. 2008). (Right) Increase in ANPP with increasing MAT for North America coastal marshlands dominated by *Spartina alterniflora*. Note the strong effects of increased temperature in an ecosystem that always has abundant water (Modified from Kirwan et al. 2009)

For example, at high temperatures, respiration may increase more than photosynthesis to warming leading to reduced NPP. In addition, unless precipitation inputs are very high, ecosystems with high MAT are likely to experience substantial water stress (due to high evapotranspiration) and considerable evidence indicates that warming effects on NPP can be strongly influenced by water availability. In studies in which both precipitation and warming are varied, warming effects are positive with high precipitation but negative with low precipitation (Fig. 23). The latter response is interpreted as evidence that increased water stress in plants caused by warming has a much stronger negative effect than any positive effects of warming. The importance of this interaction between temperature and water availability can be further demonstrated by assessing temperature effects on NPP in wetland ecosystems where water is never limiting. Recall that for most biomes, interannual variability in temperature usually does not correlate with year-to-year variation in NPP. This is in sharp contrast to precipitation where wet vs. dry years results in high and low ANPP, respectively (see “Abiotic Controls on NPP” section above). The exception to this pattern is for wetlands where there can be a strong temperature response by ANPP over time (warmer years have higher ANPP) and space (wetlands with higher MAT at lower latitudes are more productive, Fig. 23).

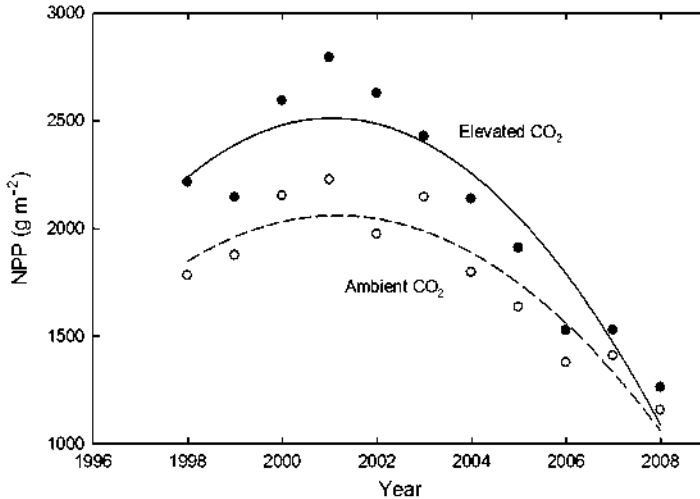
The modifying effect of water availability on NPP responses to warming is also evident when assessing NPP responses to increased CO<sub>2</sub> in the atmosphere. As noted earlier, evidence is overwhelming that many climatic changes (including warming) can be attributed to the 25 % increase in atmospheric CO<sub>2</sub> measured in the last 50 years with even greater climate change forecast for the next 100 years. Because CO<sub>2</sub> is, of course, essential for photosynthesis and NPP, ecologists have a long history of assessing the impacts of increased CO<sub>2</sub> on key ecological processes such as NPP. In general, experiments that have increased CO<sub>2</sub> to individual plants,



**Fig. 24** Examples of how the effect of increased levels of atmospheric  $\text{CO}_2$  on ANPP depends on water availability. (*Left*) Relationship between the increase in ANPP due to increased  $\text{CO}_2$  (calculated as the proportional increase in NPP in ecosystems with  $\sim 600\text{--}700$  ppm  $\text{CO}_2$  relative to ambient levels) and growing season rainfall. This relationship was developed from the results of independent field experiments in grasslands in Colorado, Kansas, and Switzerland. Note that in grasslands with low rainfall during the growing season, elevated  $\text{CO}_2$  increases ANPP by up to 50%. But this enhancement is small in grasslands that are very wet during the summer. (*Middle*) Even in a single grassland with relatively high rainfall (Switzerland), the amount of rain that fell the preceding 6 weeks prior to harvest of biomass can have a strong impact on how much elevated  $\text{CO}_2$  increases ANPP. (*Right*) Semiarid grasslands in Wyoming and Colorado show an even greater sensitivity of  $\text{CO}_2$  responses to soil moisture levels (*Left* and *middle panels* modified from Morgan et al. 2004; *Right panel* modified from Morgan et al. 2011)

and entire ecosystems, report that elevated  $\text{CO}_2$  increases NPP. This can be due to the direct stimulation of photosynthesis by the greater availability of  $\text{CO}_2$  (a  **$\text{CO}_2$  fertilization mechanism**) as well as because stomatal opening in plants is almost always reduced when  $\text{CO}_2$  is increased; this reduces transpiration and improves plant water status which can also increase photosynthesis, growth, and NPP (a **water conservation mechanism**). Field studies in grasslands clearly demonstrate the interaction between water and the effect of increased  $\text{CO}_2$  (Fig. 24). Grasslands dominated by  $\text{C}_4$  species show this interaction with water most strongly (Wyoming and Colorado grasslands in Fig. 24) since the  $\text{C}_4$  pathway is generally not subject to  $\text{CO}_2$  fertilization ( $\text{CO}_2$  concentrations are very high in the bundle sheath cells inside leaves) but stomatal sensitivity to  $\text{CO}_2$  is still evident. In this case,  $\text{CO}_2$  stimulation of ANPP only occurs under dry conditions (when water conservation matters) and not when moisture is plentiful. However, even in  $\text{C}_3$ -dominated grasslands (Swiss grasslands in Fig. 24) where direct  $\text{CO}_2$  fertilization can increase NPP, the positive effects on NPP of water conservation at high  $\text{CO}_2$  also are evident.

The high costs of conducting global change experiments, particularly those that alter  $\text{CO}_2$ , temperature, or precipitation, have resulted in a preponderance of single-factor studies. This is unfortunate because NPP in the future will be determined by multiple global change drivers impacting ecosystems concurrently. Experiments that alter single factors can certainly provide mechanistic insight for how NPP might respond to a change in a global change driver (see Figs. 10, 14, 18, 23, 24), and they can be quite valuable for parameterizing simulation models. They can also



**Fig. 25** Response of NPP of five sweetgum (*Liquidambar styraciflua*) forest stands in Tennessee grown under two CO<sub>2</sub> levels. Open circles display mean values of stands grown under ambient CO<sub>2</sub> levels (~380 ppm), while the dark circles show mean biomass production under elevated levels of CO<sub>2</sub> (~550 ppm). Note that the stimulation of NPP that occurred initially diminished over time. Despite projections of higher CO<sub>2</sub> levels in the next century due to climate change, forest NPP may be constrained by other limiting factors, such as nitrogen availability (Modified from Norby et al. 2010)

identify where co-limitation and sequential limitation (see “[Abiotic Controls on NPP](#)”) may emerge and limit NPP in the future. For example, when forest stands are exposed to elevated CO<sub>2</sub> for multiple years, NPP is stimulated initially; but in some forests, this increase in NPP is not maintained and diminishes to zero over time (Fig. 25). In this example from a deciduous forest at the Oak Ridge National Laboratory FACE (free-air CO<sub>2</sub> enrichment) site, most of the initial stimulation of NPP was associated with increased root production. But eventually N limitation of NPP constrained the response to high CO<sub>2</sub>. Thus, controls on forest NPP in the future and responses to global change will depend on which global change drivers impact any given forest. In this example, responses of forest NPP to increasing CO<sub>2</sub> (which occurs relatively uniformly globally) in forests with vs. without increased N deposition (which is a much more local and regionally variable driver) will be very different.

The cumulative and interactive effects of global environmental changes – CO<sub>2</sub>, climate, nitrogen deposition, biodiversity loss, and altered disturbance regimes – on NPP will remain a holy grail of global ecosystem biology for some time. The coincident effects of these and other global change drivers are likely to interact in complex and nonintuitive ways to influence the NPP responses of terrestrial ecosystems in the future. In lieu of the resources being available to conduct multifactor global change experiments in a variety of biomes and ecosystem types, ecosystem simulation models offer ecologists their best opportunity to explore these complex interactions and forecast how NPP patterns and controls will change in the future.

## Future Directions

- Better quantification of the proportion of total NPP that goes to belowground NPP, in particular rhizosphere carbon flux including exudation, rhizodeposition, and allocation to mycorrhizal fungi and other symbionts.
- Determination of how increasing human activities will influence disturbance regimes, land-use patterns, and vegetation structure, pattern, and composition (e.g., through the introduction of exotic species) and consequently affect NPP.
- Increased understanding is needed regarding how global environmental changes such as warming temperatures, changing precipitation regimes, increased atmospheric CO<sub>2</sub>, and greater rates of nitrogen deposition will influence global NPP in the future. Much is known of how many of these will affect NPP in individual ecosystems from single-factor experiments, but their combined effects across multiple ecosystems are highly uncertain.

---

## References

- Davidson EA, Savage K, Bolstad P, Clark DA, Curtis PS, Ellsworth DS, Hanson PJ, Law BE, Luo Y, Pregitzer KS, Randolph JC, Zak D. Belowground carbon allocation in forests estimated from litterfall and IRGA-based soil respiration measurements. *Agr Forest Meteorol.* 2002;113:39–51.
- Fridley JD. Resource availability dominates and alters the relationship between species diversity and ecosystem productivity in experimental plant communities. *Oecologia.* 2002;132:271–7.
- Gower ST, Krankina O, Olson RJ, Apps M, Linder S, Wang C. Net primary production and carbon allocation patterns of boreal forest ecosystems. *Ecol Appl.* 2001;11:1395–411.
- Haberl H, Erb KH, Krausmann F, Gaube V, Bondeau A, Plutzer C, Gingrich S, Lucht W, Fischer-Kowalski M. Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proc Natl Acad Sci USA.* 2007;104:12942–7.
- Hardiman BS, Gough CM, Halperin A, Hofmeister KL, Nave LE, Bohrer G, Curtis PS. Maintaining high rates of carbon storage in old forests: a mechanism linking canopy structure to forest function. *For Ecol Manage.* 2013;298:111–9.
- Harpole WS, Ngai JT, Cleland EE, Seabloom EW, Borer ET, Bracken MES, Elser JJ, Gruner DS, Hillebrand H, Shurin JB, Smith JE. Nutrient co-limitation of primary producer communities. *Ecol Lett.* 2011;14:852–62.
- Heisler JL, Briggs JM, Knapp AK, Blair JM, Seery A. Direct and indirect effects of fire on shrub density and aboveground productivity in a mesic grassland. *Ecology.* 2004;85:2245–57.
- Hilbert DW, Swift DM, Detling JK, Dyer MI. Relative growth rates and the grazing optimization hypothesis. *Oecologia.* 1981;51:14–8.
- Huston MA, Wolverton S. The global distribution of net primary production: resolving the paradox. *Ecol Monogr.* 2009;79:343–77.
- Huxman TE, Smith MD, Fay PA, Knapp AK, Shaw MR, Loik ME, Smith SD, Tissue DT, Zak JC, Weltzin JF, Pockman WT, Sala OE, Haddad BM, Harte J, Koch GW, Schwinning S, Small EE, Williams DG. Convergence across biomes to a common rain-use efficiency. *Nature.* 2004;429:651–4.
- Jones DL, Nguyen C, Finlay RD. Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant and Soil.* 2009;321:5–33.

- Jung M, Reichstein M, Margolis HA, Cescatti A, Richardson AD, Arain MA, Arneeth A, Bernhofer C, Bonal D, Chen J, Gianelle D, Gobron N, Kiely G, Kutsch W, Lasslop G, Law BE, Lindroth A, Merbold L, Montagnani L, Moors EJ, Papale D, Sottocornola M, Vaccari F, Williams C. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J Geophys Res.* 2011;116:G00J07.
- Keeling HC, Phillips OL. The global relationship between forest productivity and biomass. *Glob Ecol Biogeogr.* 2007;16:618–31.
- Kirwan ML, Guntenspergen GR, Morris JT. Latitudinal trends in *Spartina alterniflora* productivity and the response of coastal marshes to global change. *Glob Chang Biol.* 2009;15:1982–9.
- Knapp AK, Smith MD. Variation among biomes in temporal dynamics of aboveground primary production. *Science.* 2001;291:481–4.
- Knapp AK, Briggs JM, Childers DL, Sala OE. Estimating aboveground net primary production in grassland and herbaceous dominated ecosystems. In: Fahey TJ, Knapp AK, editors. *Principles and standards for measuring net primary production.* New York: Oxford University Press; 2007. p. 27–48.
- Lauenroth WK, Sala OE. Long-term forage production of North American shortgrass steppe. *Ecol Appl.* 1992;2:397–403.
- LeBauer DS, Treseder KK. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology.* 2008;89:371–9.
- Lett MS, Knapp AK, Briggs JM, Blair JM. Influence of shrub encroachment on plant productivity and carbon and nitrogen pools in a mesic grassland. *Can J Bot.* 2004;82:1363–70.
- Li J, Powell TL, Seiler TJ, Johnson DP, Anderson HP, Bracho R, Hungate BA, Hinkle CR, Drake BG. Impacts of Hurricane Frances on Florida scrub-oak ecosystem processes: defoliation, net CO<sub>2</sub> exchange and interactions with elevation CO<sub>2</sub>. *Glob Chang Biol.* 2007;13:1101–13.
- Lin D, Xia J, Wan S. Climate warming and biomass accumulation of terrestrial plants: a meta-analysis. *New Phytol.* 2010;188:187–98.
- Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, Hooper DU, Huston MA, Raffaelli D, Schmid B, Tilman D, Wardle DA. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science.* 2001;294:804–8.
- Luo Y, Gerten D, Le Maire G, Parton WJ, Weng E, Zhou X, Keough C, Beier C, Ciais P, Cramer W, Dukes JS, Emmett B, Hanson PJ, Knapp A, Linder S, Nepstad D, Rustad L. Modeled interactive effects of precipitation, temperature, and CO<sub>2</sub> on ecosystem carbon and water dynamics in different climatic zones. *Glob Chang Biol.* 2008;14:1986–99.
- Morgan JA, Pataki DE, Körner C, Clark H, Del Grosso SJ, Grünzweig JM, Knapp AK, Mosier AR, Newton PCD, Niklaus PA, Nippert JB, Nowak RS, Parton WJ, Polley HW, Shaw MR. Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO<sub>2</sub>. *Oecologia.* 2004;140:11–25.
- Morgan JA, LeCain DR, Pendall E, Blumenthal DM, Kimball BA, Carrillo Y, Williams DG, Heisler-White J, Dijkstra FA, West M. C<sub>4</sub> grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland. *Nature.* 2011;476:202–5.
- Nippert JB, Ocheltree TW, Skibbe AM, Kangas LC, Ham JM, Shonkwiler Arnold KB, Brunsell NA. Linking plant growth responses across topographic gradients in tallgrass prairie. *Oecologia.* 2011;166:1131–42.
- Norby RJ, Warren JM, Iversen CM, Medlyn BE, McMurtie RE. CO<sub>2</sub> enhancement of forest productivity constrained by limited nitrogen availability. *Proc Natl Acad Sci USA.* 2010;107:19368–73.
- Norris MD, Blair JM, Johnson LC, McKane RB. Assessing changes in biomass, productivity, and C and N stores following *Juniperus virginiana* forest expansion into tallgrass prairie. *Can J For Res.* 2001;31:1940–6.
- Oosterheld M, Loreti J, Semmartin M, Sala OE. Inter-annual variation in primary production of a semi-arid grassland related to previous-year production. *J Veg Sci.* 2001;12:137–42.

- Paruelo JM, Lauenroth WK, Burke IC, Sala OE. Grassland precipitation-use efficiency varies across a resource gradient. *Ecosystems*. 1999;2:64–8.
- Peters DPC, Yao J, Sala OE, Anderson JP. Directional climate change and potential reversal of desertification in arid and semiarid ecosystems. *Glob Chang Biol*. 2012;18:151–63.
- Rosenzweig M. Net primary productivity of terrestrial environments: predictions from climatological data. *Am Nat*. 1968;102:67–74.
- Runyon J, Waring RH, Goward SN, Welles JM. Environmental limits on net primary production and light-use efficiency across the Oregon transect. *Ecol Appl*. 1994;4:226–37.
- Ryan MG, Binkley D, Fownes JH. Age-related decline in forest productivity: pattern and process. *Adv Ecol Res*. 1997;27:213–62.
- Sala OE, Parton WJ, Joyce LA, Lauenroth WK. Primary production of the central grassland region of the United States. *Ecology*. 1988;69:40–5.
- Schuur EAG. Productivity and global climate revisited: the sensitivity of tropical forest growth to precipitation. *Ecology*. 2003;84:1165–70.
- Smith MD, Knapp AK. Dominant species maintain ecosystem function with non-random species loss. *Ecol Lett*. 2003;6:509–17.
- Tierney GL, Fahey TJ. Evaluating minirhizotron estimates of fine root longevity and production in the forest floor of a temperate broadleaf forest. *Plant and Soil*. 2001;229:167–76.
- Tierney GL, Fahey TJ. Estimating belowground primary productivity. In: Fahey TJ, Knapp AK, editors. *Principles and standards for measuring net primary production*. New York: Oxford University Press; 2007.
- Wu Z, Dijkstra P, Koch GW, Penuelas J, Hungate BA. Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation. *Glob Chang Biol*. 2011;17:927–42.
- Yahdjian L, Sala OE. Vegetation structure constrains primary production response to water availability in the Patagonian steppe. *Ecology*. 2006;87:952–62.
- Yahdjian L, Gherardi L, Sala OE. Nitrogen limitation in arid-subhumid ecosystems: a meta-analysis of fertilization studies. *J Arid Environ*. 2011;75:675–80.
- Zhao M, Running SW. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *Science*. 2010;329:940–3.

## Further Readings

- Fahey TJ, Knapp AK, editors. *Principles and standards for measuring net primary production*. New York: Oxford University Press; 2007.
- Gill RA, Kelly RH, Parton WJ, Day KA, Jackson RB, Morgan JA, Scurlock JMO, Tieszen LL, Castle JV, Ojima DS, Zhang XS. Using simple environmental variables to estimate belowground productivity in grasslands. *Glob Ecol Biogeogr*. 2002;11:79–86.
- McNaughton SJ. Compensatory plant growth as a response to herbivory. *Oikos*. 1983;40:329–36.
- Norby RJ, Hanson PJ, O'Neill EG, Tschaplinski TJ, Weltzin JF, Hansen RA, Cheng W, Wullschlegel SD, Gunderson CA, Edwards NT, Johnson DW. Net primary productivity of a CO<sub>2</sub>-enriched deciduous forest and the implications for carbon storage. *Ecol Appl*. 2002;12:1261–6.
- Running SW, Nemani RR, Heinsch FA, Zhao M, Reeves M, Hashimoto H. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*. 2004;54:547–60.
- Smith MD, Knapp AK, Collins SL. A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. *Ecology*. 2009;90:3279–89.

Rachel E. Gallery

## Contents

Introduction .....	249
Biogeography and Climate .....	250
Vegetation Structure and Phenology .....	252
Tropical Rain Forest Biodiversity .....	257
Why Are There so Many Tree Species in Tropical Forests? .....	259
Case Study: Plant Pests Maintain Tree Species Diversity .....	262
Productivity and Nutrient Cycling in Tropical Rain Forests .....	263
Threats to Tropical Rain Forests .....	265
Case Study: Oil Palm .....	267
Future Directions .....	267
References .....	269

---

## Abstract

- Occupying less than 7 % of Earth’s land surface, tropical rain forests harbor perhaps half of the species on Earth and are ecologically, economically, and culturally crucial for issues in global food security, climate change, biodiversity, and human health.
- Geographically located between the latitudes 10°N and 10°S of the equator, lowland tropical rain forest ecosystems share similar physical structure but vary in geology, species composition, and anthropogenic threats across the forests of Southeast Asia, Australia, Africa, and Central and South America.
- Mature tropical rain forests are stratified by multiple canopy and understory layers, and physiognomic properties include evergreen broadleaf tree species, a preponderance of species with large leaves to aid with sunlight capture in the light-limited understory, and leaf properties such as entire margins and drip tips that channel water efficiently from the leaf surface.

---

R.E. Gallery (✉)

School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

e-mail: [rgallery@email.arizona.edu](mailto:rgallery@email.arizona.edu)

- Lianas are increasing in abundance and biomass in a number of tropical rain forests. The additive effects of an increase in liana biomass are correlated with a reduction in tropical forest carbon (C) storage, a value that is currently not considered in global vegetation models.
- Most rain forest tree species do not grow, flower, or fruit year-round. Peaks in leaf flushing, flowering, and fruiting coincide with the high irradiance and low water stress associated with the onset of the wet season. This synchrony is common and largely driven by resource availability, though biotic explanations for synchrony include selection to attract pollinators or seed dispersers and to avoid herbivory and seed predation.
- With few exceptions, species richness across the tree of life is highest in equatorial tropical regions and decreases towards the poles. Tropical rain forests harbor approximately two thirds of the estimated 350,000–500,000 extant flowering plant species on Earth, with high rates of endemism and large numbers of rare species.
- Numerous evolutionary and ecological hypotheses to explain the origin and maintenance of high biological diversity in tropical forests have garnered support and include biogeographic history, evolutionary mechanisms of adaptation and speciation, range size and distribution constraints, and ecological mechanisms promoting species coexistence.
- Continental drift, climate constraints, and long-distance dispersal are responsible for some of the similarities and differences in species across tropical regions. Familial similarity among forests in Amazonia and Southeast Asia can be as high as 50 %, while independent diversification and species radiation mean that much fewer genera (around 10 %) are shared.
- Gradients in climate, parent material and soil age, topography and landscape stability, and atmospheric deposition result in strong heterogeneity in soil nutrient availability from local to regional scales. Soil order, which is generally correlated with soil fertility as a strong predictor of aboveground net primary productivity in tropical forests.
- Tropical forests account for approximately 40 % of terrestrial net primary productivity (NPP), store half of Earth's vegetative C stocks but less than 10 % of its soil C stocks. The relationships between rainfall, temperature, soil fertility, and NPP are complex and require more experimental manipulations to tease apart the interactions.
- Intact tropical forests are net C sinks, but the uptake of C ( $1.1 \pm 0.3 \text{ Pg C year}^{-1}$ ) in intact tropical forests is counteracted by the emissions from tropical biome conversion – a net C source to the atmosphere of  $1.3 \pm 0.2 \text{ Pg C year}^{-1}$  that results in a tropical biome net C balance of approximately zero.
- Stronger El Niño Southern Oscillation (ENSO) effects are increasing the frequency and severity of droughts, fires, hurricanes and cyclones, and flooding events. Recovery of aboveground biomass, species composition, and forest structure all depend on the type and severity of disturbance and its effect on soil fertility.



- Greater use of remote sensing imagery from satellites, airborne Light Detection and Ranging (LiDAR) data, and unmanned drones will allow accurate tracking of disturbance and C stocks as well as monitoring of phenology, foliar canopy chemistry, individual species identification, and biodiversity estimates from local to regional scales.
- The tropical biome is undergoing significant change. Understanding the drivers and impacts of these changes will require sustained advances across multiple disciplines. Ultimately as a society, we are left asking what is the capacity of our remaining and regrowing tropical rain forests to adapt to long-term anthropogenic and climate change and what can we do to moderate these effects while nourishing a healthy human population?

---

## Introduction

Along with their extraordinary biodiversity and predominant influences on global carbon (C), nitrogen (N), and water cycles, tropical rain forests provide powerful inspiration that has driven biological inquiry for centuries. Theories in biogeography, ecology, and evolution by natural selection crystallized through the South America and Southeast Asian journeys of Alexander von Humboldt, Charles Darwin, Alfred Russel Wallace, and Johannes Eugenius Bülow Warming – considered by some to be the founder of tropical ecology. From the lowland rain forests of Venezuela into the Andes, von Humboldt recorded the change in vegetation with climate, drawing the first conclusions that laid the groundwork for the field of biogeography. Both Darwin and Wallace developed their ideas of evolution by natural selection through their observations of exceptional species diversity in South America and Southeast Asian rain forests. Current research questions in tropical rain forest plant ecology comprise determining the origins and maintenance of such extraordinary genetic, species, and habitat diversity; the factors that regulate net primary productivity (NPP) of intact and disturbed tropical forests; and the consequences of the loss and conversion of these forests on global biogeochemical cycles, water cycles, and ecosystem services.

Occupying less than 7 % of Earth's land surface, tropical rain forests harbor perhaps half of the species on Earth and are ecologically, economically, and culturally crucial for issues in global food security, climate change, biodiversity, and human health. Tropical rain forests share a particular combination of climate parameters, floristic composition, forest structure, and plant physiognomy. Though they differ in geology and climate patterns such as intensity of El Niño Southern Oscillation (ENSO) events, tropical rain forests face the common threats of deforestation, land use conversion, invasive species, and changing climate that require the same dedication to conservation and management practices that best suit the unique socioeconomic and cultural characteristics of each region.

Current global, multi-institutional networks, such as the Center for Tropical Forest Science (ctfs.si.edu), monitor the growth and survival of approximately 4.5 million trees and 8,500 species in forests around the world to understand forest function, diversity, and sustainable management to inform natural resource policy and build capacity in the face of climate and land use change.

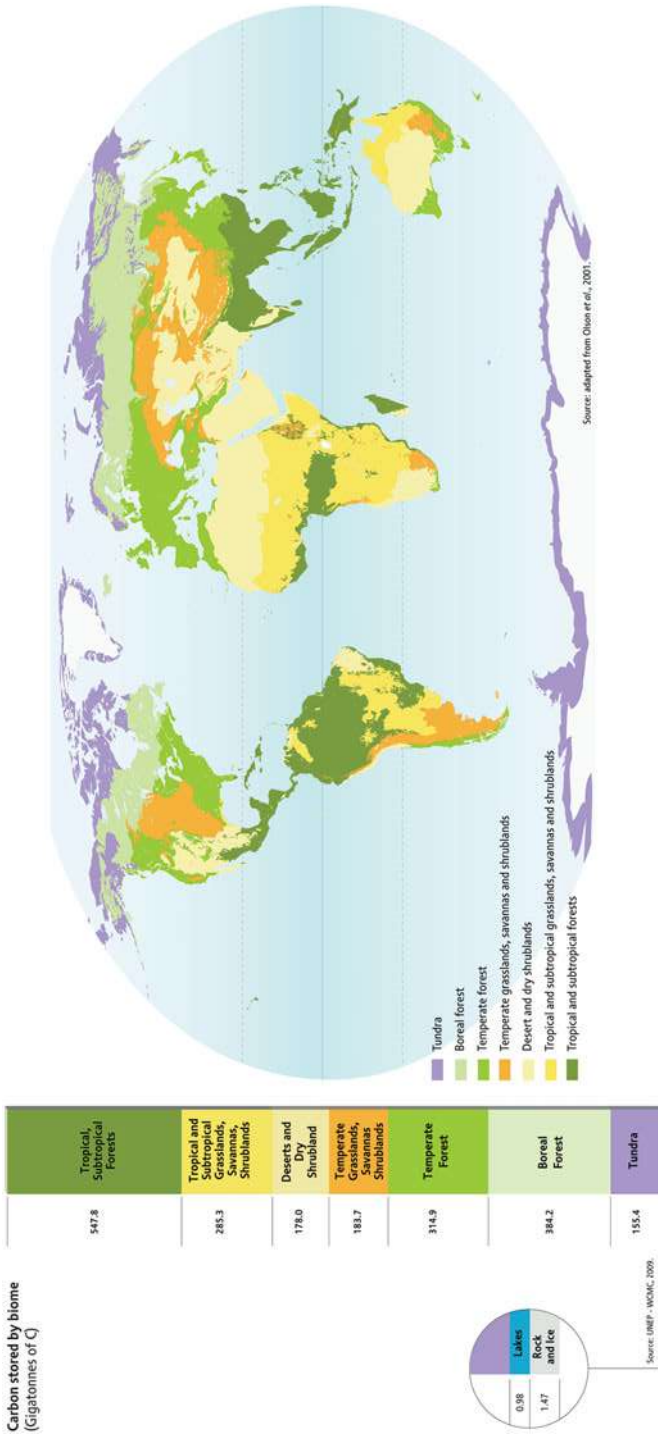
---

## Biogeography and Climate

Geographically located between the latitudes 10°N and 10°S of the equator, lowland tropical rain forest ecosystems share similar physical structure but vary in geology, species composition, and anthropogenic threats across the forests of Southeast Asia, Australia, Africa, and Central and South America (Fig. 1). Approximately 50 % of tropical rain forests are found in the Neotropics, primarily in the Amazon and Orinoco basin with patches in Central America, the Caribbean, and along the Atlantic coast of Brazil. African rain forests are mainly located in the Congo basin extending to the west coast and remnant forests remain in Madagascar. The Australian tropical realm (Oceania) includes Australia, New Guinea, and the Pacific Islands. During his travels, Alfred Russel Wallace noted distinct faunal, though not necessarily floral, differences between Australia and Southeast Asia and the “Wallace line” denotes this boundary. The severely fragmented areas of South and Southeast Asian rain forests account for less than 30 % of rain forests worldwide and are found in India, Sri Lanka, mainland Southeast Asia, the Malay Peninsula, and Indonesia.

The climate of lowland tropical rain forests is warm, humid, and relatively stable. Tropical rain forests are characterized by mean annual temperatures ranging from 23 °C to 28 °C, with mean monthly temperatures no less than 18 °C and rarely exceeding 35 °C. Diurnal temperature fluctuations typically exceed mean monthly ranges, with annual temperature ranges of less than 5 °C. Tropical biomes do not generally experience frost, even at high elevations, and tropical plants and animals do not tolerate freezing. Local variation in rainfall is much higher than temperature variation. Mean monthly precipitation exceeds 60 mm, and annual precipitation can exceed 10 m in aseasonal, evergreen rain forests such as the northwestern region of Colombia known as the Chocó. Peak rainfall typically correlates with the intertropical convergence, which lies over the equators during the two equinoxes. In semi-evergreen forests with seasonal variation in precipitation resulting in distinct rainy and dry seasons that drive plant phenological responses, mean annual rainfall is lower, with dry season months characterized by greater evaporative potential than precipitation. Average humidity in the forest understory is approximately 80 % with higher diurnal variation in the canopy.

Biomes within tropical latitudes are distinguished by differences in elevation and the seasonal patterns of rainfall that create a gradient of vegetation from wet, aseasonal rain forest at high and low latitude to seasonal forest, scrub, savanna, and desert. While there is phylogenetic overlap among the plants of tropical rainforests,



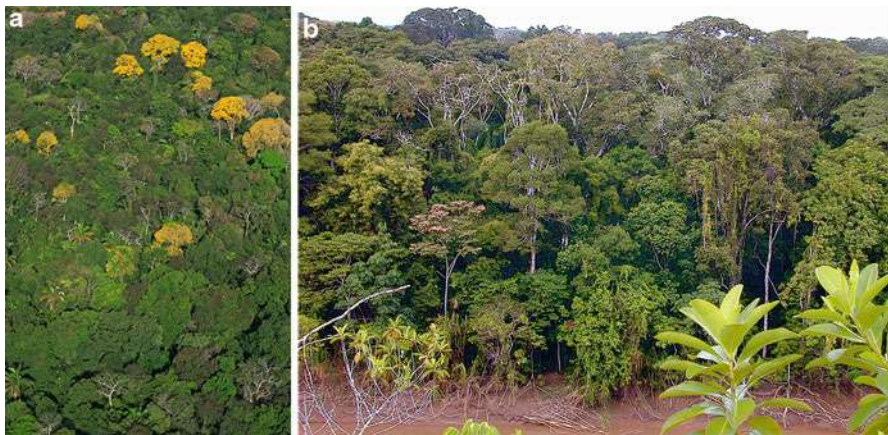
**Fig. 1** Tropical forest distributions and carbon stored by biome (in gigatonnes). Tropical and subtropical forests store more C than any other biome (Reprinted with permission (Riccardo Pravettoni, UNEP/GRID-Arendal, [http://www.grida.no/graphicslib/detail/carbon-stored-by-biome\\_9082](http://www.grida.no/graphicslib/detail/carbon-stored-by-biome_9082)))

tropical montane forests, and tropical deciduous forests, the environmental variables driving ecosystem processes and plant adaptations such as fog, in the case of montane forests, and fires and drought in seasonally dry tropical forest are sufficiently different from tropical rain forests and are beyond the scope of this chapter.

## Vegetation Structure and Phenology

Vegetation characterization of tropical rain forests can be defined by structural and physiognomic properties that are strongly influenced by physicochemical edaphic factors. Mature tropical rain forests are stratified by multiple canopy and understory layers. The distinct vertical profile of tropical rain forests generally includes emergent trees that arise above the canopy, high upper canopy trees with average height of 30–40 m, low tree sub-canopy, shrub understory, and ground layer of herbaceous plants and ferns (Fig. 2a, b).

Aside from bamboos, grasses are uncommon in most tropical rain forest understories. Epiphytes and woody vines called lianas that rely on trees for structural support to reach the forest canopy are conspicuous, as are tree buttresses that support trees by providing stability in shallow tropical soils (Fig. 3a, b). Approximately three quarters of the world's fern species and half of the world's bryophytes (mosses, liverworts, and hornworts) are found in tropical forests. Physiognomic properties include evergreen broadleaf tree species, a preponderance of species with large leaves to aid with sunlight capture in the light-limited understory, and



**Fig. 2** (a) Aerial photo of a Neotropical rain forest canopy. The brilliant yellow crowns display the synchronous flowering of *Tabebuia guayacan* (Bignoniaceae) trees. Emergent trees rise above the forest canopy and palm trees and various tree architectures are apparent. The range in hue of individual crowns depicts variation in foliar chemistry and water content (Photo credit Christian Ziegler). (b) Cross section of a lowland Amazon rain forest in Manu National Park, Peru, shows a distinct vertical profile from understory shrubs to emergent trees. River erosion exposes roots (Photo credit Kyle Dexter)



**Fig. 3** (a) The buttress of this *Ficus* (Moraceae) tree in Corcovado National Park, Costa Rica, provides support and stability in shallow tropical forest soils (Photo credit Andrea Vincent). (b) Woody lianas rely on trees for structural support to reach the forest canopy. Liana abundance and biomass are increasing in a number of tropical rain forests, including the La Selva Biological Station, Costa Rica, where this photo was taken, with significant implications for tree community diversity, gap dynamics and forest structure, and tropical forest nutrient cycling (Photo credit Eloisa Lasso)

leaf properties such as entire margins and drip tips that channel water efficiently from the leaf surface. Cauliflory, the development of flowers on tree trunks and main branches, is common in aseasonal tropical understory trees and facilitates pollination by non-volant insects or animals. Not surprisingly, the percentage of deciduous tree species increases with increasing seasonality. Across the strong precipitation gradient along the Isthmus of Panama, deciduous trees account for less than 5 % in more aseasonal forests on the Atlantic to a quarter of tree species in the forest communities on the Pacific side.

Competition for light, water, and nutrients varying over heterogeneous landscapes generate and shape ecophysiological adaptations in plants. Equatorial solar radiation levels are high, and canopy leaves and leaves exposed to direct sunlight experience very different irradiance and humidity than understory leaves. Greater than 99 % of sunlight is absorbed and reflected as the light passes through the forest canopy, resulting in low light intensity and quality in the forest understory where competition for light is high and certain plants can rapidly respond to the patchworks of light created by sunflecks. Life history strategies across the light demanding to shade tolerant spectrum include, at the one end, pioneer species with high



photosynthesis and respiration rates and low wood density to slow growing, well-defended, high wood density species that can persist in the understory until a gap forms overhead. Species are aligned across a competition–colonization continuum along a multitude of axes including seed size and dispersal, leaf lifespan, and population turnover that together highlight tradeoffs in resource allocation and reproductive strategies. Water limitation controls transpiration and photosynthesis, and tropical trees can transpire several hundred liters of water a day, which emphasizes the importance of reducing cavitation risks during low water availability. Of course environmental tolerances to temperature and water availability drive global patterns of plant distributions, and within tropical forests interspecific differences in drought tolerance have been shown to determine plant species distributions at local scales and across the strong rainfall gradient of the Isthmus of Panama. Among the soil nutrients that affect plant productivity, phosphorous (P), which is rapidly mobilized by chemical and microbial activity, is often limiting in highly weathered tropical soils. A more detailed discussion of biogeochemistry and plant productivity can be found in section “[Productivity and Nutrient Cycling in Tropical Rain Forests](#)”.

While lianas are found in temperate rain forests, their predominance and diversity in tropical rain forests are notable, as is the trend that they are increasing in abundance and biomass in a number of tropical rain forests. Contributing up to 45 % of woody stems and 35 % of species richness in a tropical forest community, lianas significantly reduce tree growth rates through direct competition, more than double tree mortality risks, and increase gap size and severity through canopy connectivity, and the capacity for lianas to alter successional pathways in tropical rain forests is only beginning to be understood (van der Heijden et al. 2013). An increase in liana biomass has serious implications for tree community diversity, gap dynamics and forest structure, and tropical forest nutrient cycling. For example, lianas reduce tree growth and survival in the slower-growing, higher wood density trees that support them, which, along with changing gap regimes, shift species composition towards faster-growing trees with lower wood density. While accurate predictions require more data, the additive effects of an increase in liana biomass are correlated with a reduction in tropical forest C storage, a value that is currently not considered in global vegetation models.

Little is known about cambial phenology – the seasonality of stem growth – in tropical rain forest trees. Our lack of understanding of the triggering factors of cambial dormancy in tropical rain forest trees has led to the long-standing assumption that tropical trees do not form annual growth rings (Jacoby 1989; Worbes 2002). Furthermore, the complex wood anatomy characteristic of the majority of tropical tree species has long steered dendrochronologists away from tropical regions. In recent decades, however, distinct annual growth ring boundaries, often consisting of marginal parenchyma bands and induced by cambial dormancy, have been detected in multiple lowland tropical rain forest species. As a result, an increasing number of reliable, climate-sensitive tree-ring chronologies are now available based on trees from various tropical biomes across Asia, the Amazon region, and Africa. These chronologies reflect seasonally fluctuating

climatic conditions that typically consist of distinct dry seasons but can also consist of periodical flooding (Schongart et al. 2004). In regions with a bimodal rainfall distribution (e.g., eastern Africa), trees can exhibit a bimodal pattern of cambial activity, and two growth rings can be found per year. Water availability is a major driver of phenological periodicity in seasonal tropical rain forests, and leaf phenology is generally synchronized with the seasonality of soil water content and tree water status. In deciduous trees, leaf fall typically occurs at the end of the dry season and leaf flushing in the wet season. Deciduousness, however, is species and site specific and can be a function of tree canopy status, with canopy and emergent trees generally showing a more distinct phenological seasonality and deciduousness than understory trees (see Fig. 2a). There is plasticity in this trait; some species have seasonal leaf fall at dry sites but are evergreen at sites with less moisture stress.

Though the climate of tropical rain forests has more tempered seasonality relative to other ecosystems, most rain forest tree species do not grow, flower, or fruit year-round. Periods of leaf flush, bud burst, flowering, fruiting, and senescence that are related to climate conditions and day length (photoperiod) are considered phenological responses, the proximate and ultimate causes of which have been studied from individual variation within populations to community and guild-level patterns. In seasonal tropical rain forests, peaks in leaf flushing, flowering (see Fig. 2a), and fruiting coincide with the high irradiance and low water stress associated with the onset of the wet season. This synchrony of events is common within communities and largely driven by resource availability, though biotic explanations for synchrony include selection to attract pollinators or seed dispersers and to avoid herbivory and seed predation (van Schaik et al. 1993). The synchronous flowering of canopy emergent tree species such as *Dipteryx panamensis* (Fabaceae) is visible in high-resolution satellite images, which enable individual tracking and have revolutionized the study of remote and large tracks of forests. Synchronous supra-annual flowering and mast fruiting that may lead to seed predator satiation are defining features of the Dipterocarp forests of Southeast Asia, with Borneo housing the greatest diversity of Dipterocarpaceae that are increasingly threatened by extensive logging and land conversion. Bamboos also wait decades between synchronized flowering before dying back. Monocarpic or semelparous trees that reproduce only once are uncommon, though examples can be found in the Neotropical genera *Tachigali* (Fabaceae) and *Spathelia* (Rutaceae) and the genus *Harmsioplanax* (Araliaceae) in tropical Asia. Wind pollination is relatively rare in tropical rain forests and many coevolutionary pollination, and seed dispersal relationships have developed between plants and insects, birds, bats, fish, and mammals.

In this chapter on tropical rain forest plant ecology, I would be remiss not to highlight a few of the archetypal associations between tropical plants and the organisms that rely on them for food and habitat. Each of the examples detailed below are pantropical and emphasize the extraordinary complexity of ecological systems. They also demonstrate the coevolution of symbiotic relationships between plants, insects, and fungi for protection, nutrient acquisition, and pollination.



**Fig. 4** Ant-plant (myrmecophytic) symbioses are a pantropical phenomenon involving greater than 100 plant genera and 40 ant genera. In this photo taken in Santa Rosa National Park, Costa Rica, the *Acacia* (Fabaceae) species form mutualistic associations with ants in the genus *Pseudomyrmex* (Formicidae) (Photo credit Andrea Vincent)

Ant-plant (myrmecophytic) symbioses are a pantropical phenomenon involving greater than 100 plant genera and 40 ant genera, whereby ants living within specialized structures of the plant – called domatia – defend the plant against herbivory and pathogen attack (Fig. 4; reviewed in Heil and McKey 2003). These often-obligate symbioses are incredibly effective with ants receiving food – namely, Beltian bodies and nectar – and habitat and plants receiving a full-time security force. Whereas herbivores and pathogens have counter adapted many strategies for overcoming plant chemical defenses, the resident ants of myrmecophytes earn their keep by effectually defending plants from their pests.

Generally considered keystone species in tropical forests, figs of the genus *Ficus* (Moraceae), (Fig. 3a) range in growth form from small shrubs to climbers to canopy trees and epiphytic parasites (e.g., strangler figs). Fig fruits are a reliable year-round and nutritious food source for numerous frugivores, and the fig keystone status stems from their role in sustaining frugivore communities when other food resources are limiting. Most notable is the intimate mutualism between figs and their tiny, obligate wasp pollinators (Agaonidae, Chalcidoidea). Phylogenetic evidence supports the hypothesis that this mutualism arose once approximately 87 million years ago. The long-standing view of a unique one-to-one species-specific pollination syndrome, however, has been challenged by recent progress in phylogenetic studies of figs and their pollinating wasps (reviewed in Herre et al. 2008). Fig species pollinated by two or more wasp species suggest that fig and pollinator speciation are not always tightly linked. Non-pollinating fig wasps are common and these parasites exploit this mutualism in diverse ways that might also drive fig adaptations. Finally, figs have some of the most effective long-distance dispersal



of any tropical tree species, with dispersal ranges of hundreds of square kilometers driven by fig wasp-mediated gene flow and seed dispersal via the numerous fig frugivores.

Mycorrhizal associations between plant roots and symbiotic fungi are pervasive and not unique to tropical rain forests; greater than 90 % of plant families form mycorrhizal associations. While ectomycorrhizal tree species are less common, both endo- and ectomycorrhizal fungi are found in tropical forests worldwide, and trees can host both groups of symbionts simultaneously. Arbuscular mycorrhizas (AM; Glomeromycota) are endomycorrhiza whose hyphae enter plant cells and produce vesicles or arbuscules that increase the surface area of contact between the plant root and fungus to facilitate nutrient transfer. AM fungi are cosmopolitan with broad host ranges though different plant species responses to mycorrhiza communities can influence the competitive outcome among seedlings. Ectomycorrhizas (EM) are found across fungal phyla (Basidiomycota, Ascomycota, Zygomycota) and their species number in the thousands compared to only hundreds of arbuscular mycorrhizal species. EM hyphae sheath the root and an extensive hyphal network, called a Hartig net, runs between plant cells within the root cortex. Tree species with EM are less common than those with AM, but all species of Dipterocarpaceae form EM associations, as do species in the Fagaceae and Fabaceae subfamily Caesalpinioideae. In both types of association, carbon fixed from the plant is transferred to the heterotrophic fungus. In return both ecto- and endomycorrhizas increase root surface area, thereby improving plant nutrient acquisition of P, N, calcium, potassium and other ions that tend to be limiting in tropical soils. There is evidence that these associations also improve plant resistance to root pathogens and tolerance to drought. The host-specific effect of different mycorrhizal communities on plant growth has been proposed as a potential mechanism reducing plant community richness. Tree species hosting particular suites of mycorrhizal communities could create a positive feedback for conspecific over heterospecific juvenile recruitment. Furthermore, in certain low diversity forests the dominant tree species tends to form EM associations and it has been hypothesized that an EM network may provide recruitment advantages to EM plant species over non-EM plant species through positive feedbacks. This hypothesis requires further testing.

---

## **Tropical Rain Forest Biodiversity**

With few exceptions, species richness across the tree of life is highest in equatorial tropical regions and decreases towards the poles. Numerous evolutionary and ecological hypotheses to explain the origin and maintenance of the latitudinal gradient in biodiversity have garnered support and include biogeographic history, evolutionary mechanisms of adaptation and speciation, range size and distribution constraints, and ecological mechanisms promoting species coexistence. After decades of research on this topic it is evident that no individual explanation is sufficient to explain this conspicuous biogeographic pattern.

The current diversity and distribution of modern plant lineages has been shaped by numerous extinction (e.g., Devonian, Permian, Cretaceous) and radiation events throughout Earth's history. The retraction of tropical rain forests during the cooler, drier Pleistocene glacial periods (ca. 100,000 year per cycle) and expansion of tropical rain forests during warmer, wetter interglacial periods (ca. 10–20,000 year per cycle) created fragmented refugia in African and Australian, though recent evidence suggests not Neotropical, forests, that may have promoted lineage differentiation and allopatric speciation that contribute to the extant high tropical plant diversity. Different scales over which diversity is measured include alpha diversity (local, habitat scale), beta diversity (species turnover at landscape to regional scales), and gamma diversity (total regional species richness). Since regional diversity reflects a balance between speciation and extinction, it should be higher in larger, older areas that offer more opportunities for isolation and divergence through environmental heterogeneity as well as lower extinction probabilities through species-area relationships and millennia without major climatic shifts, in other words, in tropical rain forest biomes.

Continental drift, climate constraints, and long-distance dispersal are responsible for some of the similarities and differences in species across tropical regions. Dipterocarpaceae are dominant only in Southeast Asia, and palms (Arecaceae) and legume species in the Fabaceae are abundant in South American tropical rain forests (e.g., Fig. 2a), but not in African ones. There are, however, a number of plant families shared between South America, Africa, and Southeast Asia (from 27 to 44 in a recent global analysis of 4 ha plots by Ricklefs and Renner 2012). In contrast, independent diversification and species radiation mean that much fewer genera are shared across regions. Between 58 % and 68 % of plant families (44 families) are shared between Yasuni, Ecuador, (65 families) and Pasoh, Malaysia, (76 families), whereas only approximately 12 % (35 genera) of their 296 (Yasuni) and 259 (Pasoh) genera overlap. Some species are widely distributed with pantropical ranges, for example, *Ceiba pentandra* (Malvaceae), a canopy pioneer tree, whose range encompasses Central and South America, the Caribbean, and eastern Africa. Interestingly, the low nucleotide divergence in microsatellite chloroplast and nuclear ribosomal DNA data among Neotropical and African populations supports long-distance dispersal, and not vicariance, as the explanation for this species' range (Dick et al. 2007). Population genetic data provide a means of inferring the dispersal and historical biogeography of species. See Kraft and Ackerly (Chap. 3, ► “[Assembly of Plant Communities](#)”) for an excellent description of phylogenetic analysis and structure within and among communities.

Tropical rain forests harbor approximately two thirds of the estimated 350,000–500,000 extant flowering plant species on Earth. Floristic endemism, whose cause may be attributed to young species age, is high – especially in island systems such as Indonesia where greater than 50 % of the indigenous vascular plant taxa do not occur anywhere else. Although tropical rain forests are generally considered synonymous with diversity, within these systems tree alpha diversity varies considerably and is broadly correlated with mean annual temperature (MAT) and mean annual precipitation (MAP). Numerous studies using the CTFS forest inventory plots reveal that patterns of alpha diversity and species or familial

dominance vary across African, American, and Asian tropical rain forests from a mean of 22 species of tree  $\geq 10$  cm dbh per ha in southern India to 254 species per ha in Ecuadorian Amazon (Table 1; Condit et al. 2005). Similarly, the number of plant families represented in forest communities varies from 47 in Korup, Cameroon, to 76 in Lambir, Malaysia (Ricklefs and Renner 2012).

Local dominance by one or a few species is found in primary rain forests throughout the tropics. In the Asian tropics, the family Dipterocarpaceae (e.g., *Dryobalanops aromatica*) dominates, while many species in the leguminous family Caesalpiniaceae dominate in the African and Neotropics (e.g., *Gilbertiodendron dewevrei* in Congo, *Mora excelsa* in Trinidad, and *Peltogyne gracilipes* in Brazil). A comprehensive assessment by Ter Steege et al. (2013) of the composition and biogeography of tree communities from 1,170 inventory plots throughout Amazonia yielded the stunning discovery that a mere 227 of the roughly 16,000 tree species in this region account for half of the trees. Species of palm trees in the Arecaceae are predominant, as well as species in the Myristicaceae, Lecythidaceae, and commonly cultivated trees. Most of these so-called hyperdominant species forming “predictable oligarchies” are only dominant in certain forest types and, while they demonstrate large geographic ranges, show strong evidence of habitat specialization though a broad range of shade tolerance is represented. It is the rare species, with average abundances of  $\leq 1$  individual per hectare that drive species richness of tropical communities (Table 2). The striking discovery that a small suite of species largely drives Amazonia’s biogeochemical cycling opens areas of inquiry into the implications of species-specific effects of climate change on productivity and phenology in this region. Elucidating mechanisms that promote dominance and monodominance also provide important conceptual contrast to those explaining high species diversity.

---

## Why Are There so Many Tree Species in Tropical Forests?

What processes underlie the diversity and assembly of communities and, to paraphrase Egbert Leigh et al. (2004), why are there so many trees species in certain tropical forests? A combination of factors (historical biogeography, environmental tolerances, demographic stochasticity, and limitations to propagule dispersal) leading to neutral ecological drift (Hubbell 2001) have been proposed as the main influences over the composition and relative abundance of species in a regional species pool. Environmental heterogeneity and dispersal limitation influence species turnover among communities (beta diversity), which can be low even when alpha diversity is high. In contrast, alpha diversity may be more strongly controlled by stochastic and biological processes such as disturbance and especially pressure and specialization of pests on locally abundant hosts. Instead of dichotomous either-or explanations, it is likely that high sympatric species coexistence results from “The Ecological Theater and The Evolutionary Play,” (Hutchinson 1965) – a combination of ecological filtering and biotic interactions operating over ecological (short-term selective processes in a fixed gene pool) and evolutionary (long-term process acting on a variable gene pool) timescales.

**Table 1** Forest diversity by region from large tropical forest plots associated with the Center for Tropical Forest Science (CTFS). Lines in the table denote Southeast Asian, Neotropical, and African regions. Annual precipitation for each forest is shown in millimeters (mm), and the number of dry season months is in parentheses. Two different size classes are shown for the full plot and per hectare. Sites marked with an asterisk were < 25 ha, and data for those sites are based on the full 16 or 20 ha. Main references for each plot are footnoted (Redrawn with permission Condit et al. 2005)

	Plot size (ha)	mm annual precipitation (dry season in mo.)	Species per ha $\geq 10$ cm dbh	Species in full plot $\geq 10$ cm dbh	Species per ha $\geq 1$ cm dbh	Species in full plot $\geq 1$ cm dbh
Lambia, Borneo, Malaysia <sup>a</sup>	52	2,664 (0)	245.7	1,008	618.1	1,179
Huai Kha Khaeng, Thailand <sup>b</sup>	50	1,476 (6)	65.6	217	101.8	259
Mudumalai, India <sup>c</sup>	50	1,206 (6)	22.0	63	25.6	72
Pasoh, Peninsular Malaysia <sup>d</sup>	50	1,788 (0)	207.3	678	496.5	814
Sinharaja, Sri Lanka	25	5,074 (0)	71.2	167	142.7	205
Palanan, Philippines*	16	3,218 (4)	98.9	262	201.6	335
Barro Colorado, Panama <sup>e</sup>	50	2,551 (3)	90.7	227	168.0	301
La Planada, Colombia	25	4,087 (0)	85.0	172	150.1	219
Yasuni, Ecuador <sup>f</sup>	25	3,081 (0)	253.6	820	665.2	1,104
Luquillo, Puerto Rico <sup>g*</sup>	16	3,548 (0)	42.2	87	77.6	140
Korup, Cameroon	50	5,272 (3)	85.4	307	235.1	494
Ituri, D.R. Congo <sup>h</sup> :						
Lenda (monodominant)	20	1,674 (2)	49.1	211	166.0	365
Edoro (mixed)	20	1,785 (2)	67.0	212	172.2	380

<sup>a</sup>Lee et al. (2002)

<sup>b</sup>Bunyavejchewin et al. (2001)

<sup>c</sup>Sukumar et al. (1992)

<sup>d</sup>Manokaran et al. (1992), Condit et al. (1996b, 1999)

<sup>e</sup>Hubbell and Foster (1983), Condit et al. (1996a, 1999)

<sup>f</sup>Romoleroux et al. (1997), Valencia et al. (2004)

<sup>g</sup>Zimmerman et al. (1994), Thompson et al. (2002)

<sup>h</sup>Makana et al. (1998)

**Table 2** Species rarity and dominance by region. Percent of rare species (those with  $\leq 0.3$  individuals per ha) at each of the plots and relative abundance of the dominant species. Both are given as mean  $\pm$  95 % confidence limits, based on replicate 20-ha subquadrats. Confidence limits for Congo sites could not be calculated, since the plots were only 20 ha; for sites marked with an asterisk, the estimates are based on the full 16 ha and also lack confidence limits. Dominant species for each site is listed along with authority and family (Redrawn with permission Condit et al. 2005)

Plot	% Rare species	% Dominance	Dominant species
Lambia, Borneo, Malaysia	14.9 $\pm$ 3.7	2.6 $\pm$ 1.0	<i>Dryobalanops aromatica</i> Gaertner (Dipterocarpaceae)
Huai Kha Khaeng, Thailand	44.8 $\pm$ 1.5	10.0 $\pm$ 5.2	<i>Croton oblongifolius</i> Roxb. (Euphorbiaceae)
Mudumalai, India	41.7 $\pm$ 4.8	22.8 $\pm$ 6.5	<i>Kydia calycina</i> Roxb. (Malvaceae)
Pasoh, Peninsular Malaysia	19.2 $\pm$ 3.5	2.7 $\pm$ 0.3	<i>Xerospermum noronhianum</i> Blume (Sapindaceae)
Sinharaja, Sri Lanka	16.6 $\pm$ 0.9	12.1 $\pm$ 0.4	<i>Humboldtia laurifolia</i> M. Vahl (Fabaceae)
Palanan, Philippines*	37.9	5.6	<i>Nephelium lappaceum</i> Poiret (Sapindaceae)
Barro Colorado, Panama	25.6 $\pm$ 2.7	15.7 $\pm$ 1.9	<i>Hybanthus prunifolius</i> Schulze-Menz (Violaceae)
La Planada, Colombia	24.2 $\pm$ 2.9	15.6 $\pm$ 0.1	<i>Faramea coffeoides</i> C.M. Taylor (Rubiaceae)
Yasuni, Ecuador	31.1 $\pm$ 0.6	3.1 $\pm$ 0.1	<i>Matisia oblongifolia</i> Poeppig & Endl. (Malvaceae)
Luquillo, Puerto Rico*	40.7	19.6	<i>Palicourea riparia</i> Benth. (Rubiaceae)
Korup, Cameroon	29.2 $\pm$ 2.6	8.3 $\pm$ 1.5	<i>Phyllobotryum spathulatum</i> Müll. Arg. (Salicaceae)
Ituri, D.R. Congo:			
Lenda (monodominant)	48.4	45.0	<i>Scaphopetalum dewevrei</i> Wildem. & Th. Dur. (Malvaceae)
Edoro (mixed)	52.2	41.8	<i>Scaphopetalum dewevrei</i>

When intraspecific interactions are more negative than interspecific interactions, species are at a relative advantage when rare and disadvantage when common. This has a stabilizing effect on species diversity. Interspecific trade-offs in species dispersal and competitive abilities result in niche partitioning along the competition–colonization continuum of traits. Niche partitioning and compensatory mortality (e.g., Janzen–Connell negative density-dependent effects and low recruitment near conspecifics) are therefore among the significant factors that favor the sympatric coexistence of tree species by preventing species dominance and

competitive exclusion of species from the community. They maintain alpha diversity within communities by reducing interspecific competition or through density-dependent pest regulation of plant populations. Pervasive dispersal and recruitment limitation, whereby a species does not successfully establish in all sites it is capable of occupying, further reduce the extirpation of less competitive species in a community.

Forest disturbances such as tree fall gaps create light and nutrient heterogeneity that generate niche opportunities allowing tree species coexistence across the continuum of light-demanding “pioneer” to longer lived, better defended shade-tolerant species. Tree fall gaps are colonized in a number of ways that can alter regeneration or successional pathways. Light-demanding pioneer species germinate readily from soil seed banks when the high light quality and temperature conditions from gaps arise. The rapid growth rate of these species results in a developing understory that leads to favorable microsites for other species to recruit. Recruitment from the seedling bank is equally common. Shade-tolerant seedlings and saplings persisting in the understory for decades are also able to exploit the high light environment of gaps and respond with rapid growth rates. Vegetative propagation, clonal shoots, and lateral growth from vines and lianas are also pathways for gap colonization, and plant recruitment and growth rates thin and slow as competition for light increases. Despite the importance of forest gaps, there is little evidence that variations in adaptation to disturbance account for the high alpha tree species diversity of tropical rain forests. Disturbance is nevertheless one of several factors that add to seemingly unpredictable microclimatic conditions within tropical forests.

---

### **Case Study: Plant Pests Maintain Tree Species Diversity**

As mentioned above, plant–pest interactions are considered one of the predominant mechanisms allowing high species diversity to be maintained in tropical tree communities. The long-standing Janzen–Connell hypothesis suggests that specialized pests such as insects and pathogens maintain high plant diversity by causing increased mortality in areas of high conspecific plant density (negative density dependence), thereby preventing species dominance. A recent experimental test of this hypothesis shows that fungal plant pathogens, but not insects, have a community-wide role in maintaining seedling diversity in a Neotropical forest (Bagchi et al. 2014). In a 17-month experiment, researchers compared the diversity of the seed rain to the diversity of seedlings germinating in adjacent control, fungicide, and insecticide-treated plots. The diversity of germinating seedlings was higher than that of the seed rain, suggesting an important recruitment filter at the seed-to-seedling stage. Among plots, plant species richness was reduced by 16 % in plots treated with fungicide. There was no change in species richness in plots treated with insecticide though a change in relative abundance of plant species indicates a disproportionate effect of insects on certain plant species. The original assumption of specialized pests driving the negatively density-dependent mortality

thought to regulate populations (see Janzen 1970; Connell 1971), however, does not seem to hold either for plant–phytophage or plant–pathogen interactions in tropical forests. Polyphagy in insects (Novotny et al. 2002) and fungi (Gilbert and Webb 2007) is the more common strategy in species-rich communities with high numbers of locally rare species. Nevertheless, plant preferences of pests and the variation in plant responses to common pests appear to be sufficient to facilitate coexistence among plants as described in the Janzen–Connell hypothesis.

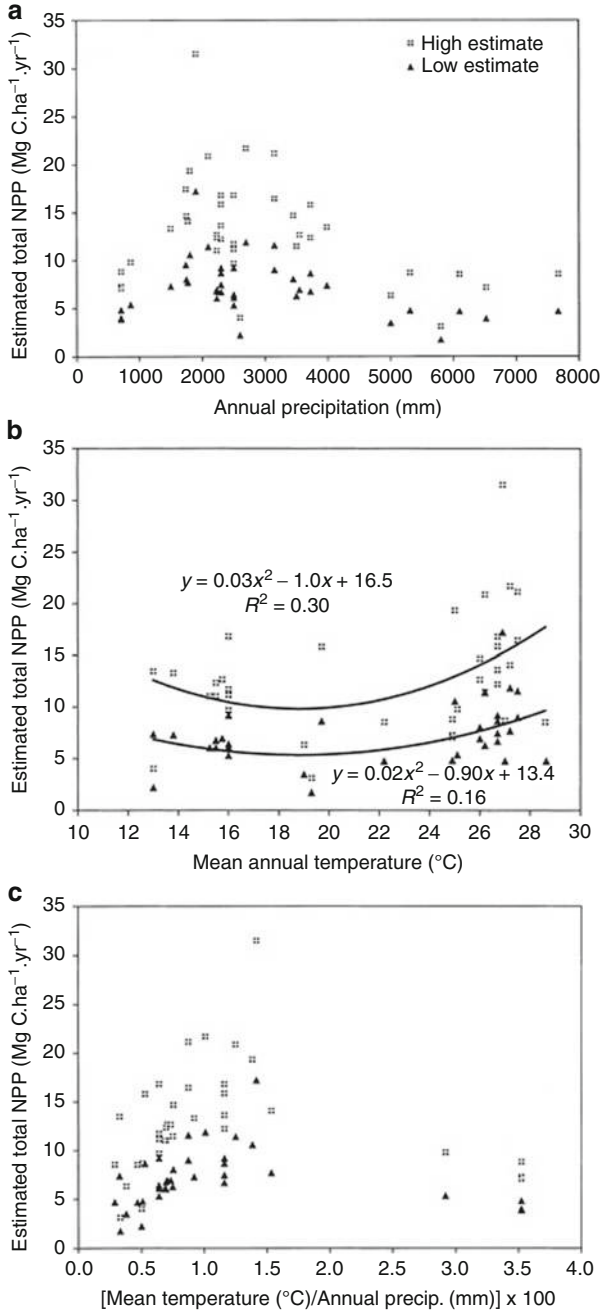
---

## Productivity and Nutrient Cycling in Tropical Rain Forests

Gradients in climate, parent material and soil age, topography and landscape stability, and atmospheric deposition result in strong heterogeneity in soil nutrient availability from local to regional scales. Tropical rain forests encompass a gradient of soils ranging from young, N-poor Alfisols whose nutrients are primarily derived from parent material to older, highly chemically weathered Ultisols and Oxisols (Townsend et al. 2008). Widespread Ultisols, or “red clay soils” due to their accumulated clay minerals in the B-horizon, are acidic with low fertility and cation exchange capacity; however, their clay content gives them greater nutrient-holding capacity than Oxisols. The highly weathered, nutrient poor, acidic Oxisols are dominated by aluminum and iron oxides and have low humus and clay content. Less common are volcanic Andisols, found in areas such as Hawaii and the infertile white sand Spodosols of Amazonia.

Tropical forests are typically characterized by rapid recycling of nutrients through the action of ants, termites, fungi, and other soil microbes, with dead organic matter decomposing over the scale of weeks compared to years in more temperate zones. Productivity and decomposition of necromass are tightly coupled in tropical forests and can be controlled by a number of different limiting nutrients. For example, denitrification often exceeds N fixation resulting in significant N losses. A meta-analysis of 81 lowland tropical rain forest sites showed soil order, which is generally correlated with soil fertility, to be a strong predictor of above-ground NPP. Through this analysis, Cleveland et al. (2011) found that soil P availability controls the tropical C cycle directly and indirectly through constraints on N turnover and N availability and the subsequent effects on photosynthetic rates. NPP can be limited by temperature, moisture, or nutrient availability, and higher elevation forests are generally less productive than lowland forests because of a combination of these limiting factors. Although Hawaiian forests show a strong increasing trend in NPP with increasing rainfall, the controls of NPP are not simple or linear. The relationships between rainfall, temperature, and NPP estimated from 39 different tropical forests were complex; both low and high MAT were associated with high NPP, and therefore the ratio of MAP to MAT was a better predictor of NPP (Fig. 5; Clark et al. 2001).

Tropical forests store approximately half of Earth’s vegetation carbon stocks but less than 10 % of Earth’s soil carbon stocks (see Fig. 1). In tropical forests there is as much carbon stored in live biomass as there is in soils, in contrast to other biomes



**Fig. 5** The relationships between low and high estimates of NPP for 39 old-growth tropical forest sites around the world and (a) annual precipitation ( $P$ ), (b) mean annual temperature ( $T$ ), and (c) the ratio  $T/P \times 100$  (Reprinted with permission Clark et al. 2001)



where soils are the dominant C store. Although there are seasonal patterns of plant growth in the tropics, high solar radiation and a relatively stable warm, wet climate provide more consistently suitable conditions for growth than drier and colder regions. Consequently, tropical forests account for approximately 40 % of NPP. An estimated 60 % of tropical forests are classified as secondary or degraded forests (Chazdon 2003), meaning tropical deforestation has considerable implications for Earth's carbon cycle.

There is evidence that aboveground biomass production is increasing in the forests of South America, Africa, and Asia, though notably not Australia. The primary mechanisms driving this trend are thought to include increased resource availability through the effect of rising atmospheric CO<sub>2</sub>, air temperature, and solar radiation on NPP, and forest recovery from past disturbances. The contrasting pattern in Australian tropical rain forests is linked to the magnitude, frequency, and scale of natural disturbances such as cyclones and strong droughts from El Niño events. Intact tropical forests are net C sinks, but the uptake of C ( $1.1 \pm 0.3 \text{ Pg C year}^{-1}$ ) in intact tropical forests is counteracted by the emissions from tropical biome conversion – a net C source to the atmosphere of  $1.3 \pm 0.2 \text{ Pg C year}^{-1}$  that results in a tropical biome net C balance of approximately zero (Malhi 2010). However, there are few studies under ambient or elevated CO<sub>2</sub> conditions where the net C uptake of tropical forests has been quantified, and the role of tropical forests in Earth's C cycle, while critical, is far from understood.

---

## Threats to Tropical Rain Forests

Population growth in tropical developing countries, large-scale agriculture for food and biofuels, industrial logging, construction of roads and dams, and oil and gas development are among the most significant anthropogenic threats to tropical old growth rain forests and the biodiversity they contain. Current estimates put global tropical deforestation rates at greater than 15 million hectares per year with the highest contemporary deforestation rates recorded in Southeast Asia (Laurance et al. 2011). In 1988, Norman Myers introduced the biodiversity hotspot concept in an effort to define regions of utmost importance for biological diversity conservation. Defined as threatened regions that harbor a high diversity of endemic species, the 34 biodiversity hotspots currently identified by Conservation International (expanded from 25 in Myers et al. 2000) contain over 50 % of the world's endemic plant species yet account for less than 3 % of Earth's terrestrial cover. The tropical hotspots in most urgent need of protection and sustainable management include forests of Madagascar, Philippines, Atlantic coastal forest of Brazil, the Caribbean, Indo-Burma, and Western Ghats/Sri Lanka (Sodhi et al. 2007), which will require economic incentives and feasible sustainable alternatives to deforestation.

The Millennium Ecosystem Assessment projects that 11–22 % of 2000 tropical forest cover will disappear by 2050 (Table 3; Asner et al. 2009). Forest fragmentation is arguably no less a threat to tropical forests than whole-scale deforestation. Harder to quantify, fragmented patches of forest within a matrix of anthropogenically

**Table 3** Approximate geographic extent of contemporary forest cover, deforestation, and selective logging by region in the humid tropical forest biome. Values are in km<sup>2</sup>, with percentage of biome extent also given<sup>a</sup> (Redrawn with permission Asner et al. 2009)

Region	Total biome extent (km <sup>2</sup> )	Area with 0–50 % forest cover, 2005 (km <sup>2</sup> ) <sup>b</sup>	Area with 50–100 % forest cover 2005 <sup>b</sup> (km <sup>2</sup> )	Forest area cleared 2000–2005 <sup>c</sup> (km <sup>2</sup> )	Selective logging <sup>d</sup> (2000s) (km <sup>2</sup> )
Africa	2,918,511	1,085,941 (37.2 %)	1,832,569 (62.8 %)	14,972 (0.5 %)	561,153 (19.2 %)
Asia/Oceania	7,191,529	5,234,293 (72.8 %)	1,957,236 (27.2 %)	93,955 (1.3 %)	1,777,963 (27.2 %)
Central America/Caribbean	685,840	501,415 (73.1 %)		184,425 (26.9 %)	9,687 (1.4 %)
South America	8,826,966	3,194,632 (36.2 %)	5,632,334 (63.8 %)	156,001 (1.8 %)	1,603,166 (18.2 %)
Total	19,622,846	10,016,282 (51.0 %)	9,606,564 (49.0 %)	274,615 (1.4 %)	3,978,379 (20.3 %)

<sup>a</sup>Percentage of regional biome extent is in parentheses, except in the column totals (last row), where percent refers to the global biome extent. Differences in the composition, spatial extent, temporal scale, and quality of the available data make it difficult to quantitatively compare rates of deforestation and selective logging. They are listed here to provide a general global perspective on the magnitude of reported or detected contemporary changes among these land-use processes

<sup>b</sup>Forest cover in 2005 calculated as 2000 forest cover minus losses from 2000 to 2005 with data from Hansen et al. (2008). Percent forest cover is based on percent within each 500 m grid cell, followed by conversion to vector format for global calculations

<sup>c</sup>Calculated from Hansen et al. (2008)

<sup>d</sup>Logging does not represent actual harvested trees, but rather regional forest areas in which timber operations occur

manipulated landscapes are susceptible to small island effects such as the loss of species diversity through unsustainable coexistence in shrinking patches, loss of population genetic diversity through restricted migration and shrinking population size, and nonnative and disturbance-adapted species invasions that alter community diversity and successional pathways. Strong edge effects in forest patches increase tree mortality of drought-sensitive species and from physical exposure to increased winds that cause blow down. Deposition of dust and aerosols rich in N and P from surrounding agriculture and development alters plant growth rates. Increased evaporation, decreased soil moisture, and the accumulation of litter increase susceptibility to fires, and, indeed, contemporary fire occurrence in tropical forests is largely associated with forest edges (Cochrane 2003). Nevertheless, these forest mosaics are the future of tropical regions, and thoughtful management can benefit agriculture as well as preserve forests and their ecosystem services that contribute to water quality and global food supply (e.g., pollinators).

Stronger ENSO effects are increasing the frequency and severity of droughts, fires, hurricanes and cyclones, and flooding events. Historical records and charcoal in soil profiles show that tropical forest fires, even in wetter forests, are not unprecedented. Fire is considered endemic but rare in most tropical rain forests,

with return intervals of hundreds if not thousands of years (Cochrane 2003). Drought is a major driver of fires. The El Niño drought in the years of 1997–1998, for example, burned tens of thousands of kilometers of forest in Brazil and Borneo, and the projected drying of parts of the tropics will greatly increase forest susceptibility to fire. Recovery after hurricanes and other disturbances that primarily affect canopies is faster than recovery after disturbances that heavily disturb soils and vegetation such as bulldozing, overgrazing, and severe fires (Chazdon 2003). Recovery of aboveground biomass, species composition, and forest structure all depend on the type and severity of disturbance and its effect on soil fertility.

---

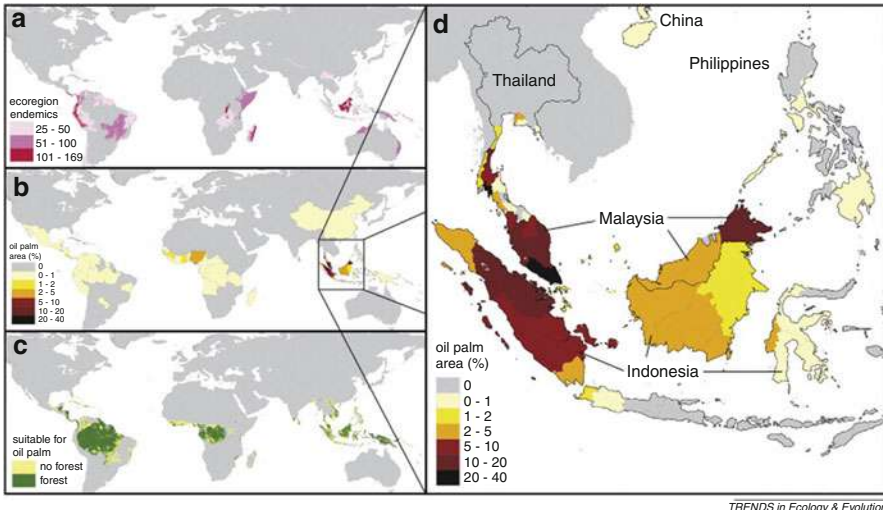
### Case Study: Oil Palm

Agriculture expansion, while necessary for supporting a healthy growing world population, is currently occurring at the expense of tropical rain forests with catastrophic consequences for global biodiversity and carbon and water cycles. Oil palm (*Elaeis guineensis*), one of the world's most rapidly expanding crop, is grown across more than 13.5 million ha in lowland tropical areas with Malaysia and Indonesia supplying greater than 80 % of global production (Fig. 6; Fitzherbert et al. 2008). With rising demand for vegetable oils and biofuels, there is no evidence that the rapid trajectory of oil palm production will abate. With a 25-year rotation cycle, oil palm monocultures are defined by uniform tree structure, low canopy, and sparse understory that support a paucity of vertebrate and invertebrate diversity. In a literature review, Fitzherbert et al. (2008) found that only 15 % of species recorded in primary forest were also present in oil palm plantations. Presence does not equate with a sustainable population, and oil palm plantation features cannot support the tropical forest fauna that tend to be of highest conservation concern. Accordingly, the predominant species in oil palm plantations tended to include non-forest specialists and nonnative invasive species – especially ants and pests. Of equal concern are the long-standing consequences of monoculture plantations such as oil palm on reduction in soil fertility, reduction in soil microbial diversity and function, and the consequent reduction in potential for native plant community recovery. Figure 6 outlines current oil palm production areas as well as areas that are suitable for oil palm production expansion – at the expense of tropical deforestation or not. Increasing demand for certified sustainable oil palm that is not produced through forest conversion is but one strategy for mitigating the impacts of oil palm on tropical forests, but it is an action that each of us can take.

---

### Future Directions

The tropical biome is undergoing significant change. Understanding the drivers and impacts of these changes will require sustained advances across multiple disciplines. Ultimately as a society, we are left asking what is the capacity of our remaining and



**Fig. 6** Global distribution of oil palm and potential conflicts with biodiversity: (a) areas of highest terrestrial vertebrate endemism (ecoregions with 25 or more endemics are shown), (b) global distribution of oil palm cultivation (harvested area as percentage of country area), (c) agriculturally suitable areas for oil palm (with and without forest), and (d) oil palm-harvested area in Southeast Asia. In (b) and (d), Brazil, Indonesia, Malaysia, the Philippines, and Thailand are subdivided by province, but other countries are not. Data are for 2006, except for the Philippines and Thailand, where 2004 data are the most recent available (Sources: (a) World Wildlife Fund (2006) WildFinder: online database of species distributions, version Jan-06, <http://www.worldwildlife.org/wildfinder>; (b, d) world: <http://faostat.fao.org>; Brazil: <http://www.ibge.gov.br/estadosat>; Indonesia: <http://www.deptan.go.id>; Malaysia: <http://econ.mpob.gov.my/economy/annual/stat2006/Area1.7.htm>; Philippines: [http://www.bas.gov.ph/downloads\\_view.php?id=127](http://www.bas.gov.ph/downloads_view.php?id=127); Thailand: <http://www.oae.go.th/statistic/yearbook47/indexe.html>; (c) forest area: European Commission Joint Research Centre (2003) Global Land Cover 2000 database, <http://www-gem.jrc.it/glc2000>; oil palm suitability: updated map from G. Fischer, first published in Fischer, G. et al. (2002) Global Agro-Ecological Assessment for Agriculture in the 21st Century: Methodology and Results, International Institute for Applied Systems Analysis and Food and Agriculture Organization of the United Nations) (Reprinted with permission Fitzherbert et al. 2008)

regrowing tropical rain forests to adapt to long-term anthropogenic and climate change and what can we do to moderate these effects while nourishing a healthy human population? Below is an incomplete list of potential research emphases.

- Continued observation of tropical plant natural history is needed to inform ecology and taxonomy, advance phylogenetic hypotheses, and expand our database of described tropical species.
- Long-term, multifactorial experiments are needed to identify the mechanisms explaining high species coexistence and identify the relative importance of altered climate (temperature and precipitation), elevated CO<sub>2</sub>, aerosol deposition, and land cover change on tropical NPP and C storage.
- Emphasis on tropical plant physiology measurements and scaling from leaf-level to stand-level processes will better constrain our estimates of NPP and tropical forest contributions to the global carbon cycle.

- Hypothesis-driven high-throughput sequencing surveys exploring metabolomic, transcriptomic, and proteomic pathways will provide insights into how tropical plants and microorganisms will respond to environmental change.
- Greater use of remote sensing imagery from satellites, airborne Light Detection and Ranging (LiDAR) data, and unmanned drones will improve monitoring of remote and large tracts of impenetrable forests. High-fidelity carbon maps such as the one generated for the entire country of Panama (Asner et al. 2013) will allow accurate tracking of disturbance and C stocks – a first step towards providing much-needed data to support economically driven climate change mitigation activities such as the United Nations Reducing Emissions from Deforestation and Forest Degradation (REDD) program.
- Expanded uses of these information-rich remote sensing datasets will improve tracking and monitoring of phenology, foliar canopy chemistry, individual species identification, and biodiversity estimates from local to regional scales. For example, spatially explicit phenological records can serve as a useful proxy for historic temperature and seasonality values.
- Fostering research synergies across disciplines and engaging stakeholders will lead to better understanding of the socioeconomic drivers of tropical deforestation and conversion, promote understanding of tropical forest ecosystem services, and put in place a framework for governance and regulation of sustainable forest product extraction and bioprospecting.

---

## References

- Asner GP, Rudel TK, Aide TM, Defries R, Emerson R. A contemporary assessment of change in humid tropical forests. *Conserv Biol.* 2009;23(6):1386–95.
- Asner GP, Mascaro J, Anderson C, Knapp DE, Martin RE, Kennedy-Bowdoin T, van Breugel M, Davies S, Hall JS, Muller-Landau HC, Potvin C, Sousa W, Wright J, Birmingham E. High-fidelity national carbon mapping for resource management and REDD+. *Carb Balance Manag.* 2013;8(1):1–14.
- Bagchi R, Gallery RE, Gripenberg S, Gurr SJ, Narayan L, Addis CE, Freckleton RP, Lewis OT. Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature.* 2014;506(7486):85–88.
- Bunyavejchewin S, Baker PJ, LaFrankie JV, Ashton PS. Stand structure of a seasonal dry evergreen forest at Huai Kha Khaeng Wildlife Sanctuary, Western Thailand. *Nat Hist Bull Siam Soc.* 2001;49:89–106.
- Chazdon RL. Tropical forest recovery: legacies of human impact and natural disturbances. *Perspect Plant Ecol Evol Syst.* 2003;6:51–71.
- Clark DA, Brown S, Kicklighter DW, Chambers JQ, Thomlinson JR, Ni J, Holland EA. Net primary production in tropical forests: an evaluation and synthesis of existing field data. *Ecol Appl.* 2001;11(2):371–84.
- Cleveland CC, Townsend AR, Taylor P, Alvarez-Clare S, Bustamante M, Chuyong G, Dobrowski SZ, Grierson P, Harms KE, Houlton BZ, Marklein A, Parton W, Porder S, Reed SC, Sierra CA, Silver WL, Tanner EVJ, Wieder WR. Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan-tropical analysis. *Ecol Lett.* 2011;14(9):939–47.
- Cochrane MA. Fire science for rainforests. *Nature.* 2003;421(6926):913–9.
- Condit R, Ashton P, Balslev H, Brokaw N, Bunyavejchewin S, Chuyong G, Co L, Dattaraja HS, Davies S, Esufali S, Ewango CEN, Foster R, Gunatillek N, Gunatillek S, Hernandez C,

- Hubbell S, John R, Kenfack D, Kiratipayoon S, Hall P, Hart T, Itoh A, LaFrankie J, Liengola I, Lagunzad D, Lao S, Losos E, Magard E, Makana J, Manokaran N, Navarrete H, Mohammed Nur S, Okhubto T, Perez R, Samper C, Hua Seng L, Sukumar R, Svenning JC, Tan S, Thomas D, Thompson J, Vallejo M, Villa Munoz G, Valencia R, Yamakura T, Zimmerman J. Tropical tree  $\alpha$ -diversity: results from a worldwide network of large plots. *Biol Skr.* 2005;55:565–82. ISSN 0366-3612. ISBN 87-7304-304-4.
- Condit R, Ashton PS, Manokaran N, LaFrankie JV, Hubbell SP, Foster RB. Dynamics of the forest communities at Pasoh and Barro Colorado: comparing two 50 ha plots. *Philos Trans Ser B.* 1999;354:1739–48.
- Condit R, Hubbell SP, Foster RB. Changes in a tropical forest with a shifting climate: results from a 50 ha permanent census plot in Panama. *J Trop Ecol.* 1996a;12:231–56.
- Condit R, Hubbell SP, LaFrankie JV, Sukumar R, Manokaran N, Foster RB, Ashton PS. Species-area and species-individual relationships for tropical trees: a comparison of three 50 ha plots. *J Ecol.* 1996b;84:549–62.
- Connell JH. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: den Boer PJ, Gradwell GR, editors. *Dynamics of numbers in populations.* The Netherlands: PUDOC, Wageningen; 1971. p. 298–312.
- Dick CW, Bermingham E, Lemes MR, Gribel R. Extreme long-distance dispersal of the lowland tropical rainforest tree *Ceiba pentandra* L. (Malvaceae) in Africa and the Neotropics. *Mol Ecol.* 2007;16(14):3039–49.
- Fitzherbert EB, Struebig MJ, Morel A, Danielsen F, Brühl CA, Donald PF, Phalan B. How will oil palm expansion affect biodiversity? *Trends Ecol Evol.* 2008;23(10):538–45.
- Gilbert GS, Webb CO. Phylogenetic signal in plant pathogen–host range. *Proc Natl Acad Sci.* 2007;104(12):4979–83.
- Hansen MC, Stehman SV, Potapov PV, Loveland TR, Townshend JR, DeFries RS, Pittman KW, Arunarwati B, Stolle F, Steininger MK, Carroll M, DiMiceli C. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proc Natl Acad Sci.* 2008;105(27):9439–44.
- Heil M, McKey D. Protective ant–plant interactions as model systems in ecological and evolutionary research. *Annu Rev Ecol Evol Syst.* 2003;34:425–53.
- Herre EA, Jandér KC, Machado CA. Evolutionary ecology of figs and their associates: recent progress and outstanding puzzles. *Annu Rev Ecol Evol Syst.* 2008;39:439–58.
- Hubbell SP, Foster RB. Diversity of canopy trees in a neotropical forest and implications for conservation. In: Sutton SL, Whitmore TC, Chadwick AC, editors. *Tropical rain forest: ecology and management.* Oxford: Blackwell Scientific Publications; 1983. p. 25–41.
- Hubbell SP. *The unified neutral theory of biodiversity and biogeography.* Princeton: Princeton University Press; 2001.
- Hutchinson GE. *The ecological theater and the evolutionary play.* New Haven: Yale University Press; 1965. p. 1–139.
- Jacoby GC. Overview of tree-ring analysis in tropical regions. *Iawa Bull.* 1989;10:99–108.
- Janzen DH. Herbivores and the number of tree species in tropical forests. *Am Nat.* 1970;104:501–28.
- Laurance WF, Camargo JL, Luizão RC, Laurance SG, Pimm SL, Bruna EM, Stouffer PC, Williamson GB, Benitez-Malvido J, Vasconcelos HL, Van Houtan KS, Zartman CE, Boyle SA, Didham RK, Andrade A, Lovejoy TE. The fate of Amazonian forest fragments: a 32-year investigation. *Biol Conserv.* 2011;144(1):56–67.
- Lee HS, Davies SJ, LaFrankie JV, Tan S, Yamakura T, Itoh A, Ashton PS. Floristic and structural diversity of 52 hectares of mixed dipterocarp forest in Lambir Hills National Park, Sarawak, Malaysia. *J Trop Forest Sci.* 2002;14:379–400.
- Leigh EG, David P, Dick CW, Terborgh J, Puyravaud JP, Steege H, Wright SJ. Why do some tropical forests have so many species of trees? *Biotropica.* 2004;36(4):447–73.
- Makana JR, Hart TB, Hart JA. Forest structure and diversity of lianas and understory treelets in monodominant and mixed stands in the Ituri Forest, Democratic Republic of the Congo. In:

- Dallmeier F, Comiskey JA, editors. Forest biodiversity diversity research, monitoring, and modeling. Paris: UNESCO, the Parthenon Publishing Group; 1998. p. 429–46.
- Malhi Y. The carbon balance of tropical forest regions, 1990–2005. *Curr Opin Environ Sustain*. 2010;2(4):237–44.
- Manokaran N, LaFrankie JV, Kochummen KM, Quah ES, Klahn J, Ashton PS, Hubbell SP. Stand table and distribution of species in the 50-ha research plot at Pasoh Forest Reserve. Kepong, Malaysia: Forest Research Institute of Malaysia; 1992.
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000;403(6772):853–8.
- Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, Drozd P. Low host specificity of herbivorous insects in a tropical forest. *Nature*. 2002;416(6883):841–4.
- Ricklefs RE, Renner SS. Global correlations in tropical tree species richness and abundance reject neutrality. *Science*. 2012;335(6067):464–7.
- Romoleroux K, Foster R, Valencia R, Condit R, Balslev H, Losos E. Especies leñosas (dap >1 cm) encontradas en dos hectáreas de un bosque de la Amazonía ecuatoriana. In: Valencia R, Balslev H, editors. *Estudios Sobre Diversidad y Ecología de Plantas*. Quito: Pontificia Universidad Católica del Ecuador; 1997. p. 189–215.
- Schongart J, Junk WJ, Piedade MTF, Ayres JM, Huttermann A, Worbes M. Teleconnection between tree growth in the Amazonian floodplains and the El Niño-Southern Oscillation effect. *Glob Chang Biol*. 2004;10:683–92. doi:10.1111/j.1529-8817.2003.00754.x.
- Sodhi NS, Brook BW, Bradshaw CJ. *Tropical conservation biology*. Oxford, UK: Blackwell; 2007.
- Sukumar R, Dattaraja HS, Suresh HS, Radhakrishnan J, Vasudeva R, Nirmala S, Joshi NV. Longterm monitoring of vegetation in a tropical deciduous forest in Mudumalai, southern India. *Curr Sci*. 1992;62:608–16.
- ter Steege H, Pitman NC, Sabatier D, Baraloto C, Salomão RP, Guevara JE, Phillips OL, et al. Hyperdominance in the Amazonian tree flora. *Science*. 2013;342(6156):1243092. doi:10.1126/science.1243092.
- Thompson J, Brokaw N, Zimmerman JK, Waide RB, Everham III EM, Lodge DJ, Taylor CM, Garcia-Montel D, Fluet M. Land use history, environment, and tree composition in a tropical forest. *Ecol Appl*. 2002;12:1344–63.
- Townsend AR, Asner GP, Cleveland CC. The biogeochemical heterogeneity of tropical forests. *Trends Ecol Evol*. 2008;23(8):424–31.
- Valencia R, Foster RB, Villa G, Condit R, Svenning JC, Hernández C, Romoleroux K, Losos E, Magförd E, Balslev H. Tree species distributions and local habitat variation in the Amazon: a large forest plot in eastern Ecuador. *J Ecol*. 2004;92:214–29.
- van der Heijden GM, Schnitzer SA, Powers JS, Phillips OL. Liana impacts on carbon cycling, storage and sequestration in tropical forests. *Biotropica*. 2013;45:682–92.
- van Schaik CP, Terborgh JW, Wright SJ. The phenology of tropical forests: adaptive significance and consequences for primary consumers. *Annu Rev Ecol Syst*. 1993;24:353–77.
- Worbes M. One hundred years of tree ring research in the tropics – a brief history and an outlook to future challenges. *Dendrochronologia*. 2002;20:217–31.
- Zimmerman JK, Everham EMI, Waide RB, Lodge DJ, Taylor CM, Brokaw NVL. Responses of tree species to hurricane winds in subtropical wet forest in Puerto Rico: implications for tropical tree life histories. *J Ecol*. 1994;82:911–22.

## Further Reading

- Cemusak LA, Winter K, Dalling JW, Holtum JA, Jaramillo C, Körner C, Leakey ADB, Norby RJ, Poulter B, Turner BL, Wright SJ. Tropical forest responses to increasing atmospheric CO<sub>2</sub>: current knowledge and opportunities for future research. *Funct Plant Biol*. 2013;40:531–51.

- Hubbell SP. Tropical rain forest conservation and the twin challenges of diversity and rarity. *Ecol Evol.* 2013;3(10):3263–74.
- Laurance WF, Sayer J, Cassman KG. Agricultural expansion and its impacts on tropical nature. *Trends Ecol Evol.* 2014;29(2):107–16.
- Lewis SL, Lloyd J, Sitch S, Mitchard ETA, Laurence WF. Changing ecology of tropical forests: evidence and drivers. *Annu Rev Ecol Evol Syst.* 2009;40:529–49.
- Molbo D, Machado CA, Sevenster JG, Keller L, Herre EA. Cryptic species of fig-pollinating wasps: implications for the evolution of the fig–wasp mutualism, sex allocation, and precision of adaptation. *Proc Natl Acad Sci.* 2003;100(10):5867–72.
- Pitman NC, Terborgh JW, Silman MR, Núñez VP, Neill DA, Cerón CE, Palacios WA, Aulestia M. Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology.* 2001;82(8):2101–17.
- Schemske DW, Mittelbach GG, Cornell HV, Sobel JM, Roy K. Is there a latitudinal gradient in the importance of biotic interactions? *Annu Rev Ecol Evol Syst.* 2009;40:245–69.
- Wright SJ. The future of tropical forests. *Ann N Y Acad Sci.* 2010;1195(1):1–27.



Russell K. Monson

## Contents

What Is a Forest? .....	275
The Climate and Phytogeography of Temperate Forests .....	275
The Geologic Origins of Temperate Forests .....	277
Temperate Forests and the Concept of Ecological Succession .....	278
Temperate Forest Carbon Cycling .....	281
Temperate Forest Net Primary Productivity .....	282
Temperate Forest Nitrogen and Phosphorus Cycling .....	284
Temperate Forest Water Cycling .....	285
Plant Functional Traits in Temperate Forest Trees .....	287
Temperate Forests and Disturbance .....	289
Future Directions .....	294
References .....	295

---

## Abstract

- “Temperate” forests occur at the mid-latitudes, between 23.5 and 66.5° N and S, where they cover approximately 20 % of the available land area and are characterized by distinct seasonal climate cycles.
- Temperate forests are dominated by plants with a woody, treelike growth form, and they produce relatively closed canopies (60–100 % areal canopy coverage). Temperate forests occur across a broad range of climate zones, including those with moist, warm summers (e.g., deciduous forests in North America and Europe) and dry, cool summers (e.g., montane and subalpine forests in North America, South America, and Europe).
- Temperate forest ecosystems exhibit carbon to nitrogen (C:N) ratios that are higher (often >100–200) than other temperate-latitude ecosystems, due to

---

R.K. Monson (✉)  
School of Natural Resources and the Laboratory for Tree Ring Research, University of Arizona,  
Tucson, AZ, USA  
e-mail: [russell.monson@colorado.edu](mailto:russell.monson@colorado.edu)

the exceptionally high C:N ratio of wood (often  $>300$ ). As a result, temperate forests are capable of storing high quantities of carbon that are assimilated from the reservoir of atmospheric  $\text{CO}_2$ .

- Classic, historical concepts in ecology, such as “succession,” have been developed from studies of temperate forest ecosystems. Forest succession refers to decadal-scale transitions in community composition. Each shift in community composition causes changes in the forest microenvironment, which in turn causes further changes in community composition. Traditionally, this pattern of progressive change in community composition and associated feedbacks to forest microenvironment was viewed within a highly deterministic framework. More recently, ecological concepts, such as “gap theory,” have emerged from the older concepts of succession and have been developed with greater emphasis on stochasticity. Both succession and gap theory has contributed greatly to our understanding of the causes of natural and anthropogenic changes to the species composition of temperate forest ecosystems.
- Nitrogen and phosphorus (N and P) are cycled through temperate forest ecosystems through a process of coupled recycling involving serial relationships between plants and soil microorganisms. N or P that is deposited to the soil through litter production is transformed from organic to inorganic forms through microbial mineralization, producing nitrate and phosphate ions, which can then be re-assimilated by plants and used to construct new organic biomass. Leaching of phosphate and nitrate from forest soils (especially nitrate in temperate forests) prior to re-assimilation by plants represents an important nutrient loss process and often limits forest biomass production.
- Root-fungal symbioses, called mycorrhizae, are well developed in temperate forest ecosystems. The hyphal biomass from the fungus radiates from associated roots and increases the capacity for trees to capture nitrate and phosphate prior to leaching and, in some cases, allows trees to take up organic nitrogen (such as small proteins or single molecules of amino acids). The acquisition of organic forms of nitrogen (and to some extent phosphorus) “short-circuits” the conventional form of biogeochemical cycles (alternating between plants and microbes) and increases the efficiency of nutrient retention in the ecosystem.
- Most water that is cycled through forests is used to sustain a favorable energy balance. Evapotranspiration from forests facilitates the loss of heat that is absorbed as net radiation (from the sun and sky) and returns water to the atmosphere, thus sustaining the terrestrial water cycle.
- Trees in temperate forests (especially in North America and Europe) have been exposed to increasing physiological stress in recent decades due to the increased frequency of drought and high temperatures. These stresses have the potential to reduce forest growth and may be responsible for the observed weakening of forest carbon sinks globally. Climate-induced stress, in turn, exposes temperate forests to an increased frequency of epidemic insect outbreaks and associated high rates of herbivory, as well as shorter fire return cycles. The combination of abiotic and biotic stress is likely responsible for an increase in observed mass tree mortality in temperate forests of the Northern Hemisphere.

- Greater frequencies of alternations between years with extreme climates (e.g., wetter-than-average years followed by drier-than-average years) have the potential to convert less damaging surface fires into more damaging crown fires due to the buildup of beneath-canopy fuels and greater connectivity of lower-elevation grasslands that border higher-elevation forests, during wet years, followed by greater ignition potential during dry years.
- Given changes in the Earth's climate system that increase the threat to fire- and insect-induced mass tree mortality in temperate forests, effective management of these ecosystems has become even more urgent and important to responsible stewardship of our natural resources. Future management efforts should be designed on a solid foundation of scientific knowledge about forest succession, forest biogeochemistry, and the natural relations of forests to fire return cycles and cycles of higher and lower levels of insect herbivory.

---

## What Is a Forest?

The term “forest” has been used since at least the Middle Ages, when William the Conqueror consolidated much of the knowledge about his newly acquired lands in the Domesday Book of 1086 CE (Common Era). Royal forests, as listed in the Domesday Book, referred to unbounded lands intended to raise wild animals that could be hunted by the monarch and other members of the royal family. Forests were not classified according to ecological or botanical attributes, but rather as legal entities afforded protection by laws and management. Forests at this time included grasslands, woodlands, heathlands, and even agricultural fields.

In more recent times, the term “forest” has been associated with woodlands, and most dictionary definitions include reference to a “high density of trees.” The US National Vegetation Classification Scheme, which is produced through oversight by the US Federal Geographic Data Committee (an interagency committee led by representatives of the US Geological Survey), distinguishes “forests” from “woodlands.” Forests are areas with trees forming overlapping crowns with 60–100 % areal coverage. Woodlands are more open, with 25–60 % crown coverage. Even with these rather precise definitions, however, ecologists will often use the term more loosely, for example, in describing the great kelp “forests” in the coastal oceans of temperate and polar regions.

---

## The Climate and Phytogeography of Temperate Forests

In this chapter, I will develop the concept of “forests” according to a deeper set of ecological attributes, and I will focus on temperate forests. Temperate forests occur between 23.5 and 66.5° N and S latitudes, extending from subtropical biomes to boreal biomes – covering the so-called “mid-latitudes.” Forests cover approximately 20 % of the available land area within these mid-latitude bands. Unlike tropical forests, temperate forests occur in climate zones with distinct seasonality.

Summers tend to be warm but moist enough to support the relatively high water demands of the tree life-form, and winters tend to be cool to cold, but also moist. At higher elevations, winter moisture is deposited as snow. Limited access to liquid water and cold winter temperatures force trees in higher-elevation temperate forests to minimize metabolic activity until the time of spring or summer snowmelt. Even evergreen, coniferous trees in these ecosystems tend to downregulate their metabolic activities in the winter to levels just sufficient to sustain basal respiration. Thus, in high-elevation, temperate coniferous forests, winter forest carbon balances are characterized by deficits, causing the forests to be net carbon sources, at least seasonally. In lower-elevation coastal forests winters are cool to cold, but moisture continues to be deposited as rain. In these coastal “rainforests,” evergreen trees remain metabolically active during the winter, and wintertime photosynthesis can represent a significant fraction of annual net primary productivity.

Temperate forests include broadleaf and needleleaf tree forms. Temperate forest trees tend to have long generation times, lasting multiple decades, compared to plants in other mid-latitude ecosystems (though temperate alpine plants can also have multi-decadal life spans). As a result, forests tend to migrate slowly across the landscape in response to climate changes. This creates disequilibrium between climate change and forest distribution. Emerging from the Last Glacial Maximum, and midway into the Holocene Era, poleward forest migrations in the Northern Hemisphere have been estimated as 2–2.5 km yr<sup>-1</sup> (based on pollen records; Davis 1989), which is considerably slower than the 6–17 km yr<sup>-1</sup> estimated for non-tree species during the current, Anthropocene warming (Parmesan and Yohe 2003; Chen et al. 2011). The slow migration of temperate forests in response to a rapidly changing climate poses interesting questions as to how temperate forest ecosystems will adjust to future human-influenced climate regimes.

Forests tend to have a unique nutrient stoichiometry, particularly with regard to C:N ratios; because of the high C:N ratio of wood (often 300 or higher), whole-tree C:N ratios (wood, leaves, and roots) tend to be between 100 and 200 for temperate forest trees (Norby et al. 1999). Herbaceous plants often exhibit C:N ratios that are less than 100 and often less than 50 (Lebreton and Gallet 2007). Because of their high C:N ratios, temperate forests account for much of the net carbon dioxide assimilated from the atmosphere during each year’s growing season. Based on the Food and Agriculture Organization of the United Nations, as of the year 2000 CE, temperate broadleaf forests and temperate evergreen forests covered approximately 400 million and 100 million ha of the Earth’s surface, respectively. Taking into account all temperate forests on the globe, it is estimated that net primary productivity for this biome type is ~8 Pg C yr<sup>-1</sup> (Saugier et al. 2001). This rate of carbon uptake is approximately 13 % of the total global rate of photosynthetic carbon sequestration and similar to the total annual anthropogenic CO<sub>2</sub> emissions due to fossil fuel combustion. Thus, temperate forest ecosystems represent a vital component of the Earth’s carbon budget and must take a central role in discussions of global carbon cycle management.



**Fig. 1** Proposed landscape for Devonian temperate forests 385 mya. The forest is composed of cycads, tree ferns, and Aneurophytaleans and is likely to be one of the oldest temperate forests yet discovered. This landscape drawing was based on fossils uncovered in Schoharie County, New York, and the drawing was produced by Frank Mannolini of the New York State Museum, Albany, New York (Reproduced here with their kind permission. Copyright remains with the New York State Museum)

---

## The Geologic Origins of Temperate Forests

In 2012, discovery of a fossil forest in Schoharie County, New York, near Albany, established a new record for the oldest forest, 385 mya, midway through the Devonian geologic period (Fig. 1). Mid-latitude forests of this era were wet, swampy, and generally warmer than in the Modern (Common) Era, dominated by *Eospermatopteris* trees and woody vines in the extinct, spore-bearing Aneurophytalean group. Forest sub-canopy cover consisted of lycopsids (club mosses), which exhibited a treelike growth form. Tree ring analysis from fossil *Archaeopteris* trunks, a Devonian derivative of the Aneurophytalean group, has shown that despite a generally warmer mean climate, annual climate cycles in these ancient New York forests were seasonal, with distinct summer-winter transitions in growth. Thus, the earliest temperate forests appear to have emerged within 75 million years after appearance of the earliest known terrestrial plants in the mid-Ordovician. In terms of geological time, it did not take long for terrestrial plants to move from simple small growth forms to more complex and large treelike growth forms.

Angiosperm trees, such as those that dominate Holocene temperate forests, most likely evolved approximately 90 mya, during the late Cretaceous period. This would have included taxa in the Ulmaceae (including the elms) and Fagaceae (including the oaks, beeches, and chestnuts), as well as the Nothofagaceae (including trees in the genus *Nothofagus*, which dominate many Southern Hemisphere temperate forests). The original taxa in these groups (with the exception of *Nothofagus*) most likely evolved in tropical or subtropical forests and migrated

northward to establish temperate forests during Cretaceous warming. In addition to being comprised of these angiosperm clades of tropical origin, the earliest temperate forests in the Northern Hemisphere most likely included northerly derived evergreen and deciduous gymnosperm species such as *Larix* and *Taxodium*. Many of the mixed temperate forests in the Northern Hemisphere retain this combination of angiosperms and gymnosperms in the Modern Era, as exemplified by the beech-spruce forests in Europe and the oak-pine forests in eastern North America.

---

## Temperate Forests and the Concept of Ecological Succession

In general terms, ecological succession refers to the decadal-scale transitions that occur in plant community composition, driven by environmental transitions (in both plant and animal components of the environment) that occur as a direct feedback from the coupled changes in communities. That is, changes in community composition cause changes in the local environment, which in turn cause further changes in community composition. From 1900 until the latter part of the twentieth century, the concept of ecological succession represented a central organizing principle in the field of plant community ecology. Observations in temperate forests had a major role in the development of this concept. In an address that was read to the Middlesex Agricultural Society in 1860, Henry David Thoreau established forests as the iconic subject of successional theory. He stated:

I have no time to go into details, but will say, in a word, that while the wind is conveying the seeds of pines into hard woods and open lands, the squirrels and other animals are conveying the seeds of oaks and walnuts into the pine woods, and thus a rotation of crops is kept up. I affirmed this confidently many years ago, and an occasional examination of dense pine woods confirmed me in my opinion. It has long been known to observers that squirrels bury nuts in the ground, but I am not aware that any one has thus accounted for the regular succession of forests.

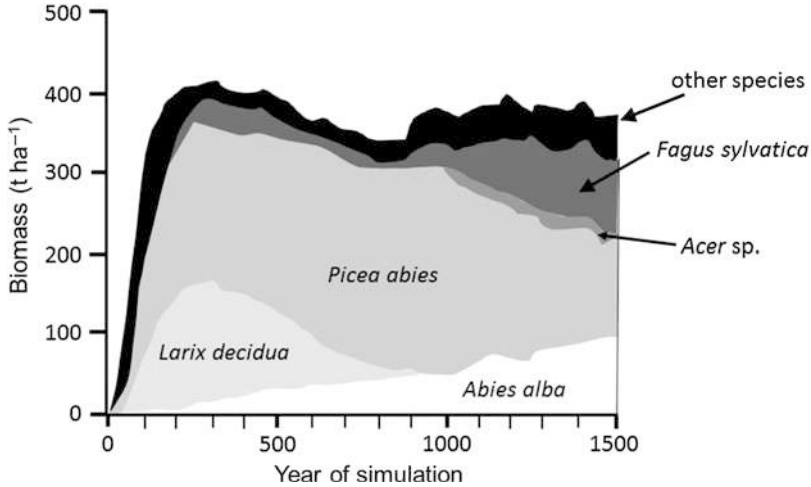
The title of Thoreau's essay was *The Succession of Forest Trees*. Thoreau's essay focused on the role of animals as they move among plant community types in establishing the seed bank for future changes in forest community composition. During the early part of the twentieth century, Frederic Clements and Henry Gleason debated openly about the nature of plant community changes and in particular forest succession. Clements viewed ecological succession as deterministic, a time-dependent process of species replacements, responding to changes in forest microclimate and soil fertility, which ended in the so-called climax community. Successional sequences could occur on newly developed substrates as they were mineralized, such as volcanic ash or rock (primary succession), or they could be reset following disturbances to established communities, such as following a stand-replacing fire, or clear-cut logging (secondary succession). In either case, the climax community could be predicted through observation of other "stable" communities in the same climate and geographic regimes. The climax state was predominantly controlled by climate, soil fertility, and their interactions with the

established adaptive traits of plants. Clements viewed plant assemblages as being consistently repeated from site to site within a climate zone, with time-dependent transitions among assemblages occurring in synchrony and with internally controlled articulation, similar to the ontogenetic transitions that occur in a maturing organism. In fact, Clement's concept of a forest, with its deterministic pattern of community development, was often referred to as "superorganismic." In contrast, Henry Gleason, having studied communities across ecological gradients, had concluded that community assemblage dynamics are not entirely predictable. Rather, species associate with one another on an individualistic basis. In Gleason's view, there is no inherent organizing force that orchestrates a predictable outcome to succession. Ecological communities cannot be viewed in terms of organismic ontogeny.

Established temperate forest communities were viewed as iconic examples of the Clementsian ecological climax. The fact that removal of a forest would lead to eventual establishment of a similar forest was cited frequently as the basis for climax ecology. In the Clementsian view, pioneer species adapted to open habitats (high light, low soil moisture and fertility, and low atmospheric humidity) would reclaim a site shortly after disturbance (in the case of secondary succession) and "prepare" the site for the second so-called sere (successional assemblage). Each successive sere would be dominated by species with progressively greater tolerance of shade, higher soil moisture, and humidity, and they would be able to effectively compete for, and recycle, soil nutrients, eventually leading to a climax sere that is self-perpetuating. Certain taxa, such as pines, were commonly identified as components of early successional seres, whereas oaks and beeches were identified as components of late successional seres. Succession is still, to this day, discussed as a key ecological principle, particularly when referring to community change, though it is now discussed within the context of nondeterministic processes that can be altered depending on each site's history and access to specific plant taxa (e.g., with phylogenetic context). This new form of the concept of succession allows for stochastic dynamics such as occurs with varying degrees of disturbance, historic patterns of biogeography, climate change, and human intervention.

In the late 1960s and early 1970s, ecologists began considering forest community dynamics at smaller scales, capable of capturing some of the individualistic nature of forest succession as envisioned by Gleason. This led to a theoretical framework known as "gap theory" and a group of models known as "forest gap models." Within this theoretical framework, disturbances caused by the mortality of individual, large trees created an opening in the forest canopy and set off a sequence of localized successional responses. The key attribute of forest gap models is that they track individual trees from birth to death in small patches of the forest (e.g., 1 ha in area). Forest gap models have now been expanded to describe dynamic processes in entire forest stands, mostly from a process (growth, photosynthesis, allocation) perspective. Once parameterized for past or current environmental conditions, these types of models can be used to predict forest successional patterns (Fig. 2).

Much of our knowledge about forest succession in specific regions of the world has been constructed from pollen and fossil (both micro and macro) records, especially from peat lands. For example, such approaches have led us to the



**Fig. 2** Forest succession patterns predicted for a native beech-dominated forest near Davos, Switzerland, as provided by the forest dynamics model, ForClim, ver. 2.9. The pattern demonstrates the expected shifts in community composition during secondary succession, with the deciduous coniferous species, *Larix decidua*, emerging as a pioneer species, and giving way to eventual dominance by *Abies alba* (white fir) and *Fagus sylvatica* (common beech). The model simulation begins with the current climate at Year 0 and progresses through a series of future climate scenarios for a period of one-an-a-half millennia (Redrawn from Bugmann (2001))

conclusion that in Central Europe, successional patterns are typified by initial stands of hazel (*Corylus* sp.), followed by oak (*Quercus* sp.), linden (*Tilia* sp.), and alder (*Alnus* sp.), which eventually give way to the shade-tolerant beech (*Fagus sylvatica*) and Norway Spruce (*Picea abies*). At higher elevations, spruce domination of forest stands are often preceded by larch (*Larix decidua*), and areas of domination by pine can occur, especially on thinner, sandier soils. The successional sequence of European temperate forests, especially in lowland forests, was accelerated during the Holocene by anthropogenic deforestation, especially during the Middle Holocene (7,000–5,000 years ago) when the climate of the Northern Hemisphere exhibited an extended warm-temperature anomaly. Cutting of forests for energy and shelter tended to shift oak-dominated early seres to beech-dominated later seres more quickly. Once established, the shade-tolerant beech did not permit reestablishment of mixed forests dominated by oak. In fact, the dark understory environment beneath beech forests during the nineteenth century led to the naming of the Black Forest region in southwestern Germany. Later deforestation of beech forests in Europe (e.g., during the nineteenth century) provided the opportunity for management through the planting of spruce, a faster growing species, and therefore more valuable for silviculture. Thus, many of the old-growth beech forests that dominated European forests since the Middle Holocene have been replaced by Norway Spruce stands. There are now efforts underway in some parts of Europe to reestablish beech as a climax species.



## Temperate Forest Carbon Cycling

Like all ecosystems on or near the Earth's surface, forests achieve their structural and functional complexity at the expense of solar energy flowing through the photosynthetic processes of autotrophic organisms. Photosynthetic energy capture is used to produce biomass, which is primarily composed of the elements carbon and oxygen, obtained from atmospheric carbon dioxide, and hydrogen, obtained from water. The assimilation of  $\text{CO}_2$  in plant chloroplasts is catalyzed by an enzyme known as ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which is required in relatively high concentrations in the chloroplast. In the leaves of forest trees, up to 35 % of the nitrogen that is present can be accounted for in Rubisco protein. In pure crystalline form, Rubisco-active sites exist at a concentration of approximately 10 mM. Although some storage proteins exist at concentrations this high, Rubisco is unique in being a catalytic protein present at such high concentrations. In fact, Rubisco is the most abundant protein on earth, and most of the nitrogen required by plants is for the purpose of producing this protein and capturing atmospheric  $\text{CO}_2$  during photosynthesis. In addition to having an adequate catalytic mechanism for the capture of  $\text{CO}_2$ , plant leaves must maintain an energy balance that allows leaf temperatures to remain favorable for catalytic function. The primary water usage for plants is not to provide substrates for the photosynthetic assimilation of  $\text{CO}_2$ , but rather to cool leaves through evaporative latent heat loss, allowing them to sustain temperatures favorable for metabolism. These physiological requirements for plant function provide the foundation for the cycling of carbon, nitrogen, and water through forest ecosystems, and it is on these three biogeochemical cycles that I will focus the next few paragraphs. Temperate forest carbon cycle budgets can be succinctly defined as the balance between gross primary production (GPP) and ecosystem respiration ( $R_e$ ), with the difference between these two fluxes representing net ecosystem production, or NEP. Thus, mathematically  $\text{NEP} = \text{GPP} - R_e$ . Net ecosystem production reflects the molar equivalent of carbon that is "sequestered" within the ecosystem, initially as autotrophic biomass and eventually as plant litter and soil organic matter. Thus, in studies of the capacity for ecosystems to extract  $\text{CO}_2$  from the atmosphere, it is NEP that is most relevant. Gross primary production represents the molar equivalent of carbon extracted from the atmosphere through photosynthesis. Rubisco is the enzyme that catalyzes the entry of atmospheric  $\text{CO}_2$  into GPP, and this catalysis is controlled by the availability of a solar photon flux to drive the energetics of photosynthesis, the acquisition of soil nitrogen to produce Rubisco and other enzymes associated with photosynthesis, the uptake of soil water and exposure to atmospheric humidity that will be adequate to sustain a favorable leaf energy balance, and a stomatal conductance that facilitates the inward diffusive flux of atmospheric  $\text{CO}_2$ . Thus, GPP is connected to the forest environment through many different abiotic variables.

Ecosystem respiration reflects contributions from both the heterotrophic decomposition of soil organic matter (i.e., the products of "old" GPP) and the oxidation of compounds produced through recent photosynthetic activity

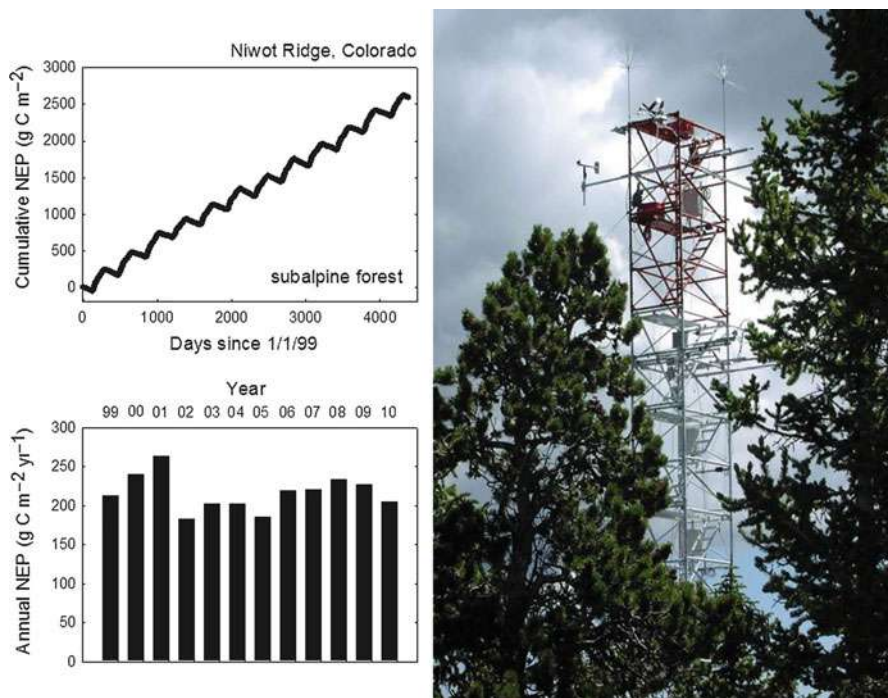
(i.e., the products of “new” GPP). The latter component of  $R_e$  is often partitioned further into the  $\text{CO}_2$  efflux from aboveground plant tissues (often included as a component of net primary productivity, NPP) and that from roots and soil microorganisms that are symbiotically associated with roots (e.g., mycorrhizal fungi and rhizospheric bacteria). It is difficult to clearly distinguish the dependencies of soil respiration on recent GPP versus older soil organic matter, as respiratory substrates exist along a continuum of ages. In some studies, the components have been distinguished through the experimental “girdling” of all trees in a forest stand or through large-scale labeling of photosynthetic products in forest trees using isotope tracers (Högberg and Read 2006). Tree girdling essentially chokes the flow of photosynthetic products from the leaves (or needles) to the soil, thus eliminating the soil respiration component linked to recent photosynthetic activity. Labeling of the forest with  $^{13}\text{CO}_2$  has been accomplished with giant tents to contain the applied label, followed by time-dependent tracing of the paths taken by labeled photosynthetic compounds and the kinetics by which  $^{13}\text{CO}_2$  is released from the labeled photosynthetic products through the processes of plant and microbial respiration. In the final accounting, all  $\text{CO}_2$  released by ecosystem respiration is dependent on the rate of GPP, as this determines the rate by which carbon substrates enter the ecosystem and are used by plant or microbial cells as respiratory energy sources. The rates, at which these substrates are utilized, however, are subject to modification according to abiotic factors, such as temperature and moisture availability.

The net carbon uptake of forest ecosystems is typically measured using towers that extend above the canopy and have instruments attached that are capable of measuring the statistical covariance between the vertical wind speed (up versus down) and the  $\text{CO}_2$  concentration in the atmosphere near the canopy surface. This is the so-called eddy covariance (or eddy flux) approach (Fig. 3). Using this approach, a continuous record of the cumulative rate of biological carbon sequestration can be measured for the forest, including all carbon stored in the trees and soil. Using this approach, one can study the effect of climate variation on forest carbon uptake across long (decadal) time scales. It is through this type of study, combined with computer models of ecosystem processes, that insight is being gained into the feedbacks between climate change and forest carbon uptake.

---

## Temperate Forest Net Primary Productivity

Temperate forest ecosystems are capable of sequestering carbon from the atmosphere at relatively high rates due to their high amounts of leaf surface area. However, in order to achieve high rates of carbon assimilation, forest leaf tissues must be capable of operating near their physiological optima. In temperate coastal forests, where winter precipitation often falls as rain, forests can retain the capacity for high rates of net primary productivity through the winter; this is typified by the coastal forests of the Pacific Northwest in the USA. The high amounts of rainfall and cool, but above-freezing, winter temperatures allow some coastal forests on the western slope of the Cascade Mountains in Oregon to exhibit net primary



**Fig. 3** *Left panel.* A 12-year record of net ecosystem productivity (NEP) measured in a subalpine forest in Colorado, USA. The “sawtooth” record shows seasonal variation in the cumulative NEP, with decreases in NEP shown for winter (the forest continues to respire but GPP is near zero) and increases in NEP shown for the growing season (GPP exceeds  $R_e$  resulting in forest carbon sequestration). In the *lower left panel*, the annual sum of forest carbon sequestration is shown for the 12-year time series. *Right panel.* A picture of the flux tower used to measure NEP from the Niwot Ridge subalpine forest. Instruments near the top of the tower record the turbulent fluxes of  $\text{CO}_2$ , and a profile of mean  $\text{CO}_2$  concentration measurements is made along the length of the tower to account for  $\text{CO}_2$  that is retained within the canopy below the turbulent flux instruments. This technology is often referred to as recording the “eddy flux” and “ $\text{CO}_2$  storage flux,” respectively, and the sum of these values provides an estimate of NEP

productivities approximating  $1 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ , and these high rates of productivity are reached before 30 years of secondary regrowth following logging (Van Tuyl et al. 2005). In contrast, forests on the drier and colder eastern slope of the Cascade Mountains reach maximum net primary productivities of approximately  $0.3 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ , and these rates are only achieved after 80–100 years of forest regrowth. For reference, oak-hickory forests in the southeastern USA exhibit annual average net primary productivities of approximately  $0.8 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ , and maple-beech forests in the northeastern USA exhibit annual net primary productivities of approximately  $0.6 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ . Net primary productivities for European mixed hardwood forests range from  $0.6$  to  $1.1 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ . Net primary productivities in Southern Hemisphere temperate forests are in the same range as the median values for their Northern Hemisphere counterparts ( $0.3$ – $0.5 \text{ kg Cm}^{-2} \text{ yr}^{-1}$ ).

Overall, temperate forests represent relatively large carbon sinks. Even compared to tropical forests, temperate forests can store large quantities of carbon. Past estimates of tropical forest net primary forest productivity have been highly variable among sites and years but have generally fallen within the range 0.2–2 kg m<sup>-2</sup> yr<sup>-1</sup> (Clark et al. 2001). Thus, while temperate forest productivity is within the lower range of that estimated for tropical forests, it can be as high as 50 %, depending on the forest site and year of consideration.

---

## Temperate Forest Nitrogen and Phosphorus Cycling

Nitrogen and phosphorus are crucial elements for the sustenance of plant metabolism, being required for protein and nucleic acid production, in the case of nitrogen, and for the production of nucleic acids, energy-rich adenylates, and membranes, in the case of phosphorus. Nitrogen and phosphorus enter temperate forest ecosystems principally through the chemical weathering of inorganic mineral surfaces, in the case of phosphorus, and through biotic fixation, in the case of nitrogen. Once nitrogen or phosphorus are incorporated into ecosystems from these primary sources, they can be recycled through litter deposition and subsequent microbial mineralization of organic N and P compounds back to inorganic forms, such as nitrate and phosphate, which can then be re-assimilated by plants and used to construct new organic biomass. Thus, plants and soil microorganisms cooperate sequentially in the conventional forms of the nitrogen and phosphorus cycles.

Although the serial nature of these cycles makes them appear as orderly processes in which plants and microbes interact in a type of cooperativity, there is actually considerable competition within the interaction. Forest trees have evolved efficient ways for roots to locate patches of nutrient-rich soil and effectively assimilate those nutrients before they are leached to deeper soil layers or captured by neighboring trees and associated microorganisms. Similarly, soil microorganisms have evolved efficient means to incorporate nutrients into their biomass, a process known as nutrient immobilization. Generally, competition between trees and microbes is limited because the roots of trees tend to be specialized for the uptake of inorganic forms of N and P, the products of microbial mineralization. However, recent research has shown that this conventional wisdom is too simple to explain processes in many ecosystems.

In some forests, especially those in which microbial mineralization is slowed by abiotic constraint (e.g., low-temperature and/or short growing seasons) or in which microbial immobilization of N is encouraged by high soil C/N ratios, plants can effectively compete with soil microorganisms for the uptake of amino acids or small proteins, thus bypassing the conventional sequence of plant-microbe-plant in forest nitrogen cycles (Lipson and Näsholm 2001). In some boreal forest ecosystems, the uptake of organic nitrogen occurs in both tree and understory species and appears to be the dominant mode of N uptake for plants. Much of the organic N uptake by forest trees is facilitated by mycorrhizal associations with fungi, with trees passing organic C to hyphae and hyphae passing organically derived N to roots (Lambers et al. 2008).

This mutual exchange allows both the tree and fungus to sustain favorable C:N ratios. However, the relationship between roots and fungi in the mycorrhizal exchange of C and N is mitigated to some extent by the availability of soil N and the requirements for fungi to immobilize a required minimal fraction. When soil N is relatively more abundant, fungal transfer of organic N to tree roots is favored, but when soil N is relatively less abundant, immobilization by mycorrhizal fungal hyphae is favored. Discovery of the use of organic N by both plants and soil microorganisms has changed our conventional view of forest nitrogen cycles. Whereas plant N nutrition was once viewed as being dependent on microbial mineralization, it is now recognized that plants and microorganisms interact directly, in both symbiotic and competitive ways, to partition organic soil N. These interactions will continue to be the focus of forest N cycling for years to come.

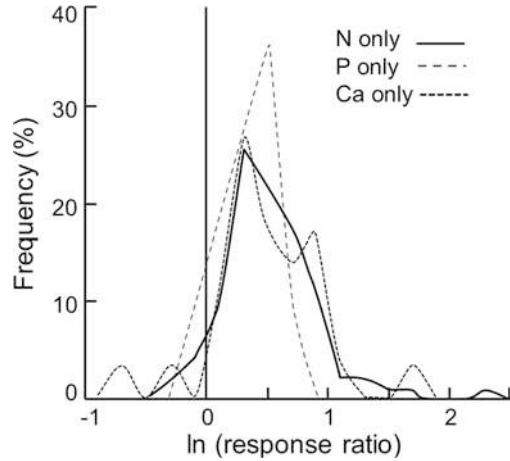
Ignoring the obvious constraints by low soil temperatures and limitations to microbial biomass, conventional wisdom holds that the maximum rate of primary production in temperate forests is ultimately limited by the availability of soil N (Reich et al. 1997). Other factors, such as soil water limitations, high air temperatures, cloudy weather, and low atmospheric humidity, can limit rates of productivity over short time scales, but over longer, decadal time scales, soil N availability will set a clear upper limit on NPP. The constraint of soil N limitation has placed selective pressures on temperate forest trees to evolve efficient rates of N recycling between the soil and plant and to retain and store N in plant tissues prior to seasonal or multi-seasonal leaf senescence. The question as to what, in turn, determines the rate of soil N availability has been addressed in conceptual models (Vitousek and Field 1999). One prominent limitation that has the potential to determine rates of N fixation in terrestrial ecosystem (and aquatic ecosystems, for that matter) is P availability. N-fixing organisms are often limited by P, and P weathers from parent minerals at relatively slow rates. Thus, a cascade of controls can be proposed for the long-term limitation of temperate forest productivity, extending from P limitations to N fixing pioneer species, N limitations to later successional species, and ultimately light limitations to carbon fixation rates as canopies close. The role of multiple nutrient constraints in limiting primary productivity is demonstrated in the meta-analysis of over 200 fertilization studies in temperate deciduous forests conducted by Vadeboncoeur (2010) (Fig. 4). Most of the studies included in that analysis showed positive growth responses to the addition of phosphorus, calcium, and nitrogen. These types of results indicate that the nutrient limitations to primary productivity in forests are complex in their nature and interactions. It may be too simple to rely on statements in the conventional wisdom, such as “nitrogen limits productivity in temperate forests and phosphorus limits productivity in tropical forests.”

---

## Temperate Forest Water Cycling

Depending on climate, soil type and local topography, and vegetation characteristics, some fraction of precipitation and water runoff from upslope areas will be transported back to the atmosphere through evaporation directly from the soil

**Fig. 4** Frequency of response ratios for a meta-analysis of 208 fertilization studies of North American deciduous forests. The “response ratio” is the ratio of net primary production in the presence of an experimental fertilization treatment relative to control plots with no treatment. The *vertical line* indicates a response ratio of 1.0 for reference (Redrawn from Vadeboncoeur (2010))



surface and transpiration from leaves (and branches to a lesser extent). The combined evapotranspiration from a forest is driven by energy inputs and biological attributes of the vegetation – a truly “biophysical” process. Energy that is absorbed from the sun and atmosphere by a forest must be partitioned into various energy loss processes, or it will contribute to an increase in forest surface temperatures. This is the nature of thermodynamics and the requirement for conservation of energy. The energy loss mechanisms that are available to forests include radiative heat loss (according to Kirchhoff’s law), sensible heat loss (through conduction of heat to the atmosphere or deeper soil layers), or latent heat loss (through evapotranspiration). The loss (through photosynthesis) or gain (through respiration) of energy is minimal compared to the processes of reradiation, sensible and latent heat loss. The tendencies for a forest to lose heat through sensible or latent transfers are to some extent, mutually exclusive. As a forest loses latent heat, its surfaces will cool, which in turn reduces the capacity for radiative heat loss and, assuming that canopy surfaces are warmer than the atmosphere, sensible heat loss. As a forest canopy conducts sensible heat to the atmosphere, less energy will be available to drive latent heat loss, and radiative heat loss will decrease. During periods of ample soil moisture, evapotranspiration rates are likely to be highest, and sensible heat losses will concomitantly decrease in importance. In contrast, during periods of drought, latent heat losses will be more limited, and the importance of sensible heat loss is likely to increase.

Temperate forest canopies, especially at high elevations, can have significant influences on ecosystem hydrology and the delivery of water resources to the watersheds that support human communities. Forest canopies intercept and retain a significant fraction of rain, especially during events with smaller rain drops and lesser drop velocity. In those cases, intercepted precipitation can be directed back to the atmosphere through evaporation from leaf surfaces, thus reducing canopy throughfall and decreasing delivery to the soil. During the winter, canopies reduce snowpack depth within the forest, thus storing less water for subsequent melt and

runoff in the spring. The negative effect of canopies on beneath-canopy snowpack occurs because snow that is intercepted by canopies can sublimate directly to the atmosphere and thus the snow never reaches the soil. Sublimation is the conversion of water directly from the solid to the vapor phase. In general, snowpacks in forested areas are up to 40 % lower than those for neighboring open areas (Varhola et al. 2010).

In high-elevation forests, canopies also affect snowmelt rates in the spring through influences on the snow energy balance. Canopies block shortwave solar radiation from reaching beneath-canopy snowpacks, emit long-wave (thermal) radiation to the snowpack, reduce near-surface wind speed and associated sensible heat transfer to the snowpack, and deposit darkly colored leaves and branches to the snow surface, thus decreasing surface albedo. Some of these influences increase the flux of energy to the melting snowpack (thermal radiation and decreased albedo), and some decrease it (interception of solar radiation and decreased sensible heat transfer). Considering all of these influences together, the net effect of a forest canopy is to reduce the energy available to melt snow, often by up to 70 % (Varhola et al. 2010). These complex influences of forest canopies on the timing and rate of snowmelt have become especially relevant recently given that large stands of pine forest have been killed by mountain pine beetles in Western North America. Many of the beetle-infested regions are also regions that direct water to streams and rivers used by human communities. The Colorado River, for example, which supplies water to major metropolitan centers, such as Los Angeles, Las Vegas, and Phoenix, is largely supplied by mountain forest watersheds. In general, it appears that large-scale mortality in western pine forests increases the potential for spring snowmelt and thus delivers water earlier in the summer to watersheds and rivers (Boon 2009).

Temperate forests exhibit broad variation in the fractions of annual productivity supported by rain versus snowmelt water. Forests in coastal regions often experience climates with less seasonal variation in temperature and above-freezing winter temperatures. In those forests, rain is the predominant form of precipitation used to drive forest productivity, irrespective of season. In higher-elevation forests, snowmelt water during the early part of the growing season can become more important than rain delivered later in the growing season. In one study using the oxygen and hydrogen isotopes of xylem water extracted from tree branches, it was determined that subalpine trees in Colorado rely on snowmelt water to drive more than 50 % of their primary productivity well into the autumn and even during the late-summer rainy season (Hu et al. 2010). Apparently, the trees in this forest possess deep roots that access stored snowmelt water from deep soil layers and rely less on shallower roots capable of utilizing summer rain.

---

## Plant Functional Traits in Temperate Forest Trees

Consistent correlations between leaf form and function have been reported for plants across broad taxonomic groups occupying a broad range of biomes. These correlations are typically studied within the context of “plant functional traits” (PFTs), and much of the insight that has been added to this line of study has come



from temperate forest tree species. The formulation of PFTs is founded on the recognition that natural selection works within populations of plants to produce convergent and predictable patterns in form and function. Because many attributes are linked in their effect on fitness, evolutionary modification of one attribute is likely to cause a change in the fitness value of a second attribute, and these coupled influences are likely to vary depending on environmental and growth habit context. As an example, tree species that exhibit the evergreen growth habit tend to have longer-lived leaves with lower metabolic rates, compared to species that exhibit the deciduous growth habit. The concept of PFTs has also been extended to leaves within a single tree; shade leaves tend to have longer life spans, lower N concentrations, and lower rates of metabolism, compared to sun leaves. A key area of research that is currently underway is to separate the effects of genetics versus environment on the patterns of coupled trait influences in plants from numerous types of biomes.

In the same way that aboveground suites of plant functional traits have been recognized as constraining physiological function in predictable ways, belowground traits have been recognized recently as predictable predictors of biogeochemical, ecosystem processes (Phillips et al. 2013). As an example, let's return to the topic of mycorrhizal associations between roots and fungi. Mycorrhizal symbioses are often classified as arbuscular mycorrhizal (AM) or ectomycorrhizal (ECM). Arbuscular mycorrhizae involve the penetration of fungal hyphae into cortical root cells, where the hyphae become highly branched and induce reorganization of the organelles and cytoskeleton of the cell in ways that enhance the potential for reciprocal exchanges of carbohydrates and inorganic nutrients. Ectomycorrhizal hyphae do not penetrate root cells, but instead form a dense network between the epidermis and cortex, which then extends out into the soil. On the outside surfaces of the roots, ECM hyphae can form a dense coat, or covering, called the mantle, which exists at a higher biomass per unit of root length than the hyphae of AM. Furthermore, ECM fungi are capable of exuding exoenzymes to the soil, which catalyze reactions capable of mineralizing organic compounds from soil litter and producing inorganic ions as well as small organic compounds capable of resorption by the plant. AM fungi tend to not exude such enzymes and instead rely on the inorganic ions secreted by decomposer bacteria – essentially functioning in a manner similar to fine roots. These different mycorrhizal associations represent belowground plant functional traits and cause forest stands dominated by one type or the other to exhibit distinct patterns of carbon and nitrogen cycling. Trees associated with AM fungi tend to also produce leaf litter that decomposes rapidly, releasing mineralized ions rapidly for fungus and plant resorption. Trees associated with ECM tend to produce leaf litter that decomposes more slowly. With their suite of exuded exoenzymes, ECM fungi are capable of obtaining their ions and organic compounds from older, more recalcitrant, soil organic matter. Thus, in ECM trees rapid litter decomposition rates are not as crucial to sustaining a favorable tree nutrient balance. This framework, whereby belowground and aboveground traits are constrained by common biogeochemical constraints, may be useful to extending the concept of plant functional traits in ways that couple atmospheric and soil processes in ecosystems and predict specific biogeochemical patterns and processes.



## Temperate Forests and Disturbance

Forests experience relatively frequent disturbances, both natural and human induced, across multiple scales, ranging from gaps due to the death of single trees, loss of entire landscapes due to fire or insect outbreaks, and selective tree removal during logging. Disturbance sustains forests in a state of disequilibrium, or quasi-equilibrium, and thus underlies many of the time-dependent dynamics in community composition that are not associated with natural succession. In this section, I will focus on three disturbances and their relation to temperate forest dynamics during the approximately 12,000 years since the last glacial maximum (i.e., the Holocene): wild fire cycles, epidemic herbivory, and logging.

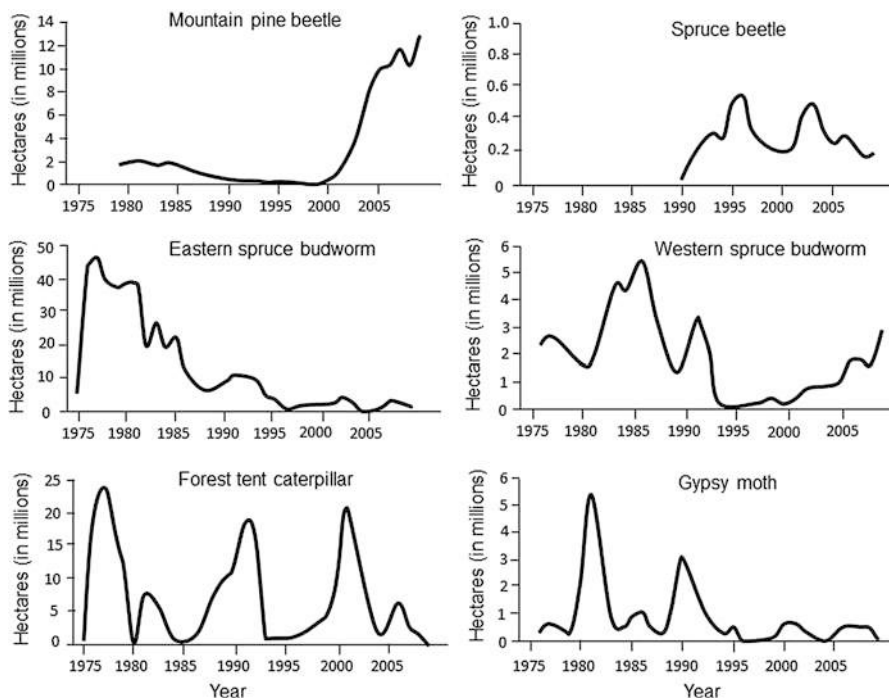
Wild fire has always existed in temperate forest ecosystems, and the presence of this disturbance should be viewed as a natural influence on forest dynamics. Fire is most damaging to a forest if it moves from the ground, where it is fueled by leaf (needle) and branch litter, to the crown, where it has the potential to destroy growth meristems and kill the tree. Ground fires scorch low-lying branches and tree boles, but as long as the growth meristems of the trees are not exposed to lethal temperatures, trees often recover. Some trees, such as some of the pines that have evolved in open ecosystems with frequent ground fires (e.g., *Pinus ponderosa* and *Pinus banksiana*), have evolved thick layers of bark and serotinous cones which provide them with unique adaptations that enable survival and recovery from frequent ground fires. The thick bark layers help insulate phloem tissue in the tree boles and thus maintain the capacity for downward transport of sugars from shoot to roots. Serotinous cones have scales that protect seeds and are sealed shut with a thin layer of “sticky” resin. When exposed to the heat of a ground fire, the resin will liquefy, the scales will open, and seeds will be distributed. (Crown fires burn at temperatures that destroy seeds in serotinous cones.) Serotiny has evolved as a trait with variable levels of expression. In some species, such as *Pinus contorta* (lodgepole pine), serotiny has been observed to vary depending on elevation and forest stand age, decreasing with elevation and increasing with stand age (Schoennagel et al. 2003).

Analysis of charcoal and “black carbon” particles from ocean sediments at the margins of continents has revealed that, globally, fire return frequency increased as a result of the spread of human civilization and development of human industry since the mid-Holocene (Carcaillet et al. 2002; Thevenon et al. 2010). It is probable that in local forested areas of Europe, the fire return frequency increased significantly since about 6 kyr BP and tripled during the late Holocene (1.8–0.6 kyr BP) as human societies moved through the Bronze Age and expanded agricultural activities, including the clearing of fields through burning. As towns and cities were established, and forest timber was increasingly used as a human resource (for both housing and energy), interest in the fire return frequency, and recognition of its role as a threat to natural resources, began to increase. In North America, the history of forest fire management is well documented and has been studied extensively. In the USA, beginning in 1905 with the establishment of the US Forest Service, fire suppression on publicly owned lands emerged as a primary policy mandate.

The need for suppression was progressively reinforced through the great fires of 1910 that burned over 1.2 million ha in the Western USA. By the late 1950s, evidence began to accumulate, especially through research at the Southern Forest Service Fire Laboratory in Macon, Georgia, that fire suppression actually increased the threat of catastrophic crown fires. In other words, the risk of losing forests as a natural resource was greater when fire suppression was practiced, than when it was not. Fire suppression causes the accumulation of understory fuels that facilitate the transformation of ground fires to stand-replacing, crown fires. Devastating fires in 1988 in Yellowstone National Park catalyzed creation of the US National Fire Plan, which replaced fire suppression as the national fire strategy, with a different strategy that emphasized canopy thinning and prescribed burns.

Recent studies using tree ring width and fire scars to reconstruct past histories of fire frequency and its relation to climate have revealed that oscillatory climate modes in the Earth system, particularly those associated with sea surface temperature, have a primary influence on fire frequency (Kitzberger et al. 2007). Patterns of sea surface temperature oscillations are complex and variable among geographic regions. However, one of the clear patterns that emerged is that in the Western USA, years of warmer-than-normal sea surface temperature in the Tropical Pacific Ocean (El Niño Southern Oscillations) often causes wet winters that facilitate an increase in fuel load. Often, El Niño years are followed by years with cooler-than-normal sea surface temperature in the Tropical Pacific (La Niña Southern Oscillations), which produce drier-than-normal winters and summers. Dry weather during la Niña years can increase flammability of the high fuel loads produced during the preceding El Niño year. Longer-term modes in sea surface temperature, such as the Pacific Decadal Oscillation (PDO) and Atlantic Multidecadal Oscillation (AMO), can interact with El Niño and La Niña oscillations to affect wildfire synchrony and frequency. These connections between the Earth's climate system and forest wildfires have been best studied in montane coniferous forests of the Western USA. Tree growth in these forests has decreased significantly during the period 1979–2008, compared to mean changes in tree growth between 1896 and 2008 (Williams et al. 2010). The trend of reduced forest growth was positively correlated with increased temperature and increases in the frequency of negative (drought) precipitation anomalies during this same period. As temperatures continue to rise and droughts become more frequent, and extreme climate modes oscillate more frequently, these montane forests are likely to experience even further decreases in growth.

Past studies on fossil plants have revealed that in general, during past geologic eras of elevated CO<sub>2</sub>, such as during the middle Eocene (approximately 45 mya) when atmospheric CO<sub>2</sub> concentrations were in the range of 800 ppmv (compared to current concentrations near 400 ppmv), rates of insect herbivory generally increase. This is because the C:N ratio of plant tissues increases, and insects must consume biomass at greater rates to obtain the nitrogen that often limits their growth and fitness. There is also increasing evidence that increased temperature (especially winter temperature) and more frequent droughts impose a stress on forests that makes them more susceptible to insect herbivory and mass mortality. These stresses increase the potential for insect larvae to overwinter in cold winter forests and thus



**Fig. 5** Recent histories of insect outbreaks in temperate forest ecosystems of North America (Redrawn from Hicke et al. (2012))

emerge with greater populations densities during the spring and summer, and they compromise the capacity for trees to produce resinous (carbon based) and other toxic (e.g., nitrogen based) herbivore defenses. Insect outbreaks in the temperate forests of North America are episodic and have occurred several times during recent decades (Fig. 5). One recent case of extreme disturbance in Western North American forests that has been linked to climate changes is that of the epidemic outbreak of mountain pine beetles (Hicke et al. 2012).

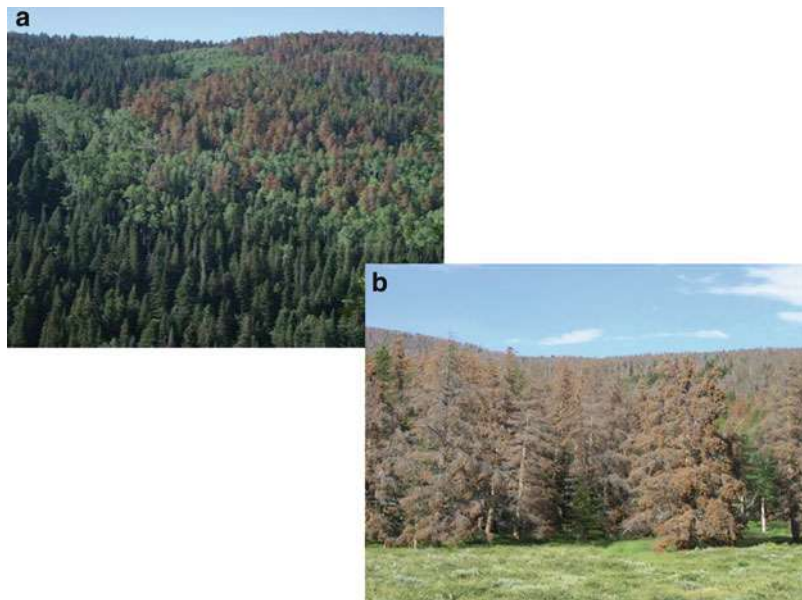
The most recent mountain pine beetle epidemic originated in British Columbia, Canada, at the beginning of the current century, and it has spread rapidly through the Rocky Mountain region of the Western USA. It is important to note that this insect (*Dendroctonus ponderosae* Hopkins) is native to montane pine forests in Western North America and has emerged in small outbreaks in past decades. The current outbreak, however, is especially large compared to previously recorded events. The mountain pine beetle associates with fungal pathogens that are carried by the beetles as they burrow into the bark of an infected pine tree and eat the phloem tissue of the tree, creating infection “galleries” or “tunnels” that are observable as the bark is peeled away. The fungal spores that are carried into the tree’s vascular tissue by burrowing beetles often germinate and proliferate within the phloem tissue, thus further disrupting the flow of sugars from shoots to the roots,

**Fig. 6** Map showing the natural range of the mountain pine beetle that infects pine forests in Western North America (area indicated by blue shading) and the approximate boundaries of the current epidemic of mountain pine beetle attack (area outlined by red dashed line)



as well as penetrating into the xylem tissue and blocking water flow from roots to shoots. Thus, an infected tree suffers from an inability to effectively transport sugars to the roots to support respiration and growth and water to the shoot to support transpiration. Trees typically die within 1–2 years after infection.

The current outbreak of mountain pine beetle in *Pinus contorta* (lodgepole pine) forests in Western North America has caused over 14 million ha of forest to shift to a state of mass mortality (Fig. 6). Infected stands typically exhibit over 60–80 % mortality of adult trees (Fig. 7). Two important secondary consequences have been predicted to result from this outbreak: (1) large areas of the Northern Hemisphere which had previously served as sinks for the uptake of atmospheric CO<sub>2</sub> will serve as sources of respired CO<sub>2</sub> for numerous decades into the future as the dead needles and wood decompose, and (2) forest fires will increase in frequency and coverage as the dead needles and wood serve as fuel. However, studies of ecosystem processes in beetle-damaged forests have revealed evidence that can be used to argue against the likelihood of both of these long-term impacts. Recent research has shown that widespread mortality of forest trees reduces the emission of CO<sub>2</sub> from soil respiration to the atmosphere because the sugars normally transported belowground and used to support both root and associated microbial respiration (often referred to as autotrophic respiration) are reduced. Furthermore, even after needles are deposited to the soil as litter, decomposition and the return of needle carbon to the atmosphere appears to be limited, at least during the initial decade after forest death, by an as-yet-to-be-identified resource or process (Moore et al. 2013). Thus, while carbon budget models predict a large carbon source for beetle-killed forests, these model predictions have not yet been validated by observations. In the case for increased



**Fig. 7** (a) A lodgepole pine and aspen-dominated forest in the Rocky Mountains of the Western USA showing the result of localized infection by mountain pine beetle. This forest is in the initial stages of an epidemic outbreak. Aspen trees appear as the lighter shade of *green*. (b) Lodgepole pine-dominated forest in the late stages of a mountain pine beetle infection in the Rocky Mountains. This forest stand is in the transition between what is commonly referred to as the *red* (earlier) and *grey* (later) stages of an infection. Note the presence of some live trees within the stand. Mortality in these stands is typically between 60 % and 80 %

fire frequency in beetle-killed forests, most studies have not shown this prediction to be borne out by observations, at least for the past two decades (Black et al. 2013). It is probable that live pine trees, with their high needle and wood concentrations of flammable terpene compounds, pose as great or greater fire risk to forests, than the increased dead wood and needle deposition in forests with high mortality rates.

One disturbance that has been chronic, and anthropogenic, for at least the past 5 millenia, is human logging of temperate forests. Wood is a natural resource that humans have used for a long time as both a fuel source and for the construction of shelter. Deforestation in some parts of the world have increased over the past several decades (e.g., the tropics), mostly for purposes of land use and the development of grazing-based animal husbandry and agriculture. However, in those regions, such as Europe and North America, where temperate forests were once used for high rates of log harvesting, removal of wood biomass has lagged behind forest regrowth for at least the past five decades. This has allowed for increased carbon sinks in the Northern Hemisphere. There is now concern that these sinks will approach saturation and weaken. The exact causes of this weakening are numerous and include climate changes and increased production of oxidant pollutants, such as ozone, which tends to inhibit plant growth.

Selective logging in forests tends to reduce amounts of forest stand biodiversity – removing certain species, which are often codominant in their representation in the community, and opening opportunities for other codominant species for competitive advantage. The logging-induced shift toward reduced biodiversity reaches a maximum when entire native forest stands are replaced by monoculture tree plantations. Tree plantations are increasing in their spatial coverage across the globe due to higher demands for bioenergy (through cellulosic ethanol production) and recognition of the high carbon sequestration capacities of forests. In temperate biomes, pine, spruce, poplar, and sweetgum have become preferred species for use in forest plantations. In general, a positive correlation has been recognized between ecosystem biodiversity and primary productivity. This relationship has been documented across ecosystem types ranging from grasslands to temperate forests. Greater biodiversity results in higher rates of biogeochemical cycling, which includes the processes of nitrogen and phosphorus cycling, and is enhanced by the higher diversity of soil microorganisms that are associated with a higher diversity of trees. In studies that have controlled for variance in climate across numerous temperate forests, while investigating the effect of stand diversity on wood production, a general positive correlation is found between production and diversity; this effect of diversity on wood production reached a maximum increase of 24 % when comparing native forests versus monospecific plantations in Europe (Vila et al. 2013). It is clear that the management of temperate forests through selective logging has high potential to alter the amount and sustainability of wood extraction as a natural resource.

In general, temperate forest disturbance, both natural and anthropogenic, tends to destabilize biogeochemical cycling of carbon and nutrients and have an overall effect of decreasing forest growth and decreasing the delivery of ecosystem goods and services. There is increasing recognition that effective management of temperate forests in the future will require greater focus on the causes of disturbance, synergistic interactions among multiple disturbances occurring in parallel or serially, and patterns and rates of forest recovery from disturbance.

---

## Future Directions

- Determination of how temperate forest carbon sinks are responding to directional climate shifts, particularly with regard to decreasing snow packs in certain regions at mid-latitudes, and increasing rain in other regions. The continental-scale redistribution of precipitation will undoubtedly impact carbon sequestration processes in temperate forest ecosystems.
- Determination of how nitrogen deposition to forest ecosystems influences natural biogeochemical cycles involving nitrogen, carbon, and phosphorus.
- Determination of interactions among disturbances on temperate forest biogeochemical cycling and the implications for such interactions on the future capacity for these ecosystems to provide water and take carbon out of the atmosphere, two essential services provided to humanity by temperate forest ecosystems.

## References

- Black SH, Kulakowski D, Noon BR, DellaSala DA. Do bark beetle outbreaks increase wildfire risks in the Central US Rocky Mountains? Implications from recent research. *Nat Areas J.* 2013;33:59–65.
- Boon S. Snow ablation energy balance in a dead forest stand. *Hydrol Process.* 2009;23:2600–10.
- Bugmann H. A review of forest gap models. *Clim Change.* 2001;51:259–305.
- Carcaillet C, Almquist H, Asnong H, Bradshaw RHW, Carrion JS, Gaillard MJ, Gajewski K, Haas JN, Haberle SG, Hadorn P, Muller SD, Richard PJH, Richoz I, Rosch M, Goni MFS, von Stedingk H, Stevenson AC, Talon B, Tardy C, Tinner W, Tryterud E, Wick L, Willis KJ. Holocene biomass burning and global dynamics of the carbon cycle. *Chemosphere.* 2002;49:845–63.
- Chen I-C, Hill K, Ohlemüller R, Roy DB, Thomas CD. Rapid range shifts of species associated with high levels of climate warming. *Science.* 2011;333:1024–6.
- Clark DA, Brown S, Kicklighter DW, Chambers JQ, Thomlinson JR, Ni J, Holland EA. Net primary production in tropical forests: an evaluation and synthesis of existing field data. *Ecol Appl.* 2001;11:371–84.
- Davis MB. Lags in vegetation response to greenhouse warming. *Clim Change.* 1989;15:75–82.
- Hicke JA, Allen CD, Desai AR, Dietze MC, Hall RJ, Hogg EH, Kashian DM, Moore D, Raffa KF, Sturrock RN, Vogelmann J. Effects of biotic disturbances on forest carbon cycling in the United States and Canada. *Glob Chang Biol.* 2012;18:7–34.
- Högberg P, Read DJ. Towards a more plant physiological perspective on soil ecology. *Trends Ecol Evol.* 2006;21:548–54.
- Hu J, Moore DJP, Burns SP, Monson RK. Longer growing seasons lead to less carbon sequestration by a subalpine forest. *Glob Chang Biol.* 2010;16:771–83.
- Kitzberger T, Brown PM, Heyerdahl EK, Swetnam TW, Veblen TT. Contingent Pacific-Atlantic Ocean influence on multicentury wildfire synchrony over western North America. *Proc Natl Acad Sci USA.* 2007;104:543–8.
- Lambers H, Raven JA, Shaver GR, Smith SE. Plant nutrient-acquisition strategies change with soil age. *Trends Ecol Evol.* 2008;23:95–103.
- Lebreton P, Gallet C. Plant communities are biochemically organized. *Acta Botanica Gallica (in French).* 2007;154:573–95.
- Lipson D, Näsholm T. The unexpected versatility of plants: organic nitrogen use and availability in terrestrial ecosystems. *Oecologia.* 2001;128:305–16.
- Moore DJP, Trahan NA, Wilkes P, Quaife T, Stephens BB, Elder K, Desai AR, Negrón J, Monson RK. Persistent reduced ecosystem respiration after insect disturbance in high elevation forests. *Ecol Lett.* 2013;16:731–7.
- Norby RJ, Wullschlegel SD, Gunderson CA, Johnson DW, Ceulemans R. Tree responses to rising CO<sub>2</sub> in field experiments: implications for the future forest. *Plant Cell Environ.* 1999;22:683–714.
- Parmesan C, Yohe G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature.* 2003;421:37–42.
- Reich PB, Grigal DF, Aber JD, Gower ST. Nitrogen mineralization and productivity in tree stands on diverse soils. *Ecology.* 1997;78:335–47.
- Saugier B, Roy J, Mooney HA. Terrestrial global productivity. San Diego: Academic; 2001. 573 p.
- Schoennagel T, Turner MG, Romme WH. The influence of fire interval and serotiny on postfire lodgepole pine density in Yellowstone National Park. *Ecology.* 2003;84:2967–78.
- Thevenon F, Williamson D, Bard E, Anselmetti FS, Beaufort L, Cachier H. Combining charcoal and elemental black carbon analysis in sedimentary archives: implications for past fire regimes, the pyrogenic carbon cycle, and the human-climate interactions. *Glob Planet Change.* 2010;72:381–9.
- Vadeboncoeur MA. Meta-analysis of fertilization experiments indicates multiple limiting nutrients in northeastern deciduous forests. *Can J Forest Res.* 2010;40:1766–80.

- Van Tuyl S, Law BE, Turner DP, Gitelman AI. Variability in net primary production and carbon storage in biomass across Oregon forests – an assessment integrating data from forest inventories, intensive sites, and remote sensing. *For Ecol Manage.* 2005;209:273–91.
- Varhola A, Coops NC, Weiler M, Moore RD. Forest canopy effects on snow accumulation and ablation: an integrative review of empirical results. *J Hydrol.* 2010;392:219–33.
- Vila M, Carrillo-Gavilan A, Vayreda J, Bugmann H, Fridman J, Grodzki W, Haase J, Kunstler G, Schelhaas M, Trasobares A. Disentangling biodiversity and climatic determinants of wood production. *PLoS One.* 2013;8, e53530.
- Vitousek PM, Field CB. Ecosystem constraints to symbiotic nitrogen fixers: a simple model and its implications. *Biogeochemistry.* 1999;46:179–202.
- Williams AP, Allen CD, Millar CI, Swetnam TW, Michaelsen J, Still CJ, Leavitt SW. Forest responses to increasing aridity and warmth in the southwestern United States. *Proc Natl Acad Sci.* 2010;107:21289–94.

## Further Reading

- Box EL, Fujiwara K. *Temperate forests around the Northern Hemisphere.* Heidelberg: Springer; 2013. 280 p.
- Frelch LE. *Forest dynamics and disturbance regimes: studies from temperate evergreen-deciduous forests.* Cambridge: Cambridge University Press; 2008. 280 p.
- Kurz WA, Dymond CC, Stinson G, Rampley GJ, Neilson ET, Carroll AL, Ebata T, Safranyik L. Mountain pine beetle and forest carbon feedback to climate change. *Nature.* 2008;452:987–90.
- Turner MG, Romme WH, Gardner RH. Prefire heterogeneity, fire severity, and early postfire plant reestablishment in subalpine forests of Yellowstone National Park, Wyoming. *Int J Wildland Fire.* 1999;9:21–36.
- Wappler T, Labandeira CC, Rust J, Frankenhauser H, Wilde V. Testing for the effects and consequences of mid-Paleogene climate change on insect herbivory. *PLoS One.* 2012;7, e40744.



Darren R. Sandquist

## Contents

Introduction .....	299
Desert Formation Affects Desert Diversity .....	300
The Abiotic Environment Underlying Desert Productivity .....	301
Precipitation and Drought .....	302
Functional Diversity and Responses to the Environment .....	302
Ecological Groupings of Desert Plants .....	302
Photosynthesis in a Water-Limited Environment .....	303
Adaptive Forms and Functions Related to Desert-Plant Water Relations .....	311
Biotic-Mediated Processes Are Critical for Nutrient Balance in Deserts Plants .....	317
Desert Biodiversity and Community Composition .....	319
Species Diversity Can Be Surprisingly High in Deserts .....	319
Population and Community Dynamics Are More Complex than Expected .....	321
Disturbance, Global Changes, and Future Challenges .....	324
Disturbances Pose Significant Challenges in Low Productivity Ecosystems .....	324
Nonnative Species Are a Major Threat to Desert Communities .....	324
Other Global Changes also Threaten Desert Regions .....	325
References .....	325

---

## Abstract

- There is no single definition of “desert,” but it is widely agreed that deserts are arid because they receive little precipitation and experience high evaporation annually. These factors result in low soil water availability that severely limits plant productivity. Thus, another feature of deserts is low vegetation cover.
- Although all deserts are dry, there is extreme abiotic and biotic variability among the world’s deserts – perhaps more so than for any other biome. This arises in part from the varied causes of desert formation, their disjunct distributions, and their independent floral histories.

---

D.R. Sandquist (✉)

Department of Biological Science, California State University, Fullerton, CA, USA

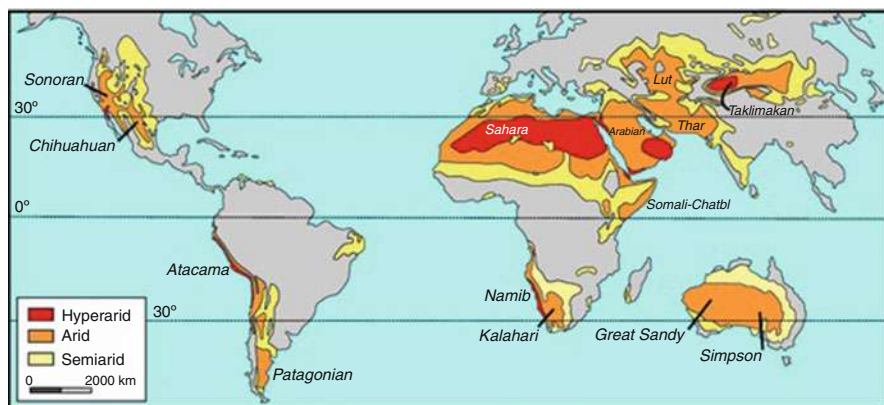
e-mail: [dsandquist@fullerton.edu](mailto:dsandquist@fullerton.edu)

- High spatial and temporal variability of the abiotic environment present challenges to desert life that has important implications at both the ecological and evolutionary scales. Besides limited water, other abiotic factors play important roles in desert ecosystems. Temperatures can be extremely high, but in some deserts low temperatures also constrain productivity. Resources, such as nitrogen, are also generally low in deserts, so that even when water is available, plant productivity may be relatively constrained.
- Most if not all life forms are present in desert ecosystems, regardless of the classification scheme used. Perennial shrubs dominate most desert landscapes, but in any single habitat trees, grasses, annuals, stem succulents, or leaf succulents may be the dominant form.
- From studies of desert plants, researchers have identified many adaptive functions at the ecophysiological level. These emerge from a plant's need to grow and survive through extreme drought, high solar radiation, and high temperatures, as well as through wide fluctuations in all of these abiotic factors.
- Plants exhibiting the succulent syndrome (which includes water storage, extensive surficial roots, and often CAM photosynthesis) are well adapted for life in warm arid ecosystems. Succulent plants are key components of many desert communities but they are rarely the dominant life form and are entirely absent from some deserts.
- Nutrients usually limit desert productivity during periods when water is available. Low external nutrient input results in decomposition by both abiotic and biotic processes playing a major role in nutrient availability. Other biotic-mediated processes, such as microbial nitrogen fixation and fungal root associations, are critical to maintaining favorable nutrient balance in desert plants.
- In spite of low productivity, deserts have surprisingly high biodiversity and endemism. Climate variability, geographic isolation, geologic history, and edaphic anomalies are among the primary drivers for greater-than-expected plant biodiversity.
- Biotic interactions were once thought to be rare in deserts and thus not important in desert community dynamics. In recent decades however, intra- and interspecific competition and facilitation have been clearly identified as important drivers in shaping desert plant communities. Arid and semiarid ecosystems are now widely used to test theories about the interplay between competition and facilitation.
- Deserts have always been susceptible to soil disturbances by nonnative ungulates and human activities. The profound effects on soil and nutrient losses are difficult to restore. In contrast, deserts were once considered relatively resistant to alien plant invasion, but recent spread of nonnatives has led to altered biogeochemical cycles and increased fire disturbances. In some cases, the changes have led to type conversion of vegetation. New pressures, such as renewable energy development, underscore the need for a solid scientific understanding of plant functions and ecosystem processes in arid and semiarid ecosystems.

## Introduction

Desert is the biome classification for terrestrial regions of Earth that are climatically arid and have low vegetation cover. Additionally, the climate of such regions is often highly variable across seasons and years. While there is no single index that is used universally to define deserts, a simple one, proposed by Meigs (1953), is based only on precipitation, whereby extremely arid regions experience at least 12 months without rainfall, arid regions receive <250 mm rainfall annually, and semiarid regions receive 250–500 mm rainfall annually (Fig. 1). Boundaries based on this index do a good job delimiting deserts across the globe and correspond closely to boundaries used in other classification systems (e.g., Ezcurra 2006). But aridity is not simply based on the amount of water derived from precipitation; it also depends on the loss of that water, which affects its availability for plant productivity. A more inclusive definition of aridity comes from the comparison of water loss via evapotranspiration (ET) versus water input from precipitation (P). The ratio  $P/ET$  is a commonly used index of aridity (e.g., UNESCO 1977) defining hyperarid zones as having  $P/ET < 0.03$  and arid zones having  $P/ET$  of 0.03–0.20. Although this definition does not significantly change the global boundaries of deserts as compared to other indices, such as Meigs', it does provide a more biologically relevant measure of aridity in terms of water availability for plant use.

Other environmental parameters, such as timing and intensity of rainfall, seasonal temperatures, and soil texture, to name a few, can also play a role in affecting the aridity of deserts, albeit at smaller spatial and temporal scales. These additional factors affect the abiotic heterogeneity within deserts that contributes to the surprising functional diversity of plants found in desert ecosystems. This chapter explores the diversity of desert plants from an ecological context. It begins with a short review of desert formation and abiotic variability as a foundation for understanding the causes of biotic diversity among and within deserts. Then, the diversity of desert vegetation is explored from a functional context through the community level.



**Fig. 1** Global distribution of nonpolar arid lands based on Meigs' (1953) classifications

It ends with some considerations of how the desert biome has changed due to human activities and how it may change with future global changes.

## Desert Formation Affects Desert Diversity

Approximately one-third of the terrestrial biosphere can be classified as desert (Fig. 1), but beyond the common feature of being arid, there is extreme variability among and within deserts in terms of abiotic properties and thus biotic composition. One reason for this variability is that desert formation varies across global, regional, and local scales. With the exception of polar deserts, most large deserts are found in the horse latitudes, near 30°N and 30°S (e.g., Sahara desert; Fig. 1). This is the result of global atmospheric circulation patterns known as *Hadley cells*. These cells are fluid masses of circulating air driven by energy from the absorption of solar radiation near the equator (where solar radiation is greatest on average). This radiation generates masses of rising warm air that are moist with water from evaporation, but as an air mass rises, it expands and cools, and the water vapor within it condenses to form clouds and rain. Most of the water is lost from this air mass as it reaches higher altitudes. The now cool and dry air mass cannot sink back to the Earth at the equator owing to the continual convection of warm air. Instead, it is deflected to the north and south, where it begins to sink near the 30°N and 30°S latitudes. As the cool dry air mass sinks, it is compressed and warms, which allows it to absorb additional moisture that it may encounter. This contributes to a reduction of cloudiness in the latitudes around 30°N and 30°S and increases penetration of solar radiation to the land surface. The combination of dry air and high radiation results in low precipitation and high evaporation in these high-pressure *subtropical latitude deserts*. Most of the deserts on our planet are influenced in part by this global pattern of air circulation.

Hadley cell circulation predisposes the latitudes around 30°N and 30°S to being arid, but processes at regional and local scales may also contribute to, or even be fully responsible for, the formation of deserts. At a regional scale, some deserts form in the interior of continents – commonly called *mid-continent deserts* (e.g., Taklimakan desert of central Asia). These exist because they are far from the ocean, which is the primary source of moisture for rainfall. As an air mass containing water evaporated from the ocean moves across land, rain falls more or less continuously; thus, by the time the air mass reaches the most interior region of a continent, most of the moisture has already precipitated. The result is a vast area of arid land within the interior of the continent.

Other major causes of desert formation occur on more localized scales. The most common are *rain-shadow deserts*, which form on the leeward side of high mountain ranges (e.g., Great Basin Desert leeward of the Sierra Nevada and Cascade mountains). These mountains force moisture-rich air masses upward, thereby decreasing the pressure and temperature of the air mass moving across the land. Water vapor within the air condenses causing rainfall on the windward mountain slopes, but as these air masses move over the range and descend to lower elevation on the

opposite side, the air pressure and temperature rise (similar to Hadley cells). The limited moisture left in the descending air mass is prevented from precipitating because of the increasing pressure and temperature. Low rainfall, warm air, and high solar radiation resulting from the presence of a mountain range result in arid areas in the mountain's "shadow."

*Coastal deserts* form where very cold ocean waters occur at the surface and adjacent to a relatively warm continental margin (e.g., the Namib Desert occurs where upwelling brings cold water to the surface along the western coast of Africa). The interaction of the ocean, air, and land is complex in these systems, but in general the cold surface waters cause air masses that overlie them to cool. This decreases evaporation and reduces the capacity of the overlying air to hold water vapor, causing condensation and offshore precipitation. Sometimes the condensation forms fog, which may be drawn onto land, but as the fog blankets the land, it too warms and evaporates back into vapor. Because of this phenomenon, coastal deserts may also be known as *fog deserts*. Coastal deserts are among the driest in the world sometimes experiencing years without measurable rainfall (e.g., Namib and Atacama). In fact, fog is typically the most reliable source of water for productivity in these deserts, and many plants of these deserts show adaptations for capturing and taking up fog-derived water (e.g., *Nolana mollis*).

Besides the different formation processes, highly varied ages among the world's deserts also contribute to the diversity among them. Some deserts appear to be extremely old (e.g., Namib Desert >55 myo) giving rise to high diversity and endemism through many generations of evolution. Other deserts are very young (e.g., Mojave Desert ~11,000) and are strongly impacted by migration processes from regional biota. This can also lead to high diversity, especially if the desert forms at the intersection of multiple ecological regions. Biodiversity among deserts is also a result of their disjunct distribution. That is, deserts of the world are largely separated from each other compared to other biomes. As such, evolutionary processes within them have taken place largely in isolation from each other.

---

## The Abiotic Environment Underlying Desert Productivity

As already noted, there is extreme abiotic and biotic variability among the Earth's deserts. Polar deserts are at one extreme of this variability. Although they are arid, the reason they sustain little life is mostly due to very low temperatures and a limited growing season. (Polar desert plants are reviewed in the Chap. 13, ► "[Plants in Arctic Environments](#)" of this volume and are not further included herein.) For the rest of the world's deserts, plant production is limited by a general lack of resources. The most ubiquitous, of course, is the lack of water, but other resources (e.g., mineral nutrients) may also be limiting, especially during periods when water is abundant. Other abiotic factors also have important impacts on desert plants and their function. Intense solar radiation, high and low air temperatures, saline soils, and strong winds are but a few of the abiotic stresses that regularly impact desert plants. Furthermore, in desert environments some of the highest

spatial and temporal variability of these abiotic factors is found as compared to anywhere globally. Biotic interactions and their affect to desert plant communities have historically been considered less important than abiotic factors; however there are many examples of important biotic interactions, both plant-plant and plant-animal, that influence desert vegetation, especially in terms of community structure, pollination, and plant recruitment.

## Precipitation and Drought

All discussions of desert abiotic factors begin with precipitation and for obvious reasons. Water is the most limiting resource for productivity in deserts. But the amount of water is only one of the many water-related factors affecting plant function in deserts. The precipitation form (rain, snow, fog), intensity, and timing all affect its ultimate availability and use by plants. Furthermore, the absence of water and duration of that absence (i.e., drought) have substantial effects on desert plant ecological functions and evolutionary responses. Because water limits lifetime growth and reproduction in desert plants, all face the challenge of balancing carbon gain against water loss. This trade-off results, for the most part, because the primary path for carbon uptake is the same as for water loss – both flux through the stomata. This trade-off appears to have driven many of the adaptive changes in physiology, morphology, and behavior seen in desert plants.

---

## Functional Diversity and Responses to the Environment

### Ecological Groupings of Desert Plants

Desert plant activity is limited first and foremost by low water availability, but more specifically it is the pulsed nature of water availability and periods of severe water limitation between these pulses that have most strongly impacted the evolution and ecology of desert biota. Not surprisingly, many classifications of desert plants focus on patterns of activity through rain and drought cycles. The simplest of this function-based classification scheme is to group plants along a continuum from drought avoiding to drought tolerating. Avoiders do not experience the stress of drought because either they have mechanisms that circumvent it or they become inactive (including dying, as in the case of annuals). Tolerators maintain activity through the drought albeit at a substantially reduced level.

This simplified scheme is sometimes difficult to use across the broad range of desert species with highly varied adaptive responses to drought, and a number of related classification systems persist in the literature. In an early and still widely used scheme proposed by Kearney and Shantz (1912) and later modified by Shantz (1927), annuals are considered *drought escaping* because they are active only during favorable conditions and absent during drought. *Drought-enduring* plants are present during drought, but become inactive usually during the early stages of

water stress. Drought-deciduous shrubs (plants that do not die, but do lose their leaves during drought) are drought enduring. There is overlap in Shantz' (1927) definitions of *drought evading* and *drought resisting*, which appears to have generated some confusion in the literature. Both types are active during drought, but *drought-evading* plants typically have higher growth per unit water used (i.e., higher water-use efficiency) due to adaptive traits that reduce water loss and prolong the growing period. Reduced transpiration due to stomatal regulation coupled with morphological features such as stomatal pits, leaf hairs or waxes, and small leaf sizes is often found in drought evaders. Kearney and Shantz (1912) also classified plants with extensive root systems into the drought-evading category. *Drought-resisting* plants have persistently low-to-moderate levels of activity through periods of low water availability as well as during more favorable periods. Reduced transpiration is the norm for drought resisters, but they can tolerate very low water potentials, often via osmotic regulation. Succulent plants and some of the most successful desert perennial shrubs (e.g., creosote bush) fall into this category.

Categorizing plants in terms of their functional attribute is useful only to a limited extent, and there are many examples of taxa that exhibit properties of more than one category. For example, one could argue that creosote bush exhibits both drought-resisting and drought-enduring characteristics (small leaves with resinous excretions to reduce transpiration). For this reason, a popular alternative is to group desert plants based on life forms. These forms usually include annuals, perennial grasses, deciduous shrubs, evergreen shrubs, CAM succulents, and deep-rooted trees (phreatophytes). Smith et al. (1997) took this approach in their summary of North American desert plant ecophysiology, and many others have applied it as a way to simplify the presentation of the complex diversity of form and function found in desert species. Interestingly, these diverse life forms are present in a broad cross section of desert taxa suggesting that the mechanisms for dealing with aridity and heat, or the ability to form them through natural selection, are fundamental to many lineages.

## Photosynthesis in a Water-Limited Environment

*The Desert Plant Dilemma: Balancing Carbon Gain and Water Loss.* There is a long history of research on the ecophysiology of desert plants and a number of valuable reviews of such studies. Smith et al. (1997), for example, provide a comprehensive review of plant ecophysiology in North American deserts, and many studies from the Namib Desert are present in von Willert et al. (1992). Most ecophysiological investigations of desert plants emphasize photosynthetic gas exchange, plant water relations, and the link between them, but these emphases are reasonable given that maximizing carbon gain and minimizing water loss are the prevailing challenges in desert systems.

In arid ecosystems any extraneous plant water loss has the consequence of reduced production and, in extreme circumstances, potentially plant death. Photosynthesis relies on CO<sub>2</sub> uptake through stomata, but this process incurs the unavoidable loss of water via the reverse path (transpiration), thus resulting in a trade-off that forms the

foundation for many of the adaptive characteristics seen in desert (and other) plants. In deserts, plant water loss from leaves is exacerbated by high water vapor pressure differences (VPD) between the leaf and the air. When water is abundant, as during the growing period for annuals, the water lost via transpiration may be inconsequential, especially compared to its role in reducing leaf temperatures. For plants that remain active during the drought period, mechanisms that help reduce water loss should be favored. The most straightforward mechanism for reduction of water loss during periods of drought is to reduce the size of the stomatal opening, thereby decreasing conductance of water from the leaf. But this comes at a cost; it reduces uptake of  $\text{CO}_2$ . In addition, for most desert plants, a reduction in transpiration results in a potentially dangerous increase of leaf temperature (see discussion of energy balance in chapter ► “Plants in Alpine Environments”). Over the years, myriad fascinating examples of morphological, physiological, and behavioral mechanisms have been identified that help desert plants avoid the full consequences of these trade-offs. In general these can be grouped into ways of improving photosynthesis relative to water loss, decrease dependence on transpiration for energy balance, and ways to take up or save more water.

### Photosynthetic Pathways

Pick up almost any book about photosynthesis and entire chapters can be found about C<sub>3</sub>, C<sub>4</sub>, and CAM photosynthesis. Indeed, the ecology and biochemistry of these three photosynthetic pathways differ so greatly that they warrant entire volumes. Rather than review the three photosynthetic pathways in detail, the attributes of each that are important for their presence in deserts are highlighted; then their distribution and how the different pathways correspond to variability among these arid ecosystems are explored. All three pathways are present across the deserts of Earth, but as might be expected, their abundances differ in relationship to the environments of each desert.

Of the three pathways, C<sub>3</sub> photosynthesis is the most widespread globally, and the same is true across deserts. However, net carbon gain of C<sub>3</sub> plants is negatively affected by photorespiration, which goes up with increasing temperatures. This is one reason C<sub>4</sub> and CAM plants may have a competitive advantage over C<sub>3</sub> plants in hot deserts (Ehleringer and Monson 1993). In deserts with cooler temperatures during the growing season, the disadvantage of photorespiration is significantly lower, thereby reducing the relative benefit of the C<sub>4</sub> pathway. The C<sub>4</sub> pathway also requires two additional ATP to fix  $\text{CO}_2$  (compared to C<sub>3</sub>), making it best suited to high-light environments. As expected from these fundamental differences, the greatest abundance of C<sub>4</sub> plants in deserts is where temperatures and light are high and water is available during warm periods.

C<sub>4</sub> plants have high water-use efficiency (carbon gain vs. water loss) because the  $\text{CO}_2$ -concentrating mechanism of the C<sub>4</sub> pathway maintains higher internal  $\text{CO}_2$  concentrations relative to stomatal conductance and thus transpiration. However, the need for water during the warm growing season prevents most C<sub>4</sub> plants from being well adapted to drought conditions. In contrast, CAM plants have extremely high water-use efficiency. They benefit from the same  $\text{CO}_2$ -concentrating



mechanism found in C4 plants, but additionally keep their stomata closed during the day when evaporative demand (i.e., VPD) is high, and then open them for CO<sub>2</sub> exchange during the lower-VPD hours of darkness.

Other attributes of CAM species benefit their tolerance of drought. As previously described, many CAM plants are succulent, having tissues that store water for use during drought, but CAM is not restricted to succulent plants. Likewise, not all succulents are CAM (e.g., many leaf-succulent shrubs of the Succulent Karoo are C3). Some CAM species can switch to C3 photosynthesis when environmental conditions are favorable, especially when water availability is high, and many can use C3 photosynthesis during a small fraction of the regular daily CAM cycle.

CAM plants increase in abundance in hot deserts that have some degree of water limitation during the warm season. This limitation may stem from an absence of precipitation or from an ephemeral and unpredictable precipitation regime. But, most CAM species are sensitive to freezing temperatures and thus absent from cold (high-elevation or high-latitude) deserts.

In North American deserts, the relative abundances of C3, C4, and CAM along a north-to-south gradient of increasing temperature and summer rainfall reflect the typical pattern among global deserts. In the winter-rain-dominated Great Basin cold desert, CAM and C4 plants are largely absent except in saline habitats (see “Desert Halophytes”). The proportion of CAM and C4 species increases slightly in the Mojave Desert to the south, where annual temperatures are warmer but winter rains still dominate. In both of these deserts, C3 species greatly outnumber C4 and CAM species. Even further south, and at overall lower elevations, CAM species become an important part of the Sonoran Desert flora, with some taxa (e.g., Cactaceae) showing remarkable morphological as well as taxonomic diversity. Summer rains are abundant here but spatially and temporally variable. C4 species, especially grasses, also become a more integral part of the flora in the Sonoran Desert but normally in the higher elevations where rainfall is more abundant and predictable. The southernmost North American desert, the Chihuahuan, has an abundance of CAM and C4 species related to the higher annual temperatures and summer rainfall of this desert. CAM agaves and cacti are more speciose here and can be the dominant taxa of some Chihuahuan communities. C4 grasses can likewise dominate vast areas of the Chihuahuan, especially where rainfall is relatively plentiful. But, both C4 grasses and CAM species are often not the dominant plants on heavily calcareous soils that occupy many parts of the Chihuahuan. Here they are replaced by C3 shrubs (primarily creosote bush and tarbush) – a shift that probably reflects poor retention of shallow water on such substrates. This pattern illustrates the potential for local edaphic effects to modify climate patterns that would otherwise favor certain ecophysiological syndromes over others.

### Leaf Energy Balance

Many adaptive traits at the leaf level are related to energy balance because (1) maintaining a favorable leaf temperature is important for photosynthesis and (2) the most efficient means of heat dissipation is by *latent heat transfer* which is due to transpiration. When moisture is abundant, the consumption of water for

latent heat transfer poses few, if any, problems. But when water is limited, which often corresponds to the warmer periods of the year, reliance on latent heat transfer presents a challenge. This challenge appears to have driven functional diversification and adaptation at the leaf level among many desert plants, as well as in other ecosystems. (For a more detailed review of energy balance, see the chapter ► [“Plants in Alpine Environments”](#).)

### **Small Leaves Decrease Leaf Temperature and Transpiration**

Reduced leaf size is one of the most widespread morphological adaptive features seen in desert plants. It seems intuitive that because there is less surface area on smaller leaves, water loss will be lower, but this is not necessarily true. Water loss from a leaf is dependent on transpiration rate, which is an area-standardized measurement (e.g.,  $\text{mmol H}_2\text{O m}^{-2} \text{ leaf s}^{-1}$ ). A priori, small and large leaves can have the same transpiration rate, in which case a canopy of many small leaves will lose the same amount of water as one with fewer large leaves (i.e., the total surface area is the same). For small leaves to be adaptive in terms of water loss, they must instead have a lower transpiration rate, which, as explained below, they usually do. Small leaves also do not heat up to the same extent as larger leaves. These two properties go hand-in-hand, and since heat and water limitations are two of the greatest challenges for desert life, it is not surprising that small leaves are common in the desert flora.

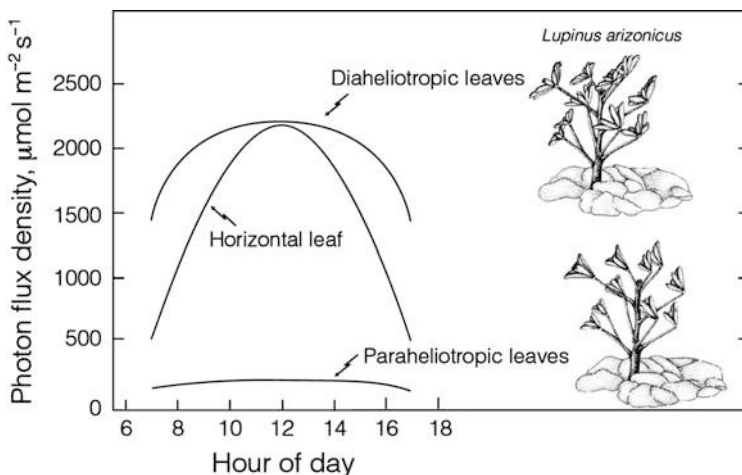
The primary reason smaller leaves stay cooler, and subsequently have lower transpiration than larger leaves, is that they have a reduced boundary layer for heat transfer. A smaller boundary layer means that heat transfer from the leaf to the surrounding air (i.e., *convective heat transfer*) is more rapid. Thus, as the leaf heats up from absorption of radiation from the sun and surrounding objects, higher convective heat loss keeps the leaf temperature closer to the air temperature ( $\Delta T$ ). Convective heat loss means that the plant is less dependent on *latent heat transfer*, via transpiration, for maintaining a favorable leaf temperature. But additionally, a lower  $\Delta T$  also reduces the vapor pressure difference (VPD) between the leaf and air, which also lowers transpiration.

Lower leaf temperature may also benefit the leaf in terms of photosynthetic rate since the lower temperature is likely closer to the thermal optimum for photosynthesis. Recall also that lower temperatures reduce photorespiration in C3 plants.

For many species, leaf sizes can vary across seasons and years, with smaller leaves produced during warmer periods or during drought. Such adjustment are crucial in plants that persist through periods of water shortage and high temperatures, underscoring the importance of another adaptive function in desert plants – acclimation.

### **Leaf Angles and Leaf Movement Affect Light Interception**

Another beautiful example of acclimation in desert plants is leaf movement known as *heliotropism* (meaning sun orienting). Some desert species, mainly annuals, display *diaheliotropic* leaf movement (orientation perpendicular to sun rays) and *paraheliotropic* leaf movement (orientation parallel to sun rays), although not all species do both. The former maximizes interception of solar radiation whereas in



**Fig. 2** Interception of solar radiation (measured as photon flux density) by diaheliotropic and paraheliotropic leaves during daylight hours. Arizona lupine (*Lupinus arizonicus*) of the Mojave and Sonoran Deserts can switch from fully diaheliotropic during periods of favorable soil moisture to fully paraheliotropic during water-stressed periods – or combine both dia- and paraheliotropism during a single day. For comparison, interception by a non-heliotropic horizontal leaf is also shown. A vertical leaf (not shown) would have an inverted curve from the horizontal leaf (Redrawn with permission from J. R. Ehleringer)

the latter minimizes it (Fig. 2). Diaheliotropism ensures that photosynthesis is rarely, if ever, light limited, which is beneficial over the short growing season of desert annual plants. The increased heat load owing to high incident solar radiation is balanced by high transpiration, which may explain why diaheliotropism is largely limited to annuals.

An interesting example of heliotropism is found in Arizona lupine (*Lupinus arizonicus*), an annual of the legume family (Fabaceae) that displays both dia- and paraheliotropism. During the warm and dry late-growing season, Arizona lupine displays diaheliotropism during the early morning hours, but as soil water declines and temperatures increase later into the day, leaves switch to being paraheliotropic (Fig. 2). This switch substantially reduces interception of direct solar radiation, which reduces photosynthesis but also decreases leaf heat load and transpiration.

Heliotropism is not restricted to annuals but is much less common in other life forms. One woody genus in which it appears is *Prosopis* – also a member of the legume family. Although the heliotropic species of *Prosopis* have extensive root systems, some being phreatophytic (described below), they may still experience daily cycles of water stress especially in non-riparian habitats (e.g., sand dunes). These daily cycles of water stress result in switches between dia- and paraheliotropic leaf movements, as seen in Arizona lupine.

Most desert plants do not have heliotropic leaf movement, but leaf angles can still play an important role in light interception and energy balance. Many desert species have vertically biased, nonrandom leaf angles. Although fixed, these leaf



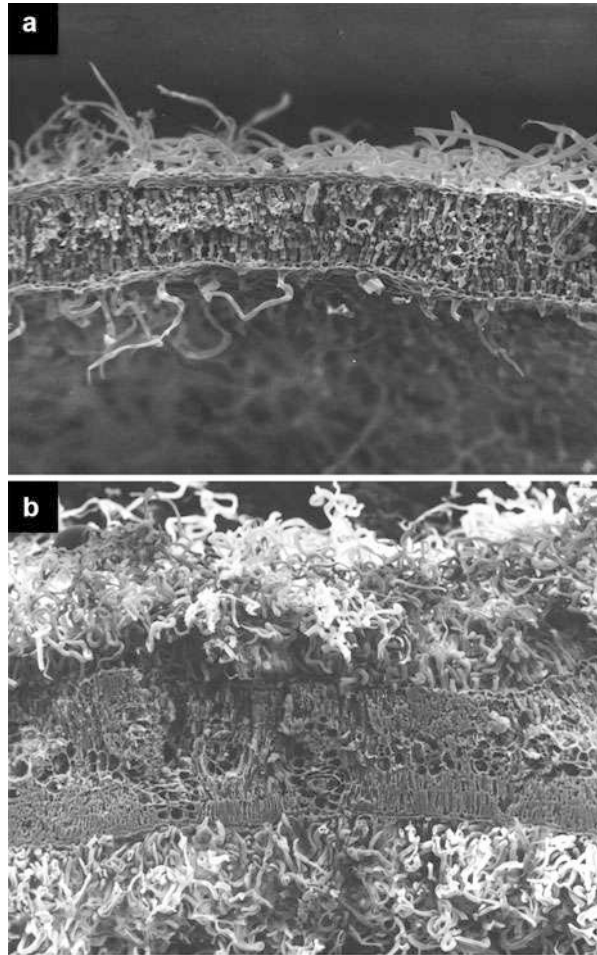
**Fig. 3** *Copiapoa cinerea* ssp. *columna-alba* of the Atacama Desert, Chile, grow with a northward orientation that helps maintain warm temperatures on the apical meristem during the cool period of the year but reduces heat load during the hot season (Photo: D. R. Sandquist)

angles function much the same as switching between dia- and paraheliotropism (Fig. 2). That is, they maximize light capture in early morning and late afternoon hours, when air temperature and VPD are lower, but avoid direct solar radiation in the more severe midday hours. Nonrandom leaf orientations may also reduce self-shading, with the angles being specific not just to daily radiation changes but also to seasonal changes. Such orientation benefits are also found in photosynthetic stems, including those of succulent species. An example of this is seen in the cactus *Copiapoa cinerea* ssp. *columna-alba* from the Atacama Desert of Chile. The succulent stems of these plants orient due-north giving the comical appearance of a small cactus army marching towards the equator (Fig. 3). Ehleringer et al. (1980) showed that this orientation facilitates apical warming for growth during the cool/wet parts of the year and reduces radiation (thus heat load) during the driest part of the year.

### Reflective Leaf Surfaces Decrease Absorption of Solar Radiation

The multiple benefits of reducing direct solar radiation suggest that other leaf properties should serve this function in desert plants. Indeed, there are a number of traits that do so at the leaf surface, reflective waxes and leaf hairs being among the most common. A well-studied example of this is found in brittlebush (*Encelia farinosa*, Asteraceae), a drought-deciduous shrub of the Mojave and Sonoran Deserts. Leaves produced by this species can have a thick layer of trichomes (leaf hairs) that strongly reflects solar radiation. Notably, the thickness of the trichomes, and thus the amount of reflectance, depends on the level of water stress experienced by the plant. Leaves produced early in the rainy season are generally large and have

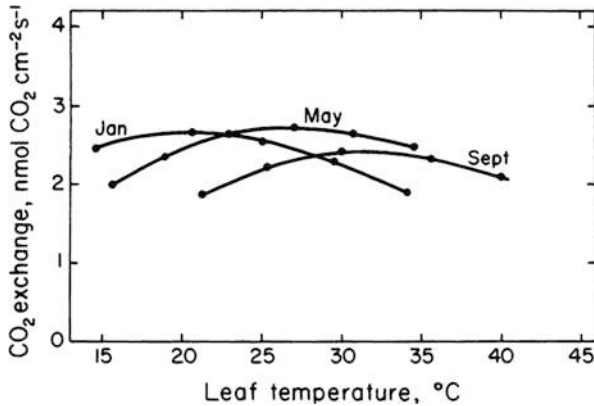
**Fig. 4** Micrographs of brittlebush (*Encelia farinosa*) leaves from the Mojave Desert. (a) Leaves produced early in the growing season when soil water availability is favorable have low trichome densities. (b) Leaves produced later in the season, when water stress has increased, have a dense trichome layer (Photos: J. R. Ehleringer)



few trichomes (Fig. 4a). These leaves absorb ~80 % of the solar radiation incident on their surface, but the heat load resulting from this radiation is easily balanced by transpiration during this wet period of the year. As the season progresses, and soil water decreases, new cohorts of leaves are produced which have increasing trichome densities (Fig. 4b). The higher densities lower radiation absorption to as little as 40 %, which attenuates excessive heat load and, importantly, reduces dependence on transpiration as the plants enter the drought period. As one would expect, the lower light absorption also decreases photosynthesis, but acclimation through increased trichome development allows plants to remain active much longer into drought, thereby compensating for the decrease of photosynthesis.

#### **Biochemical Acclimation Changes Thermal Optimum of Photosynthesis**

Rather than maintaining a narrow range of leaf temperatures for optimal photosynthesis, an alternative is to change the optimum temperature. (One might call this,



**Fig. 5** Temperature acclimation of photosynthesis (measured as CO<sub>2</sub> exchange) by creosote bush (*Larrea tridentata*) in Death Valley, California. Regardless of season, leaf photosynthesis spanned a broad range of temperatures, but the photosynthetic optimum temperature changed from ~20 °C in January to ~32 °C in September, reflecting temperature changes of the environment (Mooney et al. 1978, reproduced with permission of American Society of Plant Biologists)

“if you can’t beat them, join them.”) In a number of desert species, biochemical adjustments do just this. Such physiological acclimation results in changes of the optimum temperature for photosynthesis that closely match seasonal differences in ambient temperatures (Fig. 5). *Thermal acclimation of photosynthesis* is found primarily in evergreen plants across many forms (e.g., shrubs, grasses, succulents, and ferns) and appears to be uncommon in annuals and drought-deciduous perennials, presumably because these two growth forms do not experience the breadth of leaf temperatures that evergreen species do.

Creosote bush (*Larrea tridentata*, Zygophyllaceae) is often cited as the quintessential thermal acclimating desert species, showing temperature optima changes in the field from 20 °C in January to 32 °C in September. Importantly, these changes could be replicated in reciprocal transplant and controlled temperature experiments, thereby confirming the response to be acclimation based specifically on temperature.

### Non-leaf Photosynthetic Structures

Photosynthetic stems and twigs are present in plant species throughout the world, but this trait is of special interest in deserts where highly modified green stems are found in a number of drought-deciduous, microphyllous, and aphyllous woody species. Furthermore, unlike most species from other biomes, stem photosynthesis in these desert plants often contributes significantly to net carbon gain. The syndrome is most well studied in the deserts of North America, especially for the microphyllous tree species of the Sonoran, and although the ultimate cause of stem photosynthesis may be debated, most studies have demonstrated that stem photosynthesis confers a number of ecophysiological benefits for growth in arid and hot conditions.



When water is abundant, the majority of plants with photosynthetic stems flush a cohort of small leaves that have high transpiration and photosynthetic rates. As drought ensues leaves are abscised but carbon gain continues in the photosynthetic stems. These stems usually have lower net photosynthetic rates but greater water-use efficiency than leaves. Some studies have also shown stems to have higher temperature optima for photosynthesis than leaves. These trends suggest that these highly modified photosynthetic stems are well adapted for operation during dry and potentially hot conditions of deserts, enabling year-round carbon gain for many species and potentially facilitating more rapid responses to pulses of water availability.

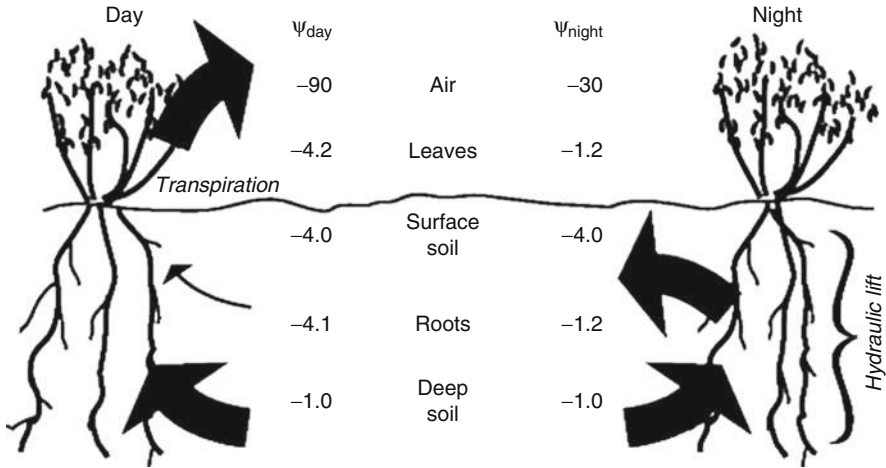
The costs associated with stem photosynthesis (e.g., construction costs and lower carbon assimilation) may be high compared to photosynthetic leaves, but those costs appear to be outweighed by the benefits. For some species the contribution of photosynthetic stems and twigs to annual plant carbon gain is important, as it can exceed 70 % (Szarek and Woodhouse 1978) and extend carbon uptake by 7 months (Tinoco-Ojanguren 2008). Furthermore, stems play other structural roles that should also be considered in the benefits, as not all plants with green photosynthetic stems engage in exogenous gas exchange. Instead, these species benefit from stem photosynthesis through the re-fixation of respired CO<sub>2</sub>, which may help maintain reserves of stored carbohydrates.

## **Adaptive Forms and Functions Related to Desert-Plant Water Relations**

### **Roots that Increase Water Uptake**

Maintaining a favorable water balance is a clearly one of the primary challenges to living in water-scarce desert environments. Thus, it is not surprising to find a number of adaptive traits related to increased water uptake and the prevention of water loss. One of the simplest solutions for achieving greater uptake is to have a large rooting system, but this is not as common as one might expect in desert plants, probably because the rooting zone eventually dries and the maintenance costs of a large root area would be unsustainable. More effective strategies present in desert plants include rapid production of new roots in response to rainfall (described below for cacti) and development of roots that exploit more favorable microhabitats in the soil. One example of the latter are plants that produce long roots capable of accessing the more permanent water supply found in the saturated soil zone (i.e., permanent water table). These deeply rooted species are called *phreatophytes*, and many of the most deeply rooted plants in the world are phreatophytes from arid or semiarid regions. For example, *Boscia albitrunca* (shepherd's tree) of the Kalahari semidesert in Africa has the deepest roots ever measured, at 68 m.

Phreatophytes are found in most deserts, possibly because water tables tend to be deeper in deserts than other ecosystems and the phreatophytic habit confers such a significant fitness advantage. In North American deserts, phreatophytic species have the highest primary production and standing biomass of these ecosystems.



**Fig. 6** Hydraulic lift is the process of water being moved from areas in the soil with high water potential to areas with low water potential via plant roots. This occurs at nighttime when stomata are closed and transpiration of water from the leaves is shut down. The translocated soil water may serve as a reserve for plant uptake the next day when transpiration resumes

Such high productivity is partly due to phreatophytes being largely decoupled from surface drought conditions. Such decoupling allows phreatophytes to be productive throughout rainless periods, even when those rainless periods extend for years – as for *Prosopis tamarugo*, a deeply rooted species of the hyper-arid Atacama Desert of Northern Chile.

Deep roots are also commonly found in plants that display a functional process described as *hydraulic lift* (or *hydraulic redistribution*). Popular accounts of this process describe it as “self-watering” by plants, whereby water from zones of high water potential (usually deep soil) is nocturnally redistributed through roots to zones of low water potential (i.e., shallow soils) and stored there until daytime when the plant takes up the stored shallow water for transpiration. This phenomenon was first quantified in the ubiquitous Great Basin Desert shrub *Artemisia tridentata* (big sagebrush) and coined “hydraulic lift” because water movement was in an upward direction. It has since been found in other desert species with roots that experience a hydraulically heterogeneous soil profile. In spite of the apparent fitness value of hydraulic lift, its presence has not been widely examined in most ecosystems, including deserts.

Hydraulic lift (Fig. 6) is driven by water potential differences that develop in the soil profile during the day. Root densities are typically greatest in shallow soil and decrease with depth, as such transpiration depletes water in the shallow layers to a greater extent than that at depth (especially if the deep roots are near saturated soils). Evaporation also contributes to a higher loss of water from shallow soils. The result is a water potential gradient in the soil profile that is bridged by the roots of the plant. At night, when stomata close and transpiration is greatly reduced, water continues to move within the plant along the residual water potential gradient.



Movement continues from the roots to the shoots until the shoot water potential comes to equilibrium with the root water potential. Below ground, in a similar manner, water fluxes from the deep roots in moist soil (high water potential) to the shallow roots in soil that has dried through the day (low water potential). The surprising aspect of this process is that the water “leaks” out of the shallow roots and into the surrounding dry soil, meaning that water movement in these shallow roots reverses. The hydraulically lifted water accumulates in the shallow root zone overnight, where it is then available for uptake by the plant the next day – when transpiration resumes and the shallow root flux again reverses.

As might be expected, water redistribution by roots can also occur in the opposite direction – inverse hydraulic lift. The movement of water from wet upper soil layers (such as after a monsoon rain) into dryer deep layers may benefit root growth into deep soil and can redistribute water away from access by shallow-rooted competitors. Inverse hydraulic lift has been demonstrated for a number of desert plants and across life forms, including Kalahari dune grasses and a Chihuahuan tree (Arizona walnut). The facultative phreatophyte, *Prosopis velutina* (velvet mesquite) of the Sonoran Desert has even been shown to engage in both hydraulic lift and inverse hydraulic lift.

Recently, hydraulic lift was shown to facilitate greater nutrient availability to plants. This results from microbial activity in shallow root zones being stimulated by hydraulically lifted water exuded by the plant. This important discovery adds another dimension (“self-fertilization”) to the value of hydraulic lift.

### **Preventing Loss of Water Transport**

Water for transpiration is moved by negative pressure (i.e., tension) from the roots to stems and leaves via the xylem. (Imagine the xylem as a straw, through which water is sucked from soil to air.) When soil water is abundant, little tension is required to move the water through this transpiration stream, but during drought, greater and greater tensions are needed to pull water up this transpiration column. Higher tensions increase the probability of breaking the water column due to *cavitation*, the change of water in a xylem conduit from liquid to vapor. Water-stress-induced cavitation occurs when xylem tensions becoming so great that they exceed the capillary forces that sustain water-filled conduits in the xylem. Once a conduit fills with vapor, it can no longer transport water. Cavitation is normal in most plants, and refilling of the conduit is possible in some species, but when cavitation is rampant throughout the xylem, water transport is substantially reduced. Because desert plants typically operate under conditions of low soil water availability, they regularly face high xylem tensions. For this reason, desert plants would be expected to have adaptive anatomical features that help resist cavitation to a greater degree than plants from more mesic habitats, and this appears to be the case.

Broad surveys indicate that plants of from arid and semiarid regions are less susceptible to water-stress-induced cavitation than those from more mesic habitats. One mechanism for this greater resistance is smaller pit pores, but this pattern is only strong for perennial evergreens – the group of plants that typically remain active during the drought period. For other life forms, short-term reductions of



**Fig. 7** Leaf and stem succulent species dominate the vegetation of the Vizcaíno region of the Sonoran Desert. Centered in this photo is the elephant-stemmed *Pachycormus discolor*. The large columnar cactus to the left is Cardón, *Pachycereus pringlei*, and the tall slender plant to the right is Boojum tree (*Fouquieria columnaris*). A number of other succulents from the Agavaceae and Cactaceae families are also present in this scene (Photo: D. R. Sandquist)

transpiration by stomatal closure help to prevent cavitation. Alternatively, some life forms are only active when cavitation is not likely (e.g., annuals, drought-deciduous perennials). In fact, resistance to water-stress-induced cavitation is not strongly associated with mean annual precipitation (MAP) within deserts and may actually increase at the lower end of MAP owing to greater reliance pulsed water availability and the high water fluxes needed for growth during the ephemeral periods of water availability.

### Storing Water: Succulent Plants

To many, succulent plants, especially cacti, are synonymous with desert life, even though these “denizens of the desert” are actually absent or rare in the most arid deserts. They are also rare in high-elevation and high-latitude deserts because succulent plants have a low tolerance for freezing temperatures. Nonetheless, succulence is a successful syndrome for desert survival that nicely illustrates the coupling of morphological and physiological functions. It is also hard to deny that desert regions favorable to abundant succulent plant growth create some of the most intriguing landscapes on the planet (e.g., Vizcaíno region of the Sonoran Desert Fig. 7, Karoo region of South Africa).

Succulence refers to the fleshy and relatively thickened tissues of a plant that store water which can be used during periods of water stress (Von Willert et al 1992; Egli and Nyffeler 2009). The requirement of being able to store water is important to this definition because there are plants that appear succulent but wilt or die when exposed to drought. These are not succulent in spite of being water rich and fleshy.

For truly succulent plants, the degree of succulence varies greatly across species, but overall it is adaptive because it allows a plant to become temporarily decoupled

from unfavorable environmental conditions (e.g., low soil water, high soil salinity) and facilitates growth and survival through such periods. Cacti are arguably the most recognizable of the succulent species with their fattened stems and spiny armor, but succulence is present in 30 of 50 plant orders (Eggle and Nyffeler 2009), most of which are represented in arid and semiarid habitats. These taxa show great diversity of form and physiology in spite of having the similar ultimate function of attenuating water stress.

In places, succulents may dominate the biomass or diversity of a desert. The arid and semiarid regions of southern Africa that include the Namib Desert and Succulent Karoo harbor the highest diversity of succulents – equaling approximately 1/3 of the estimated ~10,000 succulent species globally. Parts of the Sonoran Desert are so influenced by the presence of succulent species that Forrest Shreve relied heavily on them for delineating four of six vegetational subdivisions (Shreve and Wiggins 1964). The Arizona Upland is *crassicaulescent* (succulent stem cacti), the Central Gulf Coast is *sarcocaulous* (fleshy stem trees), the Vizcaíno region is *sarcophyllous* (succulent leaf), and the Magdalena region is *arbocrassicaulescent* (tree and stem succulent). These divisions also highlight the most common groupings of succulence among species: *leaf succulents* (e.g., aloe and yucca), *stem succulents* (e.g., most cacti), and *caudiciform succulents*, whose succulent parts may include non-photosynthetic portions of the stem, the upper part of the root and the root proper (e.g., many *Euphorbia* sp.).

Given the pulsed nature of rainfall in most arid ecosystems, rapid uptake of large quantities of water is important for succulent species. To accomplish this, many succulents have extensive root systems that are often only a few centimeters below the soil surface (e.g., cholla and barrel cacti). Another adaptive feature of the roots of some succulent species is the very rapid formation of new roots when water is present. These *rain roots* form within a couple days of wetting and die once the soil is again dry. Thereafter the main root system is impermeable to water uptake and water loss throughout the dry period, which can last for many months. Both shallow roots and rain roots provide a mechanism for succulent plants to rapidly take advantage of ephemeral water availability and small rain events that wet just the upper soil layer.

In leaf and stem succulent plants, water is typically stored in the vacuoles. Thus, another feature of succulent plants is the presence of very large vacuoles in the succulent tissues, occupying up to 90 % of the cell volume. These vacuoles also serve another purpose in many succulent species, storage of organic acids associated with CAM photosynthesis. Most succulent species display some degree of CAM photosynthesis (although there are many without any CAM activity). The combination of CAM and succulence represents a structure-function relationship that is remarkably well suited for life in warm and arid environments.

### **Desert Halophytes Face Two Challenges: Water and Salt Stress**

Plants growing near the base of a watershed or drainage basin would normally be expected to have higher water availability, but in deserts, such basins typically also have highly saline soils. High soil salinity is common throughout arid regions due to

high evaporation rates that exceed precipitation input. Dissolved solutes are not leached from these soils; instead, they are concentrated near the soil surface as water evaporates. In low-lying basins, salts are transported with rain runoff from the surrounding elevations and then further concentrated by evaporation. Over many years, this process results in extremely saline “playas” or salt basins. The center of most basins has such high salinity and fine soil particles that no vegetation can establish or survive, but along the margins where particle sizes are larger and salinity is not as extreme, the plant community is usually unique, composed of species that can survive relatively high salinity. Such salt-adapted plants are known as *halophytes*, meaning “salt plants.”

Plants that live in saline habitats but have mechanisms to prevent uptake of salts through the roots are called salt avoiders or excluders. These are not true halophytes because they always grow best in the absence of salinity. In general, salt excluders are not particularly common in deserts because the process of salt exclusion leads to increasingly greater soil salinity in the rooting zone.

True halophytes take up salt minerals (primarily  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cl}^-$ ) through the roots and into the plant tissues; thus they face the challenge of preventing physiological dysfunction and possible cell death caused by the toxicity of high salt concentrations. Controlled balance of cell ionic concentrations through rapid growth and synthesis of compatible organic solutes (i.e., osmotic adjustment), coupled with compartmentalization of the salt ions are keys to salt tolerance in halophytes. Another challenge to growth in saline soils is that salinity causes soil water potential to be lower, making it more difficult for plants to take up water. For halophytes, however, the uptake of salts into the roots facilitates water uptake by lowering the root water potential, thereby counteracting the problem of lower soil water potential.

Some halophytes actually have lower growth in nonsaline soils than in those with modest levels of salinity (i.e., 50–250 mM NaCl) (Flowers and Colmer 2008). All, however, must have mechanisms to prevent the toxic ramifications of high salt concentrations in living tissues. *Salt accumulators* prevent these negative effects by sequestering salts in the vacuoles or other cell structures, thus eliminating interactions between the salts and cytoplasmic components and membranes. Many salt accumulators are succulent because they rely on large vacuoles for this purpose. Examples of succulent salt accumulators are common in the Chenopodiaceae family (e.g., *Salicornia*, *Suaeda*, and *Allenrolfea*), but also from this family are species in the genus *Atriplex* that sequester salts in modified epidermal hairs (salt bladders). Interestingly, the salt bladder can serve an additional beneficial function. As water evaporates from the bladder, the salts precipitate from solution and become white. This increases the albedo of the leaf, which, like the leaf hairs of *Encelia farinosa*, increases leaf reflectance of solar radiation, attenuates heat load, and reduces transpiration (Mooney et al. 1977).

Another mechanism to avoid the toxic effects of high salinity is to excrete the cellular salts onto the outer leaf or stem surface. *Salt excretors* are found across plant functional groups and taxa (e.g., salt cedar tree, *Tamarix*, and salt grass, *Distichlis*). Many rely on specialized salt glands to excrete the cellular salts, where once on the surface, the salt is either washed or blown off or eliminated when the leaf abscises.

## **Biotic-Mediated Processes Are Critical for Nutrient Balance in Deserts Plants**

Although water is the resource that most limits desert plant productivity, nutrients have often been shown to constrain productivity when water is not limiting, even over short periods of time (e.g., during annual plant growth). Nutrient limitations have been documented in many deserts through experimental supplements of water and nutrients (especially nitrogen); however, not all species within a desert respond equally to nutrient supplements – and some do not respond at all (e.g., perennial grasses in Chihuahuan). Owing to such differential responses, the interplay of water and nutrient limitations can have a distinct impact on community composition in deserts.

Nutrient availability in desert soils is typically low and both spatially and temporally heterogeneous. As in many systems with low nutrient availability, plant tissues in deserts have high retention of nutrients, and although resorption of some nutrients can be much higher in desert plants than is typical, plant litter nonetheless returns more nutrients to the soil than any other input. As such decomposition plays a critical role in desert nutrient availability and cycling. In contrast to more mesic ecosystems, decomposition of surface biomass in deserts is dominated by abiotic processes, namely, photodegradation by UV light followed by physical fragmentation by wind or rain. Subsequent burial of degraded tissues completes decomposition via biological processes. Subsurface decomposition is almost entirely biological and can proceed at rates comparable to those in mesic ecosystems. However, the majority of biotic decomposition is controlled by moisture and is therefore episodic (pulse driven) in most desert systems.

Spatial variability of nutrients is also characteristic of desert ecosystems. In most deserts, *islands of fertility* form around the base of shrubs and trees owing to the accumulation of nutrients and its feed-forward effect. Litter that falls from the plant, as well as capture of litter and dust blown across the landscape, contributes to the buildup of nutrients at the base of the plant. (In some cases a coppice mound will also develop from particle accumulation below the plant and erosion around the plant.) Higher nutrient presence, as well as the shading effect of the plant, usually facilitates growth of annuals around the plant base. When these annuals die, their litter will further add to the nutrient island. Burrowing animals are also common at the base of such plants due to the cover provided by the plant, the friability of soils, and the food sources present within the island (e.g., herbaceous plants, seeds, and insects). The burrows influence water infiltration that can improve growth of the shrub or tree, while animal waste and decomposition may further contribute to the nutrients beneath the plant. Another place where nutrients commonly accumulate is surface depressions. Here, litter and soil accumulate due to particulate transport in runoff water and blowing wind.

As in other ecosystems where nutrients may limit primary productivity, biotic fixation of atmospheric nitrogen represents an important nutrient input in deserts. As such, it is not surprising that a few plant taxa having symbiotic relationships with nitrogen-fixing bacteria are common and widespread in arid and semiarid systems.

Among the most widespread are trees in the legume family (Fabaceae), which form nitrogen-fixing associations with *Rhizobium* and *Bradyrhizobium* bacteria. These trees, and the input of nitrogen due to their presence, are important components of desert communities across both the western and eastern hemispheres (e.g., *Acacia* in African and Middle East deserts; *Prosopis* in North and South American deserts). Nitrogen-fixing associations may also be formed between actinomycetes and plants and between free-living bacteria and plants that release root secretions into the rhizosphere surrounding roots. The latter is often accompanied by a rhizosheath, an anatomical specialization of the root that facilitates development of the bacterial association. The importance of these alternative nitrogen-fixing associations in deserts is poorly understood at present.

*Biological soil crusts* play many important roles in arid and semiarid ecosystems, including nitrogen input through nitrogen fixation. Biological soil crusts are an autotrophic microbiotic community composed of cyanobacteria (and other bacteria), green algae, lichens, mosses, and microfungi. Organisms in these microcommunities grow together as a mat or mound that integrates with particles in the top few millimeters of the soil via a network of cyanobacteria and fungal filaments. All arid and semiarid regions of the world have biological soil crusts, and in some places they occupy up to 70 % of the surface cover. In places where such crusts are present, plants often have greater overall biomass and higher tissue nitrogen concentrations (e.g., tissue nitrogen is 9–31 % higher for plants growing among biological soil crust in the Great Basin Desert).

One function of biological soil crusts that contributes to plant nutrition is fixation of atmospheric nitrogen by cyanobacteria and lichens of these microcommunities. This nitrogen is made available to plants through both decomposition of dead biomass and leaking of nitrogen from the cyanobacteria and lichen. For example, the cyanobacteria *Nostoc* has been shown to lose up to 80 % of its fixed nitrogen. This nitrogen enters the soil mostly as  $\text{NO}_3^-$  and is readily available for plant uptake. However, fixation and release of nitrogen are highly variable within and between deserts depending on the species composition of the crust, the soil moisture levels, and the soil temperatures. Biological soil crusts also contribute carbon to the soil microbial communities of deserts, thereby benefiting decomposition and other microbial-mediated processes that impact plant nutrition.

Biological soil crusts may also affect desert plant communities because of their impact on soil water availability, seed germination, and plant establishment. Soil water is typically greater in the presence of biotic crust because it slows the surface movement of water, which allows greater time for infiltration and may reduce evaporation from the subsurface. These benefits are best realized after large or prolonged precipitation events. Small pulses of rain may only wet the biotic crust without ever percolating into the subcrust soil. A number of studies have also shown biological soil crusts to improve or, at worst, not affect germination and establishment of native plants. In contrast, many alien species have reduced germination and establishment on biotic crusts. Such findings imply an evolutionary response of native desert plants to the presence of biological soil crust, but few hypotheses based on this context have been tested.

*Mycorrhizal fungi* facilitate uptake of water and nutrients of desert plants in the same manner as they do for other species, and they appear to be as widespread among desert plant families as for those of other ecosystems. Most desert mycorrhizae are of the arbuscular type. They improve uptake of water and nutrients because the extensive network of fungal hyphae greatly increases the functional surface area for uptake while exploiting a greater soil volume than do roots alone. *Dark septate endophytic fungi* is another group of fungi that form associations with desert plants. This group appears to be equally wide spread in deserts as mycorrhizae, perhaps more so. Their prevalence has led most authors to conclude that they play an important role in desert ecosystems and for desert plants (including for water or carbohydrate storage) yet their exact role has been difficult to elucidate.

---

## Desert Biodiversity and Community Composition

### Species Diversity Can Be Surprisingly High in Deserts

On a global scale, species diversity generally correlates with ecosystem productivity; thus deserts would be expected to have very low biodiversity, limited by scarce resources and severe climate conditions. In contrast to these constraints, however, deserts have high spatial variability of the physical environment, which generally facilitates increased species diversity and endemism. Indeed, with the exception of hyper-arid deserts, plant diversity and endemism in the world's deserts are surprisingly high. Species diversity on a local scale (alpha-diversity) can be remarkably high, as in the Negev desert where over 100 species per 0.1 ha can be found in places. In many deserts, annual plant species contribute greatly to this diversity. This is partly due to the greater number of annual individuals that can be supported in any given area compared to perennial plants but also because annuals can escape (as seeds) the constraints of resource limitations and severe climate and then grow and reproduce during periods of high resource availability and low stress.

A number of deserts have notably high biodiversity and endemism. The Chihuahuan desert of North America has nearly 3,500 plant species, including many *edaphic endemics* – species that are restricted to specific substrate types (in this case unusually widespread gypsum soils). Endemism can also result from evolution under unique climate conditions, such as that related to fog in the central Namib Desert.

The Succulent Karoo desert stands out as one of the world's most speciose regions, with over 5,000 plant species and nearly 2,000 endemics. But unlike other deserts, annual species do not dominate the biodiversity of this desert. Instead, the Succulent Karoo, as the name suggests, is home to the world's greatest diversity of succulent species (Fig. 8), including over 1,700 leaf succulents. The region also harbors a great diversity of bulb and bulb-like *geophyte* species (~600). In contrast, there are only 35 tree species. Both alpha-(local)diversity (74 species per 0.1 ha) and beta-(species turnover)diversity (1.5 per 100 m) are high in the Succulent Karoo. The relatively mild climate of the Succulent Karoo region may contribute



**Fig. 8** The Succulent Karoo is considered a global biodiversity hot spot. It harbors over 5,000 plant species including about one-third of the world's succulent species (Photo courtesy of A. G. West)

to its high biodiversity. Winter low temperatures are not extreme and neither are summer high temperatures; the latter are often buffered by cool coastal fogs and dew, which also serve to reduce drought severity and duration. Rainfall, though low (150 mm on average), comes in winter, and unlike many other deserts, it is relatively predictable.

### **Vegetation of Unique Habitats Increases Local and Regional Biodiversity**

There are many unique habitats in deserts that increase plant biodiversity owing to properties quite different from usual desert environments. Some, such as riparian corridors and oases, are anomalous water-rich havens in a sea of drought. Others, like *sky islands*, harbor biota that are not typical desert dwellers, but nonetheless interact with and influence desert plant communities. The concept of a sky island is not restricted to desert systems, but it was first applied to the Madrean sky island mountains found in the Sonoran and Chihuahuan deserts of southwestern North America. Sky island vegetation is composed of taxa that are not desert specific, although desert taxa are often present. Instead, sky island communities are a complex mix of remnant species from the past when the region was less arid, species that have migrated to the mountains in spite of the surrounding desert barrier, and species that have evolved in situ as a result of the isolation. Not surprisingly, biodiversity on sky islands tends to be greater than the lowlands of the surrounding desert. The Madrean sky islands, in fact, are part of the Madrean pine-oak woodland global biodiversity hot spot. Elsewhere, sky islands may not receive enough rainfall to support vast assemblages of vegetation, but nonetheless form plant communities that are distinct from those of lower elevations.



The central Sahara massifs, for example, receive enough water to form shrubland and grassland communities that are different from the surrounding desert and often contain endemic taxa – and where water persists for long periods of time, unique montane wadis communities form.

## **Population and Community Dynamics Are More Complex than Expected**

Plant communities of arid and semiarid ecosystems have often been used as a canvas for examining species coexistence and community dynamics. This may not seem surprising given that such systems usually have relatively few dominant species of only two or three functional types and a strongly limiting resource (water). But this simplicity is misleading, as in recent decades the role of biotic interactions has been more carefully scrutinized in arid and semiarid environments, with both theoretical and empirical evidence growing for its importance.

Perhaps understandably, early studies of desert plant communities placed most emphasis on abiotic controls. In contrast to more mesic systems, deserts have very low biomass and plants are usually so widely spaced that competition seems of relatively little importance to vegetation composition and structure. Furthermore, prominent desert ecologist, such as Forest Shreve working in the Sonoran Desert during the first half of the twentieth century, were immersed in the debate of succession theory, dominated at the time by Frederic Clements. Biotic interactions (i.e., competition) are implicit to the Clementsian theory of succession – yet to workers like Shreve, desert communities appeared to have little species succession, if any. The alternative explanation for succession, raised by Henry Gleason and later Robert Whittaker, of individualistic responses by species to environmental factors seemed more tenable for arid systems. Thus, abiotic controls were widely embraced in the desert literature, while biotic interactions were largely dismissed.

## **The Role of Competition in Shaping Desert Communities**

Over the past few decades, a number of experimental and observational studies have brought biotic interactions to the forefront of desert community ecology. Abiotic factors still play a dominant role in broad-scale patterns of diversity, but increasingly biotic factors are being identified as drivers of community- and population-level dynamics. Competition has been implicated as a driver of plant spatial patterns in deserts, in that observations of *regular* (equal) spacing of plants across a landscape are interpreted as a result of strong competition that minimizes interactions. Although disputed, this interpretation has been reported for both intra- and interspecific plant patterns. But just as often, desert plant patterns are *random*, an indication of neutral overall interactions, or *clumped* (also called contagious) indicating potential facilitation between plants or clustering of plants within favorable microhabitats. (Although the latter implies an absence of competition, seasonal changes in resource availability can result in temporal variability of competition.)

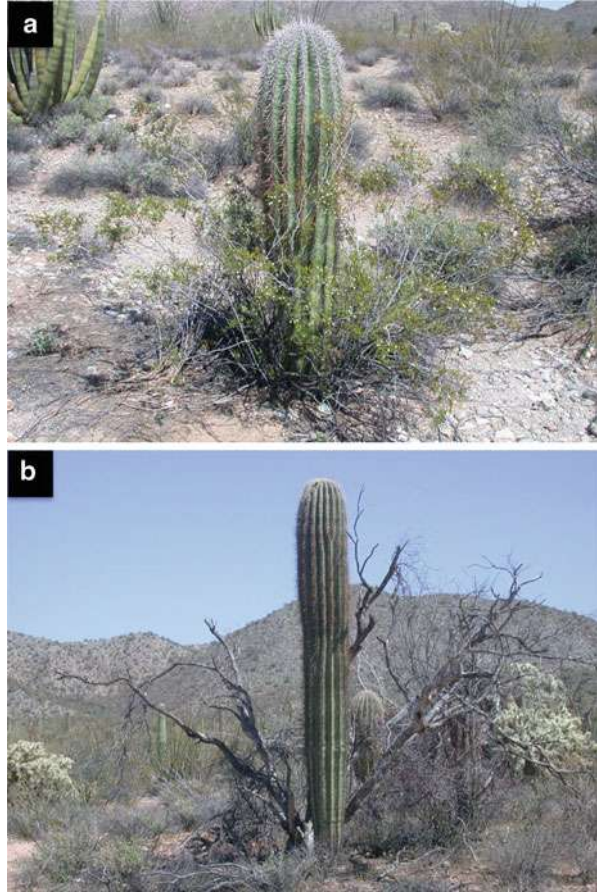
More direct evidence of competition has been elucidated from field manipulations of plants and plant resources. One early and widely cited example includes *Larrea tridentata* (creosote bush) and *Ambrosia dumosa* (white bursage), two of the most ubiquitous species of North American deserts. Widespread coexistence of these two species led to questions about belowground resource competition. In a series of studies, Fonteyn and Mahall (1978, 1981) used different combinations of plant removals (e.g., removal of *Larrea* only, *Ambrosia* only, neither, or both) to demonstrate both the presence of interspecific competition for water and that intraspecific competition was weaker than interspecific competition. Later studies by this group identified two mechanisms for these outcomes: inhibition of root growth mediated by an apparent exudate from *Larrea* roots (*allelopathy*) and avoidance of overlapping growth due to physical contact between roots of *Ambrosia*. Since these studies, a number of other neighbor-removal experiments in deserts have confirmed that competition, especially for water, is common both within and among species and for plants showing regular, clumped, and even random distributions.

Interspecific competition is one mechanism expected to lead to *resource (or niche) partitioning* between species. Studies on coexistence among desert plants have been instrumental in testing and, in many cases, verifying this concept, and although resource partitioning may not lead to full elimination of competition, it helps to minimize it. One widely used framework for such investigations is the “Walter two-layer hypothesis” attributed to German ecologist Heinrich Walter (1939, reviewed in Ward et al. 2013). The two-layer hypothesis predicts that species may coexist by partitioning belowground water resources such that one species relies primarily on shallow soil water and the other on deeper soil water. Originally proposed to explain coexistence of savanna grasses (shallow rooted) and trees (deeper rooted), this model has proved robust in deserts (reviewed by Ward et al. 2013). Coupled with an understanding of phenological differences among species, the Walter two-layer model has also proved valuable for understanding different interspecific responses to amount and seasonality of precipitation in deserts.

### **Other Species Interactions in Desert Plant Communities**

Seedling establishment in deserts is generally rare and sporadic owing mostly to the great spatial and temporal variability of favorable soil water conditions. When establishment does occur, seedlings of some species are found under the canopy of mature plants more often than expected (i.e., nonrandomly). This plant-plant interaction is called a *nurse plant* or *nurse-protégé* association. The reasons for this pattern can vary among species and may also change through time, but at the establishment stage, this association is one of *commensalism*, whereby the nurse plant facilitates establishment of the protégé, but is not affected by its presence. Nurse plant associations are more often reported in arid and semiarid environments than elsewhere, supporting the tenet that facilitation is most common in stressful environments. However, the relationship often changes as the protégé grows out of the difficult establishment period and eventually becomes a competitor with the nurse plant.

**Fig. 9** (a) A young saguaro (*Carnegiea gigantea*) growing from within the canopy of its nurse plant, creosote (*Larrea tridentata*). (b) Facilitation by the nurse plant may change to competition as the protégé saguaro matures, potentially leading to death of the nurse plant. The nurse tree here is palo verde (*Cercidium microphyllum*) (Photos: D. R. Sandquist)



Nurse plant associations are found among many plant families and thus do not appear to be phylogenetically restricted; however, in deserts the association strongly benefits establishment of CAM succulent species of the Cactaceae family. In the establishment period, tender CAM seedlings are sensitive to many environmental and biotic forces that are ameliorated by the presence of the nurse plant. The nurse canopy reduces direct solar radiation and high temperatures by shading and attenuates low overnight and winter temperatures. Surface water availability may also be greater beneath a nurse canopy due to shading or from water supplemented by hydraulic redistribution. Nurse plants also offer physical protection from herbivory and strong winds, the latter of which may also cause desiccation of young seedling plants. It is likely that a combination of these factors results in the nurse-protégé relationships found among desert cacti and other species.

The saguaro cactus (*Carnegiea gigantea*) of North America's Sonoran Desert is one of the most well-studied protégé species of the nurse-protégé syndrome (Fig. 9). Saguaro has multiple nurse species, but the palo verde tree (*Cercidium*

*microphyllum*) appears to be most common. Saguaro is protected from herbivores by palo verde, but studies have also demonstrated the importance of the microclimate under palo verde for facilitating saguaro establishment. For example, by decreasing nighttime loss of longwave radiation, temperatures above small saguaros under palo verde canopies are up to 10 °C greater. This appears to contribute to a more northerly distribution of saguaro than would be expected in the absence of a nurse plant association.

---

## **Disturbance, Global Changes, and Future Challenges**

### **Disturbances Pose Significant Challenges in Low Productivity Ecosystems**

Any ecosystem with low productivity and episodic recruitment will be slow to recover from disturbance. Deserts are no exception and in fact represent an extreme example of this tenet. For that reason, studies of responses to disturbance and recovery in arid and semiarid regions provide critical information about human impacts on ecosystem processes. Among the many types of anthropogenic disturbances that occur in desert systems, grazing of domesticated animals is one of the oldest. An obvious link exists between livestock presence and biomass decline due to herbivory, but grazing also causes soil disturbances that can reduce nutrients, increase erosion, and destroy beneficial biotic crusts. Such disturbances can also lead to loss of biodiversity and increased invasion by nonnative plant species. On a more positive note, much attention is now being given to determining sustainable practices of grazing in arid and semiarid systems.

### **Nonnative Species Are a Major Threat to Desert Communities**

Owing to the harsh growing conditions of deserts, they are considered relatively resistant to invasion by nonnative species; however there are numerous examples of successful and widespread invasion in arid and semiarid ecosystems. One of the most troubling consequences of nonnative spread in desert ecosystems is the increase of fire where alien grasses invade desert shrubland. Low biomass and relatively large bare spaces between plants in a natural desert shrublands mean that fires rarely spread if started. Invasive grasses add a fine-textured fuel to the system that often occupies the shrub interspaces and once ignited easily carries fire from shrub to shrub (i.e., artificially increasing fire spread). Furthermore, because many grasses are adapted to recovering from fire and desert shrubs are not, grass cover increases at a much greater rate following fire than that of shrubs. Subsequent fires may eliminate shrubs altogether, resulting in an *ecosystem-type conversion* from native shrubland to alien grassland. Such conversion has pronounced impacts on biogeochemical cycles that are very difficult to return to preinvasion levels, even with intensive restoration efforts.

## Other Global Changes also Threaten Desert Regions

Many other impacts of human activities affect deserts, including global warming, elevated CO<sub>2</sub>, altered rainfall patterns, air and soil pollution, and habitat fragmentation. Owing to the underlying complexity inherent to arid and semiarid ecosystems (e.g., high spatial and temporal climate variability), ecologists are just beginning to understand the long-term consequences of these impacts. The stochastic nature of plant recruitment and mortality in deserts also hinders our ability to understand effects that are small or gradual through time (e.g., increases in temperature). Only through very long-term observations or detailed modeling efforts can one begin to understand the future of desert vegetation in the changing climate on Earth. Nonetheless, important information is beginning to emerge. For example, in a 10-year study of elevated CO<sub>2</sub> in the Mojave Desert, no changes were seen for plant cover, diversity, or richness; however using remotely sensed data over a longer time frame and larger area, the effects of increased atmospheric CO<sub>2</sub> appear to have caused a modest increase in arid-land plant cover.

Human pressures on deserts are as great as ever. Ironically, these problems are increasingly at-odds with other environmentally favorable activities, such as the growing demand for lands with high solar radiation or strong winds for development of renewable energy projects. Such conflicts mean that high-quality, scientific understanding of these landscapes is more important than ever. This understanding will enable intelligent management decisions that allow sustainable use but minimize inevitable losses of desert biota and mitigate impacts on important ecosystem processes.

---

## References

- Eggle U, Nyffeler R. Living under temporarily arid conditions – succulence as an adaptive strategy. *Bradleya*. 2009;27:13–36.
- Ehleringer JR, Monson RK. Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu Rev Ecol Syst*. 1993;24:411–39.
- Ehleringer J, Mooney HA, Gulmon SL, Rundel P. Orientation and its consequences for *Copiapoa* Cactaceae in the Atacama Desert, Chile. *Oecologia*. 1980;46:63–7.
- Ezcurra E, United Nations Environment Programme. Division of Early Warning and Assessment. Global deserts outlook. Nairobi: United Nations Environment Programme; 2006.
- Flowers TJ, Colmer TD. Salinity tolerance in halophytes. *New Phytol*. 2008;179:945–63.
- Fonteyn PJ, Mahall BE. Competition among desert perennials. *Nature*. 1978;275:544–5.
- Fonteyn PJ, Mahall BE. An experimental analysis of structure in a desert plant community. *J Ecol*. 1981;69:883–96.
- Kearney TH, Shantz HL. The water economy of dry-land crops, Yearbook of the United States Department of Agriculture 1911. Washington, DC: Government Printing Office; 1912.
- Meigs P. World distribution of arid and semi-arid homoclimates. In: UNESCO, editor. Reviews of research on arid zone hydrology. Paris: United Nations; 1953.
- Mooney HA, Ehleringer J, Björkman O. The energy balance of leaves of the evergreen desert shrub *Atriplex hymenelytra*. *Oecologia*. 1977;29:301–10.
- Mooney HA, Björkman O, Collatz, GJ. Photosynthetic acclimation to temperature in the desert shrub *Larrea divaricata*. *Plant Physiology*. 1978;61:406–10.

- Shantz HL. Drought resistance and soil moisture. *Ecology*. 1927;8:145–57.
- Shreve F, Wiggins IL. *Vegetation and flora of the Sonoran Desert*. Stanford: Stanford University Press; 1964.
- Smith SD, Monson RK, Anderson JE. *Physiological ecology of North American desert plants*. Berlin: Springer; 1997.
- Szarek SR, Woodhouse RM. Ecophysiological studies of Sonoran Desert plants 4. Seasonal photosynthetic capacities of *Acacia greggii* and *Cercidium microphyllum*. *Oecologia*. 1978;37:221–30.
- Tinoco-Ojanguren C. Diurnal and seasonal patterns of gas exchange and carbon gain contribution of leaves and stems of *Justicia californica* in the Sonoran Desert. *J Arid Environ*. 2008;72:127–40.
- UNESCO. *Map of the world distribution of arid regions: explanatory note*. Paris: UNESCO; 1977.
- von Willert DJ, Eller BM, Werger MJA, Brinckmann E, Ihlenfeldt H-D. *Life strategies of succulents in deserts: with special reference to the Namib Desert*. New York: Cambridge University Press; 1992.
- Ward D, Wiegand K, Getzin S. Walter's two-layer hypothesis revisited: back to the roots! *Oecologia*. 2013;172:617–30.

## Further Reading

- Evenari M, Noy-Meir I, Goodall DW. *Hot deserts and Arid Shrublands*. New York: Elsevier; 1985.
- Ward D. *The biology of deserts*. Oxford/New York: Oxford University Press; 2009.
- Whitford WG. *Ecology of desert systems*. San Diego: Academic; 2002.

## Web Resources

- <http://worldwildlife.org/biomes/deserts-and-xeric-shrublands>
- <http://www.eoearth.org/view/article/168410>

Matthew J. Germino

## Contents

Introduction .....	328
Alpine and Subalpine Areas Are Valuable for Studying Climate Responses .....	328
Alpine and Subalpine Areas and Vegetation Provide Key Ecosystem Services in a Warming Climate .....	329
Outline for This Chapter .....	330
General Description of Alpine Vegetation .....	330
The Environmental Template for Alpine: Soils and Climate .....	334
Soils .....	334
Microclimate and Energy Balance .....	335
Intra-alpine Site Variability .....	343
Physiological Responses .....	345
Generalized Stress Response and Growth Strategies .....	345
Carbon and Nitrogen Storage .....	345
How Do Specific Climate Stresses Occur, and What Are the Physiological Responses? ....	348
Temperature Stress .....	348
CO <sub>2</sub> Availability and Photosynthetic Assimilation .....	349
Radiation Stress .....	351
Desiccation Stress .....	353
Linking Microsite, Plant Form, and Physiology in Alpine Plants .....	353
Patterns of Tree Establishment .....	354
Ecological Significance of Microclimate Amelioration and Facilitative Interactions ....	356
Future Directions .....	360
References .....	361

---

### Abstract

- Alpine and subalpine plant species are of special interest in ecology and ecophysiology because they represent life at the climate limit and changes in their relative abundances can be a bellwether for climate-change impacts.

---

M.J. Germino (✉)

Forest and Rangeland Ecosystem Science Center, US Geological Survey, Boise, ID, USA

e-mail: [mgermino@usgs.gov](mailto:mgermino@usgs.gov); [germmatt@isu.edu](mailto:germmatt@isu.edu)

- Perennial life forms dominate alpine plant communities, and their form and function reflect various avoidance, tolerance, or resistance strategies to interactions of cold temperature, radiation, wind, and desiccation stresses that prevail in the short growing seasons common (but not ubiquitous) in alpine areas.
- Plant microclimate is typically uncoupled from the harsh climate of the alpine, often leading to substantially warmer plant temperatures than air temperatures recorded by weather stations.
- Low atmospheric pressure is the most pervasive, fundamental, and unifying factor for alpine environments, but the resulting decrease in partial pressure of CO<sub>2</sub> does not significantly limit carbon gain by alpine plants.
- Factors such as tree islands and topographic features create strong heterogeneous mosaics of microclimate and snow cover that are reflected in plant community composition.
- Factors affecting tree establishment and growth and formation of treeline are key to understanding alpine ecology.
- Carbohydrate and other carbon storage, rapid development in a short growing season, and physiological function at low temperature are prevailing attributes of alpine plants.
- A major contemporary research theme asks whether chilling at alpine-treeline affects the ability of trees to assimilate the growth resources and particularly carbon needed for growth or whether the growth itself is limited by the alpine environment.
- Alpine areas tend to be among the best conserved, globally, yet they are increasingly showing response to a range of anthropogenic impacts, such as atmospheric deposition.

---

## Introduction

### **Alpine and Subalpine Areas Are Valuable for Studying Climate Responses**

The alpine zone contains low-statured, non-arboreal vegetation that is distinct from lower-elevation, subalpine vegetation, such as forests and occasionally shrub or grasslands. The highest elevations that vascular seed plants occur at are above 6,000 m in the Himalayas to near 3,000 m lower in high-latitude, maritime-influenced mountains such as in New Zealand to much lower elevations nearer to polar latitudes having arctic influences. Much of the area within a particular alpine area may be unvegetated, particularly at the higher elevations or more exposed sites. Ground cover may consist of bare rock or soil and with occasional herbs or dwarf shrubs nestled into features that collect snow. Although alpine areas comprise only about 3 % of the land on earth, they are distributed across nearly all latitudes and are highly appreciated for a range of values and ecosystem services they provide to humans. As a result, alpine areas receive considerable attention given their scarcity.



The study of plants in the alpine and subalpine has played a foundational role in fields such as plant ecophysiology and ecology. Elevation gradients and topography create a pervading physical template for alpine ecosystems. Plant species reach their low-temperature climate limit in alpine areas and ecosystem effects of climate are relatively transparent and tractable in the alpine compared to many other habitat types. The traits of plant species in alpine environments epitomize selection for stress resistance, versus traits that enhance ruderal or competitive abilities in more disturbed or temperate growing conditions and complex blends of these strategies that occur in other habitats. A number of important theories on plant-climate relationships, such as on microclimate, stress physiology, resource storage, population genetics, and facilitation in plant communities, have roots in classic studies conducted in alpine environments.

Alpine areas are typically near the upper reaches of mountains, and much of the global alpine area is in relatively small patches referred to as “sky islands” subtended by distinctly different, subalpine ecosystems. Considerable heterogeneity in microclimate, snow cover, vegetation, and soils typically occurs within alpine areas as an outcome of topographic relief and extreme climate. As a result, boundary or edge effects and flow between alpine and surrounding ecological zones are relatively important for alpine ecology (Seastedt et al. 2004). Montane and particularly alpine areas also provide the opportunity for upward migration and refugia for species as climate warms (Grabherr and Pauli 1994).

### **Alpine and Subalpine Areas and Vegetation Provide Key Ecosystem Services in a Warming Climate**

Alpine zones and plant life in them are iconic for human appreciation for nature and biodiversity in particular. A relatively high proportion of alpine area has conservation status, and many of these landscapes are relatively pristine, at least compared to lower elevations that have been impacted by development, disturbances such as altered fire regimes, and invasive species. These factors all contribute to the suitability of alpine areas for evaluating plant responses to climate.

The mountains that alpine areas occupy are increasingly valued for their role in regional hydrology, globally. High-mountain physiography generates an orographic effect that frequently results in relatively greater precipitation compared to surrounding, lower-elevation landscapes. Cooler temperatures cause a significant proportion of precipitation in most alpine areas to occur as snow. Snowpack is among the most important of natural reservoirs for sustaining inland water bodies, and the evapotranspiration potential of alpine vegetation affects runoff toward lower elevation. The ecology of streams, the economies of irrigated agriculture and hydropower, and the vitality of civilizations in the vast continental drylands of the earth have, in many cases, evolved to capitalize on the predictability of water provisioning from alpine areas. Diversions such as canals and irrigation ditches are common in mountains and frequently extend into alpine basins and accompany dams and reservoirs.

Direct anthropogenic impacts to alpine plant communities are associated with mining and widespread livestock grazing. In many regions such as the semiarid Western USA, subalpine areas hold a disproportionately large potential capacity to support livestock given their small area, compared to the desert rangelands. In recent episodic droughts (and with future desertification), the compromised livestock capacity of lower-elevation rangelands increased pressure for livestock grazing in alpine areas, and the legacy is evident in signs of erosion or erosion control (terracing) in alpine areas such as the Teton, Wasatch, and Manti Ranges in the Rocky Mountains (Ellison 1954). Deliberate burning and soil erosion are disturbances in alpine pastures of the Himalaya and in paramo of the Andes. More typically, the impacts of historic grazing of alpine areas, which are scarcer since the conservation efforts of the 1900s, are likely evident only through yet-to-be-done studies of species changes. Exotic plant invasions are not as commonly reported in alpine communities as they are in temperate ecosystems.

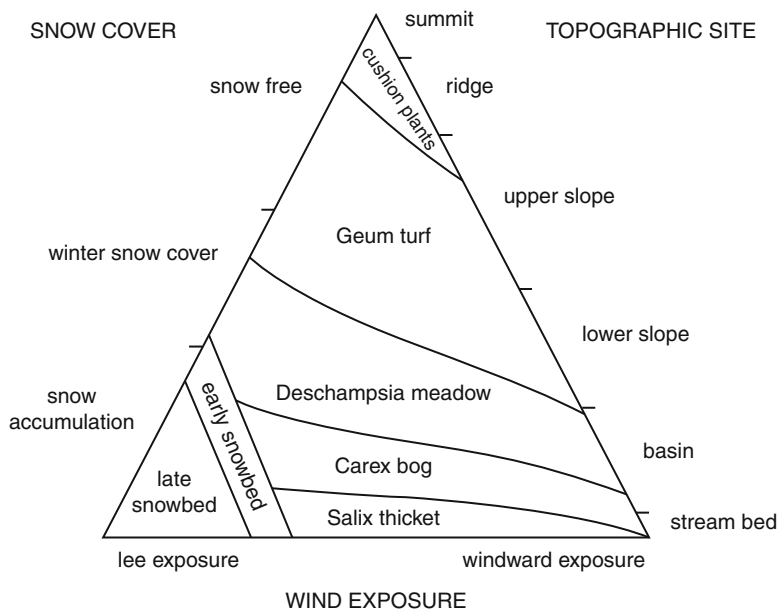
## Outline for This Chapter

The literature on alpine and subalpine vegetation is vast and expanding at a rapid rate with a recent surge in interest on plant-climate relationship. An exhaustively comprehensive overview is not possible in this chapter, but several new syntheses have become available in the last decade or so (Bowman and Seastedt 2001; Körner 2003; Nagy and Grabherr 2009; Lutz 2012). Instead, the focus is to present an overview of the significance of alpine vegetation, an overview of the diversity of alpine plants, basic principles of the climate experienced by alpine plants, how climate factors create stress in alpine plants, and what are basic physiological responses of plants to climate stresses in the alpine. The focus on climate and ecophysiological responses is justified by the relative importance of individual plant responses to climate to the ecology of alpine plant communities.

---

## General Description of Alpine Vegetation

Alpine plants are overwhelmingly perennial angiosperms that are either dwarf shrubs or herbs that are leafy or occur in a mat-like cushion forms (Billings and Mooney 1968, Billings 1974). Annuals are scarce or absent, which likely results from unfavorable conditions for reproductive processes such as seed production and especially seedling survival. In spite of the strong selection for stress resistance, and a general correlation of high ploidy levels and stress resistance in land plants, there is no indication that alpine species have unusually high ploidy levels (Körner 2003). The origin of alpine floras is unique in that many high-mountain areas escaped glaciation (i.e., were “nunataks”), but the sky-island effect likely contributes to endemism and diversity of species assemblages from one alpine to the next (Billings 1974). Many alpine species are also widespread, having global distributions and occurring also in arctic areas or other cool and mesic habitats (Billings 1974).



**Fig. 1** A classic example of plant community variation attributed by plant or community type or genus along elevation (topography), wind, and snow cover gradients in the Beartooth Mountains of Wyoming. *Geum* is a widespread genus of leafy and rhizomatous herbs commonly called “avens,” *Deschampsia* is a grass, *Carex* is a sedge, and *Salix* is willow (Reproduced from Billings and Mooney (1968), with permission)

Alpine areas commonly consist of a sharply contrasted mosaic of different plant communities that vary in their vegetation type along topographic, wind, and snowmelt gradients (Fig. 1).

A high proportion of the species diversity in alpine and subalpine areas can be attributed to herbs that have meristems belowground, including leafy herbs such as the circumboreal alpine sorrel (*Oxyria digyna*), glacier buttercup (*Ranunculus glacialis*), or bulb-forming glacier lily of the Rocky Mountains (*Erythronium grandiflorum*, Fig. 2). Subsurface meristems in these herbs enable avoidance of harsh frosts just before and after snowmelt. Leaves of leafy herbs are generally one to a few cm in width, and they tend to be oriented upright if larger, especially in drier, more exposed locations.

Herbs that have a cushion-like habit, such as the circumboreal moss campion (*Silene acaulis*) and alpine saxifrage (*Saxifraga oppositifolia*), can also be significant to local species richness. Cushion plants are low-statured with small microphyllus leaves that are typically upright in orientation and often within a few cm of the soil surface on windy ridges. Grasses, sedge, and rush species can be highly variable among alpine areas, but some common species include spike trisetum (*Trisetum spicatum*), alpine bluegrass (*Poa alpina*), and in wetter areas *Deschampsia cespitosa*. Bunchgrasses are a major component of paramo grasslands



**Fig. 2** Representative plant forms of alpine zones. *Top left*: an herb with cushion form (*Phlox*) nested into a rock cranny. *Top right*: frozen leaves of the leafy herb *Erythronium grandiflorum* at sunrise, surrounded by snow in a late-lying snowbed that the shoot had emerged through in the previous days. Lower: krummholz with flagged stems emerging on *Picea engelmannii* and *Abies lasiocarpa* at treeline in the Medicine Bow Mountains of Wyoming, USA (Photo credits MJ Germino and W Bowman)

in the Andes. Unique “giant rosettes” are common in tropical or subtropical alpinines, such as the silverswords of Hawaii (*Argyroxiphium*), *Lobelia* of Africa, and *Espeletia* of South America (Rundel et al. 1994). Whereas most other alpine vegetation has small leaves oriented in upright positions, these giant rosettes can have very large and hairy leaves. In *Lobelia*, leaves fold over buds to insulate them at night.

Where trees occur in timberline zones, a frequent pattern with increasing elevation is that (1) large unforested gaps are found in conterminous forest and then at higher elevations, (2) trees become “islands” dispersed within subalpine or alpine vegetation (meadows), then (3) the timber-like structure of trees is lost and near true alpine, and (4) any trees present may instead appear in low-prostrate-like growth forms known sometimes as “krummholz” (German for twisted wood). In many mountains, this transition from forest to alpine can occur over many meters or kilometers, and in others it occurs as a relatively sharp transition, with a change from forest to alpine occurring within just a few meters. The uppermost elevations supporting trees in their timber-like form (e.g., several m or taller) are referred to as timberline, and the uppermost elevations supporting trees in their reduced, often

krummholz form, are referred to as treeline. Extensive mats of krummholz mats that are many square meters in aerial extent and often <1 m height can be formed by species capable of adventitious roots, which emerge from stems pressed to the ground by snow loading. Many refer to the core forest, shrubland, or grasslands that have few if any alpine species but occur just below alpine, as “the subalpine.” However, where the gradient between these subtending communities and alpine forms an ecotonal gradient (rather than sharp cline), such as in broad timberline and treeline zones, alpine species can co-occur abundantly with “subalpine” species. In this chapter, the subalpine zone is defined as that area where alpine and low-elevation species such as trees intergrade – except where references are made specifically to “subalpine forest,” “subalpine shrubland,” etc. Semantics on zonation are not trivial because alpine areas are dominated by gradients, and the semantics affect the efficacy of cross-comparisons among alpine areas.

Tree communities in many subalpine areas of the northern hemisphere are codominated by spruce and fir (*Picea* and *Abies*), which are species capable of forming particularly dense and extensive mats of krummholz. Other timberlines in this zone may have 5-needled pines (*Pinus*) that can have short stature but typically do not have the very dense packing of foliage and expansive mat formation via adventitious rooting in spruce and fir. Broadleaf trees form some high-latitude treelines in Scandinavia (*Betula*) and in the tropical or low latitudes of the southern hemisphere, often with a single species or genus spanning the expansive ranges of forest elevations in a locality. Examples include ohia (*Metrosideros polymorpha*) of Hawaii, snow gums of Australia (*Eucalyptus*), Polylepis of the Northern Andes, and lenga or beech (*Nothofagus*) of the southern Andes and New Zealand.

Biomass and leaf-area indices (LAI, # leaf layers per unit ground area or  $m^2/m^2$ ) and standing crop of biomass are usually relatively low in alpine areas compared to other biomes that receive the same amount of precipitation but are warmer, but leaf-area density (LAD,  $m^2/m^3$ ) tends to be relatively high in alpine vegetation (Körner 2003). For example, LAI can range from 0.5 to 2 and biomass of about 1 to occasionally over 3  $kg/m^2$  for herbaceous alpine meadows. Cushion plants may have a low LAI near 1 yet high LAD. Depending on the species, trees in the subalpine frequently will have a high density of foliage within their crown. For example, krummholz forms of spruce and fir at treeline can have an LAI of 12. The productivity of alpine herb communities is severalfold less than communities with similar physiognomy at lower elevation on a per hectare basis, but the productivity is no less in the alpine if unvegetated ground area and length of growing season are normalized in the comparison.

The landscape patterning of plant community types tend to have compelling linkages to ecophysiological processes. For example, a tendency for trees to be clustered where they occur within the timberline zone is described below. Clustering of species within the landscape can indicate favorable environments, such as patches of suitable soils, but also can result from plant-to-plant interactions such as facilitative or nurse-plant effects, which factor prominently in the ecology of alpine areas.

## The Environmental Template for Alpine: Soils and Climate

The global diversity of latitudes and oceanic influences on mountains and their alpine areas result in some considerable differences in climate, but nonetheless there are climate similarities that are important. At temperate latitudes ( $\sim 40\text{--}44^\circ$ ), alpine/treeline elevations can range from 1,600/1,200 m near coasts to 3,500/3,400 m in inland, continental areas. In tropical areas with little to no winter dormancy, these elevations are 4,400/3,500 m for humid regions and 4,100/3,200 m drier regions.

Given this diversity in elevation among alpine areas, the most pervasive and generalizable environmental difference of alpine areas compared to lower-elevation ecosystems is reduced atmospheric pressure. Accompanying reduced pressure are lower partial pressures of physiological gases (decreases 8 kPa/km in dry regions and 3 kPa/km in wet regions, starting from  $\sim 100$  kPa at sea level), but increased diffusion rates (Smith and Johnson 2009). A number of other factors that are associated with reduced atmospheric pressure are predicted by the ideal gas law, such as lower temperatures, correspondingly more precipitation falling as snow, and less atmospheric attenuation of radiation. However, the temperature, moisture, and radiation actually experienced by alpine plants are so strongly modified by local factors (e.g., axes in Fig. 1) that these climate parameters can increase or decrease as plants experience them at greater elevations within an alpine zone.

One of the more generalizable and important features of alpine environments is the relatively short growth season, bound by winter snow cover or drying or chilling in late summer/fall. Tropical alpins can have year-long growing seasons, or nearly so. Snowmelt generates an intense period of wetting in which productivity can be viewed as energy-limited. Warmer temperatures of summer and reduced precipitation in summer for many continental alpine areas lead to drier growth period where productivity can become water-limited (by both water supply and especially demand, as described below).

Growing seasons can range from one to sometimes 6 months depending on geographic region, microsite, and whether species are deciduous or evergreen or even up to 12 months in the tropics. Alpine areas also have the full spectrum of diurnal and seasonal variation in radiation and corresponding temperature regimes that occur with latitude (i.e., less seasonality in day length at low latitudes). A typical average annual precipitation and temperature of alpine, just above treeline, is  $\sim 1,000$  mm/year of precipitation and average annual temperature a degree or so below zero C. The lower limit of alpine areas tends to fall approximately where the mean temperature of the warmest summer month is  $10^\circ\text{C}$  and temperatures of core elevations of alpine areas range from about 3 to  $8.5^\circ\text{C}$ .

### Soils

Alpine soils are generally shallow and not as strongly stratified as lower-elevation soils that have a longer and more history of weathering, biotic inputs, and stability.

Where soils in alpine areas are derived of the local geologic substrate, they are frequently only partially developed and thus have a high proportion of rockiness and coarse textures. Many alpine areas have soils derived from aeolian inputs (from air), and local fluvial processes and redistributed (and more weathered) fine soil particles within basins, generating pockets of deeper and finer-textured soils in which soil fertility can be partly attributed to upwind or upstream sources. Most alpine areas have a mosaic of soil conditions that covary with the plant community variation. In Fig. 1, one might expect a gradient of decreasing soil depth, fraction of fine particles, and soil fertility from the bottom to top of the triangle. Soil conditions can be exceptionally patchy in alpine and treeline environments as a result of these physical processes and feedbacks from plants on the soils beneath them.

Unlike polar tundra, permafrost is not a common and pervasive factor for the zones of alpine areas that support abundant vegetation. High mountains occasionally have pockets of permafrost at the upper limits of plant life, at depths ranging from near surface to 0.5–1 m. Permafrost in the alpine can result from contemporary hydroclimate, or it can be a vestige of previous glaciers or climatic conditions. Freeze-thaw action causes considerable turbation and forms polygons where coarse textures are sorted, or frost hummocks, or promotes downhill soil creep called solifluction that are examples of cryopedogenesis which have significant effects on plant community structure. Ice crystals that are several cm or longer can protrude through soil following frost events and cause considerable disturbance to plant roots.

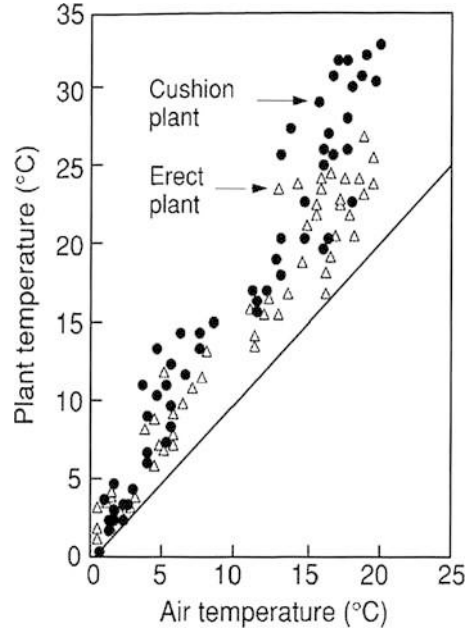
Where organic matter is present in alpine soils, it is frequently in coarse forms, which reflects low-temperature inhibition of microbial decomposition processes, in addition to low inputs of plants. Soil organic matter can range widely among microsites (up to 5–50 % of soil mass) and is generally relatively high in the subalpine and unsurprisingly low on exposed ridges with low plant abundance and also low C/N. Much of the carbon in alpine ecosystems resides in soil, reflecting a small standing stock of plant carbon and slow turnover of plant litter in soil. Alpine soils can have total carbon and nitrogen contents that are similar to low-elevation ecosystem that also have similar soil depths and textures, but these nutrients may be more bound in organic forms and are only slowly mineralized in alpine. Uptake of organic nitrogen, either directly in some cases or more generally through mycorrhizae (soil fungi attached to roots), is likely key to alpine plants. Nitrogen content of leaves tends to increase along elevation gradients, reaching 4–5 % of leaf dry mass for some high alpine herbs as a result of conservation strategies such as reserve formation, efficient resorption and recovery from senescing tissue, and accumulation (Monson et al. 2006; discussed further below).

## Microclimate and Energy Balance

Climate is a dominating factor for alpine plants, and so is emphasized in this chapter. Consideration of the microclimate of alpine plants, and its relationship to site climate, is particularly important for understanding alpine environments.



**Fig. 3** Daytime microclimate temperatures for cushion plants and leafy herbs in the alpine, as cited in Korner (2003)



Air temperature is perhaps the most central climate parameter used to distinguish the climate of alpine areas, at least at coarse scales, but microclimate relates most directly to alpine plant ecology.

Microclimate is the temperature, radiation, and wind experienced by plant tissues such as leaves or flowers. Leaf and plant microclimate is typically very different from the climate of a site. For example, during the day, temperatures of leaves and stems can be elevated considerably above the temperatures of the air surrounding these tissues, particularly when wind speeds are low (Fig. 3). Soil surfaces or leaves in cushion plants or krummholz can become up to 15 °C warmer than the surrounding air under these conditions. On clear-sky and windless nights, these same leaf and soil surfaces can become several to 10 °C cooler than the surrounding air.

Temperature gradients between alpine cushion plants and the soil and air temperatures around them are commonly used to demonstrate the concept of microclimate and to illustrate that plant form can ameliorate the harsh microclimate to enable optimum growing conditions that could never be appreciated from merely relying upon on weather station data to predict ecosystem activity.

The degree to which plant microclimate is coupled to site climate and particularly air temperature differs as a function of climate and plant form: cloudiness and windiness. Both increase the coupling, such that taller plants or plants with small and sparsely arranged leaves have temperatures that are more similar to the surrounding air. Plant tissues that are covered with snow tend to have a temperature of the snow (near 0 °C) and they are not subjected to wind or radiation stresses that prevail above snow. Thus, snow cover is a major factor affecting the climate that plants actually experience in the alpine.



Temperature is a unifying factor both for alpine plant ecology and for relating the interactive effects the main climate factors affecting plants. The ecological effects of low temperatures in the alpine have been addressed many times, and a number of studies found correlations of soil-temperature thresholds or minimum winter temperatures to treeline patterns (Körner 2003, Harsh et al. 2009). Three energy balance parameters help to relate the alpine climates to plants and their temperature regime: (1) *radiative heat exchange*, including both the visible short-wave radiation in sunlight and the long-wave radiation that primarily has thermal influences; (2) *sensible heat exchange*, including conductive heat flow from plant surfaces that are in contact with soil, snow, or water and the more prevalent convective or wind-affected heat flow from plant surfaces to air; and (3) *latent heat exchange* where heat energy is exchanged when water undergoes phase changes from solid (ice, frost), liquid, and vapor phases.

A key point for plants, particularly alpine plants, is that all three of these modes of heat exchange can either add or remove heat energy from the plant. When they have a net effect of adding energy, the plant will be warmer than the surrounding air, which is nearly always in daytime under sun exposure. It is less well appreciated but nonetheless important that a net removal of energy by these heat exchange processes causes plant surfaces to become cooler than the surrounding air. A net heat loss from plants to the surrounding environment and corresponding cooling of the plant below surrounding air temperatures typically occurs during night, but can also occur when high moisture availability results in high transpiration and latent heat loss.

The degree to which leaf and air temperatures are coupled, and the manner in which radiative, convective, and latent heat exchanges interact with one another in regulating plant temperature, can be appreciated from a conceptualization of the energy balance equation, as follows:

$$\text{Leaf (surface)temperature} \approx \text{Air temperature} * \left( \frac{\text{Net radiation} - \text{Latent heat}}{\text{Convection}} \right)$$

The equations for radiative, latent, and convective heat exchange all relate a flux (of photons, water molecules, or heat) to temperatures with coefficients that translate temperatures into energy units, such as Watts. In a nutshell, this equation tells us that the *difference* between the temperatures of leaves or other plant surfaces and the surrounding air is increased by net radiative (e.g., sunlight) or latent heat exchange (e.g., transpiration) and is minimized by convective heat change (wind). The actual energy balance equation, which is solvable, essentially states that under steady state conditions (i.e., leaves or whatever surface of interest are not warming or cooling over time and that all components of the energy balance are in equilibrium), radiative, latent, and sensible heat exchanges must sum to zero. Metabolically generated heat and heat storage sometimes need to also be entered into consideration for alpine plants, for the rare cases where electron flow in mitochondria is not coupled to NADH reduction and for cases where large water stores occur in succulent plants or large stems.

Below, each component of the environment of alpine plants is addressed as it first affects the energy balance of alpine plants and then secondly is related to the alpine site and nonthermal aspect of plants.

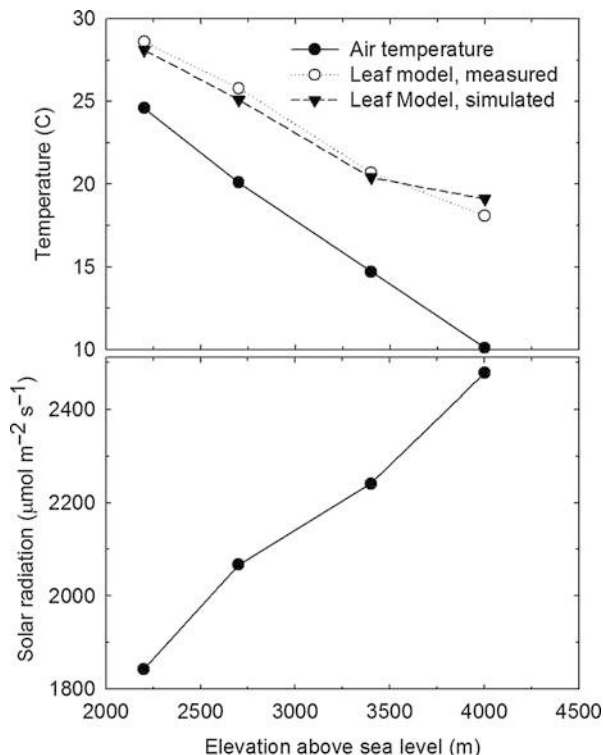
## Temperature

The low temperatures of alpine areas result from the “adiabatic lapse rate,” which refers to the decrease in temperature per increment of elevation as a result of the effect of decreasing atmospheric pressure (and correspondingly, lower density of molecules per unit air volume) and pressure-volume-temperature interactions (i.e.,  $PV = rRT$ ). Lapse rates differ regionally, but as simulated they range from 3 °C/km to 6 °C/km for relatively humid or dry conditions, respectively (Smith and Johnson 2009). In addition to these broad elevation gradients in temperatures, there can be more localized temperature gradients that result from cold-air drainage. Cold air holds less vapor, and vapor is relatively light among the molecules in air. Thus, cold air is dense and either tends to settle near ground or drains along the same watershed paths that water follows. Under clear-sky and windless conditions at night, air temperatures around low-statured alpine herbs can be 5 °C or more below the air temperature at several m height above ground, as recorded by typical weather stations.

The prevailing climate of a given alpine zone can be challenging to measure, given the considerable variability in topography associated with alpine areas. Representativeness of available data is a concern. The temperature measured in a radiation-shielded weather box or gridded models of temperature at ~1 km pixel resolution that are parameterized by and designed to predict air temperature at 2 or more m height is the typical data available. Few weather stations are positioned in a way that can give temperatures representative of the alpine and subalpine zone of interest here (the vegetated zone above forest), and many alpine area patches are not well represented by ~1 km gridded climate models. Thus, considerable uncertainty must exist in our ability to actually know the temperature or climate that prevails upon alpine plants, except that plant and soil temperatures are typically near 0 °C when snow covered.

Which temperature is of interest? The temperatures most commonly used to characterize alpine areas are primarily average annual and minimum temperatures, which are useful in gauging likelihood of snow, but most alpine vegetation is at least partially covered in snow and so climate during the snow-free season is of interest. Minimum (nighttime) temperatures during the snow-free growing season are germane to the majority of alpine plants (except some cushion plants that may become uncovered in winter or trees protruding above snow). Daytime temperatures during this period will relate most directly to the bulk of physiological processing. Alpine areas frequently have exceptionally high diurnal temperature variation. For example, leaf temperatures might increase up to 35 °C (e.g., -7 °C to 28 °C) as air temperatures warm from around 0–18 °C from sunrise to sunset, for leafy herbs, cushion plants, or krummholz on clear days in early summer or fall. It is common for alpine plants to be covered with white surface frost before sunrise, even while air temperatures are reported to be > 0 °C by weather stations. With increasing elevation into the alpine, any surface is likely to warm more above air temperature during days (Fig. 4).

**Fig. 4** Solar radiation (400–700 nm) and corresponding air and leaf model temperatures across an elevation gradient into alpine (3,500–4,000 m). Model temperatures were measured directly and were simulated using the energy balance equation (Regraphed from data from Smith and Johnson (2009))



## Radiation

Solar and long-wave radiation balances in alpine compared to lowland environments may vary according to cloud cover tendencies, and some alpine areas have a greater incidence of clouds than the surrounding lowlands. The reduction in atmospheric molecules, aerosols, or particulate matter at high elevations results in greater radiation exchange in alpine areas (Fig. 4), except where alpinos occur within cloud bands (immersion or high elevation).

### Shortwave, Solar Radiation

In the absence of confounding factors, sunlight availability increases with elevation (Fig. 4). Ironically, some of the highest solar radiation levels ever observed happen to be in continental alpine areas when large cumulonimbus thunderhead clouds form on otherwise clear-sky days. In this condition, sunlight can reflect off of the large white cloud walls and add to the already bright direct beam of sunlight. Snow banks that linger into summer also reflect a considerable amount of sunlight, resulting in maximum solar radiation (for a surface normal to the sun) in the visible wavelengths near  $3,000 \mu\text{mol m}^{-2} \text{s}^{-1}$ . The reflected sunlight onto other faces of the plant adds to this amount. Another significant means by which solar radiation (and temperature) is appreciably greater in alpine areas occurs when closed, lower-elevation basins (such as in the Great Basin, USA) develop wintertime temperature

inversions that effectively trap both cold air and haze over lower elevations while extended high-pressure systems and a corresponding clear-sky and low wind condition prevail on alpine areas.

The solar radiation intercepted by alpine plants is strongly affected by their leaf and plant form. Leaves on most species in the alpine tend toward steeply inclined, upright orientations, which leads to a reduction in the intensity of sunlight intercepted, at least at midday and at midsummer. The reduction in sunlight energy intercepted compared to the amount available is a function of the cosine of the leaf inclination angle (Lambert's cosine law), such that a leaf with a 50° inclination at 45° latitude might intercept less than 1/3 of available sunlight at midday but yet might increase interception in the morning or evening. Leaf orientation is highly plastic in many species, and many herbs and short-needled conifers exhibit steeper leaf orientations in the alpine compared to microsites at lower-elevations or nearer-to-forest canopies. On the other hand, some snowbed herbs and pines show less leaf inclination (e.g., *Caltha leptosepala*, *Pinus flexilis*), but these species often have relatively high physiological tolerance to bright sunlight compared to species like *Abies lasiocarpa* that exhibit both sensitivity to sunlight and steep leaf angles (Germino and Smith 2000). Alpine leaves are also usually relatively thick and have a lower specific leaf area (cm<sup>2</sup>/g), which are attributes common in plants of other sunny and stressful environments (e.g., deserts). Alpine leaves often have leaf hairs (trichomes) that can impart a light color and thus high albedo.

These morphological adjustments tend to be coordinated with anatomical features that affect sunlight as it propagates into leaves (Smith et al. 1997). Multiple layers of mesophyll cells are common in leaves of alpine plants and act to increase photosynthetic capacity per unit leaf area. Differentiation of mesophyll into relatively longer and column-like "palisade" cells tightly packed toward the epidermis aid in channeling light deeper into the leaf. Upright leaves in the alpine tend to have greater isobilateral symmetry, meaning that there is less distinction between the upper and lower surfaces of leaves than horizontal leaves, such as forbs of temperate mesic environments. In many alpine species that have upright leaf orientations, palisade cells occur on both sides of the leaf (ab- and adaxial), and this is accompanied by more even distributions of stomata and other features.

### Long-Wave Radiation Balance

One of the least well-appreciated but very significant aspects of alpine plant microclimate is the long-wave radiation (also called infrared or thermal) exchange between leaves and the environment. The net balance of long-wave radiation exchange is often negative for leaves and thus constitutes an important cooling mechanism that can make a habitat with low air temperatures even cooler for plant surfaces.

All objects emit and absorb long-wave radiation as a function of their temperature, according to the Stefan-Boltzmann equation. Consider a broad leaf of an alpine herb at night that is oriented horizontally with a temperature of 0 °C, and soil temperatures beneath it are the same temperature. The leaf emits and receives radiation from both soil and the sky. The leaf and soil emit the same radiation to

one another, creating a null balance of long-wave radiation exchange for the leaf in its lower hemisphere. In the upper hemisphere, the leaf is emitting the same amount of radiation to the sky as toward the soil ( $300 \text{ W/m}^2$  in each direction), but the clear nighttime sky above the leaf has an effective temperature that can be  $\sim -50^\circ\text{C}$  that might equate with only about  $150 \text{ W/m}^2$  received by the plant from the sky. Radiation from the sparse molecules in air originates from many km above the earth's surface, from the cool atmosphere. The lower molecular and particulate density of the air and atmosphere of high-elevation alpine areas further reduces the incoming radiation relative to lower elevations (Jordan and Smith 1995). The net outcome is net negative radiation balance at night ( $-150 \text{ W/m}^2$ ) that causes the leaf to be cooler than air when the energy balance is at steady state. For a leafy alpine herb with  $\sim 2$  cm-wide leaves, the leaf might be several degrees cooler than air under this negative long-wave radiation balance condition if there is little wind. Larger herb leaves and krummholz shoots can be up to  $\sim 10^\circ\text{C}$  cooler under this condition. Such "radiation cooling" frequently causes plant, flower, and other surfaces to exhibit radiation frost at any month of the year, even when site air temperatures are  $>0^\circ\text{C}$ , and this is particularly important with the cool nights that occur throughout the growing season in the alpine. Long-wave radiation balances can also be negative during day, but incident solar radiation is so large (e.g.,  $1,000 \text{ W/m}^2$ ) that physiological impacts of long-wave radiation become significant primarily at night, particularly clear nights (clouds are much warmer than a clear sky).

Leaf orientation affects net long-wave radiation balances, in addition to interception of sunlight, in ways that greatly affect leaf microclimate and particularly frost occurrence in the alpine. For example, the horizontal leaf with a  $-150 \text{ W/m}^2$  net radiation balance described above might have a net radiation balance closer to  $0 \text{ W/m}^2$  in its upright position, if surrounded by other plants or landscape features. In the case of grass canopies, which contain many leaf blades with relatively upright orientations, the very top of the leaf blades and canopy is exposed to the sky, develops a negative long-wave radiation balance, and can cool well below air temperature, and the resulting frost effects are known to inhibit other species as they emerge above the grass canopy (e.g., trees in subalpine meadows of Australia or Rocky Mountains; Ball 1994; Smith et al. 2003). Notably, the upright leaf orientation in many alpine plants is not always accompanied by other forms of sunlight avoidance such as the high albedo (low absorbance) of many desert leaves (e.g., upright but low-albedo leaves of *E. grandiflorum* in Fig. 2). This may indicate that long-wave radiation balances are relatively influential for alpine leaf morphology, which is a prospect that would require further investigation.

### Day Length and Seasonality

The strong effects of radiation balance at day and night result in appreciable differences in the thermal regime of alpine areas at different latitudes. Tropical alpine areas do not benefit from the additional hours of sunlight and warming that occur at mid to upper latitudes and instead have more hours of radiation cooling at night. Cold soil temperatures tend to linger through the entire growing season in tropic mountain ranges, and there are fewer occurrences of the extreme winter

conditions and dormancy that prevail at higher latitudes. At higher latitudes, long days lead to greater intensity of seasonal warming and a sharp transition from energy-limited growth conditions in early spring (i.e., sunlight or heat-limited) to water-limited conditions as temperature limitations essentially disappear at mid-summer, except for periodic nighttime frosts. Also, whereas lingering snow tends to insulate many alpine plants from the cool nights of spring, there is much greater incidence of intense nighttime frost in autumn, as long nights resume at mid to high latitudes.

### **Latent Heat Exchange and Water**

Water balance for plants is a function of water supply (availability and uptake), water demand of the surrounding environment (loss through transpiration), and water storage. With the exception of trees and large succulent plants of the topical alpine areas, water storage can generally be ignored for alpine plants. Water availability is often (but not always) higher in the alpine compared to lower elevations, though this varies among dry continental compared to wet maritime mountains and it certainly varies strongly within alpine areas due to snow drifting and other topographic and edaphic effects (e.g., soil texture). One of the most consistent aspects of alpine environments is a tendency for greater transpiration than for low-elevation plants that results largely from the increased vapor diffusion rates with reduced atmospheric pressure and also from increased evaporative demand (Leuschner 2000; Smith and Johnson 2009). Leaf-to-air vapor deficits increase with elevation, particularly when adiabatic lapse rates are relatively low (i.e., under humid conditions), mostly due to greater insolation and correspondingly larger leaf-to-air temperature gradients (Fig. 4).

Dewfall and frost deposition are relatively important aspects of alpine plants, and their frequency in both wetter maritime and even drier continental mountain ranges results in part from the nocturnal cooling of leaves below air temperature to reach dew points. Frost or dewfall is common on leaf surfaces in the early morning in the alpine. Just as evapotranspiration removes ~40 kJ/mol of heat from leaves, dewfall adds this same heat into leaves. Frost formation adds 6 kJ/mol. In addition to affecting leaf energy balance, condensation can have a significant effect on processes such as photosynthesis, given that carbon dioxide needed for photosynthesis diffuses 10,000 times slower through water than air. Hence, hydrophobicity, trichomes, and other leaf surface traits are important attributes of many alpine leaves, as revealed by repulsion of water drops experimentally suspended over alpine leaves (Smith et al. 1997).

The ratio of runoff: precipitation is typically high in alpine areas, owing to intense pulses of water input (snowmelt, often coincident with spring rains) draining topography (ie, mountain tops), low abundance of evapotranspiring leaf area per unit land area, and to soils that are frequently coarse, shallow, and well-drained. Aside from snow banks as a reservoir for water, alpine areas have relatively low soil water storage, and there has been relatively little exploration of differences in rooting as a factor influencing water relationships of alpine plants.

## Convection

The aerodynamic shapes of alpine plants, or their tendency to seek microsite shelter from wind, indicate that wind can be a significant aspect of the climate (Fig. 2). There is a greater incidence of extreme exposure to winds due to the landscape prominence of mountains and reduced standing crop and plant canopy would otherwise impose frictional drag on airflow and protect much of the vegetation within the canopy from wind speeds. However, some alpine areas can also be sheltered from wind and wind does not increase with elevation *per se* (Körner 2003).

Convective heat exchange is a function of wind speed and the temperature difference between plant surfaces and air temperature. Alpine plant forms strongly and often strikingly affect the wind or convection actually experienced by leaves, particularly for cushion plants or krummholz (Fig. 2). Convective heat exchange is inversely related to the aerodynamic boundary layer of the leaf, plant, plant canopy, and, in cases, the tight connection of many alpine plant species to protective microsite features such as rocks (“crannies”). The boundary layer can be considered a buffer zone of air in which air conditions are mutually affected by the leaf or other surface and the bulk air of the surrounding environment.

An important conceptual consideration for convection is that the energy balance of plants is affected by boundary layers that are nested into other boundary layers. For example, narrow cylindrical leaves (e.g., conifers) that have a steep orientation may have little boundary layer by themselves, but the dense packing of such needles into shoots, then shoots onto the whole plant, and plants into krummholz mats or tree islands results in considerable boundary layer resistance – each level of organization has its own contribution to the overall resistance. Leaves of alpine cushion plants or krummholz may be small and by themselves have small boundary layers, but they are affected by large boundary layer resistances of the crown and the soil surface that the foliage is so close to.

Wind can also cause mechanical damage to plants, causing loss of shoots or leaves or reproductive parts. During winter, snow that remains on the ground for extended periods can have crystal metamorphosis that generates granules with sharp edges, which abrade exposed plant tissue when blown across the landscape at high speeds (Smith et al. 2003). The dense clustering of leaves or stems into the individual plant or canopy (e.g., tree islands) can help optimize other aspects of wind, by reducing wind enough to trap snow in ways that protect the buried plant from extreme temperatures or snow abrasion.

## Intra-alpine Site Variability

Most efforts to understand alpine plant community patterns cannot escape incorporating the variability in plant communities and populations that correspond to the mosaic of microclimate and soil conditions associated with these landscape factors.

Tree islands in the subalpine transition from forest to treeline represent a major way that plants alter the microclimate for themselves and surrounding



landscape and vegetation. As shown below, the degree of alteration is not trivial and the resulting ecological feedbacks have important outcomes for alpine ecology. Trees affect the microclimate around themselves in several ways, by providing shade from sunlight, which affects microclimate and photosynthesis of the shorter neighboring vegetation (Ball 1994; Germino and Smith 2000). As an example of microclimate effects, in mid to upper latitudes, it is common to see triangular areas to the north (northern hemisphere) or south (southern hemisphere) that retain frost or snow when the rest of the landscape has melted, often for days or weeks.

At night, tall-statured trees increase the amount of long-wave radiation incident upon the plants and soil surrounding them, which increases surface temperatures, particularly minimum daily temperatures at night. Trees also affect wind flow, providing a bluff body effect that can cause appreciable snow drifting in their lee that inspired the term “snow glades.” The snow drifts from a single tree island can extend nearly 100 m in length across subalpine or alpine-like meadows and can reach depths of 10 m or so. Furthermore, snowdrifts can endure months following the melt out of the surrounding landscape. For example, in Western North America, most snowmelt in the portion of alpine areas having relatively abundant plant cover occurs by late May through June, occasionally into July, whereas large snowdrifts even within the lower portions of the timberline zone may persist into or through August, and in some years may even have some of the snow bank present when snowpack begins to accumulate between September to November. Snow cover insulates plants from prevailing climate, provides an important source of water during the growing season, and stimulates microbial activity such as the pathogenic *Herpotrichia* that smothers vegetation in late-lying snow banks with a coating that resembles tar.

Topographic features such as hills or cliffs can have some of the same solar shading, long-wave enrichment, and bluff body and snow drifting effects as tree islands but at a greater range of scales. Additionally, the drainage effects of topography on both water and cold air can have very large effects on the microclimate patterns of entire alpine landscapes. For demonstration, a useful exercise is to consider what the coldest location within an alpine landscape might be, given the microclimate and energy balance considerations described above. It would likely be a microsite that has high sky exposure (i.e., is distant from tree islands or large cliffs that occlude the view of sky) that is also in a closed (i.e., no outlet draining) basin with slopes that are steep enough to drain cold air to the microsite but not so steep as to block the sky view. In the absence of cloud cover that would moderate the long-wave radiation balance or strong winds and mixing of air, air and surface temperatures near ground could easily be 10–20 °C cooler than microsites with opposite conditions (e.g., on a ridge with sky-occluding features). This cold spot would also receive high sunlight during days and probably would have relatively warm midday conditions, and snow drifting and soil moisture would also differ. These drainage and sky exposure considerations can help explain phenomena such as inverted treelines, in which slopes above alpine-like (or subalpine) meadows are forested. As demonstrated below, the physiology of alpine plants is very much linked to the combinations of these different microclimate factors, perhaps more so than to any single factor itself.



## Physiological Responses

### Generalized Stress Response and Growth Strategies

The environmental challenge for plants in an alpine environment is to rapidly utilize available snowmelt for growth during a short and/or cool growing season. Research has asked how plant uptake of carbon and soil resources and growth processes are impacted or adapted to short and cold alpine growth conditions. The short growth season and prospects for alternation of favorable and less favorable growing seasons (e.g., very short growing season or one with extended water deficit) have led to three key aspects of plant adaptation to the alpine.

First, anticipatory development is common in many herbs, in which buds are preformed in the fall prior to winter dormancy, enabling rapid development upon spring or summer snowmelt. Unlike plants from temperate environments, vegetative and reproductive growths tend to be synchronous, although preformation can occur in either type of meristem. The prevalence of bud preformation should have the effect of decoupling growth of alpine herbs from the weather prevailing in any given year. Reliance on bud preformation limits the ability of alpine plants to adjust their development to current conditions, however.

Second, rapid shoot emergence is subsidized by carbohydrates and nitrogen acquired in previous growing seasons and stored in large root systems or below-ground storage organs. With these advantages, species such as marsh marigold (*Caltha leptosepala*) and several buttercups (*Ranunculus* sp.) are well known to begin development in the relatively low-light and near-freezing conditions under snow, frequently developing through snow and completing much of their life history nearly in contact with the snow retreating from around them (Billings and Mooney 1968).

Third, alpine herbs can exhibit relatively high rates of resource uptake when conditions are optimal, and they are furthermore uniquely able to sustain uptake and growth under cool conditions that characterize the growing season. Interestingly, the highest elevation herbs tend to have a leafy and not cushion physiognomy and slow growth associated with it, indicating selection upon them for rapid capitalization of growth opportunity at the highest reaches of plant life.

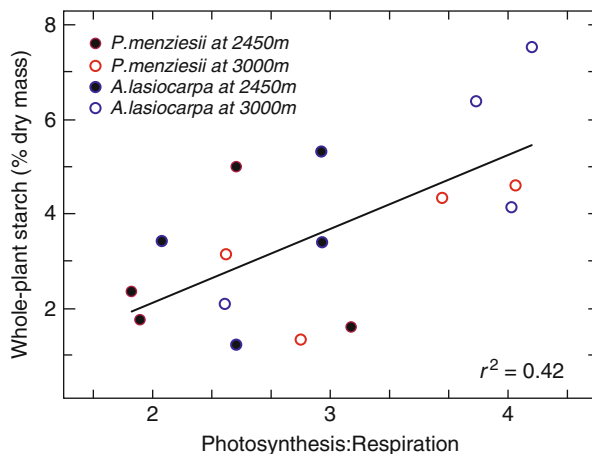
### Carbon and Nitrogen Storage

Storage of carbon and nitrogen is part of a generalized strategy for stress resistance, and it is a prevalent theme in alpine and treeline ecology. Carbohydrates generally accumulate during the growing season and are then translocated to stems, roots, or other storage locations for overwintering, especially in herbaceous perennials but even also in evergreens. In some situations, there is little ambiguity about how and why carbohydrate or other nutrient translocation and storage like this occur. For example, storage mechanisms are intrinsic to the life-history strategy of geophytes (herbs that have underground bulbs, tubers, or rhizomes), in the alpine and a wide

range of other habitats. The storage pools are quickly depleted upon release from winter dormancy, when allocation to rapidly expanding new tissue and to respiration exceeds photosynthetic gain. The importance of this translocation-storage mechanism is evident for snowbed herbs like *Caltha leptosepala*, which do not exhibit appreciable storage formation in microsites where growing season length is truncated by deep snow banks but do accumulate sugar and starch where growing seasons are longer (reviewed in Billings and Mooney 1968). Seasonal redistribution of carbon from shoots to roots has also been observed in evergreen plants, although their foliar carbon content can also increase as plants acclimate to the onset of drought and particularly winter cold. Carbohydrates also have direct roles in stress responses; simple and complex sugars (e.g., fructans and raffinose) correspond well with acclimation to chilling in leaves of *Dactylis glomerata* and other alpine herbs (Monson et al. 2006). Sugar and other osmotic compounds decrease the temperature required for ice formation (antifreeze) and decrease the tendency for water to be lost from cells, thereby protecting against desiccation.

There has been considerable emphasis on evaluating nutrient pool sizes, especially of carbohydrates in their starch or sugar form, as a means for identifying processes limiting productivity of alpine or treeline species. Many studies have revealed increases in carbohydrates or nitrogen content per unit leaf area with increasing elevation into alpine or treeline zones, in herbs or trees (Körner 2003). Many of these studies found greater concentrations of carbohydrates at higher elevations, leading to the suggestion that chilling at treeline does not limit carbon uptake in trees, but rather their ability to use carbon for growth processes (Körner 1998). These studies relied on estimates of the percent of dry mass that was available carbohydrate, i.e., “nonstructural” or “mobile” sugars or starch in leaves or stems and occasionally roots. Although it is convenient (and common) to view carbohydrate concentrations as if they are merely a passive outcome of carbon sources (photosynthesis) and sinks (growth, respiration, and losses through root exudation or tissue shedding; Ryan 2011), active regulation of carbohydrate pools is reflected in variation in concentrations of carbohydrates among alpine and subalpine plants (Bansal and Germino 2008; Wiley and Helliker 2012). The concurrence of growth reductions and elevated stores of carbon or nitrogen may result from an inability to use the resources for growth, but growth may also be reduced to ensure the formation of the reserves. Strategies like this might be expected for long-lived perennials in which rare years of very poor net carbon flux might select for reserve formation abilities, and active reserve formation could certainly be stimulated by the same factors that directly affect growth and all carbon source and sink processes.

Active storage creates reserves at the expense of growth or other processes, whereas passive storage is accumulation with no apparent cost to the plant. Sugars of many alpine plants and specialized molecules such as cyclic polyols (cyclitols) accumulate following shoot expansion, while photosynthesis is at seasonal maximum, following a pattern indicative of storage formation (Monson et al. 2006). In alpine herbs of the Caryophyllaceae, cyclitols confer protection against late-season drought stress (and probably also nighttime freezing) and their concentrations



**Fig. 5** Relationship of nonstructural carbohydrates, specifically starch, and whole-shoot photosynthesis and respiration in the short-term history of tree seedlings planted across the full breadth of a treeline ecotone (Regraphed data from Bansal and Germino (2008)). Sugars were not related to photosynthesis: respiration, and starch is generally considered a storage form of carbohydrate and allocation to growth was not an appreciable aspect of the carbon balance during the time increments evaluated. With more than half of the variability in % starch unexplained by the passive balance of photosynthesis-respiration, it appears likely that active regulation such as reserve formation may be occurring. Species were *Pseudotsuga menziesii*, which does not normally occur at 3,000 m, and *Abies lasiocarpa*, which normally spans the full alpine gradient. The values shown are the means from different sampling dates spanning the entire snow-free growing season

correspond to successful establishment, but they do not appear significant for reserve storage. In contrast, cyclitols in *Artemisia scopulorum* appear to be significant aspects of reserves for future growth. “Luxury” accumulation under high resource availability can also occur with no benefit, as indicated by reductions in future nitrogen uptake alpine bistort (*Bistorta bistortoides*) following fertilization and no net growth increases (Monson et al. 2006).

For the few cases where robust accounting for carbon sources, sinks, and pool sizes have been accomplished across elevation gradients to treeline, only a portion of the carbohydrate pool appeared to be under source: sink control and the rest either under some blend of active regulation (Fig. 5). These conceptual advances for carbohydrates at treeline have been applied more recently to other plant species and habitats, most notably the role of carbon starvation in recent drought-induced tree mortality and forest dieback (McDowell 2011).

Several additional complications in the evaluation of carbohydrate pool sizes are notable for alpine studies. Recent research has placed an emphasis on starch and sugar in leaves, stems, and roots, but there is considerable diversity in the types of molecules involved in storage, and in the organs where storage occurs, and the phenological pattern of storage formation and depletion (Körner 2003). For example, storage of carbon as lipids in leaves and stems is also known in evergreen shrubs such as *Ledum groenlandicum* (Billings and Mooney 1968). Moreover, the

distinction between carbon and nitrogen that is in a pool deemed to be “nonstructural” is clouded by recycling of hemicelluloses from cell walls or abundant and carbon- and nitrogen-rich enzymes like RUBISCO (the enzyme catalyzing CO<sub>2</sub> assimilation, ribulose-1,5-bisphosphate carboxylase-oxygenase). Recycling makes the constituent molecules available for other uses. How growth relates to carbon and nitrogen pool sizes is not known, nor is it often clear when a plant has truly become depleted or limited by an internal resource pool. Whereas most studies in alpine ecology have tended to have a binary view of structures as either substrate sources or sinks, or resource pools as being in either a structural or nonstructural form, or storage as being either actively or passively controlled, it is more likely that gradations exist for each. Considerable research advances for alpine plant ecology may lie with a framework revised along these lines, and the outcomes will improve use of cost-benefit analyses or even mass balance approaches to assess plant responses to the alpine environment (Monson et al. 2006).

---

## **How Do Specific Climate Stresses Occur, and What Are the Physiological Responses?**

### **Temperature Stress**

Special cold-stratification requirements for germination or ability to germinate at particularly low temperatures are generally not evident for alpine plants. Instead, the abilities to assimilate carbon and grow at low temperatures and to have opportunistic spurts of rapid growth and development during optimal conditions appear to be generalizable and unique attributes of alpine plants (Billings and Mooney 1968).

With decreasing temperature, the following cascade of physiological responses occurs in plants or results in adaptive responses as in many alpine species. With chilling, enzymatic activity is decreased in accordance with kinetics, such as the kinases and carboxylases in respiration and photosynthesis. With deeper freezing, vascular processes cease and cytosolic immobilization and deactivation, as well as possible damage, occur. Freezing is a profound form of desiccation in plants, and the impacts of freezing can involve loss of hydraulic conductivity due to the formation of embolisms and cavitation in xylem. The effect of chilling is usually to cause the plant to perform outside of its optimal temperature range for growth, generally resulting in less growth. Freezing and especially freeze-thaw cycles contribute to loss of growth opportunities, persistent embolisms that can result in loss of conductivity, and rupture or damage of cell walls that either diminish future growth potential under optimal conditions or can lead to death. Frost heaving and needle ice (protrusion of ice crystals) in soil can damage roots. At midsummer, warm temperatures can exacerbate the leaf-to-air vapor deficit, stimulating desiccation stress.

Acclimation and adaptation can expand the breadth of the temperature optimum of photosynthesis, respiration, stem elongation, etc., resulting in less impacts of

chilling to growth. Many alpine plants, specifically herbs, express a high capacity to sustain resource uptake and growth at very low temperatures. Snowbed herbs, known as geophytes, are frequently observed sprouting and beginning stem elongation under snowpack, where temperatures are 0 °C (or less, due to nighttime cooling of the surface). Snowbed herbs can achieve maximal photosynthesis within minutes of having been thoroughly frozen (Germino and Smith 2000). In the subalpine, seeds of conifers such as *Picea engelmannii* and *Abies lasiocarpa* commonly germinate and have cm of growth while imbedded in snow banks that persist into summer growing season, clearly utilizing their carbon reserves and growing at temperatures near 0 °C though seedling establishment may not result.

Freezing, specifically the formation of ice, occurs at lower temperatures in some alpine plants as a result of freezing point depression or supercooling. By withdrawing water or adding osmotically active molecules derived of carbohydrates or other ions, the temperature required to cause the apoplast or symplast to freeze can be decreased considerably, by a few to 20 °C or more. Plants that supercool withdraw nucleating agents, thereby inhibiting ice formation to very low temperatures. Membrane flexibility can also be adjusted by the fraction of unsaturated C-C bonds in the lipids comprising the plasma membrane, allowing the plant to avoid disruption and leakage across the plasma membrane or cell wall. Raffinose and other carbohydrates can adhere to and stabilize membranes leaves undergoing freeze-thaw cycles. Rapid dissolution of emboli in xylem elements that become cavitated during freeze-thaw cycles is also a key adaptation.

The high diurnal fluctuation in temperatures of the alpine are associated with low nighttime minimum temperatures. Nighttime frosts affect flowering of alpine and subalpine herbs (Inouye 2008). Experimental enrichment of long-wave radiation and corresponding increases in nighttime temperature have generated significant changes in alpine plant communities and in flowering and species shifts (Harte and Shaw 1995).

## CO<sub>2</sub> Availability and Photosynthetic Assimilation

A common question is whether the reduced concentration of CO<sub>2</sub> per unit air volume that accompanies reduced atmospheric pressure at high-elevation alpine areas causes reduced photosynthesis. There is consensus that such a limitation is small or unlikely to occur (Körner 2003). Whereas the diaphragm-lung inhalation in humans is sensitive to the abundance of oxygen and therefore elevation, the passive diffusion of CO<sub>2</sub> into leaves for photosynthesis is a function of how much less the concentration of CO<sub>2</sub> is inside the leaves compared to the surrounding air and the resistance (= 1/conductance) of CO<sub>2</sub> across this gradient through stomata. Specifically, a diffusion or Fick's law model predicts the flux density of net photosynthesis as the product of the concentration gradient of CO<sub>2</sub> between air and leaves ( $C_a$  and  $C_i$  in molar concentration units or  $P_a$  and  $P_i$  in partial pressure units) and the stomatal conductance to CO<sub>2</sub>. Conductance to CO<sub>2</sub> diffusion into stomata and leaves is predicted to be enhanced by a higher rate of molecular

diffusion under reduced atmospheric pressure. If alpine plants “held”  $P_i$  similar to concentrations of lowland settings while the lower  $P_a$  of the alpine prevails, then alpine plants would be expected to have intrinsically less photosynthesis despite considerable offsets from the higher molecular diffusivity (as in the simulations of Smith and Johnson 2009). There is no reason to expect the total gas pressure in leaves to be uncoupled from ambient based on physical principles. However, changes in  $P_i/P_a$  have been detected along elevation gradients to the alpine, and they have tended to be a decrease in the ratio (Körner 2003, e.g., Cordell et al. 1998).

An important source of evidence for changes in  $P_i/P_a$  across elevation gradients into alpine and subalpine areas has been the ratio of  $^{13}\text{C}:^{12}\text{C}$  of plants, reported relative to Pee Dee Belemnite as  $\delta^{13}\text{C}$ , which relates isotopes to  $P_i/P_a$  as follows:

$$\delta^{13}\text{C plant} = \delta^{13}\text{C air} - (4.4 + 22.2 P_i/P_a)$$

where 4.4 and 22.2 are fractionations for diffusion in air and through stomata.  $P_i$  can be reduced relative to  $P_a$  (increasing the gradient) by either increasing the biochemical demand for  $\text{CO}_2$  inside the leaf or restricting the supply of  $\text{CO}_2$  into the leaves with partial stomatal closure. For example, the biochemical demand for carbon can increase if leaves have a greater concentration of nitrogen, which usually is associated with more protein enzymes for carboxylation and photosynthetic fixation of  $\text{CO}_2$ . Alternatively, supply of  $\text{CO}_2$  can be reduced if stomata partially close under incipient water stress.

The universal pattern across the elevation gradients appears to be less discrimination against the heavy isotope at higher elevations, which, in several cases, corresponded with greater leaf nitrogen and stomatal density, and a reduced specific leaf area ( $\text{cm}^2/\text{g}$ ) that encompasses more mesophyll layers per unit leaf area (Cordell et al. 1998; Körner 2003). These reductions in  $P_i/P_a$  (increase in  $P_a - P_i$ ) at higher elevation have been observed into the subalpine for trees like ohia in Hawaii and conifers of Western North America and for a range of alpine herbs globally. Comparisons of the response of photosynthesis to step increases in  $C_i$  (known as the A- $C_i$  response) are generally much steeper for alpine plants at low  $C_i$  values, which indicates both high carboxylation efficiency and thus a strong biochemical demand for  $\text{CO}_2$  (Körner 2003). Taken together, these findings suggest that the resource allocation, form, and function of leaves in the alpine favors reduced  $P_i/P_a$  by a relatively strong demand for  $\text{CO}_2$  in spite of increasing capacity for diffusive supply of  $\text{CO}_2$  in stomatal conductance at greater elevations. Leaf thickness and the dense anatomical arrangement of mesophyll typical of alpine herbs might also exacerbate the gradient from  $P_i$  to chloroplasts. These considerations suggest no reductions in the  $\text{CO}_2$  gradient from the low-atmospheric-pressure air of the alpine into the site of photosynthesis, compared to lowland conditions.

A second perspective on the question of carbon substrate limitation to photosynthesis in the alpine is based on a Michaelis-Menten model for photosynthesis. This approach recognizes that the primary site of carbon assimilation can also assimilate oxygen, effectively resulting in competition for the binding site in the

RUBISCO reaction. Oxygen is expected to follow similar diffusion constraints as CO<sub>2</sub>. The oxygenase reaction, known as photorespiration, is traditionally viewed as detracting from plant productivity (but, the specter of photoinhibition indicates potential adaptive roles for photorespiration, discussed below). With respect to low pressure effects on photosynthesis, the oxygenation reaction in photorespiration is favored at warmer temperatures, whereas carboxylation is favored at cooler temperatures. Simulations suggest that the reduction of photorespiration in the alpine contributes to the small or absent reduction in photosynthesis relative to low elevations (Terashima et al. 1995).

If CO<sub>2</sub> abundance were limiting in the alpine, then photosynthetic pathways that concentrate CO<sub>2</sub> for carboxylation processes among the higher plant taxa, i.e., C4 or CAM photosynthesis, might be more prevalent. These pathways evolved in response to photorespiration when CO<sub>2</sub> was less abundant in the geologic past. *Sedum* and a number of other succulent plants capable of CAM photosynthesis are occasionally found in alpine areas, but their expression of CAM photosynthesis is rare (Körner 2003). Generally, alpine areas are dominated by C3 photosynthesis, which, aside from the low-temperature sensitivity of additional enzymes in C4 or CAM photosynthesis, may be further evidence that strong CO<sub>2</sub> limitations do not occur in the alpine.

Finally, respiration is a major aspect of net photosynthesis and could influence carbon balance across elevation gradients (net photosynthesis = gross photosynthesis – dark respiration – photorespiration). Respiration appears to have a narrower temperature optimum than photosynthesis, which is expected to result in greater reductions in respiration than photosynthesis in cool alpine conditions. In fully developed leaves, dark respiration frequently decreases more with elevation than does photosynthesis, at least for single species spanning treeline and alpine environments (Körner 2003; e.g., Bansal and Germino 2008). There is less evidence that respiration differs for alpine compared to lowland plants, in general. These considerations further suggest that net photosynthesis likely would not decrease as much as any decreases in gross photosynthesis at higher elevations and alpine conditions. The relationship of gross photosynthesis, photorespiration, and dark respiration has yet to be comprehensively evaluated with field measurements across elevations in alpine and treeline zones.

## Radiation Stress

Radiation has a number of positive and negative effects that can appear paradoxical. The positive long-wave radiation balance warms leaves in the cool alpine, but the more common negative long-wave balance cools minimum temperatures in an already cool environment. Visible shortwave (solar) radiation drives photosynthesis and warms leaves, but can also cause photochemical problems (Ball 1994). Ultra-violet radiation causes photochemical damage by causing somatic mutations of DNA, but there is no evidence that alpine plants are damaged more by UV than lowland plants. The minimal impact of UV may be due to effective screening of UV



at the epidermis, by solutes and by pigments such as the red anthocyanin that is common in alpine and most sunny habitats.

The basic response of photosynthesis to sunlight in alpine plants generally follows attributes of leaves adapted to sunny environments, which includes the anatomical and biochemical traits described above. The photosynthetic response to step changes in sunlight usually reveals that photosynthesis saturates at relatively high sunlight levels in alpine plants.

High sunlight levels can also have negative effects on photosynthesis, leading to a condition known as photoinhibition (Ball 1994). Photoinhibition refers to the light-dependent reductions in photosynthesis that tend to occur when plants are subject to other stresses, particularly low temperatures. When low temperatures cause a reduction in enzymatic activity, the processes in which recently assimilated carbohydrates are reduced (i.e., the dark reactions, Calvin cycle) and their export from chloroplasts for use in growth become slower. The enzymatic dark reactions are the primary “consumers” of the reducing equivalents (ATP, NADH) produced in the light reactions, the so-called Z-scheme of chlorophyll reaction centers and transmembrane proteins (thylakoid membranes in chloroplasts) that produce ATP and NADH from the energy derived from sunlight. The outcome is a net imbalance of supply and usage of sunlight excitation energy supply in photosynthetic carbon reduction (supply of ATP and NADH). Under these conditions, the electron transport chain of proteins becomes so reduced (saturated with electrons) that it can no longer accept electrons from photosystem II, where the sunlight energy harnessed from the chlorophyll antennae is used to split water and elevate the energy level of the resulting electrons such that they are able to flow through the electron transport toward the end products ATP and NADH. The excess excitation energy in the photosystem II complex can be dissipated through reradiation as chlorophyll fluorescence, which is directly measurable but is not considered a significant adaptive mechanism. The excess excitation energy can also be safely dissipated back into the chlorophyll antennae complex, provided that compounds on the thylakoid membranes known as xanthophylls are in a de-epoxidated molecular configuration. This safe dissipation is a reversible condition known as non-photochemical quenching (photochemical quenching leads to ATP and NADH production). When sunlight energy harnessed by leaves cannot be adequately dissipated by chilling-inhibited dark reductions or by non-photochemical quenching, the excess excitation can lead to photooxidative damage and ultimately nonreversible (or slowly reversible) damage and reduced photosynthesis and growth.

The photochemical challenge for alpine plant life has been the subject of a number of studies that have served to illustrate the ecological relevance of molecular processes in the chloroplast or mesophyll. The reversible downregulation of photosynthesis is evident for many alpine species in diurnal photosynthesis or chlorophyll fluorescence patterns. Non-photochemical quenching via fluorescence or concentrations of xanthophylls in sunny microsites do not show a clear increase with elevation, suggesting that although xanthophyll configuration may more consistently allow non-photochemical dissipation of sunlight energy, alpine plants



may not necessarily have greater capacity for this protective means (Germino and Smith 2000). One of the most important means for avoiding photoinhibition in alpine species may simply be the ability for photosynthesis to continue at low temperatures, which sustains the consumption of photochemical energy and minimizes excess absorbed excitation energy. There is greater evidence for alpine plants to have more antioxidants as a means for mitigating damage associated with nonreversible impacts of photoinhibition (Germino and Smith 2000). Reports of state transitions (spatial changes between reaction centers that redistribute and balance excitation energy), photorespiratory capacity, the Mehler reaction (electrons from water ultimately reduce peroxide), and other processes have been evaluated in alpine plants, creating an intriguing body of research that continues to illustrate the ecological role of sunlight in the alpine. Low-temperature photoinhibition illustrates the manner in which multiple stresses can interact in alpine environments, in ways that are not simply additive.

### **Desiccation Stress**

Evapotranspiration can be modeled using a diffusion or Fick's law analog to that presented above for photosynthesis, as the product of the leaf-to-air vapor concentration (or partial pressure) gradient and of the conductance of the leaf-to-water vapor transport into the bulk air. The high rate of molecular diffusion in the low atmospheric pressure of the alpine causes high leaf conductance to water vapor (just as it enhances stomatal conductance to CO<sub>2</sub>). When combined with high leaf-to-air temperature and thus vapor gradients for cushion plants, krummholz, or large-leaf herb, transpiration potential is high (Smith and Johnson 2009). The osmotic pressure of alpine leaves tends to be less (having more solutes) than lower-elevation environments, which may suggest adaptation to drier conditions but could also be linked to active regulation for carbohydrate reserve formation or freezing avoidance. A fundamental difference between semiarid basins and alpine areas, however, is the relatively greater water supply typical of alpine areas.

---

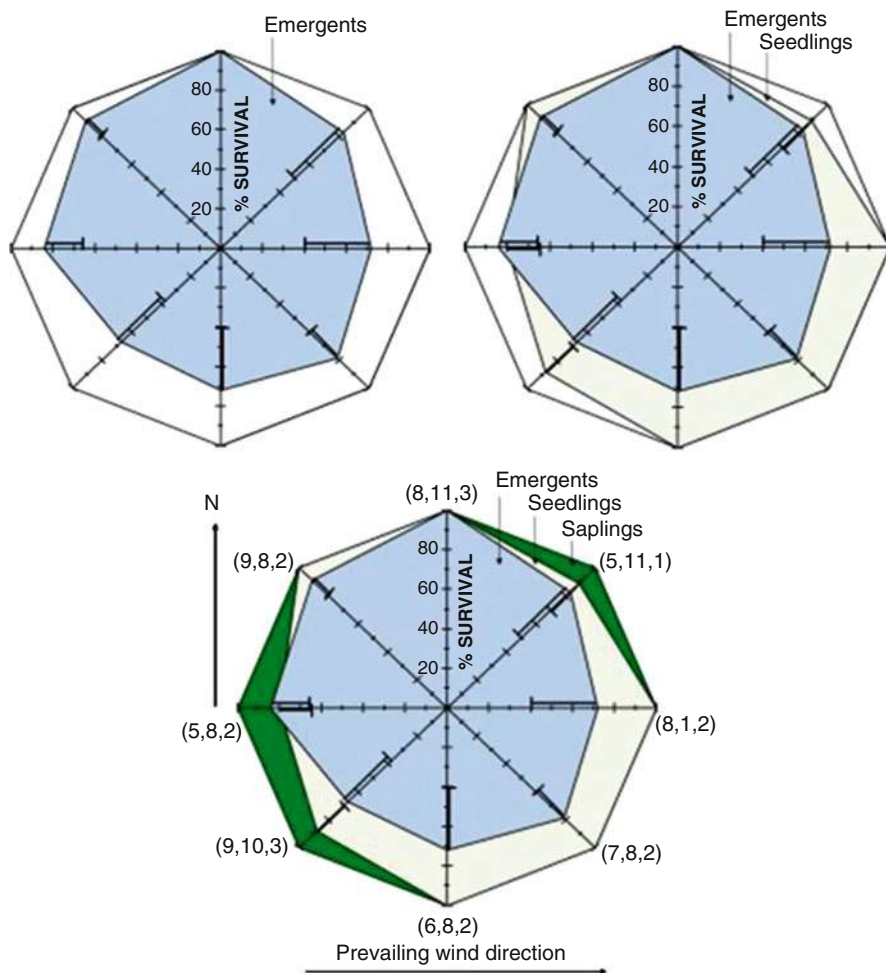
### **Linking Microsite, Plant Form, and Physiology in Alpine Plants**

Two research themes are presented here to help synthesize the major biophysical constraints, the physiological responses and the ecological outcomes on the landscape. Firstly, the patterns and underlying processes associated with tree establishment in subalpine and alpine environments are informative because the limitations to tree success in this ecological zone can illustrate factors that true alpine vegetation have overcome. Second, whereas trees by definition are outside their range limit in our zone of interest, an opposite scenario is herbs that appear adapted to one of the coldest and most photoinhibitory conditions in the alpine – the snowbed environment.

## Patterns of Tree Establishment

Tree establishment above forest elevation limits requires tree seed dispersal beyond the forest, germination, successful seedling establishment, and maturation – but these demographic steps do not appear to contribute equivalently to tree abundances and population dynamics within the timberline and treeline zones. This ecological boundary appears to be exhibiting a considerable amount of change, globally. The change is not evident so much in adult trees that have been present for centuries of climate variability. Instead, a considerable increase in new establishments has coincided with regional warming trends in the timberline and treelines that are expansive (not the sharp ecotones described above). In many regions (Western North America, Ecuador, Hawaii, Pyrenees, Australia; Ball 1994), the formation of tree islands is suggestive that new tree establishments have been clustered or have occurred where trees had established previously. There has been little or no evidence that the effect is due to greater seed deposition around trees or preferential germination near trees. Instead, seedling survival just after germination, when very high mortality can occur for long-lived species (e.g., >90 or even 99 %), indicates that seedlings are culled from microsites away from mature trees after germination, as shown in the Rocky Mountains, USA. Photosynthesis and chlorophyll fluorescence were also greater for seedlings near trees compared to away from trees. As described above, trees have an important microclimate effect in providing shade, warmer nights, and snow cover alteration. The patterns of natural establishment or survival and photosynthesis of experimental seedlings near trees (Fig. 6) suggest daytime shade and warmer nights best characterize safe sites for seedling establishment. Somewhat like trees, a number of reports show that herb cover also positively affects tree seedlings.

Experimental shading of newly germinated *Picea* and especially *Abies* seedlings in alpine-like meadows near treeline led to substantial increases in photosynthesis, which is corroborated for *Polylepis* in the Northern Andes, *Eucalyptus* in Australia (Ball 1994), and other treelines. Similarly, experimental increases in minimum nighttime temperatures for *Picea* and *Abies* also increased photosynthesis, but the greatest increases occurred with a combination of warming and shading. Notably, passive open-sided chambers were used to increase minimum temperatures by about 2 °C (via an increase in long-wave radiation from the sky of 50 W/m<sup>2</sup>), which had little effect on daytime microclimate but decreased the frequency of nights during the growing season that had plant surface temperatures below 0 °C by 35 % at this treeline location. Surface frosts could occur nearly every week of the growing season, in spite of much warmer temperatures registered by meteorological stations, and the coincidence of this “summer nighttime-frost line” with the location of treeline is compelling. Furthermore, the increase in long-wave radiation and degree of warming simulated are similar to predicted greenhouse warming effects. Nighttime frost and bright sunlight are also linked in the clear-sky conditions that occur frequently during summertime high-pressure systems. Nighttime radiation frosts are typically followed by days with very bright sunlight. A number of studies provide further evidence to suggest the combination of cold nights and bright



**Fig. 6** Tree seedling survivorship around tree islands in a treeline ecotone of Wyoming, USA, of *Picea engelmannii* and *Abies lasiocarpa*. Each axis is a cardinal direction, with the origin being the canopy edge of a tree island. Increments along each axis are annual survival. The outer line shows 100 % survival. Blue shows survivorship of newly germinated (emergent) seedlings in their first summer after germinating; gray shows seedlings are plants in the 2nd to 5th year of growth, and saplings are plants older than 5 years but less than 20 cm tall. Numbers in parentheses are sample size for emergents, seedlings, and saplings, respectively (Regraphed from data reviewed in Germino and Smith (2000))

sunlight generates symptoms of low-temperature photoinhibition, and the correlation of this proposed mechanism to establishment implies a role for carbon limitation to tree establishment.

Greater carbohydrate concentrations in trees near treeline has prompted the proposal that the treeline environment poses more challenges to carbon use rather than uptake (Körner 1998), but young tree seedlings are faced with needing to gain

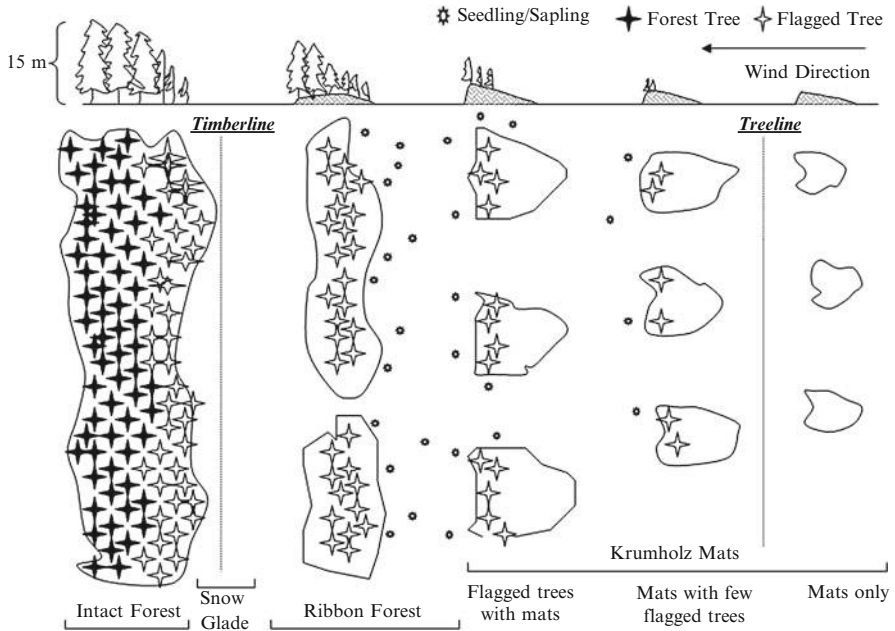
manifold increases in carbon from year to year in order to have normal growth and development and could therefore conceivably have greater carbon limitation (Bansal and Germino 2008). Depletion of carbohydrates to near 0 % of dry mass has not been observed in tree seedlings at treeline, but positive associations among survivorship, photosynthesis, and carbohydrate concentrations have been established, suggesting tenability of the hypothesis but a need for evidence other than carbohydrate concentrations. Alternative hypotheses for tree seedling affinities for tree islands, such as lack of availability of the appropriate ectomycorrhizae for tree seedlings in herbaceous meadows, or altered soil properties near trees, were explored by reciprocal soil transplanting but yielded no compelling support.

### **Ecological Significance of Microclimate Amelioration and Facilitative Interactions**

Many decades of literature has asserted competitive interactions between trees and meadow vegetation. In recent years, observations such as for tree seedlings in timberlines and other species-species affinities suggest positive interactions, or nurse-plant effects that are referred to as facilitation, as prevailing in the alpine and subalpine environment. The stress gradient hypothesis predicts that positive interactions among plants increase in likelihood for plant communities that endure greater physical stress. There is now widespread support for this hypothesis based on examples of one demographic stage favoring growth of one another, or one species favoring another, in alpine (Callaway et al. 2002) or treeline ecotones (Germino and Smith 2000; Smith et al. 2003).

A large, more significant question is how the balance of positive and negative interactions plays out over time, as recipients of facilitation increase in size and possibly exert competitive pressure on their facilitators, for example. Once a tree seedling has survived one to a few years, its chances of survival thereafter increase dramatically and its reliance upon nurse effects from neighbors is relaxed. As seedlings mature, they add foliage that is configured in ways that enable the individual plant to alter the way in which its climate is coupled to the surrounding environment. Seedlings have little to know boundary layer of their own, and are dependent on the shelter provided by trees or overtopping herbs. As seedlings grow above this surrounding protection, their accumulation of needles that are generally upright in orientation and clustered together enables an optimization of sunlight and temperature regime, in which needle arrangements optimize sunlight intensity while the clustering of needles reduces wind speeds and convective cooling during days, decreasing the potential for low-temperature photoinhibition to occur.

A model for forest development in alpine-treeline ecotones suggests that the mutual microclimate amelioration among timber-like trees provides a landscape with many more safe sites for establishment (Fig. 7; Smith et al. 2003). This tree-environment interaction is a distinct positive feedback between species and their environment and epitomizes how positive interactions among plants can have broader ecological significance, such as to land cover boundaries. The scheme in



**Fig. 7** Gradient in tree abundance, form, and corresponding seedling establishment from forest to treeline (left to right), in this case for an east-facing slope with conifers in the Rocky Mountains, USA. By extension, this schema could also show forest development over time for a fixed location (right to left; from Smith et al. 2003, with permissions)

Fig. 7 illustrates that greater seedling establishment occurs in the vicinity of flagged trees that are large and dense enough to ameliorate the nighttime frost and bright sunlight on the ground around themselves. The “snow glade” on the left results from exceedingly deep snow drifts. From right to left, the individual conifer, as krummholz, ameliorates its own microclimate, but wind shear and damaging winter snow abrasion strip leaf cuticles, and leaf desiccation and tissue loss inhibit any upward growth. The krummholz (Fig. 2) form results from a wind-pruning mechanism whereby sharp snow crystals blown across the surface of the snow (saltation) abrade the cuticle that would otherwise inhibit desiccation. In most springs, red foliage on the top of krummholz indicates where previous growth was not protected by the snow drift that forms in the dense foliage (and thus boundary layer) of the krummholz. As an aside, the factors sculpting the form and thus height growth of trees at treeline indicate that summertime growth processes (e.g., carbohydrate storage) need to be considered in light of other processes regulating plant stature. Regardless of the mechanism affecting height, tree height itself is a major factor affecting the surrounding landscape. In the next row left (Fig. 7), when meristems are able to grow rapidly enough in a good year to get meristems above the saltating snow zone, then “flagged stems” of greater stature occur. Mutual microclimate amelioration among krummholz/flag plants can then occur

as a population or community-level phenomenon (3rd–4th row). In the intact forest, the windwardmost row of trees is strongly wind-affected and inner trees are protected.

Interspecific differences in the resistance to low-temperatures and bright sunlight are insightful for community successional patterns in treeline ecotones. Subalpine fir (*Abies lasiocarpa*), with its broadly displayed needles and little tolerance of frost-induced photoinhibition, is the most restricted to “safe sites” under trees, whereas Engelmann spruce (*Picea engelmannii*) has cylindrical needles that tend to be more upright in orientation and a slightly greater resistance to frost-induced photoinhibition and less dependency on safe sites (Germino and Smith 2000). Five-needled pines such as whitebark pine are considered to have pioneering capabilities, able to establish in herbaceous meadows. Once established, whitebark pine can facilitate the establishment of subalpine fir beneath it, creating a successional pattern. Decades of fire suppression by forest managers are considered to have led to undesirable competitive pressure of subalpine fir on whitebark pine, which is a major concern for grizzly bears and other wildlife that eat its nutritious seeds, and compound dieback that has been occurring with mountain pine beetle and white-pine blister rust impacts (Tomback et al. 2001). The pending loss of species like whitebark pine from mountain ranges is likely to affect the manner in which treeline movements and afforestation occur with respect to climate, because the combined specializations of a pioneering and “infilling” species would seem to expedite any forest advance in the treeline ecotone.

### Leafy Herbs of Snowbeds

The vulnerability of tree seedlings to the interactive effects of frost and bright sunlight of alpine environments is contrasted by a high resistance in snowbed herbs. Two herbs are notable in this regard and offer insight on different avenues for adaptation, using a combination of plant form, physiology, and microsite selection. *Erythronium grandiflorum* and *Caltha leptosepala* both emerge from late-lying snow banks in alpine areas throughout much of the Rocky Mountains, USA. In the Snowy Range of Wyoming USA, these snow banks are usually on east-facing (leeward) slopes with a small basin beneath them that accumulates water and cold air (Germino and Smith 2000, and citations therein for following results).

Although they can occur within short distances (e.g., as short as 1 m distance) from one another around the same snow banks, some profound differences exist between them. *E. grandiflorum* frequently occurs on the drier ridges along the N, W, or S sides of the snow banks, tending to be 2 m in elevation above where *C. leptosepala* occurs, which are wet (often saturated) topographic depressions usually on the east or downhill sides of snow banks. *E. grandiflorum* has two leaves with a steep orientation from horizontal that had an average 2.5 cm width and 12.5 cm length, whereas *C. leptosepala* occurs in bunch-like clusters of individuals where many leaves (often up to ~20/plant) are broad (average of 6.6 cm in width × 8 cm in length) and were oriented more horizontally, and there is a greater incidence of mutual shading among leaves.

Leaf temperatures of *C. leptosepala* are substantially more extreme than for *E. grandiflorum*, with leaf temperatures regularly becoming ~15 °C warmer than air

(measured at 1 m height) during day and 8 °C cooler than air at night, leading to leaf temperatures below 0 °C on 70 % of nights during its growing season. In sharp contrast, leaf temperatures of *E. grandiflorum* deviated only 2–4 °C from air temperature and had temperatures <0 °C on less than 40 % of nights during the growing season. Interestingly, minimum leaf temperatures occurred not before sunrise as is commonly thought for all plant life, but instead occurred after sunrise for *C. leptosepala*. This appeared to be the result of sustained frost formation before sunrise and even in the morning twilight while the sun was below the horizon, followed by very large latent heat losses when thick layers of frost melted and evaporated within minutes of very bright sunlight exposure.

Differences in leaf temperature between the species are attributable nearly equally to differences in microclimate air temperature and their differences in leaf form, as determined by measuring leaves in different orientations and by experimentally altering leaf angles. Compared to the broad and flat leaves of *C. leptosepala*, the slender, upright leaves of *E. grandiflorum* have higher nighttime temperatures because their net long-wave radiation balance is greater (exchanging radiation with surrounding plants and topography instead of sky), their convective heat exchange is greater due to narrower leaves (upright leaves also more efficiently drain away the air that cools next to them), and leaves are elevated into warmer air layers as a result of the steep inclination from horizontal. Eliminating these morphological advantages leads to up to 6 °C cooler leaves.

These snowbed species exhibited a high resistance to frost. No difference in carbon gain occurred following nights with or without frosts in *E. grandiflorum*, but a 35 % decrease in photosynthesis for *C. leptosepala* for several hours in the mornings following its more severe blend of chilling and bright light. Both species exhibited about 8 % reductions in their sunlight-use efficiency from morning to midday, reflected in decreases in quantum yield (mol/mol of CO<sub>2</sub> gained per photon of sunlight absorbed, at low sunlight) and by chlorophyll fluorescence. Such losses in photosynthetic efficiency are not expected to be significant when leaves are saturated with sunlight, and instead the lower post-frost carbon gain of *C. leptosepala* corresponds with the greater mutual shading in its crown and canopy and reports that it can proceed through a growing season without appreciable carbon reserve formation. Nonetheless, chlorophyll fluorescence did not reveal that either species had any unusual ability to utilize non-photochemical pathways of dissipating bright sunlight even while leaves were nearly 0 °C in the morning, and photochemical damage was never evident. Instead, a high capacity to sustain productive use of sunlight energy for photosynthesis, thereby avoiding excess sunlight energy absorption, allows these herbs to occupy one of the severest combinations of environment and life history for chilling and bright sunlight in nature.

The restriction of *C. leptosepala* to the wet but cold depressions may be linked to its hydraulic requirements and may be further enabled by its high leaf elasticity for enduring freeze-thaw events, and ability to withstand inundation and the associated hypoxia, as could be revealed with future research. These types of research efforts may help reveal trade-offs in temperature resistance and water requirements that



can help explain species' niches and landscape patterning of communities (Fig. 1) as they have resulted from recent climate and may change with the onset of novel climates and combinations of temperature and water.

---

## Future Directions

With the specter of global warming, alpine ecosystems are increasingly valued as bellwethers for early and relatively unconfounded impacts of global warming (Smith et al. 2009; Pauli et al. 2003). Low-elevation ecosystems tend to be more impacted by the complex interactions of disturbances, invasive species, and species change resulting from biotic factors such as competition and pathogens, for examples. Furthermore, the broader distribution of the alpine biome across latitudes and continents should provide a basis for comparison and generalization for warming impacts. Alpine environments can better serve such a role if accurate predictions can be made for where and when vegetation change is likely, within and among alpine areas. These predictions would be possible with an understanding of climatic, landscape, and biotic factors that increase or decrease the vulnerability to climate change. Furthermore, predicting where and when the change is likely to occur will help in devise appropriate monitoring systems that are key for adaptive management. For example, the alpine-treeline ecotone is considered a bellwether for climate impacts, but how and where should change be evaluated? Young tree seedling abundances at alpine-treeline may be a relatively sensitive indicator of incipient change, considering that the tight coupling of tree seedling establishment in alpine-treeline ecotones to long-wave radiation balance, nighttime frost (and thus greenhouse effect), and annual climate variability contrasts with older treeline trees, in which fewer changes are apparent even over centuries of climate variability. Considerations like this help identify demographic stages, particular species, and points within the landscape that might be more likely to express change that portends larger ecosystem and landscape transformations. Notably, tree seedling establishment has increased in many alpine-treeline ecotones globally (Harsh et al. 2009), but there is considerable and unexplained variation about this tendency among mountains. Vegetation management tends to occur at scales closer to a particular alpine site, and an understanding of why the variability occurs will enable translation of the information in broad-scale vulnerability assessments back to the land management decisions that are the fulcrum for human adaptation to climate change.

How can the information needed to enable this cross-scale understanding be provided? Systematic and uniform sampling across mountain ranges, globally, such as in the GLORIA project (Pauli et al. 2003), may provide key data for intercomparison, meta-analysis, and synthesis. Second, identifying where generalizations about the mechanisms governing alpine and treeline responses to climate (i.e., uphill advance of forest into alpine) as they vary globally can or cannot be made is a key step toward scaling up information from the many plot- or single-site studies (Smith et al. 2009). This is a direction pioneered by Ch Körner that is



challenging, but the data and information needs are feasible and will lead to key insights.

At the landscape scale, a promising research frontier is on how connectivity for movement of species or genes influences alpine and treeline change, i.e., biogeographic and evolutionary controls. At the ecosystem level, alpine areas are increasingly influenced by atmospheric deposition of nitrogen and particularly dust, and how will these changes modify alpine responses to warming? At the plant community level, can concepts like positive, facilitative associations among species be used to predict local shifts in assemblage? Will the emergence of novel climate conditions affect the vulnerability of alpine areas to upward migration of exotic species as a result of their dispersal advantages and opportunism? At the organismal level, the precise manner in which alpine climates limit plant species, such as trees, is not resolved. There is considerable debate on whether carbon stores, i.e., mobile carbohydrates, can be used to identify rate-limiting processes for treeline change, which illustrates just one frontier in the broader quest to understand physiological limitation as it applies to understanding plants at their climate limit. Ecophysiological theory suggests that plants operate such that their growth is not limited so strongly by any one particular factor or process, and the compensatory responses that generate balance can occur throughout the whole plant. Research addressing limitation has tended to emphasize a select few processes that we are poised to measure given available instrumentation, and the assessments are rarely at the unit of selection on the landscape: i.e., the whole plant. Application of molecular tools, such as high-throughput genetic approaches to characterizing transcriptomes and thus the full breadth of how alpine plants respond to changes in their environment, is a likely path forward.

---

## References

- Ball MC. The role of photoinhibition during tree seedling establishment at low temperatures. In: Baker NR, Bowyer JR, editors. Photoinhibition of photosynthesis from molecular mechanisms to the field. Oxford: BIOS Scientific; 1994. p. 365–76.
- Bansal S, Germino MJ. Carbon balance of conifer seedlings at timberline: relative changes in uptake, storage, and utilization. *Oecologia*. 2008;158:217–27.
- Billings WD, Mooney H. The ecology of arctic and alpine plants. *Biol Rev*. 1968;43:481–529.
- Billings WD. Adaptations and origins of alpine plants. *Arct Alp Res*. 1974;6:129–42.
- Bowman WD, Seastedt T, editors. Structure and function of an alpine ecosystem: Niwot Ridge, Colorado. New York: Oxford University Press; 2001.
- Callaway RM, Brooker RW, Choler P, Kikvidze Z, Lortiek CJ, Michalet R, Paolini L, Pugnaire FI, Newingham B, Aschehoug ET, Armasq C, Kikodze D, Cook BJ. Positive interactions among alpine plants increase with stress. *Nature*. 2002;417:844–8.
- Cordell S, Goldstein G, Mueller-Dombois D, Webb D, Vitousek PM. Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. *Oecologia*. 1998;113:188–96.
- Ellison L. Subalpine vegetation of the Wasatch plateau, Utah. *Ecol Monogr*. 1954;24:89–184.
- Germino MJ, Smith WK. Differences in microsite, plant form, and low-temperature photoinhibition in alpine plants. *Arct Antarct Alp Res*. 2000;32:388–96.
- Grabherr G, Pauli MGH. Climate effects on mountain plants. *Nature*. 1994;369:448.

- Harsh MA, Hulme PE, McGlone MS, Duncan RP. Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecol Lett.* 2009;10:1040–9.
- Harte J, Shaw R. Shifting dominance within a montane vegetation community: results of a climate-warming experiment. *Science.* 1995;267:871–82.
- Inouye DW. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology.* 2008;89:353–62.
- Jordan DN, Smith WK. Energy balance analysis of night-time leaf temperatures and frost formation in a subalpine environment. *Agr Forest Meteorol.* 1995;77:359–72.
- Körner C. A re-assessment of high elevation treeline positions and their explanations. *Oecologia.* 1998;115:445–59.
- Körner C. *Alpine plant life: functional ecology of high mountain ecosystems.* 2nd ed. Berlin: Springer; 2003.
- Leuschner C. Are high elevations in tropical mountain arid environments for plants? *Ecology.* 2000;81:1425–36.
- Lutz C, editor. *Plants in alpine regions: cell physiology of adaptation and survival strategies.* Wien/New York: Springer; 2012.
- McDowell NG. Mechanisms linking drought, hydraulics, carbon metabolism, and vegetation mortality. *Plant Physiol.* 2011;155:1051–9.
- Monson RK, Rosenstiel TN, Forbis TA, Lipson DA, Jaeger III CH. Nitrogen and carbon storage in alpine plants. *Integr Comp Biol.* 2006;46:35–48.
- Nagy L, Grabherr G. *The biology of alpine habitats.* New York: Oxford University Press; 2009.
- Pauli H, Gottfried M, Reiter K, Grabherr G. High mountain summits as sensitive indicators of climate change effects on vegetation patterns: the “multi summit-approach” of GLORIA (global observation research initiative in alpine environments). *Adv Glob Chang Res.* 2003;9:45–51.
- Rundel PW, Smith AP, Meinzer FC, editors. *Tropical alpine environments: plant form and function.* Cambridge: Cambridge University Press; 1994.
- Ryan MG. Tree responses to drought. *Tree Physiol.* 2011;31:237–9.
- Seastedt TR, Bowman WD, Caine TN, McKnight D, Townsend A, Williams WM. The landscape continuum: a model for high-elevation ecosystems. *BioScience.* 2004;54:1111–21.
- Smith WK, Johnson DM. Biophysical effects of altitude on plant gas exchange. In: De la Barrera E, Smith WK, editors. *Biophysical plant ecology: perspectives and trends.* Mexico: Universidad Nacional Autónoma de México Press; 2009. p. 257–80.
- Smith WK, Vogelmann TC, Bell DT, DeLucia EH, Shepherd KA. Leaf form and photosynthesis. *BioScience.* 1997;47:785–93.
- Smith WK, Germino MJ, Hancock TE, Johnson DM. Another perspective on altitudinal limits of alpine timberlines. *Tree Physiol.* 2003;23:1101–12.
- Smith WK, Germino MJ, Johnson DM, Reinhardt K. The altitude of alpine treeline: a bellwether of climate change effects. *Bot Rev.* 2009;75:163–90.
- Terashima I, Masuzawa T, Ohba H, Yokoi Y. Is photosynthesis suppressed at higher elevations due to low CO<sub>2</sub> pressure? *Ecology.* 1995;76:2662–8.
- Tomback DF, Arno SF, Keane RE, editors. *Whitebark pine communities: ecology and restoration.* New York: Island Press; 2001.
- Wiley E, Helliker B. A re-evaluation of carbon storage in trees lends greater support for carbon limitation to growth. *New Phytol.* 2012;195:285–9.

Kim M. Peterson

## Contents

Introduction .....	364
Boundaries of the Arctic .....	366
Desert and Tundra .....	366
Arctic Flora .....	369
Plant Life in the Cold .....	370
Permafrost .....	377
Roots .....	378
Nutrients .....	380
Moisture and Vegetation Patterns .....	381
Geomorphic Processes .....	382
Cryoturbation, Needle Ice, and Other Soil Disturbance .....	385
Arctic Climate Change .....	386
Future Directions .....	386
References .....	388

---

## Abstract

- From a biological perspective, there is no universally accepted definition of the Arctic, but Arctic plants are generally considered to be those living in tundra and polar deserts beyond the northern climatic limits of forests, i.e., generally north of the boreal zone. The boundary between boreal forests and the Arctic is often broad and ambiguous.
- Arctic plants exist along a global continuum of decreasing floristic diversity with increasing latitude. This gradient starts well outside of the Arctic and continues within the Arctic to the northernmost reaches of land.
- Arctic plants come in a wide variety of forms. Mosses, lichens, and low-growing woody and herbaceous perennials characterize Arctic vegetation. Trees, succulents, ferns, and annual plants are rare or absent from most Arctic

---

K.M. Peterson (✉)

Department of Biological Sciences, University of Alaska Anchorage, Anchorage, AK, USA

e-mail: [km.peterson@uaa.alaska.edu](mailto:km.peterson@uaa.alaska.edu)

plant communities. Combinations of mosses, lichens, sedges, grasses, and dwarf woody shrubs dominate most Arctic tundra, and miniature flowering plants dominate the polar deserts.

- Adaptations of Arctic plants to cold and short growing seasons as well as other aspects of their physical environment are evident in their morphologies, physiologies, and life histories. Arctic plants are also adapted to their biotic environment
- Extremely low temperatures are less characteristic of the Arctic than they are of some other regions, but the Arctic is consistently cold, resulting in permafrost and direct and indirect environmental challenges to plants. During short growing seasons Arctic plants utilize seasonally thawed soils above the permafrost and tolerate frozen soils in winter.
- Low temperatures affect the availability of mineral nutrients, frequently limiting the growth and productivity of Arctic plants. Usable soil is limited by permafrost, and low temperatures retard soil genesis, microbial activity, and uptake by roots. Birds and mammals play a key role in nutrient redistribution and the creation of local sites with high fertility.
- Arctic vegetation patterns are closely correlated with moisture and steep local moisture gradients are characteristic of the Arctic. Although the Arctic is climatologically a desert, few Arctic plants experience water stress.
- Moisture affects thermal characteristics and oxygenation of soils, which in turn affects decomposition rates and the availability of mineral nutrients. Patterns of moisture are strongly influenced by topography due to the combined effects of low precipitation, low evaporation, and water ponding due to permafrost.
- Mechanical stresses associated with freezing and thawing of soils and substrates shape the habitats of Arctic plants. Geomorphic processes unique to cold regions produce vegetational patterns and can lead to cyclic plant succession.
- The climate of the Arctic is dynamic, and changes in past plant communities have occurred on a wide variety of time scales. It is very difficult, if not impossible, to anticipate the effects of a changing climate on the Arctic due to the diversity of plants and habitats and due to nonlinear interactions between environmental factors within Arctic ecosystems.

---

## Introduction

The Arctic is a cold treeless expanse of plains, hills, and mountains, including the northernmost parts of continental Eurasia and North America and numerous high-latitude islands, the largest being Greenland. Collectively these lands surround the Arctic Ocean. Even though the Arctic Ocean is variously ice covered, like other oceans it ameliorates climate, reducing extremes of temperature. Despite these maritime effects, the Arctic is cold and climatically dry. Low temperatures and

limited heat resulting from low solar angles in summer and darkness in winter keep the Arctic frozen much of the year. The sun seems to be forever rising or setting during the brief growing season and vanishes all together for extended periods during Arctic winters. Water is frozen much of the year creating potential physiological drought, and precipitation is generally low throughout the year. Despite a lack of Arctic precipitation associated with persistent polar high pressure, locally moist or even wet habitats are common throughout the Arctic.

Low temperatures and a general lack of heat profoundly affect the ecology of Arctic plants. Arctic plants face a host of challenges, including freezing temperatures, short growing seasons, limited soil fertility, episodic herbivory, and low pollinator frequencies. As a result the Arctic flora is small relative to other ecosystems and represents the end of a latitudinal gradient in floristic diversity that begins high in the tropics and declines to a minimum in the Arctic.

Arctic plants share requirements for light, carbon dioxide, mineral nutrition, and water common to all plants, and they must be able to meet these requirements within the unique constraints imposed by the Arctic. Few of the species able to persist in the Arctic are restricted entirely to Arctic ecosystems. The geographic distributions of many Arctic plant species extend outside the Arctic to high mountains, bogs, or boreal landscapes, and plant species confined to the Arctic often have close relatives in alpine or boreal areas. This is to say that many plant species found in the Arctic can, and often do, grow well outside of the Arctic, but the reverse is not true, i.e., relatively few plants found outside the Arctic are able to grow and persist in the Arctic. The Arctic environment is a selective filter, admitting a small flora, requiring plants to tolerate short cold growing seasons and long frozen winters.

Despite decades of research, many questions of climate change and potential effects upon the Arctic remain unanswered. Arctic ecosystems encompass a broad diversity of habitats: Deserts, semideserts, ice caps, glaciers, rock fields, dry tundras, moist tundras, wet tundras, shrublands, heaths, bogs, marshes, salt marshes, and aquatic communities are part of the diversity found in the Arctic. There is no single Arctic vegetation, but a matrix of distinct environments with distinct vegetational assemblages, each with differing susceptibilities to environmental change (Crawford 2008).

In Arctic landscapes, the magnitude of microhabitat distinctions is large and so fine-grained that moving a plant a few centimeters might easily put it into a habitat type for which it is ill adapted. Differences as great as those found between distinct ecosystems in factors such as temperature, soil aeration, soil moisture, snow cover, soil fertility, length of the growing season, depth of the thaw, competition, and rates of herbivory can frequently be traversed in a single step repeatedly across entire Arctic landscapes. To understand this unique aspect of Arctic plant ecology, it is necessary to understand how climate and geomorphology interact to produce unique Arctic vegetation patterns. Understanding the web of feedback interactions between landscape, moisture, mineral nutrition, and vegetation helps us to understand the complexity of these landscapes and the difficulty of making predictions regarding climate change in the Arctic.

---

## Boundaries of the Arctic

Everyone agrees the Arctic includes the northernmost landscapes of the world, but exactly where the boundary may be found has been a source of debate (CAFF 2001). Considering the earth to be a sphere, the Arctic Circle, at latitude  $66^{\circ}33'$  North, corresponds to the theoretical latitude experiencing sun continuously above or continuously below the horizon on the solstices. This is a geometric consequence of the earth's axis being inclined at an angle of  $23^{\circ}27'$  from perpendicular to the plane in which the earth orbits around the sun (called the plane of the ecliptic). Because the earth is not a perfect sphere, actual solar observations at any particular latitude vary slightly from theory; however, the Arctic Circle, an imaginary line, represents a simple definition of the Arctic. This latitude, however, does not correspond particularly well with the distributional patterns of Arctic plants or of climatic phenomenon generally associated with the Arctic.

Arctic tundra is characterized by a lack of trees, and ecologists sometimes consider the northernmost limit of trees to be the southern boundary of the Arctic. Tree lines, however, are never actually lines, and the transition from tree-dominated taiga to treeless tundra is often a very broad zone sometimes spanning hundreds of kilometers. Further complicating tree line's usefulness as a boundary is that the tree line concept is variously interpreted as the continuous forest boundary, the limits of merchantable timber, the limits of certain-sized individuals, or the limits of dwarf individuals of species that typically form trees in warmer climates. Each of these represents a different, but equally complex and dynamic, boundary. The patchwork of forest and tundra along the Arctic ecotone is indistinct, but it provides a valuable sense of the complexity of defining an Arctic boundary.

Alternatively, the boundary of the Arctic is sometimes considered strictly climatologically, with a  $10^{\circ}\text{C}$  mean temperature for the warmest month (July) being commonly considered a useful boundary of the Arctic. Thermal boundaries such as the  $10^{\circ}\text{C}$  mean July isotherm correspond reasonably well with tree line, but this approach also results in a complex and dynamic boundary. The ancient philosopher Aristotle is credited with the observation that "nature abhors a vacuum"; modern ecologists might add the corollary that nature abhors boundaries, especially an Arctic boundary. The broad ecological boundary of the Arctic is both porous and fractal, and many of the plants and plant adaptations discussed in this chapter permeate whatever boundary is used.

---

## Desert and Tundra

Just as the Arctic boundary is complex, so too are vegetation and environmental boundaries and gradients within the Arctic. The Arctic is vast and includes several distinct vegetation zones. Ecologists frequently distinguish between the High Arctic dominated by polar deserts and the Low Arctic dominated by moist and wet tundra. Alternatively, a distinction is sometimes made between the mountainous Arctic and the lowland Arctic. These broad types can be further divided into a

range of vegetation types at both regional and local scales that frequently reflect patterns of soil moisture. Different kinds of deserts, sedge grasslands, bogs, and shrub-dominated communities persist in the Arctic. Names such as polar desert, polar semidesert, moist tundra, wet tundra, tussock tundra, shrub tundra, coastal tundra, and bog are commonly used to differentiate between broad classes of vegetation found within the Arctic.

Polar deserts are unlike other deserts in that the sparseness of vegetation results from a lack of heat as opposed to a lack of water (Bliss 1997). The growing season simply does not last long enough for plants to exhaust the soil moisture available from snowmelt. Density of the vegetation is correlated with differences in growing season temperature and length, and shifts in vegetation density over decades or even centuries occur in response to climatic trends. Few plants are capable of establishment and growth during the extremely short growing season, and seed production is not always possible. Some communities depend upon exceptional years for seed production and establishment of new individuals into the population, while other communities may never set seed and depend entirely upon seeds dispersed from distant populations or upon vegetative reproduction for recruitment of new individuals into populations.

Among plants capable of persisting in the polar deserts, *Saxifraga oppositifolia* (purple saxifrage) has adapted to a range of habits, with local ecotypes even adapted to habitats where the short growing season of the High Arctic is further shortened by late melting snow banks; it does this by increasing metabolic rates and speeding shoot growth at the expense of accumulating energy and water reserves characteristic of other plants in nearby habitats. Different ecotypes of *Saxifraga oppositifolia* also adjust their relative sexual and vegetative reproductive strategies by habitat. Vegetative reproduction and pseudoviviparity (producing bulbils or plantlets rather than seeds) are notable adaptations to shortened growing seasons common in Arctic plant species.

In polar deserts, microhabitats are important to seedling establishment. Microrelief in soil patterns associated with frost activity can create patterns of seedling recruitment and survival that persist as vegetation patterning in an otherwise barren landscape (Fig. 1). Centimeters or even millimeters of physical relief create microclimates useful to plants, with slight variations in solar input, temperature, moisture, and snow cover defining the limits of potential habitats.

At the other extreme of Arctic vegetation, wet landscapes are densely covered with sedges, grasses, mosses, and forbs (Fig. 2). Growing seasons are longer in the Low Arctic, and productivity is less limited by heat than by competition for light and mineral nutrients (Brown et al. 1980). Water from snowmelt frequently remains ponded on the surface for much or all of the growing season, and emergent aquatic vegetation and ponds occupy the lowest areas. Soil aeration is poor and root respiration requires aerenchyma (air passages) in the roots of plants in the wettest sites. Locally better-drained areas are typically home to woody vegetation, particularly dwarf willows, but many species of dwarf shrubs may be found.

Much lies between the extremes of polar desert and wet coastal tundra, including mesic or moist tundra, tussock tundra, and shrub tundra. These types are the most

**Fig. 1** Polar desert on Cornwallis Island in the Canadian High Arctic showing patterned vegetation dominated by lichen (*Cetraria*) and purple saxifrage (*Saxifraga oppositifolia*) (Photo credit L. C Bliss)



**Fig. 2** Wet coastal tundra near Barrow, Alaska, dominated by sedge (*Carex aquatilis*) and grass (*Arctophila fulva*)



productive among Arctic vegetation with productivity ultimately limited by the availability of mineral nutrients. Tussock tundra, dominated by the well-studied sedge *Eriophorum vaginatum* (Tussock Cottongrass), covers much of Alaska's North Slope. An accumulation of tightly packed stems forms a ball, or tussock, with which this species constructs its own habitat. Elevated above the surrounding surface, tussocks capture more of the low-angled light. Snow accumulation in the inter-tussock hollows provides moisture augmentation while still allowing the tussocks to emerge from the snow early in the growing season. *E. vaginatum*, along with other species of *Eriophorum*, have the unique adaptation of producing an entirely new root system each year. This annual root system allows root tips, which are the points of uptake of water and nutrients, to follow the thawing soil down over the course of the growing season (as opposed to being distributed throughout the soil where many remain frozen until soils thaw).



## Arctic Flora

The current flora of the Arctic consists of between 2,000 and 2,500 vascular plants and is recent, being largely a product of the Quaternary Period. The Arctic was forested during the Tertiary Period, and many of the temperate trees of Eastern Asia, Eastern North America, and Europe show relationships to one another that attest to their Arcto-Tertiary origins. With Arctic cooling, these forests retreated from the Arctic and vegetation similar to that currently in the Arctic began to appear about three million years ago. Multiple sources are likely for the current flora, including temperate mountains, especially those of Eurasia, and as much of the Arctic remained ice-free throughout glacial periods of the Pleistocene, both local and distant refugia (areas free of ice where plants persisted) complicate past migratory patterns.

Despite the variability within Arctic vegetation, the flora of the Arctic is highly conserved, i.e., the same or closely related species comprise similar vegetation types throughout the Arctic (Polunin 1960). Such circumpolar floristic similarity is strongest in the Arctic but is also characteristic of the boreal zone (Hulten 1968). A pattern of increasing floristic affinity with latitude is understood as a consequence, in part, of plant migrations between the old and new worlds (Eurasia and North America) and the relative youth of the tundra biome.

The area currently occupied by the shallow Bearing Sea between the Russian Far East and Alaska has been land covered with terrestrial vegetation multiple times in the past (most recently during the last glacial maximum), allowing plants and animals to migrate between the continents of Eurasia and North America. This ephemeral continental connection is called the Bering Land Bridge, and the region including adjacent lands is often referred to as Beringia. Plant and animal migrations associated with Beringia help us understand the current high degree of floristic similarity throughout the Arctic. In addition to such migration, some plants may have established (or maintained) circum-Arctic affinities through long-distance dispersal. Spores of mosses are known to travel long distances in the atmosphere, even reaching the jet stream where they can circle the earth. Animals, especially birds, are effective agents of seed dispersal, and the Arctic has many migratory birds nesting during summer.

Perhaps unique to the Arctic is long-distance plant dispersal by ice. Plants growing (and even blooming) have been observed atop glacial ice rafted across the Arctic Ocean. Icebergs born from Arctic mountains are sometimes discharged into the Arctic Ocean bearing soils or gravels containing plants or seeds. Massive, these ice islands sometimes persist for decades locked in Arctic Ocean sea ice as it circulates in the prevailing clockwise currents (as seen from above the pole). Ice islands occasionally find foreign shores bearing immigrant plants as passengers. Arctic salt marshes show an interesting pattern of distribution of the grass *Puccinellia phryganodes* with distinct regions having been colonized by plants with differing numbers of chromosomes, some fertile and some sterile (although capable of vegetative reproduction by prostrate stems or stolons). Stolons of this salt marsh grass embedded in sea ice have reportedly been recovered and grown, providing a potential mechanism for distribution of this widespread Arctic species.

## Plant Life in the Cold

Plants in the Arctic are characteristically low-growing. Being small or prostrate affords the protection of snow cover from desiccating winter winds and cold and maintains plants in a microclimate with the warmest summer temperatures (Crawford 2008). Herbaceous plants commonly die back to the ground surface or to just below the ground surface during winter. Common woody plants in the Arctic appear stunted or prostrate in comparison to their temperate counterparts and are generally dependent upon snow cover during winter to help protect their dormant buds. Woody plants include both deciduous and evergreen forms (these evergreens are not conifers which are generally restricted to the Arctic boundary, but are flowering plants like *Ledum groenlandicum*, *Vaccinium vitis-idaea*, *Empetrum nigrum*, and *Rhododendron lapponicum*). Mosses and lichens are common components of nearly all plant communities in the Arctic, spore-bearing vascular plants are rare, but the genera *Equisetum* (horsetail) and *Lycopodium* (clubmoss) are represented in locations throughout the Arctic. Most habitats are dominated by flowering plants, including grasses, sedges, dwarf shrubs, prostrate shrubs, and a variety of perennial herbaceous life forms including cushion plants and rosettes. Conspicuously absent (or extremely rare) are trees, succulents, and annual plants.

Cold is generally associated with the Arctic. Cold refers to a condition of low temperatures; a related, but subtly different, concept is a lack of heat energy. Although low temperatures and a lack of heat are responsible for sculpting many aspects of Arctic environments, extreme low temperatures are more characteristic of continental climates such as the boreal forests of Siberia and North America and the steppes and prairies of Mongolia and the Dakotas. These ecosystems frequently experience lower winter temperature extremes than the Arctic. So too many high elevation mountain ecosystems experience lower temperatures than those frequently encountered in the Arctic. It is not the extremely low temperatures that characterize Arctic cold; it is the consistency of low temperatures. The high summer temperatures found in boreal and temperate continental climates are lacking in the Arctic where it remains relatively cold all year round.

To understand the environments of Arctic plants, it is important to be able to distinguish between temperature and heat and to understand the interrelationship between them. Temperature may be thought of as reflecting the average kinetic energy of the atoms or molecules within a substance. Temperature is important to plants as it influences the rates of processes like diffusion and chemical reactions. Chemical reactions depend directly upon temperature, and in biological systems, enzyme activity is influenced by temperature. Heat is a form of energy and can move from one substance to another. Temperature predicts the direction of heat flow between substances (heat flows from high temperature to low temperature). Heat may be measured in units appropriate to measuring energy such as Joules, although thermal energy is frequently measured in calories, a calorie being the amount of thermal energy necessary to raise the temperature of 1 g of water 1 °C. Twice as much heat is required to raise the temperature of 2 g of water 1 °C; thus, temperature is not a measure of heat.

Most substances will warm more quickly than water with an equivalent input of heat. The amount of heat required to raise the temperature of a substance compared to the same amount of water is called the heat capacity of the substance. Liquid water is thus a standard with a heat capacity of one, which is approximately twice that of either ice or water vapor and more than four times that of dry air. Plants are primarily composed of water and thus require more heat per unit mass to elevate their temperature 1° than does an equivalent mass of surrounding air, dry soil, or dead plant material. Because water has such a high heat capacity, moisture in the environment is an important control over the thermal behavior of soils, and flowing water can efficiently transport heat.

Changes of state of a substance also involve gaining or losing heat. In the Arctic, water exists in all three states: solid, liquid, and gas. Melting ice or vaporizing water requires thermal energy input that is not reflected in a change in the temperature. If heat is added to ice, it will initially warm to thawing temperature as predicted by the heat capacity of ice, and then it will continue to absorb heat without a corresponding rise in temperature until all of the ice is melted. Once melted, continued addition of heat will warm the water, eventually to the point of vaporization, whereupon the temperature will again remain constant, despite the continued addition of heat, until all of the water is evaporated. Continued heating at that point will elevate the temperature of the vapor. As noted above, the rates of temperature rise in ice, water, and vapor are not equal, being a function of their distinct heat capacities, but this heat addition results in a change in temperature and is termed sensible heat. When heat is added without a corresponding change in temperature, such as when ice is thawing to liquid water or when water is evaporating to vapor (both cases of a change of state), the heat consumed is termed latent heat. Water has high latent heats associated with its phase transitions compared to most other substances. Freezing and thawing are an integral aspect of the environment of Arctic plants, and the corresponding latent heat requirement is a significant energy requirement. High latent heat of fusion for water further implies a correlation between moisture and the thermal characteristics of the environment.

Heat moves by conduction, convection, and radiation or through latent heat exchange. All are important in understanding the thermal environment of Arctic plants and plant temperatures. Conduction is the transfer of thermal kinetic energy between materials in contact with each other. Heat absorbed at the soil surface must be conducted into the soil. Substances differ in their ability to conduct heat (thermal conductivity) and since they also differ in their heat capacities, the amount of heat required to elevate the temperature of soils may differ from one stratum to another. The rate at which heat moves through the soil is a function of both the thermal conductivity and the heat capacity and is termed the thermal diffusivity. Not surprisingly the major variable controlling thermal diffusivity in Arctic soils is soil moisture content. As heat moves through the soil to the depth of frozen material, additional heat is required to melt ice in the soil before heat can be conducted to deeper depths.

Heat is lost to the atmosphere from the soil surface and from plants via convection (as adjacent air is heated and rises away from the surface) and by latent heat

**Fig. 3** The hairy new growth of woolly lousewort (*Pedicularis lanata*, a.k.a. *Pedicularis kanei*) helps keep the warmth from sunlight from being lost to the wind



loss (as water evaporates at the surfaces). The soil and plant surfaces also exchange heat via radiation. Heat is absorbed primarily from solar radiation, and heat is lost, since all objects emit radiation as a function of their temperature. The emitted radiation is invisible as it is in wavelengths too long to be seen, but thermal imaging can reveal that warmer objects are brighter in these invisible wavelengths than cooler objects. Emission is a function of the temperature of the surface, which along with evaporation and convection may be influenced by the depth of a layer of relatively calm air held near the surface called the boundary layer (with the thickness of this layer largely being a function of the roughness or smoothness of the surface and the wind speed). Many Arctic plants appear fuzzy or hairy as a consequence of their morphological adaptations to increase their boundary layer and reduce heat loss (Fig. 3).

The low temperatures and limited heat in Arctic environments present both direct and indirect challenges to plants. Plants must germinate, metabolize, grow, and reproduce at tissue temperatures lower than those of plants in most other ecosystems. Low tissue temperatures both above- and belowground require plants to adjust enzymes, adjust membranes, and enhance transport processes to compensate. Altering enzymes to operate at low temperatures (largely through genetic adaptation) is not always possible, and some morphological and physiological strategies are missing in the Arctic. For example, the C4 photosynthetic pathway is not found at all in the Arctic. Some plants compensate for lowered specific enzymatic activity by increasing the total amount of enzyme (as is common with the photosynthetic enzyme RuBisCO), but this in turn generates higher demands for nitrogen to build enzymatic proteins and potentially contributes to nutrient stress. Membrane permeability can be increased at low temperatures by making the lipids more fluid. This is done by decreasing the degree of hydrogen saturation of the fatty acid tails in the membrane phospholipids, i.e., increasing the number of double bonds in the hydrocarbon chains in the fatty acid tails of these phospholipids. This is especially important in roots to allow efficient uptake of mineral nutrients as soils thaw.

In addition to the challenges of growth and reproduction at low temperatures, Arctic plants face potential damage due to freezing. Freezing can damage cells by mechanical disruption such as rupturing membranes. Water expands as it freezes, and a growing ice crystal can puncture membranes or even split cell walls. Freezing can also damage cells through disrupting metabolism. Water loss to growing ice is a form of drought stress, and with insufficient liquid water, enzymes denature. The functional shape of enzymes and other proteins depends upon their tertiary structures being maintained by the effect of water on the hydrophobic and hydrophilic portions of the amino acid chain. Drying due to freezing allows proteins to lose important properties associated with their shape.

Arctic plants deal with freezing temperatures by controlling the freezing process or by avoiding freezing. Most species greatly reduce the amount of tissue maintained over the winter period through senescence of leaves and other tissues. Perenniating tissues (those remaining alive over the winter) must deal with freezing temperatures. Strategies include (1) controlling where ice forms, such as allowing ice to form in intracellular spaces but avoiding ice formation inside cells; (2) increasing the concentration of solutes such as proteins or sugars to reduce the freezing point of the cell solution; and (3) eliminating ice crystal nucleation allowing supercooling. Supercooling is remaining in an unfrozen (or uncrystallized) state at temperatures below freezing. This is possible as the initiation of ice crystal growth requires a starting point called a nucleating agent. This is an unstable condition, and water in a supercooled state can crystallize very rapidly if a nucleating agent is introduced. Plants and many soil organisms have a variety of adaptations to prevent ice crystals from forming in cytoplasm.

Winter cold presents a hazard of desiccation. Water in plants may move from cytoplasm to accumulate as ice in intercellular, intracellular, or external locations. Given plant roots and rhizomes are completely imbedded in a matrix of frozen soil and snow during winter, they have no opportunity to replace lost water. Avoidance of desiccation injury helps explain high tissue turnover (exfoliating leaves and fine root material and in some cases senescing all but a small area of perennating tissue at the base of the stem). Snow cover in winter may help provide some plants with protection from the desiccating and cooling effects of wind, and some plants like the evergreen *Cassiope tetragona* are restricted to areas that provide adequate snow cover during winter. Arctic plants face a trade-off between winter protection by snow cover and a delayed initiation of the growing season due to the time in spring required to melt overlying snow.

In addition to low temperature effects, plants are challenged by the limited amount of seasonal heat available. One way to visualize this is as a limitation to the length of the growing season. Spring in the Arctic sees the return of the solar radiation, but thaw, or breakup as it is commonly called in the Arctic, comes closer to summer. A useful approximation is to consider that roughly half of the annual input of solar radiation is used to melt the winter's accumulation of snow and ice. The growing season for many areas in the Arctic will begin about the same time the sun has reached its apex and begins its apparent southward migration again. The total annual primary productivity must be accomplished between the summer solstice and

the initiation of freezing sometime in September or August. The growing season is generally less than 100 days throughout the Arctic, and closer to 60 days in extreme locations. Arctic plants must break dormancy, grow, flower, ripen and disperse seeds, and enter dormancy within a window of 2 or 3 months. At the beginning of the growing season, much of their root system may still be locked in frozen soil, and at the end of the growing season, failing light and accumulating snow may arrest aboveground activities even though soils at depth have not completely refrozen.

Breakup comes slowly, but it happens suddenly. Much solar radiation is reflected from the snow surface, but thermal radiation from clouds is effectively absorbed. Some solar radiation is transmitted through the snow where it may be absorbed by the vegetation before the snow melts. Moss and lichen photosynthesis begins under the snow, and under some conditions vascular plants become active before the snow is completely gone. Plant cytoplasm and soil solution have freezing points below 0 °C due to the freezing point depression of their solutes, and thus, metabolic activity can begin under the snow. The lichen *Cetraria nivalis* has been shown to achieve net photosynthesis at -5 °C. As the snow surface melts, heat is moved down into the snow and eventually to the surface as melt water percolates down into the snow where it refreezes, giving up its latent heat of fusion and in the process warming the deeper snow. The entire snow mass rises to melting temperature at approximately the same time. Within a period of a few days the snow changes to liquid and the landscape becomes covered with flowing water. Flooding is characteristic of Arctic landscapes at breakup. Especially dramatic is the flooding on the coastal plains, as the water flows across the surface, being unable to penetrate the frozen ground, resulting in widespread submersion of the vegetation.

A crude measure of heat available for vegetation is the number of degree-days (or growing degree-days) accumulated in a particular site. Although various definitions of degree-days have found utility in various studies, a simple formulation is to simply sum the average temperatures for each day for which the average temperature is above zero centigrade. Thus, 10 days at 1 °C equals 1 day at 10 °C equals 10°-days. While crude, if measured at the precise location of plant tissues, this provides a means of comparison between habitats that is somewhat correlated with the heat available to the plants, and it is slightly more nuanced than simply reporting the length of growing season. Extreme Arctic habitats may have fewer than 300°-days. Compare this to a requirement of nearly 3,000°-days for a corn crop to reach maturity, and it quickly becomes apparent that availability of heat represents a limitation to Arctic plants.

An explanation for the general lack of annual plants in the Arctic is that there is simply not enough time to complete a life cycle from seed to seed during the short growing season. An exception is *Koenigia islandica* L. a tiny annual which appears to function as an annual throughout much of the Arctic. Perhaps it is successful due to its extremely small stature, limiting the requirements for growth. Plants functioning as biennials outside of the Arctic, such as *Cochlearia officinalis*, have been observed to increase the number of seasons used to reach maturity in the Arctic, essentially functioning as perennials. Most Arctic vascular plants are relatively long-lived perennials, and it is thought that multiple growing seasons allow the

**Fig. 4** Pygmy buttercup, *Ranunculus pygmaeus*, blooms within a few days of snowmelt as a consequence of forming the flower buds the previous fall



accumulation of resources necessary for growth and reproduction. Iteroparity (reproducing repeatedly during a lifetime) is far more common in Arctic plants than is semelparity (reproducing only once at the end of life).

Some plants solve the problem of a short growing season by eliminating the need for seed germination. Plants accomplish this through pseudoviviparity or the production of bulbils. Plants like *Poa vivipara*, *Polygonum viviparum* (Alpine bistort), and *Saxifraga cernua* (drooping saxifrage) essentially eliminate the need for seed germination by dispersing miniature plantlets ready to establish themselves without the time required for seeds to break dormancy and germinate.

A common adaptation among Arctic plants is preformed flower buds, accounting for the great speed with which flowers first appear in species such as *Ranunculus pygmaeus* Wahlenb. or *Pedicularis kanei* (Fig. 4). By forming buds the previous growing season, many species flower as they break dormancy. While preformed flower buds are also found among plants outside of the Arctic, the speed with which Arctic plants can flower following thaw is particularly impressive with the first flowers appearing on the tundra a few days after the snow disappears.

Plants generally grow near the ground in the Arctic. Prostrate shrubs, cushion plants, and other low-growing forms find warmer air temperatures in summer near the ground and in winter are generally afforded some protection from low



**Fig. 5** Several species of Arctic poppies track the sun with parabolic flowers, concentrating heat on their reproductive parts. This both rewards pollinators who come to bask in the heat and warm flight muscles and provides elevated temperatures for reproductive metabolism during pollen tube growth



temperatures and winter desiccation by snow cover. Taller shrubs are generally restricted to areas where deep snowdrifts accumulate in winter. Even slightly elevated areas may be blown nearly free of snow in winter, and plants in such habitats are generally among those clinging most closely to the soil surface, and it is here lichens are most abundant.

Production of flowers and ovules and the maturation of seeds may exhibit higher heat requirements than vegetative growth. As a consequence, some plants rarely flower and reproduction is primarily vegetative for some species in Arctic habitats. Others possess adaptations to preserve or augment the heat available for sexual reproduction. Flower development may be close to the surface where thermal conditions are best and plants may form cushions, reducing their exposure to wind. Hairs or fuzz, retarding convective and evaporative heat loss, may protect buds and flowers (Fig. 3). In some cases flower shapes may be parabolic to concentrate reflected radiation onto the pistil and plants such as some members of the genera *Dryas* (avens) and *Papaver* (poppy) combine parabolic flowers with solar tracking to maximize the heat available in the flower (Fig. 5). This adaptation not only benefits the developing ovules but also may serve as a thermal reward to pollinators.

Challenges of the Arctic environment are mostly physical, with biological adaptations less apparent than in species-rich biomes, but two aspects of the Arctic



biotic environment have received attention: herbivores and pollinators. Principal Arctic herbivores include musk ox, caribou (and reindeer), ground squirrels, hares, microtines (lemmings and voles), ptarmigan, moose, beavers, and insects (Nuttall 2004). As productivity in the Arctic is low, the larger herbivores traverse large distances and sporadically visit grazing or browsing areas. Small herbivores are more chronic, but their population numbers fluctuate dramatically; thus, herbivory by both large and small herbivores tends to be episodic. Common morphological adaptations against herbivores such as spines or thorns are generally lacking in Arctic plants, but woody plants in particular have evolved chemical defenses against herbivores. Most commonly these are general defenses such as tannins or other secondary compounds that reduce the digestibility of the plant tissues (and hence reduce the utility to the herbivore of eating the plant).

Pollinators are infrequent in the Arctic in comparison to other regions, but the large burrowing bumblebee *Bombus polaris* is a notable pollinator throughout the Arctic. Flies, moths, thrips, and even mosquitoes act as pollinators in the Arctic, but many Arctic plants self-pollinate, use wind pollination, or reproduce mainly vegetatively. Plants of the genus *Pedicularis* (louseworts) are dependent upon pollinators and appear to have evolved staggered flowering times; this is adaptive both in the sense of reducing foreign pollen transferred to stigmas and in the coevolutionary sense of helping ensure the survival of nectar-dependent pollinators like *Bombus* by ensuring a continuous food supply throughout the summer.

---

## Permafrost

Heat is primarily added to the Arctic by solar radiation and is extracted by emitted thermal radiation from the ground surface back to space. Warm objects emit more radiation than cold objects, but all objects continually emit thermal radiation. The balance of incoming and outgoing radiations over time (assuming no other energy exchanges) determines the temperature of the soil-vegetation surface (and via heat conduction it also determines the temperatures below the surface of the soil). At the ground surface the temperature varies daily (and seasonally) as solar input varies, but at a depth of a few meters, such variations in temperature are very small, being essentially averaged over time by the process of heat conduction and storage. At a few meters depth, the substrate temperature remains close to the long-term mean temperature at the surface. In temperate locations this accounts for the observation of moderate constant temperature in caves, with caves feeling cooler than outside air in summer and warmer than outside air in winter. In the Arctic, the mean temperature is below freezing, resulting in a condition at depth known as permafrost.

Permafrost is not a kind of substance; rather it is a thermal characteristic of a substance. Any substance that remains frozen (i.e., below 0 °C) for a period of 2 years or more is considered to be permafrost. Permafrost may pertain to various kinds of rock, sand, soil, or ice. Permafrost results from the fact that the average annual temperature at the surface is below 0 °C, and insufficient heat is conducted into the substrate to raise temperatures at depth above freezing. Most terrestrial

landscapes in the Arctic are underlain by permafrost, and because of this lands in the Arctic are said to be in the zone of continuous permafrost. It is not uncommon for permafrost to extend deeply into the underlying parent materials, in some cases as much as hundreds of meters. At some point the interior heat of the earth begins to have a larger effect than the history of surface temperatures, and this point defines the bottom of the deep permafrost. In some areas permafrost may be actively growing deeper due to a recent history of colder climatic conditions, and in other areas permafrost may be a relic of past climatic conditions and be slowly disappearing. Of importance to plants is the depth at which the top of the permafrost is found, or more precisely the depth of the soil column that is thawed at any particular time, as this is frequently less than the depth to the permafrost. The depth of thaw reaches a maximum near the end of summer. The perennial maximum depth of thaw defines the top of the permafrost (as permafrost is defined as remaining frozen 2 or more years). The zone of thawing soil atop the permafrost (the active layer) represents the maximum soil volume available to support plant growth.

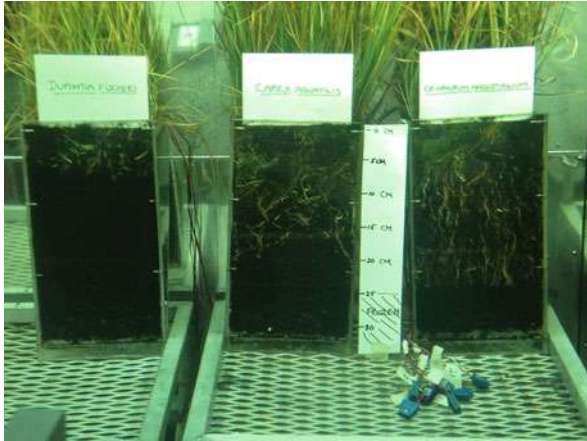
Active layer depths vary throughout the Arctic and locally primarily as a consequence of the ability of various soils to hold and conduct heat. The chief determinant of soil thermal properties is the amount of water in the soil, with water increasing both the heat capacity and thermal conductivity of soils. Active layer depths also vary as a consequence of variations in the amount of solar radiation absorbed by the surface. Surfaces differ in their reflectivity (albedo) with surfaces like peat or vegetation absorbing more energy than surfaces like snow or sand which reflect as much as 90 % or more of the radiation striking them (high albedo). Ponds and lakes absorb more heat, than vegetated surfaces, contributing to deeper thaw under water bodies. In some cases this can result in the formation of a *talik*, or thaw bulb, which is a pocket of thawed material that persists throughout the winter under the frozen surface and above the permafrost. Most Arctic landscapes freeze solid in the winter with the active layer freezing down to the permafrost.

While the bottom portion of the active layer may remain thawed well after the surface has begun to refreeze at the end of the growing season, at some point in the fall or winter, the entire soil column becomes frozen and remains frozen until thaw is initiated at the soil surface at the beginning of the next growing season. Arctic plants are faced with the challenge of living in a soil medium that freezes completely during the winter. As the above- and belowground periods of thaw do not entirely match, Arctic plants have adopted a variety of means to obtain minerals from soils. Freezing of soils also affects plant habitats indirectly through mechanical and geomorphic processes that create and modify local habitats.

---

## Roots

Arctic plant roots grow in cold soils relative to plants of other biomes. Soil temperatures are close to air temperatures at the soil surface but decrease to 0 °C at the bottom of the active layer, which throughout most of the Arctic is just a few decimeters.



**Fig. 6** While superficially similar aboveground, these sedges and grass exhibit morphological differences belowground. On the *left*, the grass *Dupontia fisheri* roots are shallow in the warmest soil near the surface, while the roots of the sedge *Carex aquatilis* explore deeper and colder soils and on the *right* the sedge *Eriophorum angustifolium* sends its annual root system deepest directly against the thawing front of the soil (Photo from an experiment by Gaius Shaver)

The depth of thawed soil at the beginning of the growing season is just a few centimeters at the time of snowmelt, usually in June, and takes a month or more to reach the full depth of the active layer. Thus, plant roots are generally below 10 °C and some are near 0 °C for the entire growing season. At low temperatures diffusion to root surfaces and the permeability of root membranes are lowered. Low soil temperatures also affect soil microorganisms, including beneficial mycorrhizae and rhizosphere organisms. Mycorrhizae are common among Arctic plants, helping to overcome some of the nutrient limitations imposed by Arctic habitats.

Arctic plants exhibit distinct rooting strategies to cope with the challenges of cold soils (Fig. 6). Differences in root longevity, mass, and depth are apparent, even among species that are superficially similar aboveground. At one extreme exist plants of several species of the genus *Eriophorum*, including *Eriophorum vaginatum*, a widespread dominant plant of the low Arctic, possessing an annual root system. These plants perennate from the thickened base of their stems where they store annual reserves accumulated during the growing season, and both aboveground and belowground structures senesce at the onset of winter. A new root system and leaves are produced from the stem base at the beginning of the growing season. The roots of *Eriophorum* are less dense than those of other graminoids and relatively short lived compared with roots of other sedges that may persist many years, but there are two clear benefits of a disposable root system: First, at the beginning of the growing season, there is no unproductive investment of root structure frozen into soils, and second, the root tips, which are the principal point of absorption for the root system, can follow the melting zone, maximizing root tips in the zone most likely to be richest in nutrients.

In contrast, the grass *Dupontia fisheri* maintains a highly branched fibrous root system near the soil surface. This strategy concentrates root tips in the warmest and earliest thawed portion of the active layer. Rooting strategies reflect constraints other than temperature, and species in bogs or saturated soils frequently exhibit aerenchyma, internal passages that allow oxygen to diffuse along roots in poorly aerated soils.

---

## Nutrients

Arctic plants generally experience a limited supply of nutrients, particularly nitrogen and phosphorus ions. A chief cause of low nutrient supply is the limited volume of soils available to plants. Near-surface permafrost can lead to a situation in which vegetation experiences a condition similar to pot-bound greenhouse plants, i.e., the entire soil volume available to plants is already exploited by roots and additional root growth does not yield additional nutrients. Add to this the fact that low temperatures potentially place demands on plants for high levels of nutrients, and it is not surprising that competition for mineral nutrition is a common feature of tundra vegetation. Low temperatures also retard the decomposition (or mineralization) of plant materials, and there is little or no input of minerals from the atmosphere or weathering.

Nutrient availability varies considerably between habitats, and some sites, often associated with animals, are nutrient rich. Birds and mammals concentrate nutrients, particularly nitrogen around nests, dens, or other areas frequently used. Such areas are often easily spotted due to the color and luxuriance of the vegetation. In the mountainous Arctic, areas below cliffs favored by nesting birds, especially sea birds, are particularly fertile. In lowland tundra even small hummocks stand out in the flat landscape and are utilized by snowy owls as hunting perches. These are easily recognized by the vigor and greenness of plants and by the accumulation of owl pellets (wads of hair and bones of small mammals regurgitated by the owls). The dens of ground squirrels and Arctic foxes present fertile habitats for plants, and dens characteristically support plant species with high nutrient requirements such as *Arctagrostis latifolia* that is otherwise rare or lacking in the surrounding tundra. Calcium may be added to soils by antlers. In Caribou, both sexes have antlers that are annually shed, and calving grounds in particular receive a source of calcium, which may be an important plant requirement in acidic tundra.

Plants in most habitats are limited by the availability of nitrogen or phosphorus. Nitrogen is generally available to plants in the form of nitrate or ammonia, with plants in anaerobic soils more likely to be able to effectively use ammonia. In the Arctic, it was first discovered that plants can also potentially take up amino acids directly, and this can contribute significantly to the nitrogen requirements of Arctic plants. Biological nitrogen fixation (converting atmospheric molecular nitrogen to a form usable by plants) is accomplished in wet and moist Arctic habitats by cyanobacteria and lichens, particularly by the lichen *Peltigera*. Due to the restricted availability of nutrients, Arctic plants recover large amounts of nutrients from

senescing tissues. The species *Petasites frigidus* (Arctic Colt's foot) is capable of recovering in excess of 80 % of the nitrogen and nearly 90 % of the phosphorus from its senescing leaves.

Inputs of minerals from weathering of rock and mineral soil are low due to low temperatures. Soil development in general is slowed in the Arctic, and many soils are quite rocky. Wet tundra soils and bogs are often characterized by peat accumulations that occupy most of the active layer. In sites where organic matter accumulates, permafrost generally aggrades, i.e., the additional accumulation of organic materials insulates the soil resulting in upward growth of the permafrost. Here, plants must rely on atmospheric inputs, biological fixation, and minerals released by decomposition (mineralization). Over time the nutrient economies of such sites gets tighter, resulting in decreased productivity.

Phosphate ions are subject to leaching and are moved across the landscape by flowing water at breakup. Low-lying areas are thus generally higher in available phosphorus than are adjacent soils, helping to account for different species in adjacent habitats. Windblown sediments from river bars, along with animals, help move phosphorus back up the hydrological gradient. Mycorrhizae are important to the nutrient economies of Arctic plants.

---

## Moisture and Vegetation Patterns

Precipitation is low throughout the Arctic and generally is well below the threshold for climatic distinction as desert. Global atmospheric circulation deposits air onto the poles, creating polar high pressure. As the air flows south across the Arctic, it is affected by the earth's rotation, and winds rapidly gain a dominant eastward component. This cold dry air warms as it moves, making precipitation even less likely. One must, however, make a distinction between climatologically defined desert and desert vegetation. Although desert vegetation dominates the High Arctic, most Arctic vegetation is not a desert, and many parts of the Arctic are actually dominated by wetlands. One might ask how it is possible to have wet tundra or bogs in areas that are climatologically classified as desert. The answer is low rates of evaporation and little or no percolation of water into deep substrates. These are a consequence of the Arctic cold and permafrost.

The combination of low precipitation, low evaporation, and low percolation produces steep gradients in soil moisture that are controlled primarily by local drainage. Elevated areas tend to be dry and lowlands tend to be wet. Even the flat coastal plains show dramatic differences in soil moisture associated with microrelief. Elevation differences of a few decimeters or even centimeters often result in a distinction between dry and saturated soils or terrestrial and aquatic systems. Essentially the water table in much of the tundra is perched at or very near the soil surface due to the inability of water to penetrate the ice-sealed permafrost substrate below. Areas raised or depressed a few centimeters make a profound difference to plants, and local patterns of vegetation strongly resemble the patterns of microrelief.

**Fig. 7** The redistribution of snow by wind leaves some microhabitats free of protective winter snow cover, while other areas such as stream margins may be covered with many times the average snow depth



Where elevation relief is greater, the redistribution of moisture by blowing snow plays a significant role in determining soil moisture and vegetation patterns. Snow drifts into depressions and blows away from higher areas (Fig. 7). This further enhances the correlation between local relief, soil moisture, and vegetation.

Arctic hydrology is dominated by surface runoff since deep drainage is hampered by permafrost. Spring thaw, or breakup, is usually a dramatic event with much flooding, despite relatively little snow on the landscape. Snowmelt and runoff are sudden and brief, leaving saturated and partly thawed soils in its wake. Patterns of summer precipitation vary across the Arctic and may currently be changing, but in general little precipitation occurs until fall when rains may precede modest snow accumulation. Arctic annual precipitation totals are generally less than 20 cm and in many areas are less than 10 cm of water equivalent. Despite this meager precipitation, poorly drained Arctic landscapes are covered with lakes and ponds of various sizes, and wetland vegetation is more common than deserts.

---

## Geomorphic Processes

Geomorphic processes sculpt local relief. Some processes are similar to those found outside the Arctic, while others are more typical of the Arctic. No matter what geomorphic processes are involved, the importance of local microrelief in the Arctic is paramount to understanding local habitats. Local relief, moisture, and vegetation are inextricably interlinked.

Annual freezing and thawing of soils leads to several phenomena that collectively create patterns of microrelief. Chief among these is the growth of ice wedges. Frozen soils and substrates cool and contract during winter. Wherever the matrix of upper permafrost materials is cemented together by ice, the substrate is inelastic, and annual cooling produces tension cracks up to 2-m deep. As seen from above, these cracks form a polygonal pattern similar to those produced in miniature by

**Fig. 8** Low center polygons in a wet coastal tundra landscape. The relatively wet centers and troughs are frequently just a few decimeters lower than the relatively dry polygon rims that are pushed up by ice wedge growth



drying mud. Snow, water vapor, and melt water enter these contraction cracks before summer heat again expands the substrate. Water in its various forms entering the permafrost is retained as ice and prevents re-expansion of substrate materials into their former position, creating pressures that deform the substrate and raise ridges in the overlying soil.

Repeated patterns of cracking lead to annual increments of ice accumulation with approximately 1 mm added to the width of growing ice formations that are wedge-shaped in vertical cross section (expansion and contraction are greatest near the surface and diminish to nothing deeper than a couple of meters, leading to the cross-sectional wedge shape). On the surface, forces of expansion deform the ground and elevate soil materials adjacent to growing ice wedges creating a pattern of raised ridges adjacent to either side of the polygonal ice wedge networks. At the same time some slumping of materials directly over the growing wedges forms a network of troughs. Such ice wedge growth leads to very common Arctic patterns of relief known as low-centered polygons (Fig. 8). These polygons are frequently up to 10 or more meters across with flat central areas surrounded by raised rims and separated from adjacent polygons by troughs that mark the location of the underlying ice wedges. Rims and troughs are frequently less than a meter wide each (but may exceed 2 m in width) and generally reflect the width of the underlying ice wedge. Centers of low-centered polygons are poorly drained due to the surrounding rim, and, as rims continue to grow, polygon centers may become ponds. Polygon troughs are generally wet but are generally integrated with patterns of regional drainage.

Erosion of low-center polygons typically generates another common landscape type called high-center polygons. Centers do not change much (if any) in elevation, but the surrounding troughs deepen, and rims collapse into the deepening troughs, leaving the polygon centers high and separated from each other only by enlarging deep troughs. The melting of ice wedges and the collapse of adjacent materials facilitate such erosion as melt water is drained away. Such features are frequently found adjacent to streams, lakes, and other areas where lateral drainage is

**Fig. 9** Thermokarst erosion where tundra meets the sea in northern Alaska, exposing soil ice and permafrost



augmented by local landscape relief. Other important Arctic surface patterns include frost medallions, stone circles, nets and stripes, frost mounds, pingos, palsas, and raised beaches.

Familiar patterns associated with wind and water are also found in the Arctic, riparian habitats, and the influence of rivers is similar in the Arctic to other ecosystems with the added impact associated with heat transport. Major rivers of the Arctic originate outside of the Arctic and flow north into the Arctic Ocean, bringing nutrients, sediments, and heat. Even smaller rivers entirely within the Arctic tend to flow north and carry heat. Riverbeds, valleys, and deltas exhibit deeper thaw depths than surrounding landscapes and may be free of permafrost. Cutbanks, meanders, oxbows, bars, islands, and braided channels produce familiar landscape patterns. Landscape depressions associated with rivers and streams frequently fill with snow in winter, protecting streamside (riparian) plant communities, often willows, allowing them to grow much higher than other tundra plants.

Areas of sand dunes are found within the Arctic, and wind-blown sand and loess have been even more widespread in the past, deepening sediments over broad areas and obscuring polygons.

Thermokarst is a form of erosion involving the collapse of surface materials associated with the thawing of soil ice (Fig. 9). The formation of high-center



polygons described above is a type of thermokarst erosion. Thaw lakes represent another Arctic thermokarst feature important in coastal plains and lowlands. Ice accumulation by ice wedge growth and other ice-incorporating mechanisms can result in the upper permafrost material being composed primarily of ice. When ice is a major constituent, such permafrost is said to be ice rich and prone to thermokarst. When such materials thaw, the surface soils are no longer supported and slump. Where drainage is restricted and water remains, a pond or lake may form. Since water absorbs heat better than vegetation (lower albedo), thermokarst ponds and lakes continue to grow through heating adjacent ice-rich permafrost. Such growth frequently ends when ponds or lakes either partially or completely drain themselves as their growth intersects a stream or other drainage pathway.

Plants colonize the exposed bottoms of the former lakes, and ice accretion may begin anew in the underlying sediments, leading eventually to renewed potential for thermokarst and lake formation. This cycle, known as the thaw lake cycle, may take thousands of years to complete but renews the surface and initiates plant succession. The thaw lake cycle plays an ecological role similar to that of fire in many other ecosystems. Some plants are better adapted to seed dispersal than others, and just as in ecosystems dominated by fire, it is these plants that tend to become pioneer species in the thaw lake succession. Other plants are better able to contend with the competition and they come to dominate over time. As is commonly the case with fire, the newly exposed surfaces tend to have a relatively high mineral nutrient availability that declines as succession proceeds. Plant succession in the Arctic is frequently associated with natural geomorphic disturbance. Succession tends to proceed from pioneers with good seed dispersal eventually to dominant plants that can tolerate competition for light or nutrients.

---

### **Cryoturbation, Needle Ice, and Other Soil Disturbance**

Among the indirect effects of cold presenting challenges to Arctic plants, mechanical pressure resulting from the expansion of freezing water is among the most apparent. Various expressions of such mechanical pressures include needle ice, frost boils, ice wedge polygons, ice lenses, sorted stone circles, palsas, pingos, and more. One of water's unique properties is that it is less dense in crystalline form than it is in liquid form, meaning that it expands as it freezes. The actual increase in volume is close to 8 % resulting in significant forces for displacement of soils. Germination of spores or seeds may be disturbed and recurring freeze/thaw activity in some areas such as frost boils may prevent plant colonization all together.

Colonization of bare mineral soils by plants is difficult due to the formation of needle ice, which is effective at moving seeds, seedlings, and even small stones, elevating them temporarily above the soil surface. Repeated movements of stones lead to netlike patterns of sorted stone or stone circles in sparsely vegetated areas. Finer sediments too can be moved by frost, and frost boils or frost medallions similarly prevent plant colonization. Disturbance of vegetated surfaces can lead to the initiation of frost scars, which persist through frost action, resisting vegetation reestablishment.

Even in areas of continuous plant cover, contraction cracks may disrupt plant growth, breaking rhizomes and limiting the vegetative spread of individual plants.

---

## Arctic Climate Change

The climate has noticeably changed over the past few decades throughout much of the Arctic, but the trends are not consistent in duration, intensity, or direction. The Arctic is susceptible to climatic oscillations and past variations in climate have occurred on many time scales. Climatic oscillations intrinsic to the Arctic are superimposed upon any longer-term trends that may be more global in nature. Concern over global climate change and the perception that the Arctic is both vulnerable and more likely to experience changes has led to efforts to predict the consequences of a changing climate on Arctic plants and ecosystems. While simple to state, this question is currently unanswerable. One might anticipate that warming in the Arctic would be a straightforward matter to understand, but there is no single Arctic ecosystem and responses are certain to be varied and heterogeneous. There are many feedbacks in Arctic ecosystems that can lead to counterintuitive results (Chapin et al. 1991). Plants do not experience the effects of warming independent of other environmental changes. Warming, for instance, changes precipitation patterns. Moisture and nutrient availability are not simple functions of temperature and precipitation but also depend upon depth of thaw and regional drainage. Snow accumulation insulates the soil from heat loss, but it also retards the onset of the growing season. Increasing vegetation density has a cooling effect upon soils that can cause permafrost to grow upward limiting rooting volume, etc. Not unlike the economy, with its many feedbacks and counterintuitive effects, it is very difficult to make credible predictions about the future state of Arctic systems.

The Arctic is a large area and contributes to the overall global carbon budget. Peat accumulations and carbon-rich soils of the tundra point to past carbon sequestration and the potential for positive or negative carbon exchange with the atmosphere. While only 5 % of the earth's terrestrial surface, the Arctic contains 14 % of its soil carbon. Most arctic landscapes currently appear to be very nearly carbon neutral, with soil moisture (anoxia) trumping temperature in those areas where net carbon accumulation may be occurring. Predicting the fate of Arctic soil carbon depends upon an ability to predict the future surficial hydrology of the tundra, a challenge even greater than predicting climate. Biogenic methane production is likewise controlled more by Arctic soil moisture conditions than by temperatures. The key question is not whether or not the Arctic is warming, but whether or not it is drying.

---

## Future Directions

The Arctic has long been considered a fragile ecosystem, subject to disturbance by development and climate change. Such concerns have led to continuing questions regarding the susceptibility of Arctic organisms and ecosystems to human

perturbations, and current research in the Arctic is commonly focused on answering such questions. Much has been learned, and continues to be learned, about the Arctic, and it continues to be at the forefront of understanding and integrating ecological knowledge with other relevant sciences necessary to achieve an understanding of the entire system. The Arctic represents a globally significant resource for the study of intact natural ecosystems, for despite development, the transformation of indigenous cultures, and environmental contaminants from outside the Arctic, much about the Arctic remains nearly pristine. This quality, along with a concomitant aesthetic value, compels our conservation interests, while the development of resources within the Arctic and the interests of local populations frequently reflect the economic values of Arctic landscapes. The question is not will the Arctic change, but how fast and to what degree and to what extent can science inform this process. This leads to a continuing need for research to address the practical problems of the management of natural resources and for basic field research in the Arctic to address theoretical issues of evolutionary and ecosystem biology. These basic and applied research efforts are not unrelated, as currently there is no general systems theory capable of providing a suitable framework for the kinds of questions that need to be addressed regarding development alternatives in the Arctic. That is to say assumptions such as the fragility of Arctic ecosystems have no actual basis in theory. Intact ecosystems of the Arctic represent a unique opportunity to develop more robust general theories of ecosystems.

The Arctic, while relatively pristine, is experiencing pollution, particularly persistent organic pollutants (POPs) which accumulate in the Arctic through a process of global distillation. The effects of these compounds in Arctic organisms and ecosystems, as well as of other potential pollutants, are a pressing research need.

Despite decades of research, much remains to be understood about the dynamics belowground in Arctic ecosystems. Continuing research challenges include seeking a better understanding of roots (production, distribution, and turnover), interactions between microbial communities and roots, seasonal belowground dynamics, and the physiochemical properties of soil solutions in the rhizosphere. Much has been learned regarding the carbon and nutrient dynamics of Arctic plants and ecosystems, but spatial extrapolation of this understanding and the validation of predictive models are incomplete. Studies continue on regional patterns of phenology, species migrations, dynamics of plant communities, and concerns over invasive species. Taxonomic revisions, and systematic studies using modern techniques, combined with palenology and geomorphology are exciting areas leading to new interpretations of the history and past dynamics of Arctic landscapes.

Indigenous peoples of the Arctic have known and used its plants and ecosystems for millennia. Involvement of indigenous peoples in continuing research, and integration of existing knowledge of the many cultures that inhabit the Arctic, is generally recognized as an important goal of the continuing development of Arctic science. There are great opportunities for current and future students of Arctic plant ecology, but new approaches and ways of inquiry involving Arctic residents need to go well beyond ethnobotany to meet this challenge.

## References

- Bliss LC. Arctic ecosystems of North America. In: Wielgolaski FE, editor. *Ecosystems of the world 3. Polar and alpine tundra*. New York: Elsevier; 1997. p. 551–683.
- Brown J, Miller PC, Tieszen LL, Bunnell FL. Arctic ecosystem, US/IBP synthesis series, vol. 12. Stroudsburg: Dowden, Hutchinson and Ross; 1980.
- Chapin III FS, Jefferies RL, Reynolds JF, Shaver GR, Svoboda J, Chu EW. *Arctic ecosystems in a changing climate: an ecophysiological perspective*. San Diego: Academic; 1991.
- Conservation of Arctic Flora and Fauna (CAFF). *Arctic flora and fauna: status and conservation*. Helsinki: Edita; 2001. p. 272.
- Crawford RMM. *Plants at the margin: ecological limits and climate change*. Cambridge: Cambridge University Press; 2008. [Hardcover].
- Hulten E. *Flora of Alaska and neighboring territories: a manual of the vascular plants*. Stanford: Stanford University Press; 1968.
- Nuttall M. *Encyclopedia of the Arctic*. 3rd ed. New York: Routledge; 2004.
- Polunin N. *Circumpolar Arctic flora*. Oxford: Oxford at the Clarendon Press; 1959.

## Further Reading

- Conservation of Arctic Flora and Fauna (CAFF). *Arctic flora and fauna: status and conservation*. Helsinki: Edita; 2001. p. 272. [http://www.caff.is/publications/view\\_document/167-arctic-flora-and-fauna-status-and-conservation](http://www.caff.is/publications/view_document/167-arctic-flora-and-fauna-status-and-conservation)
- Crawford RMM. *Plant survival in a warmer Arctic*. In: Crawford RMM, editor. *Plants at the margin: ecological limits and climate change*. Cambridge: Cambridge University Press; 2008.
- Lee JA. *Arctic plants: adaptations and environmental change*. In: Scholes JD, Barker MG, editors. *Physiological plant ecology (39th symposium of British Ecological Society)*. Oxford; Malden, MA, USA: Blackwell Science; 1999.
- Pielou EC. *A naturalist's guide to the Arctic*. Chicago: University of Chicago Press; 1995.

John Blair, Jesse Nippert, and John Briggs

## Contents

Introduction .....	390
General Characteristics and Global Distribution of Grasslands .....	392
Basic Biology and Ecology of Grasses .....	398
Morphology .....	398
Population Dynamics .....	401
Physiology .....	402
Roots .....	404
Grasslands, Drought, and Climate Change .....	406
Fire in Grasslands .....	408
Grazing in Grasslands .....	412
Potential Threats to Grassland Conservation .....	416
Grassland Restoration .....	418
Future Directions .....	420
References .....	421

---

## Abstract

- Grasslands are one of Earth's major biomes and the native vegetation of up to 40 % of Earth's terrestrial surface. Grasslands occur on every continent except Antarctica, are ecologically and economically important, and provide critical ecosystem goods and services at local, regional, and global scales.
- Grasslands are surprisingly diverse and difficult to define. Although grasses and other grasslike plants are the dominant vegetation in all grasslands, grasslands also include a diverse assemblage of other plant life forms that contribute to their species richness and diversity. Many grasslands also support a diverse animal community, including some of the most species-rich grazing food webs on the planet.

---

J. Blair (✉) • J. Nippert • J. Briggs  
Division of Biology, Kansas State University, Manhattan, KS, USA  
e-mail: [jblair@ksu.edu](mailto:jblair@ksu.edu); [nippert@ksu.edu](mailto:nippert@ksu.edu); [jbriggs1@ksu.edu](mailto:jbriggs1@ksu.edu)

- Grasslands allocate a large proportion of their biomass below ground, resulting in large root to shoot ratios. This pattern of biomass allocation coupled with slow decomposition and weathering rates leads to significant accumulations of soil organic matter and often highly fertile soils.
- Climate, fire, and grazing are three important drivers that affect the composition, structure, and functioning of grasslands. In addition to the independent effects of these factors, there are many interactions among grazing, fire, and climate that affect ecological patterns and processes in grasslands in ways that may differ from the independent effects of each driver alone.
- Grasslands occur under a broad range of climatic conditions, though water is generally limiting for some part of the year in most grasslands. Many grasslands experience periodic droughts and a dormant season based on seasonal dry or cold conditions.
- Grasslands are sensitive to climate variability and climate changes. There are well-documented shifts in the distribution of North American grasslands in response to past droughts, and both observational data and experiments suggest that grasslands will be affected by future changes in rainfall and temperature.
- Fire is a common occurrence, particularly in more mesic grasslands, due to the large accumulations of dry, highly combustible fine fuel in the form of dead plant material. Fire affects virtually all ecological processes in grasslands, from the physiology of individual plants to the landscape-level patterns, though the effects of fire vary with grassland productivity and the accumulation of detritus.
- All grasslands are grazed or have experienced grazing as a selective force at some point in their evolutionary history. The ecological effects of grazing vary with climate and plant productivity, and the associated evolutionary history of grazers in different grasslands.
- Grasslands have been heavily exploited by humans, and many temperate grasslands are now among the most threatened ecosystems globally. Widespread cultivation of grasslands was the major land-use change that impacted grasslands historically, while multiple global changes drivers (i.e., altered fire and grazing regimes, woody plant encroachment, elevated CO<sub>2</sub>, invasive species, fragmentation) contribute to the contemporary loss of grasslands.
- Grassland restoration aims to recover the diversity and ecosystem services that grasslands provide. While restored grasslands may attain productivity comparable to native grasslands and sequester carbon for extended periods, they typically support much less diversity than comparable native grasslands. Recovery of soil communities and properties is often very slow.

---

## Introduction

Grasslands and other grass- and graminoid-dominated habitats (e.g., savanna, open and closed shrubland, and tundra) occur on every continent except Antarctica (though some grasses do occur there) and occupy about 30–40 % of Earth's land surface. They cover more terrestrial area than any other single biome type.

The extent and diversity of grasslands and related habitats is reflected in their ecological and economic importance at local, regional, and global scales. For example, grasslands provide critical habitat for a diverse array of plants and animals. Grassland soils store tremendous quantities of carbon and other key nutrients and play a major role in global biogeochemical cycles. There is also a long and complex relationship between grasslands and humans. Modern humans are thought to have originated in the open grasslands and savannas of Africa, and grasslands have provided the template and biological raw material for the development of modern agriculture and associated human societies. The fertile soils that developed under many grasslands have been plowed and the nutrients mined to support agricultural production. Domesticated grasses, such as corn, rice, wheat, oats, and sorghum, have become some of our most important agricultural crops, and barley was used by Neolithic humans to produce one of the first known alcoholic drinks. Grasses are not only consumed directly by humans, but they also support the production of domestic livestock for human use. More recently, several species of grasses are being widely used or considered as feedstock for biofuel production (e.g., *Panicum virgatum*, *Miscanthus* spp.). It is estimated that as many as 800 million people worldwide rely directly on grasslands for their livelihoods (White et al. 2000), and virtually everyone uses grassland products (food, fiber, fuel) in their daily existence. In total, it is clear that grasses and grasslands have played an important role in the history of humans and will continue to do so in the future.

Grasslands have also played an important role in the development and testing of ecological theory, such as assessing relationships between species richness and ecosystem function and as model systems for assessing the impacts of global changes, including responses to chronic N deposition, elevated CO<sub>2</sub> concentrations, and climate change. This is due, in part, to the relative ease of performing manipulative experiments in grasslands, the sensitivity of grasslands to perturbations, and the relatively rapid responses they often exhibit to these manipulations. In fact one of the longest running field experiments in the world is the Park Grass Experiment at the Rothamsted Experimental Station in England. This experiment was established in 1856 with the original goal of assessing the effects of various nutrient amendments on grass yields. The experiment has since been used to address a broad range of fundamental questions in ecology and evolutionary biology (Silvertown et al. 2006).

Grasslands also include some of the most endangered ecosystems on the planet, such as the tallgrass prairies of North America and other temperate grasslands (Hoekstra et al. 2005). In addition to the historical loss of grasslands to agricultural expansion, grasslands today are threatened by a broad array of environmental changes, including climate change, elevated atmospheric carbon dioxide concentrations, increased nitrogen deposition, invasive species, habitat fragmentation, degradation due to overgrazing, change in natural disturbance regimes (e.g., fire suppression), and woody plant expansion. Conserving, and in some cases restoring, these ecosystems will require a solid foundation of ecological knowledge. This chapter focuses on the ecology of grassland ecosystems and provides the reader with an introduction to grassland plants and the major abiotic and biotic factors that

influence the structure and functioning of grassland ecosystems. Our goal is to present a sufficiently broad coverage to familiarize readers with the variation that exists in different grasslands from different parts of the globe, combined with more detailed information and specific examples of key ecological processes from a few well-studied grassland ecosystems, including the mesic tallgrass prairies of North America where the authors have extensive experience.

---

## General Characteristics and Global Distribution of Grasslands

A simple, all-encompassing definition of grasslands is surprisingly difficult to come by, and grasslands have been defined and distinguished from other biome types in many different ways. One defining feature of grasslands is that they are dominated or codominated by graminoid vegetation, including the true grasses (family Poaceae) and other grasslike plants including sedges (Cyperaceae) and rushes (Juncaceae). Defined narrowly, grasslands are ecosystems characterized by a relatively high cover of grasses and other graminoid vegetation in an open, often rolling, landscape with little or no cover of trees and shrubs. However, the term grassland can also be used in a broader sense to encompass ecosystems with a significant grass cover interspersed with varying degrees of woody vegetation, including relatively open savannas and woodlands (e.g., the cerrados of South America) and some deserts and shrub grasslands (also referred to as steppes) that include a significant cover of grasses interspersed with succulent plants and/or shrubs. In this context, grasslands can vary in the relative abundance of grasses and other plant life forms, such as trees and shrubs. In fact, the cover of woody vegetation is increasing in many grasslands globally, as discussed later in this chapter, and there is often disagreement about how to delimit grasslands from other vegetation types that include significant grass cover mixed with other herbaceous and/or woody vegetation.

Although grasses provide the matrix in which other plant species co-occur, grasslands include other plant life forms, such as annual and perennial forbs (non-graminoid, nonwoody plants), shrubs, and trees. The matrix-forming species in most of the world's major grasslands are perennial grasses that are relatively long-lived and that can reproduce either sexually or asexually via belowground meristematic tissue (belowground buds), though a few grasslands are dominated by annual species that must reproduce from seed each year (e.g., California and other annual grasslands). Some grasslands are dominated by grass species that produce individual tillers evenly distributed across the soil and often joined by underground stems called rhizomes (i.e., rhizomatous or "sod-forming" grasses), while other grasslands are dominated by species that produce densely packed clumps of tillers that are distinct from one another and often separated by bare soil spaces (i.e., caespitose or bunchgrasses; Fig. 1).

The graminoid flora of grasslands can be quite species rich (Fig. 2). For example, the Konza Prairie Biological Station (a tallgrass prairie research site in eastern Kansas, United States) supports more than 100 species of grasses and sedges.



**Fig. 1** Contrasting growth forms of grasses. The foreground is dominated by the caespitose grass *Bothriochloa bladhii*, an exotic species native to parts of Africa, Eurasia, and Australia. The more even cover of grasses in the background includes the rhizomatous native tallgrass prairie grasses *Andropogon gerardii* and *Sorghastrum nutans* (Photo by John Blair)

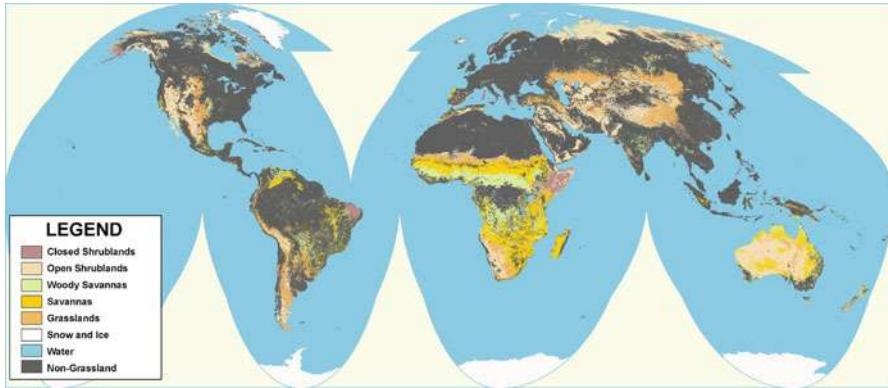


**Fig. 2** Although grasslands are often dominated by a small number of grass species, they often co-occur with a diverse assemblage of other grasses, as well as forbs and woody plant species. As a result, high floristic diversity is characteristic of many grasslands, such as the North American tallgrass prairie pictured here (Photo by Dan Whiting)



Yet this prairie, like most other grasslands, is dominated by just a few species of grass that comprise the majority of grass cover and contribute the bulk of annual plant productivity. For example, at Konza Prairie *Andropogon gerardii*, *Sorghastrum nutans*, and *Schizachyrium scoparium* comprise about 70 % of total plant cover and up to 90 % of the aboveground net primary productivity (ANPP), particularly in frequently burned and ungrazed areas. In fact, many grassland types are described by their dominant species (e.g., bluestem prairie). However, despite the general prevalence of graminoid plant cover, different types of grasslands are surprisingly diverse in the richness and cover of non-grass species. Using the Konza Prairie example, the grasses co-occur with over 400 species of forbs and woody plants, which provide much of the floristic diversity characteristic of the prairie.

The global distribution of grasslands is extensive, with widespread representation of grasslands on every continent except Antarctica (Fig. 3). Although grasslands are

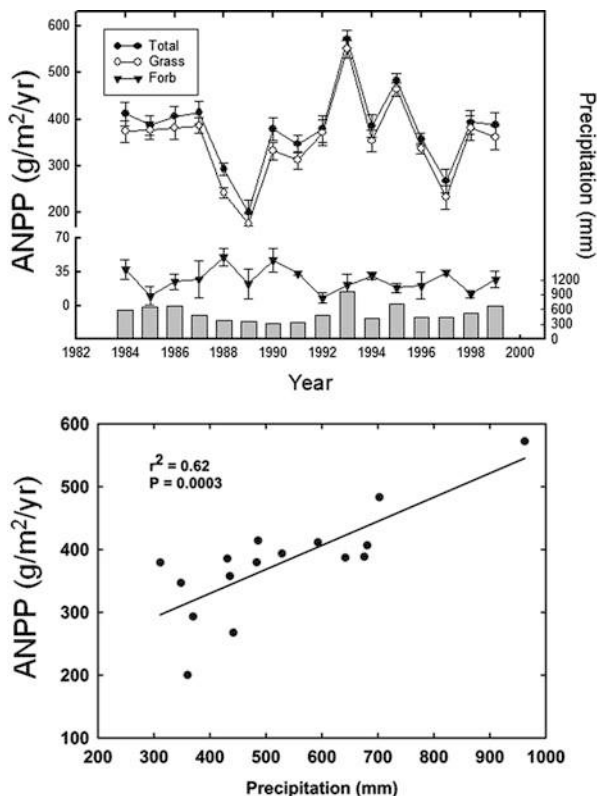


**Fig. 3** Global distribution of grasslands and other ecosystem types dominated by grasses or graminoid vegetation (Reproduced from White et al. 2000)

currently absent from Antarctica, a grass species (Antarctic hairgrass, *Deschampsia antarctica*) does occur on the Antarctic Peninsula and surrounding islands surrounding, where recent warming is thought to be promoting the spread of this native grass. Major grasslands in the temperate regions of the world include the steppes of Eurasia, the velds of southern Africa, the pampas of Argentina, and the prairies of North America (Archibold 1995). Grasslands and savannas also occur within the subtropics and tropics, such as the mesic grasslands of Florida, the bushvelds of Africa, and the campos and llanos of South America, and in areas with a Mediterranean climate (dry summers and relatively warm, wet winters). Grasslands can be found in coastal areas near sea level, and in montane regions at elevations up to 4,500 m (e.g., neotropical páramos and temperate montane meadows or parks). Intensively managed, human-planted, and maintained grasslands (e.g., pastures, lawns, etc.) occur worldwide as well, though these are not discussed further in this chapter.

As might be expected with such widespread distribution, grasslands occur under a very broad range of mean annual temperature and rainfall. The climates of grasslands vary from temperate to tropical with annual rainfall ranging from about 250 mm/year in arid grasslands to well over 1,000 mm/year in mesic grasslands. Mean annual temperatures vary from near 0 °C to around 26 °C. While there are many significant correlations between mean annual precipitation and the properties of grasslands, such as aboveground net primary productivity, rooting depth, and soil organic matter accumulations, these relationships are often more complex than they might first appear. Grasslands often experience very high intra- and interannual variability in rainfall, and comparisons with other biomes indicate that grasslands are more responsive to variation in rainfall amounts than are most other biomes (Fig. 4). This may occur because the relatively high density of plants and associated meristematic tissue (growing points) in grasslands results in greater growth potential when water is available, relative to more arid

**Fig. 4** *Top*: Long-term record of aboveground net primary productivity (ANPP) (mean  $\pm$  SE,  $n = 20$ ) for grasses (primarily  $C_4$  species) and forbs ( $C_3$  herbaceous plants) with corresponding growing season (April–Sept) precipitation amount in an annually burned mesic grassland in NE Kansas (Konza Prairie LTER site). *Bottom*: Positive relationship between grass ANPP and growing season precipitation (mm) based on the data in *top panel* (From Nippert et al. 2006)



ecosystems, and because wetter forests and woodlands are not as limited by water availability. These results suggest that grasslands may be especially sensitive to changes in precipitation amounts or timing in an altered future climate. Seasonality of precipitation, in addition to total annual amount, is also critical in grasslands. For example, in North America the area around Washington, DC, is dominated by eastern deciduous forest, and the annual precipitation is  $\sim 102$  cm, which is very similar to the annual precipitation amount ( $\sim 100$  cm) near Lawrence, KS, which is dominated historically by tallgrass prairie. In spite of similarities in total rainfall amount, the seasonal distribution of rainfall is very different with over 60 % of the rainfall occurring in the growing season (April to September) and with drier late summer months in Lawrence, KS, whereas the precipitation is more evenly distributed throughout the year in Washington, DC. The importance of seasonal patterns of rainfall in grasslands is apparent in the numerous studies that have used climatic data and concurrent measurements of ecological processes to identify specific times of the year (critical climate windows) when precipitation has the greatest effect on processes such as plant productivity or grass reproductive effort. There are also significant interactions between rainfall amounts and temperature, and the ratio of precipitation to the potential evapotranspiration (PET) is often a better predictor of

ecological properties and process rates than is mean annual precipitation alone. Of course, the ability of soils to hold and supply water is also critical, and soil water dynamics are affected not only by rainfall quantity and intensity but also by physical characteristics of the soil, such as soil texture and porosity. At local scales, soil water dynamics in grasslands are often highly correlated with plant physiological processes, plant productivity, and soil microbial activity.

Climatically determined grasslands are those that result from prevailing climatic conditions, as opposed to planted grasslands (pastures or lawns) or those that represent intermediate successional stages. A characteristic feature of climatically determined grasslands is that they are subject to periodic droughts, which contributes to the accumulation of highly flammable plant detritus and the occurrence of periodic fires. Many of the world's most extensive grasslands occur in the interior regions of the continents, where annual rainfall amounts are relatively low and irregularly distributed across the year. Some of these grasslands lie between more arid deserts and more mesic forests and woodlands, while others occur in the rain shadows of major mountain ranges. The continental climates of these regions are often marked by extremes in seasonal temperatures (hot summers and cold winters), to which the plants and animals living there are adapted. For example, at Konza Prairie in the Central United States, the mean monthly temperature varies from a January low of  $-3^{\circ}\text{C}$  to a July high of  $27^{\circ}\text{C}$ . In temperate grasslands with such continental climates, a significant proportion of annual rainfall often coincides with the warm growing season, and plant dormancy is a mechanism for surviving low winter temperatures. Many grassland animals also become dormant or migrate to avoid harsh winter conditions. In grasslands with a Mediterranean climate, such as those in the Central Valley of California, dormancy is driven by summer droughts, and the growing season coincides with seasonal rainfall that occurs in the relatively warm winter months. Tropical grasslands also exhibit distinct seasonality based on cyclic annual rainfall patterns, though annual temperatures vary less than in temperate grasslands. Dormancy still occurs, but in this case it is a response to annual dry seasons that alternate with the rainy growing season as a result of annual movement of tropical low pressure system boundaries. Soils of tropical grasslands may also be less fertile than comparable temperate grassland soils as a result of faster weathering rates under warm year-round temperatures and soils that are much often much older than in temperate grasslands. Many tropical grasslands also have a greater density of woody shrubs and trees than do temperate grasslands.

Although many climatically determined grasslands experience seasonal water deficits and periodic droughts that preclude the establishment of forests in those regions, some mesic grasslands, such as the tallgrass prairies of North America or the sourvelds of South Africa, occur in regions where the climate could support woodland, shrubland, savanna, or even forest vegetation. In these cases, the persistence of grasslands often depends on recurring disturbances, such as fire and grazing. Such grasslands may be best thought of as disturbance-dependent communities, where periodic fires, droughts, and the activities of grazers are necessary to keep grasslands from transitioning to other ecosystem types.

In fact, it is generally recognized that climate, fire, and grazing are three key factors that are responsible for the origin, maintenance, and structure of the most extensive natural grasslands on Earth. Although the relative importance of fire in structuring grassland communities tends to be greatest in the most mesic and productive grasslands, which also burn at more frequent intervals and with greater fire intensities do to large accumulations of fine fuel in the form of aboveground grass litter, fires do occur at varying frequencies in most grasslands, including shortgrass steppe and even desert grasslands. In addition, most grasslands coevolved with large grazers, and herbivory is an important process affecting ecological processes at levels ranging from the physiology of individual plants through population and community dynamics to ecosystem processes and landscape patterns. Although there are some similarities with respect to the effects of fire and grazing (i.e., both can be considered disturbances that remove aboveground plant biomass and free up resources), there are importance differences in their effects on soil resources and plant communities, as well as some important interactions between fire and grazing in grasslands. The effects of fire and grazing, and their interactions, are discussed in more detail in later sections of this chapter.

A final characteristic feature of grasslands is a relatively high allocation of plant biomass belowground (a high root to shoot ratio) and proportionally large inputs of plant root litter relative to surface litter. Relatively high belowground plant inputs coupled with relatively slow decomposition rates due to periods of water limitation can lead to large accumulations of organic matter and nutrients in the soil. In addition, the limited rainfall characteristic of most grasslands reduces the rate of weathering and leaching of critical plant nutrients from the rooting zone of grassland soils. The resulting high fertility of grasslands soils is one of the reasons they have been so widely exploited for agricultural purposes. The accumulation of soil organic matter is generally positively correlated with water availability, which stimulates plant productivity more so than decomposition, such that the most productive grasslands also tend to store the most organic matter and nutrients in the soil. Although grasslands can occur on a variety of different soil types, the archetypal dark, rich soils characteristic of many grasslands are known as Mollisols in the US Soil Taxonomy system or as a Chernozem in the World Reference Base for Soil Resources. These are the dark, rich soils that formed under the prairie of North America and the steppes of Europe and that have now largely been cultivated for use in agricultural production. Grasslands can also occur on other soil types, too. Many tropical and subtropical grasslands occur on soils that are geologically much older and therefore more highly weathered than most temperate grassland soils. These soils may be more depleted in cations and have lower phosphorus availability than temperate grassland soils. One unique association between soils and grasslands are the serpentine grasslands. Serpentine soils have a unique chemical composition due to the type of parent material from which they formed. Serpentine soils generally have high levels of magnesium and other metals and low concentrations of calcium. The flora growing on these soils is often very different from surrounding soils growing on more typical soils. In many cases, serpentine grasslands include species that are uncommon in other habitats.

## Basic Biology and Ecology of Grasses

Grasslands are species-rich ecosystems with a variety of life forms including annual, biennial, and perennial plant species. The defining plant species are the grasses, but these ecosystems also contain a diverse assemblage of other plant types, including forbs (herbaceous non-grasses), sedges, wetland plants, and woody plants (shrubs and trees). The high rates and amount of growth by grasses in grasslands may be attributable to their unique morphology and physiology. As noted earlier in this chapter, many grasslands are “disturbance-rich” ecosystems, existing in locations that typically experience frequent, wide swings in weather (daily, weekly, monthly), a variable climate over longer periods of time (periodic extended droughts), and forces like fire and the activities of large grazers that alter the landscape. Grasses have adapted to these forces over evolutionary time, and their unique morphology, developmental patterns, and physiology make them well suited to the grassland environment.

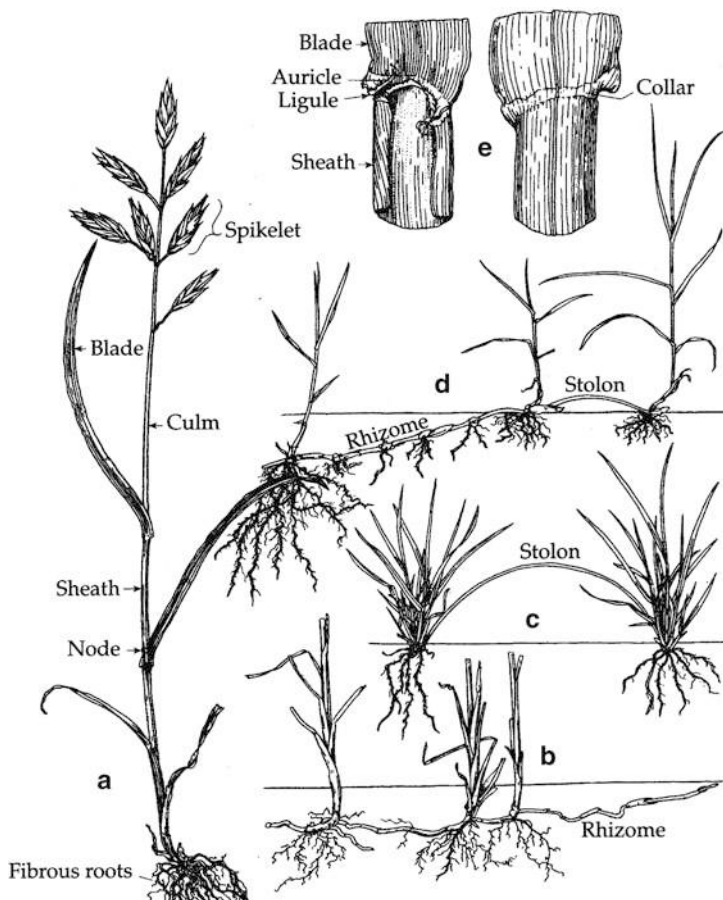
### Morphology

The aboveground portion of grasses is organized into tillers – individual shoots growing from the base of the plant. Tillers may be vegetative or reproductive and consist of one or more repeating units called phytomers, which are the basic building blocks of all grass shoots. Each phytomer consists of a node and internode with an axillary bud, cylindrical sheath, and leaf blade (Fig. 5).

Tillers are initiated from undifferentiated cellular tissue (meristematic tissue) that typically exists just beneath the soil surface. This is an important feature in an environment that includes periodic disturbances that remove tissues above the soil surface (i.e., fire and grazing). Additional meristematic tissue in grasses is also located at the intersections where leaves attach to the tiller (intercalary meristems). Thus, the oldest portion of a grass leaf is at the tip of the leaf and the top of the plant, and the youngest portion of a leaf is nearest the stem or the soil surface. For this reason, when grass blades are eaten, the actively growing plant tissues (intercalary or basal meristems) are left to produce new growth to replace removed leaf tissue. The presence of protected meristematic tissue belowground also allows grasses to survive and regrow when grazed or when fire removes aboveground tissues. This is an important mechanism giving grasses an advantage in environments with recurring droughts and fires or high grazing pressure (Fig. 6).

An individual grass plant generally consists of multiple joined tillers, but different grass species show great variation in the way tillers are aggregated as they expand from their origin. Two general classifications of tiller aggregation apply to most grasses: bunch-forming (caespitose or tussock) forms that are common in more arid grasslands and sod-forming (rhizomatous) grasses found more commonly in mesic grasslands (see Fig. 1). Sod-forming grasses utilize stolons (aboveground stems running along the ground surface) or rhizomes (belowground stems that occur just beneath the soil surface) to expand laterally through the

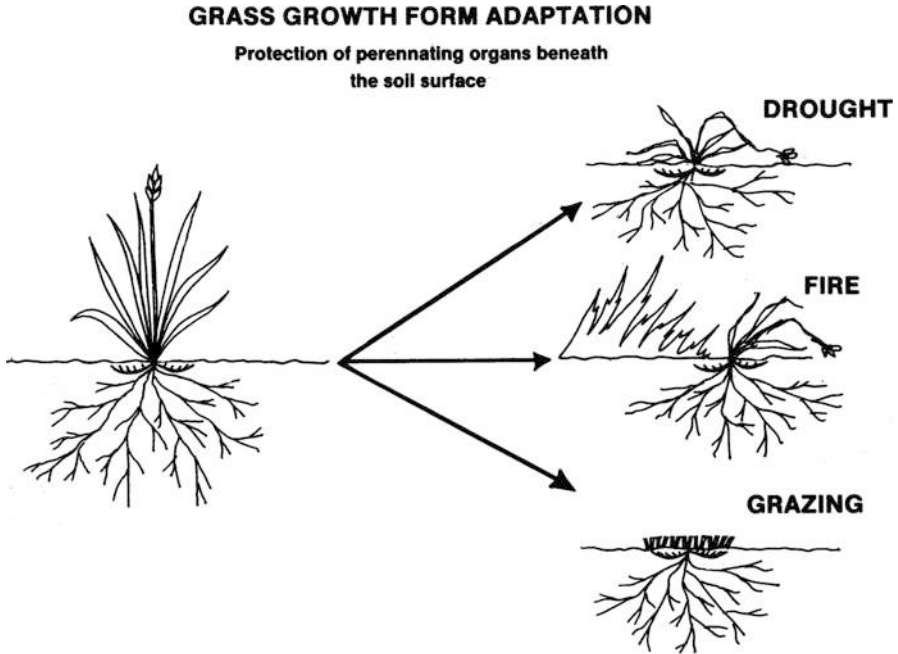




**Fig. 5** Structure of the grass plant: (a) General habit (*Bromus unioloides*); (b) rhizomes; (c) stolon; (d) rhizome and stolon intergradations (*Cynodon dactylon*); and (e) the leaf at the junction of sheath and blade, showing adaxial surface (left) and abaxial surface (right) (Reproduced from *Common Texas Grasses. An Illustrated Guide* by F. W. Gould by permission of the Texas A&M University Press)

asexual production of new tillers (see Fig. 5). Bunch-forming grasses cluster the production of new tillers around a central stem without rhizome or stolon production. Annual plants and the bamboos are obvious exceptions to these two tiller classification schemes, as annual plants complete their life history within a single growing season, and bamboos can produce very large wood-like stems.

Grass leaves are narrow, parallel veined, and characterized by thick-walled cells that provide rigidity and support that allows them to remain upright despite environmental (i.e., wind) or biotic (trampling) forces. Grasses also have specialized cells (bulliform cells) that permit leaf rolling during periods of water deficit or high-light stress, and some species have specialized tissues with air channels

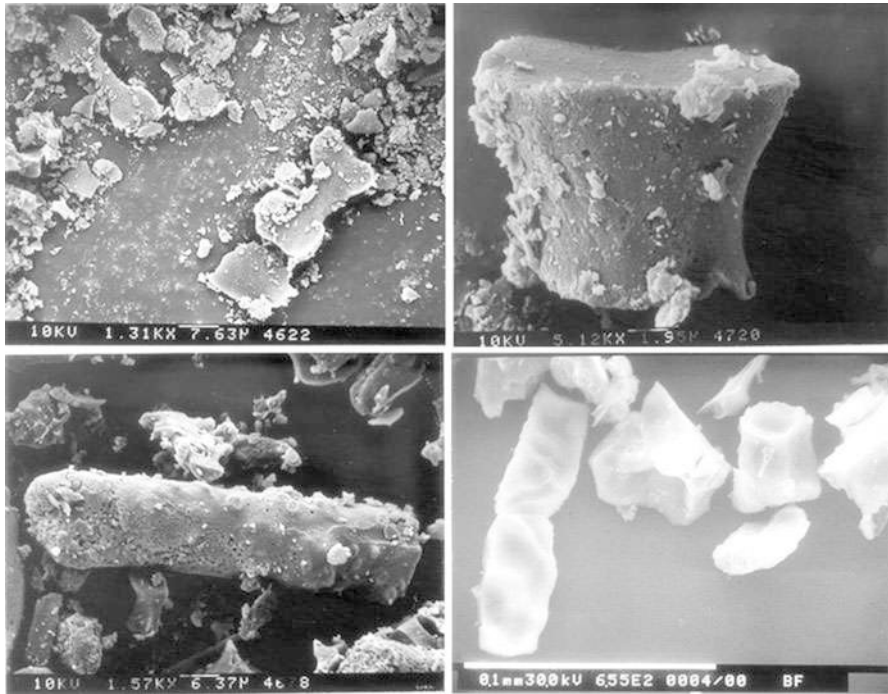


**Fig. 6** Belowground location of perennial meristematic tissue contributes to ability of grasses to survive and regrow following loss of aboveground biomass (From Anderson 1990)

(aerenchyma) that facilitate growth in water-logged soils. Another feature of grass leaves is the presence of biogenic deposits of silica in structures known as phytoliths, which provides structural rigidity and contributes to defense against herbivores. The physical structure of a phytolith is typically distinct within a species or taxonomic group (Fig. 7), and phytoliths recovered from soils and buried sediments have been used to determine the historic presence of grasses and to reconstruct past plant communities. Phytoliths breakdown slowly, allowing them to persist in the soil for long periods of time. For this reason, phytoliths are a useful paleo-ecological tool for assessing changes in grassland species composition over centuries and millennia.

Because biogenic silica produced by grasses may weather at rates different from soil silica pools, the presence of large amounts of biogenic silica in soils can alter weathering rates (Blecker et al. 2006). In addition to its role in structural rigidity of plant parts, silica deposits within grass tissues wear down an herbivore's teeth over the lifetime of the animal. It is now generally accepted that the evolution of abrasion-resistant teeth in many modern grazing animals was an evolutionary response to tooth-wearing effects of a diet high in grass. This also suggests that the grasses and their megaherbivore grazers are highly coevolved. In fact, grass phytoliths have been found in fossilized dinosaur dung from the Late Cretaceous (65–70 MYA), indicating that a long evolutionary relationship of grasses and their herbivores (Prasad et al. 2005).





**Fig. 7** Scanning electron micrographs of phytoliths. *Upper left*, *Andropogon gerardii*; *Upper right*, *Bouteloua gracilis*; *Lower left*, *Festuca* sp.; *Lower right*, *Stipa comata* (Photos from E.F. Kelly)

## Population Dynamics

Population dynamics of grassland plants are the product of the demography of the species living there, including life-history traits such as reproductive effort, germination and survivorship, and patterns of mortality. Temperate grasslands can be divided into two main types based on the life-history characteristics of the dominant grass species – the annual grasslands (i.e., California grasslands) and the perennial grasslands (i.e., tallgrass prairie). All grasses are flowering plants (Angiosperms) and nearly all are wind pollinated with a (relatively) simplified floral structure. Within the annual grasslands, recruitment of new individuals from year to year is based exclusively on sexual reproduction and germination of seeds by annual (i.e., monocarpic) grass species. Seed production and viability are critical parameters affecting population dynamics, and the soil seed bank is an important reservoir of new individuals. Annual grass species vary in the longevity of seeds in the soil seed bank, germination cues, rates of growth, and generation time. In contrast, recruitment of new individuals and population dynamics of perennial grasses are influenced much less by sexual reproduction and seed dynamics (production, viability, germination, and growth), but rather are a product of asexual

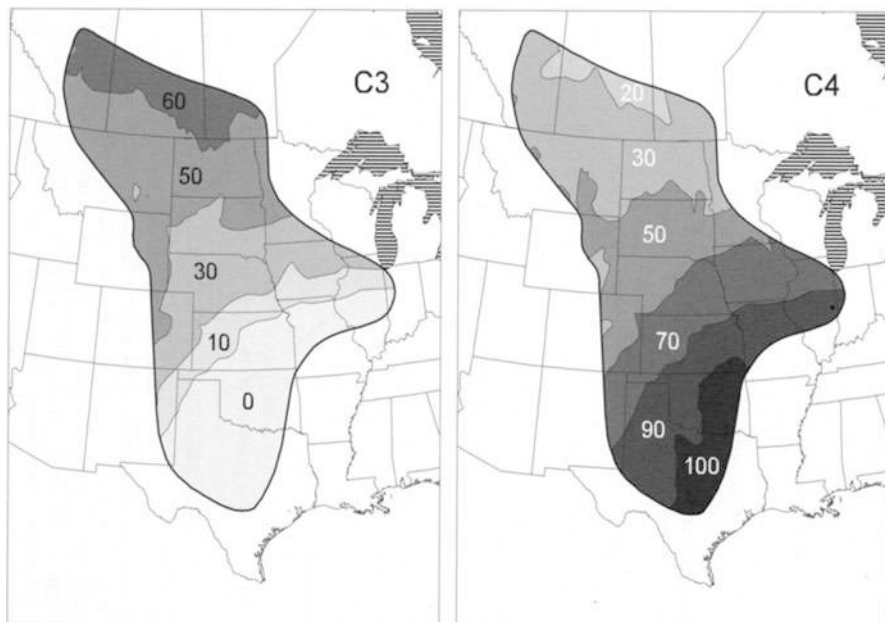
reproduction, and the recruitment of new “individuals” (really new tillers) is via clonal stems from existing tillers (Benson and Hartnett 2006). For these perennial grass species, rhizomes and associated belowground buds are the primary means of reproduction, and recruitment of individuals from seeds tends to be very low, except under specific circumstances such as large soil disturbances. Belowground “bud banks” in perennial grass species can be very responsive to changing environmental conditions or to disturbances such as fire and grazing, and this may be an important mechanisms underpinning spatial and temporal variability in the population dynamics and productivity of grasses (Dalgleish and Hartnett 2009).

## Physiology

In addition to the morphological adaptations outlined above, grasses possess a suite of physiological traits that facilitate growth in environments that experience periodic or episodic drought, high light intensity, extremes in temperature, and pulses in nutrient availability. One of the most fundamental physiological characteristics of different grass species is the type of photosynthetic pathway used, and this is another way to distinguish between major grassland types. Throughout the world today, tropical, subtropical, arid, semiarid, and warm temperate grasslands are typically dominated by grasses that use a  $C_4$  photosynthetic pathway (warm-season grasses), while grasses using the  $C_3$  photosynthetic pathway (cool-season grasses) are more common in cooler grasslands at higher latitudes or higher elevations.

Most vascular plants (and ~50 % of all grass species) use the  $C_3$  photosynthetic pathway.  $C_3$  photosynthesis occurs in leaf mesophyll cells where the enzyme Rubisco catalyzes a reaction fixing a low-energy carbon source (atmospheric  $CO_2$ ) to a five-carbon sugar (ribulose biphosphate), to form two molecules of a higher energy three-carbon organic acid (3-phosphoglycerate). With energy derived from the light reactions of photosynthesis, 3-phosphoglycerate is ultimately reduced to a single six-carbon sugar (glucose) that forms the metabolic template for all subsequent anabolic pathways in the plant. However, Rubisco is a nonspecific catalyst and can also catalyze the reaction of  $O_2$  with the five-carbon backbone, ultimately resulting in a net loss of energy to regenerate ribulose biphosphate (a process termed photorespiration, which results in a net loss of fixed carbon). The affinity by Rubisco for  $O_2$  over  $CO_2$ , and therefore photorespiration, increases at higher temperatures and during geologic periods with low atmospheric  $CO_2$  concentrations. These selective pressures are likely to have driven the evolution of the  $C_4$  photosynthetic pathway.

$C_4$  photosynthesis is a more recent physiological and morphological modification of the  $C_3$  pathway, having evolved over 50 different times and in many locations on Earth (Strömberg 2011).  $C_4$  photosynthesis provides a growth rate advantage in the high-light and high temperature environments typical of many grassland regions worldwide. In  $C_4$  photosynthesis,  $CO_2$  is initially captured by the enzyme phosphoenolpyruvate carboxylase (PEP-C) in leaf mesophyll cells to form a four-carbon acid (oxaloacetate). Oxaloacetate is transported into specialized morphological tissues



**Fig. 8** Grasses with the  $C_4$  photosynthetic pathway are more abundant in warmer grasslands of central US grasslands, whereas  $C_3$  grasses show the opposite pattern. Similar patterns occur on other continents, indicating that differences in biochemical pathways of C fixation play a strong ecological role in the distribution and success of grasses (From Lauenroth et al. 1999)

named bundle sheath cells that typically surround the leaf conductive tissue. Once in the bundle sheath, oxaloacetate is decarboxylated, releasing  $CO_2$  for Rubisco to fix and sugars to be formed using the  $C_3$  photosynthetic pathway. The primary benefit of the  $C_4$  photosynthetic pathway is the ability to concentrate  $CO_2$  within the bundle sheath essentially eliminating the likelihood of photorespiration and maximizing the reaction kinetics of carboxylation by Rubisco. As such, the efficiency of energy capture and conversion into carbohydrates is maximized, and efficient photosynthesis can be performed in environmental conditions that otherwise would have high photorespiration (i.e., dry, hot, high-light environments). The advantage of  $C_4$  grasses in warmer climates is evident in the proportions of  $C_4$  versus  $C_3$  grass species across latitudinal gradients (Fig. 8).

The  $C_4$  photosynthetic pathway has multiple secondary benefits for the grass species that use this pathway.  $C_4$  photosynthesis results in a higher instantaneous water use efficiency (ratio of  $CO_2$  gained to water lost) because PEP-C has a higher affinity for  $CO_2$  than does Rubisco. This allows grasses using the  $C_4$  pathway more flexibility in regulating stomatal openings to reduce water vapor lost from the leaves via transpiration while maintaining adequate internal  $CO_2$  concentrations for photosynthesis as soils dry down, relative to  $C_3$  grasses. The high affinity of PEP-C for  $CO_2$  also allows  $C_4$  plants to photosynthesize at higher levels than

C<sub>3</sub> plants when atmospheric CO<sub>2</sub> concentrations are low. As a result, it has been hypothesized that the C<sub>4</sub> photosynthetic pathway may have evolved in response to declining atmospheric CO<sub>2</sub> concentrations during glaciation events of the Earth's history. Finally, because the efficiency of Rubisco is maximized in the high CO<sub>2</sub> environment inside the bundle sheath, less total Rubisco is required to maintain a given rate of carbon assimilation compared to C<sub>3</sub> photosynthesis. For this reason, the photosynthetic nitrogen use efficiency (PNUE) (ratio of C gained per unit N mass) is higher in C<sub>4</sub> plants, allowing for greater productivity in N-limited environments, including many temperate and tropical grasslands.

## Roots

As noted previously, most grasslands are characterized by a large investment in root biomass and a high root to shoot ratio (Fig. 9). However, the root systems of different grasslands are highly variable in terms of species-specific patterns, total biomass invested, types of roots produced, and distribution throughout the soil profile. Many grass species share similar characteristics – fine roots that are highly branched, fibrous in nature, and concentrated in the upper soil profile (top 20–50 cm).

In contrast, the coexisting woody and herbaceous forb species in grasslands have root types that vary widely in terms of root types (fibrous, taproots, etc.), root depth distribution, and root to shoot biomass allocation. For this reason, most of our ability to generalize on the drivers of root structure and function in grasslands has been focused on the grasses. However, it is important to note that differences in rooting systems between the grasses and many forbs and woody plants may allow for differential use of soil resources, such as water and nutrients, and these differences can contribute to coexistence of different life forms in grasslands, as well as changes in the relative abundance of grasses and other plant life forms under changing environmental conditions. This concept of niche differentiation among grasses and woody plants was first described by Heinrich Walter and is known as “Walter’s two-layer hypothesis” (Walter 1971). This hypothesis was originally intended only for the semiarid savannas of the Southern Hemisphere, but the main concepts tend to apply to grasslands worldwide; grasses have a relatively fixed strategy of water uptake focused on surface soils, while woody plants have more plastic water uptake strategies and typically use considerably more water from deeper soil depths compared to grasses (Nippert and Knapp 2007).

The amount of root biomass varies markedly among grass species in different grassland types (mesic – semiarid – annual grasslands) as well as within a single site according to interannual variability in climate, topography, soil type, site management (fire and grazing frequency), and by depth in the soil profile. For many grassland types, the dominant grass species have very high root to shoot ratios (>3) illustrating a greater allocation of carbon to growth belowground versus aboveground. While nearly all grasslands are characterized by relatively large investments in belowground versus aboveground growth, this is typically greatest

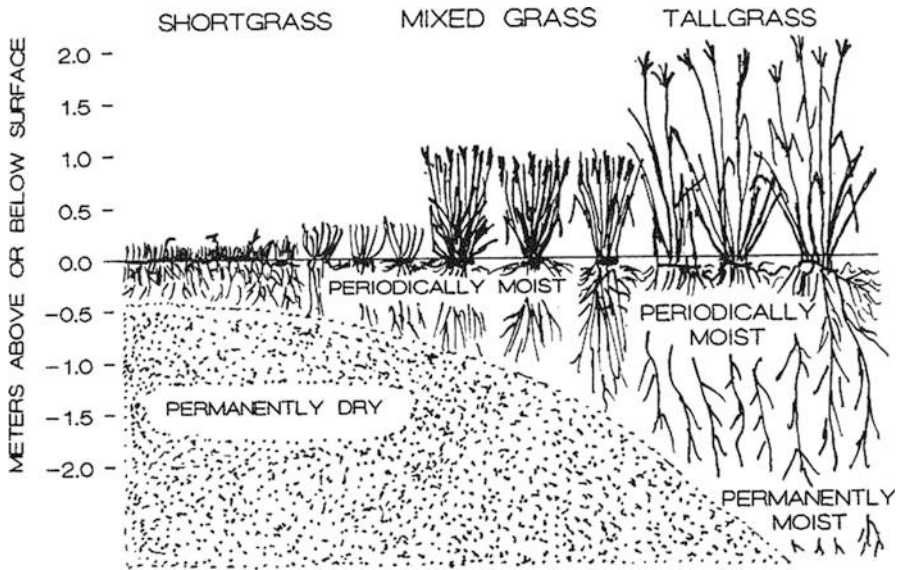
**Fig. 9** Exposed root biomass along an eroded stream bank at the Konza Prairie Biological Station (Photo by Jesse Nippert)



in grasslands with high water or nutrient limitation. In general, dry years (or adverse environmental conditions) tend to reduce overall grass growth including a reduction in root production. However, adverse environmental years tend to reduce the growth of shoots more than the growth of roots in most grasslands, though studies in the montane grasslands of Yellowstone National Park suggest that roots may be more sensitive to drought than shoots in some grasslands (Frank 2007). Changes in root production in response to disturbance tend to be mixed, varying according to ecosystem type and disturbance legacies. In tallgrass prairies that have been grazed or recently burned, root production can decrease by ~25 %, as grasses tend to allocate growth towards new leaf and stem production aboveground. The greatest reduction in root biomass production in these scenarios is in the uppermost soil layers (top 10 cm). In some other grasslands, increases in root turnover in the presence of grazers have been reported.

In addition to high relative belowground biomass (around 700–1,000 g m<sup>-2</sup> in mesic grasslands), the roots of many grasses extend deep into the soil profile (>2 m deep in mesic grasslands such as tallgrass prairie). Most grasses do not possess a tap root, but rather have long fibrous roots that taper with depth. The average depth distribution of roots in grasslands is generally correlated with mean annual precipitation and the depth distribution of water in the soil profile. Thus, the roots of grasses in arid grassland are much shallower than those in mesic grasslands (Fig. 10). Despite the presence of deep roots in some grasslands, the distribution of root biomass generally declines with soil depth, and majority of the biomass and total root length is concentrated in the upper soils.

The presence of grass roots at significant depths within the soil led early grassland ecologists to hypothesize that these roots served as a mechanism for drought avoidance. This hypothesis presumed that during periods of drought, deep roots would facilitate water uptake from deep soil zones recharged by infiltration from winter precipitation and maintain plant growth despite low water availability in surface soils. A closer examination of the unique physiology and morphology of



**Fig. 10** Regional gradients in rainfall affects the distributions of major grassland types as well as mean root depth and root productivity, which in turn affect soil organic matter storage and other soil properties and processes (From Seastedt 1995)

grass roots has shown that drought tolerance is a more likely strategy used by many grass species (Nippert et al. 2012). For example, in soils with very low soil moisture, grasses can maintain carbon uptake despite tremendous negative physical pressures within the vascular tissues of the roots, stems, and leaves (up to  $-14$  MPa, or nearly 58 times the pressure of automobile tires!). The ability to withstand these pressures without collapse is facilitated by vascular tissues with a greater number of vessels each with a smaller diameter. Thus, while many grasses can be deeply rooted, the small vessel number and diameter limits the total amount of water that can be transported from deeper soil depths, compared to the high root biomass and total root length present in surface soils. The unique physiology, morphology, and distribution within grassland soils provide a significant advantage for grass roots compared to forbs and woody plants to tolerate long periods of low water availability during drought.

## Grasslands, Drought, and Climate Change

Despite the adaption of many grassland species to periodic water deficits, grasslands are sensitive to both short-term climatic variability (e.g., variability in rainfall patterns within and between years) and longer-lasting changes in climate (e.g., multiyear droughts or directional changes in prevailing climate). One of the most well-documented grassland responses to severe drought comes from the Central



Plains region of North America in the early twentieth century. The early 1930s marked the beginning of a series of successive droughts that resulted in very little rainfall over much of the Central Plains and extreme reductions in soil moisture in the top meter of soil. This period, known as the Great Drought, was characterized by low precipitation (persistent reduction by ~50 % than average), higher wind speeds, low humidity, and maximum air temperatures that were ca. 5–6 °C above average maximum values during the summer months (Weaver 1968). The combination of extended severe drought conditions and widespread unsustainable agricultural practices led to the Dust Bowl and the widespread loss of top soil throughout much of the southern and central Great Plains. Prior to the Great Drought, Prof. John E. Weaver at the University of Nebraska-Lincoln spent 5 years surveying the community composition of 60,000 sq. miles throughout the central Great Plains (Weaver and Fitzpatrick 1934). This survey provided the basis for assessment of changes imposed by the continued drought later in the decade, and Weaver provided the most detailed assessment of the role of drought on grassland community structure ever performed.

Initially, the first stages of the drought (1930–1931) resulted in little change in grassland community composition (Weaver 1968). However, as the drought continued from 1934 to 1940, it had profound consequences for grassland productivity and community composition. In the eastern areas dominated by tallgrass prairie, the initial and most dramatic response to the drought was the desiccation and widespread mortality of the dominant species, primarily big bluestem, *Andropogon gerardii* (then classified as *Andropogon furcatus*); little bluestem, *Schizachyrium scoparium* (then classified as *Andropogon scoparius*); Indian grass, *Sorghastrum nutans*; and Kentucky bluegrass, *Poa pratensis* (Weaver and Albertson 1939). The loss of cover of the dominant species resulted in the exposure of much bare ground (estimates range from 36 % to 100 % reductions in basal area of plant cover in the permanent quadrats studied by Weaver (1968)). The drought eventually impacted the entire grassland community, with high rates of mortality for forbs, woody species, and ruderal species. An increase in cover was reported by those species adapted to drier grasslands to the west (mixed-grass and shortgrass prairie – including western wheatgrass, *Agropyron smithii*; side-oats grama, *Bouteloua curtipendula*; and needlegrass, *Stipa spartea*). Changes in the relative cover of species (from tallgrass to shortgrass prairie species) did not occur by immigration of individuals or seeds, but rather by changes in cover of species that were present, but less abundant (<1 % of cover), prior to the drought (Weaver and Albertson 1939). In all, the replacement of “true prairie” (i.e., tallgrass prairie) by mixed-grass and shortgrass prairie species occurred over an extensive range (~150 mile wide band) and within a period of 7 years. While community replacement did occur (from bluestems to xeric species), large reductions in basal cover (>50 %) persisted. The dramatic changes recorded during the Great Drought are best expressed by Weaver (1944, pp. 128–129):

The drought has shown clearly that nature has richly endowed True Prairie with many species, some of which are best adapted to cover the soil, enrich it, and hold it against the forces of erosion during moist climatic cycles. Others which are then found in such small

amounts that they seem almost a non-essential part of grassland rapidly increase to great abundance and become of great importance when a severe drought cycle occurs. This is what happened in the 1934–1940 drought and must have occurred many times in the historical and geological past, although no written record has been made.

Once the long period of drought ended, bare ground was colonized by ruderal (i.e., early successional) species common to disturbance (Weaver 1944). Stands of western wheatgrass, needlegrass, and buffalo grass (*Buchloe dactyloides*) that had increased during the drought remained resistant to immediate invasion for the first few years after drought (although species composition and cover ultimately returned to pre-drought conditions in the decades to follow). In regions where the bluestem cover was reduced, but not lost altogether, recovery to pre-drought abundance occurred within several years via rhizome extension into bare patches. Finally, for many of the original dominant perennial grasses (bluestems) as well as the forb species, recovery occurred via dormant rhizomes, root crowns, bulbs, and corms that persisted in the soil for the duration of the drought (without production of aboveground stems or leaves). Originally classified as “dead” years before, these individuals reinitiated growth 2–3 years following the drought from their decade of belowground “dormancy” [term used by Weaver – 1944]. Thus, the recovery of the tallgrass prairie was spatially and temporally varied – with quick recovery (~years) in locations where species persisted at low abundance but slow recovery (~decades) in locations where bare patches allowed the development of new grassland communities or replacement by mixed-grass or xeric prairie species.

The responses of grasslands to historic droughts may provide some insights into possible responses to future climate changes. Many climate change predictions for regions currently occupied by grasslands include more extreme weather patterns and increased temperatures, which may combine to reduce soil water availability and increase plant stress. Past responses to drought suggest that such climate changes may result in mortality and reduced cover of species adapted to wetter climates and possible replacement of those species with other adapted to drier conditions. Such changes in climatic conditions and species distributions would also be accompanied by changes in a suite of ecological processes, such as primary productivity, decomposition, nutrient cycling, soil formation, and species interactions. The degree to which species distributions and community boundaries shift in under a future climate may depend on the rate at which climate changes occur, the severity of those changes, and whether those changes are transient or represent a more permanent shift in prevailing climates.

---

## Fire in Grasslands

Grasses produce shoots that when senescent or dormant leave behind fine combustible fuel in the form of surface plant litter (detritus) and standing dead grass biomass. The accumulation of highly flammable plant litter coupled with periods of drought, relatively open landscapes, and windy conditions is highly conducive to large-scale fires (Fig. 11). As a result, fire is (or was) an important force in many



**Fig. 11** Although fire can be a destructive force in some ecosystems, fire is an important natural disturbance in many grasslands. Historically, fire was particularly important in moist, productive grasslands, such as North American tallgrass prairie, and it remains an important tool for the preservation of these grasslands today (Photo by Eva Horne)



grasslands around the world, though the frequency and intensity of fire varies as a function of precipitation (or soil water availability) and aboveground productivity. Historically, many grassland fires originated as a result of lightning strikes or due to the activities of aboriginal humans. Once ignited, fire could sweep relatively unimpeded through large areas of open grassland that lacked natural fire breaks, and fires are generally thought to have been widespread and common in many of the extensive grassland regions around the world. The higher productivity of more mesic grasslands would have promoted more rapid and larger accumulations of combustible fuel, and so fires were likely more frequent in mesic than arid grasslands. However, even desert grasslands can burn once sufficient fuel accumulates, and some arid grasslands are more often now as a result of introduced annual grasses that promote more frequent fires.

The intensity of grassland fires vary, depending on such factors as fuel load (accumulated biomass), fuel condition (compaction, moisture content, etc.), relative humidity, wind speed, and topography. Grassland fires can be very intense and can generate sufficient heat aboveground to damage the aboveground shoots of woody plants (“top kill”) or even kill entire trees. However, because these fires tend to move rapidly and much of the fuel is above the ground, most of the heat is concentrated aboveground and temperatures peak quickly as fire passes. Heat transfer into the soil is generally small, and soil heating into the range that is biologically damaging ( $>60^{\circ}\text{C}$ ) occurs only at the surface. Thus, the belowground buds and meristematic tissues of the grasses and many other grassland plants are well protected against even the most intense grass fires. This is an important contrast to other ecosystems (e.g., forests and woodlands), where the effects of fire are often associated with an immediate negative impact on plant mortality and even the effects of soil heating on loss of soil organic matter and nutrients and changes in soil microbial communities. For grasslands, many of the most significant effects of fire are indirect and result from changes in the postfire environment, rather than the effects of the fire per se. Recovery from a fire event in grasslands in

terms of new plant growth and accumulated aboveground biomass is generally very rapid, especially for mesic grasslands. Recovery in more arid or desert grasslands may take considerably longer.

Changes in natural regimes and/or fire suppression have been implicated as one of the major drivers of contemporary land-cover change in many grasslands worldwide. In many instances, this is a function of a reduction in the frequency or intensity of fires relative to their historical occurrence and subsequent increases in woody plant cover or, in some cases, the conversion of grasslands to shrublands, woodlands, or forest. However, there are also cases where increasing fire frequency is the driver of land-cover change, such as the positive feedbacks between grass cover and fire associated with the spread of invasive fire-prone grasses into ecosystems that were historically less susceptible to fire (e.g., the spread of cheatgrass (*Bromus tectorum*) throughout Western US shrublands). Prescribed fire has also become an important management tool in many grasslands, such as tallgrass prairies where it is used to limit the growth of woody plants and to promote the growth and vigor of the dominant C<sub>4</sub>, or warm-season, grasses. Because of its importance in the development and persistence of tallgrass prairie, research on the effects of fire has been a major emphasis of the Konza Prairie Long-Term Ecological Research Program. Fire alters many aspects of prairie ecosystem structure and functioning. At Konza Prairie, over 20 years of data on the effects of different fire regimes, including annual spring burning and infrequent burning (every 10–20 years), has been amassed. Below examples from these studies have been used to illustrate some of the ecological effects of grassland fires.

Although fires can occur at anytime of the year, dormant season fires are generally most common in grasslands. In tallgrass prairie, burning at the end of winter dormancy (i.e., early spring) is a common management practice. Spring burning generally increases total plant productivity by stimulating growth of the warm-season grasses, particularly in times (wet years) or locations (deeper soils) with adequate soil water available. This is due primarily to the removal of the large amount of plant detritus (up to 1,000 g m<sup>-2</sup>) that accumulates in the absence of the fire and the changes in microclimate and soil resource availability induced by the removal of detritus (Knapp and Seastedt 1986). This detritus acts as a mulch layer, insulating the soil surface and greatly limiting light availability for emerging plants. The removal of this accumulated surface detritus and standing dead biomass alters the energy environment and microclimate of the soil. Direct solar inputs to the soil increases soil temperatures as much as 20 °C in the early spring, relative to comparable unburned grasslands. The warmer temperatures promote earlier emergence and more rapid spring growth, especially for the dominant warm-season grasses. In most years, these changes in the soil microclimate promote the growth of the dominant warm-season grasses, as long as there is adequate water in the soil profile. However, removal of the detrital layer also enhances evaporation from the soil surface, and in dry years or shallow soils, this can reduce productivity following fire. This is also a reason that the effects of fire on plant productivity vary across precipitation gradients, with positive effects in wetter grasslands and neutral or negative effects in drier grasslands.

In tallgrass prairie and other mesic grasslands, the enhanced growth of the grasses also increases their ability to compete for limiting resources with other plant species, leading to another effect of frequent fires – a reduction in overall plant species richness and diversity due to reductions in the abundance and cover of many subordinate species, including the cool-season graminoids and the forbs that provide much of the biodiversity in tallgrass prairie. Thus, frequent burning generally increases plant productivity, but lowers plant diversity, at least in ungrazed prairie. The presence of grazers that preferentially graze on warm-season grasses can offset this effect and changes the relationship between fire and plant diversity, as discussed in the next section.

In addition to its more apparent effects on prairie vegetation, fire alters nutrient cycling processes in these grasslands (Blair et al. 1998). The most important effects involve changes in the cycling of nitrogen. Nitrogen (N) is an essential plant nutrient which often is in short supply relative to plant demand, and the availability of N limits plant productivity in many ecosystems. Based on fertilizer studies, N availability has been shown to limit plant productivity in tallgrass prairies. However, N limitation is not a universal characteristic of tallgrass prairie and, in fact, depends on management practices, such as fire and grazing, and on other external factors, such as climate and topography. In addition to its effects on plant productivity, N availability can alter competitive interactions among plant species and, therefore, plant community composition. Nitrogen availability is a major determinant of plant nutritional quality for herbivores, and the N content of plant litter influences rates of litter decomposition and therefore the storage of organic matter in tallgrass prairie soils. Understanding how N cycling processes are altered by different land-use practices, such as burning, is an important prerequisite to understanding and predicting grassland ecosystem responses to these practices.

When plant detritus burns, some nutrients are lost with the smoke and gases, while others are released and deposited in the ash. Much of the nitrogen contained in surface detritus and plants is volatilized, or converted to gaseous forms, in the heat of a prairie fire, while other heavier elements such as phosphorus and many cations are simply deposited in the ash. The volatilization of nitrogen by fire is the major pathway by which nitrogen is lost from the prairie (especially ungrazed prairie), and frequent fires represent a substantial loss of the prairie's nitrogen capital. Nitrogen cycling in frequently burned prairie is further altered by the responses of the grasses, which produce more root biomass and produce plant tissue which is lower in N content, or which has a higher C/N ratio. The increased input of organic matter with a wider C/N ratio stimulates nitrogen immobilization by soil microbes, leading to even greater N limitation under frequent burning regimes. Thus, the loss of N, along with the increased growth of the grasses, greatly reduces the amount of available N in the soil and increases N limitation for the plants growing in frequently burned prairie. An important question is how a frequently burned prairie can maintain higher productivity than unburned prairie, in spite of increased N limitation. This appears to be, in part, to the increased abundance of warm-season grasses and the high efficiency with which these grasses utilize N, giving them a competitive advantage over other coexisting plant types.

## Grazing in Grasslands

Grazing is a form of herbivory in which herbaceous plants (grasses and forbs) are consumed by herbivores (Fig. 12). This process differs from browsing in which the leaves and woody twigs are consumed from trees and shrubs. Grazing is, or was historically, an important process in nearly all grasslands and is considered a key factor affecting species composition and biomass production in grassland ecosystems. The relationship between grazers and grasslands has developed over millions of years, and it is likely that grazers and grasslands ecosystems coevolved. Grazers promote heterogeneity in grasslands by selectively consuming some species while leaving others, through trampling, soil compaction, soil tunneling, and redistribution of nutrients.

Grazing occurs both aboveground (leaves and stems) and belowground (fine roots and root hairs) by a wide variety of animal herbivores from microscopic invertebrates to the large mammalian megafauna. In general, while a relatively low density of the largest grazers (e.g., bison, wildebeest, zebra) can consume a significant proportion of plant biomass, many small rodents or numerous invertebrates can have comparable impacts within the same grassland when their densities are high enough. Grazers can have a tremendous impact on grasslands through their effects on plant populations and community composition, on energy flow and nutrient cycling in grassland ecosystems, and on landscape-level heterogeneity and movement of materials (McNaughton 1985; Knapp et al. 1999). Although some grasslands (the tallgrass prairies of North America or the Serengeti grasslands of Africa) appear to be well adapted to relatively high grazing intensities, other grasslands can be quickly degraded by overgrazing. When managed in an unsustainable fashion (e.g., overgrazing), large ungulates can significantly impact grassland health and sustainability.

Spatial and temporal patterns of activity by grazers can be greatly affected by fire and grazing by large herbivores and, in turn, can greatly alter the effects of fire in grasslands (Fig. 13). These interactive effects of fire and grazing are especially important in mesic temperate and tropical grasslands. Many large grazers are attracted to recently burned areas, as the removal of detritus and the emergence of new grasses provides a high-quality grazing areas. Intensive grazing in these areas can lead to selection for high-quality grazing tolerant grasses and the formation of a “grazing lawn.” At the same time, increased grazing intensity in burned areas removes aboveground biomass that would otherwise accumulate and serve as fuel for future fires. As a result, fire and grazing in extensive grasslands can be spatially and temporally dependent on each other and can transform the grassland landscape into a dynamic mosaic of shifting patches that vary in time since fire, grazing intensity, and fuel accumulation (Fuhlendorf and Engle 2011). This spatiotemporal interaction of fire and grazing has been referred to pyric herbivory, a term that highlights the codependence of fire and grazing in many natural grasslands. This same principle is the basis of a proposed alternative management practice called patch-burn grazing, which is designed to mimic the interaction of fire and grazers to promote greater heterogeneity and habitat for wildlife in grasslands managed for production of domestic grazers (i.e., cattle).

**Fig. 12** Large vertebrate grazers, such as these North American bison (*Bison bison*), can modify the plant species composition and the flow of energy and resources within grassland ecosystems (Photo by Matt Whiles)



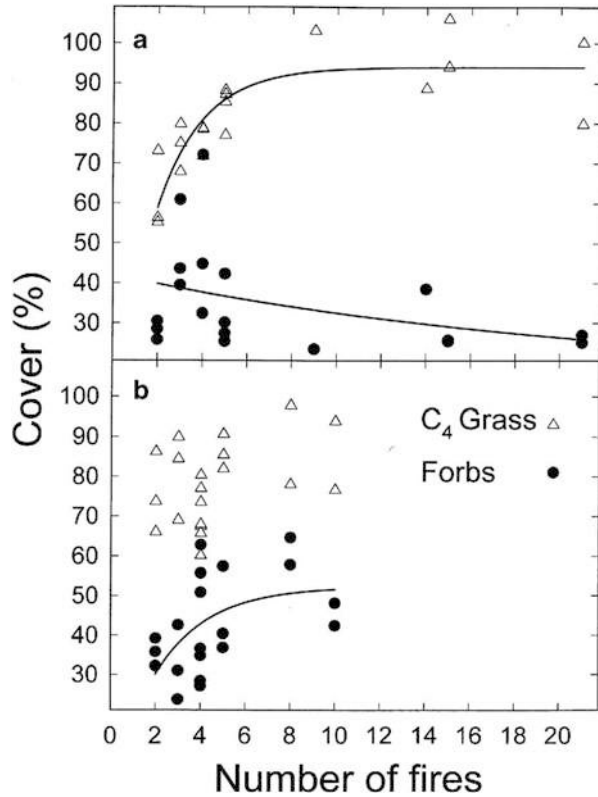
**Fig. 13** There are often significant interactions between fire and the activities of grazers, as illustrated by the patchy nature of fire in areas that are grazed by bison. Grazers are often attracted to fresh grass regrowth in areas that were previously burned, but the activities of grazers can reduce fine fuel to carry future fires resulting in a mosaic of burned and unburned patches, as shown here (Photo by John Briggs)



As noted in section “[Fire in Grasslands](#),” fire in ungrazed mesic grasslands often reduces heterogeneity and lowers species diversity by removing detritus, reducing woody plant cover, and promoting the dominance of grasses that respond positively to fire. However, large ungulate grazers selectively feed on many of these same grasses.



**Fig. 14** Effects of fire frequency on the abundance (cover per 10 m<sup>2</sup>) of warm-season (*C*<sub>4</sub>) grasses and forbs in ungrazed tallgrass prairie (A) and in tallgrass prairie grazed by bison (B). In ungrazed prairie, more frequent fire greatly increases the abundance of the dominant *C*<sub>4</sub> grasses and decreases the abundance of forbs, which results in lower overall plant diversity. However, the presence of grazers offsets these effects and increases the relative abundance of forbs even with more frequent fires (From Collins et al. 1998)



Thus, grazing can offset the reduction in species diversity that results from frequent burning of productive grasslands such as tallgrass prairie by reducing grass dominance and increasing plant species diversity in areas that have been burned (Fig. 14). In xeric grasslands, on the other hand, grazing may lower species diversity particularly by altering the availability of suitable microsites for forb species. These effects are strongly dependent on grazing intensity. Overgrazing may rapidly degrade grasslands to systems dominated by weedy and nonnative plant species.

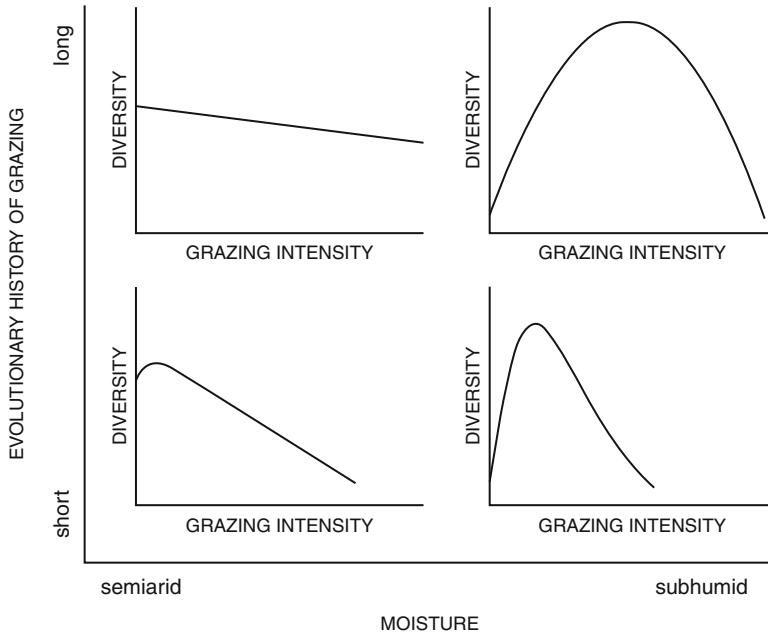
Most grazers are highly selective in the plants they consume. This selectivity results in a landscape with heterogeneous species composition and patchy nutrient distributions. Plants that lose tissues to grazing must use assimilated carbon and nutrients to regrow leaves (or roots), leaving less palatable species to grow taller and increase in number. Many large grazers such as African buffalo, North American bison, or domesticated cattle primarily consume the grasses, allowing less abundant forb species to increase in abundance and new species to colonize the space that is made available. In more productive grasslands adapted to the activities of grazers, grazing can be an important management tool to increase biodiversity when managed at appropriate stocking rates.

Grazers also accelerate the conversion of plant nutrients from forms that are unavailable for plant uptake to forms that can be readily used. Essential plant nutrients, such as nitrogen, are bound for long periods of time in unavailable (organic) forms in plant foliage, stems, and roots. These plant parts are slowly decomposed by microbes, and the nutrients they contain are only gradually released in plant-available (inorganic) forms. This decomposition process may take several years. Grazers consume plant tissues, process this material inside the gut, and excrete nutrients that are available for uptake by plants back onto the landscape. This nutrient processing happens rapidly compared to the slow decomposition process, and nutrients are excreted in high concentrations in small patches. Thus, grazers may increase the availability of potentially limiting nutrients to plants as well as alter the spatial distribution of these resources.

Some grasses and grassland plants can compensate for aboveground tissue lost to grazers by growing faster after grazing has occurred. Thus, even though ~50 % of the grass foliage may be consumed by large grazers, when compared to ungrazed plants at the end of the season, the grazed grasses may be only slightly smaller, the same size or even larger than ungrazed plants. This latter phenomenon, called “overcompensation,” has not been shown in all grassland ecosystems, but the ability of grasses to compensate partially or fully for foliage lost to grazers is well established. Compensation occurs for several reasons including an increase in light available to growing shoots in grazed areas, greater nutrient availability to regrowing plants, and increased soil water availability (because less water is being lost via leaf transpiration compared to an ungrazed dense plant canopy).

As with fire, the impact of grazing on grasslands and the ability of grasslands to tolerate heavy grazing depend upon where the grassland occurs (usually more mesic grasslands can recover more quickly than arid grasslands) as well as the growth form of the grasses within the system: caespitose (bunch-forming grasses) versus rhizomatous grasses. But another key factor determining the ecological responses of grasslands to grazing is the evolutionary history of the grassland (Fig. 15). In general, grasslands with a long evolutionary history of grazers, as in Africa and North and South America, are very resilient to grazing. The evolution of this resilience may reflect the migratory nature of most herds of large grazing mammals. Historically, herds of thousands (and up to millions) of grazers moved across African and North American landscapes in response to seasonal cues and availability of resources. While the impact of these large herds has (or had in the case of North America) a tremendous impact on the grasslands, the animals spend only a small period of time within a given location, allowing for periods of recovery before the next grazing event.

Due to the ability of grasses to cope with high rates of herbivory, many former natural grasslands are now being managed for the production of domestic livestock, primarily cattle in North and South America and Africa, as well as sheep in Europe, New Zealand, and other parts of the world. Grasslands present a vast and readily exploited resource for domestic grazers. However, if not managed properly, grasslands can be easily overexploited with subsequent land degradation, nutrient loss, and susceptibility to invasion by undesirable plant species.



**Fig. 15** Response of grassland plant communities to grazing intensity as a function of moisture gradients and grazing evolutionary history (From Milchunas et al. 1988)

## Potential Threats to Grassland Conservation

Although grasslands are the natural vegetation of much of the Earth's terrestrial surface, many grassland communities and ecosystems are among the most impacted and endangered in the world. Why is this? In many parts of the world, grasslands are the natural vegetation on some of the most ecologically productive lands with high levels of soil nutrients and an open rolling topography conducive to cultivation or ranching. Consequently, many grasslands around the globe have been cultivated and converted to agriculture use or are intensively managed for the production of domestic livestock. As a result, both the spatial extent of native grasslands and the quality of remaining grasslands is declining. This is due primarily to human-induced modifications such as agriculture, excessive or insufficient fire, livestock grazing, fragmentation, and invasive plants and animals. Precise estimates of the areal extent of these changes are difficult to come by as there is no international organization tracking grasslands and because of the difficulty in identifying what is grassland and what is not. In addition, it is known that all croplands were developed from either forests or grasslands. In that respect, since areas of cropland are expanding, it can be assumed that on the whole, grassland areas are continuing to decline. On the other hand, large areas of tropical rainforests are being cleared to



provide pasture for livestock. Therefore, grasslands – at least in the form of pastures – may be expanding in some localized areas.

Recent estimates suggest that a large percentage of the Earth's total grazing land has been degraded to the point that it has lost some of its animal carrying capacity. Even though the damage from overgrazing is spreading, the world's livestock population continues to grow in step with increases in the human population and a growing demand for meat that accompanies increased wealth; thus, grasslands will continue to deteriorate. As world population increased from 2,500 million in 1950 to over 7 billion, the world's cattle, sheep, and goat populations have also grown exponentially. As a result of overstocking and overgrazing, grasslands in much of Africa, the Middle East, Central Asia, the northern part of the Indian subcontinent, Mongolia, and much of northern China are deteriorating. While grazing was once a pastoral activity that involved people moving with their herds from place to place, it has become a far more sedentary undertaking. The result is an increase in grassland degradation worldwide.

In addition to grazing, grassland environments are the basis for major agricultural areas worldwide. Historically, the major threat to the mesic grasslands of the United States (and world) was cultivation of soils and conversion to row-crop agriculture. Although the conversion of grasslands to agriculture continues today (especially with increased demand for biofuels; Fargione et al. 2008), some of the most significant losses of grasslands now are related to changing land management coupled with other global change phenomena. Temperate grasslands are important from both agronomic and ecological perspectives. As mentioned earlier, many of the most productive temperate grasslands in North America and elsewhere are considered to be endangered ecosystems. For example, in the United States, up to 99 % of native tallgrass prairie ecosystems in some states have been plowed and converted to agricultural use or lost due to urbanization. Similar but less dramatic losses of mixed and shortgrass prairies have occurred in other areas.

While the loss of native grasslands due to agricultural conversion and urbanization is ongoing in many locations around the world, another major threat is the dramatic increase in shrubs and trees (many of them native species) now occurring in many grasslands (Briggs et al. 2005). Increases in the abundance and cover of native woody plant species in areas that were historically grass dominated can occur as a result of expansion of woody plant cover within grasslands as well as encroachment of woody species into grasslands from adjacent ecosystems. In many cases, these are tree species that were historically present in grasslands, but at a relatively low abundance. In other cases, grasslands are being invaded by nonnative woody plants. Recent increases in cover and abundance of woody species in grasslands and savannas have been observed worldwide, with well-documented examples from North America, Australia, Africa, and South America. In North America, this phenomenon has been documented in mesic tallgrass prairies of the eastern Great Plains, in subtropical savannas of Texas, in desert grasslands of the southwest, and in shrub steppes of the upper Great Basin. Some of the purported drivers of increased woody plant cover include changes in climate, increased

atmospheric CO<sub>2</sub> concentrations, elevated nitrogen deposition, altered grazing pressure, and changes in disturbance regimes, such as the frequency and intensity of fire. Although the drivers of woody plant expansion may vary for different grassland types, the consequences for grassland ecosystems are strikingly consistent. In most areas, the expansion of woody species increases above ground biomass and thus aboveground carbon storage, but at the same time reduces biodiversity of native grassland fauna and flora. However, the full impact of woody plant encroachment on grassland environments remains to be seen.

Another contemporary threat to native grasslands is the increase of nonnative grass species. For example, in California it is estimated that an area of approximately 7,000,000 ha has been converted to grassland dominated by nonnative annuals primarily of Mediterranean origin. Conversion to nonnative annual vegetation was so fast, so extensive, and so complete that the original extent and species composition of native perennial grasslands are unknown. In addition, across the Western United States, invasive exotic grasses are now dominant in many areas and these species have a significant impact on natural disturbance regimes. For example, the propensity for annual grasses to carry and survive fires is now a major element in the arid and semiarid areas in western North America. In the Mojave and Sonoran deserts of the American Southwest, in particular, fires are now much more common than they were historically which may reduce the abundance of many native cactus and shrub species in these areas. This annual-grass-fire syndrome is also present in native grasslands of Australia and managers there and in North America are using growing season fire to try to reduce the number of annual plants that set seed and thus reduce the population, usually with very mixed results.

---

## Grassland Restoration

Given the ecological importance and extensive loss or degradation of grasslands globally, it isn't surprising that grassland restoration has become increasingly important and widespread, especially in locations where substantial areas of native grasslands have been lost as a result of land-use or land-cover change. Grassland restoration often takes place on formerly cultivated lands and involves reintroduction of native species characteristic of grasslands in that particular region. However, there are other types of grassland restoration, including restorations that target reductions in woody plant cover in areas that have experienced woody plant encroachment or those that target the removal of invasive species and their replacement with native grassland species. The motivation for these restoration efforts varies from restoring native plant biodiversity, to restoring ecosystem processes that provide environmental benefits (e.g., limiting soil erosion and improving water quality, sequestering carbon), to providing suitable habitat for regional native fauna. There are multiple difficulties associated with restoring grassland communities and ecosystems, fragmentation of historically extensive areas of intact grassland, loss of genetic diversity of grassland plant and animal populations, and insufficient area to include some of the drivers that were historically important in

shaping grasslands, such of landscape-level patterns of fire and grazing. Nevertheless, there are widespread efforts to restore native grassland diversity and ecosystem functioning.

Much research has focused on restoring temperate grasslands in North America, particularly in the tallgrass prairie region where the cover of native tallgrass prairie has declined 82–99 % since the 1830s, primarily as a result of cultivation for agricultural use. Dispersal of native grasslands plants into abandoned agricultural fields is very limited, and many areas targeted for grassland restoration are isolated from potential native seed sources. As a result, restoration of these grasslands typically begins with the introduction of seeds or transplants of native plant species. One of the earliest attempts to restore tallgrass prairie on ex-arable land began in the 1930s at the Curtis Prairie in Madison, WI. Since then numerous prairie restorations have been initiated at a range of spatial scales, and recent decades have seen a sharp increase in efforts to restore prairie for both conservation and research purposes. In fact, restored grasslands are being used to address a variety of basic and applied ecological questions, such as the relationship between species diversity and ecosystem function, the role of resource heterogeneity in structuring plant communities, or the role of dominant species in community assembly (Baer et al. 2003, 2005). It has even been suggested that restoration can serve as an “acid test” of our understanding of community assembly.

Reestablishing the dominant grass species in restored grasslands is relatively easy. However, it is difficult to establish and maintain many of the less common species that provide the majority of biodiversity in native prairies. As a result, restored grasslands generally have much lower diversity than comparable native grasslands. Even when initial seed mixtures include a diverse assemblage of subdominant and rare forbs, establishment of these species may be poor. In addition, the cover of the dominant warm-season grasses tends to increase over time in many restored grasslands, with a concurrent loss of rarer species, such that diversity declines over time. Overseeding (adding additional seeds to restored grasslands) is sometimes used in an effort to overcome potential dispersal limitations and enhance recruitment of new species in older restorations. However, the underlying reasons for loss of diversity are unclear, and additional studies are needed to assess the relative importance of dispersal limitations, interspecific competition, resource heterogeneity, herbivory, or other factors on limits to diversity in restored grasslands.

The restoration of grasslands on former agricultural soils can provide other benefits, including reduced soil erosion, greater nutrient retention, and providing a sink for atmospheric CO<sub>2</sub>. One of the well-documented effects of cultivation is the loss of a significant proportion of carbon stored in the form of soil organic matter. Cultivation of grasslands reduces inputs of plant-derived new organic matter and the disruption of soil structure coupled with improved aeration greatly increased microbial mineralization of stored soil carbon. As a result, grasslands can lose from 20 % to 50 % of their organic carbon content within a few decades of cultivation. Eventually, these cropland soils come to a new equilibrium soil C content that is much lower than the grassland soils they replaced. However, if these fields are

removed from cultivation and restored with perennial grasses and forbs, the soil carbon pools will increase as new perennial root systems redevelop, new C inputs are added to the soil, and soil structure begins to reform. Several studies have documented significant rates of carbon accrual, generally in the range of 20–60 g C m<sup>-2</sup> year<sup>-1</sup>, and suggested that these rates could persist for decades until a new equilibrium is reached. It is important to point out, however, that although some soil C (and N) pools in restored prairie may approach those of native prairie within a few decades, it may take much longer for other soil properties (e.g., soil aggregate structure or soil microbial communities) to recover.

---

## Future Directions

Below are a few suggestions regarding future research directions that are particularly relevant to grassland conservation and management. This is not an exhaustive list, but rather meant to stimulate further discussions about the scope and directions of future research required for an improved understanding of grassland ecology and the maintenance/conservation of these ecosystems around the world.

- It is essential to develop a mechanistic understanding of how grasslands are responding and will respond in the future, to multiple global change phenomena, including changes such as enhanced N deposition, altered climate, and elevated CO<sub>2</sub> changing land use and land cover. Additional multifactor experiments are needed to address the interactions of global changes driver that occur in combination. Better forecasting of potential responses to environmental changes will improve both conservation goals and the sustainable use of grassland resources.
- A better understanding of the factors that affect the success of grassland restoration efforts is needed. While many studies have focused on deterministic factors, such as site preparation, seed sources, and seeding rates, additional studies that address the relative importance of stochastic factors (e.g., climatic variability, in establishment years) are also needed. This information will be critical for designing more effective methods of restoring grassland in areas where they have been degraded or extirpated.
- Effective management and conservation of grasslands will require a better understanding of social and economic drivers. One example of a newly emerging threat is the increase in restrictions on the use of grassland fires for management and conservation due to human health concerns. There is a need to explore other methods to minimize the negative effects of burning (e.g., impacts of smoke on air quality) in areas where fire is essential for maintaining grassland flora and fauna or perhaps ways to “simulate” some of the major ecological effects of fire to achieve desired management goals.
- Understanding the abiotic and biotic conditions that result in variable responses to grazing in different grasslands has both basic and applied significance. Many studies report contrasting effects to grazing, for example, with respect to root productivity and belowground carbon allocation. Similar conflicting results have been reported to for a suite of other responses. The occurrence of grazing in most

grasslands, and increased reliance on rangelands as a source of food for a growing human population, increases the importance of understanding grassland-grazer interactions and designing more sustainable means of managing grasslands for multiple goals in a changing environment.

- Linking theory to conservation, grasslands may serve as the first terrestrial ecosystem in the development of “warning signs” that signify a pending transition to an alternate ecosystem attractor (state shift). These warning signs would allow land managers and conservationists to employ adaptive management techniques to avoid the rapid conversion of grassland to shrubland or grassland to degraded states.

---

## References

- Anderson RC. The historic role of fire in the North American grassland. In: Collins SL, Wallace LL, editors. *Fire in North American tallgrass prairies*. Norman: University of Oklahoma Press; 1990.
- Archibold OW. *Ecology of world vegetation*. London/New York: Chapman and Hall; 1995.
- Baer SG, Blair JM, Knapp AK, Collins SL. Soil resources regulate productivity and diversity in newly established tallgrass prairie. *Ecology*. 2003;84:724–35.
- Baer SG, Collins SL, Blair JM, Fiedler A, Knapp AK. Soil heterogeneity effects on tallgrass prairie community heterogeneity: an application of ecological theory to restoration ecology. *Restor Ecol*. 2005;13:413–24.
- Benson E, Hartnett DC. The role of seed and vegetative reproduction in plant recruitment and demography in tallgrass prairie. *Plant Ecol*. 2006;187:163–77.
- Blair JM, Seastedt TR, Rice CW, Ramundo RA. Terrestrial nutrient cycling in tallgrass prairie. In: Knapp AK, Briggs JM, Hartnett DC, Collins SL, editors. *Grassland dynamics: long-term ecological research in tallgrass prairie*. New York: Oxford University Press; 1998.
- Blecker SW, McCulley RL, Chadwick OA, Kelly EF. Biologic cycling of silica across a grassland bioclimate sequence. *Global Biogeochem Cycles*. 2006;20, GB3023. doi:10.1029/2006GB002690.
- Briggs JM, Knapp AK, Blair JM, Heisler JL, Hoch GA, Lett MS, McCarron JK. An ecosystem in transition: causes and consequences of the conversion of mesic grassland to shrubland. *BioScience*. 2005;55:243–54.
- Collins SL, Knapp AK, Briggs JM, Blair JM, Steinauer E. Modulation of diversity by grazing and mowing in native tallgrass prairie. *Science*. 1998;280:745–7.
- Dalgleish HJ, Hartnett DC. The effects of fire frequency and grazing on tallgrass prairie productivity and plant composition are mediated through bud bank demography. *Plant Ecol*. 2009;201:411–20.
- Fargione JE, Hill J, Tilman D, Polasky S, Hawthorne P. Land clearing and the biofuel carbon debt. *Science*. 2008;319:1235–8.
- Frank DA. Drought effects on above and below ground production of a grazed temperate grassland ecosystem. *Oecologia*. 2007;152:131–9.
- Fuhlendorf SD, Engle DM. Restoring heterogeneity on rangelands: ecosystem management based on evolutionary grazing patterns. *BioScience*. 2011;51:625–32.
- Hoekstra JM, Boucher TM, Ricketts TH, Roberts C. Confronting a biome crisis: global disparities of habitat loss and protection. *Ecol Lett*. 2005;8:23–9.
- Knapp AK, Seastedt TR. Detritus accumulation limits productivity of tallgrass prairie. *BioScience*. 1986;36:662–8.
- Knapp AK, Blair JM, Briggs JM, Collins SL, Hartnett DC, Johnson LC, Towne EG. The keystone role of bison in North American tallgrass prairie. *BioScience*. 1999;49:39–50.

- Lauenroth WK, Burke IC, Gutmann MP. The structure and function of ecosystems in the central North American grassland region. *Great Plains Research* 1999;9:223–59.
- McNaughton SJ. Ecology of a grazing ecosystem: the Serengeti. *Ecol Monogr.* 1985;55:259–94.
- Milchunas DG, Sala OE, Lauenroth WK. A generalized model of the effects of grazing by large herbivores on grassland community structure. *Am Nat.* 1988;132:87–106.
- Nippert JB, Knapp AK. Linking water uptake with rooting patterns in grassland species. *Oecologia.* 2007;153:261–72.
- Nippert JB, Knapp AK, Briggs JM. Intra-annual rainfall variability and grassland productivity: can the past predict the future. *Plant Ecol.* 2006;184:65–74.
- Nippert JB, Wieme RA, Ocheltree TW, Craine JM. Root characteristics of C<sub>4</sub> grasses limit reliance on deep soil water in tallgrass prairie. *Plant and Soil.* 2012;355:385–94.
- Prasad V, Stromberg CA, Alimohammadian H, Sahni A. Dinosaur coprolites and the early evolution of grasses and grazers. *Science.* 2005;310:1177–90.
- Seastedt TR. Soil systems and nutrient cycles of the North American prairie. In: Joern A, Keeler KK, editors. *The changing prairie.* New York: Oxford University Press; 1995.
- Silvertown J, Poulton P, Johnston E, Edwards G, Heard M, Biss PM. The park grass experiment 1856–2006: its contribution to ecology. *J Ecol.* 2006;94:801–14.
- Strömberg CAE. Evolution of grasses and grassland ecosystems. *Annu Rev Earth Planet Sci.* 2011;39:517–44.
- Walter H. *Ecology of tropical and subtropical vegetation.* Edinburgh: Oliver & Boyd; 1971.
- Weaver JE. Recovery of midwestern prairies from drought. *Proc Am Philos Soc.* 1944;88:125–31.
- Weaver JE. *Prairie plants and their environment: a fifty-year study in the Midwest.* Lincoln: University of Nebraska Press; 1968.
- Weaver JE, Albertson FW. Major changes in grassland as a result of continued drought. *Bot Gaz.* 1939;100:576–91.
- Weaver JE, Fitzpatrick TJ. *The prairie.* *Ecol Monogr.* 1934;4:109–295.
- White R, Murray S, Rohweder M. Pilot analysis of global ecosystems (PAGE): grassland ecosystems. Washington, DC: World Resources Institute (WRI); 2000.

## Further Reading

- Axelrod DI. Rise of the grassland biome, Central North America. *Bot Rev.* 1985;51:163–201.
- Beerling DJ, Osborne CP. The origin of the savanna biome. *Glob Chang Biol.* 2006;12:2023–31.
- Borchert JR. The climate of the central North American grassland. *Ann Assoc Am Geogr.* 1950;40:1–39.
- Collins SL, Wallace LL, editors. *Fire in North American tallgrass prairies.* Norman: University of Oklahoma Press; 1990.
- French N, editor. *Perspectives in grassland ecology. Results and applications of the United States international biosphere programme grassland biome study.* New York: Springer; 1979.
- Gibson DJ. *Grasses and grassland ecology.* New York: Oxford University Press; 2009.
- Havstad KM, editor. *Structure and function of a Chihuahuan desert ecosystem: the jornada basin long-term ecological research site.* New York: Oxford University Press; 2006.
- Knapp AK, Briggs JM, Hartnett DC, Collins SL, editors. *Grassland dynamics: long-term ecological research in tallgrass prairie.* New York: Oxford University Press; 1998.
- Lauenroth WK, Burke IC, editors. *Ecology of the shortgrass steppe: a long-term perspective.* New York: Oxford University Press; 2008.
- McClaran MP, Van Devender TR. *The desert grassland.* Tucson: University of Arizona Press; 1997.

- Oesterheld M, Loreti J, Semmartin M, Paruelo JM. Grazing, fire, and climate effects on primary productivity of grasslands and savannas. *Ecosyst World* ISSU. 1999;16:287–306.
- Osborne CP. Atmosphere, ecology and evolution: what drove the Miocene expansion of the C4 grasslands? *J Ecol*. 2008;96:35–45.
- Risser PG, Birney EC, Blocker HD, May SW, Parton WJ, Weins JA. *The true prairie ecosystem*. Stroudsburg: Hutchinson Ross; 1981.
- Sala OE, Parton WJ, Joyce LA, Lauenroth WK. Primary production of the central grassland region of the United States. *Ecology*. 1988;69:40–5.
- Samson F, Knopf F. *Prairie conservation in North America*. BioScience. 1994;44:418–21.
- Weaver JE. *North American prairie*. Lincoln: Johnsen Publishing; 1954.

Anna R. Armitage

## Contents

Introduction .....	426
History .....	427
Stressors .....	427
Salt Marshes .....	431
Zonation .....	431
Case Study: Plant-Animal Facilitation in a New England Salt Marsh .....	433
Mangroves .....	437
Mangrove Stress Adaptations .....	437
Zonation .....	440
Case Study: Plant-Animal Interactions on Mangrove Islands in Florida .....	443
Future Directions: The Salt Marsh-Mangrove Ecotone: A Developing Field .....	444
Ecosystem Functions and Services .....	445
Water Quality .....	447
Nutrient Cycling and Storage .....	447
Erosion Control and Surge Buffer .....	448
Nursery Habitat .....	449
Recreation .....	449
Management Issues and Strategies .....	450
Development .....	450
Sea Level Rise .....	451
Freshwater Diversion .....	452
Eutrophication .....	452
Policy .....	452
Restoration .....	453
Future Directions: Integrating Science and Restoration .....	454
References .....	454

---

A.R. Armitage (✉)

Department of Marine Biology, Texas A&M University at Galveston, Galveston, TX, USA

e-mail: [armitaga@tamug.edu](mailto:armitaga@tamug.edu)



---

**Abstract**

- Coastal wetlands are plant communities at the land-sea interface. Two common types of coastal wetlands are salt marshes and mangrove swamps. Marshes are dominated by nonwoody grasses and shrubs; mangrove swamps are dominated by trees.
- The global distribution of salt marshes and mangroves is governed by temperature; most mangrove species cannot tolerate freezing temperatures, so they grow in warmer tropical and subtropical latitudes. Marshes are more common in cooler temperate latitudes.
- Salt marshes and mangroves overlap in some subtropical regions; these areas may experience shifts in species composition in response to climate change. The dynamics and ecological consequences of these shifts are important topics for future research.
- Plants in coastal wetlands are adapted for abiotic stressors including prolonged inundation, which causes soil anoxia, and high salinity.
- Salt marshes exhibit predictable zonation patterns, where the distribution of species within a site varies with small changes in elevation. These zonation patterns are driven by species-specific adaptations to abiotic stressors and by interspecific competition. Zonation patterns in mangrove swamps are more variable.
- Coastal wetlands provide a variety of ecosystem services to human communities: wetlands can improve water quality, store nutrients, and buffer against erosion and storm surge and provide nursery habitat for commercially and recreationally important fishery species.
- Current management issues in coastal wetlands include encroaching suburban and agricultural development, sea level rise, nutrient enrichment and eutrophication from agricultural runoff and treated sewage discharge, and freshwater diversion.
- The policies regulating development on coastal wetlands are complex and dynamic. Restoration is the most common approach to mitigate for anthropogenic impacts. An understanding of wetland ecology is crucial to making wise decisions concerning the nature and direction of restoration projects.

---

**Introduction**

Coastal plant communities are broadly defined as those habitats shaped by terrestrial and marine influences. Many, though not all, coastal habitats can be defined as **wetlands**; the ecology and management of those habitats are covered in this chapter. Wetlands are defined by the United States Army Corps of Engineers by the presence of three features: (1) **wetland hydrology**, inundation or saturation for at least part of the growing season; (2) **hydric soils**, soils that are anoxic (containing little or no oxygen for at least part of the growing season; this condition usually develops when soils are inundated with water); and (3) **hydrophytic vegetation**, vegetation adapted to wet conditions.

The coastal wetlands covered in this chapter are often located within estuaries. An **estuary** is a semi-enclosed body of water where freshwater from rivers or streams mixes with oceanic waters, creating **brackish** (slightly salty) conditions. Tidal movement and riverine freshwater input are variable, causing spatial and temporal variations in salinity (Fig. 1). Freshwater input supplies estuaries with sediment, organic matter, and critical nutrients such as nitrogen, phosphorus, and iron. Tidal marine input brings in animal larvae and other essential nutrients such as sulfate and bicarbonate. The combination of these freshwater and marine inputs makes estuaries highly productive habitats.

---

## History

Many early human cultures lived in harmony with wetlands, using these productive habitats to obtain food, fuel, and shelter. However, beginning in the 1700s, and perhaps even earlier, many agriculture-based cultures viewed wetlands as fallow areas with no cultivation value and as breeding grounds for disease-carrying insects. For decades, wetlands were drained for agriculture or cleared and filled for development. By the mid-twentieth century, the resultant wetland losses totaled more than 50 % worldwide; up to 80 % of that loss may be attributable to agricultural expansion (Dahl 1990).

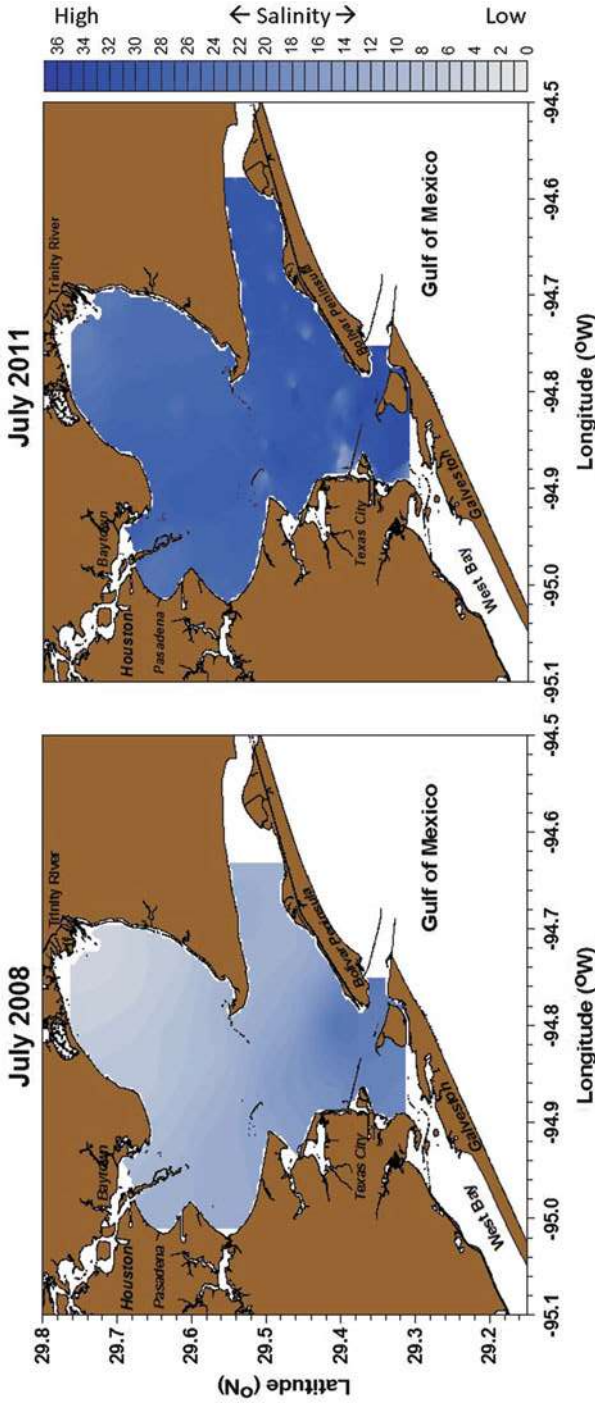
By the 1970s, however, the links between wetland habitats and vital coastal ecosystem services – fishery support, erosion control, water quality improvement – had become better understood. The rate of development slowed, impacts became better managed, and restoration began in earnest. Now, the need to protect and manage these habitats has emerged as a top priority in coastal management. These ecosystem services and management and restoration challenges will be discussed in more detail later in the chapter.

---

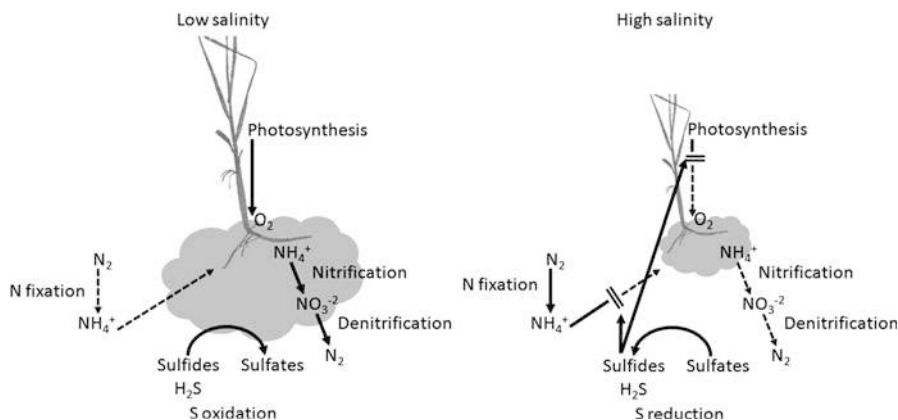
## Stressors

Freshwater and marine inputs can augment estuarine productivity, but those inputs also create abiotically stressful conditions. Plant communities are particularly strongly influenced by salinity and flooding, which is usually accompanied by **anoxia** (no oxygen) or **hypoxia** (low oxygen) in the soil.

Most plants in coastal wetlands are **halophytes** – tolerant of high salt levels. Halophytes can withstand some amount of salt in their tissues, but even the most halophytic species must be able to avoid excessive salt accumulation. High concentrations of salt ions can have many negative impacts on plants: salt ions can be toxic, create an osmotic imbalance that prevents uptake of water even when inundated, and repel and prevent uptake of positively charged nutrients like  $\text{NH}_4^+$  (ammonium). At the ecosystem level, saline coastal wetlands often have lower plant biomass but faster rates of decomposition, which in turn yields slower rates of



**Fig. 1** *Left:* typical salinity gradient in the Galveston Bay estuary (Texas, USA), depicting lower salinity (*light blue shades*) near the riverine inputs and higher salinity (*dark blue shades*) near the marine input. *Right:* salinity gradient in the bay during an exceptional drought in 2011 (Data provided by the Galveston Bay Estuary Program)

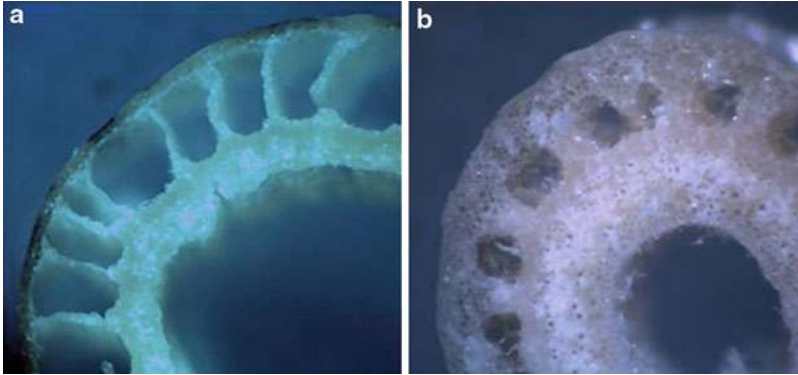


**Fig. 2** Simplified conceptual model depicting the relationships between plant productivity, salinity, oxygen availability, and nitrogen and sulfur cycling in wetland sediments. The *gray cloud* represents the relative size of the oxygenated rhizosphere. *Solid arrows* represent active processes; *dashed arrows* represent inhibited or reduced processes

nitrogen accumulation relative to freshwater and brackish wetlands (Craft 2007; Craft et al. 2009). Furthermore, potential **denitrification**, or the conversion of nitrate ( $NO_3^-$ ) to biologically inert nitrogen gas ( $N_2$ ), is generally lower at higher salinities (Fig. 2). This is a critical step in the removal of nitrogen from wastewater in treatment wetlands (this topic will be discussed in more detail later in the chapter).

Most halophytes have some mechanism, such as storage in vacuoles or high concentrations of glycolipids and sterols in cell membranes, to help halophytes **exclude** salt from metabolically active parts of cells. Other common salt avoidance mechanisms include **secretion**, where salt is excreted from the plant through specialized glands, usually on the leaves; **storage**, where plants concentrate salt in the bark or older leaves that are then sloughed off or dropped; **succulence**, where plants store water to dilute internal salts; and **external exclusion**, where plants produce waxy substances such as **suberin** to block salt uptake through the root epidermis.

Coastal wetlands can be inundated by tides for extended periods of time, and plant and animal respiration quickly use up the biologically available oxygen in tidal flood waters. This causes **hypoxic** or **anoxic** conditions in wetland soils; these conditions may be temporary or can persist for weeks or longer. Low oxygen conditions facilitate the decomposition process, where bacteria reduce sulfate ( $SO_4^{2-}$ ) to hydrogen sulfide ( $H_2S$ ). The production of sulfides generates a “rotten egg smell” in many wetlands. This is a natural process, but sulfides can be toxic at high concentrations or inhibit nutrient uptake by vascular plants (Fig. 2). To reduce sulfide production and oxygenate wetland plant roots, a common adaptation in wetland plants is **aerenchyma** tissue. Aerenchyma refers to internal spaces that



**Fig. 3** Rhizome cross sections of two wetland plants, showing the hollow spaces forming the aerenchyma tissue. (a) *Spartina alterniflora*, a low-elevation grass species with extensive aerenchyma. (b) *Spartina patens*, a mid- to high-elevation grass species with less aerenchyma tissue (Photo credit A.R. Armitage)

extend from the leaves to the roots, providing a low-resistance internal pathway for the transport of oxygen from the leaves above the water to the submerged tissue (Fig. 3). Aerenchyma forms from the collapse of cortex cells in “programmed cell death” (**apoptosis**). Through aerenchyma, oxygen is transported to the roots to be used for metabolic processes. The subsequent oxygenation of the **rhizosphere** (zone surrounding the roots of plants) can lower sulfide production and reduce sulfide toxicity. If, however, sulfide production is extremely high, aerenchyma can become occluded by **callus** tissue (cells that grow over wounds), leading to plant dieback events.

Both salinity and low soil oxygen levels can potentially impact nitrogen cycling in coastal wetlands, largely because some steps in the nitrogen cycle are oxygen dependent, and others require anoxic conditions. The simplified conceptual diagram in Fig. 2 illustrates some of the key interactions among salinity, oxygen levels, and the nitrogen cycle. High salinity is linked to lower primary productivity, thus lowering oxygen production and transport to the rhizosphere. Lower oxygen levels in the rhizosphere facilitate the anaerobic reduction of sulfate to hydrogen sulfide. Hydrogen sulfide ( $H_2S$ ) is toxic at high concentrations, which further reduces productivity and creates a feedback that maintains anoxic conditions. **Nitrogen fixation**, the conversion of atmospheric nitrogen ( $N_2$ ) to ammonium ( $NH_4^+$ ), is an anaerobic process that occurs at a relatively rapid rate in most anoxic wetland soils. However,  $H_2S$  blocks ammonium uptake, further reducing productivity and contributing to the feedback loop that maintains anoxic soil conditions. Denitrification is also lower at high salinity, in part due to the salt-mediated inhibition of **nitrification**, an **aerobic** (oxygen dependent) process that converts ammonium into nitrites and then nitrates (Fig. 2).

Different types of wetlands are typically defined by the character of their plant communities. **Swamps** are wetlands dominated by trees or shrubs; **marshes** are primarily composed of herbaceous, nonwoody vegetation such as grasses, rushes,

sedges, and forbs. Both swamps and marshes can occur in marine and freshwater habitats; this chapter will focus on two common types of coastal marine wetland communities: **salt marshes** and **mangrove swamps**.

---

## Salt Marshes

Salt marshes are defined as those marshes subjected to regular tidal flooding by salt water. Salt marshes occur in estuaries and along marine coastlines, primarily in temperate latitudes. In tropical regions, the short-stature grasses and forbs in salt marshes are generally outcompeted by the taller vegetation in mangrove forests, which will be addressed later in this chapter. A typical salt marsh can be subdivided into several zones based on elevation relative to sea level (Fig. 4). Each of these zones varies in salt and flooding stress; the plants in each of these zones are adapted to those conditions.

### Zonation

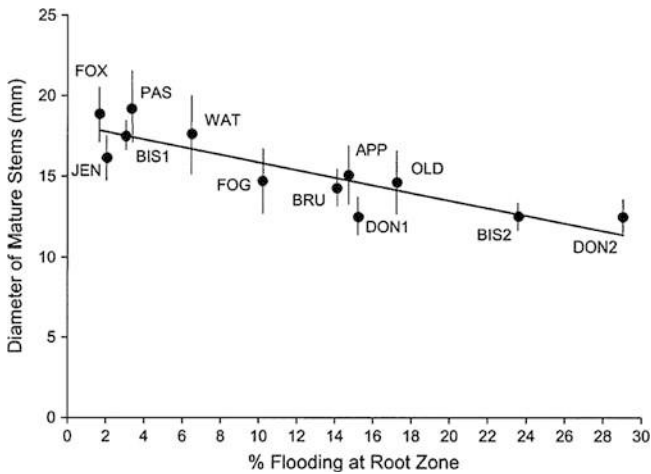
The border zone between salt marshes and nontidal upland habitat is characterized by plants that can grow in moderately saline soils but are intolerant of flooding. Plants in this “*marsh border*” zone along the high tide line often lack aerenchyma tissue, making them sensitive to flooding and associated soil anoxia. For example, the marsh elder, *Iva frutescens*, a typical marsh border plant in the Gulf of Mexico, experiences reduced growth and higher mortality if the roots are inundated for as little as 8 % of the growing season (Fig. 5; Thursby and Abdelrhman 2004).

Below the marsh border zone is a large zone broadly often referred to as *high marsh*. This zone covers a relatively wide elevation range that encompasses a variety of flooding regimes. In this zone, salts tend to accumulate in the soils due to regular but brief tidal flooding followed by evaporation, especially in the more seaward region of the zone. Soil salinities can be more than double that of ambient floodwater. Despite this stressor, plant diversity tends to be high relative to lower elevations (Fig. 6), in part because there are many different adaptations to salt stress. Few plants in this zone are tolerant of prolonged flooding – many have reduced or absent aerenchyma (Fig. 3b).

The lowest vegetated elevation zone in a salt marsh is the *low marsh*. Soil salinity is close to that of ambient floodwater. Plant species in this zone must be able to produce extensive aerenchyma in order to withstand prolonged flooding (Fig. 3a). Few plant species can survive the anoxic conditions associated with extensive flooding, so the low marsh zone has relatively low plant diversity. On the east and Gulf coasts of the United States, the low marsh zone is dominated by *Spartina alterniflora* (Fig. 4). This grass species occurs in all tidally flooded zones of salt marshes, but it grows taller at lower elevations than at higher elevations (Fig. 7). The mechanisms driving this morphological variation are complex; genetic differences and environmental influences both contribute to tall- and short-form morphology.



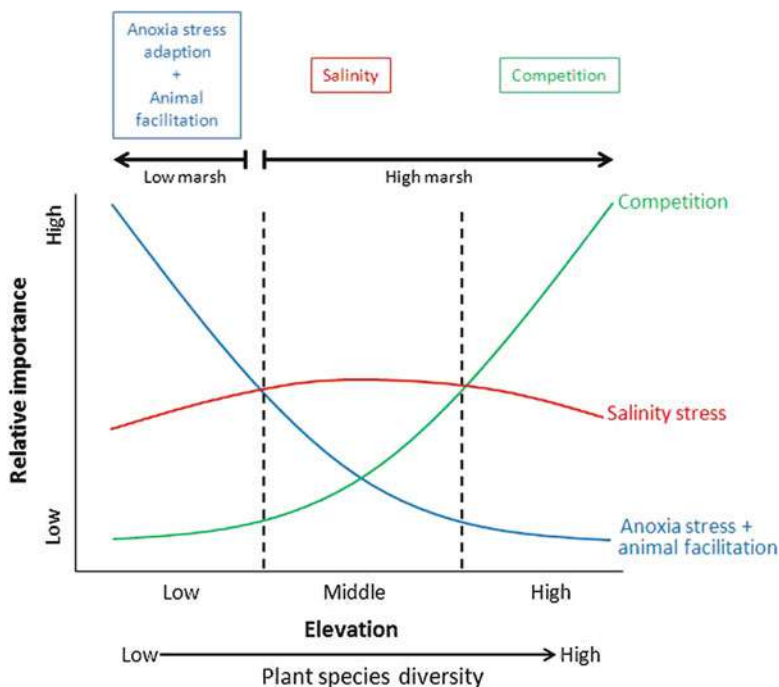
**Fig. 4** Zonation patterns in a salt marsh. In this picture, the marsh border zone is dominated by the marsh elder, *Iva frutescens*. The high marsh zone is comprised of grasses such as *Spartina patens* (marsh hay; lighter green) and the rush *Juncus roemerianus* (black rush; darker green). The low marsh zone is dominated by the grass *Spartina alterniflora* (smooth cordgrass) (Photo credit A.R. Armitage)



**Fig. 5** Excerpt from Fig. 7 in Thursby and Abdelrhman (2004). Relationship between mean stem diameter for older stems of *Iva frutescens* and the duration of flooding (as percent of growing season) at the root zone (10 cm below soil surface). Percent flooding values are based on elevation measurements made near the same location that the stem samples were taken. Vertical bars are  $\pm 2$  SE. The means are of 10 stems except for Fox Hill Cove (FOX) and Jenny Creek (JEN) ( $n = 30$ ) and Mary Donavon Marsh-1 (DON1) ( $n = 20$ ); ( $p < 0.01$ ) (Reprinted with permission from Springer-Verlag)

Within marsh zones, a microhabitat called a **salt pan** can form. Salt pans are unvegetated or sparsely vegetated patches, usually in the high marsh, that are characterized by very saline soil. There are several mechanisms for the formation of salt pans (Boston 1983). For example, **wrack** (floating organic debris) deposition





**Fig. 6** Simplified conceptual model depicting the relative importance of abiotic stressors and biotic interactions at different elevations within salt marshes. The predominant factor in each elevation zone is highlighted in the boxes at the top of the graph

can cover underlying vegetation (Fig. 8a). When it is eventually washed out following a high **spring tide** (during full or new moon phases), the ground underneath will be devoid of vegetation. Alternative mechanisms of salt pan formation include ice scouring, which can remove large clumps of marsh vegetation in the winter, or waterlogging in small topographic depressions, which can cause mortality of established plants. In all cases, after initial formation of the bare patch, evaporation will rapidly raise soil salinity, often to more than twice as high as ambient seawater. High salinity will depress seed germination and inhibit plant invasion, preventing recolonization and maintaining the salt pan microhabitat for long periods of time. Vegetation in salt pans is typically restricted to a few individuals of extremely salt-tolerant species (e.g., *Sarcocornia* spp.) and blue-green algae (cyanobacteria) (Fig. 8b). Although these microhabitats have little vegetation, they provide important roosting habitat for many coastal bird species (Fig. 8c).

### Case Study: Plant-Animal Facilitation in a New England Salt Marsh

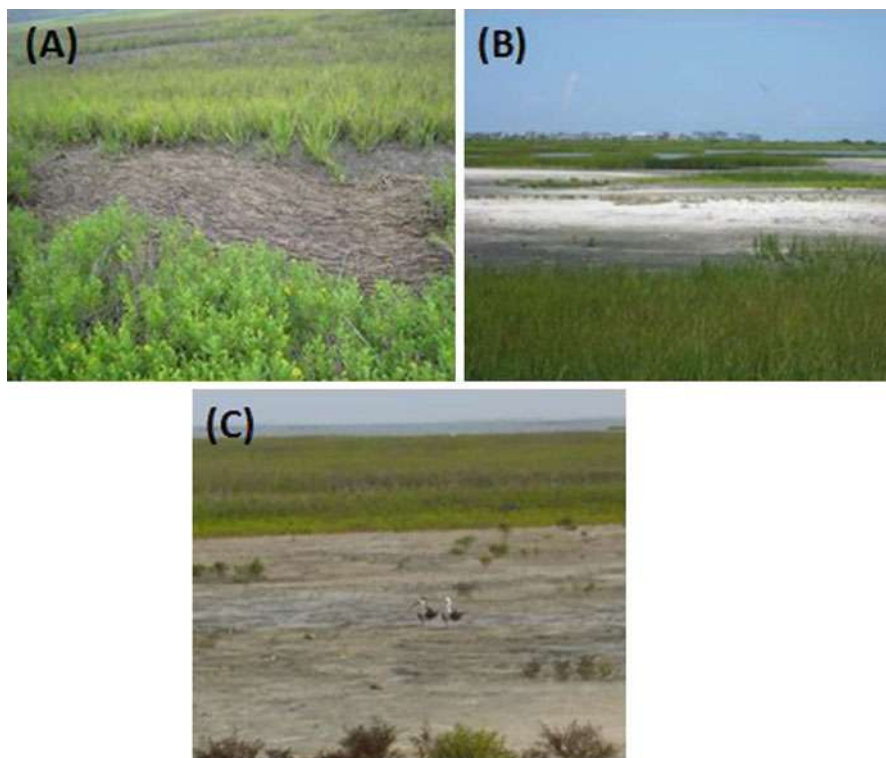
Salt marshes in New England are dominated by smooth cordgrass, *Spartina alterniflora*. This species is particularly well adapted to frequently flooded low





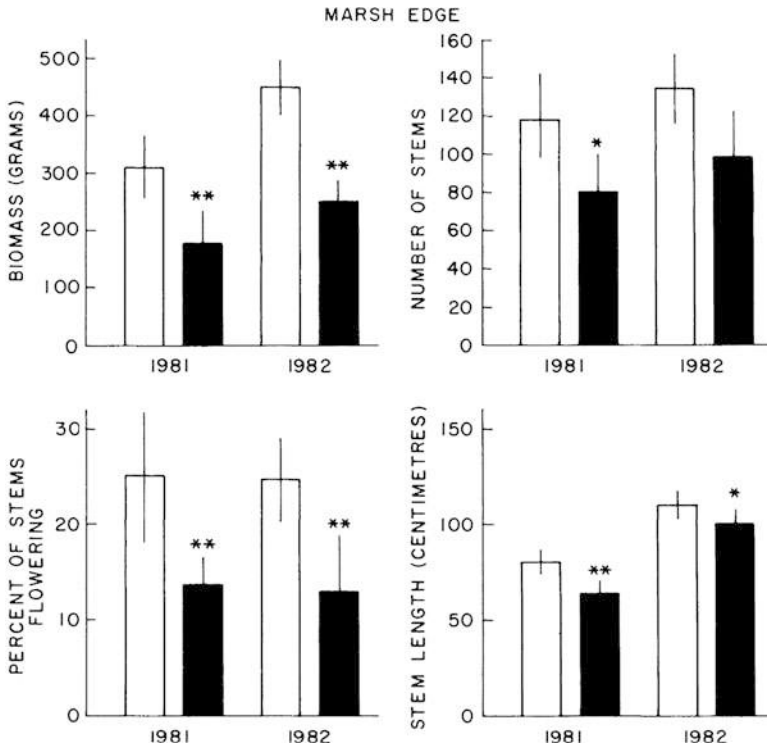
**Fig. 7** Tall and short forms of *Spartina alterniflora* (Photo credit A.R. Armitage)

elevations, where it co-occurs with several marsh fauna species. The marsh grasses and fauna have a close **facultative mutualistic** relationship, where each benefits from the other, though they do not completely rely on each other for survival. One common faunal group in salt marshes is comprised of fiddler crabs (*Uca* spp.), which excavate extensive burrows. In a set of experiments, Bertness (1985) removed crabs from high-density, low-elevation zones and added crabs to low-density, high-elevation zones. These experiments revealed several mutualistic interactions between crabs and smooth cordgrass. Crab burrowing activity oxygenates the sediment, augments drainage, and increases the decomposition of organic matter, all of which increase smooth cordgrass above- and belowground productivity. Crabs benefit from this association as well – smooth cordgrass roots substantially increase the integrity of crab burrows. This positive feedback between smooth cordgrass and fiddler crabs is strongest within the low marsh elevation, just above the marsh vegetation-water interface. Burrows excavated at the marsh edge, in softer, wetter sediment with few roots, will rapidly collapse. At high marsh elevations, denser root mats interfere with the ability of fiddler crabs to excavate burrows. Therefore, the strength of the fiddler crab-smooth cordgrass facilitation is greatest at the upper edge of the low marsh, where there is a maximized mutual benefit for plants (anoxia stress is alleviated) and crabs (burrow integrity is increased).



**Fig. 8** (a) Wrack deposition in the high marsh zone of a salt marsh. Wrack has accumulated between stands of *Borrichia frutescens* (sea oxeye daisy, with yellow flowers) and short-form *Spartina alterniflora*. Previously covered patches that have turned into salt pans are visible in the background. (b) Fully formed salt pan with sparse succulent vegetation and cyanobacterial mats (visible as blackened patches on the soil). (c) *Black skimmers* (*Rynchops niger*) roosting in a salt pan (Photo credit A.R. Armitage)

Another common animal in New England salt marshes is the ribbed mussel (*Geukensia demissa*). These bivalves require a surface for attaching anchoring filaments, and smooth cordgrass stems and roots provide a suitable substrate (Bertness 1992). Mussels can be particularly dense along the seaward edge of the tall smooth cordgrass zone. The anchoring filaments bind smooth cordgrass stems together, which in turn increases sediment stabilization and decreases erosion. Mussels deposit waste products that provide nutrients for plant growth (Jordan and Valiela 1982), resulting in increased aboveground and belowground productivity (Fig. 9; Bertness 1984). Mussels also benefit from this association – mussel growth and survivorship is higher for mussels in smooth cordgrass beds (Stiven and Kuenzler 1979). Smooth cordgrass benefit mussels by providing an attachment substrate and may also supply organic matter as an indirect food source (Bertness 1984).



**Fig. 9** Excerpt from Fig. 3 in Bertness (1984). Summary of aboveground *Spartina alterniflora* parameters in mussel manipulation experiments done on the marsh edge during the 1981 and 1982 growing seasons. Control quadrats; mussel removal quadrats ( $\pm$ SE) (All data are for 0.25-m<sup>2</sup> quadrats). \* $P < .05$ , ANOVA in comparison to control within years. \*\* $P < .01$ , ANOVA in comparison to control within years (Reprinted with permission from the Ecological Society of America)

### Summary: Salt Marshes

In summary, zonation in salt marshes is driven by abiotic stressors, interspecific competition, and facultative mutualistic plant-animal interactions. The variation in the relative importance of these factors across salt marsh elevation zones is summarized in the conceptual model in Fig. 6. In the low-elevation zone, prolonged inundation and associated soil anoxia limit plant assemblages to a few species, though facilitative plant-animal interactions somewhat ameliorate this stress. Salinity stress is the primary abiotic stressor at higher elevations. Many of low-elevation plant species can survive at higher, less stressful elevations, but are competitively excluded from those less stressful habitats. This pattern was succinctly described by ecologist Mark Bertness (1991): "Zonation patterns are maintained by competitive dominants restricting the distribution of competitive subordinates to physically stressful habitats."

## Mangroves

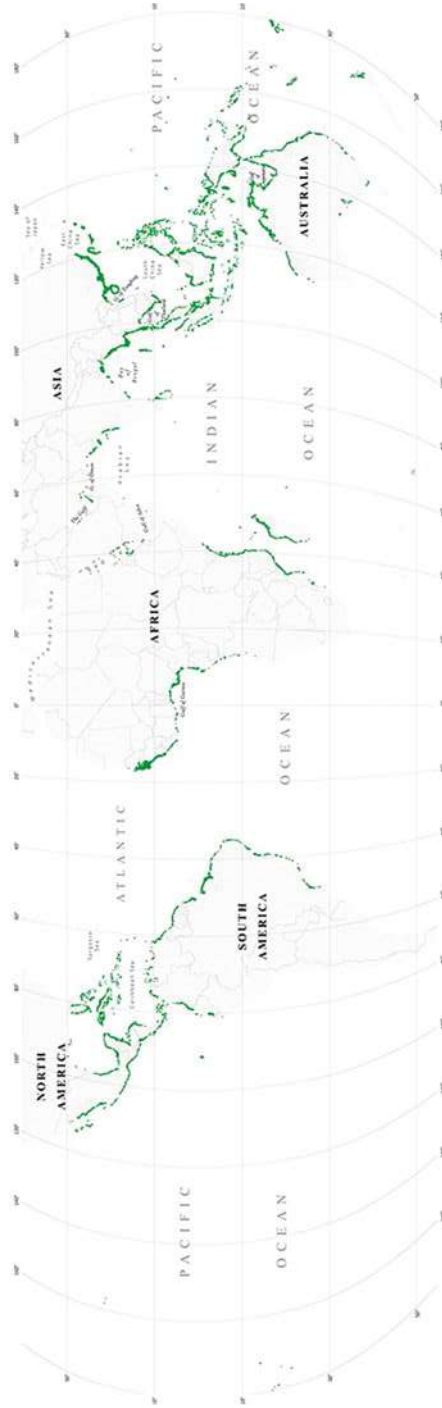
Mangrove swamps are dominated by **halophytic** (salt tolerant) trees that live at the land-sea interface. In an example of convergent evolution, mangrove species evolved from non-mangrove plant lineages independently many different times. In fact, mangroves occur in over 30 families of dicots (class Magnoliopsida). Therefore, trees that are called “mangroves” are not necessarily closely related in an evolutionary sense. Mangrove species differ in their stress adaptations and in their degree of stress tolerance. However, most mangroves are intolerant of freezing temperatures, which limits their distribution to tropical and subtropical latitudes (Fig. 10; Giri et al. 2011). There are over 65 species of mangroves worldwide, with the highest diversity in the Indo-Pacific and Indian Oceans; about four species occur in North America and the Caribbean.

### Mangrove Stress Adaptations

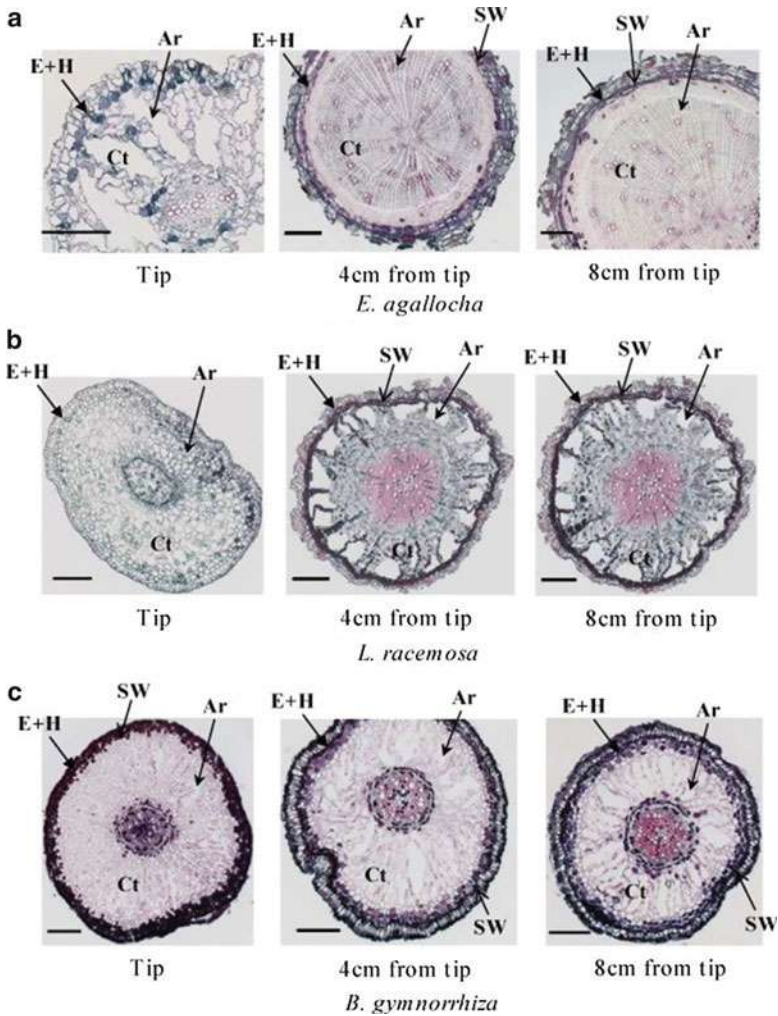
Mangroves and salt marshes experience similar abiotic stressors, particularly high salinity and prolonged flooding. In addition to the general adaptations described earlier, many mangroves have specialized structural modifications that facilitate survival in these harsh tidal coastal environments.

To adapt to saline waters, the roots of many mangroves are **suberized**. **Suberin** is an extracellular glycerolipid polymer found in the cell walls of many plant species. In plant roots, suberin is found at the hypodermis, where it blocks **apoplastic** (extracellular) transport into the root, and at the endodermis, where it limits transport into the stele. In mangroves, it is particularly concentrated in the epidermis and hypodermis of roots (Fig. 11; Pi et al. 2009). It forms a thick, waxy layer that effectively blocks apoplastic salt uptake by the plant; in some species, over 90 % of the salt in seawater can be excluded by this substance. The suberin layer also reduces radial oxygen loss from the roots (Pi et al. 2009) and may lower transpiration rates and increase water-use efficiency (Baxter et al. 2009).

Another adaptation to salinity found in about a third of all mangrove species is **vivipary**, which is a reproductive strategy where there is substantial development of the zygote while still attached to the parent tree. In some mangrove species, the seed embryo will penetrate through the fruit pericarp and grow to a considerable size before dispersal, producing characteristic **propagules** with elongated **hypocotyls** (embryonic trunks) (Fig. 12a). In other species, the zygote does not penetrate the **pericarp** (fruit wall) before dispersal, but the hypocotyls will emerge shortly after release from the parent (Fig. 12b). Among dicots, true vivipary – sexual development on the parent tree – is relatively rare and occurs mostly in mangroves. About 30 of the 33 plant species known to exhibit true vivipary are mangroves (Elmqvist and Cox 1996). **Pseudovivipary** – asexual development on the parent tree – occurs in several other groups of plants in extreme climates with high abiotic stress levels, such as deserts or alpine environments (Elmqvist and Cox 1996). In all cases, this jump start on seedling development helps protect young plants from the abiotic stressors in the environment



**Fig. 10** Excerpt from Fig. 1 in Giri et al. (2011). Mangrove forest distributions of the world – 2000 (Reprinted with permission from Blackwell Publishing Ltd)



**Fig. 11** Excerpt from Fig. 4 in Pi et al. (2009). Cross sections of root tip, basal zone (4 cm from the root tip), and mature zone (8 cm from the root tip) of *Excoecaria agallocha*, *Lumnitzera racemosa*, and *Bruguiera gymnorhiza* (cross sections with thickness of 10  $\mu\text{m}$  were made and photographed, scale bars equal to 200  $\mu\text{m}$ ; E+H epidermis and hypodermis, Ar aerenchyma air spaces, Ct cortex, SW suberized walls) (Reprinted with permission from Elsevier BV)

by facilitating rapid establishment soon after dispersal. In mangroves, vivipary protects new, vulnerable seeds from salt water stress, allows nutrient uptake from the parent plant under low salt stress, and reduces chloride inhibition of germination. Propagules can float after being released from the parent tree, facilitating long-distance dispersal. Rooting is initiated when favorable habitat is encountered.

A striking morphological characteristic of many mangroves is their complex **aerial root** structures, which primarily function as adaptations to flooded conditions.





**Fig. 12** (a) Propagules of the red mangrove, *Rhizophora mangle*, still attached to the parent tree. (b) Rooted propagule of the black mangrove, *Avicennia germinans* (Photo credit A.R. Armitage)

Aerial roots that extend from the mangrove trunk are termed **prop roots**, and those that protrude upward from lateral belowground roots are called **pneumatophores** (Fig. 13). The aerial portions of these “roots” are covered with large pores called **lenticels**. Air is taken up through the lenticels and transported through the **aerenchyma** tissue to the belowground root system (Fig. 11), thus delivering the oxygen necessary for root cellular metabolism in otherwise hypoxic or anoxic soils.

## Zonation

In concept, intertidal zonation patterns are dictated by physiological responses of each species to abiotic stressors that vary along tidal gradients. Mangroves are somewhat plastic in their internal and external morphology, so some species can occur at a range of elevations, and zonation patterns are variable within and among geographic regions of the world. A wide variety of factors, including shoreline topography, tidal and freshwater influence, salinity, and sediment characteristics, influence mangrove distribution along elevation gradients. Thom (1984) identified no fewer than eight distinct geomorphic and biological settings that have unique mangrove zonation patterns. This section will focus on some of the most common types of mangrove tidal “zones,” with specific emphasis on the species common to Caribbean mangrove swamps.

The land-sea interface, often referred to as **fringe** mangrove habitat, is characterized by permanently flooded soils, giving the plants constant exposure to salt water.

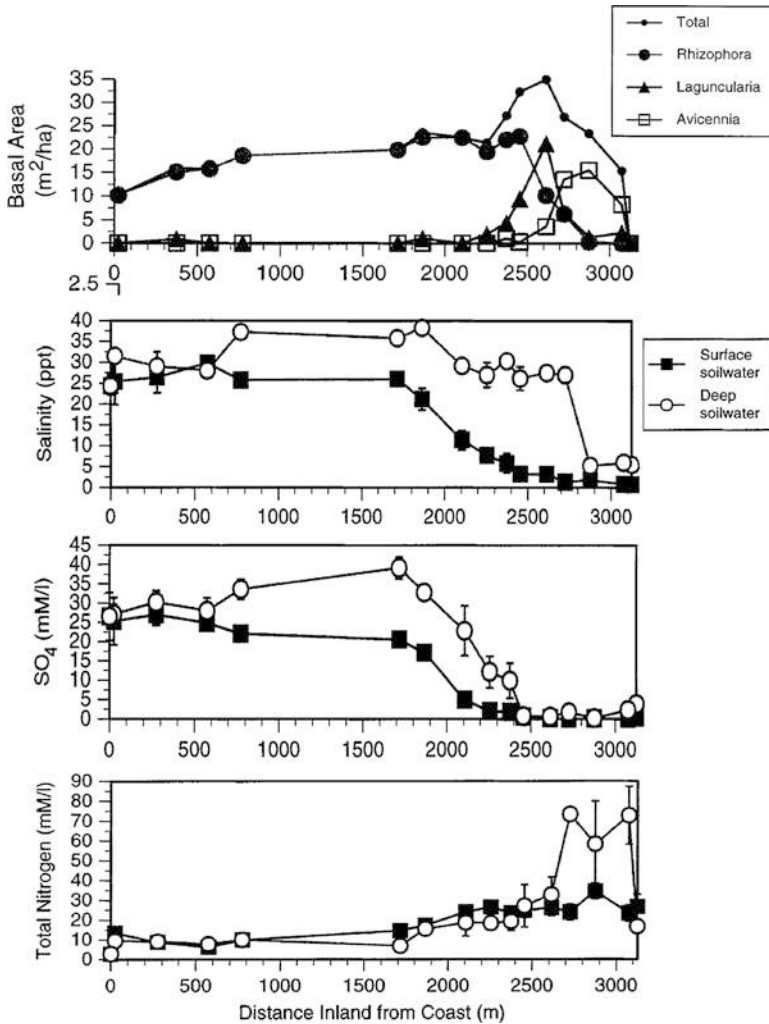


**Fig. 13** Mangrove aerial root structures. (a) Prop roots on a juvenile red mangrove, *Rhizophora mangle*. (b) Pneumatophores extending upward from lateral roots of a juvenile black mangrove, *Avicennia germinans* (Photo credit A.R. Armitage)

The soils generally have low oxygen content, though they are not necessarily anoxic (McKee 1993). Oxygenic phototrophs such as diatoms and other eukaryotic algae inhibit nitrogen fixation, thereby maintaining low soil nitrogen content in fringe mangrove soils (Fig. 14; Lee and Joye 2006). This aerobic activity also facilitates sulfide oxidation, reducing the buildup of toxic sulfides (Fig. 14; Sherman et al. 1998). In the Caribbean, the red mangrove (*Rhizophora mangle*) dominates this fringe habitat. With its characteristic, prominent prop roots (Fig. 13a), red mangroves form an iconic image of the Caribbean coastline. Prop roots are covered with lenticels and contain aerenchyma tissue, enabling red mangroves to survive in permanently flooded soils. Red mangroves also have heavily suberized roots that can block up to 99 % of salt uptake from the flooding seawater. The long, thin propagules characteristic of red mangroves (Fig. 12a) are an additional adaptation to the salt water environment.

The zone above the fringe habitat is difficult to succinctly characterize. In some areas, this zone is called a **transition** habitat that contains a mix of species. In other areas, this drier habitat is called a **basin** habitat and is dominated by just one or two species. In general, the flooding duration in mid-elevation habitats is relatively short, facilitating the diffusion of oxygen from the atmosphere into the soils. As in the fringe habitat, nitrogen and sulfide accumulation rates are relatively low (Fig. 14). The shorter flood periods allow mangroves in this zone to have somewhat reduced aerial root structures. In the Caribbean, black mangrove (*Avicennia germinans*) is characteristic of this zone. The pneumatophores of this species can





**Fig. 14** Excerpt from Figs. 1, 3, and 5 in Sherman et al. (1998). Changes in mangrove and soil characteristics with increasing distance from the shoreline (Reprinted with permission from Springer-Verlag)

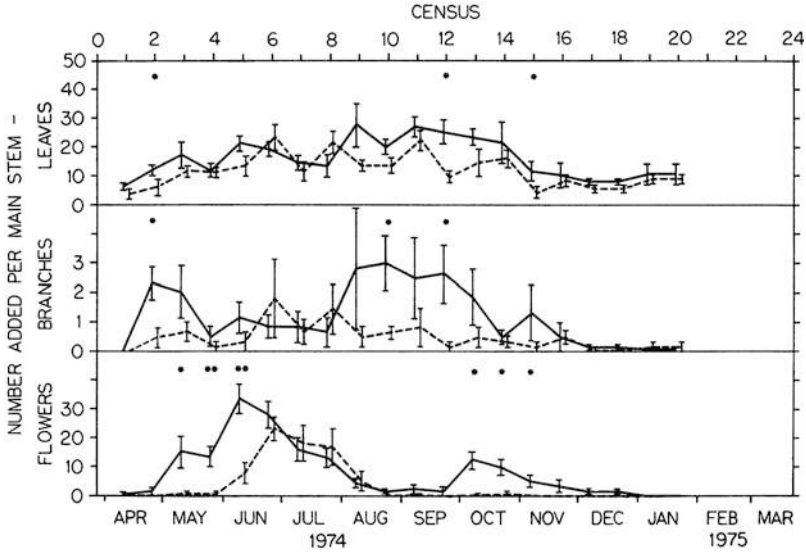
extend upward out of the ground for several meters away from the primary tree trunk (Fig. 13b). Like prop roots, pneumatophores have aerenchyma and lenticels to facilitate gas exchange and root aeration. Black mangrove roots are suberized, but not as heavily as red mangrove roots. Black mangroves manage excess salt uptake by secreting salt through numerous small salt excretion glands scattered across leaf surfaces. The production of small but numerous propagules (Fig. 12b) facilitates seedling survival in saline soils.

The highest elevations in mangrove swamps sometimes transition to terrestrial or freshwater habitat, but in other cases, they are characterized as **dwarf** mangrove habitat. Dwarf habitat is essentially basin habitat that is so infrequently flooded or is otherwise abiotically stressful that the trees are stunted in height. In these habitats, soils are generally anoxic, facilitating sulfate reduction and the accumulation of sulfides (Fig. 14). Nitrogen fixation also occurs in the anoxic soil, increasing total nitrogen concentration in the soil. If this habitat is occasionally tidally influenced, then the soils will be saline. In general, the duration of flooding is relatively short, so mangroves at this elevation have more adaptations for managing salt than for flooding. In the Caribbean, white mangroves (*Laguncularia racemosa*) are characteristic of this zone, though they can occur at lower elevations as well. White mangroves can develop small pneumatophores or reduced prop roots if prolonged flooding occurs, but are frequently found at higher elevations and without aerial roots. White mangroves usually occur in saline soils, so they have moderately suberized roots and large salt excretion glands on the leaves. Like many other mangrove species, white mangroves produce propagules to reduce salt stress on seedlings.

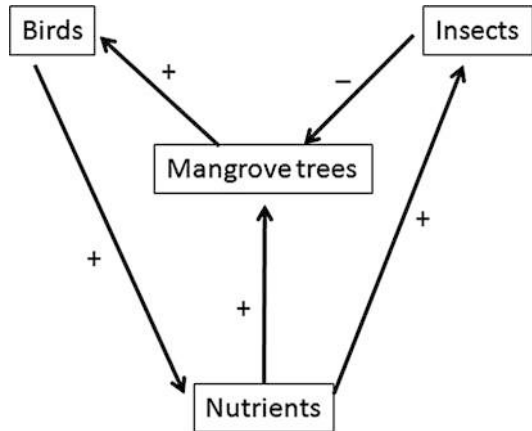
### **Case Study: Plant-Animal Interactions on Mangrove Islands in Florida**

Small islands with dense stands of red mangroves are common along Caribbean and Florida coastlines. Some of these islands are used as rookeries by nesting birds (e.g., herons, egrets, pelicans, cormorants). During the nesting season, copious amounts of guano (bird feces) are deposited on the rookery islands, and mangroves take up some of the excess nutrients. Trees on the enriched rookery islands produce more branches and flowers than trees on non-rookery islands (Fig. 15; Onuf et al. 1977). This example illustrates how important nonconsumptive relationships can be in structuring plant communities. In this case, mangroves provide birds with nesting habitat, and the birds benefit the plants by supplying nutrients for growth. The indirect interaction between birds and plants demonstrates that bottom-up forces, in this case resource supply, can influence both plant and bird fitness and productivity (Fig. 16).

The plant-animal interactions in this community become more complex when other community members, such as insects, are considered. Leaf production on trees in rookeries is not always augmented as much as might be expected based on the amount nutrient supply from guano. This is largely due to higher herbivory pressure on rookery islands – insects prefer the guano-enriched leaves, and herbivory can be up to four times higher than on non-rookery islands. Ultimately, increased mangrove productivity from nutrient enrichment is mitigated by nutrient-induced herbivory. This case study shows how complex interactions between bottom-up (resource availability, e.g., nutrient supply) and top-down (consumption, e.g., herbivory) forces can structure plant communities (Fig. 16).



**Fig. 15** Excerpt from Fig. 4 in Onuf et al. (1977). Mean numbers ( $\pm$  SE) of leaves, branches, and flowers added per 1-cm diam. main stem in high- (solid line) and low- (dashed line) nutrient areas. Differences between sites were significant by *t*-tests ( $df = 10$ ) for dates where \* ( $p < .05$ ) or \*\* ( $p < .01$ ) appear in the upper part of the figure (Reprinted with permission from the Ecological Society of America)



**Fig. 16** Simplified conceptual diagram depicting the interaction between top-down and bottom-up forces influencing mangroves on islands that are used as rookeries

### Future Directions: The Salt Marsh-Mangrove Ecotone: A Developing Field

Mangroves are not tolerant of freezing temperatures; this temperature sensitivity limits mangroves to tropical and subtropical latitudes (Fig. 10). Many families of salt marsh species can tolerate a wide range of weather conditions, but on

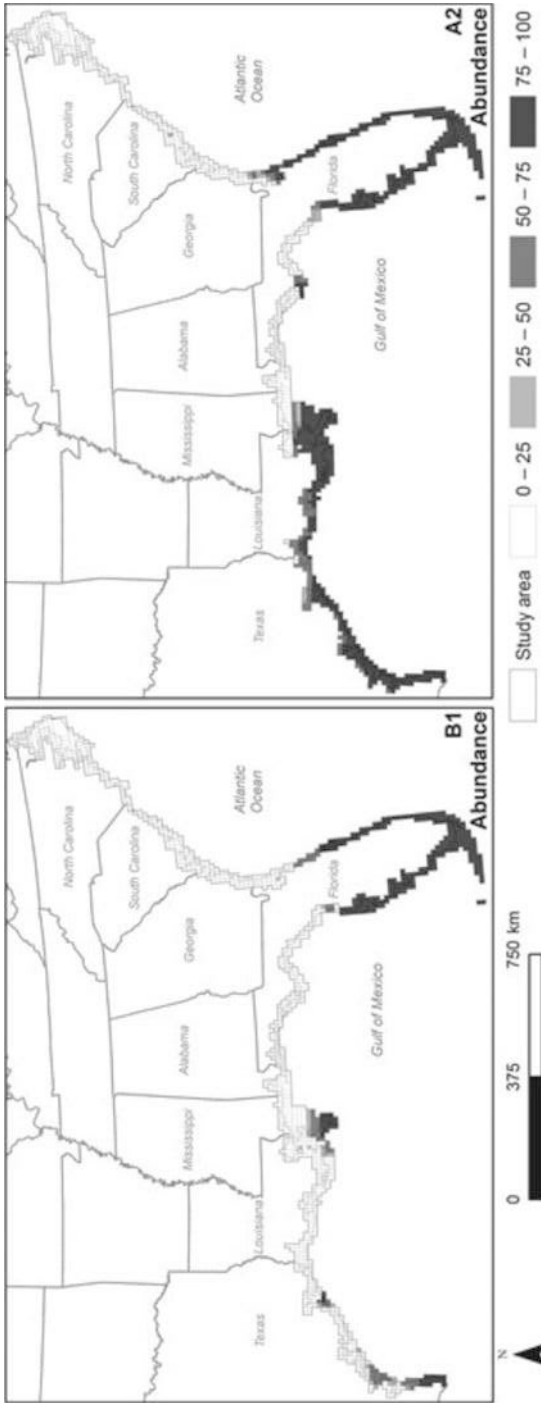
tropical coastlines, smaller salt marsh species are outcompeted by dense, tall mangrove canopies. In some subtropical areas, there is a transition zone – an **ecotone** – between marsh and mangrove habitats. These ecotones occur in temperate areas of Australia, New Zealand, and the southern continental United States. Mangrove-marsh ecotones are dynamic habitats – mangroves often expand into salt marshes during periods with warm winters and contract during periods with hard freezes. This dynamic is primarily driven by temperature, but many other factors influence mangrove-marsh distribution as well, including rainfall, salinity, sea level, propagule supply, and interspecific competition. For example, *Spartina alterniflora* can outcompete newly sprouted black mangrove propagules (McKee and Rooth 2008), but if the mangrove seedlings survive through a few growing seasons, the established tree will begin to displace the surrounding marsh grasses and forbs.

Current research suggests that mangrove distributions may continue to expand in response to climate change. For example, models predict that an increase in winter minimum temperatures of 2–4 °C may lead to black mangroves replacing salt marsh on nearly all of the Texas and Louisiana coastlines by the year 2100 (Fig. 17; Osland et al. 2013). Other climate-related factors that may increase mangrove expansion rates include rising sea level due to glacial melting and thermal expansion. As little as 10 cm of sea level rise over the next 100 years will likely result in substantial mangrove expansion in all Gulf of Mexico states; sea level rise may cause mangroves to displace over 10,000 ha of coastal marsh in both Florida and Louisiana (Doyle et al. 2010). Climate change scenarios that include increasing atmospheric carbon dioxide concentration and changing herbivore populations will also likely influence mangrove-marsh dynamics, though these interactions are complex. Elevated CO<sub>2</sub> alone may not be sufficient for mangrove seedlings to outcompete marsh plants, but if there is also low herbivory pressure and sufficient nitrogen supply, then elevated CO<sub>2</sub> may accelerate mangrove growth (McKee and Rooth 2008). The exact role of each of these factors, and how they interact with each other, is a rapidly growing field of study in coastal plant ecology. Furthermore, appropriate management of coastal resources depends on our understanding of the ecological implications of this shift in plant communities. Will this change in plant species composition alter the **ecosystem services** that wetlands provide, such as fishery nurseries, erosion control, or water quality improvement? Key ecosystem services of coastal plant habitats are described in the next section.

---

## Ecosystem Functions and Services

Coastal plant communities provide a unique suite of ecosystem functions and associated ecosystem services. **Ecosystem functions** are characteristic exchanges or processes within an ecosystem, such as primary productivity, energy flow, or nutrient cycling. **Ecosystem services** are ecosystem functions that benefit human-kind. Human valuation of ecosystem functions is complex and based on many



**Fig. 17** Excerpted from Fig. 6 in Osland et al. (2013). Predictions of mangrove forest relative abundance (i.e., percentage of tidal saline wetlands dominated by mangrove forests) under alternative future (2070–2100) winter climate projections: left panel, mangrove forest relative abundance with an ensemble *B1* scenario climate; right panel, mangrove forest relative abundance with an ensemble *A2* scenario climate. Note that these predictions apply just to the tidal saline wetland habitat within each cell and not the entire cell. Climate scenarios are defined by the Intergovernmental Panel on Climate Change (Reprinted with permission from Blackwell Publishing Ltd.)

factors, including the provision of food for sustenance, monetary gain, aesthetic value, and clean air and water. Several ecosystem functions of coastal plant communities that are particularly valued by humankind are highlighted in the following section.

## Water Quality

Coastal plant communities are widely recognized for their capacity to improve nearshore water quality. This plant-mediated improvement of water quality is termed **phytoremediation**. Coastal wetlands are not stagnant water bodies – many have slow but directional water flow from inland sources to nearshore habitat. Some wetlands are specifically constructed to manage water flow between terrestrial and marine ecosystems – these are called **treatment wetlands**. The plants in natural and treatment wetlands provide frictional resistance, slowing down water flow, thus facilitating the removal of nutrients, bacteria, and other pollutants through a variety of mechanisms. When wetland plants lower water velocity, this facilitates the **settlement** of suspended solids and adhered contaminants. Settlement is the primary mechanism for removal of organic solids (i.e., sewage waste) from water moving through coastal wetlands. Many nutrients, especially ammonium, nitrate, and phosphate, can be removed from the water through direct **uptake** by plants and bacteria, which then use these nutrients for metabolic processes. Bacteria in wetland soils can transform ammonium into nitrate (**nitrification**) and then into  $N_2$  gas (**denitrification**). Nitrogen gas can then **volatilize** (evaporate or diffuse) from the water into the atmosphere. Some nutrients, particularly inorganic forms of phosphorus, can become tightly bound to clay particles in a process called **adsorption**. These phosphorus-clay complexes are largely biologically inert, and as the clay particles settle to the benthos, the phosphorus is functionally removed from the water column.

## Nutrient Cycling and Storage

Coastal wetlands play critical roles in many global nutrient cycles; nitrogen and carbon cycles are among the most important (Vitousek et al. 1997). The anoxic soils in coastal wetlands harbor **nitrogen-fixing** bacteria that convert atmospheric nitrogen (nitrogen gas,  $N_2$ ) to organic forms such as ammonium. Nitrogen fixation is the primary mechanism by which inert pools of atmospheric nitrogen become biologically available for plant uptake. Other bacteria in wetlands facilitate **denitrification**, the conversion of nitrate to nitrogen gas ( $N_2$ ). Denitrification is important for controlling the export of organic nitrogen out of wetlands – without denitrification, nitrogen will stay in biologically available organic forms. Excess nitrogen will eventually be exported out of wetlands through rivers and streams to nearshore ecosystems, potentially exacerbating **eutrophication** (see “[Management Issues and Strategies](#)” section).

Coastal wetlands also play an important role in the global carbon cycle, particularly given their potential for **carbon sequestration**. Carbon sequestration occurs when carbon assimilation is greater than carbon loss in an ecosystem. In marine environments, including coastal wetlands, sequestered carbon is referred to as **blue carbon**. Mechanisms of carbon assimilation in wetlands include photosynthesis, soil microbe assimilation, and the decomposition and burial of plant tissue. Carbon is lost from wetlands through microbial and plant respiration and through the decomposition and export of plant tissue into adjacent waterways. Natural and **anthropogenic** (human-caused) wetland loss can accelerate carbon loss and reduce sequestration potential. Changes in wetland vegetation – such as the shift from marsh- to mangrove-dominated systems – may also alter the blue carbon storage potential in wetlands; the nature of these potential changes is a currently growing field of study.

## Erosion Control and Surge Buffer

Many regions of the world are prone to large, powerful hurricanes. The east and Gulf coasts of the United States have been hit by several particularly damaging hurricanes over the last 10 years. Storm damage to coastal urban communities can potentially be lessened, to a degree, by the fringing coastal marsh and dune ecosystems. Coastal plant communities can **attenuate** (reduce) storm surge through wave energy *dispersal*, where the physical structure of the plant assemblage breaks up wave energy. In addition, wetlands can *store* large amounts of water, subsequently reducing the amount of water that travels inland to urban communities. For example, the storm surge from Hurricane Katrina, which struck New Orleans, Louisiana in 2005, entered the coastal salt marshes in less than 24 h, but after the storm, it took more than 4 days for all of the surge waters to drain back out into the Gulf of Mexico.

A commonly used rule of thumb is that each 2.7 miles of marsh (from the coastline extending inland) attenuates storm surge by 1 ft. The actual attenuation rate and inland extent of the storm surge varies among and even within storms and is influenced by the geometry of the shoreline, vegetation type, slope of ocean floor, and, perhaps most importantly, the size, speed, direction, and duration of storm. Hurricane Rita, which struck the Louisiana-Texas border as a Category 3 storm in 2005, is an excellent example of how variable real-world attenuation rates can be. Due to the track of the storm, western Louisiana experienced among the highest wind speeds (up to 120 mph), but high winds persisted for a relatively short duration. The attenuation of storm surge in the area was close to the 2.7 miles: 1 f. prediction. Eastern Louisiana, however, was exposed to the powerful winds in the northeast quadrant of the storm for nearly a full day. Even though the maximum wind speed was lower (about 90 mph), the duration of the exposure to hurricane-force winds was much longer than in West Louisiana. Coastal marshes in this area became inundated, and their attenuation capacity was essentially nullified.

Wave attenuation in mangrove forests is similarly variable. Although it is currently popular to promote mangroves as “bioshields” against storm surges and tsunamis, their role in actually reducing human casualties in the face of natural catastrophes remains somewhat controversial. Mangrove trees can undoubtedly reduce wave intensity through friction and wave disruption, but the effect is probably limited to relatively small waves. For catastrophic wave inundation events associated with tsunamis or large cyclones, most quantitative studies suggest that the risk of damage to a coastal settlement is more closely linked to distance from the shoreline than to the presence or absence of a mangrove forest (e.g., Gedan et al. 2011).

In addition to providing occasional protection against catastrophic erosion and flooding, coastal wetlands also provide day-to-day protection to the built environment on smaller spatial scales. Urban areas with higher levels of permitted wetland alteration have more frequent flooding following precipitation events. In fact, the number of permitted alterations may be a stronger predictor of flooding risk than watershed characteristics like area, slope, or population density (Brody et al. 2007).

## Nursery Habitat

A wide range of commercially and recreationally important fish and invertebrate species rely on coastal wetlands, especially salt marshes, for part or all of their life cycle. In fact, over 75 % of commercially and recreationally targeted fishery species spend at least part of their life cycle in estuarine wetlands. For example, red drum (*Sciaenops ocellatus*) is a popular sport fish on the Atlantic and Gulf coasts of the United States. This fish spawns in nearshore habitats. Larvae and juveniles reside in estuaries, foraging on small shrimp, crabs, and other larval fish in salt marshes at high tide. Shrimp fisheries are also dependent on salt marshes. In the Gulf of Mexico, brown (*Farfantepenaeus aztecus*) and white shrimp (*Litopenaeus setiferus*) spawn at sea but inhabit *Spartina alterniflora* or *Juncus* spp. marshes in the postlarval (non-planktonic) stage. These shrimp fisheries are most productive in areas with extensive estuarine marshes, like the Mississippi Delta.

## Recreation

Wetland plants provide habitat for many species of animals beyond those that directly contribute to commercial fisheries. Many recreationally fished species also rely on coastal wetlands. In the Gulf of Mexico, for example, over 80 % of recreationally targeted species spend at least some of their life in estuarine wetlands. Coastal wetlands also provide critical stopover and wintering habitat for migratory birds: on a typical winter day in any given coastal wetland in Baja California, 5,000 or more migratory shorebirds may be spotted. Some coastal



wetlands provide essential habitat for endangered species, such as the whooping crane (*Grus americana*), which forages exclusively in salt marshes in Texas in the winter. While enjoying these diverse and abundant wildlife populations, recreational fishers and birders contribute billions of dollars to coastal economies each year. In 2006, a typical year, birders alone contributed to \$82 billion in total industry output to the United States economy, primarily through purchases of lodging, transportation, food, and equipment (Carver 2009).

---

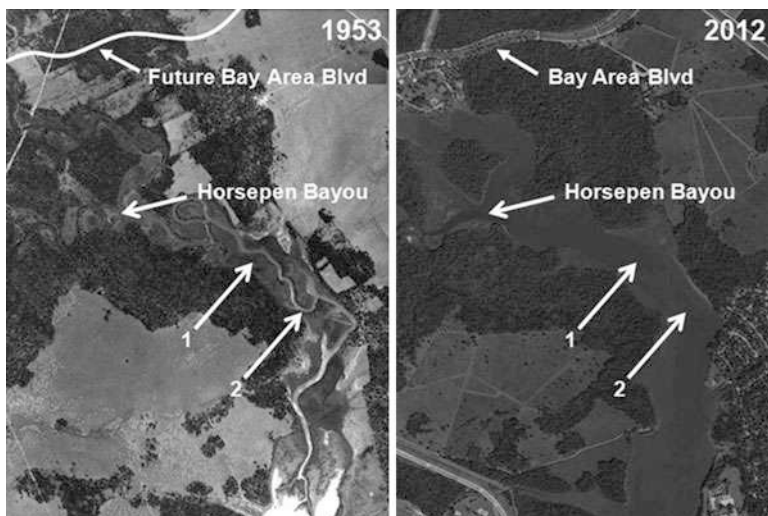
## Management Issues and Strategies

Coastal wetlands have been substantially reduced in area over the past 200 years, and many remaining wetlands are impacted or degraded. In the continental United States, almost every state has lost at least a quarter of its historical wetland area; much of this loss has occurred in coastal wetlands. This section will include a discussion of the major mechanisms of wetland loss and impacts on remaining wetlands and will conclude with a brief discussion of the dynamic and sometimes controversial legal policies that protect coastal wetlands.

## Development

Modern civilization, and accompanying urban and agricultural development, has dramatically altered coastal ecosystem landscapes. Some wetlands are filled for urban development; other developments occur in upland habitats directly adjacent to wetlands. The higher elevations of mangrove swamps are sometimes cleared to create room for urban growth or resort communities. Other mangrove swamps are cleared and excavated to create room for mariculture ponds to grow shrimp or other farmed seafood resources. Coastal marshes have been diked or drained to create agriculture fields or livestock grazing habitat; other marshes are flooded for rice farming.

In addition to directly causing habitat loss, development also increases groundwater use, which can accelerate **subsidence**. Subsidence is the gradual lowering of the sediment surface through mechanisms such as sediment compaction. Natural subsidence occurs slowly and is usually mitigated by the accumulation of sediment that enters estuaries from rivers. However, subsidence rates can be greatly exacerbated by anthropogenic activities, especially the withdrawal of groundwater. A particularly striking example of anthropogenic subsidence was documented around Houston, Texas, in the 1970s. A booming oil industry spurred population growth in the area, driving up the industrial and residential demand for groundwater. Rapid withdrawal of groundwater accelerated subsidence, and over a period of less than 10 years, many neighborhoods sunk more than half a meter. Some localized spots sank even more – up to 3 m (Fig. 18). This rapid subsidence permanently inundated tidal marshes, causing over 95 % marsh loss in a very short time period. Entire neighborhoods had to be abandoned due to chronic flooding problems.



**Fig. 18** Google Earth images of Armand Bayou (near Houston, TX) in 1953 and 2012. In the 1953 image, note the tidal marshes in Horsepen Bayou and at marker #1 and the narrow tidal channel at marker #2. By 2012, subsidence had flooded most of those features. Ongoing restoration work in Horsepen Bayou is reestablishing some of the tidal marsh features

By the mid-1980s, a wider municipal and public appreciation of the anthropogenic subsidence issue led to better management of groundwater, and subsidence rates slowed dramatically.

## Sea Level Rise

Although anthropogenic subsidence rates in many estuaries in the United States have slowed, coastal wetlands are also subject to inundation from sea level rise, which is partly driven by climate change. **Relative sea level rise** in any one particular place is determined by both **eustatic** (global) and regional changes in sea level. Eustatic changes are related to climate change, including thermal expansion and glacial melting. Regional changes in sea level are linked to local dynamics like subsidence and riverine sediment supply. Most conservative estimates, as synthesized by the Intergovernmental Panel on Climate Change, suggest that at least 50 cm of relative sea level rise will occur in many coastal regions by 2100 (IPCC 2007).

Prior to human development of the coastline, wetlands could respond to sea level rise by migrating, albeit slowly, upland. However, many coastlines are now heavily developed or “hardened”; roads, bulkheads, and other built structures, as well as natural topographic features, prevent landward migration. As a result, many wetlands are experiencing a “**coastal squeeze**,” where wetland area shrinks as rising seas and anthropogenic barriers to upland migration limit the area of elevation suitable for wetland plant growth.

## Freshwater Diversion

Recall the concept of an estuary: a body of water where fresh and salt water mixes. Many estuaries are parts of heavily developed **watersheds**, which are the areas encompassing all the lakes and rivers that eventually drain into a large water body. Demands for fresh water from urban and agricultural developments ultimately reduce freshwater input to the estuaries. What happens to an estuary when freshwater inflows decrease? The most acute impact, arguably, is an increase in salinity. These increases in salinity are likely to be exacerbated by extreme environmental events phenomena like droughts (Fig. 1). Long-term effects of high salinity could include plant or animal die-offs or shifts toward more marine species assemblages.

## Eutrophication

Plant productivity in most ecosystems is limited by particular nutrients – those nutrients that are in shortest supply relative to others, and will therefore limit organism growth. In pristine coastal habitats, **nitrogen** and **phosphorus** are typically the most limiting. There are many anthropogenic sources of these limiting nutrients, including fertilizer runoff, sewage, and livestock waste. Moderate input of anthropogenic nutrients can increase ecosystem productivity, but excessive nutrient input can cause **anthropogenic eutrophication**: the rapid buildup of organic matter. In salt marshes, plants respond to excess nutrients by accelerating aboveground production: this produces the excess organic matter that is characteristic of eutrophic conditions. However, increased aboveground production is typically matched by a decrease in belowground production (Deegan et al. 2012). Lower root biomass is linked to decreased sediment stability, which eventually results in marsh erosion and habitat loss.

## Policy

Wetlands are currently the only ecosystem with an international agreement focused on conservation and sustainable utilization. This agreement, the Ramsar Convention, was formed in 1971 by conservation groups in Europe that recognized the ecological and economic implications of widespread wetland loss. Currently, at least 163 nations are members of the convention. Central to the Ramsar Convention is the “wise use” concept: wetlands should be conserved and sustainably used for the benefit of humankind. Although the Ramsar Convention has no regulatory power, it has helped nations identify conservation priorities and define management strategies.

In the spirit of the Ramsar Convention, George H.W. Bush adopted a “**No Net Loss**” policy for the United States in 1989. The essence of this policy is that for every one acre of wetlands that is lost, at least one acre must be created or restored in its place. This policy applies specifically to **jurisdictional** wetlands, which are

**Fig. 19** A young volunteer helps the Galveston Bay Foundation plant smooth cordgrass (*Spartina alterniflora*) in a restored marsh in Galveston Bay, TX (Photo credit A.R. Armitage)



generally defined as those wetlands that fall under federal or local protection, based on the 1977 Section 404 amendment to the Clean Water Act (Kruczynski 1990). The definition of jurisdictional wetlands has narrowed and widened at times in response to sometimes contentious disputes among landowners, developers, environmental groups, and federal management agencies. These legal scuffles are complex and ongoing, but at this time, most coastal wetlands, including salt marshes and mangroves, are protected by the No Net Loss policy.

## Restoration

The No Net Loss policy stipulates that if development impacts jurisdictional wetlands, then an equivalent area of wetland needs to be restored as compensation for the impact. The process of wetland restoration is simple in concept, but challenging in practice. In concept, wetland restoration first involves creating (by excavating, filling, or leveling) an appropriate elevation for the targeted plant species. Then, plants are allowed to establish naturally or are transplanted into the site – an undertaking that often involves large groups of volunteers, who then develop a stewardship of the new habitat (Fig. 19). Once plants are established, the “Field of Dreams” hypothesis is usually implicitly or explicitly invoked: “If you build it, they will come” (Palmer et al. 1997). In this context, “they” refers to the

animals and ecosystem processes that are characteristic of reference marshes. Although it may take decades for restored wetlands to develop a full set of target conditions, and not all restoration projects are fully successful, the study of wetland restoration can better inform future projects, helping to ensure future successes.

---

## Future Directions: Integrating Science and Restoration

On a global scale, coastal wetlands have been substantially reduced in area over the past 100 years, primarily due to urban and agricultural development, hydrological alterations, and subsidence due to natural events (soil consolidation and faults) and extraction of groundwater and minerals. Although the rate of loss has slowed in recent years, coastal wetlands continue to be vulnerable to disturbance from development, storm events, and offshore oil spills. Near- and long-term management priorities focus on conserving and restoring ecosystem functions of coastal wetlands, as they provide substantial support for local and state economies. To address these management priorities, wetland restoration projects, ranging from large (>100 ha) to small (<1 ha), have been implemented in many parts of the world.

In practice, restoration usually focuses on permit stipulations, which often emphasize vegetation cover characteristics and cover ecologically short time scales (3–5 years). Vegetation cover in restored sites can be linked to some specific ecosystem functions (e.g., nutrient uptake). However, metrics that are more closely linked to long-term ecosystem health, such as nutrient storage and belowground plant biomass, rarely achieve natural levels, even decades after restoration (Craft et al. 1999). This highlights a major challenge: is there a way to improve ecosystem functions and long-term sustainability, without making restoration markedly more expensive or labor intensive? For example, will increasing the number of plant species or genetic diversity improve ecosystem functions? Can facilitative interactions among plants, animals, and microbial communities be used to augment restoration success? Will the integration of higher elevations into restoration design improve ecosystem resilience in the face of near-term sea level rise? The answers to these types of questions will vary among and even within sites and regions. That heterogeneity presents a substantial challenge for restoration practitioners: ecologically successful restoration requires an in-depth understanding of local wetland ecology. Incorporating that understanding into a restoration plan that includes both near- and long-term ecological measures of success is the ultimate goal of those who study and those who practice coastal wetland restoration.

---

## References

- Baxter I, Hosmani PS, Rus A, Lahner B, Borevitz JO, Muthukumar B, Mickelbart MV, Schreiber L, Franke RB, Salt DE. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in *Arabidopsis*. Plos Genet. 2009;5:e1000492.

- Bertness MD. Ribbed mussels and *Spartina alterniflora* production in a New England salt marsh. *Ecology*. 1984;65:1794–807.
- Bertness MD. Fiddler crab regulation of *Spartina alterniflora* production on a New England salt marsh. *Ecology*. 1985;66:1042–55.
- Bertness MD. Zonation of *Spartina patens* and *Spartina alterniflora* in a New England salt marsh. *Ecology*. 1991;72:138–48.
- Bertness MD. The ecology of a New England salt marsh. *Am Sci*. 1992;80:260–8.
- Boston KG. The development of salt pans on tidal marshes, with particular reference to south-eastern Australia. *J Biogeogr*. 1983;10:1–10.
- Brody SD, Highfield WE, Ryu HC, Spanel-Weber L. Examining the relationship between wetland alteration and watershed flooding in Texas and Florida. *Nat Hazards*. 2007; 40:413–28.
- Carver E. Birding in the United States: a demographic and economic analysis. Addendum to the 2006 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation, report 2006-4, U.S. Fish & Wildlife Service, Arlington; 2009
- Craft C. Freshwater input structures soil properties, vertical accretion, and nutrient accumulation of Georgia and U.S. tidal marshes. *Limnol Oceanogr*. 2007;52:1220–30.
- Craft C, Reader J, Sacco JN, Broome SW. Twenty-five years of ecosystem development of constructed *Spartina alterniflora* (Loisel) marshes. *Ecol Appl*. 1999;9:1405–19.
- Craft C, Clough J, Ehman J, Joye S, Park R, Pennings S, Guo HY, Machmuller M. Forecasting the effects of accelerated sea-level rise on tidal marsh ecosystem services. *Front Ecol Environ*. 2009;7:73–8.
- Dahl TE. Wetlands losses in the United States 1780s to 1980s. Washington, DC: U.S. Department of the Interior, Fish and Wildlife Service; 1990.
- Deegan LA, Johnson DS, Warren RS, Peterson BJ, Fleeger JW, Fagherazzi S, Wollheim WM. Coastal eutrophication as a driver of salt marsh loss. *Nature*. 2012;490:388–92.
- Doyle TW, Krauss KW, Conner WH, From AS. Predicting the retreat and migration of tidal forests along the northern Gulf of Mexico under sea-level rise. *For Ecol Manage*. 2010; 259:770–7.
- Elmqvist T, Cox PA. The evolution of vivipary in flowering plants. *Oikos*. 1996;77:3–9.
- Gedan KB, Kirwan ML, Wolanski E, Barbier EB, Silliman BR. The present and future role of coastal wetland vegetation in protecting shorelines: answering recent challenges to the paradigm. *Clim Change*. 2011;106:7–29.
- Giri C, Ochieng E, Tieszen LL, Zhu Z, Singh A, Loveland T, Masek J, Duke N. Status and distribution of mangrove forests of the world using earth observation satellite data. *Glob Ecol Biogeogr*. 2011;20:154–9.
- IPCC. Climate Change 2007: the physical science basis. Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK/New York: Cambridge University Press; 2007.
- Jordan TE, Valiela I. A nitrogen budget of the ribbed mussel, *Geukensia demissa*, and its significance in nitrogen flow in a New England salt marsh. *Limnol Oceanogr*. 1982;27:75–90.
- Kruczynski WL. Mitigation and the Section 404 program: a perspective. In: Kusler JA, Kentula ME, editors. Wetland creation and restoration: the status of the science. Washington, DC: Island Press; 1990. p. 549–54.
- Lee RY, Joye SB. Seasonal patterns of nitrogen fixation and denitrification in oceanic mangrove habitats. *Mar Ecol Prog Ser*. 2006;307:127–41.
- McKee KL. Soil physicochemical patterns and mangrove species distribution – reciprocal effects? *J Ecol*. 1993;81:477–87.
- McKee KL, Rooth JE. Where temperate meets tropical: multi-factorial effects of elevated CO<sub>2</sub>, nitrogen enrichment, and competition on a mangrove-salt marsh community. *Glob Chang Biol*. 2008;14:971–84.
- Onuf CP, Teal JM, Valiela I. Interactions of nutrients, plant growth and herbivory in a mangrove ecosystem. *Ecology*. 1977;58:514–26.

- Osland MJ, Enwright N, Day RH, Doyle TW. Winter climate change and coastal wetland foundation species: salt marshes versus mangrove forests in the southeastern U.S. *Glob Chang Biol*. 2013;19:1482–94.
- Palmer MA, Ambrose RF, Poff NL. Ecological theory and community restoration ecology. *Restor Ecol*. 1997;5:291–300.
- Pi N, Tam NFY, Wu Y, Wong MH. Root anatomy and spatial pattern of radial oxygen loss of eight true mangrove species. *Aquat Bot*. 2009;90:222–30.
- Sherman RE, Fahey TJ, Howarth RW. Soil-plant interactions in a neotropical mangrove forest: iron, phosphorus and sulfur dynamics. *Oecologia*. 1998;115:553–63.
- Stiven AE, Kuenzler EJ. The response of two salt marsh molluscs, *Littorina irrorata* and *Geukensia demissa*, to field manipulations of density and *Spartina* litter. *Ecol Monogr*. 1979;49:151–71.
- Thom BG. Coastal landforms and geomorphic processes. In: Snedaker SC, Snedaker JG, editors. *The mangrove exosystem: research methods*. Paris: Unesco; 1984. p. 3–17.
- Thursby GB, Abdelrhman MA. Growth of the marsh elder *Iva frutescens* in relation to duration of tidal flooding. *Estuaries*. 2004;27:217–24.
- Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH, Tilman DG. Human alteration of the global nitrogen cycle: sources and consequences. *Ecol Appl*. 1997;7:737–50.

## Further Reading

- Craft C, Megonigal P, Broome S, Stevenson J, Freese R, Cornell J, Zheng L, Sacco J. The pace of ecosystem development of constructed *Spartina alterniflora* marshes. *Ecol Appl*. 2003;13:1417–32.
- Dugan P. *Wetlands in danger: a world conservation atlas*. New York: Oxford University Press; 1993.
- Engle VD. Estimating the provision of wetland services by Gulf of Mexico coastal wetlands. *Wetlands*. 2011;31:179–93.
- Mendelssohn IA, McKee KL, Patrick Jr WH. Oxygen deficiency in *Spartina alterniflora* roots: metabolic adaptation to anoxia. *Science*. 1981;214:439–41.
- Perry CL, Mendelssohn IA. Ecosystem effects of expanding populations of *Avicennia germinans* in a Louisiana salt marsh. *Wetlands*. 2009;29:396–406.
- Rützler K, Feller IC. Caribbean mangrove swamps. *Sci Am*. 1996;274:94–9.
- Saintilan N, Rogers K, McKee K. Salt marsh-mangrove interactions in Australasia and the Americas. In: Perillo GME, Wolanski E, Cahoon DR, Brinson MM, editors. *Coastal wetlands: an integrated ecosystem approach*. The Netherlands: Elsevier; 2009. p. 855–83.

Hugh Kirkman

## Contents

Introduction .....	458
Seagrass Ecosystems .....	464
Seagrass Morphology .....	466
Economic Goods and Services Provided by Seagrass Ecosystems .....	466
Hydrodynamics and Resilience in Seagrass Ecosystems .....	467
Seagrass Grazers .....	469
Epiphytes and Epiphyte Grazers .....	469
Complex Food Webs Associated with Seagrass Ecosystems .....	470
Threats to the Future Vitality of Seagrass Ecosystems .....	472
Restoration and Recovery .....	476
Genetic Diversity .....	477
Future Directions .....	479
References .....	480

---

## Abstract

- Seagrasses are the only marine-submerged angiosperms, and there exist approximately 60 species of seagrasses, worldwide.
- Tropical and temperate seagrass ecosystems are markedly different. Temperate seagrasses are larger and beds are denser. Temperate seagrasses respond to seasons and water temperature whereas tropical seagrasses, although also responding to seasons, i.e., wet and dry, do not show growth correlations with changes in water temperature.
- Globally many seagrass beds have been lost, and many more are threatened by human activities; protection is vital. Reduced light (due to eutrophication of coastal regions and sediment disturbance) is the single most important cause of seagrass loss.

---

H. Kirkman (✉)  
Australian Marine Ecology Pty Ltd, Kensington, Australia  
e-mail: [hughkirkman@ozemail.com.au](mailto:hughkirkman@ozemail.com.au)



- Seagrass beds support numerous invertebrates and juvenile commercially and recreationally important fish and crustaceans. Many of these dependent animal communities are herbivorous but few eat seagrasses. Plants and animals growing on seagrass lead to complex food-web communities, with numerous trophic levels. Many birds and mammals use seagrass ecosystems as sources of food, despite using other coastal ecosystems for habitation.
- Seagrass beds are a sink for nutrients delivered from terrestrial runoff and detritus from seagrass beds and other marine ecosystems. These nutrients support their extensive food web. Seagrass beds are also a significant net sink for atmospheric carbon storage.

---

## Introduction

Restoration and remediation of seagrass ecosystems have not met with great success. The use of vegetative propagules as a means for reestablishment of seagrass beds has been plagued with difficulties due to mismatches between propagule sources and targeted restoration beds. Removing vegetative propagules from donor beds leads to problems of the donor beds recovering. Growing seagrass from seed is not always a viable option for restoration because of the vulnerability of seedlings and poor recruitment into unvegetated areas. Remediation of destroyed seagrass is not often successful. An understanding of levels of genetic diversity and spatial genetic structure can contribute to improved restoration outcomes by identifying the most genetically appropriate source material for restoration sites. The discoveries made recently through DNA analysis and phylogenetic affinities have also helped untangle some of the taxonomic identities of seagrass and led to better decisions as to the choice of restoration sources and materials.

The ancestors of the higher plants left the sea some 400 million years ago, but the seagrasses are the only ones to have returned to a completely submerged marine existence. This polyphyletic group of flowering plants reinvaded the sea probably about 100 million years ago in the Cretaceous (Larkum and den Hartog 1989). Our current knowledge of species affinities and phylogenetic origins is poor for this group of plants and requires urgent improvement in order to better inform management and researchers (Table 1). A stable taxonomy is a necessary base for all botanical research. Morphological and anatomical variations within the species are not systematically documented, and it is recommended that samples of material used for molecular, physiological, and morphological research are deposited in recognized herbaria.

There are about 60 species of seagrass in the world in 13 genera (Table 1). *Ruppia* and *Lepilaena* are often grouped among the seagrasses but can grow in brackish and fresh water. There are so few seagrass species globally and locally, and a large degree of endemism that the loss of one species may mean thousands of other organisms are lost. Kuo and den Hartog (2001) describe all seagrass to that date and offer a key for their identification.

**Table 1** List of seagrass species of the world. The distributions have been taken from Green and Short (2003). *The Seagrasses of the World*. There is still taxonomic activity deciding on whether some species here are real species or strains of others. Distributions too are unclear in some cases

Family	Genus	Species	Distribution
Zosteraceae	<i>Zostera</i>	<i>marina</i>	Europe, North America
		<i>caespitose</i>	Japan
		<i>caulescens</i>	North Korea and Japan
		<i>asiatica</i>	Korea and Japan
		<i>noltii</i>	East Atlantic, Baltic, Mediterranean, Black, Caspian, and Aral Seas
		<i>japonica</i>	Japan
		<i>capensis</i>	Southern Africa
		<i>capricornii</i>	Australia
		<i>muelleri</i>	Australia
		<i>mucronata</i>	Australia
	<i>novazelandica</i>	New Zealand	
	<i>Phyllospadix</i>	<i>scouleri</i>	Western North America
		<i>torreyi</i>	Western North America
		<i>serrulatus</i>	Northwestern North America
		<i>iwatensis</i>	Korea, China, and Japan
		<i>japonicus</i>	Korea and Japan
	<i>Heterozostera</i>	<i>tasmanica</i>	Southern Australia
		<i>polychlamis</i>	Southern Australia
		<i>nigricaulis</i>	Southern Australia
		<i>chiliensis</i>	Chile
Cymodoceaceae	<i>Halodule</i>	<i>uninervis</i>	Tropical and subtropical Australia, West Africa, SE Asia, India, Pacific
		<i>beaudetti</i>	Northeast Madagascar, Caribbean
		<i>wrightii</i>	Global
		<i>bermudensis</i>	Bermuda
		<i>ciliate</i>	Tobago Island, Panama
		<i>pinifolia</i>	Indo-West Pacific
	<i>Cymodocea</i>	<i>emarginata</i>	Brazil
			Mediterranean and North Africa
			Indo-West Pacific
		<i>nodosa</i>	Indo-West Pacific
	<i>Syringodium</i>	<i>rotundata</i>	Northwestern Australia
		<i>serrulata</i>	Caribbean, Florida
		<i>angustata</i>	Indo-West Pacific
		<i>filiforme</i>	Indo-West Pacific
	<i>Thalassodendron</i>	<i>isoetifolium</i>	South Western Australia
		<i>ciliatum</i>	Southern Australia
<i>Amphibolis</i>	<i>pachyrhizum</i>	Southern Australia	
	<i>antarctica</i>	Southern Australia	
	<i>griffithii</i>	South western Australia	

(continued)

**Table 1** (continued)

Family	Genus	Species	Distribution	
Posidoniaceae	<i>Posidonia</i>	<i>oceanica</i>	Mediterranean	
		<i>australis</i>	Southern Australia	
		<i>sinuosa</i>	Southern Australia	
		<i>angustifolia</i>	Southern Australia	
		<i>ostenfeldii</i>	Southern Australia	
		<i>robertsoniae</i>	Southern Australia	
		<i>coriacea</i>	Southern Australia	
		<i>denhartogii</i>	Southern Australia	
		<i>kirkmanii</i>	Southern Australia	
Hydrocharitaceae	<i>Enhalus</i>	<i>acoroides</i>	Indo-West Pacific and Australia	
Thalassioideae	<i>Thalassia</i>	<i>hemprichii</i>	Australia	
		<i>testudinum</i>	Caribbean and Florida	
Halophiloideae	<i>Halophila</i>	<i>ovalis</i>	Global	
		<i>ovata</i>	Trop. Australia, Southeast Asia	
		<i>minor</i>	Australia, SE Asia, Western Pacific	
		<i>australis</i>	Southern Australia	
		<i>hawaiiiana</i>	Hawaii	
		<i>madagascariensis</i>	Madagascar	
		<i>johnsonii</i>	Florida	
		<i>decipiens</i>	Australia	
		<i>capricorni</i>	Queensland and New Caledonia	
		<i>Halophila</i> sect. <i>Microhalophila</i>	<i>beccarii</i>	India and SE Asia
		<i>Halophila</i> sect. <i>Spinulosa</i>	<i>spinulosa</i>	Tropical Australia, Indonesia, and Philippines
		<i>Halophila</i> sect. <i>Tricostatae</i>	<i>tricostata</i>	Tropical East Australia
		<i>Halophila</i> sect. <i>Americanae</i>	<i>engelmannii</i>	Gulf of Mexico and Caribbean
			<i>baillonii</i>	Caribbean
Ruppiaaceae	<i>Ruppia</i>	<i>tuberosa</i>	Australia	
Zannichelliaceae	<i>Lepilaena</i>	<i>marina</i>	Southern Australia	

The taxa regarded as seagrasses belong to four families, viz., the Zosteraceae, the Cymodoceaceae, the Posidoniaceae, and the Hydrocharitaceae. The first three families contain only seagrasses, but the Hydrocharitaceae contains only three genera that are considered seagrasses. The other 14 genera are confined to fresh-water habitats. Two other families contain one species each, and these have not received a lot of research – *Ruppia tuberosa* and *Lepilaena marina* (Table 1). Nine of the 13 genera are dioecious.

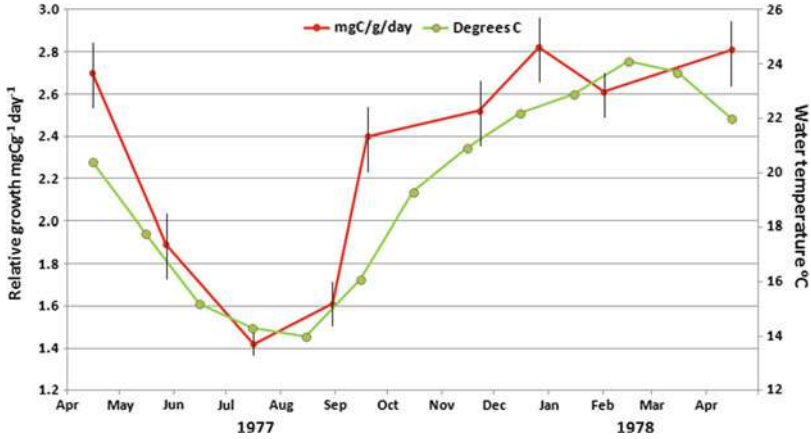
Sculthorpe (1969) gave a very comprehensive description of the morphology, physiology, and ecology of submerged aquatic plants in his definitive book. Seagrass plants have adapted to being supported by water and have nonfunctional stomates; they assimilate dissolved CO<sub>2</sub> by diffusion through the epidermis which

is the major site for photosynthesis, in contrast to terrestrial plants. Seagrasses vary in their ability to grow in low-light conditions. Most species of seagrass are adapted to lower light levels, and they have evolved gas storage organs, both of which can be considered adaptations that allow them to photosynthetically assimilate CO<sub>2</sub> at low, but sufficient rates. Seagrasses have a thin cuticle over the leaf blade and are halophytic in their physiological traits. Most can live for short periods in a wide range of salinities; the salinity of coastal seawater is about 35 parts per 1,000. Seagrasses also withstand a wide range of temperatures in the coastal waters and are capable of acclimating to seasonal and spatial variability in this environmental factor. *Zostera marina* was found to be growing healthily under ice in an embayment of the Bering Sea. Furthermore, it was living there in anaerobic conditions. Thus, these plants are quite robust in their adaptive potential! Seagrasses have become anatomically adapted to limited access to oxygen by developing aerenchymatic tissues with continuous air-filled lacunae running from leaves to roots. Oxygen is only lost to the water column during the day, but it is continuously lost from roots and rhizomes to the sediment. The oxygen produced in photosynthesis is stored in lacunal spaces of the leaves and can be recycled for use in a limited and localized rate of aerobic respiration. The loss of oxygen to the rhizosphere from root surfaces is vital to protect root tissues by oxidizing reduced toxic phytotoxins like iron, manganese, and sulfide. The oxygen released to sediments has important implications for the degradation of organic matter, acting as the terminal electron acceptor in the oxidative breakdown of organic molecules.

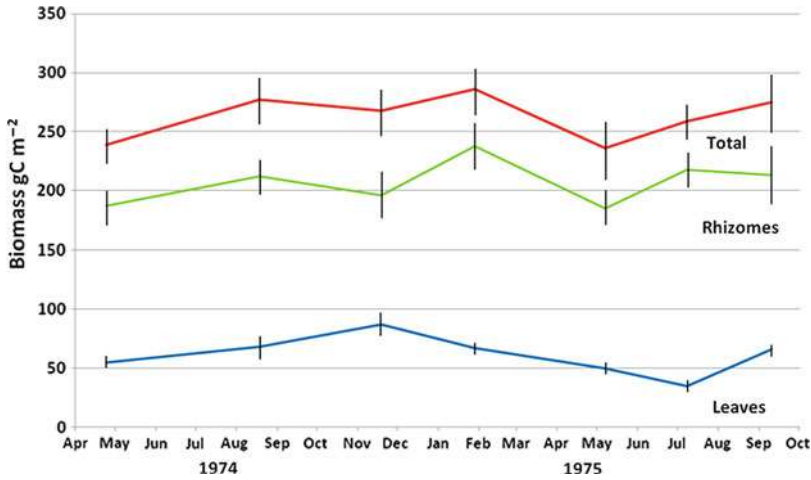
Seagrasses may be monoecious or dioecious. Pollination in the seagrasses takes place in the water column except in *Lepilaena* and *Enhalus* where pollen is released at the surface. In *Enhalus* the male flower breaks the surface and releases the floating pollen to the receptive female flower, and a number of seeds mature in a fruit that may be 5–10 cm long. The seeds germinate on release (McConchie and Knox 1989).

Seagrass ecosystems grow in coastal waters from intertidal to 50 m deep or more. This is an important statement to make at the beginning of a chapter on seagrass ecosystems. Seagrasses are limited in their distribution by light, and 50 m is about the limit that suitable light can penetrate even the clearest coastal waters. Seagrasses require an underwater photosynthetic irradiance more than 11 % of that incident on the water surface. Light is reduced by turbidity in the water, and this turbidity is determined by the content of sediment or organic matter. Light is also reduced by animals or plants growing on the seagrass plants; these epiphytes, as they are called, can often shade seagrass plants to below the photosynthetic compensation point required to sustain plants, leading to death under high nutrient conditions.

Most temperate seagrasses are seasonal having a strong growth in spring and early summer then declining in productivity in fall and winter. In a *Posidonia australis* bed growing in Port Hacking, New South Wales, Australia, the relative growth rate measured as mg of carbon per gram of leaf per day closely followed water temperature (Fig. 1). There is a steep increase in relative growth rate at the beginning of spring to a maximum at the end of summer. When the mean weight of



**Fig. 1** Average relative growth of *Posidonia australis* leaves from April 1977 to April 1978, with surface water temperature over the seagrass bed. Vertical lines are standard errors about the mean (Kirkman and Reid 1979)



**Fig. 2** Dry weight biomass of *Posidonia australis* estimated for a 15-month period. Vertical lines represent one standard error about the mean (Kirkman and Reid 1979)

leaves and rhizomes were charted separately for 15 months, there was not such a seasonal influence as there was in productivity (Fig. 2). These biomasses were the means from ten quadrats each of 0.0625 m<sup>2</sup>. These records are important because they represented measurements that could be used for monitoring seagrass condition. Obviously productivity is a more sensitive measurement to detect changes. Unfortunately these measurements are more difficult to make in the field than biomass measures, and we found later that, for large-leaved plants, shoot density

was a better measure to determine changes in seagrass condition. Shoots are considered to be a collection of leaves coming from a single node of the rhizome.

Along the tropical and subtropical coasts of Northeastern Australia, the Caribbean, Southeastern USA, Eastern Africa, and Southeast Asian grows a diverse and extensive assemblage of seagrass. These tropical seagrasses growing along continental coastlines are subjected to greater natural disturbance than those in temperate areas. Greater frequency in tropical cyclones, monsoonal extremes in seasonal freshwater runoff into coastal estuaries, and, in tropical Australia, tides to 9 m are not conducive to the establishment and growth of seagrasses. Some tropical seagrasses recover after storms or disturbance quite rapidly, within a year or so, while temperate seagrasses take longer to recover and therefore are not as resilient to disturbance. The well-established concepts of temperate seagrass ecology and habitat function are not appropriate to the diverse range of seagrass habitats in tropical Australia and parts of Southeast Asia. Seagrasses are smaller, are more ephemeral, and have more natural disturbance from dugongs, turtles, and cyclones than do temperate seagrasses. In the tropics *Halophila ovalis* regenerates by vegetative propagules — its rhizomes — while other species set many seeds. Reproductive capacity, in general, is high in tropical seagrasses. Thus, while disturbances to tropical seagrass ecosystems may be more frequent and of greater intensity, they also have greater capacity to recover following those disturbances.

In Darwin Harbour in the Northern Territory of tropical Australia, two species dominate but are ephemeral with cover changing seasonally and distribution also being variable within a year. *Halophila decipiens* and *Halodule uninervis* grow to a depth of about 4 m, but biomass and percentage cover are impossible to estimate by conventional methods due to the variability of observations. Video transects were used to assess seagrass cover and distribution. Thousands of hectares disappeared during the wet season and were replaced in the dry season. Predictions as to how much and where were not accurate. It is believed that the seed bank for each of these lies in the sediment through the wet season (July to January) when light is below compensation level and seeds germinate in April for the dry season where they grow, flower, and set seed until September/October. The waters surrounding these habitats have a very low nitrogen concentration, and the ecosystems are subjected to high disturbance.

Carruthers et al. (2002) identified four broad categories of seagrass habitat. They defined them as “River estuaries,” “Coastal,” “Deep water,” and “Reef” controlled by terrigenous runoff, physical disturbance, low light, and low nutrients, respectively.

In tropical regions seagrass is eaten by manatees, dugongs, and turtles; in contrast, in temperate seagrass beds some swans, geese, and ducks are important consumers of intertidal seagrass. The realization of the biological importance of seagrass was highlighted in the 1920s when large areas in the USA and Europe died causing a decline in commercially harvested fish and shellfish.

In the USA the bay scallop, *Argopecten irradians*, fishery in North Carolina and Chesapeake Bay collapsed following the eelgrass wasting disease of 1931–1932 that destroyed more than 90 % of seagrass. The scallops returned as eelgrass

recovered in North Carolina but not in Chesapeake Bay. In southern Florida, in the late 1980s to early 1990s, the pink shrimp, *Penaeus duorarum*, declined by 50 % when there was a 20 % loss of *Thalassia*, the main nursery (from [Butler and Jernakoff \(1999\)](#), Chap. 2).

---

## Seagrass Ecosystems

Seagrasses have diversified and spread to become dominant organisms throughout the world's shallow sediment bottoms around all continents except Antarctica, primarily in estuaries and more sheltered coastal seas. Two genera (the northeastern Pacific *Phyllospadix* and the temperate southern "sea nymph," *Amphibolis*) have even colonized rocky shores. Colonization by seagrasses profoundly changed the nature of coastal sediment systems.

Aboveground, the often dense vegetation strongly reduces the physical energy of waves and currents, creating a zone of kinetic stability within which animal communities can thrive; in addition, it provides food for herbivores and physical structure that shelters a much higher abundance and diversity of animals than do the surrounding bare sediments. The refuge value of seagrasses generally rises with its species or density complexity. Seagrass leaves provide a substratum for growth of epiphytic microalgae and sessile invertebrates and macroalgae that fuel complex food webs. This combined productivity of seagrasses and associated algae ranks seagrass beds among the most productive ecosystems on earth (Table 2).

Moreover, because much seagrass production ends up in belowground tissues and ungrazed detritus, seagrass beds are an important global sink for carbon, accounting for an estimated 15 % of net CO<sub>2</sub> uptake by marine organisms on a global scale, despite contributing only 1 % of marine primary production. Tropical seagrasses tend to support higher metabolic rates and somewhat lower net community production than temperate ones. The production-to-respiration ratio tended to increase with gross primary production exceeding 1 on average. It has been estimated that for a low global seagrass coverage of 300,000 km<sup>2</sup> from 20 to 50 Mt of carbon per year and for a high seagrass coverage of 600,000 km<sup>2</sup> from 40 to 100 Mt of carbon per year ([Duarte et al. 2010](#)) has been taken up.

Seagrass beds provide important nursery areas for juvenile fish including commercially and recreationally used fish and shrimp. For example, in the Gulf of Carpentaria in Northern Australia, juvenile *P. esculentus* (tiger prawns) live in seagrass beds and reach sexual maturity at a carapace length of around 32 mm. Although seagrass biomass in the Gulf of Carpentaria was not a consistent linear predictor of juvenile tiger prawn numbers, mean catches of both the 2–2.9 mm carapace length postlarvae and juvenile *P. esculentus* were highest when the biomass of seagrass exceeded 100 g m<sup>-2</sup>. However, these high-biomass seagrass beds contribute only 6 % to the total extent of seagrasses in the shallow waters (<2.5 m deep) of the Gulf of Carpentaria. Although the numbers of juvenile tiger prawns were lower in the low-biomass seagrass beds, because of their extent, these seagrass beds are the main nurseries for sustaining the production of the valuable

**Table 2** Comparison between average seagrass and other marine and terrestrial ecosystems. Net primary production (NPP) (Modified from Mateo et al. (2007))

Ecosystem	NPP (gCm <sup>-2</sup> /year)	Total global NPP (PgC/year)
Mangroves	1,000	1.1
Seagrass	817	0.49
Forests	400	16.4
Macroalgae	375	2.55
Crops	350	5.25
Terrestrial	200	29.6
Phytoplankton coastal	167	4.5
Phytoplankton ocean	130	43

Northern Prawn Fishery in Australia (production of all prawns in the Northern Prawn Fishery was nearly \$28.5 million or 1,627 t in 2011).

The strong network of underground rhizomes found in seagrass ecosystems stabilizes otherwise mobile sediments and filters overlying water by slowing it to allow organic matter and sediments to deposit locally, rather than being washed further offshore. Seagrass beds are nutrient sinks, accumulating detritus from the organic matter deposited in them. Nutrient cycling between sediment organic matter and seagrasses is loosely coupled in time as anaerobic decomposition is slower than aerobic decomposition. Internal cycling of nutrients in seagrass beds comes from seagrass detritus, the animals that live in it and those that ingest seagrass above- and belowground parts.

Seagrass habitats grow naturally as patches in many ecosystems, though they often form continuous coverage under ideal, conditions with rare disturbance. The area covered by them may be stable over decades under some environmental conditions. Increased stresses due to eutrophication and mechanical disturbance from storm surges have the potential to change communities from continuous to fragmented seascaapes. Changing from continuous cover to a fragmented seascape may induce positive feedbacks that increase vulnerability of these systems to even further biophysical degradation. Fragmentation of coverage has the potential to cause collapse of food webs, decrease the potential for reproductive continuity within plant and animal populations, and generally threaten biological diversity. The major abiotic factors affecting seagrass seascape structure are the following: physical disturbance from storm and wind-driven waves, the hydrodynamics surrounding the seagrass bed including how well it is protected from storms, water flow around the bed usually from tidal currents, and the size and amount of particle deposition into and around the bed including sediment that drops out of the water column and causes turbidity to increase. Biotic factors are the following: successful recruitment of propagules, clonal reproduction by vegetative propagules, herbivory by animals that eat seagrass leaves, and the abundance and diversity of associated species that rely on seagrass and provide some assistance to seagrass, e.g., animals that break down detritus. The success of predators living on seagrass grazers or on the animals and plants that live on the leaves and stems depends on a complex status that may not be there when seagrass beds are fragmented. Patches of seagrass may



not have the resources available that provide for a stable ecosystem. Nutrient cycling and availability may not be as concentrated as they were in an entire unfragmented bed, producing areas that decline below the size that can withstand storms and wave surges.

---

## Seagrass Morphology

Seagrasses are rooted plants, and many form dense mats of rhizomes in the underlying sediments which reduce the mobility of those sediments and thus stabilize components of local biogeochemical cycles. Roots are not usually supportive organs but have root hairs of variable size and density. The roots of seagrasses are adventitious and grow from the lower surface of the rhizomes, generally at the nodes. Seagrass rhizomes are usually herbaceous and monopodially. Monopodial branching occurs when the terminal bud continues to grow as a central leader shoot and the lateral rhizomes remain subordinate or irregularly branched; however, in *Amphibolis* and *Thalassodendron* the rhizome branches sympodially and becomes woody. Sympodial branching occurs when the terminal bud ceases to grow (usually because a terminal flower has formed) and an axillary bud or buds. Rhizomes are almost always buried in the sediment, and the persistent fibrous remains of old leaf sheaths usually cover the rhizomes of *Enhalus* and *Posidonia* and partially cover the rhizomes in some other genera. The coverage of decomposing leaf sheaths on rhizomes likely provides protection from physical damage as rhizomes are abraded by sediment movement. The leaf is produced either from the rhizome nodes, normally from the upper side in *Enhalus*, *Posidonia*, and the Zosteraceae, or from the apex of erect stems in *Thalassia* and the Cymodoceaceae. The leaf sheath is clearly differentiated from the leaf blade and encloses the young, developing leaves in all seagrass genera with ribbon blades. *Thalassia* and *Amphibolis* leaves and sheaths abscise together. Leaf sheaths also provide unique protective microhabitats for small invertebrates and their larvae.

## Economic Goods and Services Provided by Seagrass Ecosystems

The economic value of seagrass ecosystems has not been well documented. This may be because of the difficulty in defining the goods and services that come from a seagrass bed and then putting a value on the services. "Ecosystem services are the direct or indirect contributions that ecosystems make to the well-being of human populations" is one of many definitions used by economists to value estuarine and coastal resources.

The seagrass ecosystem resource is very valuable when considering the goods and services mentioned above, as a nursery for many species valued by the seafood industry, as a global carbon sink, for nutrient cycling and water purification and to physically stabilize coastlines. Many authors have used generic financial figures to estimate the value of ecosystem goods and services of seagrass beds; but,

practically, they vary so widely and have such broad uncertainties when considered together that it is better to gain specific value estimates for specific sites.

Even at the site scale, there is still a large number of ecosystem services that have either no or very unreliable valuation estimates. The most significant problem faced in valuing ecosystem services, including those of seagrasses, is that very few are marketed. Some of the products arising from seagrasses, such as raw materials, food, and fish harvests, are bought and sold in markets; it is easiest to place financial value on these products.

However, the valuation process, even for these products, is more complicated than it first appears. For example, one important service of seagrass beds is the maintenance of fisheries through providing coastal breeding and nursery habitat. Although many fisheries are exploited for commercial harvests sold in domestic and international markets, studies have shown that the inability to control fishing access and the presence of production subsidies and other market distortions can impact harvests, the price of fish sold, and, ultimately, the estimated value of the seagrass habitat in supporting commercial fisheries (Barbier et al. 2011). There is a need for more financial models that include higher-order economic connections and feedbacks in order to more accurately estimate the values of seagrass ecosystems. It is likely that human behavior in both financial and regulatory arenas will have to be added to such models, making it crucial that ecologists' work with economists and social scientists to develop novel modeling frameworks.

## Hydrodynamics and Resilience in Seagrass Ecosystems

Seagrass species often sort themselves into assembled communities according to hydrodynamic regimes, e.g., *Amphibolis* spp. and *Phyllospadix* spp., growing in areas of higher flow, compared to other species, in Australia and the East Pacific, respectively. In *Phyllospadix*, reduction of vascular bundles and the absence of woody or cork material allows the leaves to remain erect in the face of strong water action and mechanical stress. It is also more securely attached to its substratum, probably due to greater density in root hair growth, than many species from weaker hydrodynamic regimes. The roots and rhizomes of *Phyllospadix* also have thicker outer epidermal walls, making it better able to withstand strong wave force. The lacunae (internal air spaces) are reduced in volume in this genus, because the plants live in a highly oxic (oxygen-rich), well-mixed environment. Reduced lacunar volume likely provides for greater strength in stems. As would be expected for a plant that needs to be adapted to water motion in a turbulent surf zone, *Phyllospadix* shows more flexible (non-lignified hypodermal) leaf tissues than does *Zostera*, a species from less turbulent environments.

For *Amphibolis* a different adaptation has allowed it to grow in areas of high water movement. It has a characteristic stem and leaf cluster morphology that presents a gap in the canopy, allowing water to flow beneath the main canopy. By contrast, *Posidonia* plants have a uniform leaf shape, maintaining the same leaf width from base to tip (although an increase in canopy density will occur as leaves

emerge from their sheaths). This means that there is no gap for water to flow through and hence results in a smoothly decreasing water velocity profile. In the genus *Posidonia* there are two distinct groups: the *australis* group and the *ostenfeldii* group. The *australis* group has stout underground rhizomes that grow laterally in the sediment. This allows them to spread into unvegetated areas, but they do not have the strong hold on the sediment that is exhibited in the *ostenfeldii* group. This group can grow in strong swells and has a typical windrow appearance due to the fact that its seedlings only grow successfully on the lee side of sand ripples. When establishing, the seedlings of members of this group grow as a clump because their rhizomes grow downward once they have established on the lee side of the sand ripples, unlike the lateral pattern of growth in the *australis* group. Gradually the clump enlarges until it coalesces with others, and a full cover is achieved. The leaves of this group are also noticeably stronger than those of the *australis* group.

Exposure to hydrodynamic energy is widely considered an important environmental factor influencing seagrass species distributions; however, its influence compared to other mechanisms has not been tested in many places, and this generalization needs broader consideration. Recently, Hansen and Reidenbach (2013) have shown the importance of *Zostera marina* in reducing velocities of water over them by 60 % in the summer, when leaves were longer, and 40 % in winter compared with an unvegetated site. The seagrass bed also dampened wave heights in all seasons except winter when leaves were shortest. Shear stress was reduced in the summer so that less sediment was resuspended and plants had more light for photosynthesis. Suspended sediment was enhanced by low seagrass coverage in winter compared with an unvegetated site.

Hydrodynamic processes also influence the dispersal of seagrass seeds and vegetative fragments, as well as eggs and larvae of organisms that inhabit seagrass communities and form associated food webs, e.g., invertebrates and fish. Seagrasses baffle unidirectional tidal and oscillatory (wave-driven) currents. Plant morphology and structure affect the capacity of seagrasses to influence water flow. The capacity of seagrasses to baffle water flow and currents is linked to the accretion of sediments and increases with increasing patch structure and size. This, in turn, improves conditions for seagrass growth and recruitment, accelerating patch density and the extent of coverage. Empirical studies of temperate seagrass responses to hydrodynamics, however, have been limited to *Posidonia* spp. and *Amphibolis* spp. in Australia and *Zostera marina* in temperate USA and Europe. There is room for much broader consideration of these potential adaptations and influences on multi-trophic dynamics.

Tidal height and range influence variability in biomass and productivity in intertidal seagrass populations, e.g., those of *Zostera muelleri* in Victoria, Southern Australia, and *Halophila decipiens* and *Halodule uninervis* in turbid tropical waters of great tidal range. Low water levels (tidal heights), barometric conditions, and high temperatures can prompt prolonged atmospheric exposure and desiccation for intertidal species which may result in dieback (Seddon et al. 2000). Empirical studies on the response of seagrasses to atmospheric exposure are limited.

## Seagrass Grazers

Waterfowl are significant grazers of seagrasses consuming large amounts of rhizomes and leaves. Swans (*Cygnus atratus*) in Australia eat *Zostera muelleri* while migratory herbivores such as brant geese (*Branta bernicla*) live between the Atlantic coast of the USA from Maine to Georgia, in Alaska, California, and Mexico and feed on seagrass. In the Gulf of Mexico redhead duck (*Aythya americana*) eats *Halodule wrightii*. Swans ingest the rhizomes and leave the leaves to float off, thus affecting spatial patterns of decomposition. Dugongs (*Dugong dugon*) pull out the small plants of *Halophila*, *Cymodocea*, and *Halodule* and, in Shark Bay, Western Australia, eat *Amphibolis antarctica*. Dugongs leave circuitous trails in seagrass beds they have grazed, once again producing the potential for unique spatial patterning in community and ecosystem processes; this is considered as the possible basis for ecological interactions and stimulates seagrass growth. The green sea turtle, *Chelonia mydas*, eats seagrass and macroalgae in tropical seas. They tend to graze in “grazing plots” of *Thalassia testudinum* in the Bahamas choosing young leaves by consistent cropping. There is more digestible forage – higher in protein and lower in lignins – than ungrazed older leaves. Small fish may eat seagrass leaves, fruit, and seeds, and some small grazers, such as snails, and amphipods eat leaf tissue. Because the assimilation rate is quite low, large amounts are returned as detritus and broken down by bacteria. This interaction of vertebrates, invertebrates, bacteria, and seagrass will affect seagrass growth patterns. Some invertebrates ingest seagrass leaves, for example, leaf mining linseed isopods were found in *Posidonia* leaves with more than 90 % of leaves containing burrows. The isopods consumed mesophyll tissue and cells of the vascular bundles (Brearley and Walker 1995).

## Epiphytes and Epiphyte Grazers

Seagrass leaves provide a substratum for growth of epiphytic microalgae that fuel food webs and provide shelter for invertebrates and fishes. Mostly, the grazers on seagrass leaves eat epiphytes growing on the leaves (Fig. 3). To predict the impact of grazer-epiphyte interactions, a detailed knowledge of the main processes taking place on several spatial and temporal scales is required. Results cannot be extrapolated from one site to another, and knowledge of recruitment dynamics, the influence of species and morphology of seagrasses on epiphytes and grazers, and the dietary requirements of grazers must be determined for a full understanding of these complex interactions (Jernakoff et al. 1996).

Epiphyte biomass is enhanced by eutrophication more than seagrass biomass, providing the potential for greater optical depths of epiphytes on leaf surfaces and greater extinction of the photon flux required to drive seagrass photosynthesis. Indications of eutrophication may be excessive growth of green and red macroalgae such as *Ulva*, *Enteromorpha*, and *Gracilaria*; algal blooms of phytoplankton can also appear. In the marine environment it is nitrogen that is most limiting, so



**Fig. 3** *Posidonia australis* fruits, note the epiphytes of macroalgae and calcareous polychaetes on the healthy seagrass leaves (Photograph: H. Kirkman)

increased nitrogen stimulates opportunistic algae. The nitrogen, as nitrate and ammonium, enters coastal ecosystems through agricultural runoff, untreated sewage, urban runoff, and land-based pollution that are washed into rivers. The resulting macroalgal blooms that form in eutrophied waters may also form floating rafts, forming an optical filter over beds of underlying seagrass. Another factor that affects the load of epiphytes on seagrass leaves is self-cleaning by the leaves brushing against each other when there is water movement. Epiphytes will accumulate more on seagrasses with stems such as *Amphibolis* and *Heterozostera nigricaulis*, because the stems have been in the water column longer than the leaves. Older leaves will attract more epiphytes than younger leaves.

### Complex Food Webs Associated with Seagrass Ecosystems

Epiphyte grazers are part of a complex food web starting with the primary producers – the macroalgal and microalgal epiphytes. There may be hundreds of species of macroalgae on seagrass leaves and stems. Borowitzka et al. (1990) found over 150 species of multicellular algae and over 40 species of sessile invertebrates growing epiphytically on *Amphibolis griffithii* at three widely spaced sites in southern Western Australia. The plant epiphytes are grazed by a multitude of small invertebrates including snails and amphipods. These invertebrates are preyed on by other snails, fish, isopods, and starfish. The grazing fish are preyed on by octopus and larger fish which may be eaten by yet larger fish, sharks, seals, and humans. At the same time the seagrass leaves and aboveground parts are used for protection from predators by many organisms. The pipefish (*Stigmatopora argus*) is well camouflaged in *Posidonia* leaves (Fig. 4), and there are other



**Fig. 4** Pipefish *Stigmatopora argus* on *Posidonia coriacea* (Photograph: H. Kirkman)

invertebrates that are camouflaged, e.g., isopods, snails, and nudibranchs. Juvenile shrimp use seagrass beds as nursery areas, and the tiger prawn (*Penaeus monodon*) in the Gulf of Carpentaria in Australia is only caught in seagrass beds.

The effects of overfishing on seagrass beds can be quite devastating. A top-down trophic cascade can occur when the top-level predators are removed. The decline in large predators brought about by fishing causes an increase in small-fish predators which deplete populations of mollusc and crustacean grazers that normally reduce epiphyte loads. Thus, excessive fishing of some upper trophic level fish has the potential to cause cascading effects down the food web, which ultimately decrease productivity in the primary producers. This process may have more steps in a complex food web, but the end result is that seagrass leaves are smothered by epiphytes reducing the light falling on the seagrass leaves, and if the available light falls below the compensation point (the light level required to sustain a positive carbon balance in the plant), the plants will eventually die (Heck and Valentine 2007). The threat of a trophic cascade caused by recreational and commercial fishing should always be considered.

Under pristine conditions, the older the leaves the more epiphytes there are. In temperate regions, plants like *Posidonia* and *Amphibolis*, which have longer leaf retention times, may hold more epiphytes than the shorter-lived leaves of *Halophila*. Similarly, in the tropics *Enhalus* will hold more epiphytes than *Halodule* or *Halophila*.

The prolific diversity and abundance of motile, epibenthic, invertebrate fauna found in seagrass beds can be illustrated by beam trawls through the seagrass at night when the animals are above the substrate (Fig. 5). A beam trawl for this purpose is usually a meter wide with a roller to prevent damage to seagrass and has skids to move it easily over the seagrass vegetation. The net is usually 2 mm with a





**Fig. 5** The animals from a 50 m beam trawl through a *Posidonia australis* bed at Kangaroo Island in South Australia (Photograph: H. Kirkman)

1 mm cod end. The beam trawl is pulled along the bottom at about 2–3 km/h for 50 m collecting all the animals from 50 m<sup>2</sup>. An example of the difference between the abundance and diversity of epibenthic fauna in seagrass and on unvegetated sediment was shown in the Albany harbors in Western Australia. In Princess Royal Harbour 18, 50 m beam trawl samples on unvegetated sand caught 258 individuals from 23 species, whereas nearby, in a *Posidonia australis* bed, 3,923 individuals were caught from 68 species (Kirkman et al. 1991). The species collected were amphipods, fish, isopods, molluscs including octopus and squid, and sea cucumbers, brittle stars, and starfish in the echinoderms.

The effect of human impacts on food webs is described by Coll et al. (2011) for temperate Atlantic seagrass beds. They found that the food-web structure was similar among low-impact sites in Eastern Canada and a tropical seagrass web suggesting consistent food-web characteristics across seagrass ecosystems at different latitudes.

### Threats to the Future Vitality of Seagrass Ecosystems

Lack of light is most likely the main cause of global seagrass loss. There are several reasons for reductions in light in seagrass beds. Low light at the deeper edge of a seagrass bed is usually caused by turbidity in the water column, which stirs sediments and thus sets a limit to the depth at which the seagrasses can grow. Observations of dynamics in the position of the deeper edge of a seagrass community can be used to describe a great deal about the condition of the seagrass bed and its susceptibility to water quality. At the shallower edge prolific growth of epiphytes will shade seagrasses and reduce their potential for growth and biomass maintenance. Once again, observations of dynamics in the state of the epiphytic cover can

be used to track ecosystem vitality over time. Either the epiphytes can be monitored regularly or the border of the seagrass bed can be progressively marked and recorded. These simple measurements at the outer and inner boundaries of the seagrass bed will assist with management.

Human impacts on seagrasses are well discussed in Ralph et al. (2007). Runoff from land clearing in preparation for housing and urban construction may be the largest impact on offshore seagrass meadows. The problem is that the land is cleared for building and sometimes heavy rains wash off the topsoil because it is no longer held by vegetation. New roads and cuttings for roads are another source of sediment runoff. Both of these influences will affect water turbidity and the potential for seagrass growth by threatening light penetration to the seagrass beds.

In Western Port, Victoria, Australia, beds of the subtidal *Zostera muelleri* have been progressively reduced in coverage for the past 50 years. The causes are difficult to remediate. Erosion from clay cliffs and the shore generally and runoff from streams and drains have put sediment into the water column. The continual loss of seagrass has given rise to larger areas of unvegetated mud which is disturbed in rough weather thus adding to the suspended solids and increasing turbidity. Reducing erosion from the cliffs is expensive. Terrestrial runoff is due to poor farming practices and considerable urban development in the catchment and the loss of vegetated stabilized area continues to exacerbate the problem. Attempts are being made to grow mangrove as a sediment stabilizer outside the boundary of seagrass beds and thus reduce wave energy causing erosion.

Development of the coast by building causeways and shoreline armoring may divert water and generally destabilize beaches and shorelines. Rivers are often diverted or changed to enable the extraction of freshwater, and this may have an effect on seagrass beds by favoring one species that prefers seawater (*Heterozostera tasmanica*) over *Zostera muelleri* that has adapted to changed salinity conditions.

Physical damage to seagrass beds can occur when marinas, jetties, and boat ramps are built on or adjacent to seagrass beds, or these structures may change the dynamic hydrology (water circulation patterns) of the area, reducing onshore drift and water flow. Onshore drift is the gradual lateral shift along a beach of beach material resulting from waves meeting the shore at an oblique angle. Mining or oil and gas extraction from under seagrass beds are potentially damaging when considering freshwater flows, oil spills, and mining accidents that cause collapse of mined areas. Moorings and boat ramps add further problems for seagrass ecosystems. The moorings cut spheres in the seagrass bed by chain movement caused by tides and wind. Boat ramps lead to channels being cut in seagrass beds by boat propellers at low tides when boats are leaving or returning to the ramp. Adequate channel markers and a channel will help to prevent this. The main problem with propeller scouring is that during tidal cycles water washes in and out through these rills and these are eroded to form quite large channels in which seagrass propagules are prevented from colonization.

Human occupation of the coastal zone is accompanied by increased rates of pollution. Industrial chemicals from factories, including heavy metals, petrochemicals, and toxic compounds, are a danger to seagrass ecosystems. These pollutants



enter the sea from runoff and storm water drains. Agricultural runoff containing herbicides and insecticides can damage seagrass beds and its associated fauna.

By far the most damaging pollutant in seagrass beds is nutrients. These nutrients promote epiphyte growth that smothers the photosynthetic potential of seagrasses and reduces dissolved oxygen levels to dangerously low levels. In marine systems nitrogen excess is usually the primary culprit. Eutrophication occurs when high nutrient loads, particularly inorganic nitrogen, are taken up by opportunistic macroalgae growing on seagrass leaves. The epiphytes and dead seagrass leaves fall to the substrate and are broken down by bacteria that use up oxygen, and this anoxic sediment gives off hydrogen sulfide that kills the benthic flora. The whole seagrass ecosystem may then collapse. Food-web structure and functioning of seagrass habitats change with human impacts, and the spatial scale of food-web analysis is critical for determining results (Coll et al. 2011). The spatial scale is a relevant issue in food-web ecology in general as food webs are typically assembled in aggregated forms (cumulative or summary webs) due to limited data availability on trophic interactions.

Dredging near seagrass beds increases turbidity, and this may cause a smothering effect as well, especially if silt screens are not used. If the sediment load is very high, the effect of seagrass leaves slowing the surrounding water will cause the sediment to drop out of the water column and smother plants. Dredging should generally be carried out in the season when seagrass is least productive, for example, in temperate regions in winter, after carbohydrates and stored material have been laid down in rhizomes or, in the tropics, in the wet season when seagrass beds may die out due to low light because of high sediment loads caused by terrestrial runoff and disturbance of the substrate. They recover naturally during the dry season.

Globally, disease in seagrasses has not been identified as a major threat. After the dramatic reduction of the seagrass *Zostera marina* in the 1930s in the USA and Europe, recovery was slow and only occasionally has *Labyrinthula zosterae*, a marine slime mold-like protist been shown to cause large-scale losses. The death of seagrasses was attributed to *Labyrinthula zosterae*, but later it was established that the plants were under stress and the disease proliferated because of the low resistance of the seagrasses. Diligent monitoring of seagrass beds will alert managers to conditions that could foster secondary impacts due to disease.

Many of the seagrass beds in the USA and Europe provided insulation material from the leaves of *Zostera marina* in the 1920s. The dried leaves, usually recovered from drift on beaches, were used as insulation in sleeping bags and the walls of houses. Collections of large amounts of drift material may affect the nutrient recycling of seagrass beds. There are numerous reports of the slow rate at which seagrass beds will recover from disturbance. One of these is in Spencer Gulf in South Australia where *Posidonia australis* plants were removed to obtain the underlying fiber. This fiber was from the persistent fibrous remains of old leaf sheaths of *P. australis* and was used in clothing manufacture and for insulation in refrigeration units and steam-heating systems. It is of interest to note that although this mining was discontinued in the 1920s, the scars where dredges removed the

fiber are still visible today. This and other evidence from seismic blasting suggests that *Posidonia* spp. beds take decades to recover.

Invasive species are a problem in seagrass meadows in some parts of the world. Of particular note is the damage done by *Caulerpa taxifolia* in *Posidonia oceanica* seagrass beds in the Mediterranean Sea (Meinesz, et al. 1993). Some consideration should be given to other invasive species that may arrive, e.g., *Undaria* and *Asterias* are potential invaders that could pose problems in the future. *Undaria pinnatifida* is an edible kelp called wakame, from Japan, that has invaded seagrass beds and rocky temperate reefs. *Asterias amurensis* is the Northern Pacific seastar also from Japan that removes all organisms from reefs and is also found in seagrass beds.

The full extent of climate change effects on seagrass ecosystems has not yet been demonstrated or predicted. However, given the changes that have been noted to date in ocean temperature, salinity, acidification and aragonite saturation, sea level, circulation, productivity, and exposure to damaging UV light, we can anticipate significant degradative effects due to climate change in the future. Loss of seagrass coverage due to exposure to extremes in sunlight or heat has recently been shown in South Australia (Seddon et al. 2000).

Indirect effects of climate change on seagrass communities could occur due to intensification and increases in the frequency of tropical and subtropical cyclones. As discussed above, storms stir up sediment in shallow seas and hence reduce light to seagrass. Increased storm frequency means that there will be increased turbidity and this may reduce light to lower than compensation levels for marginal meadows at the deeper edge. Increased frequency of storms may also disturb seed beds that normally lie in the sediment, e.g., *Halophila ovalis* and *Halodule uninervis* were lost from Hervey Bay, Queensland, when two very large storms followed each other, the first destroying the seagrass and the second destroying newly germinated seedlings (Preen et al. 1995). It took about 5 years for the area to recover. More intense storms will also increase erosion of edges.

Warmer temperatures and ice cap melting are expected to raise sea levels. For seagrasses this will bring their habitats shoreward. Those seagrasses growing at the deeper edge of their habitat may be lost while the shallower margins will gain coverage. The problem is if development has used those shallower edges to the point that the seagrass can move no further up the shore, large areas will be lost. The building of sea walls, coastal roads, housing to the edge of the sea, and other development must be carefully managed with sea-level rise in mind.

Little is known about the effect of seawater temperature rising, but shifts in distribution are expected. Seagrass plants cannot move as can some invertebrates and fish as the water temperature increases. The success of a slow distributional shift will depend upon the suitability of a new habitat being available, the connectivity between seagrass beds and potential new growth areas, and the dispersal mechanisms of the propagules.

As carbon dioxide rises in the atmosphere, more is dissolved in seawater leading to ocean acidification. In seagrass ecosystems, calcareous epiphytes will be the main victims. The response of calcareous epiphytes to a fall in pH from 8.2 (seawater) to 7.7 in aquaria was a loss of all calcareous algae, and the only calcifiers

were bryozoans at pH 7.7 (Martin et al. 2008). This result may have dramatic effects on biogeochemical cycling of carbon and carbonate in coastal ecosystems dominated by seagrass beds.

## Restoration and Recovery

There is considerable confusion in the natural-resource management field about the terms “rehabilitation” and “restoration.” Dictionaries generally tend not to differentiate between the two (e.g., see Shorter Oxford English Dictionary) nor do many learned articles on the creation of new seagrass habitats, but there is a distinction worth making, especially with degraded ecosystems. “Restoration” could mean “*reversion of a degraded ecosystem to its original condition*” or “*inducing and assisting abiotic and biotic components of an environment to recover to the state that they existed in the unimpaired or original state.*” This is acknowledged as being an unlikely outcome in practice.

In contrast, “rehabilitation” describes an *acceptable improvement in ecological condition* and, in most cases, is a more realistic management objective. Rehabilitation of degraded, seagrass beds is where management interventions are expected to markedly improve the ecological condition of these systems and allow them to again deliver, in broad terms, the sorts of ecosystem services that humanity expects but are never intended to return the system to some notional “pristine” condition. From rehabilitation one could distinguish three types of management outcomes: (i) maintenance, (ii) improvement, and (iii) reconstruction. In this scheme, reconstruction broadly equates with restoration, and improvement with rehabilitation.

An associated discipline is ecological engineering, which involves restoring and creating sustainable seagrass ecosystems that have value to humans and nature. Ecological engineering should restore/rehabilitate damaged seagrass ecosystems and create new sustainable systems in a cost-effective way.

The term “mitigation” refers to the enhancement or creation of seagrass areas to compensate for permitted seagrass losses. Offsets may be used when a seagrass bed is sacrificed for a shipping channel, land claim, or development that destroys a seagrass bed and the bed cannot be restored or moved somewhere else.

Planting success may be defined in a number of ways. First, sometimes success is claimed if seedlings grow sufficiently to produce their own reproductive structures, and their canopy, covering the area planted, is similar to a nearby unaffected seagrass bed. Second, criteria, preferably measurable as quantitative values, could be established prior to the commencement of planting activities. Success can then be defined as the successful integration of plant material establishment with fishery and wildlife habitat establishment and water quality improvements. This habitat equivalence can be measured with quantitative measures such as species presence in conjunction with plant cover. The habitat measurements are compared with a proximate seagrass ecosystem. A third approach might be to set a numerical target for survival over a given period, e.g., 70 + % survival of planted seedlings or transplants after 1 year.

Environmental offsets are measures to compensate for the adverse impacts of an action on the environment. Offsets do not reduce the impacts of an action: instead they provide environmental benefits to counterbalance the impacts that remain after avoidance and mitigation measures. These remaining impacts are termed “residual impacts.” Offsets are not intended to make proposals with unacceptable impacts acceptable. In assessing the suitability of an offset, government decision making should be informed by scientifically robust information and conducted in a consistent and transparent manner.

More specifically, offsets are measures to compensate for environmental impacts on seagrass ecosystems that cannot be adequately reduced through avoidance or mitigation. Offsets for seagrass ecosystems can help to achieve long-term conservation outcomes for protected areas, while providing flexibility for proponents seeking to undertake an action that will have unavoidable environmental impacts. For example, if a seagrass area is to be dredged or claimed for development, the seagrass that is to be destroyed could be collected and planted somewhere else where seagrass was known to have previously survived and is suitable for restoration.

A major difficulty in restoring seagrass ecosystems is the difficulty of obtaining suitable propagules. Sometimes seeds are unavailable or scarce such as in the genus *Syringodium* in Australia, the USA, and Caribbean or where seeds are plentiful such as in *Zostera muelleri*, in Australia, but the germination rate is low. Some genera produce viviparous seedlings and no seeds are seen, e.g., *Amphibolis* and *Thalassodendron* in Australia. *Posidonia* produces a buoyant fruit (Fig. 3) from which a seedling falls after floating for a few days. These seedlings, although numerous, present problems when attempting to restore large areas. *Posidonia oceanica* does not regularly produce copious quantities of seedlings in the Mediterranean. *Amphibolis* and *Heterozostera* produce adventitious roots from their stems, and these are useful natural propagules when the stems break off the plant and float away to eventually sink in a suitable environment.

Seagrass transplanting is well known for its failure arising from a number of causes, such as planting at sites where seagrass had no history of growing; disturbance of the substrate by burrowing animals (bioturbation), storms, insufficient light, lack of knowledge, and experience by those transplanting; and other local reasons.

In the absence of natural recruitment, sprigs or seedlings may need to be sourced from a donor site some distance away. An understanding of levels of genetic diversity and spatial genetic structure can contribute to improved restoration outcomes by identifying the most genetically appropriate source material for restoration sites.

## Genetic Diversity

The poor knowledge of the minimal habitat requirements for seagrass growth, colonization and establishment mechanisms, genetic diversity, and reproductive

modes required to maintain ecologically successful populations hinders the development of sound management practices. The development of molecular DNA sequencing techniques over the last decade has provided new tools to examine genetic variability within and among seagrass populations. Much of the power inherent in molecular genetic data can be tapped, revealing otherwise unobtainable information at all levels of biotic hierarchy (Kendrick et al. 2005).

Alberte et al. (1994) assisted with breakthroughs in determining that populations that were morphologically distinct and may have shown different depth distributions could be distinguished by DNA fingerprinting. They also determined that *Zostera marina*, in particular, was not characterized by a high degree of clonal reproduction at spatial scales over 5 m, and they found that *Z. marina* growing in a physically disturbed bay had reduced genetic diversity. Knowing the effect that disturbance has on genetic stability can help establish mitigation and restoration criteria.

Genetic diversity in terms of greater numbers of distinct clones was positively associated with seagrass bed density, and this in turn was correlated with greater invertebrate density, nitrogen retention, and areal productivity. Higher abundances of invertebrates associated with seagrasses in more genetically diverse *Zostera* plots and the positive effects of seagrass genotypic diversity on both seagrass and grazer biomass depended on grazer species identity. Since mesograzers can have strong effects on the biomass of both epiphytic algae and seagrasses, and since seagrass genotypes vary in palatability, understanding the implications of changing diversity in seagrass ecosystems will require more detailed study of genetic and species diversity effects at multiple trophic levels. Nevertheless, the picture emerging from controlled experiments and seagrass restoration projects appears consistent: seagrass genetic diversity may be a key variable influencing seagrass productivity and community processes (Duffy et al. 2013).

There is also a positive impact of clonal diversity along an entire depth gradient on food-web complexity and density and nutrient retention. Ecosystem restoration will significantly benefit from obtaining sources (transplants and seeds) of high genetic diversity and from restoration techniques that can maintain that high genetic diversity (Reynolds et al. 2012).

Seagrasses provide convincing examples of the broader ecological importance of genetic or genotypic diversity. Higher allelic diversity within individuals increased vegetative shoot production and sexual reproduction in transplanted seagrasses, and transplant success correlated positively with the genetic diversity of individuals in the source population (Procaccini et al. 2007). More convincing was the evidence from experimental manipulations of the number of seagrass genotypes (as measured by DNA microsatellites), which demonstrated that genetic diversity within a patch can influence primary and secondary production, particularly in the face of disturbance or stress. Patches of eelgrass (*Zostera marina*) with greater numbers of clonal genotypes were more resistant to seasonal grazing by migratory geese, resulting in increased shoot density after grazing in high-diversity areas and quicker recovery to pre-grazing densities, in the more diverse areas. Genotypic (and thus phenotypic) diversity also increased the rate of recovery

from extremely high water temperatures in *Zostera marina* suggesting that this effect may be a generalized response to aboveground biomass removal. Subsequent manipulations that controlled for disturbance confirmed the positive effects of genetic diversity in the presence and absence of disturbance. Thus there is growing evidence, albeit only from *Zostera* so far, that genetic diversity within seagrass species can be important in buffering seagrasses from several types of perturbations. Genotypic diversity can have positive consequences at the community level as well.

It is only recently that one has begun to understand the genetics of seagrass plants and what a seagrass plant is. In Western Australia vast beds of *Posidonia* extend for kilometers along the coast; until now it has not been possible to say how extensive a single plant is. *Posidonia oceanica* in the Mediterranean is one of the largest, slowest growing, and longest-lived plants terrestrially or in the sea. In a recent genetic study of 40 *P. oceanica* populations across the Mediterranean, Arnaud-Haond et al. (2012) found individual clones spanning up to 15 km. Based on the plant's known growth rate, such individuals are likely to be thousands, possibly tens of thousands of years old. This was different from the high degree of clonal reproduction in *Zostera marina* shown by Alberte et al. (1994).

The discoveries made by DNA have also helped untangle some of the taxonomic identities of seagrass. It is at this point that an understanding of levels of genetic diversity and spatial genetic structure can contribute to improved restoration outcomes. Identifying the most genetically appropriate source material for restoration sites can be carried out with DNA analysis.

From molecular studies in combination with ecological and hydrological assessments, it is evident that seagrasses are resilient and have persisted in a physiologically challenging submerged environment because they have broad niches. That local persistence of seagrasses has been achieved by clonal growth and by recruitment from sexually derived propagules. Some seagrasses invest significant amounts of energy in sexual reproduction, producing seeds with a high capacity for long-distance dispersal that enables them to colonize distant new locations (Kendrick et al. 2012).

## Future Directions

There is a recent trend for widespread loss in tropical and temperate seagrass ecosystems. Large-scale declines have been reported by Hemminga and Duarte (2000) at 40 locations, 70 % of which are attributed to human induced disturbance. There are some areas that have recovered but the long-term trend is for continual global loss. Short and Wyllie-Echeverria (1996) estimated the area of seagrass lost globally at 12,000 km<sup>2</sup> or about 2 % of the area originally covered. Present losses are expected to accelerate, particularly in areas of Southeast Asia and the Caribbean where human pressure is greatest and development incentive is greater than environmental conservation. Restoration of seagrasses seems to be the greatest challenge facing ecologists. Efforts to restore seagrass need to be based on knowledge

of local conditions, the ecological state of the system prior to disturbance, and informed decisions about what should be there after restoration. The genetic investigations into clonal seagrass identity may be helpful in restoration efforts.

It is difficult to separate natural variability from human-caused disturbance. The role of disturbance and the response by seagrass species to a particular disturbance should be a major focus of long- and short-term research. Now that climate change is a component of disturbance, the investigation has become even more complex. It is recommended that monitoring of seagrass to distinguish between these causes and to answer relevant questions on management of seagrass ecosystems be carried out.

As concern increases for the state of natural resources and the degradation of the world's oceans, it is critical for countries to progress with conservation actions specifically focused on seagrass ecosystems. Guidelines for Applying the IUCN Protected Area Management Categories to Marine Protected Areas (MPA) aim to make clear what is most significant and of highest priority, and this effort will help countries more accurately detail their successes ([www.iucn.org/pa\\_guidelines](http://www.iucn.org/pa_guidelines)). These guidelines will define MPAs thus preventing the trend of fisheries advisory bodies claiming that area mechanisms exploiting fish are MPAs. About 50 % of global MPAs are considered to have been wrongly allocated because the name of the MPA, e.g., National Park and Sanctuary, has been used to determine the category, rather than the management objectives. Confusion tends to arise when sites have been incorrectly assigned on the basis of activities that occur, rather than using the stated management objectives. In recent years pressure to deliver success stories has resulted in false claims of large areas of seagrass being properly protected. It is time to be realistic about our definition of MPAs in seagrass ecosystems.

Protecting seagrass beds through education of local communities and fishers and by regulations and even enforcement will help conserve this valuable resource. Properly regulated marine protected areas will assist with conserving seagrass ecosystems with benefits to conserving biological diversity and spillover advantages to nonprotected areas.

It is time to stop pretending more areas of seagrass are protected than they actually are. Understanding which seagrass beds are protected and how they are protected is of paramount importance in promoting driving global conservation efforts. Without this information it is difficult to hold the process of determining marine protected areas in seagrass ecosystems accountable.

---

## References

- Alberte RS, Suba GK, Procaccini G, Zimmerman RC, Fain SR. Assessment of genetic diversity of seagrass populations using DNA FINGER printing: implications for population stability and management. *Proc Natl Acad Sci USA*. 1994;91:1049–53.
- Arnaud-Haond S, Duarte CM, Diaz-Almela E, Marba N, Sintés T, Serrae EA. Implications of extreme life span in clonal organisms: millenary clones in the threatened seagrass *Posidonia oceanica*. *PLoS ONE*. 2012;7(2):e30454.

- Barbier EB, Hacker SD, Chris Kennedy C, Koch EW, Stier AC, Silliman BR. The value of estuarine and coastal ecosystem services. *Ecol Monogr.* 2011;81(2):169–93.
- Borowitzka MA, Lethbridge RC, Charlton L. Species richness, spatial distribution and colonization pattern of algal and invertebrate epiphytes on the seagrass *Amphibolis griffithii*. *Mar Ecol Prog Ser.* 1990;64:281–91.
- Brearley A, Walker DI. Isopod miners in the leaves of two Western Australian *Posidonia* species. *Aquat Bot.* 1995;52:163–81.
- Carruthers TJB, Dennison WC, Longstaff BJ, Waycott M, Abal EG, McKenzie LJ, Lee Long WJ. Seagrass habitats of north east Australia: models of key processes and controls. *Bull Mar Sci.* 2002;73(3):1153–69.
- Coll M, Schmidt A, Romanuk T, Lotze HK. Food-web structure of seagrass communities across different spatial scales and human impacts. *PLoS ONE.* 2011;6(7):1–13.
- Duarte CM, Marba N, Gacia E, Fourqurean JW, Beggins J, Barron C, Apostolaki ET. Seagrass community metabolism: assessing the carbon sink capacity of seagrass meadows. *Glob Biochem Cycles.* 2010;24:GB4032.
- Duffy JE, Hughes AR, Moksnes P-O. *Ecology of seagrass communities.* Sunderland: Sinaur Associates; 2013. p. 271–97.
- Green EP Short FT. *World Atlas of Seagrasses Prepared by the UNEP World Conservation Monitoring Centre, University of California Press, Berkeley, USA.* 2003. pp 298.
- Hansen JCR, Reidenbach MA. Seasonal growth and senescence of a *Zostera marina* seagrass meadow alters wave-dominated flow and sediment suspension within a coastal bay. *Estuar Coasts.* 2013;36:1099–114.
- Heck Jr KL, Valentine JF. The primacy of top-down effects in shallow benthic ecosystems. *Estuar Coasts.* 2007;30(3):371–81.
- Hemminga M, Duarte CM. *Seagrass ecology.* Cambridge, UK: Cambridge University Press; 2000. 298 pp.
- Jernakoff P, Brearley A, Nielsen J. Factors affecting grazer-epiphyte interactions in temperate seagrass meadows. *Oceanogr Mar Biol Annu Rev.* 1996;34:109–62.
- Kendrick GA, Marba N, Duarte CM. Modelling formation of complex topography by the seagrass *Posidonia oceanic.* *Estuar Coast Shelf Sci.* 2005;65:717–25.
- Kendrick GA, Waycott M, Carruthers TGB, Cambridge ML, Hovey R, Krauss SL, Lavery PS, Les DH, Lowe RJ, Mascaró O, Vidal OM, Ooi JLS, Orth RJ, Rivers DO, Ruiz-Montoya L, Statton J, van Dijk JK. and J. Verduin, J.J. The central role of dispersal in the maintenance and persistence of seagrass populations. *BioScience.* 2012;62(1):56–65.
- Kirkman H, Reid DD. A study of the role of a seagrass *Posidonia australis* in the carbon budget of an estuary. *Aquat Bot.* 1979;7:173–83.
- Kirkman H, Humphries P, Manning R. The epibenthic fauna of seagrass beds and bare sand in Princess Royal Harbour and King George Sound, south-western Australia. In: Wells FE, Walker DI, Kirkman H, Lethbridge R, editors. *Proceedings of the Third International Marine Ecological Workshop: The Marine Flora and Fauna of Albany, Western Australia;* Perth: Western Australian Museum; 1991. p. 553–63.
- Kuo J, den Hartog C. Seagrass taxonomy and identification key. In: Short FT, Coles RG, editors. *Global seagrass research methods.* Amsterdam: Elsevier; 2001. p. 31–58.
- Larkum AWD, Den Hartog C. Evolution and biogeography of seagrasses. In: Larkum AWD, McComb AJ, Shepherd SA, editors. *Biology of seagrasses a treatise on the biology of seagrasses with special reference to the Australian region.* Amsterdam: Elsevier; 1989. p. 113–56.
- Martin S, Rodolfo-Metalpa R, Ransome E, Rowley S, Buia M-C, Gattuso J-P, Hall-Spencer J. Effects of naturally acidified seawater on seagrass calcareous epibionts. *Biol Lett R Soc.* 2008;4(6):689–92.
- Mateo MA, Cebrián J, Dunton K, Mutchler T. In: Larkum AWD, Orth RJ, Duarte CM, editors. *Seagrasses: biology, ecology and conservation. Carbon Flux in Seagrass Ecosystems.* Dordrecht: Springer; 2007. p. 159–92.



- McConichie CA, Knox RB. In: Larkum AWD, McComb AJ, Shepherd SA, editors. *Biology of seagrasses a treatise on the biology of seagrasses with special reference to the Australian region. Pollination and Reproductive Biology of Seagrasses*. Amsterdam: Elsevier; 1989. p. 74–111.
- Meinesz A, de Vaugelas J, Hesse B, Mari X. Spread of the introduced tropical marine alga *Caulerpa taxifolia* in northern Mediterranean waters. *J Appl Phycol*. 1993;5:141–7.
- Preen AR, Lee Long WJ, Coles RG. Flood and cyclone related loss, and partial recovery, of more than 100 km<sup>2</sup> of seagrass in Hervey Bay, Queensland, Australia. *Aquat Bot*. 1995;52:3–17.
- Procaccini G, Olsen JL, Reusch TBH. Contribution of genetics and genomics to seagrass biology and conservation. *J Exp Mar Biol Ecol*. 2007;350:234–59.
- Ralph PJ, Durako MJ, Enríquez S, Collier CJ, Doblin MA. Impact of light limitation on seagrasses. *J Exp Mar Biol Ecol*. 2007;350:176–93.
- Reynolds LK, McGlathery KJ, Waycott M. Genetic diversity enhances restoration success by augmenting ecosystem services. *PLoS ONE*. 2012;7(6):1–7.
- Sculthorpe CD. *The biology of aquatic vascular plants*. London: Edward Arnold; 1969.
- Seddon S, Connolly RM, Edyvane KS. Large-scale seagrass dieback in northern Spencer Gulf, South Australia. *Aquat Bot*. 2000;66:297–310.
- Short FT, Wyllie-Escheverria S. Natural and human induced disturbance of seagrasses. *Environ Conserv*. 1996;23:17–27.

## Further Reading

- Butler A, Jernakoff P. *Seagrass in Australia: strategic review and development of an R. and D. plan*. Collingwood: CSIRO Publishing; 1999. 210 pp.
- Connell SD, Gillanders BM, editors. *Marine ecology*. Melbourne: Oxford University Press; 2007. p. 595–630.
- Duarte CM, Chiscano CL. Above ground and below ground seagrass biomass vs degrees of latitude. *Aquat Bot*. 1999;65:159–74.
- Duffy JE. Biodiversity and the functioning of seagrass ecosystems. *Mar Ecol Prog Ser*. 2006;311:233–50.
- Larkum AWD, Orth RJ, Duarte CM. *Seagrasses: biology, ecology and conservation*. Dordrecht: Springer; 2007. 691 pp.
- Short FT, Coles RG. *Global seagrass research methods*. Amsterdam: Elsevier; 2001. 473 pp.
- Waycott M, Duarte CM, Carruthers TJB, Orth RJ, Dennison WC, Olyarnike S, Calladinea A, Fourqurean JW, Heck Jr KL, Hughes AR, Kendrick ARG, Kenworthy WJ, Short FT, Williams SL. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc Natl Acad Sci USA*. 2009;106(30):12377–81.

Richard J. Geider, C. Mark Moore, and David J. Suggett

## Contents

Introduction .....	485
Phytoplankton Diversity .....	486
Picophytoplankton .....	489
Nanophytoplankton .....	491
Microphytoplankton .....	491
Plankton Functional Groups and Trait-Based Phytoplankton Ecology .....	493
Ecological Roles and Functions .....	493
Emergent Biogeochemical Properties .....	494
Trait-Based Phytoplankton Ecology .....	495
Characteristics of the Pelagic Environment .....	496
Temperature, Salinity, and Density .....	496
Vertical Light Attenuation and Ocean Color .....	499
Vertical Stratification and Mixing .....	500
Vertical Nutrient Distributions .....	501
Dissolved Inorganic Carbon .....	502
Ocean Acidification .....	503
Primary Production .....	503
Net and Gross Primary Production of Marine Phytoplankton .....	505
Net Community and Export Production .....	508
Is the Oligotrophic Ocean Autotrophic? .....	509
Remote Sensing of Primary Production .....	509

---

R.J. Geider (✉)

School of Biological Sciences, University of Essex, Colchester, Essex, UK

e-mail: [geider@essex.ac.uk](mailto:geider@essex.ac.uk)

C.M. Moore

Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK

e-mail: [c.moore@noc.soton.ac.uk](mailto:c.moore@noc.soton.ac.uk)

D.J. Suggett

Functional Plant Biology & Climate Change Cluster, University of Technology, Sydney, NSW, Australia

e-mail: [David.Suggett@uts.edu.au](mailto:David.Suggett@uts.edu.au)

The Photosynthesis–Irradiance Response Curve .....	511
Phytoplankton Ecology .....	513
The Annual Phytoplankton Production Cycle in Temperate Zone Waters .....	514
Latitudinal Dependence of the Production Cycle .....	516
Nutrient Limitation .....	518
Geographical Patterns of Nutrient Limitation in the Ocean .....	519
Adaptations to Nutrient Limitation .....	521
Anthropogenic Impacts on Marine Phytoplankton .....	523
Future Directions .....	527
References .....	529

## Abstract

- Marine phytoplankton account for about 45 % of global net primary production (NPP). In addition, they perform other important biogeochemical functions including nitrogen fixation, calcium carbonate precipitation, and the production of climatically active gases such as dimethyl sulfide.
- Oceanographers employ a wide variety of platforms for studying marine phytoplankton ecology, including sampling from ships, sampling from autonomous remotely operated vehicles, and collecting observations from Earth-orbiting satellites.
- Marine phytoplankton range in size from  $<1 \mu\text{m}$  in diameter to about 1 mm in length and include representatives from at least five eukaryotic phyla together with the cyanobacteria. This wide size range and phylogenetic diversity presents challenges for quantifying and characterizing phytoplankton communities.
- Functional traits that quantify responses of growth rate, photosynthesis and nutrient uptake to temperature, irradiance, and nutrient availability provide a useful basis for understanding phytoplankton ecology.
- A variety of complementary approaches are used to measure gross and net primary production. These include measuring production of  $\text{O}_2$  and organic matter in bottle experiments and measuring diel and seasonal changes of  $\text{O}_2$  in open waters. Information obtained from satellite remote sensing of ocean color is used to calculate NPP on regional and global scales.
- The physical and chemical variables that drive NPP include temperature, nutrient availability, and solar radiation. These vary in time and space, and our understanding of this variability is largely encapsulated in the concepts of the seasonal production cycle and marine biogeochemical provinces.
- Nutrient limitation sets an upper limit to NPP over most of the ocean surface, with either inorganic iron or nitrogen being the proximate limiting element in different regions.
- The upper water column is stably stratified over much of the ocean, and pronounced vertical gradients of light and nutrients lead to depth separation of ecotypes with differing adaptations to nutrient availability and the light environment.

- Although the growth of individual phytoplankton cells is often limited by temperature and the availability of nutrients and light, biotic interactions including predation and disease often control the growth of phytoplankton populations and the species composition of phytoplankton communities.
- Anthropogenic impacts on the ocean, including nutrient loading to coastal waters, climatic forcing associated with global warming, and ocean uptake of anthropogenic CO<sub>2</sub>, are influencing the chemistry and physics of the upper ocean, with multiple potential impacts on the phytoplankton.

---

## Introduction

Phytoplankton (from Greek “phytos” meaning plant and “planktos” meaning “drifting”) are the primary producers that form the base of ocean food webs and as such play vital roles in ocean ecology. Unlike terrestrial plants, most of which are macroscopic and rooted in place, phytoplankton are microscopic unicells or colonies that float in the water. Historically, research into marine phytoplankton was prompted by a desire to understand why fish stocks vary in abundance. Current research is motivated more by the need to understand how phytoplankton affect atmospheric CO<sub>2</sub> and other climatically active gases through their roles in the oceanic and global carbon and nutrient cycles.

The total biomass of phytoplankton is only about 1 % of that of terrestrial plants. As such, phytoplankton are relatively inconspicuous, although their presence and abundance can still be detected from changes in ocean color. Remarkably, despite their low biomass, marine phytoplankton are as significant as forests and grasslands to global photosynthetic CO<sub>2</sub> fixation (Table 1). Current estimates indicate that phytoplankton account for about 45 % of global net primary production.

Our understanding of the ecology of marine phytoplankton ecology has been obtained using four complementary approaches: (i) oceanographic surveys and time series, (ii) field-based perturbation experiments, (iii) laboratory culture experiments, and (iv) numerical models (Table 2). In many cases, phytoplankton are investigated within the context of multidisciplinary programs addressing wider issues in marine ecology or biogeochemistry. Oceanographic surveys document the spatial distribution of phytoplankton at a particular time, whereas time series document the seasonal and interannual changes of phytoplankton at a fixed location. Perturbation experiments are used to test hypotheses concerning the effects of altering physical, chemical, or biological variables on phytoplankton ecology. Laboratory culture experiments are used to study the physiological ecology of phytoplankton. Numerical models are used to test our understanding of phytoplankton ecology.

This chapter describes how the approaches listed in Table 2 are used to gain an understanding of primary production and phytoplankton ecology in the open sea. The chapter starts by describing the main types of phytoplankton from both taxonomic (section “[Phytoplankton Diversity](#)”) and functional

**Table 1** Annual net primary production (NPP) of the biosphere. All values are in petagrams of C (1 Pg =  $10^{15}$  g)

Marine	NPP	Terrestrial	NPP
Phytoplankton <sup>a</sup>		Forest <sup>b</sup>	
Oligotrophic	9.2	Tropical rain forests	17.8
Mesotrophic	34.8	Broadleaf deciduous forests	1.5
Eutrophic	5.6	Mixed broadleaf and needleleaf forests	3.1
Subtotal	49.6	Needleleaf evergreen forests	3.1
		Needleleaf deciduous forests	1.4
Coastal fringe <sup>c</sup>		Subtotal	26.9
Microphytobenthos	0.3	Grasslands, shrublands, and extreme environments <sup>b</sup>	
Coral reef algae	0.6	Savannas	16.8
Macroalgae	2.6	Perennial grasslands	2.4
Sea grasses	0.5	Broadleaf shrubs with bare soil	1.0
Salt marsh	0.4	Tundra	0.8
Mangrove forest	1.1	Desert	0.5
Subtotal	5.5	Subtotal	21.5
Chemosynthesis and anoxygenic photosynthesis <sup>d</sup>	0.4	Cultivation <sup>b</sup>	8.0
Total marine	55.1	Total terrestrial	56.4

<sup>a</sup>Carr et al. (2006)<sup>b</sup>Field et al. (1998)<sup>c</sup>Duarte and Cebrián (1996)<sup>d</sup>Raven (2009)

(section “[Plankton Functional Groups and Trait-Based Phytoplankton Ecology](#)”) perspectives. It then outlines the main physical and chemical characteristics of the pelagic environment (section “[Characteristics of the Pelagic Environment](#)”). The approaches that are used to measure phytoplankton production are discussed in section “[Primary Production.](#)” Section “[Phytoplankton Ecology](#)” describes how primary production varies both spatially and temporally in the ocean and explains how primary production in the ocean is limited or regulated by physical–chemical factors and biotic interactions. Section “[Anthropogenic Impacts on Marine Phytoplankton](#)” describes how anthropogenic impacts including eutrophication, climate change, and ocean acidification influence marine primary production. Section “[Future Directions](#)” outlines unresolved issues and future research directions.

## Phytoplankton Diversity

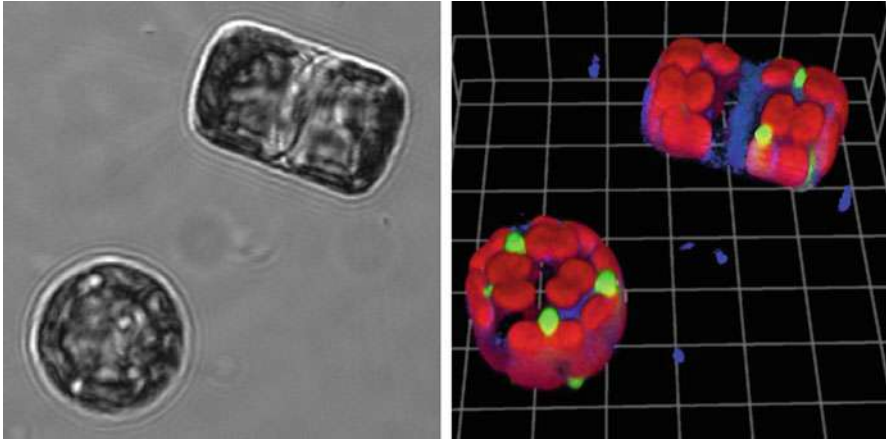
Phytoplankton cells range in size from about 0.5  $\mu\text{m}$  in diameter to  $>1$  mm in length: this is roughly the same relative difference in size as between a bumblebee (about 1 cm) and a blue whale (up to 30 m) or between a tuft of grass (50 cm high) and a redwood tree (100 m tall). The smallest phytoplankton cells have very simple spherical or elliptical shapes. The largest can be highly ornate with elaborate

**Table 2** Approaches to investigating phytoplankton ecology

Approach	Methodology		Examples
Oceanographic surveys and time series	Time-series stations	Natural community	Bermuda Atlantic time series ( <a href="http://bats.bios.edu/">http://bats.bios.edu/</a> ) Hawaii ocean time series ( <a href="http://hahana.soest.hawaii.edu/hot/">http://hahana.soest.hawaii.edu/hot/</a> )
	Transects	Natural community	Atlantic Meridional Transect ( <a href="http://amt-uk.org/">http://amt-uk.org/</a> )
	Remote sensing	Natural community	Coastal zone color scanner ( <a href="http://oceancolor.gsfc.nasa.gov/CZCS/">http://oceancolor.gsfc.nasa.gov/CZCS/</a> ) SeaWiFS ( <a href="http://oceancolor.gsfc.nasa.gov/SeaWiFS/">http://oceancolor.gsfc.nasa.gov/SeaWiFS/</a> ) Aqua/MODIS ( <a href="http://modis.gsfc.nasa.gov/">http://modis.gsfc.nasa.gov/</a> )
Field-based perturbation experiments	Microcosms (0.1–20 L)	Natural community	Nutrient limitation (Mills et al. 2004); elevated CO <sub>2</sub> and temperature (Feng et al. 2009)
	Mesocosms (1–100 m <sup>3</sup> , <a href="http://mesocosm.eu/">http://mesocosm.eu/</a> )	Natural community	Ocean acidification experiments (Riebesell et al. 2013) Eutrophication experiments (Romero et al. 2012)
	Open-water nutrient additions (25–75 km <sup>2</sup> )	Natural community	Iron fertilization experiments (Boyd et al. 2007)
Laboratory culture experiments	Phenotypic (physiological) acclimation	Algal culture	Light limitation (Falkowski et al. 1985); nutrient limitation (Sunda et al. 2009)
	Genetic adaptation	Algal culture	Adaptation to high CO <sub>2</sub> (Lohbeck et al. 2012)
Numerical models	Simulation modeling of biogeochemical and ecological processes subject to physical forcing		Nutrient, phytoplankton, zooplankton, detritus (NPZD) (Fasham et al. 1990) Plankton functional group (PFG) (Le Quéré et al. 2005) Self-assembling (Follows et al. 2007)

“spines” and/or “wings.” Given this wide range of sizes, it is often convenient to differentiate phytoplankton into size classes. One approach is to separate cells of different sizes by sequentially passing a sample through a set of filters of decreasing pore size. The most commonly employed pore sizes are 200, 20, 2, and 0.2  $\mu\text{m}$ . Colonies and very large cells that are retained on 200  $\mu\text{m}$  sieves are referred to as mesoplankton. Organisms that pass through a 200  $\mu\text{m}$  sieve but are retained on a 20  $\mu\text{m}$  pore filter are referred to as microplankton, and those that are retained by a 2  $\mu\text{m}$  pore filter but pass through a 20  $\mu\text{m}$  pore filter are referred to as nanoplankton. The cells that pass through a filter with 2  $\mu\text{m}$  pores are referred to as picoplankton.

Traditionally, quantifying phytoplankton abundance in the sea has involved collecting cells from seawater in a settling chamber or on a filter. Then, microscopy is used to count and identify individuals. Phase contrast microscopy is often used



**Fig. 1** Transmitted and fluorescence light micrographs of a centric marine diatom. The *left side* shows the brightfield image of two *Thalassiosira weissflogii* cells in different orientations, as seen using differential interference contrast. The *right side* shows a 3-dimensional reconstruction of the fluorescence signals in the same cells that arise from chloroplasts, nuclei, and lipid droplets. Red autofluorescence of chloroplast is shown in *red*, double-stranded DNA in *blue* (stained with DAPI), and neutral lipids in *green* (stained with Nile Red) (Courtesy of Philippe Laissue and Narin Chansawang (University of Essex))

for phytoplankton that have minimum linear dimensions of about 5–10  $\mu\text{m}$ . This approach is limited since many phytoplankton are difficult to differentiate from heterotrophic protozoa. In addition, many phytoplankton are considerably smaller than 5  $\mu\text{m}$  in size and lack distinguishing morphological features. This problem can be overcome using the fact that chlorophyll *a* – a pigment found in all phytoplankton – emits red light when illuminated. This chlorophyll autofluorescence allows phytoplankton to be differentiated from heterotrophic organisms. Epifluorescence microscopy allows cells that contain chlorophyll *a* to be visualized (Fig. 1), which is useful for enumerating all but the very smallest phytoplankton.

Analytical flow cytometry (AFC) is now in routine use for enumerating phytoplankton. Like epifluorescence microscopy, AFC uses the red autofluorescence of chlorophyll *a* to distinguish phytoplankton from protozoa and bacteria. Individual cells are entrained in a narrow stream of fluid so that they can be passed one at a time in front of a laser. The light that is scattered by the cells provides an indication of their size and the red autofluorescence an indication of their pigment content. Some flow cytometers have the capability to obtain images of individual cells. High-throughput counting using flow cytometry has been routine for picoplankton since the 1980s, and the upper limit to the size range of phytoplankton cells that can be readily measured by AFC continues to increase.

The remainder of this section introduces some of the most important taxa of marine phytoplankton (Table 3), starting with the smallest photosynthetic organisms, the picoplankton, and progressing upward through the size classes to the nano- and microplankton.

**Table 3** Characteristics of marine phytoplankton taxa based on information summarized by Jeffrey et al. (2011)

Kingdom	Division		Typical size (µm)	Cell covering	Flagella	
Eubacteria	Cyanobacteria	<i>Prochlorococcus</i>	<1	Organic	Absent	
		<i>Synechococcus</i>	<3			
		Unicellular diazotrophs	3–10			
		<i>Trichodesmium</i>	4–40			
Eukaryota	Alveolata (alveolates)	Class Dinophyceae (dinoflagellates)	5–2,000	Naked or cellulose plates	One transverse and one anterior flagella	
		Division Chlorophyta (green algae)	Class Chlorophyceae	10–40	Naked or cellulose	0, 2, 4, or 8 smooth flagella
	Class Prasinophyceae		1–40	Naked or organic scales	1–8 flagella	
	Division Cryptophyta (cryptomonads)	Class Cryptophyceae	6–20	Naked	Two equal flagella	
		Division Haptophyta	Class Haptophyceae	5–20	Organic or CaCO <sub>3</sub> scales	Two flagella and a haptonema
	Stramenopiles (heterokonts)	Class Bacillariophyceae (diatoms)		2–200	Silica frustule	Two unequal flagella (male gametes only)
			Class Chrysophyceae (chrysophytes)	8–15	Naked or scaled	Two unequal flagella
		Class Dictyochophyceae (silicoflagellates)	3–5	25–100 µm silica skeleton	1 or 2	
		Class Pelagophyceae (pelagophytes)	1.5–5	Naked or organic wall	One hairy forward and one smooth trailing	
		Class Raphidophyceae (raphidophytes)	50–100	Naked	One hairy forward and one smooth trailing	

## Picophytoplankton

*Prochlorococcus*, typically found in tropical and subtropical waters, are the most numerous phytoplankton in the ocean (and are potentially the most abundant photosynthetic cells on the planet!). They are often found at concentrations of more than  $10^8$  cells per liter of seawater. Despite this abundance, *Prochlorococcus* was overlooked until the mid-1980s; because of its small size (typically 0.5–0.8 µm in diameter), it could not be differentiated from



**Table 4** Contributions of different size classes and/or taxonomic groups to global phytoplankton biomass and net primary production (NPP)

	Dominant taxa	Contribution to global biomass <sup>a, b</sup> (Pg C)	Contribution to global NPP <sup>c, d</sup> (Pg C/year)
Picoplankton	<i>Prochlorococcus</i>	0.21	
	<i>Synechococcus</i>	0.10	11
	Picoeukaryotes (Prasinophytes, Pelagophytes, Prymnesiophytes, Chrysophytes)	0.44	
Nanoplankton	Prymnesiophytes, Pelagophytes, Cryptomonads	–	20
Microplankton	Diatoms	0.51	15
Diazotrophs	<i>Trichodesmium</i> , unicellular diazotrophs	0.09	0.4–1

<sup>a</sup>Contribution of different groups to mean phytoplankton biomass for the ocean as a whole is from information in the MAREDAT global synthesis (Buitenhuis et al. 2012)

<sup>b</sup>Contributions of diazotrophs to biomass are the arithmetic mean given by Luo et al. (2012)

<sup>c</sup>Contribution of different size fractions to global oceanic annual primary production is from Uitz et al. (2010)

<sup>d</sup>Contribution of diazotrophs to primary production is based on the arithmetic mean N<sub>2</sub> fixation rate of 140 Tg N/year given by Luo et al. (2012) using a C:N ratio of 6 gC/gN to convert N<sub>2</sub> fixation to C fixation

heterotrophic bacteria. In fact, *Prochlorococcus* was only discovered after highly sensitive flow cytometers were optimized for the detection of the faint, red autofluorescence from the divinyl chlorophyll *a* found in these cells. However, even with the most sensitive flow cytometers, a proportion of the *Prochlorococcus* population may still go undetected in high-light regions near the sea surface because these cells contain so little pigment that they hardly fluoresce at all.

Cyanobacteria of the genus *Synechococcus* are slightly larger (typically 0.8–1.5 μm in equivalent spherical diameter) than *Prochlorococcus*. *Synechococcus* inhabits a much wider geographical range than *Prochlorococcus*, including Arctic and coastal waters. There are pronounced gradients in absolute and relative abundances of *Prochlorococcus* and *Synechococcus* between nutrient-poor and nutrient-rich environments, with *Prochlorococcus* being most abundant in the most nutrient impoverished ocean waters.

Also present in the picophytoplankton are very small eukaryotic cells that possess a single chloroplast and are similar in size or slightly larger than *Synechococcus*. The photosynthetic picoeukaryotes are differentiated from other eukaryotic algae by their size rather than their phylogeny and include representatives from at least four algal classes: prasinophytes, pelagophytes, prymnesiophytes, and chrysophytes. Current estimates suggest that for the ocean as a whole, picoeukaryotes account for at least as large a proportion of the biomass and overall productivity of picophytoplankton as *Prochlorococcus* or *Synechococcus* (Table 4).

## Nanophytoplankton

The distinction between pico- and nanophytoplankton is in fact arbitrary since there is a continuum of cells with sizes from about 0.8  $\mu\text{m}$  in diameter to  $>5 \mu\text{m}$  in length. Flagellated nanophytoplankton make a significant contribution to primary production. These small cells are often called “microflagellates” (small flagellates) even though most are  $<10 \mu\text{m}$  in length and so should rightly be called “nanoflagellates.” Among the microflagellates, cryptomonads have been much less extensively investigated than other groups, although they can make an important contribution to phytoplankton biomass in coastal waters. Cryptomonads employ phycobilins as major light-harvesting pigments instead of chlorophyll *a/b* or chlorophyll *a/c* light-harvesting complexes found in other photosynthetic eukaryotes. Some cryptomonads are heterotrophic, and many are mixotrophic, supplementing photosynthesis by ingesting bacteria or absorbing dissolved organic matter. The haptophytes (or prymnesiophytes) are distinguished from other flagellates by possessing a haptonema, a coiled flagellum-like structure that is located between paired straight flagella. The haptonema may be used in prey capture or to escape from predators. The coccolithophorids are a subset of haptophytes that are covered in calcium carbonate plates called coccoliths. Coccolithophorids are major contributors to carbonate deposition in deep-sea sediments and, as such, affect both the pH and alkalinity of the ocean. One of the most well-studied species of coccolithophorids is *Emiliana huxleyi*, which sometimes blooms to such high abundances that it imparts a chalky white color to the sea. The water masses that contain these blooms can be seen from space!

Some unicellular cyanobacteria, also in the nanophytoplankton size class, are capable of using nitrogen gas ( $\text{N}_2$ ) as a nitrogen source. Such organisms are termed diazotrophs: from the prefix “diazo” which refers to two N atoms bonded together and the suffix “troph” which means “nourishment.” Most bacteria and all eukaryotes are unable to use  $\text{N}_2$  because they lack the enzyme nitrogenase, which is required to break the very strong N-to-N triple bond. Diazotrophs play a key role in ensuring the continued fertility of the sea by “fixing” nitrogen into compounds that can be used by other organisms. Their abundance and taxonomic composition is most often assessed from the number of copies of nitrogenase genes (*nif* genes). The potential importance of these unicellular photosynthetic diazotrophs to the N budget of the ocean has only been recognized since the beginning of this century.

## Microphytoplankton

Diatoms (class Bacillariophyceae) make the largest single contribution to global oceanic net primary production, accounting for about 30 % of the total (Table 4). Diatoms are characterized by being enclosed within a silica cell wall called a frustule. The frustule is composed of two interlocking valves, much like a Petri dish. The smallest diatoms are about 2–4  $\mu\text{m}$  in diameter and hence are a component of the nanophytoplankton (e.g., *Minidiscus trioculatus*), while the largest are

about 2 mm (e.g., *Ethmodiscus rex*). However, most diatom cells have diameters between about 10 and 100  $\mu\text{m}$ , but their effective size is often increased by spines or by forming colonies consisting of chains of cells. Although their silica frustule has contributed to their evolutionary and ecological success, it can also be their Achilles heel. Silicate, which is essential for building the frustule, can become depleted before other nutrients, often bringing diatom blooms to an abrupt end, while other phytoplankton that do not require silicate can continue to grow. Diatom blooms are often followed by rapid export of organic matter from the illuminated surface waters to the deep sea, accounting for as much of 50 % of the organic matter that sinks to the deep sea.

Dinoflagellates (division Dinophyta) are unicellular organisms that have been classified as algae by botanists and protozoa by zoologists. About half of all dinoflagellate species are heterotrophic, with the remainder being photosynthetic or mixotrophic. Dinoflagellates make a much smaller contribution to marine primary production than the diatoms, but nonetheless play important ecological roles with significant economic impacts. They are motile and as such thrive under calm conditions in stably stratified water columns. Some photosynthetic dinoflagellates obtained their chloroplasts from the secondary endosymbiosis of red or green algae, whereas others obtained chloroplasts from tertiary endosymbiosis of either a cryptomonad or haptophyte. Dinoflagellates often grow very slowly in nutrient-poor waters. Despite having low growth rates, dinoflagellates can form blooms by employing effective defenses against grazers. The scales of armored dinoflagellates are often shaped into spines or wings that provide a mechanical defense. Some species are bioluminescent, emitting flashes of light when disturbed. These flashes act as deterrents by making their zooplankton predators more conspicuous to fish. Others produce toxins that affect mammals, birds, or fishes and are dangerous to man when accumulated in seafood, such as oysters.

The filamentous cyanobacterium *Trichodesmium* is the most prominent diazotroph in the sea. It is a major contributor to the input of fixed nitrogen to the tropical ocean, particularly in the North Atlantic and Indian Oceans. *Trichodesmium* blooms are also observed in waters around Australia and in the Red Sea. *Trichodesmium* is present as individual filaments (trichomes) and also as large colonies of filaments. Surface blooms, referred to as “sea sawdust,” rise to the surface under calm conditions. When present at high abundance over large areas, these surface aggregations can be detected using satellite ocean color sensors. Also contributing to  $\text{N}_2$  fixation in the sea are diazotrophic heterocyst cyanobacteria, principally *Richelia* and *Calothrix*: these are found in symbiotic association with some large diatoms.

*Phaeocystis* is a haptophyte genus that is widely distributed throughout the ocean. *Phaeocystis* has received particular attention because it can form large colonies consisting of hundreds or thousands of cells. The gel-like matrix of the colonies is thought to store energy (polysaccharide) and nutrients (phosphate, iron), whereas the “skin” of colonies is thought to prevent infection with pathogens and present a mechanical barrier to zooplankton grazing. *Phaeocystis* blooms have been reported in Arctic and Antarctic open ocean waters as well as nutrient-enriched coastal waters.

## Plankton Functional Groups and Trait-Based Phytoplankton Ecology

The taxonomic position of an organism does not always convey unambiguous information about its ecological roles, its biogeochemical functions, or its ecophysiological traits. Consequently, ecologists often group organisms according to the roles they play, the functions they perform, and/or the traits that underpin resource acquisition and population dynamics. By analogy to plant functional types (PFTs), which are employed in models of terrestrial primary production, oceanographers have introduced the concept of plankton functional groups (PFGs). Terrestrial PFTs have proven useful in understanding terrestrial primary production because the distribution of vegetation types can be mapped using remote sensing, and the net primary production (NPP) of these vegetation types can be estimated from PFT-specific algorithms. A challenge for oceanographers has been to define plankton functional groups that are as effective in accounting for plankton NPP in the ocean as PFTs are for accounting for NPP on land. In addition to accounting for NPP, oceanographers need to account for other important biogeochemical transformations performed by phytoplankton including nitrogen fixation ( $N_2$  fixation) and biomineralization (calcification, silicification). These biogeochemical processes influence the large-scale cycling of carbon, nitrogen, phosphorus, and iron in the oceans, linking the cycling of these elements to the rest of the ecosystem and climate.

### Ecological Roles and Functions

The ecological roles and functions of organisms are related to three broad themes in ecology, namely, trophic dynamics, population dynamics, and biogeochemical cycles. These are discussed in turn. Trophic dynamics refers to the flow of energy through an ecosystem, with organic carbon often used as a surrogate for energy. The trophic functions correspond to three main nutritional types: autotrophs, phagotrophs, and osmotrophs. Primary producers are autotrophs (self-feeders) and include photosynthetic and chemosynthetic organisms. The consumers are phagotrophs (particle eaters), which include protozoa, zooplankton, and nekton in the ocean. The decomposers are osmotrophs (osmotic feeders), which absorb rather than ingest organic matter, and consist of heterotrophic bacteria and fungi. Many phytoplankton can absorb and/or ingest organic matter in addition to being able to photosynthesize and as such are mixotrophs that perform more than one trophic function. Production of detritus and dissolved organic matter that provides the food for decomposers is another trophic function, one that is performed by zooplankton and nekton during “sloppy” feeding, defecation, and excretion and by viruses when their host’s cells are killed and lyse.

Population dynamics assesses changes in the sizes of populations that arise from reproduction, death, immigration, and emigration. These population processes influence the dynamics of populations through both bottom-up and top-down

interactions. Bottom-up interactions include competition for limiting resources, sequestration of non-limiting resources, and modification of the physical environment (light penetration, pH, and redox state). Top-down interactions include predation, parasitism (fungal and parasitoid), and disease (viral and bacterial). Organisms can be classified by their interactions with other organisms as competitors, predators, prey, or symbionts.

The biogeochemical functions of organisms arise from their interactions with the physical/chemical environment. Among the most important biogeochemical functions performed by marine phytoplankton are photosynthetic CO<sub>2</sub> fixation and O<sub>2</sub> evolution; N<sub>2</sub> fixation; production of climatically active gases such as dimethyl sulfide and terpenes; and precipitation of minerals, such as CaCO<sub>3</sub> (calcium carbonate) and SiO<sub>2</sub> (biogenic silica).

Biogeochemical functions overlap with trophic functions and therefore the processes that shape population dynamics. As a result, care must be taken when classifying organisms by their roles or functions. Take, for example, coccolithophorids. These organisms are trophically primary producers. The processes that determine their population dynamics include competition with other primary producers for nutrients and the mortality due to predation and disease. The biogeochemical functions of coccolithophorids include production of oxygen and organic carbon, precipitation of CaCO<sub>3</sub>, and production of volatile organic sulfur compounds. These functions determine the roles of coccolithophorids in nutrient cycling and in the consumption and production of greenhouse gases including CO<sub>2</sub>. Other phytoplankton share the trophic role of being a primary producer with coccolithophorids, but compete with coccolithophorids for resources, such as nutrients and light, while having quite different biogeochemical functions. For example, *Trichodesmium* is a phytoplankton that fixes N<sub>2</sub> but does not precipitate CaCO<sub>3</sub>. Diatoms are phytoplankton that precipitate SiO<sub>2</sub> but do not calcify. To complicate matters, photosynthetic coccolithophorids share the biogeochemical function of precipitating CaCO<sub>3</sub> with phagotrophic foraminifera, which are consumers rather than primary producers. To complicate matters even further, many foraminifera house symbiotic photosynthetic dinoflagellates, and this symbiosis contributes to CO<sub>2</sub> fixation.

## Emergent Biogeochemical Properties

Some important biogeochemical functions result from interactions among organisms that belong to different taxonomic or functional groups. For example, export production plays an important role in Earth's climate by transferring CO<sub>2</sub> to the interior of the ocean, thereby reducing the amount of CO<sub>2</sub> in the atmosphere. Export production is the sum of losses of organic matter from the surface layer of the ocean in sinking particles, advection (including detrainment), and mixing of dissolved organic matter from source to sink regions. The first step in export production is net primary production by phytoplankton. Zooplankton then feed on phytoplankton and produce fecal pellets that sink rapidly out of the euphotic zone. Organic aggregates,

sometimes called marine snow, also contribute to export production. These aggregates form when mucus nets produced by some zooplankton become clogged and are discarded and/or when phytoplankton, bacteria, and detritus stick together to form clumps. Fecal pellets and marine snow sink more rapidly when they include dense mineral phases (e.g., calcium carbonate and biogenic silica) that are produced by coccolithophorids and diatoms. Vertical migration of zooplankton and nekton, which involves feeding near the sea surface at night and moving to deeper waters during the day to avoid predators, can also contribute to export production. For these reasons, export production is an emergent property that arises at the ecosystem level; the amount of export production cannot be inferred simply from the properties of the components of the ecosystem, but relies on how these components interact.

### Trait-Based Phytoplankton Ecology

Traits are used to qualify and quantify the abilities of organisms to perform particular ecological or biogeochemical functions. Organisms can be categorized into PFGs based on the different combinations of traits that they possess. For example, photosynthetic diazotrophs possess the abilities to fix  $N_2$  and  $CO_2$ , whereas heterotrophic diazotrophs possess the former but not the latter. Even if an organism has the potential to express a particular trait, it may not do so in all circumstances. For example, *Trichodesmium* fixes  $N_2$  when there is insufficient combined N ( $NH_4$  or  $NO_3^-$ ) in the environment.

Some traits can be assigned numerical values. Morphological traits of phytoplankton include surface area, volume, and shape. Other traits characterize how the performance of phytoplankton cells depends on abiotic environmental conditions such as temperature, irradiance, and nutrients. Performance can be assessed at the population level or the individual level and includes, for example, the population growth rate or the cell-specific photosynthesis rate. Cell size can be considered to be a “master trait” because many other traits are highly correlated with size. These include, for example, maximum growth rate, affinity for nutrient uptake, sinking rate, swimming rate, and indices of the susceptibility to being preyed upon by different types of zooplankton.

Allometry is the study of the relationship of morphological or physiological variables to the size of an organism. Two indices of size that are commonly used are the volume and the mass of an organism, which for phytoplankton are related through a power law:  $M = a \cdot (\text{Vol})^b$ , where  $M$  is the cell's mass,  $\text{Vol}$  is the cell's volume, and  $a$ ,  $b$  are empirically determined coefficients. If the cell's mass was directly proportional to its volume, the exponent of this relation would equal 1.0 (i.e.,  $b = 1.0$ ). However, the exponent for the observed dependence of cellular organic carbon content on volume is only about 0.8–0.9. This means that the cellular density of carbon declines with increasing cell size: for diatoms, this decrease is from about  $0.2 \text{ pg C } \mu\text{m}^{-3}$  for the smallest cells to  $0.05 \text{ pg C } \mu\text{m}^{-3}$  for the largest; this decrease can be attributed to the increasing percent of the

volume of large diatom cells that is occupied by the watery vacuole. For phytoplankton that lack vacuoles, such as dinoflagellates, the reduction in the cellular density of carbon with increased volume is less pronounced.

Physiological traits including the cell-specific maximum nutrient uptake rate and the affinity for nutrient assimilation also change markedly with increases of cell volume. Much of the size dependence of these traits can be explained from physical principles. For example, the rate at which nutrients diffuse to a cell should be proportional to its radius. This leads to the expectation that the half saturation constant for nutrient uptake should increase with  $(Vol)^{0.33}$ . On the other hand, the maximum rate at which nutrients can enter the cell is expected to be proportional to the number of transporters on the cell surface, and this leads to the expectation that the maximum rate of nutrient uptake will increase with  $(Vol)^{0.67}$ . Such constraints on physiology imposed by geometry and physics are most evident when a very wide range of cell size is considered. However, there is considerable variability that cannot be accounted for by cell size, and when working within a restricted range of sizes, physiological sources of variability become increasingly important.

---

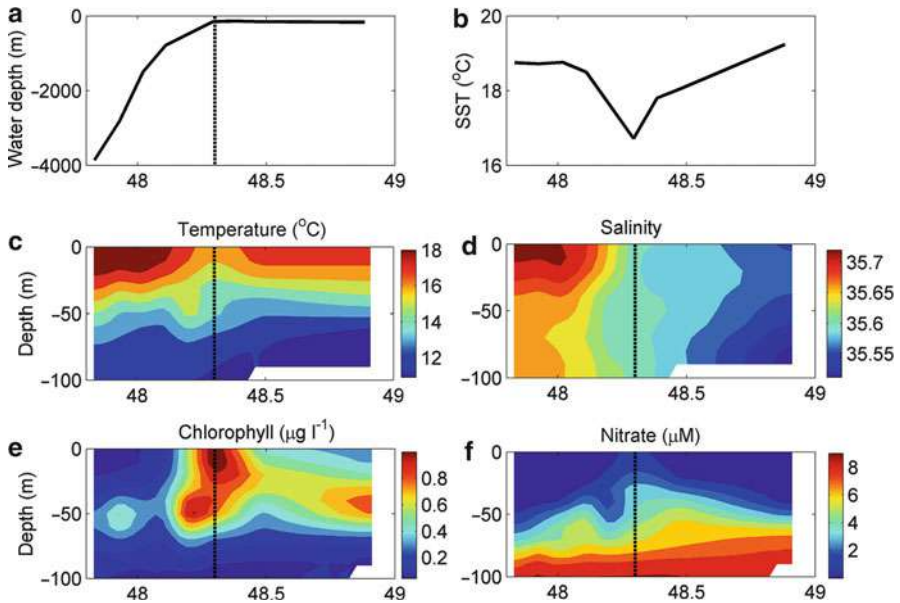
## Characteristics of the Pelagic Environment

The pelagic zone is the region between the seabed and sea surface; it includes the waters that lie over the continental shelf, called the neritic zone, and the waters of the open ocean, called the oceanic zone. The neritic zone covers about 9 % of the ocean surface and is typically less than 200 m deep. The oceanic zone accounts for the remaining 91 % and has an average depth of about 4 km.

The pelagic is a fluid environment that is shaped by ocean currents. Although it lacks geographic barriers like mountain ranges or rivers, it can nonetheless be subdivided into water masses that are separated by sharp horizontal gradients in temperature and salinity called fronts. For example, fronts located near the edge of the continental shelf separate neritic waters of shelf seas from oceanic waters (Fig. 2). These water masses differ not only in temperature and salinity but also in nutrient availability and plankton communities.

## Temperature, Salinity, and Density

The temperature of ocean surface waters varies from a minimum of  $-2\text{ }^{\circ}\text{C}$  at high latitudes to maximum values of  $\sim 35\text{ }^{\circ}\text{C}$  in some equatorial regions, as a consequence of the strong latitudinal gradient in solar heating between the tropics and the poles. The lower temperatures of the Arctic and Antarctic reduce the maximum growth rates of phytoplankton to around 20 % of rates that can be achieved in the tropics. The temperature of the deep ocean is fairly uniform at around  $2\text{ }^{\circ}\text{C}$ . This is because the deep waters form at high latitudes when cold, saline water sinks and then spreads across the ocean interior.

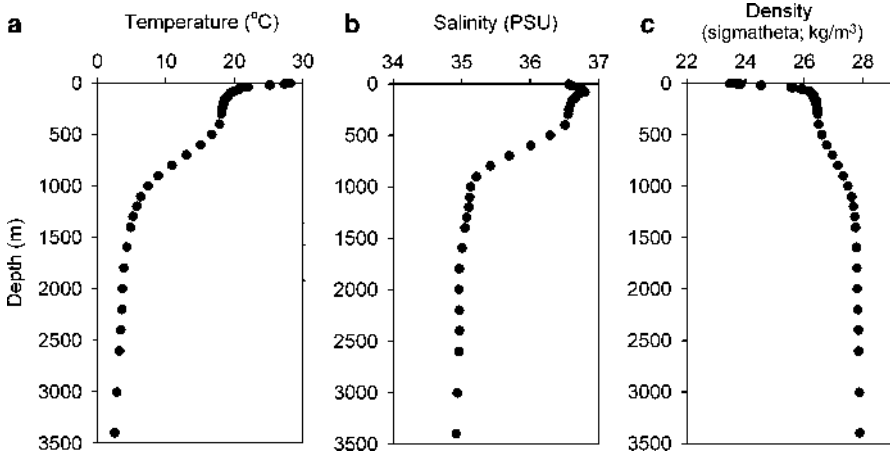


**Fig. 2** Shelf front transect. Data are collected through the water column across the transition between shallow neritic shelf waters and deep oceanic waters as indicated by the water depth (a). A marked decrease in the temperature of the sea surface (SST) is observed at the transition between the shelf and ocean (b), termed the shelf break. Vertical cross sections of temperature (c) and salinity (d) also display gradients at the shelf break, indicating the different water masses on and off the shelf. The abundance of phytoplankton, as indicated from chlorophyll *a* (e) and the availability of nutrients (f), also varies across the frontal transition

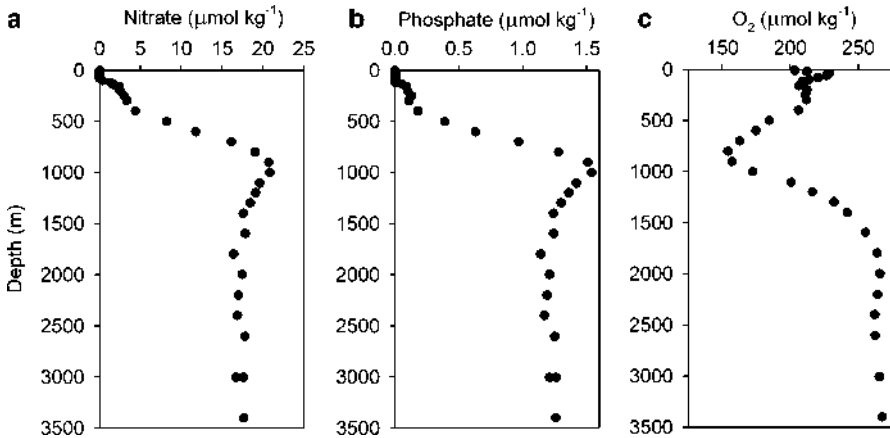
The average salt concentration of ocean waters is about 3.5 % (35 g of salt per kg of seawater). Alterations occur when rainfall and river runoff decrease salinity by dilution or when salinity increases due to evaporation or to ice formation. Salinity is reported in practical salinity units (PSU), where 1 PSU is approximately equal to 1 g of salt per kg of seawater. Salinity is sufficiently uniform in the pelagic that it has little direct influence on the physiology of phytoplankton over most of the open ocean. However, salinity affects the density of seawater, which in turn influences water motion and thus the supply of nutrients. Although salinity over most of the sea surface ranges from about 30 to 38.5 PSU, marked deviations occur near coasts where freshwater inputs are large and physical exchange with the open sea is limited. Salinity varies from <0.5 to 20 PSU in estuaries and can be low in semi-enclosed waters such as the Baltic and Black Seas. At the other extreme, salinity can reach very high values (>100 PSU) in brine-filled pockets of sea ice.

Oceanographic data is often presented as vertical profiles. These are plots of the property of interest on the horizontal axis versus depth on the vertical axis. For example, vertical profiles of temperature, salinity, and density for a location in the subtropical North Atlantic Ocean are illustrated in Fig. 3 and corresponding profiles of nutrients and dissolved oxygen in Fig. 4. Among the most conspicuous features in





**Fig. 3** Vertical profiles of temperature, salinity, and density, expressed as ‘sigma-theta’ = density - 1000 kg/m<sup>3</sup> at the Bermuda Atlantic Time-Series Station. The permanent thermocline is evident at depths below 500 m. Above this depth, temperature and density vary seasonally due to solar heating and evaporation. BATS cruise 10106; 15 July 1997. Data provided by the U.S. National Science Foundation funded Bermuda Atlantic Time-series Study Program: <http://www.bios.edu/research/hydrodata.html>



**Fig. 4** Vertical profiles of nitrate, phosphate, and dissolved oxygen at the Bermuda Atlantic Time-Series Station. BATS cruises 10106 and 10107. 15 July 1997 and 1 August 1997. Data provided by the U.S. National Science Foundation funded Bermuda Atlantic Time-series Study Program: <http://www.bios.edu/research/hydrodata.html>

these profiles are the decreases of temperature and salinity and increase of density and inorganic nutrients at depths between 500 and 1,000 m. These features persist throughout the year, and the depth zone from about 500–1,000 m is called the “permanent thermocline.” Also located in this depth zone is an oxygen minimum layer.

## Vertical Light Attenuation and Ocean Color

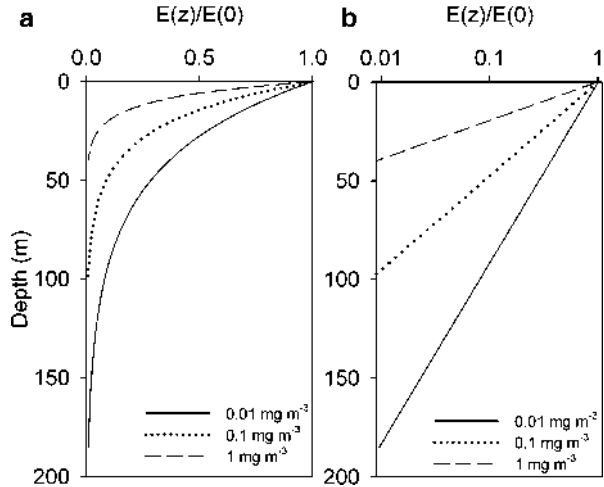
The region of the water column in which there is enough light to support net photosynthesis is referred to as the euphotic zone (from Greek for “well lit”). As a rule of thumb, the lower limit of the euphotic zone corresponds to the depth at which irradiance equals 1 % of the surface value. The maximum depth of the euphotic zone is about 150 m in the clearest open ocean regions. In coastal waters, where dissolved and particulate matter absorbs and scatters light, the euphotic zone can be as shallow as 5–10 m, whereas in very turbid conditions, it may be as shallow as 1 m or less. Since the mean depth of the ocean is about 4,000 m, most of the ocean volume is too dark to support phytoplankton growth. With the exception of a small contribution from chemosynthesis, life in the deep sea depends on the supply of organic matter from the euphotic zone. Most of this organic matter is in the form of fecal pellets and aggregates of living and detrital organic matter, which can sink at rates from tens to hundreds of meters per day.

The rate of decline of irradiance with increasing depth is approximately exponential (Fig. 5) and is given by the equation  $E(z) = E(0) \exp(-K_d z)$ , where  $z$  is the depth in meters,  $E(z)$  is the irradiance at a depth of  $z$  meters,  $E(0)$  is the irradiance just below the sea surface, and  $K_d$  is the vertical light attenuation coefficient. Although  $K_d$  varies with wavelength, it is convenient to consider that the spectrally integrated light is approximated by this equation. The attenuation coefficient is not a constant, but depends on the material that is dissolved and suspended in the sea. This includes phytoplankton, bacteria, detritus, and colored dissolved organic matter.

The color of the sea arises from the light that is reflected by water molecules and absorbed or scattered by other seawater constituents. This reflected light originates from different depths within the upper 20 % of the euphotic zone that corresponds to depths of 1–20 m, depending on the water clarity. Some wavelengths of light are attenuated more strongly than others, and as a consequence, the color of underwater light changes. Blue light is absorbed least by pure water and thus penetrates to the greatest depth in clear ocean waters. Consequently, a high proportion of blue light is reflected by clear ocean waters, giving these waters their blue color.

Differences in ocean color arise from differences in the attenuation of light of different wavelengths as a result of the optical properties of water itself and of the substances that are suspended and dissolved in water. The most important properties that contribute to variations in ocean color are the abundance and species composition of phytoplankton and the concentration of colored dissolved organic matter (CDOM). As phytoplankton abundance increases, the absorption of blue light by chlorophylls and carotenoids increases, thereby causing the wavelength that penetrates deepest to shift towards the green. Where there is a high concentration of CDOM, the wavelength that penetrates deepest is shifted even further towards the red end of the spectrum. These differences in light attenuation affect both the color of the ocean that is perceived from above and the spectral distribution of light that is available for photosynthesis at depth. The spectrum becomes

**Fig. 5** Vertical attenuation of photosynthetically active radiation (PAR) plotted on linear (a) and logarithmic (b) irradiance scales. Profiles for waters with different chlorophyll *a* concentrations were calculated from Morel's (1988) bio-optical model  $K_d = 0.121 [\text{Chl}]^{0.428}$ , where  $K_d$  is the mean attenuation coefficient for PAR and  $[\text{Chl}]$  is the chlorophyll *a* concentration within the euphotic zone



increasingly dominated by blue light in clear ocean waters and by green light in coastal waters with higher amounts of phytoplankton. These changes in ocean color are the basis for remote sensing of phytoplankton abundance using satellite-borne sensors.

## Vertical Stratification and Mixing

The density of the surface layer of the ocean declines when it is heated by solar radiation, allowing it to float on the denser water below. Surface heating during the summer in temperate and polar regions, or throughout the year in the tropics and subtropics, leads to one of the most important physical features of the pelagic, the increase of water density with increasing depth. Over most of the ocean, warm buoyant surface water floats on cold denser waters. The typical vertical profile in tropical and subtropical seas consists of a mixed layer, which has a uniform density, below which density increases rapidly with increasing depth, in a region referred to as the pycnocline (density gradient) or thermocline (temperature gradient). The mixed layer varies in depth from about 25–250 m depending on location and season.

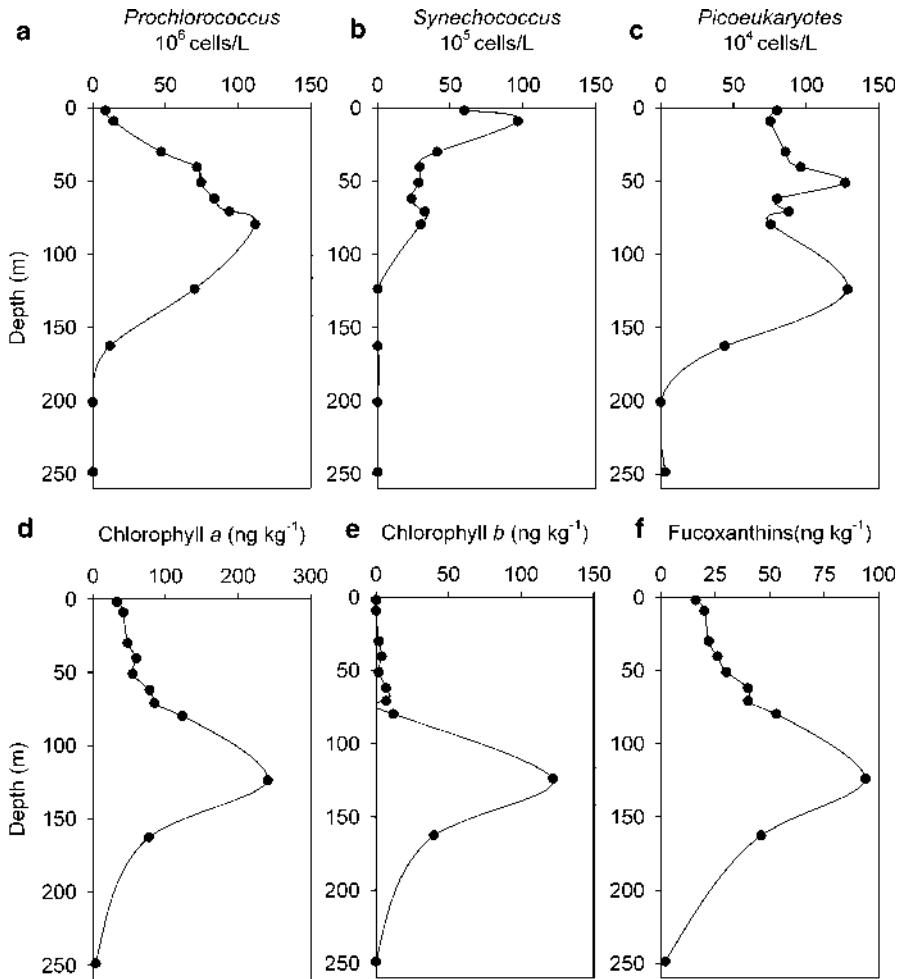
The upper 1,000 m of the ocean can be divided into a zone that is permanently stratified, above which there is a zone that stratifies seasonally. The depth of the permanent pycnocline/thermocline varies geographically, with ridges, domes, and troughs that are associated with upper ocean currents such as the Gulf Stream in the North Atlantic or the Kuroshio Current in the North Pacific. At any given location in the ocean, seasonal variations occur in both the density of water in the surface mixed layer and the depth of the seasonal pycnocline. As surface waters cool in

autumn and winter, density increases and the mixed layer deepens, eroding the top of the seasonal pycnocline. Conversely, surface waters become lighter and the mixed layer shoals when winter gives way to spring and summer. Superimposed on the seasonal changes of mixed layer depths are diel changes driven by the cycle of heating during the day and cooling at night. Deepening also occurs when high winds increase vertical mixing by introducing turbulence at the sea surface.

## Vertical Nutrient Distributions

When organic matter that is exported from surface waters is respired at depth, oxygen is consumed, and nutrients and carbon dioxide are released. A consequence of this vertical separation of net primary production from net decomposition is that nutrients and dissolved inorganic carbon are depleted near the sea surface and enriched at depth (Fig. 4). Dissolved oxygen shows the opposite vertical pattern and is often supersaturated near the surface due to net photosynthesis and undersaturated at depth due to net decomposition.

Vertical density stratification in the pycnocline is a barrier to mixing of water between the deep and surface ocean. Thus, the pycnocline restricts transport of the deeper nutrient-rich subsurface waters back to the surface and of oxygen-rich surface waters to depth. Where there is a persistent, well-developed seasonal pycnocline within the euphotic zone, the phytoplankton community in the mixed layer typically consumes all the available nutrients, resulting in subsequent production being nutrient limited. This is the case in the subtropical gyres, which cover ~60 % of the ocean surface. Here the pycnocline separates a low-nutrient/high-light surface mixed layer from deeper nutrient-rich/low-light layers. Under such conditions, sharp vertical gradients of irradiance and nutrient concentration are evident, and phytoplankton abundance and species composition may also change dramatically through the pycnocline (Fig. 6). A subsurface maximum of chlorophyll *a* concentration (the deep chlorophyll maximum, DCM), but not necessarily phytoplankton biomass, is typically found at the top of the nitrate gradient within the pycnocline. Below the peak of the DCM, phytoplankton growth is limited by light, whereas above the peak, phytoplankton growth is limited by nutrient availability. Taxonomically distinct low-light-adapted (“shade”) phytoplankton species are found within and below the DCM, whereas species found within the mixed layer must be able to cope with low nutrients and high light. The vertical structure of the phytoplankton community is best established for the smallest cells (the picoplankton), which are amenable to automated analysis using analytical flow cytometry. For example, distinct sun and shade “strains” have been identified within the genus *Prochlorococcus*. The vertical structure is destroyed during deep-mixing events, but becomes reestablished when stratification returns.

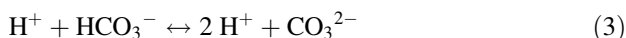


**Fig. 6** Vertical distributions of picophytoplankton and photosynthetic pigments at the Bermuda Atlantic Times-Series Station on 18 August 1992. The bottom of the euphotic zone is at a depth of about 150 m. (a) *Prochlorococcus* has a subsurface maximum of abundance; (b) *Synechococcus* has a surface maximum of abundance, whereas (c) picoeukaryotes are more uniformly distributed throughout the euphotic zone. (d) The subsurface chlorophyll *a* maximum is also a maximum for (e) chlorophyll *b* and (f) fucoxanthins (sum of fucoxanthin, 19'-butanoyloxyfucoxanthin and 19'-hexanoyloxyfucoxanthin, which are found in haptophytes, pelagophytes, and chrysophytes). Data provided by the U.S. National Science Foundation funded Bermuda Atlantic Time-series Study Program: <http://www.bios.edu/research/hydrodata.html>

## Dissolved Inorganic Carbon

The most common gases in the atmosphere ( $\text{N}_2$ ,  $\text{O}_2$ , and Ar) do not undergo chemical reactions with seawater. However,  $\text{CO}_2$  combines with water to form

carbonic acid (Eq. 1), which in turn dissociates (breaks apart) to form carbonate ions (Eq. 2) that in turn dissociate to form bicarbonate ions (Eq. 3).



As a consequence, most of the  $\text{CO}_2$  that dissolves in seawater reacts to form bicarbonate ( $\text{HCO}_3^-$ ) and carbonate ( $\text{CO}_3^{2-}$ ) ions. These different forms of inorganic carbon exist in a dynamic, thermodynamic equilibrium that is dependent on both temperature and salinity. Together these three forms comprise the dissolved inorganic carbon (DIC) pool. Bicarbonate is the most abundant, accounting for >90 % of the DIC in seawater. In contrast, aqueous  $\text{CO}_2$  is present at very low concentrations, <1 % of the DIC.

Approximately 40 % of  $\text{CO}_2$  released into the atmosphere by anthropogenic activity has dissolved in oceans as a consequence of Henry's law and the chemical reactions depicted in Eqs. 2 and 3. As these surface waters sink, DIC is transported into the interior of the ocean. Unfortunately, the rate at which the ocean is absorbing  $\text{CO}_2$  from the atmosphere is predicted to be slowing down. This means that a higher proportion of the  $\text{CO}_2$  that is currently being produced by human activities is accumulating in the atmosphere than has been the case in the past.

## Ocean Acidification

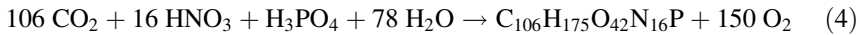
The ocean becomes slightly more acidic when  $\text{CO}_2$  dissolves in seawater because protons ( $\text{H}^+$ ) are released during the reactions that form bicarbonate and carbonate. There has been about a 30 % increase in the mean concentration of  $\text{H}^+$  in the ocean's surface waters during the past 250 years, and the rate of increase is accelerating as more fossil fuel is burned. The mean pH of the surface waters has decreased by 0.1 pH units over the past 250 years, from about pH = 8.2 to pH = 8.1, and is projected to drop to as low as pH = 7.9 by the end of the century. Although these changes may seem small, they will be accompanied by marked decreases in the concentration of carbonate ions and of the saturation state of carbonate minerals with potentially dire consequences for marine organisms that produce calcium carbonate shells (see section "[Anthropogenic Impacts on Marine Phytoplankton](#)").

---

## Primary Production

The leaves of terrestrial vascular plants are essentially "sugar factories" that produce sugars and starch during photosynthesis, with subsequent translocation from mature leaves to the roots and actively growing tissues. For these plants,

primary production is virtually synonymous with photosynthesis. This is not the case for phytoplankton. Phytoplankton are more like protein factories than sugar factories because photosynthetic  $\text{CO}_2$  fixation is very closely linked to nutrient assimilation and the synthesis of proteins and lipid in actively growing, nutrient-replete phytoplankton. The following chemical equation, which is consistent with the typical biochemical composition (lipid to protein to nucleic acid) of algae, depicts marine primary production:



This equation accounts not only for  $\text{CO}_2$  fixation but also for the assimilation of nitrate and phosphate into organic matter. Somewhat paradoxically, this stoichiometry was obtained by examining the reverse process, namely, the decomposition of organic matter in the deep sea, which leads to covariation in the concentrations of nitrate, phosphate, dissolved inorganic carbon, and dissolved  $\text{O}_2$ .

The ratio of  $\text{O}_2$  produced to  $\text{CO}_2$  fixed is called the photosynthetic quotient and is designated PQ. Equation 4 gives a PQ of  $150/106 = 1.45 \text{ mol O}_2 (\text{mol CO}_2)^{-1}$ , whereas this ratio is  $1.0 \text{ mol O}_2 (\text{mol CO}_2)^{-1}$  for synthesis of sugars. The PQ is used when comparing measurements of primary production based on  $\text{O}_2$  evolution (see section “[The Photosynthesis-Irradiance Response Curve](#)”) with those based on  $\text{CO}_2$  fixation (see section “[Net and Gross Primary Production of Marine Phytoplankton](#)”).

Oceanographers are concerned not only with gross primary production (GPP) and net primary production (NPP) but also with net community production (NCP), which takes into account the respiratory activity of heterotrophic organisms including bacteria, protozoa, and animals. The relationships among these different processes are conveniently summarized in the following equation:

$$\text{NCP} = \text{GPP} - R_A - R_H = \text{NPP} - R_H \quad (5)$$

where NCP is net community production, GPP is gross primary production,  $R_A$  is respiration by autotrophs, NPP is net primary production, and  $R_H$  is respiration by heterotrophs.

NCP is the small proportion of GPP that is not respired by phytoplankton, bacteria, protozoa, or animals. This organic matter either sinks to the deep sea in fecal pellets or detritus or accumulates in surface waters as dissolved organic matter. There is an additional category of production that oceanographers call export production. Export production is dominated by the loss of organic matter from the euphotic zone in sinking particles. When integrated over sufficiently long time and space scales, NCP should equal export production. Export production is important in ocean carbon and nutrient cycles. Specifically, by transferring carbon to deep waters, export production lowers the partial pressure of  $\text{CO}_2$  at the sea surface, thus facilitating the uptake of  $\text{CO}_2$  from the atmosphere by the ocean. Export production is closely linked to the input of nutrients into the euphotic zone, since nutrients are required to support net synthesis of biomass.

## Net and Gross Primary Production of Marine Phytoplankton

Primary production of terrestrial plants is commonly assessed from measurements of the rate of decline of  $\text{CO}_2$  in the air as photosynthesis incorporates  $\text{CO}_2$  into organic matter. This is possible because  $\text{CO}_2$  is present at very low concentrations, and a plant leaf contains a high amount of photosynthetic tissue. In contrast to the low concentration of  $\text{CO}_2$  in air, DIC is present at high concentrations in seawater. It is difficult to measure the small changes in the concentration of DIC that accompany phytoplankton NPP. To overcome this difficulty, oceanographers have devised a number of ways to use radioisotopes and stable isotopes to measure primary production. However, different approaches measure different processes, with some suitable for measuring GPP, others for measuring NPP, and still others for measuring NCP.

The most common method for measuring GPP and/or NPP involves:

- (i) Collecting seawater samples from several depths within the euphotic zone
- (ii) Dispensing these samples into bottles
- (iii) Incubating these bottles at the depths from which the samples were collected
- (iv) Measuring the uptake of  $\text{CO}_2$  or release of  $\text{O}_2$  by phytoplankton

Incubations typically last from dawn to dawn (24 h) to account for photosynthesis during the day and respiration during the day and at night.

Bottle incubations provide measurements of primary production in a given volume of seawater, but it is often more useful to know the primary production under a given area of the sea surface. This requires that the measurements obtained at different depths are added together. The normal practice is to measure primary production at between 6 and 10 depths spaced throughout the euphotic zone (Fig. 7) and then to sum the values within particular depth intervals to obtain the total for the water column.

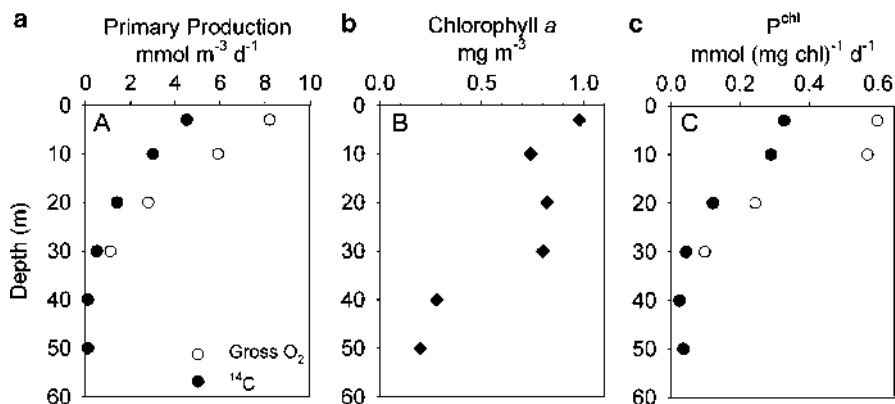
$$\text{Areal NPP} = \sum_{i=1}^N P(i) \cdot \Delta Z(i) \quad (6)$$

In this equation,  $N$  is the number of depth intervals,  $P(i)$  is the mean value of NPP within the  $i^{\text{th}}$  depth interval, and  $\Delta Z(i)$  is the width of that depth interval.

Throughout the first half of the twentieth century, oceanographers relied on the measurement of  $\text{O}_2$  evolution as a proxy for  $\text{CO}_2$  fixation. This is because  $\text{O}_2$  has a relatively low solubility in water, and very precise analytical methods for measuring  $\text{O}_2$  concentration have been available since the late nineteenth century. The principle of the light–dark bottle method is simple. Briefly,  $\text{O}_2$  is produced by photosynthesis in the light bottle, but is consumed by respiration in both the light and dark bottles. NPP is obtained by measuring the increase of  $\text{O}_2$  concentration in the light bottle, and respiration is obtained by measuring the decrease in of  $\text{O}_2$  in a darkened bottle. GPP is then obtained from the sum of the increase of  $\text{O}_2$  in the light and decrease in the dark.

In 1944, the American oceanographer Gordon Riley made the first estimate of global oceanic primary production; this estimate was based on light–dark bottle  $\text{O}_2$  production determinations. Assuming a photosynthetic quotient of 1.45  $\text{CO}_2$  fixed





**Fig. 7** Primary production in the North Atlantic. Shown are vertical profiles of (a) gross  $\text{O}_2$  evolution (open circles) and  $^{14}\text{C}$  assimilation (*filled circles*) during dawn-to-dusk incubations, (b) chlorophyll *a* concentration at dawn, and (c) chlorophyll *a*-specific primary production rates (Data are from Kiddon et al. 1995)

per  $\text{O}_2$  evolved, Riley's calculations give a value for GPP of the ocean as a whole of  $87 \pm 56 \text{ Pg C per year}$  (mean GPP of  $234 \pm 151 \text{ g C m}^{-2} \text{ year}^{-1}$ ). Most of the change that Riley observed was due to respiration in the dark bottle rather than NPP in the light. The number of observations and their geographical range were very limited and the error estimates associated with this calculation very large. In addition, Riley employed incubations that lasted 3 days in order to obtain sufficiently large changes in  $\text{O}_2$  concentrations to be detected reliably. It was clear that a more sensitive method for measuring primary production was needed.

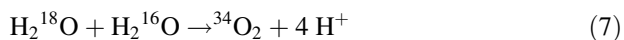
A new method for measuring primary production was introduced to oceanography in the 1950s by the Danish botanist and experimental biologist Einer Steemann-Nielsen. Steemann-Nielsen developed the first protocols for using a radioactive isotope of carbon, carbon-14 ( $^{14}\text{C}$ ), to measure marine primary production. The ease and sensitivity of the  $^{14}\text{C}$  method relative to the more cumbersome and less sensitive light–dark bottle  $\text{O}_2$  method have allowed  $\text{CO}_2$  fixation to be measured routinely many thousands of times. Current estimates of oceanic NPP (Table 1) rely on the accumulated database of point  $^{14}\text{C}$  measurements of primary productivity that have been extrapolated to the global scale using ocean color data (see section “[Remote Sensing of Primary Production](#)”). The current estimate for NPP of the ocean based on extrapolation of the  $^{14}\text{C}$  database is about  $50 \text{ Pg C year}^{-1}$  (Table 5). This may be an underestimate as it only accounts for the particulate carbon production, neglecting the carbon that is fixed into dissolved organic matter, which can be a significant proportion of the total.

The light–dark bottle oxygen exchange method provides an unambiguous measurement of NPP, but may underestimate GPP because  $\text{O}_2$  uptake by phytoplankton is often stimulated by light and therefore will not be accounted for by the consumption that is measured in the darkened bottle. Two approaches used to obtain accurate measurements of GPP rely on the stable oxygen isotope oxygen-18 ( $^{18}\text{O}$ ) and were developed and applied in the 1980s.  $^{18}\text{O}$  accounts for only about 0.2 % of oxygen in

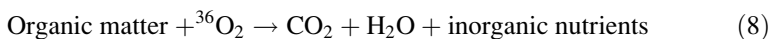
**Table 5** Summary of calculations of global marine phytoplankton primary production from ocean color data using 24 algorithms (Carr et al. (2006)). Total production is reported by region, by chlorophyll level, and by sea surface temperature. Oligotrophic (<0.1 mg chl *a* m<sup>-3</sup>); mesotrophic (0.1–1 mg chl *a* m<sup>-3</sup>); eutrophic (>1 mg chl *a* m<sup>-3</sup>)

	Area %	Mean Pg C year <sup>-1</sup>	Range Pg C year <sup>-1</sup>
<b>Region</b>			
Pacific	45	21	15.5–30.9
Atlantic	23	12.8	9.1–17.9
Indian	17	9.9	6.9–15.1
Southern	13	2.6	1.1–4.9
Arctic	1.2	0.33	0.02–1.2
Mediterranean	0.8	0.45	0.28–0.73
Total		47.1	35–60
<b>Chlorophyll concentration</b>			
Oligotrophic	26–32	9.2	4.6–14.1
Mesotrophic	65–68	34.8	24.2–48.8
Eutrophic	3–5	5.6	2.4–9.9
Total		49.6	
<b>Sea surface temperature</b>			
<0 °C	2–4	0.52	0.17–2.1
0–10 °C	13–17	5.1	2.1–8.4
10–20 °C	20	11.9	7.6–18.9
>20 °C	60	32	19.1–48.7
Total		49.5	

nature, but can be enriched to provide water and O<sub>2</sub> that contain almost 100 % <sup>18</sup>O. Oxygen that contains two <sup>18</sup>O atoms has a molecular weight of 36 (designated <sup>36</sup>O<sub>2</sub>), whereas oxygen that contains one <sup>18</sup>O and one <sup>16</sup>O atom has a molecular weight of 34 (designated <sup>34</sup>O<sub>2</sub>) and oxygen that contains two <sup>16</sup>O atoms has a molecular weight of 32 (designated <sup>32</sup>O<sub>2</sub>). Mass spectrometers are used to detect the amounts of O<sub>2</sub> with different masses. In the first approach, water that is labeled with <sup>18</sup>O (i.e., H<sub>2</sub><sup>18</sup>O) is added to a sample, and the production of <sup>34</sup>O<sub>2</sub> is measured (Eq. 7). In the second approach, <sup>36</sup>O<sub>2</sub> is added to a sample, and its consumption is measured (Eq. 8). The production of <sup>34</sup>O<sub>2</sub> from an illuminated sample that contains <sup>18</sup>O-labeled water provides a direct measurement of the gross photosynthetic O<sub>2</sub> evolution:



In contrast, a less direct approach to measuring GPP uses the consumption of <sup>36</sup>O<sub>2</sub> to obtain the rate of O<sub>2</sub> consumption. In darkness, O<sub>2</sub> is consumed by respiration of organic matter.



Additional processes contribute to light-dependent O<sub>2</sub> consumption including photorespiration and the Mehler reaction. The rate of O<sub>2</sub> consumption is added to

the net O<sub>2</sub> production in the light to obtain a value for GPP. GPP measured using stable isotopes almost always exceeds estimates obtained from the light–dark bottle technique.

One of the primary motivations for developing the <sup>18</sup>O methods was to obtain data that could be compared with <sup>14</sup>C production to resolve whether the <sup>14</sup>C method yields results that are closer to NPP or GPP. During short incubations on the order of minutes, it is anticipated that <sup>14</sup>C production will be close to GPP because little of the fixed <sup>14</sup>C will have been respired back to CO<sub>2</sub>. However, as the duration of incubations increases up to one day, more of the organic matter will be labeled with <sup>14</sup>C, and at least some of this will be respired back to CO<sub>2</sub>. Direct comparisons of <sup>14</sup>C production with gross O<sub>2</sub> production have shown that the rate of CO<sub>2</sub> fixation is often about 50 % of the rate of gross O<sub>2</sub> evolution. This difference is far larger than can be explained by the photosynthetic quotient (see Eq. 4) and has been interpreted to indicate that there may be a significant rate of light-dependent O<sub>2</sub> uptake. Several processes may account for this increase including the Mehler reaction, photorespiration, and light-stimulated mitochondrial respiration.

## Net Community and Export Production

NCP can be estimated from the net changes (increases or decreases) in the concentration of dissolved O<sub>2</sub> in a water body provided that corrections are made for exchanges across the air–sea interface and the thermocline. A crude index of NCP is the ratio of O<sub>2</sub>-to-Ar because Ar (Argon) is an inert gas that is affected only by physical–chemical processes, whereas O<sub>2</sub> is affected not only by physical–chemical processes but also by photosynthesis and respiration. High values of the ratio of O<sub>2</sub>-to-Ar relative to those predicted for pure water by thermodynamics indicate that NCP is positive, whereas low values of this ratio indicate that NCP is negative. To obtain absolute values of NCP, appropriate corrections have to be made for differences in air–sea exchange kinetics for the two gases and for mixing between different source waters with different O<sub>2</sub>-to-Ar ratios. The following equation shows the expected relationship between changes in the O<sub>2</sub>-to-Ar ratio, NCP, and air–sea gas exchange.

$$\Delta(\text{O}_2 : \text{Ar})/\Delta t = \text{NCP} - \text{air/sea gas exchange} \quad (9)$$

where  $\Delta(\text{O}_2:\text{Ar})$  is the change in the ratio O<sub>2</sub>-to-Ar during the time interval  $\Delta t$ .

A less direct way to estimate NCP is from information on the respiration of organic matter in the deep waters below the euphotic zone. This is possible because NCP is exported to the waters below the euphotic zone where more than 99 % is respired, consuming O<sub>2</sub> and releasing CO<sub>2</sub>. Since the 1950s, oceanographers have been estimating the rate of respiration in the deep ocean from information on the oxygen content and the residence time of water at different depths in the ocean.

In addition to <sup>16</sup>O and <sup>18</sup>O, there is a third stable isotope, <sup>17</sup>O, which accounts for only 0.04 % of the total oxygen. Geochemists have developed sensitive methods to

measure differences in the ratios  $^{17}\text{O}:^{16}\text{O}$  and  $^{18}\text{O}:^{16}\text{O}$  in  $\text{O}_2$ , and these measurements can be used to estimate GPP without the need to incubate samples in bottles. This triple isotope method relies on differences between the isotopic composition of  $\text{O}_2$  added to the ocean by photosynthesis,  $\text{O}_2$  removed from the ocean by respiration, and  $\text{O}_2$  that dissolved into the ocean from the atmosphere. Photosynthesis produces  $\text{O}_2$  that has the same isotopic composition as seawater. In contrast respiration discriminates against the heavier isotopes and so increases the amounts of  $^{17}\text{O}$  and  $^{18}\text{O}$  relative to  $^{16}\text{O}$ , with greater increases in  $^{18}\text{O}:^{16}\text{O}$  than  $^{17}\text{O}:^{16}\text{O}$ . The isotopic composition of  $\text{O}_2$  in the atmosphere is not only affected by photosynthesis and respiration but also by the exchange of oxygen between  $\text{O}_2$ ,  $\text{O}_3$ , and  $\text{CO}_2$  in the stratosphere. As a consequence of these processes, the atmosphere has a higher ratio of  $^{18}\text{O}:^{17}\text{O}$  than seawater. Taken together, these differences allow GPP to be calculated from measurements of  $^{17}\text{O}:^{16}\text{O}$  and  $^{18}\text{O}:^{16}\text{O}$ .

### Is the Oligotrophic Ocean Autotrophic?

A major uncertainty in the open ocean carbon cycle concerns the zone that lies between about 10–40° north and south of the equator. Specifically, there is contradictory evidence about whether the nutrient impoverished oceanic ecosystems located in this zone are net producers or consumers of organic matter. Light–dark bottle measurements of primary production and comparisons of  $^{14}\text{C}$  fixation rates with bacterial respiration rates suggest that the surface waters in these regions consume more organic matter than they produce. In contrast, geochemical evidence indicates that these regions produce more organic matter than they consume. This evidence includes supersaturated concentrations of dissolved  $\text{O}_2$  in surface waters, export of organic matter in sinking particles, and estimates of rates of  $\text{O}_2$  consumption in the deep sea. A comparison of measurements of NCP obtained from  $\text{O}_2:\text{Ar}$ , GPP from oxygen triple isotope and  $\text{O}_2$  exchanges measured in bottles, suggests that GPP is likely to be underestimated in the bottles. However, the issue has not been resolved completely because several corrections need to be made when GPP is calculated from the triple oxygen technique. Specifically, as with the  $\text{O}_2:\text{Ar}$ -based estimates of NCP, accurate use of the triple isotope method requires correcting for mixing between different source waters, in particular the vertical entrainment of thermocline waters. Thus, although it is fairly well established that NCP is underestimated when samples are incubated in bottles, whether the cause of the underestimate is that respiration is stimulated or gross photosynthesis is inhibited has not been established unequivocally.

### Remote Sensing of Primary Production

The approaches described above provide measurements of primary production at fixed locations at specific times. Despite decades of research, these direct

observations are too few in number to calculate accurately the total primary production of the ocean, let alone how primary production varies geographically or how primary production changes through time. Fortunately, this problem can be redressed by using information collected using satellite remote sensing.

Satellites provide three types of data that are used when inferring primary production. These are chlorophyll *a* concentrations in the surface mixed layer, the amount of solar radiation reaching the sea surface, and sea surface temperature. Global distributions of chlorophyll *a* concentration are available from several sensors including the coastal zone color scanner (CZCS) for the 1980s and more recently the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) from 1997 to 2010 and MODIS (since 2002). Remote sensing of primary production relies on the fact that the primary production of a water column is correlated with the mixed layer chlorophyll *a* concentration. However, the correlation is not exact, which is why oceanographers use additional information including location, solar irradiance, and sea surface temperature in the calculations.

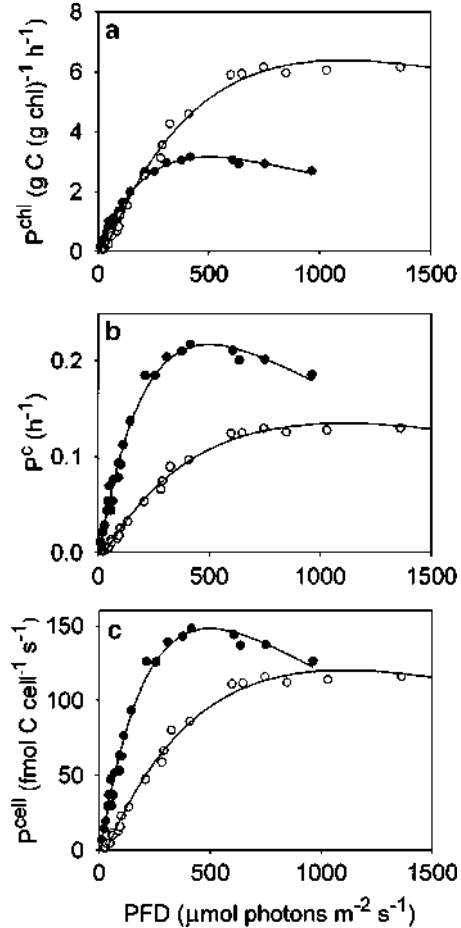
The earliest remote sensing estimates of primary production were made at local and regional scales where empirical relationships had been established between primary production and sea surface chlorophyll *a*. Extending the approach to ocean basin and global scales required that more complex algorithms be developed, several dozen of which have been devised. These differ in detail, but all rely on mixed layer chlorophyll *a* being a robust index of the depth, chlorophyll *a* content, and primary production of the euphotic zone. The basis of all algorithms is calculation of NPP from the chlorophyll *a* concentration and chlorophyll *a*-specific net photosynthesis rate:

$$\text{NPP} = [\text{chl } a] \cdot P^{\text{chl}} \quad (10)$$

where  $[\text{chl } a]$  is the chlorophyll *a* concentration and  $P^{\text{chl}}$  is the chlorophyll *a*-specific rate of net primary production. Ideally, NPP would be calculated throughout the day, taking into account the changes in solar radiation, the vertical distribution of chlorophyll *a*, and the dependence of  $P^{\text{chl}}$  on irradiance. In reality, information to justify this level of detail is not available, and many simplifications are made.

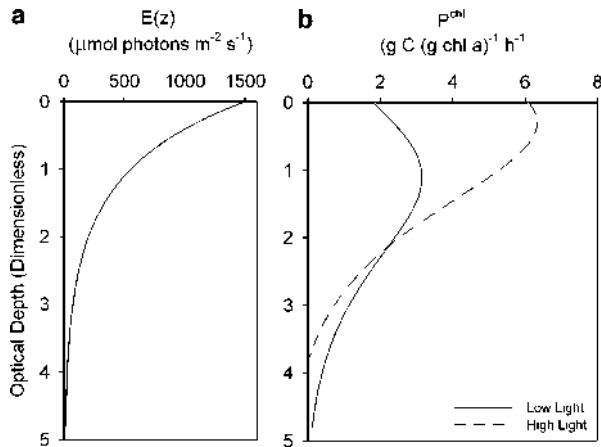
Oceanographers rely on the accumulated data base of primary production measurements to calibrate these algorithms. Since there are many ways in which this information can be combined, different scientists calculate different values of phytoplankton primary production even when using the same satellite data (Table 5). Compounding differences that arise from using different algorithms is uncertainty in the values of chlorophyll *a* and light attenuation inferred from the ocean color observations. The depth range that is “seen” by the satellite ocean color sensors corresponds to the upper 20 % of the euphotic zone. This means that 80 % of the euphotic zone is not sampled. Further restrictions on temporal coverage arise because ocean color cannot be observed under cloudy conditions. This drawback means that data is sparse for many regions.

**Fig. 8** Photosynthesis–light response curves for *Skeletonema costatum* acclimated to low light,  $50 \mu\text{mol photons m}^{-2} \text{s}^{-1}$  (filled circle), and high light,  $1,200 \mu\text{mol photons m}^{-2} \text{s}^{-1}$  (open circles). The same observations of  $\text{CO}_2$  fixation have been normalized to three different indices of biomass: to chlorophyll concentration in panel a; to organic carbon concentration in panel b; to cell abundance in panel c. Observations are for  $^{14}\text{CO}_2$  assimilation during 30-min incubations and thus approximate gross  $\text{CO}_2$  fixation (Data from the experiments reported by Anning et al. (2000))



## The Photosynthesis–Irradiance Response Curve

Many of the algorithms used in calculating NPP from ocean color explicitly incorporate algal physiology by accounting for the dependence of photosynthesis on irradiance. The relationship between the chlorophyll *a*-specific photosynthesis rate (designated  $P^{\text{chl}}$ ) and irradiance (designated  $E$ ) is one of the most widely studied aspects of phytoplankton ecophysiology. To obtain this relationship,  $P^{\text{chl}}$  is measured on samples that are incubated at different irradiances from darkness to full sunlight. The observed values of  $P^{\text{chl}}$  are then plotted against irradiance to obtain photosynthesis–irradiance or PE curve (Fig. 8). The PE curve is comprised of three regions. These are a low-light region in which the absorption of light energy limits photosynthesis, an optimal light region in which “dark” reactions limit photosynthesis, and a supraoptimal region in which photosynthesis is inhibited



**Fig. 9** Dependence of (a) irradiance and (b) chlorophyll *a*-specific photosynthesis rate on optical depth. Shown are the photosynthesis versus irradiance curves for *Skeletonema costatum* acclimated to low light and high light. The curves from Fig. 8 have been replotted versus optical depth for a surface irradiance of  $1,200 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ . Optical depth is defined as  $\ln(E(z)/E(0))$ , where  $E(z)$  is the irradiance at depth  $z$  and  $E(0)$  is the irradiance just below the sea surface. Optical depths of 2.3 and 4.6 correspond to 10 % and 1 % of surface irradiance

by further increases of irradiance. PE curves, like those illustrated in Fig. 8, can be used to construct vertical profiles of primary production provided that the surface irradiance and the vertical attenuation of light are known (Fig. 9).

Mathematical descriptions or models of the PE curve attempt to account for the influences of all the processes that affect the light dependence of photosynthesis using a small number of parameters. The minimum number of parameters required to account for the light dependence of gross photosynthesis is two. The first is the maximum photosynthesis rate when light is saturating. This parameter is designated  $P_m^{\text{chl}}$ . The second is the initial slope, which characterizes the rate of increase of photosynthesis at low light, designated  $\alpha^{\text{chl}}$ . Another term,  $E_K = P_m^{\text{chl}}/\alpha^{\text{chl}}$ , is often used to characterize whether cells are adapted to high light or low light because  $E_K$  indicates the irradiance at which photosynthesis begins to approach the light-saturated maximum rate.

The parameters of the PE curve are important photophysiological traits that influence gross photosynthesis. The initial slope ( $\alpha^{\text{chl}}$ ) accounts for the light dependence of photosynthesis at low light and is equal to the product of the chlorophyll-specific rate of light absorption  $a^{\text{chl}}$  and the maximum quantum yield ( $\phi_m$ ).

$$\alpha^{\text{chl}} = \phi_m a^{\text{chl}} \quad (11)$$

A cell's pigment content and composition, together with its size and shape, combine to determine the value of  $a^{\text{chl}}$ , as a consequence,  $a^{\text{chl}}$  varies widely between species, and it also varies with environmental conditions. The maximum quantum yield ( $\phi_m$ ) describes the greatest amount of photosynthesis that can be

achieved per unit photons absorbed. Cells “actively” alter  $\phi_m$  as they acclimate to altered environmental conditions; for example, photoacclimation to high light lowers  $\phi_m$  because “photoprotective” pigments are synthesized to dissipate more absorbed light energy and hence transfer less to photosynthesis.

The maximum value of photosynthesis ( $P_m^{\text{chl}}$ ) is observed at irradiances where light absorption no longer limits photosynthesis. What sets the value of  $P_m^{\text{chl}}$  is unclear; it may be limited by (i) the rate at which energy in the form of reductant (NADPH) and ATP is delivered to the Calvin cycle, (ii) the rate at which  $\text{CO}_2$  is incorporated into sugar phosphates by ribulose biphosphate carboxylase, or (iii) the rate at which the sugar phosphates produced by the Calvin cycle can be utilized. The rate-limiting step varies between different species and/or in response to different environmental conditions. More work to identify the mechanisms that set  $P_m^{\text{chl}}$  is essential given the importance of the light-saturated photosynthesis rate in determining phytoplankton production and the possibility of  $\text{CO}_2$  limitation of photosynthetic carbon fixation in some species.

Although  $P^{\text{chl}}$  is commonly reported by oceanographers and commonly employed in bio-optical algorithms, it is a poor predictor of phytoplankton growth. This is because there is taxonomic and phenotypic plasticity in the ratio of chlorophyll *a*-to-organic carbon. If, as is often the case, one wishes to know the specific growth rate of phytoplankton, then in addition to measuring  $P^{\text{chl}}$  one must know the ratio of chlorophyll *a*-to-carbon. This is because the three variables are related as follows:

$$\mu = P^{\text{chl}} \cdot [\text{chl} - \text{to} - \text{C}] - R_A \quad (12)$$

In this equation,  $\mu$  is the specific growth rate,  $P^{\text{chl}}$  is the chlorophyll *a*-specific photosynthesis rate, [chl-to-C] is the ratio of chlorophyll *a*-to-carbon, and  $R_A$  is the respiration rate. Consequently, the characteristics of PE curves normalized to chlorophyll *a*, organic carbon, or cell abundance differ (Fig. 8). Thus, care needs to be taken when using information from these curves as quantitative traits.

---

## Phytoplankton Ecology

The environmental factors that affect phytoplankton communities vary in time and space in predictable and unpredictable ways. One particularly important predictable pattern is in the seasonality of light and temperature in temperate and polar zones. In these regions the total biomass of phytoplankton varies widely, with periods of rapid proliferation in spring and autumn alternating with periods of decline. In contrast, phytoplankton biomass is much more stable throughout the year in subtropical and tropical waters that experience small changes in solar radiation and where seasonal forcing by light and nutrient availability is much lower. Other predictable patterns in phytoplankton community structure and primary production are associated with the large-scale ocean circulation. Superimposed on these predictable patterns is randomness in solar radiation and nutrient availability due to



changing weather and currents. For example, sustained changes of wind speed and direction in tropical waters can drive upwelling of nutrients to the surface, which in turn drives changes in primary production.

At the broadest geographical scale, the oceans can be divided into four broad domains (or “biomes”). These are:

- High-latitude polar regions (where seasonal forcing is strongest)
- Low-latitude (sub-)tropical regions (where seasonal forcing is weakest)
- Intermediate mid-latitude regions
- Coastal regions (where oceanic and atmospheric circulation patterns interact strongly with the continents)

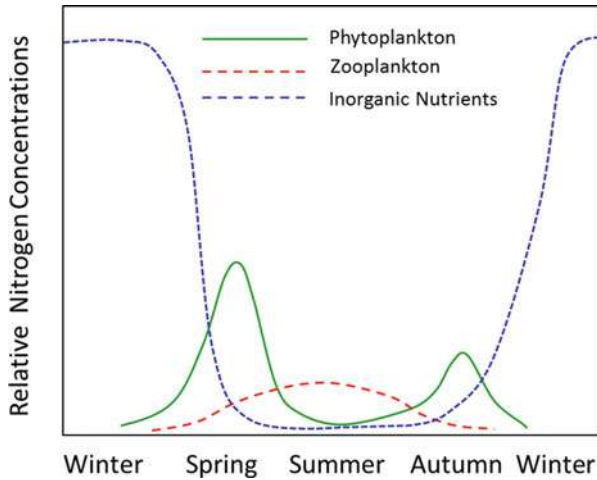
The location of physical oceanographic features, including pronounced horizontal gradients in temperature and salinity, is used to delineate different provinces within each of the four domains.

## **The Annual Phytoplankton Production Cycle in Temperate Zone Waters**

Early interest in phytoplankton ecology stemmed from a desire to understand why the abundances of commercially important fish, such as herring, varied widely from year to year. The herring fishery is seasonal, dependent on both herring population growth and migration between spawning and feeding grounds. Herring feed on plankton, which led marine biologists in Europe to investigate seasonal, year-to-year, and geographical changes in plankton abundance as a possible explanation for variations in the sizes of herring stocks. The seasonal cycle of irradiance and temperature in temperate waters was known to be accompanied by seasonal changes in plankton abundance. However, establishing cause–effect relationships to explain the seasonal plankton production cycle awaited the development of reliable, accurate, and rapid methods for measuring both the concentrations of inorganic nutrients and phytoplankton abundance.

The archetypical annual production cycle involves peaks of phytoplankton abundance in spring and autumn, with minima in summer and winter (Fig. 10). The classical explanation for this pattern is that net phytoplankton population growth is limited by low irradiance in winter and that biomass and growth are limited by low nutrient availability in summer. Peaks of phytoplankton abundance occur in spring and autumn when irradiance and nutrient availability are both sufficiently high to support population growth. Changes in the degree of vertical stratification of the water column play an important role in the annual production cycle.

During winter, when the sea loses heat to the atmosphere, surface waters cool, increase in density, sink, and displace subsurface waters, which in turn rise to the surface. Convection throughout winter brings nutrient rich water to the surface to replenish nutrient pools. However, phytoplankton cells are also mixed deeply within the water column, and so the average light level they experience is very low. As irradiance increases and air temperature rises in the spring, surface waters



**Fig. 10** Archetypical seasonal production cycle in temperate waters. The spring phytoplankton bloom occurs when solar radiation is sufficient to stabilize the water column and stimulate phytoplankton growth. Nutrient depletion and/or grazing by zooplankton brings the bloom to an end. Phytoplankton production in the mixed layer during the summer relies primarily on recycling of nutrients. An autumn phytoplankton bloom occurs when the mixed layer deepens. This bloom ends due to light limitation in deep mixed layer during winter

absorb heat and become more buoyant, and the mixed layer shoals. Consequently, the average irradiance that phytoplankton experience increases, and a spring bloom develops. The first quantitative explanation of the timing of the onset of the spring bloom was developed by Harold Sverdrup and is referred to as critical depth theory.

The possibility that the onset of the spring phytoplankton bloom occurs as a consequence of decreased grazing pressure exerted by zooplankton has recently been proposed as an alternative to the traditional theory that the bloom starts simply because the light environment becomes more favorable. The impact of zooplankton on phytoplankton populations decreases rapidly when deep mixing dilutes the abundances of both predators and prey. The reasoning behind this “dilution hypothesis” is that zooplankton will encounter phytoplankton much less frequently as both populations decrease due to dilution. In the case where the water below the euphotic zone is devoid of microorganisms, mixing equal volumes of deep water with surface water will decrease the encounter frequency, and hence the mortality due to zooplankton grazing, by a factor of four. Zooplankton populations, especially protozoan populations, may decline as a consequence of food limitation, opening up a window of opportunity for phytoplankton to escape being eaten by zooplankton when the water column begins to stabilize again.

As the spring bloom develops, much of the particulate matter sinks out of the surface layer as fecal pellets or amorphous aggregates of particulate organic and inorganic matter together with attached microorganisms. One explanation for the end of the spring bloom is depletion of nutrients associated with this export. Not all taxa are equally affected by nutrient limitation. In particular, the growth of diatoms

in the early stages of the spring depletes silicate, restricting further increases of diatom populations. This typically occurs before nitrate is depleted, allowing other phytoplankton taxa that do not require silicate, for example, the coccolithophorid *Emiliania huxleyi*, the opportunity to bloom.

However, it is also possible that the bloom will peak before nutrients are exhausted if the phytoplankton population is subjected to high rates of zooplankton grazing or by outbreaks of viral disease. Whether the spring bloom is terminated by nutrient limitation or high mortality, nutrients continue to be lost from the surface mixed layer as organic particles continue to sink below the pycnocline. Subsequently, inorganic N becomes depleted, especially at lower latitudes ( $<40^{\circ}\text{N}$ ), limiting primary production during the summer and causing the phytoplankton community to shift to flagellate and picoplankton assemblages.

During summer, phytoplankton in the surface layer rely on the recycling of nutrients, which can account for up to 80–90 % of primary production at this time of year. Consequently, primary production and heterotrophic consumption are tightly coupled during this low nutrient period. A subsurface chlorophyll *a* maximum (DCM) layer usually develops in the pycnocline at the interface between a nutrient-limited zone at shallower depths and a light-limited zone below. The phytoplankton in this layer intercept inorganic nutrients as they diffuse upwards from below, and the DCM can make a significant contribution to primary production in summer.

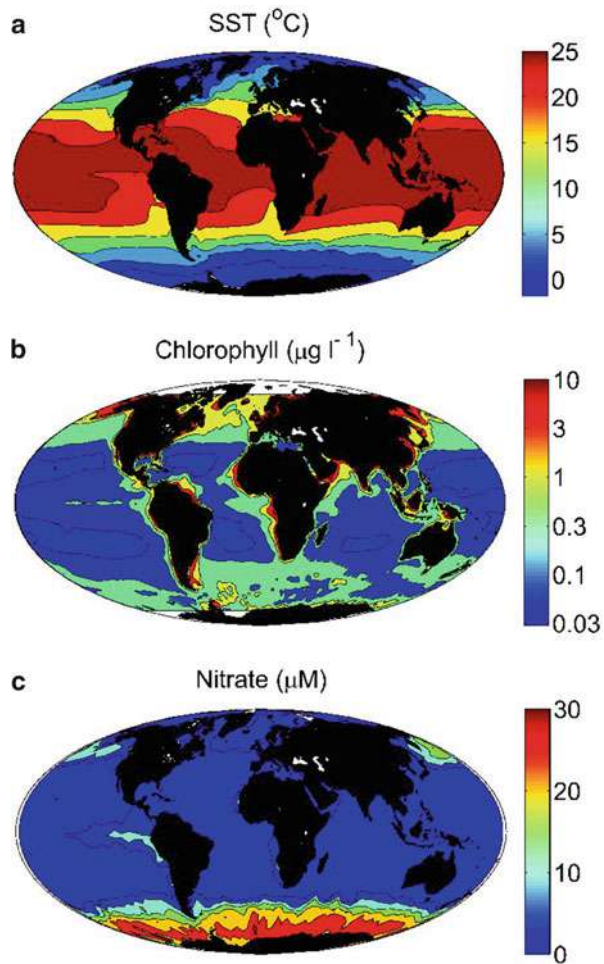
As solar radiation declines in autumn, surface waters cool, increase in density, and sink. This convective mixing erodes the pycnocline from above, transporting nutrients and phytoplankton from the DCM into the surface mixed layer. This can produce an autumn bloom, which eventually ends due to the light limitation as autumn gives way to winter.

This description of the annual phytoplankton cycle emphasizes how limitation by light and nutrients varies over the year. However, it has long been recognized that phytoplankton populations increase in abundance much more slowly than individual cells grow. The difference between the growth of individuals and the growth of populations is due to mortality. The annual production cycle remains a matter of active research because despite more than half a century of research, debate continues on the relative importance of nutrient and/or light limitation of individual growth versus mortality due to grazing and disease in controlling the size, productivity, and species composition of phytoplankton communities.

## Latitudinal Dependence of the Production Cycle

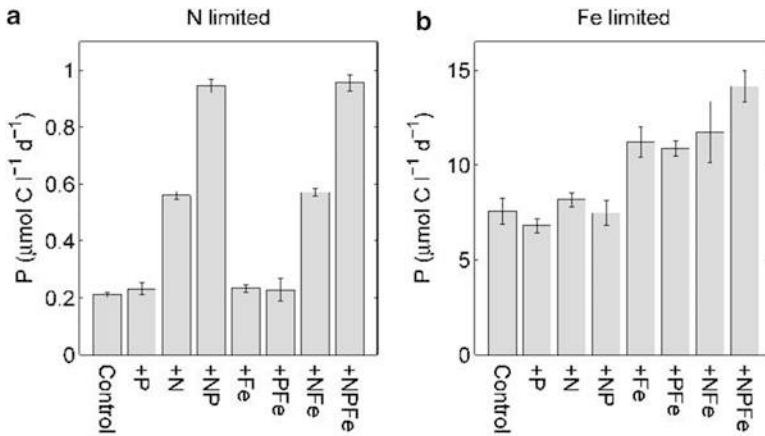
The extent of winter cooling of surface waters varies markedly with latitude, affecting the timing and extent of convective mixing, which in turn affects the various biotic and abiotic factors that influence the annual production cycle. The annual cycle of phytoplankton abundance observed in the temperate zone (Fig. 10) disappears in low-latitude tropical waters and is compressed into a single summer bloom in high-latitude polar waters. In the tropical and subtropical oceans, annual

**Fig. 11** Global maps of (a) sea surface temperature (SST), measured from satellites; (b) annual maximum sea surface chlorophyll concentrations measured from satellite ocean color; and (c) annual average sea surface nitrate concentrations compiled from multiple ship-based sampling expeditions. Data for the SST and nutrients are from the World Ocean Atlas: <http://www.nodc.noaa.gov/OC5/indprod.html>. Chlorophyll is from SeaWiFS: <http://oceancolor.gsfc.nasa.gov/SeaWiFS/>



variability in heat input is too small to generate enough cooling for convective overturning to penetrate very far into the permanent thermocline. Consequently, the convective input of nutrients to the surface is small. High solar radiation and the limited extent of convective mixing throughout the year create conditions in which phytoplankton growth and mortality remain tightly coupled resulting in low variability in phytoplankton biomass. The highly stratified regions of the subtropical and tropical oceans are characterized by year round near-surface nutrient depletion and relatively low uniform phytoplankton biomass (Fig. 11). Outside of the tropics, the timing of the phytoplankton “spring” bloom varies not only with the seasonal changes in the incident solar radiation but also with seasonal variability of mixed layer depth.

The extent of convective mixing during winter is one of the main determinants of the timing and magnitude of the bloom. Convective mixing increases at higher



**Fig. 12** Experimental data from nutrient addition bioassay experiments conducted in (a) a low-latitude N limited region of the subtropical Atlantic and (b) and higher latitudes in an Fe-limited region. Differences in primary production measured by  $^{14}\text{C}$  incorporation are measured in control samples and samples incubated with various concentrations of potentially limiting nutrients (Replotted from Moore et al. 2006)

latitudes, for example, in the North Atlantic from about 150 m at 30°N to >800 m at 60°N. Deeper convection leads to higher nutrient concentrations in surface waters during winter. Deeper convection also decreases the extent to which the growth of phytoplankton populations can be prevented by the grazing activity of zooplankton. Together these two factors (higher winter nutrient concentrations and lower top-down control of phytoplankton biomass by grazers) lead to more pronounced blooms at higher latitudes in the North Atlantic. For example, around 30°N in the North Atlantic, rather than the bloom occurring in the spring, it occurs during winter as nutrients are mixed into a well-lit surface layer. In contrast, in the regions furthest to the north (>60°N), stratification is delayed, and the main phase of the bloom occurs in summer. Year-to-year variability in weather (cloudiness and wind speed), which influences both vertical mixing and the amount of solar radiation that reaches the sea surface, can shift the timing of the bloom by several weeks.

## Nutrient Limitation

Nutrient limitation has proven to be one of the more contentious issues in phytoplankton ecology. Up until the 1980s, geochemists were convinced that phosphorus was the ultimate limiting nutrient in the sea, whereas biologists were equally convinced that the key limiting nutrient was nitrogen. Geochemists maintained that nitrogen could not be the limiting nutrient since  $\text{N}_2$  fixation would be used to obtain nitrogen when other forms were exhausted. However, biologists had shown from nutrient addition experiments that adding nitrate to samples stimulated phytoplankton growth but that adding phosphate on its own did not (Fig. 12) and so

concluded that nitrogen must be the limiting nutrient. The demonstration that iron can be a limiting factor over large parts of the ocean in the 1980s and 1990s added a new dimension to the debate between geochemists and biologists, but also helped to reconcile their differences. It is currently thought that the input of iron to the ocean limits that rate of  $N_2$  fixation, thus preventing the ocean as a whole from shifting from nitrogen limitation to phosphorus limitation.

Nutrient limitation is often inferred from correlative studies that examine the relationship between phytoplankton abundance or chlorophyll *a* concentration and inorganic nutrient distributions over time (seasonal cycle) and/or in space (vertically in water column or horizontally along a transect). In these studies, low concentrations of dissolved inorganic nutrients provide presumptive evidence of nutrient limitation. However, limitation of growth rate is not proven because recycling may be important and organic nutrients may be used. In addition, covariation in the concentrations of limiting nutrients often precludes unambiguous attribution to a single factor.

Presumptive evidence for limitation can be confirmed experimentally using bioassays (Fig. 12). These involve collection of a large volume of water, which is dispensed into a set of bottles to which the suspected limiting nutrients are added alone and in combination, incubated under appropriate light and temperature conditions and changes of biomass and other variables of interest are assessed. Although widely used, such bottle experiments are not without their critics. In particular, bioassay experiments perturb ecological processes that affect community structure such as predator–prey interactions or stimulation/inhibition of some functional groups such as diazotrophs. To allow examination of ecosystem scale responses to nutrient addition, oceanographers have turned to large-scale nutrient fertilization experiments. Briefly, a patch of water about 10–100 km<sup>2</sup> in area is enriched with the suspected limiting nutrient, and the increase of chlorophyll *a* and declines in inorganic nutrients and CO<sub>2</sub> are measured both inside and outside the patch over a period of a few days to a few weeks. An inert tracer, SF<sub>6</sub>, is added to account for advection and mixing. Such experiments have confirmed that Fe is a limiting nutrient in oceanic regions where the macronutrients such as nitrate remain high, but where chlorophyll *a* concentrations remain relatively low.

## Geographical Patterns of Nutrient Limitation in the Ocean

Much of the geographical variability in phytoplankton abundance and productivity can be attributed to variations in the rate of delivery of nutrients from the deep ocean to the euphotic zone. As discussed above (section “[Latitudinal Dependence of the Production Cycle](#)”), nutrient input to the sea surface depends on the depth and intensity of convective mixing during winter. In addition, ocean circulation contributes significantly to the regional- and global-scale patterns of nutrient availability. Of particular importance is upwelling of nutrient-rich water associated with divergences of surface currents and downwelling of nutrient-poor water

where surface currents converge. At the global scale, deep ocean waters containing high concentrations of nitrate and phosphate are upwelled to the surface of the Southern Ocean as a result of the westerly winds that circle the Antarctic continent. These waters are advected to lower latitudes by surface currents or into the permanent thermocline by subsurface currents. Subsequently, regional-scale upwelling of cool, nutrient-rich water from the thermocline occurs in coastal systems on the eastern boundaries of the low-latitude gyres and within the equatorial Pacific Ocean. These geographical patterns in resupply of deep ocean nutrients to the surface drive similarly large-scale patterns in the extent and nature of nutrient limitation, in phytoplankton distributions, and in pelagic ecology (Fig. 11).

Nitrogen, phosphorus, and silicate availability tend to be low throughout the year in the vast low-latitude subtropical and tropical oceanic regions, resulting in persistently low phytoplankton standing stocks. Exceptions to this overall pattern are observed within some coastal regions, where local upwelling can bring macronutrients to the surface. Also exceptional is the HNLC eastern equatorial Pacific where strong upwelling brings macronutrients to the surface along the equator. Away from these upwelling regions, the subtropical gyre regions which constitute >50 % of the ocean surface, and hence >30 % of the whole Earth surface, are highly oligotrophic. Dissolved inorganic forms of nitrogen ( $\text{NO}_3^-$ ,  $\text{NO}_2^-$ , and  $\text{NH}_4^+$ ) are highly depleted in the subtropical gyre systems, and bioassay experiments have confirmed that nitrogen is the proximal limiting factor for primary production in these systems.

As discussed previously, temperate and high-latitude North Atlantic waters are characterized by a seasonal cycle; light availability restricts phytoplankton growth in winter, while the lack of one or more nutrients contributes to the termination of the spring bloom. However, the marked annual cycle of macronutrient (N, P, and Si) concentrations and phytoplankton biomass that typifies the North Atlantic is unusual when considered in the context of the global ocean. The annual cycle of phytoplankton biomass and productivity is less pronounced in the other mid-latitude and high-latitude systems, including the Southern Ocean and the sub-Arctic North Pacific. Macronutrient concentrations remain high throughout the year in these systems, while overall peaks in phytoplankton biomass (or chlorophyll) are relatively low. Consequently, such regions are frequently termed high-nitrate, low-chlorophyll (HNLC) regions. In these HNLC regions, the concentrations of micronutrients, in particular Fe, are severely depleted.

The potential for Fe availability to play a major role in controlling phytoplankton production in these HNLC regions had been suspected for more than half a century; however it wasn't until the 1980s that John Martin and colleagues provided the first evidence in support of this hypothesis. Both bottle-enrichment experiments (Fig. 12b) and experimental releases of dissolved Fe into the ocean have demonstrated unequivocally that phytoplankton photosynthesis and growth responds positively to the addition of Fe in the HNLC regions of the Southern Ocean, the eastern equatorial Pacific, and the sub-Arctic North Pacific. Studies of naturally iron-enriched coastal systems, for example, around sub-Antarctic

Islands, provide further support to this theory, and the Fe-limited status of the HNLC systems is now widely accepted.

The existence of the HNLC systems can be understood by considering the chemistry of dissolved Fe in seawater. Fe is highly insoluble and readily sticks to particles in well-oxygenated seawater. Consequently, while inorganic nitrogen and phosphorus are returned to the dissolved pool when organic matter decomposes, iron remains attached to particles. These sink to the seabed, removing iron and leaving behind an excess of dissolved nitrate and phosphate. Physical transport of deep waters back to the surface supplies large amounts of the macronutrients nitrate and phosphate, but very little Fe. It is therefore unsurprising that net growth of phytoplankton depletes Fe before the macronutrients can be consumed, leading to the development of Fe-limited systems. From this context, it is relevant to ask “Why do the macronutrients N and P ever become depleted to the point where they become limiting?”

The answer lies in considering the sources of Fe to the upper layer of the oceans. The main inputs are from the Fe released from anoxic coastal sediments and from dust generated in from arid regions and blown across the oceans by the wind. These sources deliver large amounts of Fe to the lower latitudes of the North Atlantic, which is one of the few Fe-replete ocean basins. In contrast, delivery of Fe to the HNLC regions is insufficient to provide all of the Fe needed by phytoplankton to fully utilize all macronutrients because the major HNLC regions are distant from the largest dust sources. However, another important factor is that all of the HNLC systems are characterized by high rates of deep mixing and/or wind-driven upwelling, which replenish macronutrients. Hence, a large annual supply of Fe would be required to fully remove all the macronutrients from these systems. Conversely, under conditions where the resupply of subsurface nutrients is slower, such as within the stable highly stratified subtropical ocean gyres, dust and other fluxes of Fe are sufficient to enable phytoplankton to fully utilize all the macronutrients.

The overall pattern of nutrient limitation at large scales can be summarized as follows: iron is the limiting element in the upwelling dominated HNLC regions which comprise around 30–40 % of the oceans. Nitrogen is the limiting element over most of the remainder of the ocean, dominated by the downwelling subtropical gyres. Exceptional is the Mediterranean Sea in which  $N_2$  fixation and primary production appear to be P limited, particularly in the eastern basin.

## **Adaptations to Nutrient Limitation**

Phytoplankton have evolved a range of adaptations for coping with nutrient limitation. One unifying selective pressure is related to the advantage of small cell size under conditions where diffusion limits the transport of nutrients to the cell surface. This can be readily understood by considering how nutrient flux towards a cell and cellular biomass are related to cell size. For simplicity, assume that the cell is spherical and that biomass is proportional to cell volume. At low nutrient concentration, the supply of nutrients to the cell is proportional to its radius ( $r$ ), but the



requirement for nutrients increases as the radius cubed ( $r^3$ ), and thus the growth rate will decrease as the inverse of radius squared:

$$\mu = (a S)/r^2 \quad (13)$$

In this equation,  $\mu$  is the growth rate,  $S$  is the concentration of the limiting nutrient, and “ $a$ ” is a constant that accounts for the diffusion coefficient of the limiting nutrient in water and the intracellular concentration of that nutrient per unit cell volume.

These considerations (Eq. 13) suggest that when nutrients are at extremely low concentrations, *Prochlorococcus* cells with a radius of about 0.25  $\mu\text{m}$  should have the potential to grow 4 times faster than *Synechococcus* cells with a radius of 0.5  $\mu\text{m}$  and 16 times faster than picoeukaryotes cells with a typical radius of 1  $\mu\text{m}$ . Thus, even within the picoplankton, there is scope for cell size to modify growth rates by as much as 16-fold. Direct measurements of the growth rates of these three groups indicate that picoeukaryotes do indeed grow slower than *Prochlorococcus* and *Synechococcus*, but not by as much as this simple calculation predicts. This suggests the growth rates of *Prochlorococcus* and *Synechococcus* are not likely to be severely limited by diffusion of nutrients to the cell surface. This is unlikely to be the case for larger phytoplankton in the nanoplankton or microplankton, because all other factors being equal, a cell with a 10  $\mu\text{m}$  radius will have a 100-fold lower affinity for limiting nutrients than a cell with a 1  $\mu\text{m}$  radius and hence the large cell would be at a considerable disadvantage when competing for nutrients. Although small size can reduce diffusion limitation, it can come at a price. For example, in order to achieve its small cell size, *Prochlorococcus* has reduced the size of its genome, including dispensing with the ability to take up nitrate.

At the opposite extreme of the size range, large phytoplankton can take advantage of their ability to migrate up and down through the water column to acquire nutrients. Some phytoplankton migrate between the surface mixed layer where nutrients are low but irradiance is high and the thermocline, where nitrate and phosphate concentrations are high, but irradiance is low. The migration rate is related to cell size, with larger cells being able to migrate more rapidly. Phytoplankton that undertake vertical migration to tap this deep pool of inorganic nutrients include organisms that can swim (dinoflagellates) and those that can regulate their buoyancy (*Trichodesmium* and large diatoms). Accumulation of starch, which has a density much higher than that of water, occurs at high irradiance. Starch provides cells with an energy store and also ballast that adds to density, thus aiding sinking. The starch is metabolized in the low-light environment of the pycnocline to provide the energy to assimilate nutrients and exclude heavy ions. This contributes to buoyancy allowing cells to float back into the mixed layer. Very small cells are unable to make use of this strategy because the maximum rate at which they can move vertically is too slow.

Uptake and assimilation of organic nutrients is another adaptation to limiting concentrations of inorganic nutrients. This commonly involves the use of hydrolases and amino acid oxidases on the cell surface to cleave phosphate and

ammonium from organic molecules that are dissolved in seawater. Organisms that use this strategy also express high-affinity nutrient transporters to insure uptake of the ammonium and phosphate released by these enzymes. Other phytoplankton can obtain nutrients by ingesting particles including bacteria and smaller phytoplankton cells.

As previously discussed, nitrogen fixation is employed by *Trichodesmium* and other diazotrophs to obtain nitrogen. However, diazotrophs require P and Fe in addition to N, and these nutrients likely limit N<sub>2</sub> fixation over large parts of the ocean. High *Trichodesmium* abundances and high N<sub>2</sub> fixation rates in the North Atlantic Ocean occur downwind of the Sahara Desert and the semiarid Sahel regions of Northern Africa due to deposition of wind-borne dust blown that contains high amounts of Fe. *Trichodesmium* still requires P, which it can obtain from hydrolysis of dissolved organic phosphorus compounds.

---

## Anthropogenic Impacts on Marine Phytoplankton

Impacts to the natural environment by human activities have been recognised since documentation of bioaccumulation of pesticides in top predators in the 1950s and of stratospheric ozone depletion in the 1970s. Farming, deforestation, fishing, and industrial activity are among the drivers of changes in ocean biodiversity, nutrient cycles, and climate. This section outlines some of the anthropogenic impacts on marine phytoplankton at regional and global scales. In some cases, such as coastal eutrophication, there have been clear and dramatic impacts. In other cases, including global warming and ocean acidification, the impacts that have occurred to date have been relatively subtle. Continued global warming is expected to profoundly influence marine phytoplankton ecology, mainly through changes in ocean circulation and vertical mixing. Ocean acidification is expected to affect calcifying organisms including coccolithophorids. Other impacts on pelagic food webs have arisen from the devastation of the populations of large pelagic predator populations by overfishing. The abundances of top predators have been reduced by 80–90 % over vast areas of the ocean by intensive fishing.

One of the first of the global-scale anthropogenic impacts on marine primary production to be investigated was whether increased UV-B radiation reaching the Earth's surface in the Arctic and Antarctic due to stratospheric ozone depletion has reduced the net primary production of high-latitude marine ecosystems. These ecosystems can support large populations of crustaceans (krill), fish, and marine mammals. Loss of stratospheric ozone over polar regions in spring has been documented since the late 1970s. This loss was catalyzed by accumulation of chlorofluorocarbons that are used as refrigerants and propellants. Loss of ozone since the 1970s has allowed UV-B radiation to increase by 130 % under the springtime Antarctic ozone hole. Most research suggests that the inhibiting effects of natural levels of UV-B radiation are already large and that the increases of UV-B due to ozone depletion have been marginal. Some estimates suggest that primary production in the spring may be reduced by as much as 8 %, but others suggest

**Table 6** Preindustrial, current, and projected future inputs of nitrogen, phosphorus, and silicate to the ocean. Values are in Tg of N, P, or Si per year. The wide range of values between the studies indicates the considerable uncertainty in these estimates

		Gruber (2008); Bennett et al. (2001)		Duce et al. (2008)		Seitzinger et al. (2010)		
		Preindustrial	1990	1860	2000	1970	2000	Future (2030)
Nitrogen	N <sub>2</sub> fixation	135 ± 50	135 ± 50	–	60–200	–	–	–
	River discharge	30	80 ± 20	–	–	37	43	41–48
	Atmospheric deposition	6	50 ± 20	10–30	38–96	–	–	–
Phosphorus	River discharge	8	22	–	–	5.9	6.6	8.4–8.5
	Atmospheric deposition	1	1	–	–	–	–	–
Silicate	River discharge	–	–	–	–	142	144	136–138

much lower effects. The assessment of the inhibition of primary production under the ozone holes is complicated by difficulty in accounting for nonlinear effects of UV-B and the interaction of UV-B with visible radiation in phytoplankton cells that are subjected to vertical mixing.

The nutrient load to the ocean has increased dramatically over the past 300 years as a result of population growth and intensification of farming practices. Some estimates suggest that nitrogen and phosphorus inputs have increased by two to three times above preindustrial levels, although there is considerable uncertainty as calculations vary by about twofold (Table 6). Increased phosphorus and nitrogen loading has not been evenly spread across the ocean. For example, loads to Chesapeake Bay have increased sixth- to eightfold and loads to the North Sea by about 10 times. Coincident with increased nutrient loading have been increases in the incidence of harmful algal blooms (HABs) in coastal waters. Some HAB species produce toxins, which can kill fish, shellfish, marine mammals, and/or seabirds. Algal blooms can “harm” ecosystems in other ways. Persistent low oxygen (hypoxic) conditions are found where O<sub>2</sub> is depleted due to decomposition of organic matter that has sunk from the surface to bottom waters and sediments. These “dead zones” are found in the Gulf of Mexico under the Mississippi River plume, off the east costs of Asia and North America and in coastal waters of Northern Europe. At the same time that anthropogenic N and P inputs have increased, changes in the terrestrial and freshwater nutrient cycling have led to a decrease in the inputs of silicate. The increased nitrogen-to-silicate ratio that rivers deliver to coastal waters has shifted the composition of phytoplankton communities away from diatoms and toward flagellates, often decreasing the nutritional quality and palatability of the phytoplankton. The input of nitrogen to the ocean from the atmosphere has also increased due to emission of NO and NO<sub>2</sub> accompanying combustion of fossil fuels and emission of NH<sub>3</sub> during the production and use of fertilizers.

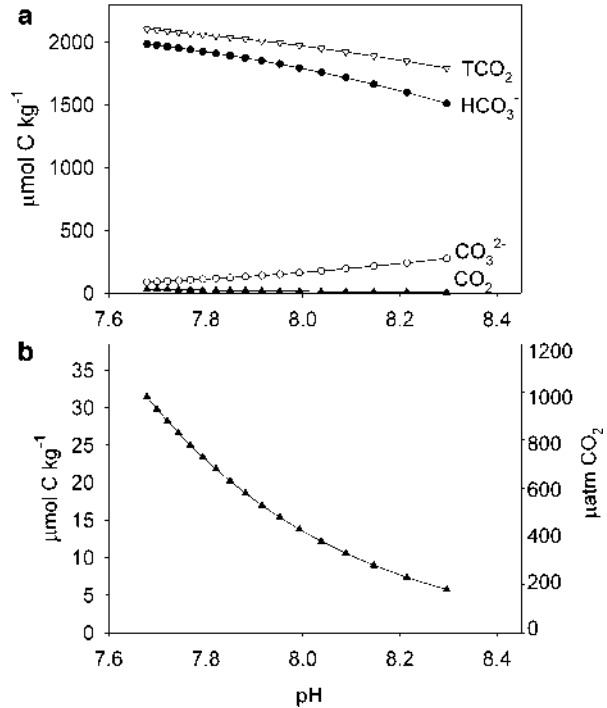
Plumes of air polluted with these nitrogen compounds extend far downwind of major population centers. Anthropogenically produced nitrogen compounds are being deposited over almost all areas of the open ocean, with about 75 % of the nitrogen deposition in regions that are nitrogen limited, and the input of anthropogenic nitrogen into these regions is already approaching 50 % of the natural input due to nitrogen fixation.

The average temperature of the atmosphere has increased by about 1 °C in the past 150 years. Most climate scientists attribute this to the increased concentrations of CO<sub>2</sub> and other greenhouse gases in the atmosphere. The increase in air temperature would have been dramatically larger were it not for the moderating influence of the oceans over this time period. The oceans have absorbed about 40 % of the CO<sub>2</sub> released through burning fossil fuels and deforestation. This has slowed the buildup of atmospheric CO<sub>2</sub>, which nonetheless is already over 1.4 times higher than preindustrial levels. In addition, the oceans absorb a large amount of heat that would otherwise warm the atmosphere; average sea surface temperature (SST) has increased by about 1 °C since 1880, and the interior of the ocean is also warming.

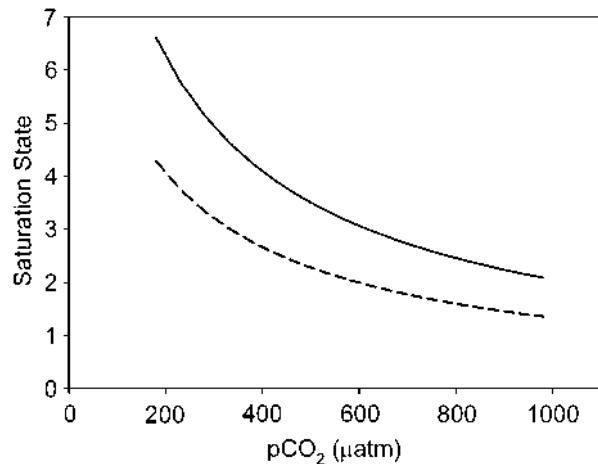
Ocean warming has already affected the geographical distributions of plankton. For example, there has been a well-documented northward shift in the distributions of boreal and temperate copepod species in the North Atlantic Ocean. Ocean warming will be accompanied by changes in ocean circulation and seasonal cycles of stratification and mixing. The spatial extent of the subtropical gyres is expected to expand, and the intensity of vertical mixing is likely to decrease. These regions are characterized by year round or seasonally low macronutrient concentrations. Consequently, increases in the area of these regions are likely to be accompanied by a decline of oceanic net primary production. The flux of organic matter out of the surface to the deep ocean (export production) is also likely to decline as stratification increases in the future. Although these processes will perturb the cycling of carbon through the marine system, the feedbacks on atmospheric CO<sub>2</sub> are not simple to predict. They may be relatively minor, as the decreased export of organic carbon should be balanced, in part, by a decrease in the return of CO<sub>2</sub> and other forms of inorganic carbon from the deep ocean to the surface.

Changes in atmospheric circulation and in the hydrological cycle, which are accompanying global warming, are likely to affect the availability of iron to phytoplankton. Changes in the areas of arid regions and changes in atmospheric circulation will affect the amounts of iron delivered to different ocean basins by the wind. Primary production will be stimulated if more iron is delivered to the iron-limited HNLC regions. Increased iron inputs to the nitrogen-limited subtropical gyres may also stimulate primary production by reducing the extent to which iron limits N<sub>2</sub> fixation. Unfortunately, our understanding of the feedbacks in the climate system is still too rudimentary to accurately predict how transport of atmospheric dust to the oceans will change in a warming planet. Thus, it is also not possible to predict the effect on marine phytoplankton. However, signals in the geological record suggest that significant changes in oceanic primary production that have occurred in the past were related to changes in Fe inputs.

**Fig. 13** Changes in the chemical speciation of inorganic carbon of seawater as a function of pH (seawater scale) for atmospheric  $\text{CO}_2$  concentrations ranging from 180 to 1,000 ppm  $\text{CO}_2$  by volume. Calculations are for a temperature of  $20^\circ\text{C}$ , a salinity of 35 practical salinity units, and an alkalinity of  $2.2\text{ mmol kg}^{-1}$ . Calculations were made using CO2SYS (van Heuven et al. 2009). (a) Dissolved  $\text{CO}_2$  (filled triangles); carbonate (open circles); bicarbonate (filled circles); total inorganic carbon (inverted open triangles). (b) Same data as in (a) for  $\text{CO}_2$ , replotted on an expanded scale



**Fig. 14** Changes in the saturation state of two forms of calcium carbonate in seawater as a function  $\text{CO}_2$  concentrations. Calculations are for a temperature of  $20^\circ\text{C}$ , a salinity of 35 practical salinity units, and an alkalinity of  $2.2\text{ mmol kg}^{-1}$ . Calculations were made using CO2SYS (van Heuven et al. 2009)



The declining pH of the ocean due to invasion of the  $\text{CO}_2$  produced by man's activities is called ocean acidification (see sections "[Dissolved Inorganic Carbon](#)" and "[Ocean Acidification](#)") is called ocean acidification (OA). Ocean acidification is significantly altering the chemistry of seawater, including pH and  $\text{CO}_2$  (Fig. 13) and calcium carbonate saturation state (Fig. 14). Critically, the current rate of pH

change is 100 times faster than the natural rates of pH change that have occurred in the past. The potential influences of these changes on phytoplankton photosynthesis and calcification have been investigated with laboratory monocultures and mesocosm experiments.

Laboratory investigations on a small number of marine phytoplankton species indicate that the response of growth rate to  $\text{CO}_2$  is most pronounced at  $\text{CO}_2$  levels that are significantly lower than present-day values. Further increases of  $\text{CO}_2$  are expected to have a negligible impact on the growth rate of most species. This lack of response is likely due to the presence of carbon-concentrating mechanisms (CCMs) that insure sufficient  $\text{CO}_2$  enters phytoplankton cells to meet the requirements for photosynthesis. Species in which growth rate increases in response to elevated  $\text{CO}_2$  may lack CCMs or have inefficient CCMs. Some studies suggest that growth of some picoplankton may be stimulated by elevated  $\text{CO}_2$  whereas others suggest that it is microphytoplankton that benefit the most. However, even in these cases, the effect of doubling  $\text{CO}_2$  from current levels is often small, typically less than 10%. Nonetheless, small differences in the response of growth rate to elevated  $\text{CO}_2$  among species may still significantly affect phytoplankton community structure due to the cumulative effect of differences in exponential growth over many generations. Unlike growth rates, which are largely unaffected by ocean acidification, the rate of calcium carbonate precipitation by coccolithophorids shows a marked response. Although most studies show either no effect or a slight inhibition of growth rate of coccolithophorids in elevated  $\text{CO}_2$ , calcification usually declines in response to OA, and the ratio of calcification to photosynthesis declines as a consequence.

The insights from laboratory monoculture experiments do not allow assessment of how OA affects species interactions, including competition for nutrients and predator–prey dynamics. To address these issues, researchers have examined intact plankton communities via experimental manipulations of “closed” systems (shipboard microcosms or in situ mesocosms) or observations of “open” systems made along natural pH/p $\text{CO}_2$  gradients. Open system observations take advantage of the fact that low-pH seawater is found “naturally,” for example, upwelling of intermediate waters along the western North American continental margin and volcanic  $\text{CO}_2$  vents in the Mediterranean and Indo-Pacific. These studies on intact communities have demonstrated that community structure responds to manipulation of pH and p $\text{CO}_2$ . Nonetheless, results of these studies remain highly variable, thus limiting our ability to predict reliably the possible effects of increasing  $\text{CO}_2$  and OA on phytoplankton productivity and ocean nutrient cycling.

---

## Future Directions

Major unsettling of the earth–atmosphere–ocean system – including global warming, ocean acidification, and cultural eutrophication – is impacting marine ecosystems. Currently, a predictive understanding of how these changes will affect phytoplankton communities and productivity is lacking. Thus, a major focus for

ongoing and future research will be to document the changes in marine ecosystems that are arising from anthropogenic activity and to develop a mechanistic understanding of why these changes are taking place. The goal is to obtain enough knowledge to be able to make informed projections of the future state of marine ecosystems and of the role of these ecosystems in global biogeochemistry. The major questions include: How will phytoplankton species adapt to changing ocean temperature and pH? How will phytoplankton communities be reorganized by the responses to these changes? How will these changes in phytoplankton ecology affect ocean biogeochemistry, for example, through release of climate reactive trace gases? How will changes in phytoplankton influence higher trophic levels, for example, impacting on fisheries yields, and how will overfishing affect phytoplankton ecology?

Satellite remote sensing allows us to measure how phytoplankton biomass varies across the ocean. Calculating primary production from this information depends on algorithms, which in the past have been developed from calibration against  $^{14}\text{C}$  measurements. Ideally, these algorithms should instead be derived from first principles and then tested against the  $^{14}\text{C}$  measurements. Unfortunately, our understanding of the fundamental biological processes driving phytoplankton growth and productivity (and how they are regulated by the environment) lags behind our capability to measure biomass. Therefore, research needs to be undertaken to better understand the ecophysiology of phytoplankton photosynthesis and the ecology and evolution of phytoplankton communities.

To date, most studies of phytoplankton ecophysiology have tended to examine one factor at a time, holding others constant. Although such studies can be useful for gaining the most straightforward scientific insight, in an oceanic environment where several factors naturally change simultaneously, it will be necessary to conduct multifactorial investigations. However, because the number of experimental treatments that can potentially be investigated increases exponentially with the number of different interacting factors under consideration, the design of such studies needs to be informed by a clear understanding of how factors may covary in both natural and anthropogenically perturbed systems.

The challenges of the multifaceted marine environment are particularly acute when considering the biotic interactions that affect competition and succession. Environmental change may simultaneously influence multiple trophic levels and the interactions between them. In particular, sources of mortality remain relatively underexplored when compared to the bottom-up processes of resource limitation. Mortality can arise from grazing by zooplankton and protozoa and/or by infection by viruses and pathogenic bacteria. How these other components of the ecosystem respond to climate change will no doubt be less predictable than those that will take place in the physical-chemical environment.

Genomic, transcriptomic, and proteomic approaches have the potential to contribute to increasing our mechanistic understanding of the linkage between the physiology of phytoplankton and their reciprocal interactions with the oceanic environment. High-throughput sequencing is already revealing the high taxonomic and metabolic diversity of marine phytoplankton alongside the complex integrated

changes in cellular activity which can occur as a result of changing environmental conditions. The growing use of in-depth genotyping and phenotyping of whole microbial communities using meta-genomic, transcriptomic, and proteomic techniques should provide further insights into the mechanisms by which the environment selects for different genotypes and how the activities and interactions between the organisms characterized by these genes subsequently influence the cycling of nutrients, energy, and carbon through oceanic systems. Moving forward, the development of transformable genetic systems will likely provide unprecedented information on the function of individual genes and gene products within selected phytoplankton taxa.

In summary, developing a predictive understanding of how and why phytoplankton communities and primary production vary in space and time is a prerequisite for predicting how future changes in ocean physics and chemistry due to global warming and ocean acidification will affect the roles that phytoplankton play in the marine carbon cycle and marine food webs. Superficially, primary production is a simple concept; but the deeper understanding that oceanographers are now seeking demands addressing the complex interplay of biochemical, physiological, and ecological processes.

---

## References

- Anning T, MacIntyre HL, Pratt SM, Sammes PJ, Gibb S, Geider RJ. Photoacclimation of the marine diatom *Skeletonema costatum*. *Limnol Oceanogr.* 2000;45:1807–17.
- Bennett EM, Carpenter SR, Caraco NF. Human impact on erodible phosphorus and eutrophication: a global perspective. *BioScience.* 2001;51:227–34.
- Boyd PW, et al. Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. *Science.* 2007;315:612–17.
- Buitenhuis, Li WKW, Vaulot D, Lomas MV, Landry MR, Partensky F, Karl DM, Ulloa O, Campbell L, Jacquet S, Lantoiné F, Chavez F, Macias D, Gosselin M, McManus GB. Picophytoplankton biomass distribution in the global ocean. *Earth Syst Sci Data.* 2012;4:37–46. doi:10.5194/essd-4-37-2012. [www.earth-syst-sci-data.net/4/37/2012/](http://www.earth-syst-sci-data.net/4/37/2012/)
- Carr M-E, et al. A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Res.* 2006;53(Pt II):741–70.
- Duarte CM, Cebrián J. The fate of marine autotrophic production. *Limnol Oceanogr.* 1996;41:1758–66.
- Duce RA, et al. Impacts of atmospheric anthropogenic nitrogen on the open ocean. *Science.* 2008;320:893–7.
- Falkowski PG, Dubinsky Z, Wyman K. Growth irradiance relationships in phytoplankton. *Limnol Oceanogr.* 1985;30:311–21.
- Fasham MJR, Ducklow HW, McKelvie SM (1990) A nitrogen based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Systems* 48:591–639.
- Feng Y, et al. Effects of increased pCO<sub>2</sub> and temperature on the North Atlantic spring bloom. I. The phytoplankton community and biogeochemical response. *Mar Ecol Prog Ser.* 2009;388:13–25.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science.* 1998;281:237–40.
- Follows MJ, Dutkiewicz S, Grant S, Chisholm SW. Emergent biogeography of microbial communities in a model ocean. *Science.* 2007;315:1843–6.



- Gruber N. The marine nitrogen cycle: overview and challenges. In: Capone DG, Bronk DA, Mulholland MR, Carpenter EJ, editors. *Nitrogen in the environment*. 2nd ed. Amsterdam: Elsevier; 2008. p. 1–50.
- Jeffrey SW, Wright SW, Zapata M (2011) Microalgal classes and their signature pigments. In *Phytoplankton pigments characterization, chemotaxonomy and applications in oceanography* (Eds. Roy S, Llewellyn CA, Egeland ES, Johnsen G). Cambridge University Press. PP. 3–77.
- Kiddon J, Bender ML, Marra J (1995) Production and respiration in the 1989 North Atlantic spring bloom: An analysis of irradiance-dependent changes. *Deep-Sea Res.* 42:553–576.
- Le Quéré C, et al. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob Change Biol.* 2005;11:2016–40.
- Lohbeck KT, Riebesell U, Reusch TBH. Adaptive evolution of a key phytoplankton species to ocean acidification. *Nat Geosci.* 2012;5:1–6.
- Luo YW, et al. Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst Sci Data.* 2012;4:47–73. doi:10.5194/essd-4-47-2012. [www.earth-syst-sci-data.net/4/47/2012/](http://www.earth-syst-sci-data.net/4/47/2012/)
- Mills MM, Redame C, Davey M, LaRoche J, Geider RJ. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature.* 2004;429:292–4.
- Moore CM, Mills MM, Milne A, Langlois R, Achterberg EP, Lochte K, LaRoche J, Geider RJ. Iron limits primary productivity during spring bloom development in the central North Atlantic. *Glob Change Biol.* 2006;12:626–34.
- Morel A. Optical modelling of the upper ocean in relation to its biogenic matter content (case-I waters). *J Geophys Res-Oceans.* 1988;93:10749–68.
- Raven JA. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol.* 2009;56:177–92.
- Riebesell U, Gattuso JP, Thingstad TH, Middelburg JJ. Arctic ocean acidification: pelagic ecosystem and biogeochemical responses during a mesocosm study. *Biogeosciences.* 2013;10:5619–26.
- Romero E, Peters F, Marrase C. Dynamic forcing of coastal plankton by nutrient imbalances and match-mismatch between nutrients and turbulence. *Mar Ecol Progr Ser.* 2012;464:68–87.
- Seitzinger SP, Mayorga E, Bouwman AF, Kroeze C, Beusen AHW, Billen G, Van Drecht G, Dumont E, Fekete BM, Garnier J, Harrison JA. Global river nutrient export: a scenario analysis of past and future trends. *Glob Biogeochem Cycles.* 2010;24:GB0A08. doi:10.1029/2009GB003587.
- Sunda WG, Shertzer KW, Hardison DR. Ammonium uptake and growth models in marine diatoms: Monod and Droop revisited. *Mar Ecol Progr Ser.* 2009;386:29–41.
- Uitz J, Claustre H, Gentili B, Stramski D. Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. *Global Biogeochem Cycles.* 2010;24, GB3016. doi:10.1029/2009GB003680.
- van Heuven SD, Lewis PE, Wallace DWR. MATLAB Program Developed for CO<sub>2</sub> System Calculations. ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee; 2009.

## Further Reading

- Arigo KR. Marine microorganisms and global nutrient cycles. *Nature.* 2005;437:349–55.
- Barton AD, Pershing AJ, Litchman E, Record NR, Edwards KF, Finkel ZV, Kiørboe T, Ward BA. The biogeography of marine plankton traits. *Ecol Lett.* 2013;16:522–34.
- Boyd PW, Strzepek R, Fu F, Hutchins DA. Environmental control of open-ocean phytoplankton groups: now and in the future. *Limnol Oceanogr.* 2010;55:1353–76.
- Cullen JJ, Boyd PW. Predicting and verifying the intended and unintended consequences of large-scale ocean iron fertilization. *Mar Ecol Progr Ser.* 2008;364:295–301.

- Cullen JJ, Franks PJS, Karl DM, Longhurst A. Physical influences on marine ecosystem dynamics. In: Robinson AR, McCarthy JJ, Rothschild BJ, editors. *The sea, Biological-physical interactions in the ocean*, vol. 12. Boston: Harvard University Press; 2002. p. 297–336.
- Day TA, Neale PJ. Effects of UV-B radiation on terrestrial and aquatic primary producers. *Annu Rev Ecol Syst.* 2002;33:371–96.
- Diaz RJ, Rosenberg R. Spreading dead zones and consequences for marine ecosystems. *Science.* 2008;321:926–9.
- Doney SC, Fabry VJ, Feely RA, Kleypas JA. Ocean acidification: the other CO<sub>2</sub> problem. *Ann Rev Mar Sci.* 2009;1:169–92.
- Falkowski PG, Barber RT, Smetacek V. Biogeochemical controls and feedbacks on ocean primary production. *Science.* 1998;281:200–6.
- Falkowski PG, Katz M, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor JFR. The evolution of modern eukaryotic phytoplankton. *Science.* 2004;305:354–60.
- Finkel ZV, Beardall J, Flynn KJ, Quigg A, Rees TAV, Raven JA. Phytoplankton in a changing world: cell size and elemental stoichiometry. *J Plankton Res.* 2010;32:119–37.
- Geider RJ, et al. Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. *Glob Change Biol.* 2001;7:849–82.
- Katz ME, Finkel ZV, Grzebyk D, Knoll AH, Falkowski PG. Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. *Annu Rev Ecol Evol Syst.* 2004;35:523–56.
- Litchman E, Klausmier CA. Trait-based community ecology of phytoplankton. *Annu Rev Ecol Evol Syst.* 2008;39:615–39.
- MacIntyre HL, Kana TM, Geider RJ. The effects of water motion on short term rates of photosynthesis of marine phytoplankton. *Trends Plant Sci.* 2000;5:12–7.
- Moore CM, et al. Processes and patterns of oceanic nutrient limitation. *Nat Geosci.* 2013;6:701–10.
- Rabalais NN, Turner RE, Justic D, Diaz RJ. Global change and eutrophication of coastal waters. *ICES J Mar Sci.* 2009;66:1528–37.
- Riebesell U, Tortell PD. Effects of ocean acidification on pelagic organisms and ecosystems. In: Gattuso JP, Hansson L, editors. *Ocean acidification*. Oxford: Oxford University Press; 2011. p. 99–121.
- Thomas MK, Kremer CT, Klausmier CA, Litchman E. A global pattern of thermal adaptation in marine phytoplankton. *Science.* 2012;338:1085–8.

Andrew D. B. Leakey

## Contents

Introduction .....	534
Global Environmental Change from the Industrial Revolution to Today .....	535
Greenhouse Effect .....	535
Greenhouse Gas Emissions .....	537
Temperature and Precipitation .....	539
Forecasts of Global Environmental Change in the Twenty-First Century .....	540
Plants as Pivot Points in the Global Carbon Cycle .....	541
Plants and Ecosystem Services .....	542
Plant Responses to Elevated Carbon Dioxide (CO <sub>2</sub> ) .....	543
Introduction .....	543
Photosynthetic Responses of C <sub>3</sub> Species to Growth at Elevated [CO <sub>2</sub> ] .....	544
Respiration Responses of C <sub>3</sub> Plants to Elevated [CO <sub>2</sub> ] .....	547
Stomatal Conductance and Water Relations of C <sub>3</sub> Plants Under Elevated [CO <sub>2</sub> ] .....	547
Biomass and Seed Responses of C <sub>3</sub> Plants to Elevated [CO <sub>2</sub> ] .....	548
Elevated [CO <sub>2</sub> ] and Water Use Efficiency .....	550
Physiological Responses of C <sub>4</sub> Species to Growth at Elevated [CO <sub>2</sub> ] .....	550
Plant Responses to Temperature .....	551
Introduction .....	551
Photosynthetic and Respiratory Responses to High Temperature .....	551
Cellular Responses to High Temperature .....	554
Crop Reproductive and Yield Responses to High Temperature .....	556
Carbon Cycling Responses to High Temperature .....	557
Plant Responses to Drought .....	558
Introduction .....	558
Stomatal, Photosynthetic, and Respiratory Responses to Drought .....	559
Plant Dehydration, Osmotic Adjustment, and Hydraulic Failure .....	560
Whole-Plant Physiological Plasticity and Adaptations to Drought .....	561
Crop Yield and NPP Responses to Drought .....	562

---

A.D.B. Leakey (✉)

Department of Plant Biology and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

e-mail: [leakey@illinois.edu](mailto:leakey@illinois.edu)

Plant Responses to Ground-Level Ozone (O <sub>3</sub> ) .....	562
Introduction .....	562
Physiological Responses to Elevated [O <sub>3</sub> ] .....	563
Biomass and Seed Responses to Elevated [O <sub>3</sub> ] .....	565
Adaptation of Plants to Environmental Change .....	566
Plant-Based Mitigation of Environmental Change .....	567
Future Directions .....	570
References .....	571

## Abstract

- The Anthropocene is the period of Earth history since the Industrial Revolution and is defined by the impact of mankind on the environment.
- Greenhouse gas concentrations, temperatures, precipitation, and atmospheric pollutants have changed significantly from 1750 to today.
- Climatic and environmental change will accelerate in the twenty-first century.
- Plants act as pivot points in global biogeochemical cycles.
- Plants provide many important ecosystem services including food production.
- Elevated carbon dioxide (CO<sub>2</sub>) concentration enhances plant productivity.
- Rising temperature stimulates plant productivity at high latitudes but impairs plant productivity at many temperate and tropical latitudes.
- Greater drought impairs plant productivity.
- Elevated ozone (O<sub>3</sub>) concentration impairs plant productivity.
- Crop plants can be adapted to future environmental change.
- Future environmental change can be mitigated by appropriate management of plants in agricultural and natural ecosystems.

## Introduction

The term Anthropocene emerged recently to describe the period of time during which mankind has significantly impacted the function of the Earth system, i.e., biosphere, atmosphere, geosphere, ocean, and cryosphere. The use of the term is intended to reflect the fact that since the start of the Industrial Revolution (c. 1750–1850), humans have caused global environmental change comparable with events that demark past geological epochs (e.g., the Holocene as the period of ~12,000 years since the last ice age). The word Anthropocene has Greek roots, with *anthropo-* meaning human and *-cene* meaning “new.” Human-induced changes in the Earth system are occurring today at an accelerating pace and are anticipated to continue for the foreseeable future. The resulting impacts on climate as well as ecosystem goods and services are a growing challenge to human well-being. Secretary General of the United Nations, Ban Ki Moon, in 2007 described climate change as “the defining challenge of our age.” Recognition of this fact is a key driver of efforts to achieve sustainable development, i.e., where current resource use meets human needs while also preserving the environment to insure these needs can be met for future generations.

Plants mediate many key interactions between global environmental change and humans. Plants play key roles in global biogeochemical cycles. For example, the removal of carbon dioxide (CO<sub>2</sub>) from the atmosphere by plants via the process of photosynthesis modifies greenhouse gas concentration in the atmosphere and the greenhouse effect. In addition, plants play key roles in delivering the ecosystem goods and services of food, fuel, fiber, forage, clean air, and clean water. Therefore, any impacts of global environmental change on plants in natural or agricultural ecosystems will influence human well-being.

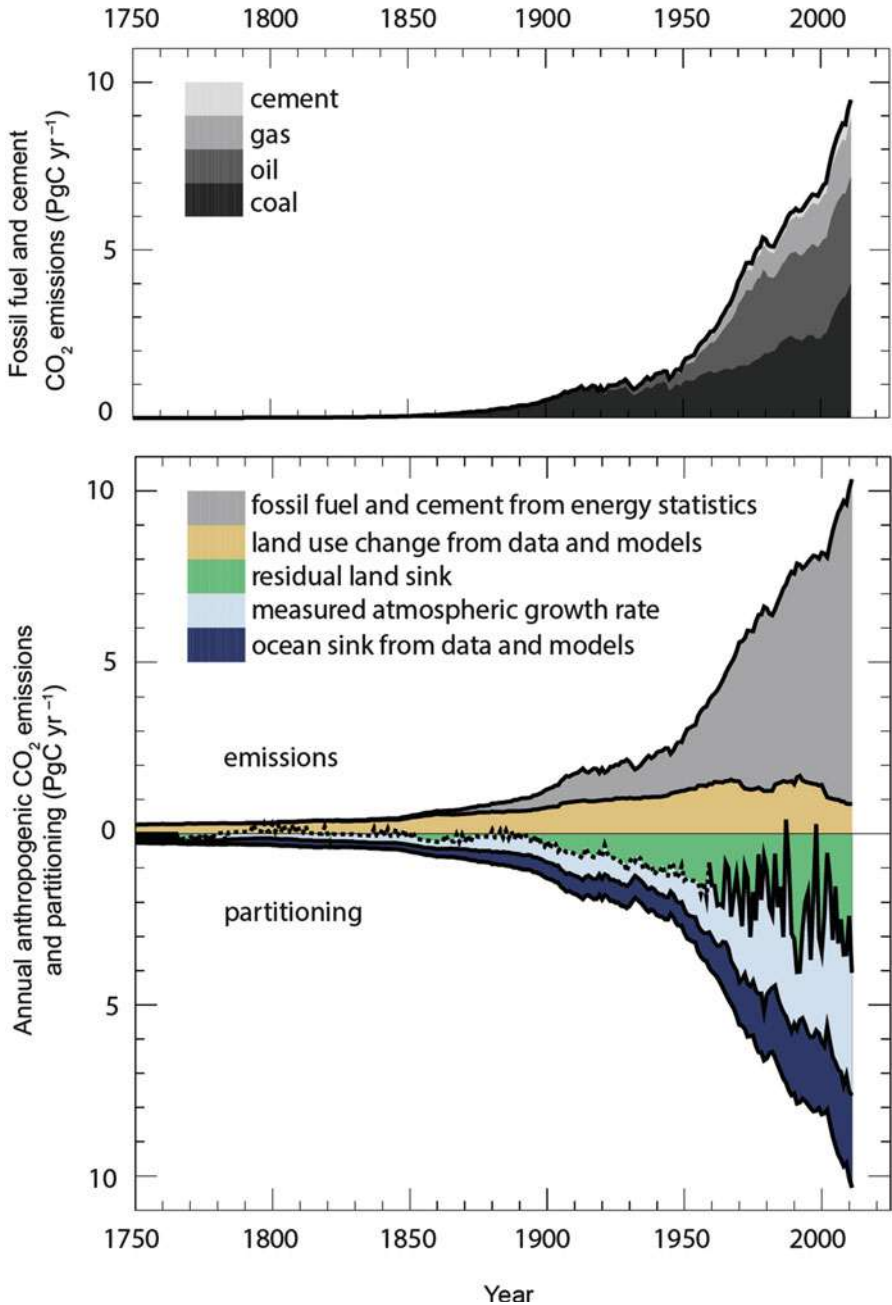
This chapter takes a plant-centric view of the Anthropocene and aims to explain (1) past and future global environmental change in the Anthropocene, (2) the role of plants in global biogeochemical cycles and food security, (3) plant responses to major elements of global environmental change (elevated CO<sub>2</sub>, temperature, drought, elevated ground-level ozone), (4) adaptation of crop plants to global environmental change, and (5) plant-based mitigation of global environmental change.

---

## **Global Environmental Change from the Industrial Revolution to Today**

### **Greenhouse Effect**

Greenhouse gases are characterized by their ability to absorb and emit infrared (thermal) radiation. This property plays a key role in maintaining conditions on Earth that are favorable for life. Short-wave radiation from the sun is not significantly absorbed by greenhouse gases as it passes through the atmosphere and is absorbed by the Earth's surface. The Earth's surface is warmed by the solar radiation and in turn emits longer wavelength infrared radiation. Some of this infrared radiation is absorbed by greenhouse gases in the atmosphere and reradiated back to the Earth's surface. This "greenhouse effect" acts to trap heat at the Earth's surface and prevents the extreme fluctuation in day/night temperatures observed on other planets without greenhouse atmospheres. The most important naturally occurring greenhouse gases are water, CO<sub>2</sub>, methane, nitrous oxide and tropospheric ozone. The increasing concentration of anthropogenic greenhouse gases (CO<sub>2</sub>, methane, nitrous oxide, ozone, halocarbons) in the atmosphere since the Industrial Revolution has strengthened this effect, causing warming. The additional trapping of solar energy is termed radiative forcing and since the Industrial Revolution is estimated to have risen in total to  $\sim 2.6 \text{ W m}^{-2}$  (i.e., 2.6 units more energy absorbed per second per square meter of Earth's surface). The contribution of individual greenhouse gases to this total varies with their ability to absorb infrared radiation, their increases in concentration over time, and their lifetime in the atmosphere. To date, rising CO<sub>2</sub> concentrations contribute more radiative forcing than the other anthropogenic greenhouse gases combined.



**Fig. 1** Reproduced with permission from Ciais et al. (2013): Annual anthropogenic CO<sub>2</sub> emissions and their partitioning among the atmosphere, land and ocean (PgC yr<sup>-1</sup>) from 1750 to 2011. (Top) Fossil fuel and cement CO<sub>2</sub> emissions by category, estimated by the Carbon Dioxide Information Analysis Center (CDIAC) based on UN energy statistics for fossil fuel combustion

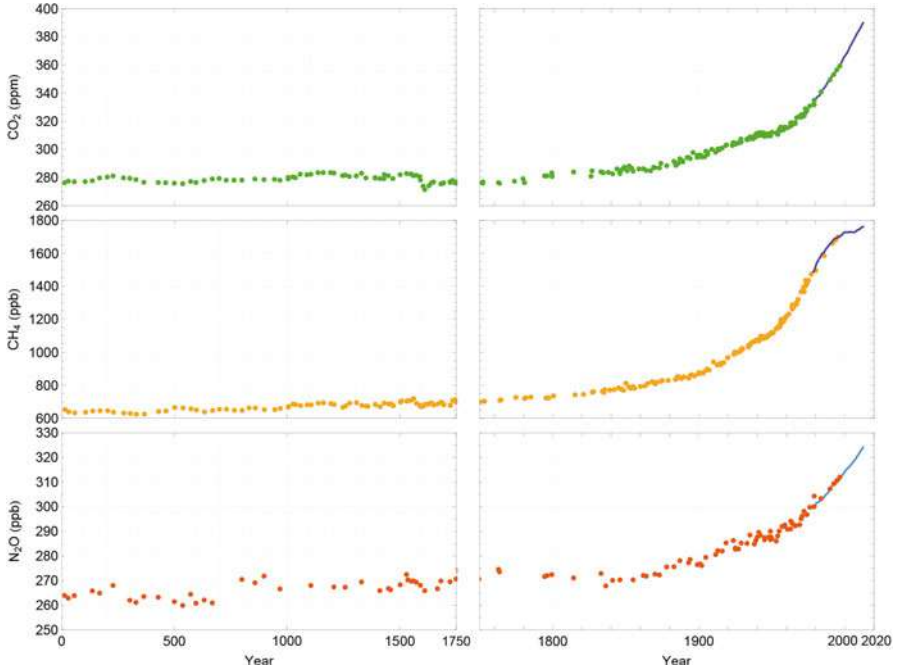
## Greenhouse Gas Emissions

The dominant driver of global environmental change has been, and continues to be, anthropogenic (human produced) greenhouse gas emissions (Ciais et al. 2013). The five major anthropogenic greenhouse gases are CO<sub>2</sub>, methane, nitrous oxide, ground-level ozone (O<sub>3</sub>), and halocarbons. These gases are produced as a result of fossil fuel burning, farming activities, and/or industrial processes. The Industrial Revolution occurred when fossil fuel (coal) was first used to power mechanized industry and transport. This initiated a period of sustained and rapid growth in living standards, the global economy, technological innovation, human population, and natural resource use – all of which accelerated greenhouse gas emissions.

Coal burning dominated CO<sub>2</sub> emissions during the 1800s and still contributes approximately one third of CO<sub>2</sub> emissions today (Fig. 1). Burning of petroleum products and natural gas became the next most important sources of CO<sub>2</sub> emissions starting in the early- and mid-1900s, respectively. The mid-1900s also saw the start of significant CO<sub>2</sub> emissions from cement production. Deforestation has been a feature of human activity for millennia, but has increased exponentially as a source of CO<sub>2</sub> emissions since the Industrial Revolution. In 2008, approximately 85 % of CO<sub>2</sub> emissions resulted from fossil fuel burning and cement production, with the remaining 15 % resulting from deforestation, particularly in the tropics. The consequence of these CO<sub>2</sub> emissions has been an increase in atmospheric CO<sub>2</sub> concentration from ~280 parts per million (ppm) in 1800 to greater than 400 ppm today (Fig. 2). This represents a very significant perturbation of the Earth system because today's CO<sub>2</sub> concentration is greater than it has been at any point in the last 20 million years. In addition to being a greenhouse gas, CO<sub>2</sub> is combined with water



**Fig. 1** (continued) and US Geological Survey for cement production (Boden et al. 2011). (*Bottom*) Fossil fuel and cement CO<sub>2</sub> emissions as above. CO<sub>2</sub> emissions from net land-use change, mainly deforestation, are based on land cover change data and estimated for 1750–1850 from the average of four models (Pongratz et al. 2009; Shevliakova et al. 2009; van Minnen et al. 2009; Zaehle et al. 2011) before 1850 and from Houghton et al. (2012) after 1850 (see Table 6.2). The atmospheric CO<sub>2</sub> growth rate (term in light blue “atmosphere from measurements” in the figure) prior to 1959 is based on a spline fit to ice core observations (Neftel et al. 1982; Friedli et al. 1986; Etheridge et al. 1996) and a synthesis of atmospheric measurements from 1959 (Ballantyne et al. 2012). The fit to ice core observations does not capture the large interannual variability in atmospheric CO<sub>2</sub> and is represented with a dashed line. The ocean CO<sub>2</sub> sink prior to 1959 (term in dark blue “ocean from indirect observations and models” in the figure) is from Khatiwala et al. (2009) and from a combination of models and observations from 1959 from Le Quéré et al. (2013). The residual land sink (term in green in the figure) is computed from the residual of the other terms and represents the sink of anthropogenic CO<sub>2</sub> in natural land ecosystems. The emissions and their partitioning only include the fluxes that have changed since 1750 and not the natural CO<sub>2</sub> fluxes (e.g., atmospheric CO<sub>2</sub> uptake from weathering, outgassing of CO<sub>2</sub> from lakes and rivers, and outgassing of CO<sub>2</sub> by the ocean from carbon delivered by rivers; see Figure 6.1) between the atmosphere, land, and ocean reservoirs that existed before that time and still exist today. The uncertainties in the various terms are discussed in the text and reported in Table 6.1 for decadal mean values



**Fig. 2** Reproduced with permission from Ciais et al. (2013): Atmospheric CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O concentrations history over the industrial era (*right*) and from year 0 to the year 1750 (*left*), determined from air enclosed in ice cores and firn air (*color symbols*) and from direct atmospheric measurements (*blue lines*, measurements from the Cape Grim observatory) (MacFarling-Meure et al. 2006)

to make carbohydrate through the process of photosynthesis. It is therefore vital to plant life, as well as the communities of animals and microbes that consume plants (alive or dead), thereby forming almost all ecosystems on Earth.

Anthropogenic methane emissions escape during the mining and processing of fossil fuels as well as come from microbial activity associated with ruminant livestock (e.g., cows), paddy rice farming, biomass burning, and landfill trash deposits. While it is difficult to attribute emissions to specific sources, methane concentrations in the atmosphere have risen from ~700 parts per billion (ppb) prior to the Industrial Revolution to greater than 1,750 ppb today (Fig. 2).

Approximately two thirds of anthropogenic emissions of nitrous oxide result from breakdown of fertilizers applied to crops. Additional sources include breakdown products of livestock excreta, combustion of transportation fuels, and industrial processes. Global fertilizer use has increased more than tenfold since the early 1900s. Consequently, atmospheric concentrations of nitrous oxide have increased from ~270 to ~320 ppb since the Industrial Revolution (Fig. 2).

Ground-level (or tropospheric) O<sub>3</sub> is a secondary pollutant and greenhouse gas produced by photochemical reactions of methane, volatile organic carbon molecules, and nitrogen oxides. O<sub>3</sub> is produced only during daylight and more rapidly at



higher temperatures, but is also highly reactive and degrades quickly. Therefore, O<sub>3</sub> concentrations are highly variable in both time and space. Rising anthropogenic emissions of methane and nitrogen oxides have caused ground-level O<sub>3</sub> concentrations to increase from a preindustrial concentration of ~10 to ~40 ppb during summer days in many parts of the world. In addition to being a greenhouse gas, O<sub>3</sub> is toxic to all life forms and significantly reduces the physiological performance and productivity of all plants.

Unlike the other greenhouse gases described above, chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) do not occur naturally and are solely the product of industrial processes. They were used extensively during the 1900s for a wide array of applications, including as refrigerants and aerosol propellants. However, CFCs and HCFCs were discovered to cause degradation of the high-level (stratospheric) O<sub>3</sub> layer that is responsible for absorbing harmful ultraviolet rays from the sun. Consequently, a global ban on production of CFCs and HCFCs has taken effect and substantially reduced emissions. As a consequence, concentrations of CFCs and HCFCs peaked in the early 1990s, but the decline in concentrations is slow due to the very long-lived nature of the molecules.

In summary, with the exception of CFCs, plants play important roles in the emission of greenhouse gases and/or are directly influenced by changes in greenhouse gases that are toxic or important resources. In addition, greenhouse gases indirectly impact plants by altering the climate.

## Temperature and Precipitation

As a consequence of radiative forcing by anthropogenic greenhouse gases, the global average surface air temperature increased 0.8 °C from 1850 to 2000. This increase in air temperature has resulted in warming of oceans to depths of up to 3,000 m and significant melting of snow and ice in ice caps and glaciers. These changes have combined to drive a rise in sea level of 200 mm since the 1800s. In addition to rising average temperatures, the last three to four decades have been characterized in many parts of the world by (1) warmer and fewer cold days and nights, (2) warmer and more frequent hot days and nights, (3) more frequent heat waves, and (4) stronger and more frequent droughts.

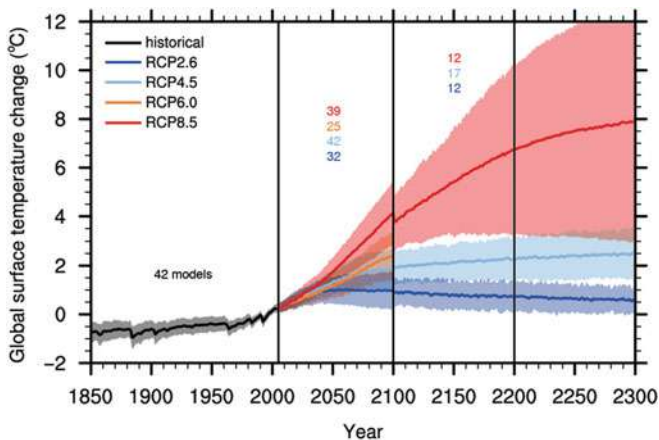
There is also evidence that global warming has started to intensify the water cycle since the 1970s. This has taken the form of more frequent heavy precipitation events, separated by longer periods of drought stress. In certain regions this has been accompanied by more intense tropical cyclone activity. However, precipitation patterns are inherently variable and more poorly understood than temperature changes, so confidence that changes in precipitation have been caused by human activity is only moderate while confidence that rising temperatures have been driven by human activity is high. Key evidence comes from analysis showing that models simulating only “natural” drivers of climate variation fail to match the rising temperatures measured around the world between 1970 and 2000, while models that simulate both “natural” and “human”

(i.e., greenhouse gas emissions) drivers of climate change correspond well with measured data.

---

## Forecasts of Global Environmental Change in the Twenty-First Century

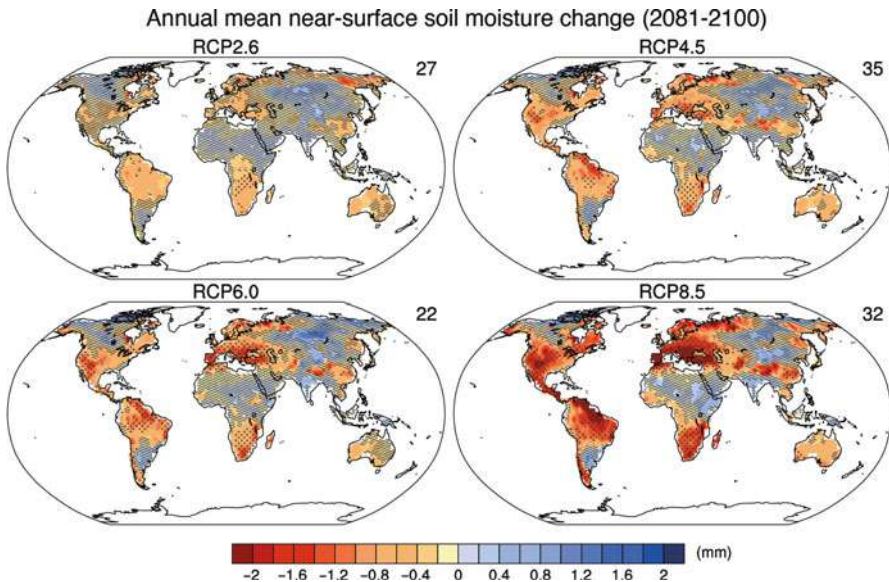
Forecasts of atmospheric greenhouse gas concentrations and climate change over the twenty-first century have been driven by a set of greenhouse gas emission scenarios generated by the Intergovernmental Panel on Climate Change and relating to different socioeconomic conditions (Ciais et al. 2013; Collins et al. 2013). Scenario A1 assumes a world of rapid economic growth and rapid introduction of new and more efficient technologies. Scenario A2 assumes a very heterogeneous world with an emphasis on family values and local traditions. Scenario B1 assumes a world of dematerialization and introduction of clean technologies. Scenario B2 assumes a world with an emphasis on local solutions to economic and environmental sustainability. These scenarios all predict that CO<sub>2</sub> emissions will increase in the early decades of the twenty-first century, after which they would follow various trajectories until by 2100 they would range from lower than 2000 (~7 Gt of C; scenario B1) to roughly double that of 2000 (~13 Gt of C; scenarios A1 and B2) or even four times that of 2000 (~28 Gt of C; scenario A2). Different CO<sub>2</sub> emission scenarios result in an increase of atmospheric CO<sub>2</sub> concentrations to ~450 to 550 ppm in 2050 and ~500 to 950 ppm in 2100. This in turn leads to a range of possible radiative forcing, from which a number of scenarios have been selected (RCP2.6, RCP4.5, RCP6.0, RCP8.5) to forecast an increase in global average surface temperature from 1990 to 2090 of between 1.8 °C and 4.0 °C (Fig. 3). Notably, even if CO<sub>2</sub> concentrations remained constant from 2000 to 2100, there would still be surface warming of 0.6 °C due to slow heat exchange of the oceans. However, warming will not be uniform over the globe. Warming is predicted to be greater with increasing latitude and greater over the land than the seas. As a consequence, snow and ice cover are expected to contract and thaw depth will increase over permafrost regions where soil is currently frozen for all or part of the year. Sea ice is predicted to contract at both poles and in some scenarios complete melt of Arctic sea ice occurs in late summer. It is very likely that heat waves and heavy precipitation events will continue to become more frequent. Annual precipitation is predicted to increase at latitudes where it is currently wetter (i.e., tropical, temperate, and upper latitudes) and decrease at latitudes where it is currently drier (subtropics). In general models agree that precipitation will be more variable, leading to greater droughts and flooding, but there is less confidence in model predictions of future precipitation than future temperatures for specific regions, decades, or seasons. The combination of greater temperatures with more variable precipitation is expected to lead to drier soils, especially in mid-continental regions (Fig. 4). Increases in emission of methane and nitrous oxides are predicted to increase ground-level ozone concentrations around the world, but particularly in Asia and the Middle East.



**Fig. 3** Reproduced with permission from Collins et al. (2013): Time series of global annual mean surface air temperature anomalies (relative to 1986–2005) from CMIP5 concentration-driven experiments. Projections are shown for each RCP for the multi-model mean (*solid lines*) and the 5 % to 95 % range ( $\pm 1.64$  standard deviation) across the distribution of individual models (*shading*). Discontinuities at 2100 are due to different numbers of models performing the extension runs beyond the twenty-first century and have no physical meaning. Only one ensemble member is used from each model and numbers in the figure indicate the number of different models contributing to the different time periods. No ranges are given for the RCP6.0 projections beyond 2100 as only two models are available

## Plants as Pivot Points in the Global Carbon Cycle

The atmosphere stores approximately 800 GtC (gigatonnes of carbon; or 800,000,000,000 t), primarily in the form of CO<sub>2</sub>. More than a quarter of this CO<sub>2</sub> pool is absorbed each year by photosynthesis performed by plants. Plants in the global carbon cycle on the land (120 GtC) and in the ocean (90 GtC). Photosynthesis uses solar energy to assimilate CO<sub>2</sub> and water into sugars, which are ultimately converted into plant biomass. Terrestrial plant biomass (550 GtC) stores almost as much carbon as the atmospheric CO<sub>2</sub> pool. On land, biomass that has been incorporated into soil forms a relatively large pool (2,300 GtC). In the oceans, after phytoplankton die they sink transporting organic carbon to deeper layers that is then preserved in sediments or decomposed into a very large pool of dissolved inorganic carbon (37,000 GtC). Plants, animals, and microbes all release CO<sub>2</sub> to the atmosphere as a by-product of generating energy and synthesizing biomass through the process of respiration. The natural carbon cycle is in equilibrium on both land and in the oceans. Plant and microbial respiration release approximately the same amount of CO<sub>2</sub> as is removed from the atmosphere through photosynthesis. As a result of the large fluxes and pools of carbon attributable to plants, they play a key role in regulating the global carbon cycle and, therefore, atmospheric CO<sub>2</sub> concentration and climate. For example, approximately ~9 GtC was released to the



**Fig. 4** Reproduced with permission from Collins et al. (2013): Change in annual mean soil moisture (mass of water in all phases in the uppermost 10 cm of the soil) (mm) relative to the reference period 1986–2005 projected for 2081–2100 from the CMIP5 ensemble. Hatching indicates regions where the multi-model mean change is less than one standard deviation of internal variability. Stippling indicates regions where the multi-model mean change is greater than two standard deviations of internal variability and where at least 90 % of models agree on the sign of change (see Box 12.1). The number of CMIP5 models used is indicated in the *upper right* corner of each panel

atmosphere by human fossil fuel burning and land-use change in 2009. While ~45 % of the CO<sub>2</sub> emissions stayed in the atmosphere, ~30 % was absorbed by land plants and ~25 % was absorbed by the oceans (Fig. 1; Ciais et al. 2013). These land and ocean sinks for CO<sub>2</sub> have significantly slowed the rate at which atmospheric CO<sub>2</sub> is rising and the climate is warming. However, the proportion of CO<sub>2</sub> emissions absorbed by photosynthesis and stored on land or at sea is declining. Determining how future global environmental change will alter the performance of plants and the control they exert on the global carbon cycle is therefore a scientific priority. While these processes cannot be actively managed in natural ecosystems, greater production of biofuels has the potential to increase carbon sequestration while reducing fossil fuel use.

## Plants and Ecosystem Services

Ecosystem services Ecosystem services Plants and ecosystem services are critical to human well-being and are classified into four major categories. (1) Supporting ecosystem services are the biogeochemical cycles as well as biological and physical





processes that drive ecosystem function. As described above, plants play a critical role in the global carbon cycle. They also influence water cycling by acting as a conduit for water to move from the soil to atmosphere through the process of transpiration. Variation in plant cover or function that alters transpiration can influence precipitation. For example, deforestation of the Amazon forest leads to reduced transpiration, which in turn reduces convective rainfall and can intensify drought. Plants also play important roles in global nutrient cycles; most prominently by interacting with microbes to perform nitrogen fixation. Through which 200 Mt of atmospheric nitrogen gas is converted each year across the globe into chemical forms in the soil and ocean that are accessible to other organisms for uptake. (2) Provisioning ecosystem services are actively harvested by us to meet demand for natural resources including food, water, timber, and fiber. Approximately 1/8th of the plant biomass produced on the plant each year is harvested for these purposes. Approximately 75 % of all calories consumed by humans come directly or indirectly (via animal feed) from the four major crops of maize, wheat, rice, and soybeans. (3) Regulating ecosystem services are processes in the Earth system that control key physical and biological elements of our environment, e.g., climate regulation, flood regulation, disease regulation, and water purification. As plants are the primary producers of all terrestrial ecosystems – i.e., they synthesize the carbon sources all animals and microbes subsequently use as energy sources – they are key to ecosystem stability and maintenance of regulating services. (4) Cultural ecosystem services reflect the aesthetic and spiritual values placed on nature as well as the educational and recreational activities dependent on ecosystems. Plants contribute to cultural ecosystem services as a result of mankind's emotional response to time spent in a forest or a beautiful garden. Overall, plants strongly influence human well-being through the services associated with both pristine, natural ecosystems (e.g., tropical rain forests or arctic tundra), and highly managed ecosystems (e.g., crop fields or urban landscapes). Consequently, the response of plants to the elements of global environmental change in the Anthropocene (Fig. 5) has and will continue to play a key role in determining the ultimate impacts on human well-being.

---

## Plant Responses to Elevated Carbon Dioxide (CO<sub>2</sub>)

### Introduction

In the majority of terrestrial plants species, the rates of photosynthetic CO<sub>2</sub> fixation ( $A$ ) and stomatal conductance ( $g_s$ ) are sensitive to changes in [CO<sub>2</sub>] that have occurred since the Industrial Revolution and are continuing today (Norby et al. 2005; Ainsworth and Rogers 2007; Leakey et al. 2009; Norby and Zak 2011; Leakey and Lau 2012). The effects of increasing [CO<sub>2</sub>] on  $A$  and  $g_s$  initiate a set of cellular and physiological responses, which typically increase biomass production and reproductive output while reducing water use and altering nutrient dynamics. Variation in sensitivity to [CO<sub>2</sub>] among genotypes, populations, species,

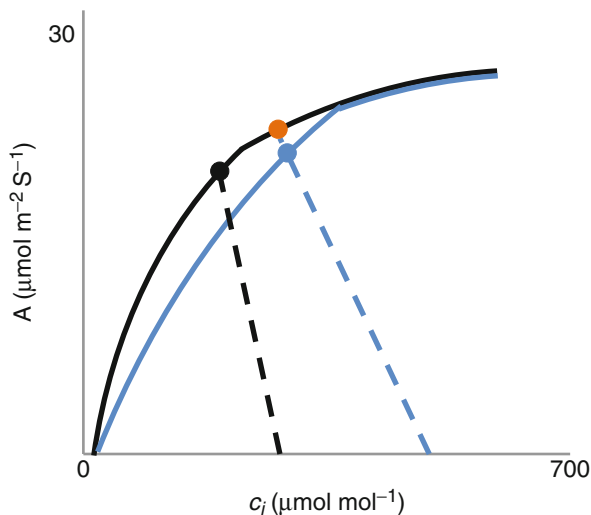
	Elevated CO <sub>2</sub>	Warming at warm locations	Warming at cool locations	Drought	Ozone	
Ecosystem 	↑	↓	↑	↓	↓	NPP Shifts in composition of species or genotypes
Whole plant 	↑↑ ↑↑ ↑/↓ -/-↓	↓↓ ↓↓ ↓↓ ↑	↑↑ ↑↑ ↑↑ ↓	↓↓ ↓↓ ↓↓ ↑	↓↓ ↓↓ ↓↓ ↑	Biomass Leaf Area Index Seed production Defense Senescence
Leaf 	↑↑ ↑↑ -/-↓ -/-↑	↓↓ ↑↑ ↓↓ ↓	↑↑ ↑↑ ↑↑ ↑	↓↓ ↓↓ -/-↓ ↓	↓↓ ↑↑ -/-↓ -/-↓	Photosynthesis Carbohydrates Respiration Nutrient status Water status
Cell 		↓↑ ↑	↑↓ ↓	↓↑ ↓	-/-↓ ↑ -/-↓	Enzyme stability ROS Membrane stability

**Fig. 5** Effects of global environmental change factors on plant processes at the ecosystem, whole-plant, leaf, and cellular scales. *Arrows* indicate direction of response (Modified from Ainsworth et al. (2012))

and functional groups creates the possibility of important ecological and evolutionary consequences over a wide range of spatial and temporal scales. There are three major photosynthetic types in higher plants: C<sub>3</sub>, C<sub>4</sub>, and CAM. C<sub>3</sub> plants are the most common and most sensitive to future, elevated [CO<sub>2</sub>]. C<sub>4</sub> plants are less common and less sensitive to elevated [CO<sub>2</sub>] but include some of the world’s most important crops and weeds, as well as the grass species that dominate the world’s tropical savannas. The response of CAM plants to elevated [CO<sub>2</sub>] has not been studied extensively, in large part because they appear to be largely insensitive to elevated [CO<sub>2</sub>]. Consequently, this section of the chapter focuses on the effects of elevated [CO<sub>2</sub>] on C<sub>3</sub> plants and then C<sub>4</sub> plants.

### Photosynthetic Responses of C<sub>3</sub> Species to Growth at Elevated [CO<sub>2</sub>]

The enzyme ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) catalyzes the initial reaction that “captures” CO<sub>2</sub> from the atmosphere and combines it with ribulose-1,5-bisphosphate (RuBP) to form a sugar in C<sub>3</sub> plants. Today’s atmospheric [CO<sub>2</sub>] limits the rate of this carboxylation reaction and the overall rate of photosynthetic CO<sub>2</sub> uptake (*A*). The impact on *A* of varying [CO<sub>2</sub>] is typically visualized as a photosynthetic CO<sub>2</sub> response (*A/c<sub>i</sub>*) curve (Fig. 6). The slope of the steep, initial portion of the *A/c<sub>i</sub>* curve provides a measure of the maximum carboxylation capacity of RuBisCO (*V<sub>cmax</sub>*). The asymptote of the curve, where



**Fig. 6** The response of leaf  $\text{CO}_2$  uptake ( $A$ ) to intercellular  $[\text{CO}_2]$  ( $c_i$ ). Curves representing the photosynthetic capacity of plants grown at ambient  $[\text{CO}_2]$  (*black solid line*) and plants that have undergone photosynthetic acclimation to long-term growth at elevated  $[\text{CO}_2]$  (*blue solid line*) are shown. The instantaneous stimulation of  $A$  when an ambient  $[\text{CO}_2]$ -grown leaf experiences greater internal  $\text{CO}_2$  supply after a shift from ambient  $[\text{CO}_2]$  to elevated  $[\text{CO}_2]$  is represented by the black and orange dots. The stimulation of  $A$  when plants are grown long term in ambient  $[\text{CO}_2]$  or elevated  $[\text{CO}_2]$  is represented by the *black* and *blue* dots. *Dashed lines* represent the decline in  $[\text{CO}_2]$  from outside to inside the leaf (supply function) associated with resistance to diffusion through stomata

$A$  approaches saturation by  $\text{CO}_2$  supply, provides a measure of the maximum capacity of photosynthetic electron transport to generate RuBp in the Benson-Calvin cycle ( $J_{\text{max}}$ ). The initial slope of the  $A/c_i$  curve is steep because any increase in  $[\text{CO}_2]$  causes a direct stimulation of  $A$  for two reasons. First, greater substrate ( $\text{CO}_2$ ) availability stimulates the rate of the RuBisCO carboxylation reaction. Second, elevated  $[\text{CO}_2]$  competitively inhibits RuBisCO from performing an oxygenase reaction that otherwise leads to photorespiration and reductions in photosynthetic efficiency. In other words, RuBisCO assimilates more  $\text{CO}_2$  and produces more photoassimilate (sugars) at elevated  $[\text{CO}_2]$  as a result of performing more carboxylation reactions and fewer oxygenation reactions. However, on the upper portion of the  $A/c_i$  curve,  $c_i$  is sufficiently great that the carboxylation reaction of RuBisCO is no longer limited by  $\text{CO}_2$  availability. Consequently, further increases in  $c_i$  only stimulate photosynthesis by competitively inhibiting the oxygenation reaction and photorespiration, with the result that the curve is less steep. The increase in  $A$  caused by suppression of photorespiration requires no additional light, water, or nitrogen, making photosynthesis more efficient with respect to all of these resources.

As temperature increases, the rate of photorespiration increases because RuBisCO tends to perform fewer carboxylations and more oxygenations. As a

consequence, the competitive inhibition of the oxygenation reaction by elevated  $[\text{CO}_2]$  stimulates  $A$  to a larger degree as temperature rises. In addition, elevated  $[\text{CO}_2]$  also causes increases in the temperature at which  $A$  reaches its maximum, i.e., optimum temperature and the greatest temperature at which positive rates of  $A$  can be maintained. These changes in the constraint of photosynthesis by temperature at elevated  $[\text{CO}_2]$  highlight the significance of considering elevated  $[\text{CO}_2]$  as a factor that modifies physiological responses to other abiotic variables and not just a factor that stimulates  $A$ .

The direct stimulation of photosynthesis by elevated  $[\text{CO}_2]$  described above occurs within seconds to minutes of leaves experiencing greater  $[\text{CO}_2]$ . When plants are grown for extended periods of weeks, months, or years under elevated  $[\text{CO}_2]$ , the phenomena of photosynthetic acclimation to elevated  $[\text{CO}_2]$  is also often observed. Acclimation is defined as the phenomenon whereby living organisms adjust to the present environmental conditions and in doing so enhance their probability of survival. These adjustments are on the timescales of less than one generation and may involve changes in physiological processes and structure. Adaptation is distinct from acclimation, meaning a genetic change that better fits the individual to the new environmental conditions. When plants acclimate to elevated  $[\text{CO}_2]$ ,  $V_{\text{cmax}}$  is lower than at ambient  $[\text{CO}_2]$  (Fig. 6). This decrease in biochemical capacity has been associated with reduced expression of genes encoding RuBisCO and other proteins of the photosynthetic apparatus. One consequence is that the stimulation of  $A$  is less after photosynthetic acclimation than would be predicted from measuring short-term instantaneous responses of  $A$  to variation in  $[\text{CO}_2]$ . Nonetheless,  $A$  of a plant measured at elevated  $[\text{CO}_2]$  after photosynthetic acclimation is still typically greater than  $A$  of a plant grown and measured at ambient  $[\text{CO}_2]$ . This dampened response of  $A$  to elevated  $[\text{CO}_2]$  is considered a metabolic optimization response due to the interaction between carbon and nitrogen metabolism in plants. For greater photoassimilate production at elevated  $[\text{CO}_2]$  to be translated into greater biomass production, all the nutrients required to build biomass must be available in sufficient quantities to match the additional carbon available and maintain appropriate tissue composition. In many soils, especially in temperate latitudes, availability of nitrogen is limiting to plant growth. Stimulation of  $A$  when plants are grown at elevated  $[\text{CO}_2]$  frequently exacerbates this nitrogen limitation. As a consequence, tissue protein and nitrogen concentrations are often lower at elevated  $[\text{CO}_2]$ . And, carbohydrates such as starch and sugars accumulate in leaves and other tissues at elevated  $[\text{CO}_2]$ . There is evidence to suggest that the plant triggers the photosynthetic acclimation response after sensing this accumulation of sugars. The subsequent reduction in the synthesis of RuBisCO during photosynthetic acclimation to elevated  $[\text{CO}_2]$  can relieve the nitrogen limitation because it makes available some of the approximately ~25 % of leaf nitrogen that is allocated to synthesis of RuBisCO. The lower the availability of nitrogen to a plant, the more photosynthetic acclimation to elevated  $[\text{CO}_2]$  is observed and the less photosynthesis is stimulated relative to ambient  $[\text{CO}_2]$ . Under extreme nitrogen deprivation the stimulation of  $A$  by elevated  $[\text{CO}_2]$  can be eliminated completely. Other conditions that restrict the capacity of



carbon sinks relative to the supply of photoassimilate also promote photosynthetic acclimation to elevated  $[\text{CO}_2]$ . For example, soybean varieties with limited seed production show greater photosynthetic acclimation to elevated  $[\text{CO}_2]$  than varieties capable of filling extra seeds at elevated  $[\text{CO}_2]$ . Or, when plants are grown in small pots that constrain their root growth, photosynthetic acclimation is also exaggerated. There is also variation among plant functional groups in this response. Trees often show a greater response than herbaceous species. Legumes are capable of nitrogen fixation through their symbiosis with Rhizobia bacteria and so manage to balance greater carbon gain with greater nitrogen assimilation at elevated  $[\text{CO}_2]$ . Therefore, photosynthetic acclimation to elevated  $[\text{CO}_2]$  occurs weakly if at all in legumes.

### **Respiration Responses of $\text{C}_3$ Plants to Elevated $[\text{CO}_2]$**

Plant dark respiration is of fundamental importance at cellular, physiological, and biogeochemical scales. At the cellular scale, respiration produces ATP, reducing power and carbon-skeleton intermediates while releasing  $\text{CO}_2$  as a by-product. At the physiological level, respiration supports maintenance and growth processes and is a key determinant of plant carbon balance. The nature of respiratory responses to elevated  $[\text{CO}_2]$  have been more controversial than that of photosynthesis or water relations. This is a significant source of uncertainty in projections of future crop and ecosystem function because 30–80 % of the carbon fixed by plants through photosynthesis can be rereleased through respiration. At the global scale, plant respiration releases five to six times as much  $\text{CO}_2$  into the atmosphere as human fossil fuel burning, so environmentally induced changes in plant respiration would feedback substantially on future rates of climate change. Various studies have concluded that leaf respiration either increases, decreases, or does not change at elevated  $[\text{CO}_2]$ . This has been explained to be a function of whether increases in photoassimilate substrate supply stimulate respiration more or less than decreases in leaf protein reduce demand for respiratory products. However, recent experiments including transcriptional data suggest that upregulated expression of respiratory genes occurs across a wide variety of herbaceous species at elevated  $[\text{CO}_2]$  and that this is associated with greater dark respiration rates in response to stimulated substrate supply, even when plant nitrogen status is low.

### **Stomatal Conductance and Water Relations of $\text{C}_3$ Plants Under Elevated $[\text{CO}_2]$**

The internal  $[\text{CO}_2]$  of the leaf ( $c_i$ ) is typically ~70 % of the atmospheric  $[\text{CO}_2]$  outside the leaf in  $\text{C}_3$  plants. For a given atmospheric  $[\text{CO}_2]$ , this “operating point” is connected by a straight-line supply function to the respective atmospheric  $[\text{CO}_2]$  on the x-axis (Fig. 6) in order to represent the ease with which  $\text{CO}_2$  can diffuse into

the leaf (stomatal conductance;  $g_s$ ). Variation of stomatal aperture provides plants with dynamic control of the trade-off between carbon gain and water use. Growth at elevated  $[\text{CO}_2]$  leads to lower  $g_s$  in almost all plants, with the exception of some conifers and beech species. This is a direct and rapid response that appears not to be modified by any acclimation of stomatal function after plants are grown at elevated  $[\text{CO}_2]$  for long periods of time.

The decrease in  $g_s$  at elevated  $[\text{CO}_2]$  acts to decrease transpiration per unit leaf area. This in turn can decrease canopy-scale transpiration and overall crop water use at elevated  $[\text{CO}_2]$  compared to ambient  $[\text{CO}_2]$ . However, the canopy-scale response is usually more modest than the leaf-scale response due to aerodynamic conductances between the leaf and the atmosphere and changes in leaf temperature that accompany changes in  $g_s$ . The relatively still air immediately next to a leaf (the leaf boundary layer) becomes more humid as the leaf transpires. This process occurring on many leaves collectively results in higher humidity within the plant canopy. This decreases the gradient in humidity from the inside to the outside of the leaf that drives transpiration. In dense, compact canopies where the air inside the canopy is rarely mixed with the bulk atmosphere, transpiration can become significantly uncoupled from stomatal conductance as a result. For example,  $g_s$  of wheat is often  $>20\%$  less at elevated  $[\text{CO}_2]$  than ambient  $[\text{CO}_2]$ , but the resulting change in canopy evapotranspiration is  $<10\%$ . Two other factors also play a role in this response. First, total leaf area per unit ground area (or Leaf Area Index, LAI) can be greater at elevated  $[\text{CO}_2]$  and offset the decrease in transpiration per unit leaf area. Second, the decrease in  $g_s$  and transpiration at elevated  $[\text{CO}_2]$  results in less evaporative cooling of the canopy and increases in leaf temperature. The internal air spaces of leaves are saturated with water vapor (i.e., relative humidity = 100%). Therefore, as leaf temperature rises there is an exponential increase in the water vapor pressure of air inside the leaf, and the gradient of water vapor pressure from inside the leaf to outside the leaf becomes greater, driving greater transpiration for a given stomatal conductance.

Reduced canopy-scale transpiration at elevated  $[\text{CO}_2]$  can ameliorate drought stress by conserving soil moisture during drying events and delaying the onset of stress. In addition, greater starting  $c_i$  at elevated  $[\text{CO}_2]$  and the nonlinear shape of the  $A/c_i$  curve mean that there is less inhibition of  $A$  by reduced  $\text{CO}_2$  supply (low  $c_i$ ) when plants close their stomata in response to drought.

### **Biomass and Seed Responses of $C_3$ Plants to Elevated $[\text{CO}_2]$**

Most  $C_3$  plants grown at elevated  $[\text{CO}_2]$  achieve greater biomass accumulation due to improved carbon gains associated with (1) direct stimulation of  $A$  and (2) indirect amelioration of stress when it occurs. The combined action of these two mechanisms is the basis for the expectation that the relative stimulation of biomass production by elevated  $[\text{CO}_2]$  will become progressively greater under increasingly drought stressed conditions. That said, at some level drought will be so stressful that any effects of elevated  $[\text{CO}_2]$  will become irrelevant because the plants are unable to survive.

Growth at elevated  $[\text{CO}_2]$  concentrations predicted to occur in the mid-twenty-first century has been shown to stimulate the annual net biomass production (defined as aboveground Net Primary Production; NPP) by approximately 20 % over a broad range of temperate forest types (Norby et al. 2005; Norby and Zak 2011). Extra biomass has been shown to take the form of extra wood or greater fine root production, depending on the forest type. This demonstrates the potential for forests to absorb more  $\text{CO}_2$  from the atmosphere as atmospheric  $[\text{CO}_2]$  rises and slow the rate of climate change relative to anthropogenic carbon emissions. However, experiments fumigating entire forest canopies with elevated  $[\text{CO}_2]$  in order to test this possibility have been restricted to young, plantation forests in temperate latitudes and relatively short periods of time (<15 years) compared to the life cycle of most trees (tens to hundreds of years). This is significant because in some forest experiments, elevated  $[\text{CO}_2]$  stimulated NPP initially, but then the response diminished and stopped after approximately a decade. This pattern has been attributed to a process called progressive nitrogen limitation. Progressive nitrogen limitation occurs when stimulation of biomass production at elevated  $[\text{CO}_2]$  locks up a large fraction of the nitrogen in an ecosystem in inaccessible pools including wood and soil organic matter. Over time, insufficient nitrogen is then available to support stimulation of biomass production by greater photoassimilate availability. In some forests exposed to elevated  $[\text{CO}_2]$ , faster release of nitrogen by microbial decomposition from soil organic matter occurs due to greater allocation of carbon from trees to the microbes. This takes the form of greater exudation of carbon-rich compounds from roots and greater carbon supply to mycorrhizae. When nitrogen cycling is accelerated at elevated  $[\text{CO}_2]$  in this manner it has prevented progressive nitrogen limitation from occurring for a decade. However, it is not clear how long such mechanisms can continue to operate. Mathematical modeling suggests that progressive nitrogen limitation is likely in most temperate forests on multi-decadal timescales. Experimental evidence for this is lacking, along with information on how mature forests as well as tropical and boreal forests may respond to elevated  $[\text{CO}_2]$ . This is a significant source of error in projections of future carbon cycling because of the large contribution of these particular forests to the terrestrial carbon sink.

In  $\text{C}_3$  crops, greater biomass production is typically associated with greater seed yield (Easterling et al. 2007; Tubiello et al. 2007; Leakey et al. 2009). Multiple components of yield can contribute to this response, although an increase in the number of seeds is usually more sensitive than an increase in the size of individual seeds. Greater seed number can result from greater numbers of seeds per pod or panicle or increases in the number of pods or panicles. On average, the major  $\text{C}_3$  crops of wheat, rice, and soybean achieve ~15 % greater yield when grown in the field at  $[\text{CO}_2]$  expected for the mid-twenty-first century versus ambient  $[\text{CO}_2]$  at the beginning of the century. However, there is significant variation around the mean driven by genetic variation among crop varieties and environmental conditions. Genetic variation in crop yield response to elevated  $[\text{CO}_2]$  could be exploited to identify key genes that control sensitivity and provides one possible route to adapting crops for improved performance in future growing conditions (Leakey and Lau 2012).

## Elevated [CO<sub>2</sub>] and Water Use Efficiency

The term water use efficiency is used to express the ratio of transpiration to carbon assimilation. Since both  $g_s$  decreases and  $A$  increases when plants are grown in elevated [CO<sub>2</sub>], plants generally become more water use efficient. The effects of elevated [CO<sub>2</sub>] on water use efficiency have been studied at various scales ranging from the leaf to the ecosystem. As is commonly found with transpiration, effects of elevated [CO<sub>2</sub>] on water use efficiency appear greater at the leaf or plant scale than is commonly seen at the canopy level. However, since the effect of elevated [CO<sub>2</sub>] on photosynthesis generally remains higher than control, even in light of photosynthetic downregulation, the effect of water use efficiency is not usually seen to decrease to the same magnitude over various scales of measurement as transpiration.

---

## Physiological Responses of C<sub>4</sub> Species to Growth at Elevated [CO<sub>2</sub>]

C<sub>4</sub> plants response to CO<sub>2</sub> are of key economic, ecological, and biogeochemical significance at a global scale. Maize, sorghum, millet, and sugarcane are all important C<sub>4</sub> crops. C<sub>4</sub> species in tropical and temperate grass ecosystems contribute approximately one quarter of global terrestrial NPP. And, 14 of the world's 18 worst weed species use C<sub>4</sub> photosynthesis. Therefore, understanding their response to elevated [CO<sub>2</sub>] and other aspects of global environmental change is important.

C<sub>4</sub> photosynthesis involves anatomical and biochemical modifications that concentrate CO<sub>2</sub> to levels five to six times greater than atmospheric [CO<sub>2</sub>], in specialized bundle sheath cells. RuBisCO is localized to these cells containing super-elevated [CO<sub>2</sub>], and as a consequence, its carboxylation reaction is favored over the oxygenation reaction. This adaptation avoids photorespiration and improves photosynthetic efficiency under conditions that otherwise promote photorespiration, i.e., high temperatures and drought stress. In addition, C<sub>4</sub> photosynthesis is typically CO<sub>2</sub> saturated at today's ambient [CO<sub>2</sub>]. Therefore, when C<sub>4</sub> plants are grown at elevated [CO<sub>2</sub>], there is no direct stimulation of  $A$  like there is in C<sub>3</sub> species (Leakey et al. 2009). Nevertheless, growth at elevated [CO<sub>2</sub>] decreases  $g_s$  and increases  $c_i$  in C<sub>4</sub> species. Lower  $g_s$  reduces canopy transpiration more consistently than in C<sub>3</sub> species because LAI is not greater in the absence of a direct stimulation of  $A$ . Lower canopy water use at elevated [CO<sub>2</sub>] can in turn slow the depletion of soil moisture during drought and delay stress. Additionally, greater  $c_i$  counteracts the reduction in  $c_i$  caused by stomatal closure (lower  $g_s$ ) during drought stress. Overall, this means that elevated [CO<sub>2</sub>] can ameliorate growth and yield losses to stress in times and places of drought.

The extent to which the amelioration of drought stress and indirect stimulation of productivity at elevated [CO<sub>2</sub>] occurs when C<sub>4</sub> plants are grown in different environmental conditions (e.g., varying water supply, nutrient availability, or temperature) is still poorly understood. Furthermore, only maize, sorghum, and miscanthus have been the subject of detailed field-based studies. This contributes to uncertainty in predictions of future ecosystem productivity and crop production.

## Plant Responses to Temperature

### Introduction

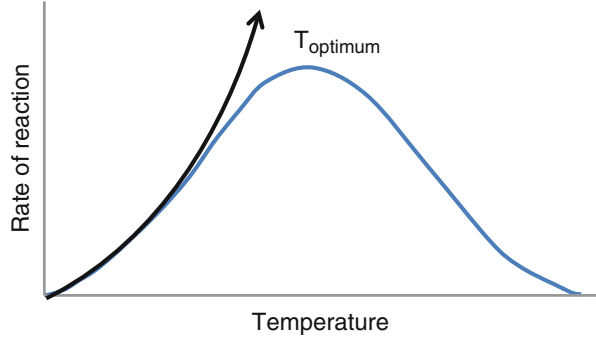
Under ideal conditions, the rate of a chemical reaction increases exponentially with increasing temperature (Fig. 7). This results from the reactants being more likely to collide and react because they are moving faster at higher temperatures. However, in biological systems the majority of reactions are catalyzed by enzymes or associated with lipid membranes. And, in these cases, the exponential increase in reaction rate continues only until a temperature is reached at which enzyme or lipid membrane functionality begins to decline. Above this temperature, defined as the temperature optimum (Fig. 7), biological reaction rates decline progressively until they eventually reach zero. Enzymes are temperature sensitive, because as proteins they rely on a variety of bonds between amino acids that form the peptide chain to maintain the conformation (or shape) of the enzyme. Correct enzyme conformation is essential to successful binding of the correct substrates and the reaction taking place. High temperatures destabilize these bonds, causing the protein to initially lose functionality. Eventually, if temperatures become high enough, the protein will become denatured, as happens to an egg when it is boiled. Likewise, lipid membranes that play key roles in many cell functions become too fluid and unstable at high temperatures. This results in leakage and instability. For example, this can interfere with establishment of the proton gradients across membranes that drive ATP synthesis in photosynthesis and respiration. The overall result is that temperature response curves of most biological processes have a characteristic hump-backed shape, often with an optimum temperature between 35 °C and 40 °C (Fig. 7) – although there is wide variation in the optimum temperature and the sensitivity of reaction rates to changes in temperature above or below the optimum.

There is significant variation in temperature response to temperature across the globe associated with latitude, altitude, continentality, seasonality, and the day-night cycle. This means that at different times and locations, global warming will be superimposed upon current temperatures that could be below, at, or above the temperature optimum for a particular biological process, e.g., plant growth. At times and locations where temperature is below-optimum, warming will increase activity. At times and locations where temperature is currently optimal or above-optimum, warming will cause moderate or rapid decreases in activity, respectively. This chapter mainly focuses on mechanisms of plant response to above-optimum temperatures.

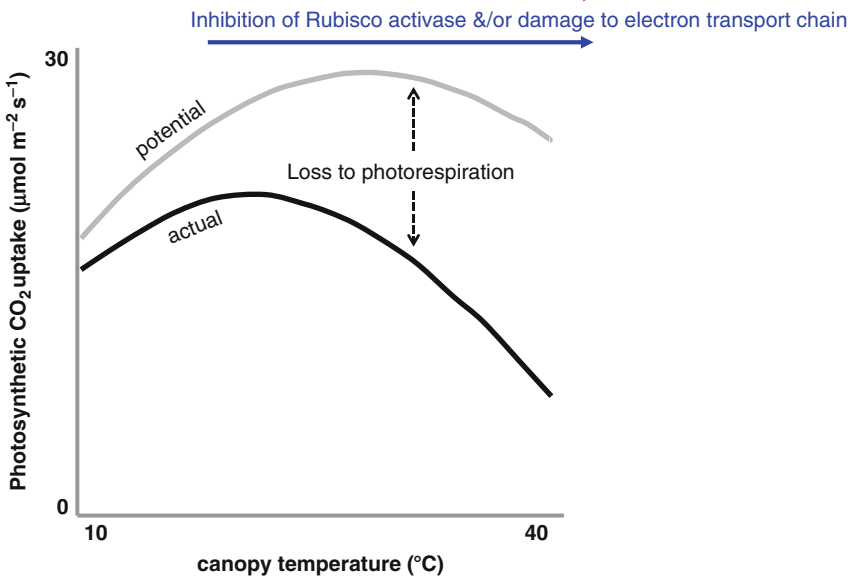
### Photosynthetic and Respiratory Responses to High Temperature

At current [CO<sub>2</sub>] and high light intensities, the activity of RuBisCO (Rubisco) is typically limiting. As leaf temperature increases from zero, the rate of the RuBisCO carboxylation reaction increases. However, the rate of the oxygenation reaction of RuBisCO increases more rapidly with rising temperature than the rate of

**Fig. 7** The response of reaction rates to temperature under either ideal conditions in a nonbiological system (*black line*) or in a biological system (*blue line*) where reaction rates are limited at high temperatures by increasing instability of important cellular components such as enzymes and lipid membranes



Increasing velocity of carboxylation, but decrease in Rubisco specificity



**Fig. 8** Temperature response of potential and actual photosynthetic CO<sub>2</sub> uptake (*A*), along with loss of CO<sub>2</sub> from photorespiration. Above the optimum temperature for photosynthesis, RuBisCO activase and/or damage to photosynthetic membranes contributes to impairment of *A*

carboxylation. This results from a change in the specificity of Rubisco for oxygen versus CO<sub>2</sub>, and a greater increase in the solubility of oxygen than CO<sub>2</sub>. As a consequence, *A* (the net fixation of CO<sub>2</sub> resulting from the balance of carboxylation by RuBisCO and other processes releasing CO<sub>2</sub>, including photorespiration and mitochondrial respiration in the light) increases initially as temperature rises, but then reaches an optimum beyond which increases in photorespiration rate exceed the rate of carboxylation (Fig. 8; Sage and Kubien 2007). In addition to this

difference in the enzyme kinetics of the carboxylation and oxygenation reactions of RuBisCO, *A* is inhibited at high temperatures by components of the photosynthetic machinery that are heat labile. Two currently competing hypotheses regarding the photosynthetic component that is most heat labile have been proposed.

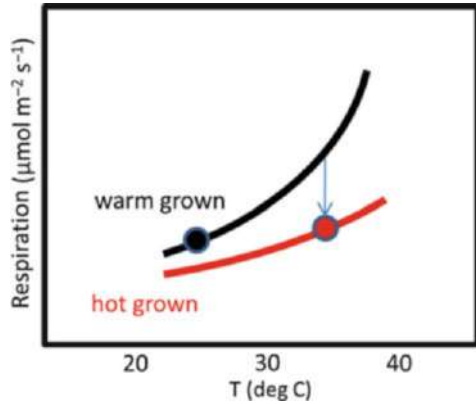
First, in species including *Arabidopsis*, black spruce, and poplar, there is evidence that the rate-limiting process for photosynthesis at above-optimum temperatures is the denaturation of RuBisCO activase. RuBisCO activase is an enzyme that normally acts to remove bound inhibitors from RuBisCO active sites and supports the carbamylation and  $Mg^{2+}$  binding necessary for binding and carboxylation of RuBp. RuBisCO activase functionality declines as temperature increases above thresholds between 22 °C and 35 °C, while the rate of processes that inactivate RuBisCO accelerates, leading to inhibited *A*. The key role of RuBisCO activase has been demonstrated through alterations in thermotolerance arising from modifications of the gene sequence in *Arabidopsis*.

Second, in species including sweet potato, cotton, tobacco, and spinach, there is evidence that the rate-limiting process for photosynthesis at above-optimum temperatures is regeneration of RuBp in the Calvin cycle. This is driven by declining electron transport rates caused by greater permeability of thylakoid membranes. In these cases, declines in RuBisCO activation with rising temperature are argued not to be limiting, but instead a regulated response to the decreases in electron transport. Notably there was no change in the high temperature tolerance of transgenic tobacco in which RuBisCO activase content was reduced by 55 %.

It is unclear what distinguishes the two proposed response mechanisms, but one strong possibility is that species, and possibly ecotypes, fundamentally differ in which component of the photosynthetic machinery and regulatory apparatus is most temperature sensitive. In many cases, genotypic variation in photosynthetic responses to temperature is not understood. But, it has been linked to diverse mechanisms including (1) polymorphisms in the RuBisCO activase gene, especially those in regions associated with ATPase activity and RuBisCO recognition; (2) variation in the number of RuBisCO activase genes carried by different species; (3) alternative splicing to generate multiple isoforms of RuBisCO activase with distinct temperature response characteristics; and (4) differences in the relative capacities for carboxylation by RuBisCO and RuBp regeneration. In addition, there is likely to be genotypic variation in the role of protein:protein interactions in stabilizing RuBisCO activase at high temperature, a function that has been proposed for a particular chaperonin protein in *Arabidopsis*.

Confounding the generalized responses described above is the impact that acclimation of the underlying photosynthetic machinery often has when plants are grown in elevated temperature. The temperature optimum of *A* is often shown to shift towards the average growing temperature thereby enhancing *A*. However, there are limits to the extent to which acclimation can relieve heat stress. In addition, plants that grow in high temperature conditions have evolved adaptations to allow them to operate under extreme conditions. In many cases, the mechanistic basis for acclimation and adaptation of the photosynthetic apparatus to high temperature is poorly understood.

**Fig. 9** Temperature responses of dark respiration where warm-grown (25 °C) plants are compared with plants grown at hot temperatures (35 °C) and displaying acclimation responses. Dots depict respiration rates at growth temperatures and the arrow represents the potential for respiratory homeostasis (Redrawn from Atkin et al. (2005))



The temperature optimum for respiration is greater than is typically experienced by leaves at night when it is dark or other tissues that are protected by bark or are underground. Therefore, for practical purposes, plant respiration in the dark can be considered to increase exponentially with rising temperatures (Fig. 9; Atkin et al. 2005). This reduces net plant carbon balance. Increased respiration of leaves at high temperature is particularly significant because leaves contribute the largest fraction of whole-plant respiration by any tissue. In addition, substantial carbohydrate reserves that would otherwise support growth or storage processes are wasted.

However, upon long-term growth at a high temperature, acclimation often occurs to diminish the stimulation of respiration rates (Fig. 9). The extent of the acclimation can vary from no change in short-term temperature response characteristics to complete homeostasis. In the latter case, respiration rates are equal before and after the transition to higher temperature. When plants develop in sustained higher temperatures, acclimation can involve greater metabolic adjustment and homeostasis. Acclimation of this type is proposed to be the result of reductions in respiratory capacity that involve lower mitochondrial density and respiratory enzyme content that can only be achieved through the production of new leaves.

## Cellular Responses to High Temperature

While moderately high temperatures result in decreased plant carbon balance and productivity due to reduced *A* and stimulated respiration, if temperature rises further, it can cause widespread dysfunction of cellular processes and relatively rapid death. For example, exposure of many crop plants to 50 °C for as little as 10 min is lethal. Plants evolutionarily adapted to high temperatures are more tolerant, such as the cacti, prickly pear, which can survive 15 min at 65 °C before being killed. In addition, certain tissues are more temperature tolerant than the plant as a whole. Pine pollen is not killed until it has experienced 70 °C for an hour and alfalfa seeds are killed by 120 °C only after 30 mins.



Heat stress causes dysfunction in the enzymes and bilayer lipid membranes that are essential to cell function. This results in increased production of ROS from the high-energy enzyme-driven reactions that take place in the thylakoid and mitochondrial membranes as part of photosynthesis and respiration, respectively. Further, ROS are produced as a part of cellular stress signaling. The result is damage to a wide range of complex molecules found in cells, including enzymes and other protein structures, lipid membranes, and nucleic acids. High temperatures also increase water loss from tissues and can result in tissue dehydration. Cellular dehydration can be an additional cause of enzyme dysfunction, particularly for the large fraction of enzymes in the cell which require sufficient water available to be solubilized.

A coordinated set of cellular responses occurs in response to sudden high temperature exposure, which operate to counteract the negative effects of heat on enzymes, lipid membranes, ROS, and dehydration (Mittler et al. 2004; Mittler and Blumwald 2010). Collectively, these are referred to as a heat-shock response, and it is characterized by transiently increased expression of genes encoding heat-shock proteins. Heat-shock proteins are found in all forms of life, including archaea, bacteria, plants, and animals. The temperature at which a plant normally grows influences the temperature at which expression of heat-shock proteins is induced. But, in general, heat-shock proteins in higher plants are induced by temperatures greater than 38–40 °C. Heat-shock proteins function by binding to a wide range of structurally unstable proteins in response to many stresses in addition to heat. The binding of heat-shock proteins helps to prevent aggregation of denatured proteins, renature proteins that are denatured, stabilize proteins as they are being translated from RNA, and modify proteins to allow membrane transport. Five classes of heat-shock proteins have been identified, based on their molecular weights (HSP 100, HSP 90, HSP 70, HSP 60, and small [sm]HSP). They are found throughout the cell. The gene expression of heat-shock proteins is controlled by transcriptional regulators called heat-shock factors. Expression of heat-shock factors and heat-shock proteins is associated with signaling that upregulates antioxidant metabolism, osmotic regulation, and changes in lipid membrane structure. For example, heat stress induces expression of cytosolic ascorbate peroxidase as part of upregulating antioxidant metabolism. Protection against dehydration is also induced by upregulation of pathways producing compatible solutes and osmolytes that stabilize complex molecules and increase osmotic potential. These small molecules include mannitol, proline, and glycinebetaine. Membrane stabilization is achieved at high temperatures by increasing the fraction of lipids that are saturated versus unsaturated. This raises the melting point in the same manner that makes butter solid at room temperature when olive oil is liquid. As a result the viscosity of the membrane can be maintained at levels that optimize its function as an ion barrier and medium that supports proteins of diverse functions.

Acquired temperature stress tolerance and acquired thermotolerance are the terms used to describe the phenomenon whereby a normally lethal temperature can be survived as a result of being initially exposed to a high, but sublethal temperature. For example, *Arabidopsis* seedlings grown at 22 °C are killed by a 120-min exposure to 45 °C. However, if the seedlings experience 38 °C for 90 min

prior to the exposure to 45 °C, they survive. Acquired thermotolerance to extreme temperature heavily depends on the heat-shock protein network.

## Crop Reproductive and Yield Responses to High Temperature

Impaired carbon gain and unwanted carbon use due to high temperature reduces the resources available to build reproductive structures and fill seeds. But, in addition, severe reproductive failure can result from direct effects of temperature on reproductive processes, including (1) early or delayed flowering, (2) asynchrony of male and female reproductive development, (3) defects in parental tissue, (4) defects to male and female gametes, and (5) impairment of seed filling (Barnabas et al. 2008). This is a serious issue because the majority of our food supply is a product of sexual reproduction in flowering plants. And, during the short time surrounding fertilization, even a single hot day or cold night can cause reproductive processes to fail for many plant species.

Moderate heat stress will often accelerate flowering, which may cause reproduction to occur before plants accumulate adequate resources (i.e., carbon or nutrient reserves) for allocation to developing seeds. Above a critical threshold, even small changes in temperature can act as cues for the induction of flowering. This response has a genetic basis that is distinct from the known genetic pathways of floral transition and appears to correlate with changes in RNA processing. There is substantial genetic variation in this response in Arabidopsis, suggesting crops could be bred to minimize it.

Temperature stress can sometimes have different effects on male and female structures, thereby creating asynchrony between male and female reproductive development. In maize (*Zea mays*), floral asynchrony is a significant problem under conditions of combined stress from heat and water deficit. There is significant genetic variation in the anthesis-silking interval (time between maturation of male and female flowers) of maize varieties. This means greater yields can be achieved under stressful conditions by selecting genotypes with a genetic predisposition for short anthesis-silking interval.

High temperature stress can shorten the period of time in which the stigmas in the flowers are receptive to pollen and thereby decrease the chances for a successful fertilization. For example, the stigmas in peach at 30 °C lose their ability to support pollen germination after 3 days, whereas at 20 °C they are viable for 8 days. Heat stress at critical stages of flower development causes ovary abnormalities in wheat and reduces the total number of ovules as well as increasing the ovule abortion rate in Arabidopsis. These are all examples of heat damage to female reproductive structures. The effects of temperature stress on male gametes are well documented for numerous plant species. Pollen maturation, viability, germination ability, and pollen tube growth can be negatively affected by heat. For example, increasing growth temperature for tomato from 28/22 °C (daytime/nighttime) to 32/26 °C caused a 50 % reduction in pollen production and a further ~66 % reduction in viability of pollen that were produced. In rice, high temperature stress at anthesis

involves abnormal pollen release from the anthers, resulting in less pollen and unviable pollen reaching the stigma. This is important because a threshold of 10–20 viable pollen grains must reach the stigma to ensure successful pollination. Genotypic variation in anther size and structure has been related to heat tolerance.

Yield losses to rising temperature have been observed over the last 30–40 years in some temperate and tropical regions and are anticipated to worsen for many crops across much of the globe. A latitudinal gradient of crop response is anticipated (Easterling et al. 2007; Tubiello et al. 2007). Where current temperatures are low at higher latitudes, warming will likely increase yields. This is partly due to higher average temperatures during the growing season and a lengthening of the growing season. However, these gains will be modest since many of these regions are not currently intensively farmed. Agricultural intensification would not be favored because the soils are often low in nutrients, but contain very high levels of organic matter that would be oxidized and released by microbial respiration as CO<sub>2</sub> if the land is converted to agriculture. In lower, warmer latitudes, losses of crop yield are expected to be greater because (1) many crops already grow near or above optimal temperatures and (2) greater humidity reduces evaporative cooling of crop canopies, resulting in greater plant tissue temperatures relative to air temperature.

The rice production area of tropical/subtropical Asia provides a case study of a cropping system that will be strongly negatively impacted by global warming. Rice yields in this region are currently being reduced by at least 30 % for every degree of increase in night temperature during seed filling above a critical threshold of 22–23 °C. In addition, high daytime temperatures are causing reproductive failure of rice at local scales. Temperature increases associated with climate change will immediately exacerbate the mechanisms currently driving yield loss, while also potentially exceeding the temperature thresholds of additional physiological processes that are important in determining yield. It appears that observed reductions in yield resulting from high nighttime temperatures can be explained, at least in part, by greater respiratory loss of carbon. Meanwhile, current-day yield losses to high daytime temperatures are most commonly ascribed to reproductive failure, with inhibition of photosynthesis expected to cause further yield loss as daytime temperatures rise in the future.

## **Carbon Cycling Responses to High Temperature**

As with crop yield, the effect of warming temperatures on terrestrial NPP will vary with latitude (Bonan 2008; Allen et al. 2010). Warming in the arctic is increasing productivity, in part by allowing greater growth of woody species further north. This will continue in the future and coincide with a shift in the boreal forest biome to the north as well. This is projected to lead to a net loss of carbon from the land to the atmosphere because the amount of carbon lost from fire and decomposition on the drying southern edge of the boreal forest as it is converted into grassland will be greater than the additional carbon fixed by expansion to the north. Accelerating rates of tree mortality in the Western USA over the last 40 years have been

attributed to result from warming that has triggered greater drought stress in interaction with greater attacks from insect and microbial pests and pathogens. This disturbance is again predicted to continue and be associated with a net loss of CO<sub>2</sub> from the biome. Finally, stem growth and NPP of tropical forests have been shown to correlate negatively with annual average daily minimum temperature, suggesting that warm nights lead to greater respiratory carbon losses. As a consequence of these trends across many regions of the world, models of the global biogeochemical cycling indicate that global warming will act to lower NPP and reduce absorption of fossil fuel emissions from human activities. In fact, some models predict an amplification loop will occur in which warming leads to loss of CO<sub>2</sub> from ecosystems, especially the Amazon forest, which in turn accelerates warming and drought, before ultimately leading to forest collapse. However, there is considerable uncertainty in the resilience of forest ecosystems to such a response and the precise tipping point of warming and drying that would be required to trigger it. Fire plays an important role in forest mortality. As trees die from stress, they increase the fuel load so that the heat and extent of fires are increased. This in turn leads to greater canopy loss and, especially in tropical areas, reduced local convective water cycling. This in turn further exacerbates drought stress. Notably, the minimum area of a tropical forest fragment required for populations of mammals and birds to be self-sustaining is similar to that necessary to support local convective water cycling. Therefore, efforts to conserve forest patches to support biodiversity may incidentally allow local climatic regulation by the ecosystem too.

---

## Plant Responses to Drought

### Introduction

Water plays a number of essential roles in the life of plants. Water is the medium in which all cellular activities occur. For example, it readily dissolves the large amounts of ions and metabolites essential for metabolism. Water is also the medium in which most metabolites and hormones are transported around plants. Plants respond to drought depend on water to a large degree for their structural integrity. Finally, the most fundamental and unavoidable resource trade-off for terrestrial plants is that in order to fix CO<sub>2</sub> through photosynthesis, they inevitably lose water. This fact means that plants incorporate <1 % of the water that they absorb. This contrasts with retention of >90 % of absorbed nitrogen, phosphorus, and potassium or 10–70 % of CO<sub>2</sub> recently fixed by photosynthesis. Water is also a major feature of a plant's energy budget. As water evaporates from leaves, it cools the canopy significantly. If transpiration is reduced, the canopy warms. If water supply or flow is restricted to the point that transpiration ceases, leaves rapidly reach lethal temperatures. Given all of these important roles for water, plants experience significant stress when it is scarce (Chaves et al. 2003; McDowell et al. 2008). For the purposes of this chapter, drought is defined as when the supply

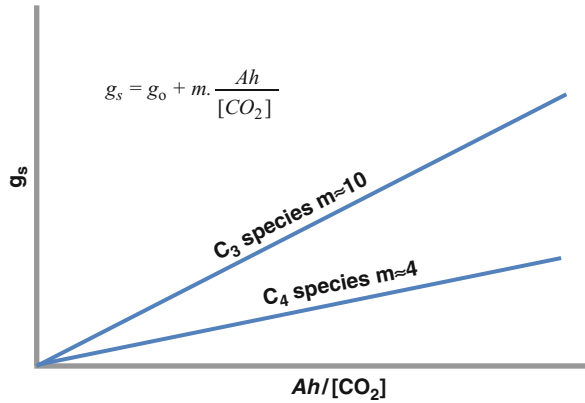
of water to a plant does not meet the demand of the plant for water. Drought stress can therefore occur as a product of soil drying and/or desiccating atmospheric conditions of low humidity and/or heat. Drought occurs with varying frequency and intensity in all biomes and is the primary limitation to crop yield as well as ecosystem productivity globally.

## **Stomatal, Photosynthetic, and Respiratory Responses to Drought**

Stomata play a key role in regulating the trade-off between carbon gain and water use in plants. Stomata are highly sensitive to signals of environmental conditions associated with drought stress, including soil moisture and atmospheric humidity. Low atmospheric humidity increases the gradient in humidity from inside to outside of the leaf, causing greater rates of transpiration for a given  $g_s$ . However, low atmospheric humidity also causes stomata to close. This reduces  $g_s$ , which slows the rate of water loss by transpiration. But, it also constrains the diffusion of  $\text{CO}_2$  into the leaf, limiting the supply of this substrate to RuBisCO and lowering  $A$ . The exact signal associated with low atmospheric humidity and the mechanism by which this triggers stomatal closure is not fully understood. However, stomatal sensitivity to atmospheric humidity, in addition to atmospheric  $[\text{CO}_2]$  and  $A$ , is captured to a remarkable degree by a simple linear equation (Fig. 10). In this equation,  $h$  is the fractional atmospheric relative humidity,  $[\text{CO}_2]$  is the atmospheric  $[\text{CO}_2]$  at the leaf surface,  $g_0$  is the y-axis intercept, and  $m$  is the slope of the line. This model and derivatives of it are the core of many models of future crop performance and ecosystem biogeochemistry.

Stomata also close in response to drying soil associated with low water supply (Wilkinson and Davies 2010). These responses include both chemical signals and hydraulic signals. Chemical signals are expected to dominate in the early stages of soil drying, while hydraulic signals predominate later under more severe drying when leaf water potential becomes increasingly negative and wilting occurs. Abscisic acid (ABA) is the predominant chemical message arriving from roots in contact with drying soil. Soil drying enhances the concentration of this hormone and increases pH in the xylem sap. The stomata of more desiccated plants and plants with greater xylem pH or cytokinin concentrations are more sensitive to ABA. However, the ABA that acts on guard cells does not originate entirely in the roots and is influenced by degradation and release of ABA in the stem and leaves. There is evidence that hydraulic signals can cause stomatal closure in the absence of ABA signals and this may involve the operation of proteins in cell membranes that act as osmosensors of altered cell turgor. The same chemical and hydraulic signals that cause stomatal closure have also been implicated in inhibiting leaf growth through interactions with other plant hormones such as ethylene and cytokinins. This is proposed to enhance drought response of the plant by reducing future canopy transpiration while also allowing greater resources to be allocated to root growth and acquisition of new water resources at greater depth in the soil.

**Fig. 10** Response of stomatal conductance ( $g_s$ ) to variation in leaf  $\text{CO}_2$  uptake ( $A$ ), humidity ( $h$ ), and  $[\text{CO}_2]$  in  $\text{C}_3$  and  $\text{C}_4$  species



Mild drought primarily inhibits  $A$  as a result of reduced  $\text{CO}_2$  supply through closing stomata, but as drought becomes more severe, changes in leaf osmotic status and chemistry also inhibit the biochemical reactions of photosynthetic metabolism. Many different processes in photosynthesis have been implicated in sensitivity to drought, but the ultimate collective effect is to reduce photosynthetic capacity in terms of both  $V_{\text{cmax}}$  and  $J_{\text{max}}$ .

Drought stress consistently reduces total plant respiration because it directly inhibits growth and impairs  $A$  needed to produce the carbohydrate substrates that supply respiration. Leaf respiration generally follows this overall trend, but there are a meaningful fraction of cases in which drought has caused no change or an increase in leaf respiration. However, changes in respiration are small in scale relative to photosynthetic responses, so net carbon balance always declines with drought. Where rates of respiration are maintained or increased under drought, it is associated with greater demand for energy from protective and repair processes.

### Plant Dehydration, Osmotic Adjustment, and Hydraulic Failure

Water moves down gradients of increasingly negative water potential. The atmosphere has extremely negative water potential ( $\sim -10$  to  $-100$  MPa) compared to plant tissues ( $\sim -0.5$  to  $-2.5$  MPa), which in turn are more negative than soil water potential ( $\sim 0$  to  $-2.0$  MPa) except under terminal drought stress. As soil dries and soil water potential becomes more negative, the gradient in water potential driving water absorption by the plant becomes diminished. However, water loss from the plant to the atmosphere continues, even if at a diminished rate because the stomata are closed. Therefore, plant tissues become dehydrated. This impairs a range of cellular processes including enzyme and lipid membrane function. In extreme examples, turgor is lost and plants wilt. Plants can respond to dehydration by increasing the synthesis of osmotically active molecules and ions – such as potassium, proline, and mannitol – which makes the osmotic potential of the tissue more negative. This in turn makes total water potential more negative and draws water into the plant from the soil by

osmosis. Many of these osmotically active molecules also act as compatible solutes, which improve protein stability under dehydrating conditions. The osmotic adjustment mechanism can maintain turgor during mild to moderate stress, but is eventually overwhelmed during intense drought. The extent to which cell water potential can decrease until the turgor-loss point is reached depends on cell elasticity. Cells with highly elastic walls contain more water at full turgor; hence, their volume can decrease more, before the turgor-loss point is reached. The elasticity of the cell walls depends on chemical interactions between the various cell-wall components.

In addition to loss of turgor, failure of water supply to meet demand can cause hydraulic dysfunction in the form of xylem cavitation. This occurs when the tension of water in xylem becomes sufficiently high that xylem sap vaporizes and dissolved air forms a bubble in the xylem vessel. This halts water flow in the xylem and damages the ability of the plant to transport water to the leaves, resulting in accelerated dehydration and stress. Most plants can repair cavitated xylem, and some do so frequently, but there is delay in return to full function. The manner in which stomata respond to water shortages influences the likelihood of cavitation. Isohydic species (e.g., soybean and sunflower) are those where a threshold water potential triggers stomatal closure to minimize further transpiration. Anisohydic species (e.g., poplar and maize) are those that are relatively insensitive to leaf water potential and whose rates of transpiration are consistently higher. These modes of action represent the extremes of a continuum of behavior. Isohydic species avoid extreme low water potentials and therefore xylem cavitation, but are more likely to suffer from carbon starvation due to limited CO<sub>2</sub> supply. Anisohydic species are more likely to experience hydraulic failure in the long term, but maintain greater carbon balance under mild stress.

## **Whole-Plant Physiological Plasticity and Adaptations to Drought**

Plant adjustments to drought include a number of whole-plant scale phenomena. Classic examples include changes in allocation of resources to root growth resulting in a shift towards deeper rooting in order to access deeper soil layers containing more moisture. Many grasses roll their leaves longitudinally in response to drought. This reduces transpiration by decreasing the canopy area intercepting solar radiation. Also, high humidity air is trapped inside the cylinder formed by leaf rolling, reducing the humidity gradient from inside to outside the leaf and reducing transpiration. Some species only have stomatal pores on the leaf surface than forms on the interior of the cylinder, increasing the benefits of this response. Leaf and stem wilting is both a consequence of tissue dehydration and a mechanism to reduce water loss, because it reduces absorption of solar radiation. In some species, particularly long-lived trees and shrubs, a more extreme version of the same response involves leaf or branch death and abscission under drought conditions. In each case, survival through reduction in water use and retention of water in key tissues is more important than the negative effects on plant carbon balance. This is an effective strategy because many plant species have considerable capacity for rehydration and recovery after soil rewetting.

Given the fundamental trade-off between carbon gain and water loss for plants, it is not surprising that evolutionary adaptations to avoid drought stress are highly diverse. They include strategies to maximize water uptake and water storage as well as minimize water use. The archetypal drought adaptation is Crassulacean Acid Metabolism (CAM) – a type of photosynthesis found in cacti and other plant groups found in desiccating environments. Plants with CAM photosynthesis circumvent the trade-off between carbon gain and water use by opening stomata at night and closing them during the day. At night, cooler temperatures mean that the gradient of atmospheric humidity from inside the leaf to the atmosphere outside is much smaller and rates of transpiration are relatively low. Meanwhile, CO<sub>2</sub> is captured at night and stored as malic acid in the vacuole. During the day, this CO<sub>2</sub> is rereleased and assimilated by RuBisCO. CAM photosynthesis is often accompanied by other drought adaptations such as water storing trunks, stems, or leaves; deep tap roots; thorns to eliminate tissue loss to herbivores; slow growth rates; or large underground storage organs. Slow growth rates represent a conservative strategy of stress tolerance by slow resource use and gain. Alternatively, some species living in dry environments avoid stress by being annuals with very rapid life cycles and long-lived seeds that remain in the soil until conditions are favorable for growth. A similar strategy of stress avoidance is found in seasonally dry forests of tropical and Mediterranean climates, which largely restrict growth to wet seasons.

## **Crop Yield and NPP Responses to Drought**

Drought stress will undoubtedly be one of the primary drivers of lower crop yield and ecosystem productivity as a result of climate change. However, it is very challenging to distinguish plant stress associated with inadequate water supply from stress associated with high temperatures because high temperatures greatly increase plant demand for water in addition to causing direct heat damage to plant tissues. In addition, while spatial variation in global warming over the twenty-first century can be predicted with reasonable confidence, models of future precipitation patterns are highly uncertain. Therefore, the predicted patterns of response crop yield and NPP response to greater drought in the future are largely a product of predicted warming over the twenty-first century (Easterling et al. 2007; Tubiello et al. 2007). These were described in the section above, on plant responses to temperature.

---

## **Plant Responses to Ground-Level Ozone (O<sub>3</sub>)**

### **Introduction**

OzoneOzonePlantsresponse to ozone (O<sub>3</sub>) in the atmosphere at ground level is a damaging air pollutant to almost all forms of life that are in contact with it, including plants. Economic losses in crop yield to O<sub>3</sub> pollution are currently estimated at \$14–26 billion per year. The productivity of natural ecosystems,



along with the ecosystem goods and services they provide, is also negatively impacted by current  $[O_3]$ . It is important to note that ground-level  $O_3$  pollution is a different environmental problem to the “ozone hole,” which is a reduction in  $[O_3]$  found high in the stratosphere that normally acts to filter out harmful UV rays from the sun.

Most ground-level  $O_3$  forms from a series of reactions that are catalyzed by sunlight between methane, volatile organic compounds (VOCs), and nitrous oxides ( $NO_x$ ). All of these gases are predominantly released into the atmosphere by human activities. There is significant spatial and temporal variation in  $[O_3]$  because it is highly reactive and forms or degrades quickly. Formation of  $O_3$  is favored by high temperatures and sunlight. Therefore, a clear diel cycle in  $[O_3]$  is typically observed with low  $[O_3]$  at night, rising  $[O_3]$  from shortly after dawn until midafternoon and then declining  $[O_3]$  in the evening. Sources of air pollutants such as vehicle exhausts and fossil fuel burning industries drive greater local ozone formation. Prior to the industrial revolution,  $[O_3]$  was less than 10 parts per billion (ppb) and this provides an estimate of “natural” background  $[O_3]$ . Today, daytime summer  $[O_3]$  regularly exceed 40 ppb in many parts of the Northern Hemisphere. However, ground-level  $O_3$  pollution is not restricted to urban areas and significant plumes of elevated  $[O_3]$  air often form or move over rural areas. And, in cities, recently formed  $O_3$  can react with  $NO_x$  precursors in a futile cycle of synthesis and degradation. This chapter mainly focuses on the responses of vegetation to  $O_3$  pollution. But, vegetation plays an important role as a sink for  $O_3$  from the atmosphere and can therefore mediate significant land-atmosphere feedbacks. The short lifetime of  $O_3$  in the atmosphere means that successful regulation of air pollution could lead to relatively rapid reductions in ground-level  $[O_3]$ . However, clean air legislation is poorly enforced in many regions of the world and ground-level  $[O_3]$  is predicted to rise on average in the twenty-first century, with significant increases projected for Asia and the Middle East. Some long-distance transport of  $[O_3]$  does occur, especially in the stratosphere, and inversion events can bring this distributed source of  $O_3$  pollution to ground level.

## Physiological Responses to Elevated $[O_3]$

The majority of damage caused to plants by  $O_3$  occurs after the gas has diffused into leaves via the stomata (Ainsworth et al. 2012; Fuhrer 2009). Therefore, the dose of  $[O_3]$  received by the plant is highly dependent on  $g_s$ . Conditions that favor greater  $g_s$ , such as greater light intensity, greater humidity, greater soil moisture availability, and greater leaf photosynthetic capacity all favor greater  $O_3$  uptake. In contrast, conditions that diminish  $g_s$ , such as low light intensity, low humidity, low soil moisture, elevated  $[CO_2]$ , and lower photosynthetic capacity, all favor lower  $O_3$  uptake.

Once inside leaves,  $O_3$  reacts rapidly with the water and dissolved molecules found on the cell walls that surround internal air spaces (the apoplast). This produces reactive oxygen species (ROS), including hydrogen peroxide, superoxide radicals, hydroxyl radicals, and nitric oxide. ROS are in turn highly reactive

molecules that can damage important classes of complex molecules found in cells, including enzymes and other protein structures, lipid membranes, and nucleic acids.

Cells rapidly sense elevated ROS levels in the apoplast and a complex signal transduction network involving plant hormones, calcium ions, and protein phosphorylation cascades is activated. As a result, the expression of defense genes is increased, leading to upregulation of antioxidant metabolism as well as cellular repair processes.

Elevated  $[O_3]$  decreases  $A$  across a wide range of species and environmental conditions. Reductions in  $A$  at elevated  $[O_3]$  are associated with reduced gene expression, protein content, and activity of RuBisCO and other photosynthetic enzymes. Lower  $A$  in turn reduces the pool sizes of sucrose and starch at elevated  $[O_3]$ . The decrease in carbon gain at elevated  $[O_3]$  is often compounded by greater rates of dark respiration. This may be due to the greater demand for energy from antioxidant, defense, and repair processes induced by elevated  $[O_3]$ . For example, there is evidence for elevated  $[O_3]$  stimulating production of apoplastic ascorbate, flavonoids, volatile terpenoids, and epicuticular waxes. In addition, elevated  $[O_3]$  commonly accelerates leaf senescence, reducing the lifetime over which a leaf can be contributing as a source of photoassimilates to the plant. The reduction in carbon supply to other growing tissue from the range of responses described above frequently leads to impaired root growth.

The decrease in  $A$  at elevated  $[O_3]$  drives a feedback mechanism resulting in lower  $g_s$  as well. Furthermore, there is evidence that  $O_3$  exposure can have a direct influence on stomatal function. This includes sluggish or insensitive stomatal responses to other environmental stimuli, including abscisic acid. This implies that plants grown at elevated  $[O_3]$  may fail to close their stomata in response to soil drying and exhaust soil moisture resources leading to greater productivity and yield losses to drought. However, other studies have reported that elevated  $[O_3]$  diminishes stress under drought by decreasing stomatal conductance and reducing plant water use. The need for greater understanding of the mechanistic basis for interactions between elevated  $[O_3]$  and drought or temperature is a key knowledge gap. On the other hand, many studies have indicated that elevated  $[CO_2]$  protects plants from  $O_3$  damage by reducing flux into the plant due to reduced  $g_s$  and by providing greater photoassimilate to fuel defense and repair responses.

It is important to note that a distinction is often drawn between plant responses to long-term exposure to moderate  $[O_3]$  (defined as “chronic” exposure of weeks to months at  $<150$  ppb) versus short-term exposure to high  $[O_3]$  (defined as “acute” exposure of minutes to hours at  $>150$ – $300$  ppb). Chronic  $O_3$  damage of the type described in the sections above is the most common scenario in the natural world and is often not evident from rapid visual inspection of leaves. Acute exposures have most commonly been applied in experimental settings. However, locations with extreme air pollution do experience  $[O_3]$  in the range that causes acute damage. Acute damage is characterized by programmed cell death and significant production of visible lesions on leaves.

## Biomass and Seed Responses to Elevated [O<sub>3</sub>]

Reduced *A* per unit leaf area is combined with reduced LAI to significantly lower NPP of crops and natural ecosystems under elevated [O<sub>3</sub>] (Ainsworth et al. 2012; Fuhrer 2009). Present day [O<sub>3</sub>] are estimated to be decreasing *A* of northern temperate tree seedlings by approximately 11 % and biomass by 7 %. Data for mature forests is scarce. But, an experimental aspen-birch-maple plantation in the upper Midwest USA responded to increasing daily season means of [O<sub>3</sub>] from 33 to 39 ppb to 49 to 55 ppb by decreasing aboveground biomass production by 13–23 % depending on the species mixtures. A similar experiment on mature beech and spruce trees reduced wood production by 44 %. However, as with other factors of global environmental change, there is still considerable uncertainty about how mature forests respond to elevated [O<sub>3</sub>], particularly in low latitudes. Nevertheless, mathematical modeling has been used to estimate that current [O<sub>3</sub>] is reducing NPP over parts of North and South America, Europe, Asia, and Africa by 5–30 %. Furthermore, it has been estimated that the negative effects of rising [O<sub>3</sub>] on global NPP could offset the stimulation of global NPP by rising [CO<sub>2</sub>].

Experimental tests of grassland ecosystem responses to elevated [O<sub>3</sub>] have been more variable than those on crops and forests. While evidence for reduced NPP at elevated [O<sub>3</sub>] has been reported, certain communities of temperate, calcareous, and alpine grasslands have been shown to be relatively insensitive. This may reflect the high diversity of grassland communities relative to the crop and plantation forest experimental systems, because shifts in relative species abundance resulting from altered competition occurred at elevated [O<sub>3</sub>] which may make the ecosystem NPP more resilient to perturbation. Nonetheless, short-term grassland experiments have observed the classic physiological responses to elevated [O<sub>3</sub>] in grasses, and impacts on NPP may become apparent on timescales longer than most experiments that have been done to date.

In wheat, lower *A* at elevated [O<sub>3</sub>] translates into lower yield primarily through reductions in seed weight. In soybean and rice, lower *A* at elevated [O<sub>3</sub>] translates into lower yield through reductions in both seed weight and the number of reproductive sinks (pods or spikelets) on the plant. A number of experimental methodologies have been used to estimate dose-response curves of crop yield, i.e., yield loss over a range of [O<sub>3</sub>]. These in turn have been used to justify air quality targets intended to protect ecosystem as well as human health in Europe. Air quality standards are either set to protect human health or are not in place for other regions of the world. It is important to note that there is genotypic variety in sensitivity to ozone within all the major crop species tested to date. This indicates the potential for the genetic basis for ozone tolerance to reside within current germplasm and be exploited in crop improvement programs that apply breeding or biotechnology to adapt crops for improved performance under elevated [O<sub>3</sub>]. On the other hand, there is evidence that current breeding strategies have not selected for improved O<sub>3</sub> tolerance in soybean or wheat over recent decades. This may in part reflect the general lack of awareness among farmers and agribusiness that [O<sub>3</sub>]

is globally responsible for \$14–26 billion in crop losses each year. This corresponds to significant crop losses in the major crops of soybean (6–16 %), wheat (7–12 %), rice (3–4 %), and maize (3–5 %). There is significant potential for these losses to grow over the twenty-first century, particularly in Asia and the Middle East.

---

## **Adaptation of Plants to Environmental Change**

Adaptation to environmental change Plants adaptation to environmental change involves responses that will reduce the sensitivity of the Earth system to changes in environmental conditions. A key factor determining human well-being in the twenty-first century and beyond will be the degree to which food and fuel crops can be adapted to future growing conditions. Adaptation may be achieved by changes in crop management as well as the development of new improved crop varieties.

There are many potential changes in management that can help crops tolerate future, more stressful growing conditions. It should be relatively simple to switch to planting alternative existing crop species/varieties that are more stress tolerant than the current crop at a given location. However, more stress tolerant species/varieties may have requirements for soil or photoperiod that do not match conditions in the growing area. Addition of greater irrigation and fertilizer could offset stress and yield loss to environmental change, but these solutions will not be sustainable in many cases. In addition, these options are not open to many farmers in the developing world, who will be amongst those most affected by climate change. On the other hand, technologies that reduce crop water use and loss, such as partial root zone drying (where only part of the root system receives irrigation so that root-to-shoot signals of soil drying are maintained and minimize  $g_s$ , while providing enough water to avoid significant dehydration), could be more widely adopted. Growing season length is already being extended due to warmer conditions early and late in the growing season. If adequate water is available, this provides a longer growth period for carbon fixation and can increase productivity. Integrated pest management practices can also be more widely adopted. Greater resistance to pest and diseases can confer greater drought tolerance when root tissue becomes healthier and can achieve greater water uptake.

Heat and drought stress have been limiting crop production since the inception of agriculture. Therefore, considerable effort has been applied to breeding heat and drought tolerant varieties of all major crops. One key approach in this process has been to exploit natural spatial variation in climate to provide locations where germplasm could be tested for tolerance to drought and heat. In addition, the first crops carrying transgenes that confer drought tolerance have recently been released from industrialized biotechnology research and development pipelines. For example, maize expressing a cold-shock protein from a bacterium has been marketed in the Central USA as a drought-tolerant product. Biotechnology arguably has the potential to more easily confer stress tolerance beyond the range found in existing crop germplasm. However, it also raises a range of socio-economic and political and

agroecological challenges. Whatever approach is used, further advances in the heat and drought tolerance of food and fuel crops are urgently needed because by late this century average growing season temperatures are predicted to exceed the most extreme years experienced currently.

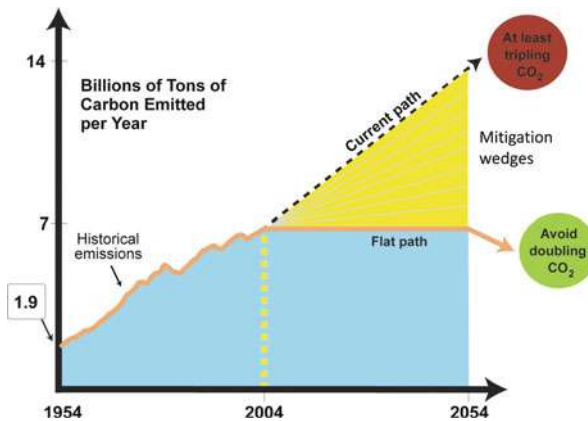
In contrast to heat and drought, rising atmospheric  $[\text{CO}_2]$  and  $[\text{O}_3]$  have only been recognized in the last few decades as factors substantially modifying crop performance. Therefore, there is not a history of farmer selection or industrialized breeding/biotechnology to maximize crop performance under elevated  $[\text{CO}_2]$  or elevated  $[\text{O}_3]$ . Nevertheless, genetic variation in crop responses to these factors suggests there is potential to gain greater benefits from elevated  $[\text{CO}_2]$  and ameliorate the negative impacts of elevated  $[\text{O}_3]$ . In addition, knowledge and modeling of the processes that limit plant metabolism and productivity under elevated  $[\text{CO}_2]$  have led to identification of targets for future crop improvement. For example, modeling and experimental data suggest that upregulated expression of the enzyme sedoheptulose-bisphosphatase can lead to enhanced *A* and productivity of  $\text{C}_3$  crops under elevated  $[\text{CO}_2]$ .

---

## Plant-Based Mitigation of Environmental Change

Mitigation of environmental change involves actions that reduce the net flux of  $\text{CO}_2$ , and other greenhouse gases, into the atmosphere from land and ocean. This includes (1) reductions in human fossil fuel burning and land-use change and (2) management of agricultural and natural ecosystems to maximize  $\text{CO}_2$  uptake from the atmosphere and long-term carbon storage (often referred to as carbon sequestration). There is considerable debate about what target is reasonable for mitigation efforts to attempt to achieve. Maintaining fossil fuel burning at current levels for the next 50–60 years has been suggested to be attainable and desirable, because it is projected to keep atmospheric  $[\text{CO}_2]$  from doubling above preindustrial concentrations (Fig. 11). To switch from the historical trend of greater  $\text{CO}_2$  emissions each year to constant emissions would require progressively larger amounts of  $\text{CO}_2$  to be kept out of the atmosphere each year. By 2060, projected annual  $\text{CO}_2$  emissions would need to be reduced by 8 Gt C. It has been proposed that 15 or more mitigation strategies, or “wedges,” that apply current or near-term technologies, could successfully be combined into a stabilization triangle to meet this goal. Three of the strategies are based around plant systems: biofuels, no-till agriculture, and reforestation.

Producing biofuels from crops rather than burning fossil fuels presents an opportunity to reduce  $\text{CO}_2$  emissions and sequester carbon in agricultural soils. Ideally, the amount of  $\text{CO}_2$  removed from the atmosphere by the crop through photosynthesis would exceed the amount of  $\text{CO}_2$  released to the atmosphere through plant and soil respiration, combustion of the biofuel product, and fossil fuel inputs involved in growing the crop and manufacturing the biofuel. In such a circumstance, the  $\text{CO}_2$  emissions saved would correspond to the total fossil fuel combustion required to perform the same amount of work as the biofuel, the fossil



**Fig. 11** Visualization of a proposal to mitigate future global environmental change by holding anthropogenic carbon emissions to the atmosphere at current levels through adoption of a set of mitigation strategies (or wedges) that will combine by 2060 to offset approximately 8 Gt C of emissions. If successful it is estimated this would hold  $[\text{CO}_2]$  at below  $2 \times$  preindustrial  $[\text{CO}_2]$  rather than continuing on a business-as-usual strategy where  $[\text{CO}_2]$  would end up at least  $3 \times$  preindustrial  $[\text{CO}_2]$  (Reproduced with permission from the Carbon Mitigation Initiative, Princeton University)

fuel inputs to make that amount of fossil fuel (e.g., oil refining), plus the net balance of carbon in the agricultural ecosystem. It is complex to determine “carbon budgets” for various fuel choices, and the discipline of life cycle analysis has emerged to calculate the biological as well as socioeconomic budget factors. It is important to recognize that biofuels have great potential as a strategy to mitigate rising  $[\text{CO}_2]$ , but this potential will only be met if the correct crops are grown in appropriate locations.

The two largest sources of biofuels produced today are maize (or corn) ethanol in the USA and sugarcane ethanol in Brazil. Maize is an annual crop requiring significant fossil fuel use in the production of fertilizers and pesticides as well as to drive the mechanized equipment used for planting, fertilizing, and harvesting. Sugarcane is a perennial crop that can regrow after harvesting each year and is typically replanted every 6–7 years. Sugarcane is also grown with lower application of synthetic fertilizers because it is perennial with higher nitrogen use efficiency and because green fertilizers from harvest straw or cover crops are applied. Both crops produce large quantities of biomass per unit land area. But, sugarcane is grown in Brazil on marginal quality land not used for grain production, while maize is grown in the USA on prime land. In the case of maize, this introduces competition for land between production of biofuel feedstock and production of maize for use in animal feed, food processing, and industrial applications. Ethanol is produced by either: (1) fermenting the sucrose stored in sugarcane stalks or (2) fermenting sugars produced by physical and chemical degradation of starch in maize kernels. The necessity for fossil fuel inputs to produce sugars from maize starch introduces inefficiency relative to using sucrose from sugarcane.

In addition, the production of ethanol is a major inefficiency in both cases. Ethanol is produced through fermentation by microbes in a sugar solution. But, ethanol is toxic to the microbes at low to moderate concentrations. This makes the manufacturing process inefficient because fermentation is done in batches, rather than as a continuous flow process. In addition, ethanol is miscible in water and so large fossil fuel inputs are needed to heat the mixture and distill off the ethanol with a high degree of purity. However, while fossil fuels are burned to distill maize ethanol, the bagasse (crushed stalks left after sucrose extraction) of sugarcane is burned as a biofuel to generate heat needed for distillation. When all the factors described above are combined, sugarcane ethanol offsets substantially more CO<sub>2</sub> emissions than maize ethanol.

As increasing areas of land are used to grow biofuels, it is critical to assess whether significant carbon stores are released into the atmosphere as a result of the land-use change. The worst case scenario is exemplified by the production of palm oil as a biofuel feedstock on land cleared out of tropical rain forest. In this case, the amount of carbon lost out of storage in the rain forest is extremely large. This creates a carbon “debt” that will take 10s–100s of years of biofuel production from palm oil to compensate for before any net reduction in CO<sub>2</sub> emissions occurs. In this case, it would be better to leave the rain forest intact and not produce the biofuels. In contrast, there are significant areas of marginal quality agricultural lands that have been abandoned that could be planted with biofuel feedstocks. In this case, CO<sub>2</sub> release can still occur, but will be more modest and can be minimized by selecting sites with primarily herbaceous vegetation and by avoiding soil disturbance. If such areas are planted with highly productive biofuel crops, then net reduction of CO<sub>2</sub> emissions can be achieved quickly. One complex scenario that should also be avoided is that of indirect land-use change. This describes the condition where biofuels are produced on land currently used for food production. The land is already in agricultural production, so there is typically not an increase in CO<sub>2</sub> emissions from local vegetation or soils. However, if food production is reduced, it can cause an economic response that encourages conversion of natural ecosystems into food production somewhere else in the world. This in turn releases CO<sub>2</sub> into the atmosphere and eliminates the benefit of the biofuel production.

In order to realize and maximize the benefits of biofuels, substantial investment is being made to test next-generation biofuel feedstocks and improve the efficiency of biomass conversion to liquid fuel. For example, perennial grasses such as miscanthus and switchgrass have been found to produce large quantities of biomass in the USA with relatively low nitrogen inputs. Miscanthus in particular is very productive due to having cold tolerant C<sub>4</sub> photosynthesis, rapid canopy closure, a long growing season, high water and nitrogen use efficiency, and resistance to biotic stress. A key additional objective is to develop methods to make liquid fuels from the cellulose. Cellulose is a polymer of glucose and makes up a very large fraction of plant biomass. But, it is much more chemically stable than starch and harder to break down into sugar. If efficient conversion can be achieved, then much larger quantities of liquid fuel could be produced per hectare of farmland. Maximizing production per hectare is key because it minimizes competition for farmland

and reduces the likelihood of unwanted land-use change from natural ecosystems to agriculture.

Most intensively farmed agricultural soils have been losing carbon from soil organic matter to the atmosphere over recent decades. This occurs in large part as a result of tillage, where physical disturbance of the soil surface and incorporation of crop residue by plows allow aerobic respiration by soil microbes to metabolize the residue and soil organic matter and release it as CO<sub>2</sub>. This CO<sub>2</sub> emission can be substantially reduced by no-till or reduced-till practices. These involve the use of alternative soil preparation and seeding equipment to sow the crop while maximizing the crop residue left on the soil surface and minimizing disturbance of the soil surface. No-till or reduced-till practices have the added benefits of reducing CO<sub>2</sub> emissions and costs associated with the number of times a tractor works a field, as well as reducing erosion and increasing nutrient retention.

Forests are a major store of carbon in the global carbon cycle. Roughly 15 % of current annual anthropogenic CO<sub>2</sub> emissions come from deforestation. Reducing deforestation and encouraging reforestation are therefore powerful potential mitigation strategies. However, these are challenging goals to achieve because most forests are unmanaged, in remote locations and in developing tropical countries where deforestation is profitable and not heavily regulated. Reducing Emissions from Deforestation and Forest Degradation (REDD) is a scheme that places monetary value on carbon stored in forests and, thereby, uses market-based economics to incentivize reforestation. The idea is to provide developing countries with financial incentives to reduce national deforestation rates below a baseline determined from historical trends or a future projection. Countries that were able to demonstrate reductions in CO<sub>2</sub> emissions from deforestation would then be able to sell carbon credits on an international carbon market. However, there are significant scientific and political challenges to implementing this plan. The greatest scientific challenge is to find methods to assess how forest carbon stocks change through time with deforestation and reforestation over vast land areas. Modeling and remote sensing approaches where aircraft or satellites gather data on forest structure and extent are being developed to address this need.

---

## Future Directions

Key knowledge gaps in understanding the role of plants in future global environmental change, and society's response to it, include

- Interactive effects of multiple factors of environmental change
- Thresholds in plant responses to environmental change
- Genetic variation in plant responses to environmental change in agricultural and natural ecosystems
- Ecological and evolutionary interactions that could amplify or negate the physiological responses to environmental change observed in individual plant genotypes



- Context-specific responses to environmental change of low and high latitude ecosystems that have been poorly studied, e.g., modification of CO<sub>2</sub> fertilization effects on tropical rain forest by variation in phosphorus supply
- Long-term responses of mature, perennial ecosystems to gradual changes in environmental conditions
- Understanding of how genotype drives phenotype so that complex multigenic traits can be engineered to give crop plants enhanced productivity and stress tolerance

---

## References

- Ainsworth EA, Rogers A. The response of photosynthesis and stomatal conductance to rising CO<sub>2</sub>: mechanisms and environmental interactions. *Plant Cell Environ.* 2007;30:258–70.
- Ainsworth EA, Yendrek CR, Sitch S, Collins WJ, Emberson LD. The effects of tropospheric ozone on net primary productivity and implications for climate change. In: Merchant SS, editor. *Annual review of plant biology*, vol. 63. Palo Alto: Annual Reviews; 2012. p. 637–61.
- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EH, Gonzalez P, Fensham R, Zhang Z, Castro J, Demidova N, Lim JH, Allard G, Running SW, Semerci A, Cobb N. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For Ecol Manage.* 2010;259:660–84.
- Atkin OK, Bruhn D, Hurry VM, Tjoelker MG. The hot and the cold: unravelling the variable response of plant respiration to temperature. *Funct Plant Biol.* 2005;32:87–105.
- Barnabas B, Jager K, Feher A. The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 2008;31:11–38.
- Bonan GB. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science.* 2008;320:1444–9.
- Chaves MM, Maroco JP, Pereira JS. Understanding plant responses to drought - from genes to the whole plant. *Funct Plant Biol.* 2003;30:239–64.
- Ciais P, Sabine C, Bala G, Bopp L, Brovkin V, Canadell J, Chhabra A, DeFries R, Galloway J, Heimann M, Jones C, Le Quéré C, Myneni RB, Piao S, Thornton P. Carbon and other biogeochemical cycles. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, editors. *Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change.* Cambridge, UK/New York: Cambridge University Press; 2013.
- Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichetef T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver AJ, Wehner M. Long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, editors. *Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change.* Cambridge, UK/-New York: Cambridge University Press; 2013.
- Easterling W, Aggarwal P, Batima P, Brander K, Erda L, Howden S, Kirilenko A, Morton J, Soussana J-F, Schmidhuber J, Tubiello F. Food, fibre and forest products. In: Parry M, Canziani O, Palutikof J, van der Linden PJ, Hanson CE, editors. *Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the fourth assessment report of the intergovernmental panel on climate change.* Cambridge, UK: Cambridge University Press; 2007. p. 273–313.

- Fuhrer J. Ozone risk for crops and pastures in present and future climates. *Naturwissenschaften*. 2009;96:173–94.
- Leakey ADB, Lau JA. Evolutionary context for understanding and manipulating plant responses to past, present and future atmospheric CO<sub>2</sub>. *Philos Trans R Soc B-Biol Sci*. 2012;367:613–29.
- Leakey ADB, Ainsworth EA, Bernacchi CJ, Rogers A, Long SP, Ort DR. Elevated CO<sub>2</sub> effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *J Exp Bot*. 2009;60:2859–76.
- McDowell N, Pockman WT, Allen CD, Breshears DD, Cobb N, Kolb T, Plaut J, Sperry J, West A, Williams DG, Yezzer EA. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytol*. 2008;178:719–39.
- Mittler R, Blumwald E. Genetic engineering for modern agriculture: challenges and perspectives. In: Merchant SBWROD, editor. *Annual review of plant biology*, vol. 61. Palo Alto: Annual Reviews; 2010. p. 443–62.
- Mittler R, Vanderauwera S, Gollery M, Van Breusegem F. Reactive oxygen gene network of plants. *Trends Plant Sci*. 2004;9:490–8.
- Norby RJ, Zak DR. Ecological lessons from free-Air CO<sub>2</sub> enrichment (FACE) experiments. In: Futuyma DJ, Shaffer HB, Simberloff D, editors. *Annual review of ecology, evolution, and systematics*, vol. 42. Palo Alto: Annual Reviews; 2011. p. 181.
- Norby RJ, DeLucia EH, Gielen B, Calfapietra C, Giardina CP, King JS, Ledford J, McCarthy HR, Moore DJP, Ceulemans R, De Angelis P, Finzi AC, Karnosky DF, Kubiske ME, Lukac M, Pregitzer KS, Scarascia-Mugnozza GE, Schlesinger WH, Oren R. Forest response to elevated CO<sub>2</sub> is conserved across a broad range of productivity. *Proc Natl Acad Sci U S A*. 2005;102:18052–6.
- Sage RF, Kubien DS. The temperature response of C-3 and C-4 photosynthesis. *Plant Cell Environ*. 2007;30:1086–106.
- Tubiello FN, Soussana JF, Howden SM. Crop and pasture response to climate change. *Proc Natl Acad Sci U S A*. 2007;104:19686–90.
- Wilkinson S, Davies WJ. Drought, ozone, ABA and ethylene: new insights from cell to plant to community. *Plant Cell Environ*. 2010;33:510–25.

---

# Plant Influences on Atmospheric Chemistry 19

Christine Wiedinmyer, Allison Steiner, and Kirsti Ashworth

## Contents

Introduction .....	574
Emissions to the Atmosphere from Plants .....	575
Biogenic VOC Emissions .....	575
Moving VOC from the Leaf into the Atmosphere .....	582
Transport Versus Chemistry .....	583
In- and Above-Canopy Turbulent Transport .....	584
Top-of-the-Canopy Fluxes .....	586
Transport in the Atmospheric Boundary Layer .....	587
Chemistry in the Troposphere .....	587
Gas-Phase Chemistry .....	588
Atmospheric Particles .....	592
Impacts on Air Quality and Climate .....	593
Climate .....	593
Air Quality .....	594
The Climate-Air Quality Conflict .....	596
Future Directions .....	597
References .....	597

---

C. Wiedinmyer (✉)

Atmospheric Chemistry Division, NCAR Earth System Laboratory, National Center for  
Atmospheric Research, Boulder, CO, USA

e-mail: [christin@ucar.edu](mailto:christin@ucar.edu)

A. Steiner

Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor,  
MI, USA

e-mail: [alsteiner@umich.edu](mailto:alsteiner@umich.edu)

K. Ashworth

Ecosystems-Atmosphere Interactions Group, Karlsruhe Institute of Technology, Garmisch-  
Partenkirchen, Germany

Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor,  
MI, USA

e-mail: [kirsti.ashworth@kit.edu](mailto:kirsti.ashworth@kit.edu)

---

**Abstract**

- Vegetation emits significant amounts of reactive gases, known as biogenic emissions, to the atmosphere.
- The most prevalent biogenic emission from plants is isoprene ( $C_5H_8$ ), but plants emit a broad suite of chemical compounds.
- Not all biogenic emissions released into a canopy reach the atmosphere because some react within the canopy or deposit onto vegetation; therefore, understanding the canopy transport is key to explaining atmospheric concentrations of these gases.
- Biogenic VOC emissions can play an important role in atmospheric chemistry and climate by impacting the concentrations of air pollutants, chemical radicals, and greenhouse gases in the atmosphere.

---

**Introduction**

Have you ever walked through a forest and noticed that “pine forest” smell? What you smell are trace gases released from the forest plants into the atmosphere. These gases are known as *biogenic* emissions or emissions released to the atmosphere from biological sources. Trace gases, such as volatile organic compounds (VOCs) and the oxides of nitrogen ( $NO_x$ ), are emitted to the atmosphere from organisms through a variety of biophysical and biochemical processes and can play an important role in local, regional, and even global atmospheric chemistry and climate.

In the mid-twentieth century, scientists began to recognize the importance of biogenic emissions to the physical states and chemical processes of the atmosphere. Went (1960) presented evidence that plants emitted organic compounds to the atmosphere, and further hypothesized that the blue haze observed in rural regions, such as over the Blue Ridge Mountains in the eastern United States, is the result of biogenically released compounds that have reacted and condensed to form atmospheric particles. Rasmussen (1970, 1972) began to identify specific organic compounds that were emitted from plant and other organism sources rather than anthropogenic (human-made) sources. Since that time, advances in measurement technologies have enabled the detection and identification of hundreds of chemical compounds that are emitted from vegetation to the atmosphere. Some of these compounds are important to atmospheric chemistry, air quality, and climate due to the magnitude of their emissions and/or their reactivity with respect to other chemical species. For example, Chameides et al. (1988) provided the first quantitative study to show the importance of biogenic VOC emissions for the production of ozone (aka photochemical smog) in the southeastern United States. Many studies have shown that controls on anthropogenic sources of pollution may be ineffective, or even counterproductive, unless biogenic VOC emissions are considered. Therefore, the understanding and quantification of biogenic emissions are critical for the development of accurate models of the chemistry of our atmosphere, air quality, and climate. In this chapter, emissions of biogenic compounds,

their exchange between the biosphere and the atmosphere, and their impacts on atmospheric chemistry and climate are explored.

---

## Emissions to the Atmosphere from Plants

Because biogenic emissions are so prevalent in the Earth System, it is critical to constrain the magnitude of such emissions and identify the various compounds that are emitted. The spatial and temporal variability in emissions is substantial, since emissions are dictated by the vegetation type and the physiological state of the vegetation, which are sensitive to seasonal and longer-term variation in weather and climate. Because they are so variable in time and space, one of the great challenges to assessing the impact of biogenic VOC emissions on the atmosphere is to accurately quantify emissions in a way that can be adjusted to various spatiotemporal scales within the Earth System.

### Biogenic VOC Emissions

Biogenic emissions include a variety of VOCs and other “inorganic” trace gases, such as nitrogen oxides (NO<sub>x</sub>). Biogenic sources dominate VOC emissions. Globally, plants emit an estimated 1,000 Tg VOC year<sup>-1</sup>. This is approximately *ten times more* than the total amount of VOC emitted worldwide from anthropogenic sources including fossil fuel combustion and industrial sources (Warneck 2000). Thus, one of the initial important concepts to establish is that biogenic emissions from the plants, microorganisms, and animals in ecosystems are far greater in terms of controlling atmospheric states and processes, compared to human-generated VOCs. An important clarification must be emphasized here. When we discuss the topic of biogenic VOCs and influences on the atmosphere, we are not including CO<sub>2</sub>, which is an inorganic compound. We are also not including methane (CH<sub>4</sub>), in our discussion. Methane is an organic compound and, thus, a legitimate component of what one might refer to as “biogenic VOCs.” However, CH<sub>4</sub> is principally produced from anaerobic soils in wetland ecosystems and from emissions from the enteric bacteria of ruminant animals. While plants can act as important conduits for the transport of soil-derived CH<sub>4</sub> to the atmosphere, they are not the primary source of CH<sub>4</sub> production. Neither the soil nor the ruminant animals fit within the frame of reference of this book and chapter, which focuses on plant processes. Thus, for the remainder of this chapter, the focus will be on important “non-methane” biogenic emissions of VOC.

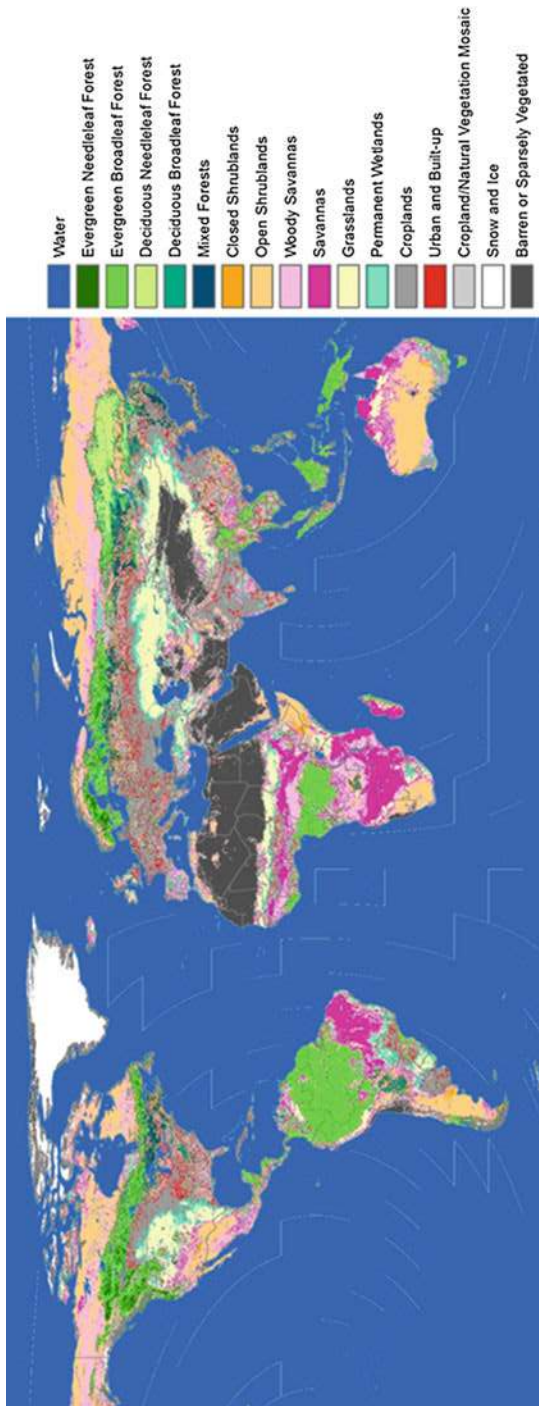
There are many VOC species that are emitted from vegetation. Table 1 lists the most prevalent biogenic gaseous emissions and their estimated global annual emission rates. The most dominant biogenic VOC emission is the unsaturated hydrocarbon isoprene (C<sub>5</sub>H<sub>8</sub>). Isoprene has two double carbon bonds and is therefore very reactive in the atmosphere. Recent estimates predict that isoprene emissions from vegetation globally are on the order of 500–600 Tg year<sup>-1</sup>, more than

**Table 1** Annual global emissions of biogenic compounds (Adapted from Guenther et al. (2012))

Compound class	Compound	Emissions (Tg year <sup>-1</sup> )
<i>Isoprene</i>	Isoprene	535
<i>Monoterpenes</i>	a-Pinene	66
	t-b-Ocimene	19
	b-Pinene	19
	Limonene	11
	Sabinene	9
	Myrcene	9
	3-carene	7
	Camphene	4
	Other monoterpenes	18
	<i>Sesquiterpenes (SQT)</i>	a-Farnesene
b-Caryophyllene		7
b-Farnesene		4
Other sesquiterpenes		11
<i>Oxygenated VOC</i>	2-3-2 methyl butanol (MBO)	2
	Methanol	100
	Acetone	44
	Ethanol	21
<i>Bidirectional VOC</i>	Acetaldehyde	21
	Formaldehyde	5
	Acetic acid	4
	Formic acid	4
<i>Stress VOC</i>	Ethene	27
	cis-3-hexenol	5
	Other stress VOC	16
<i>Other VOC</i>	Propene	16
	Butene	8
	Other VOC	8
<i>Carbon monoxide (CO)</i>		82
	<b>Total (VOC and CO)</b>	<b>1,087</b>

half of the total biogenic VOC emissions (Guenther et al. 2012). Other commonly emitted compounds are monoterpenes (compounds containing 10 carbons, C<sub>10</sub>H<sub>16</sub>) and sesquiterpenes (compounds containing 15 carbons). Biogenic VOC emissions include oxygenated compounds, alkanes, alkenes, and acidic compounds (Table 1).

In general, the spatial distribution of the emissions closely follows the spatial distribution of vegetation on the globe. Figure 1 shows the global land cover and land use distributions as observed by satellite instruments from space. Biogenic emissions are closely aligned with these types of global vegetation maps. The specific types, as well as the quantity of volatile organic compounds produced by ecosystems, are highly dependent on distributions of plant species and growth form. For example, most oak trees (*Quercus*) emit isoprene at high rates; however, pine



**Fig. 1** Global land cover and land use, as defined by the MODIS Land Cover-Type Product ([http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD\\_NUMBER=12](http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD_NUMBER=12))

trees (*Pinus*) do not emit isoprene, but they do emit monoterpenes. Isoprene can be emitted in large quantities from areas with tropical forests and deciduous hardwood forests. Monoterpene emissions are largely emitted from areas where boreal or temperate coniferous species dominate the ecosystems. This is reflected in the maps of biogenic VOC emissions shown in Fig. 2.

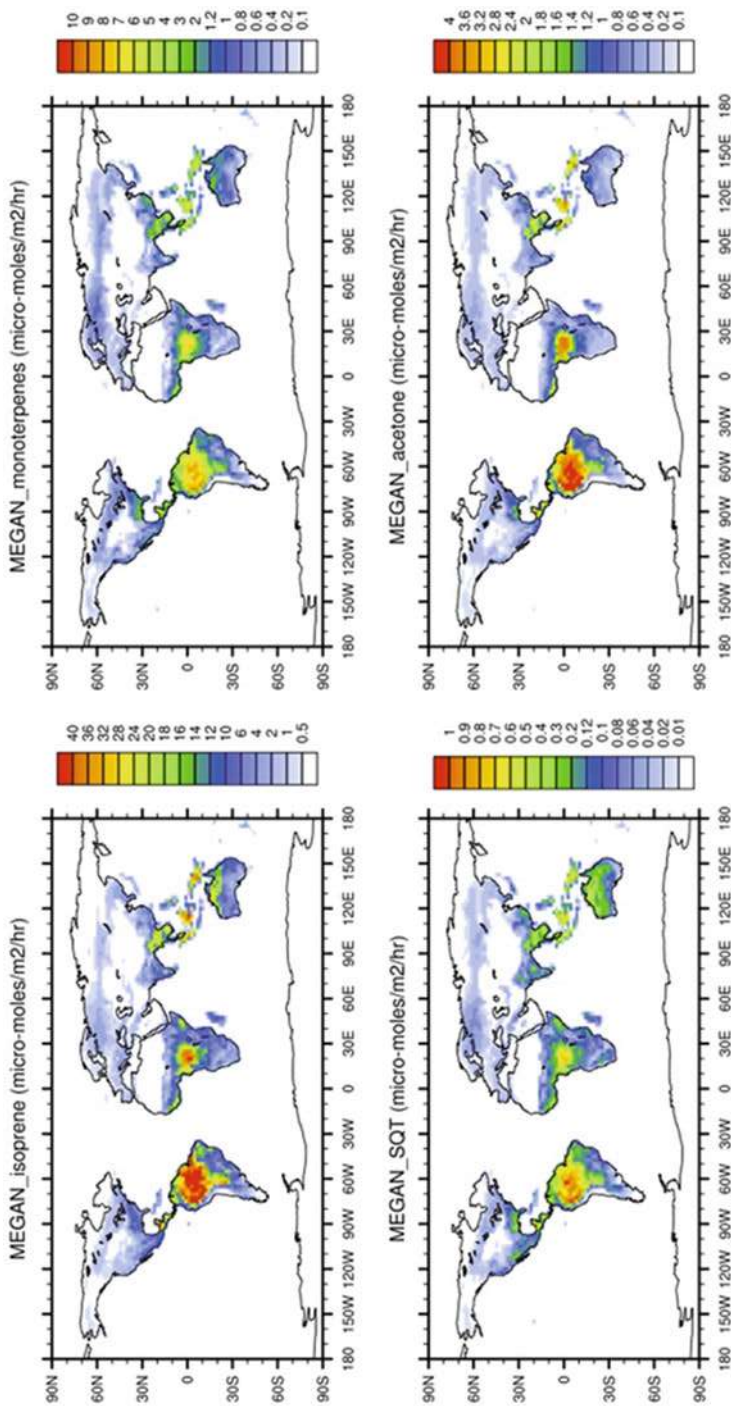
The mechanisms by which VOCs are produced and emitted also vary. Observations have shown that some biogenic VOCs, such as isoprene and some monoterpenes, are emitted as a function of environmental conditions that change on short time scales, most importantly temperature and light. The time scale for the response of these emissions is similar to that of photosynthesis, and in fact, these emissions have been shown to be metabolically connected to photosynthesis through various processes in the chloroplast. Other controls, such as leaf age and leaf area index (LAI), can impact emissions but to a smaller extent. The reasons for the emissions of biogenic VOCs vary and are in some cases not fully understood. For example, several hypotheses exist to explain the ultimate aspects of natural selection that have led to the evolution of isoprene emissions from leaves. One hypothesis is that isoprene emission provides protection from elevated temperatures or from high levels of atmospheric oxidants like ozone (Sharkey et al. 2008). Other compounds are emitted as a response to insect attack or other abiotic or biotic stresses. These stresses include light intensity, temperature, moisture availability, and exposure to ozone pollution and insect attack. Monoterpene production in leaves and needles has been attributed to protection from abiotic stresses, similar to the case for isoprene, in some species, and to protection from insect herbivory in other species. Plants may produce monoterpenes in different tissues and store them at different levels, depending on these variable adaptive roles. In some leaves, monoterpenes are produced in chloroplasts and not stored, rendering them susceptible to immediate “leakage” to the atmosphere; these compounds are thought to be most effective at protecting leaves from abiotic stresses such as extreme heat, light, and drought. In other leaves, particularly the needles of coniferous species, monoterpenes are produced in the cells of resin ducts and blisters and are stored as a means of deterring insect consumption; these compounds leak more slowly to the atmosphere and are thought to be most effective at protecting leaves from the biotic stress of herbivory.

To estimate the quantity of emissions, particularly for atmospheric chemistry and climate applications, biogenic VOC emissions are commonly represented by Eq. 1:

$$E_i = EF_i * \gamma_i \quad (1)$$

where  $E_i$  is the emission of compound  $i$  ( $\text{mass area}^{-1} \text{time}^{-1}$ ),  $EF_i$  is the potential emission rate of compound  $i$  at a set of standard conditions, and  $\gamma_i$  is an activity factor that accounts for all environmental and phenological variables that control the emissions.  $EF_i$  is also known as an *emission factor* and its value can be a function of a specific plant genus or ecosystem type. Table 2 shows the emission factors of isoprene and some selected monoterpenes for several specific tree and





**Fig. 2** Global annually averaged emission rate estimates ( $\text{mmoles compound m}^{-2} \text{ h}^{-1}$ ) of several important biogenic VOC species for using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) v2.1 and the Community Land Model version 4 (Guenther et al. 2012). [SQT sesquiterpenes]. Note the different scales for each figure

**Table 2** Emission factors ( $\text{mg compound m}^{-2} \text{ h}^{-1}$ ) of selected compounds from different plant functional types (Guenther et al. 2012)

Plant functional type (PFT)	Isoprene	Limonene	3-carene	t- $\beta$ -Ocimene	$\beta$ -Pinene	a-Pinene	Other		
							monoterpenes	$\beta$ -Caryophyllene sesquiterpenes	
Needleleaf evergreen temperate tree	600	100	160	70	300	500	180	80	120
Needleleaf evergreen boreal tree	3,000	100	160	70	300	500	180	80	120
Needleleaf deciduous boreal tree	1	130	80	60	200	510	170	80	120
Broadleaf evergreen tropical tree	7,000	80	40	150	120	600	150	60	120
Broadleaf evergreen temperate tree	10,000	80	30	120	130	400	150	40	100
Broadleaf deciduous tropical tree	7,000	80	40	150	120	600	150	60	120
Broadleaf deciduous temperate tree	10,000	80	30	120	130	400	150	40	100
Broadleaf deciduous boreal tree	11,000	80	30	120	130	400	150	40	100
Broadleaf evergreen temperate shrub	2,000	60	30	90	100	200	110	50	100
Broadleaf deciduous temperate shrub	4,000	100	100	150	150	300	200	50	100
Broadleaf deciduous boreal shrub	4,000	60	30	90	100	200	110	50	100
Arctic C3 grass	1,600	0.7	0.3	2	1.5	2	5	1	2
Cool C3 grass	800	0.7	0.3	2	1.5	2	5	1	2
Warm C4 grass	200	0.7	0.3	2	1.5	2	5	1	2
Crop	1	0.7	0.3	2	1.5	2	5	4	2



**Fig. 3** Photos of leaf enclosure measurements in the laboratory and in the field

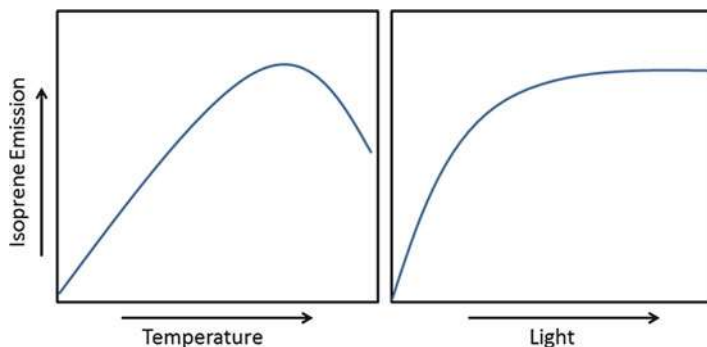
ecosystem types. As noted, different vegetation species emit different compounds and in different quantities. The emission factors of isoprene are much higher than those of other biogenic emission species. As shown here and mentioned previously, forested ecosystems, particularly those dominated by broadleaf trees, have the highest isoprene emissions. Crops and grasses have the lowest isoprene emission factors.

The emission factors in Table 2 and published elsewhere have been developed from laboratory and field measurements of leaf and branch enclosures, as well as above-canopy flux measurements. The photos of Fig. 3 show examples of leaf enclosure measurements in the laboratory and field. For these types of measurements, the concentrations of biogenic VOCs are measured in the inlet and outlet of the enclosure, and an emission factor is developed based on the increase in outlet concentrations and the mass of plant material in the enclosure.

The activity factor ( $\gamma_i$ ) in Eq. 1 represents the various controls that can regulate emissions of a specific compound (Guenther et al. 2012). This parameter includes emission response to light ( $\gamma_P$ ), temperature ( $\gamma_T$ ), leaf age ( $\gamma_A$ ), soil moisture ( $\gamma_{SM}$ ), leaf area index (LAI), and atmospheric  $\text{CO}_2$  concentrations ( $\gamma_C$ ) as

$$\gamma_i = \text{LAI} * \gamma_{P,i} * \gamma_{T,i} * \gamma_{A,i} * \gamma_{SM,i} * \gamma_{C,i} \quad (2)$$

The responses to various environmental and ecological conditions, or the individual gamma ( $\gamma$ ) values, are also dependent on the type of emitted VOC compound. Controls on isoprene emissions are dominated by leaf temperature and light exposure. Isoprene is not emitted during the nighttime when it is dark.



**Fig. 4** Schematic of isoprene emissions as a function of temperature and light

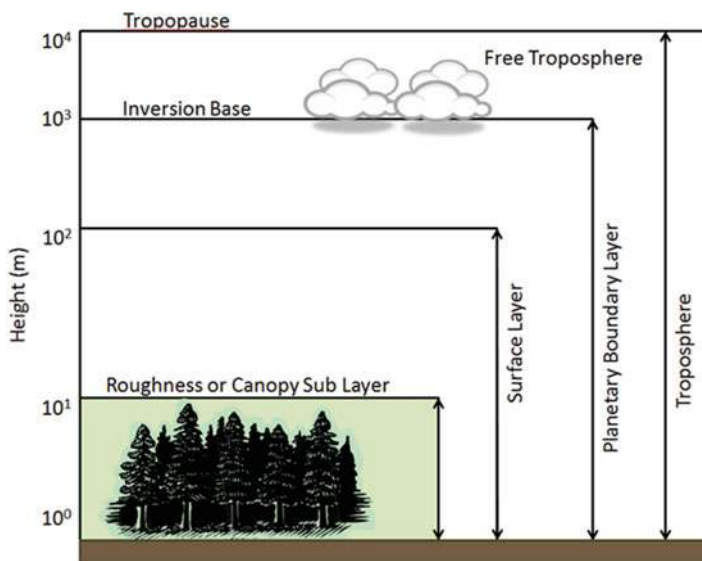
Measurements from enclosures, such as those shown in Fig. 3, can be used to evaluate the controls (e.g., light and temperature) on the emission rate. Based on laboratory and field measurements, Guenther et al. (1991) developed empirical algorithms that describe the dependency of isoprene emissions on light and temperature. These equations are still used today to predict emissions from plants. Isoprene emissions increase with increasing temperature until a maximum temperature is reached (typically  $\sim 40$  °C). At temperatures above this point, emissions decrease. Emissions also increase with increasing sunlight until a saturation point is reached, after which no further increase in emissions is observed. Figure 4 illustrates the light and temperature dependencies of isoprene emissions.

Unlike isoprene, the emissions of monoterpenes from stored reserves in resin canals and resin blisters are less influenced by available sunlight. The emissions of monoterpenes that are not produced in chloroplasts and are not stored in leaves exhibit light and temperature dependencies that are similar to those for isoprene. Emissions of monoterpenes from storage reservoirs are driven primarily by temperatures, increasing as temperatures increase. Other compounds that are emitted from vegetation include oxygenated compounds, such as ethanol and methanol. Some compounds can be both emitted from various plants and also taken up by the plants, such as acetaldehyde, formaldehyde, acetic acid, and formic acid. Therefore, they have *bidirectional* fluxes, which are dependent on the atmospheric concentrations of the specific compounds.

---

## Moving VOC from the Leaf into the Atmosphere

Once emitted from a plant, biogenic VOCs are transported from the point of emission (usually the leaf) into the canopy air space, out of the canopy, and into the lower troposphere where they can impact atmospheric chemistry and climate. Figure 5 illustrates the layers of the atmosphere with respect to biogenic VOC emission. Because the movement of these biogenic VOC molecules from the canopy layer into the planetary boundary layer determines chemical concentrations



**Fig. 5** A schematic of the layers of the troposphere relevant for biogenic VOC (Adapted from Arya (2001))

in the atmosphere, the transport of biogenic compounds is an important component to understanding and simulating the chemistry of the atmosphere. The complexity of the transport process depends on several factors, including the chemical reactivity of the individual VOC species, the height and structure of the canopy, and the influence of the canopy on meteorological conditions and turbulence in the canopy. Because the majority of biogenic VOC mass is emitted from forest canopies, this section will focus on the complexities of turbulence in a forest canopy and its impact on biogenic VOC transport.

## Transport Versus Chemistry

Biogenic VOC molecules are transported through the atmosphere via molecular diffusion, eddy transport generated by turbulence near the surface, or advection (mean wind flow). Here, we define turbulent transport of biogenic VOC as the movement of constituents due to the turbulent motions of air, often described as eddy transport. Molecular diffusion plays an important role in moving molecules out of the leaf, but once in the canopy air space, the dominant driver of motion to the atmosphere above the canopy is turbulent transport. Turbulent eddy motion is far more efficient for moving molecules large distances and is the dominant process in the canopy sublayer and the atmospheric surface layer. Advection, or the general circulation of the atmosphere, becomes increasingly important in the atmospheric boundary layer.

The transport of biogenic VOC within a canopy and out into the atmosphere is further complicated by the fast chemical reactivity of some biogenic VOCs

(e.g., isoprene and sesquiterpenes). The reactivity of each type of biogenic VOC is defined in terms of the “lifetime” of the compound ( $\tau$ ), which refers to the average length of time that a molecule will reside in the atmosphere before engaging in a chemical reaction that changes its chemical structure. Atmospheric lifetimes reflect the balance between compound emission rate and chemical reaction rate, which in turn depends on factors such as the concentrations of reactants, temperature, and the presence of catalytic surfaces that can reduce reaction activation energies. Generally, the local lifetime of isoprene ranges from a few hours to a few days (Fuentes et al. 2000), the lifetime of monoterpenes is on the order of minutes to hours, and sesquiterpenes, more complex molecules with multiple double bonds, have a shorter reactivity time of seconds to minutes. As a result, compounds with the shortest lifetimes, such as sesquiterpenes, have the potential for reaction within the canopy air space and thus may be unable to escape the canopy and enter the planetary boundary layer.

One metric to estimate the relative importance of the reactivity to the atmospheric transport is the Damköhler number, representing the ratio of the chemical lifetime of the compound to the transport time out of the forest canopy. If this ratio is low, it indicates that chemical reaction times are much longer than transport times, and most of the emitted species will be transported out of the canopy. However, as this number approaches and exceeds unity, then the chemical reactions occurring within the canopy are faster than the mean vertical canopy transport time and the compound may not be emitted to the atmosphere above the canopy. Additionally, a Damköhler number near or exceeding one also indicates that there will likely be spatial and temporal inhomogeneities of biogenic VOC within the forest canopy. These inhomogeneities in biogenic VOC concentrations, as well as the concentrations of the radicals that drive chemical reactions, can effectively lower reaction rates, a process known as *segregation* (Dlugi et al. 2010). Therefore, understanding the relative roles of transport and atmospheric chemistry is important for understanding fluxes out of the top of a forest canopy and will vary depending on the biogenic VOC in question.

The Damköhler number varies as a function of canopy structure and meteorological conditions. For example, if we assume an average canopy residence time of 3 min, and a chemical lifetime of 84 min for isoprene (Table 3; isoprene + OH reaction), the Damköhler number would be 0.04, indicating that most of the isoprene will be transported to the surface layer. However, at nighttime when canopy residence times lengthen (e.g., 10 min), a more reactive compound such as terpinolene (a sesquiterpene with a chemical lifetime of 1 min with  $\text{NO}_3$ ; Table 3) would yield a Damköhler number of 10, indicating that most nighttime sesquiterpene emissions will react before leaving the canopy.

## In- and Above-Canopy Turbulent Transport

Quantifying within-canopy turbulence can be challenging, and our current understanding is predominantly based on field observations and high-resolution

**Table 3** Calculated atmospheric lifetimes ( $\tau$ ) of selected biogenic VOCs with OH, NO<sub>3</sub>, and O<sub>3</sub> (Rate constants from Warneck and Williams (2012)). The atmospheric concentrations of OH, NO<sub>3</sub>, and O<sub>3</sub> at which the lifetimes were calculated are provided at the bottom of the Table

Compound	OH	NO <sub>3</sub>	O <sub>3</sub>
Isoprene	1.4 h	48 min	1.3 days
a-Pinene	2.7 h	5 min	4.7 h
t-b-Ocimene	37 min	2 min	44 min
b-Pinene	1.9 h	13 min	1.1 days
Limonene	51 min	3 min	1.9 h
Sabinene	1.2 h	3 min	4.8 h
Myrcene	39 min	3 min	51 min
3-carene	1.6 h	4 min	11 h
Camphene	2.6 h	51 min	18 days
b-Phellandrene	50 min	4 min	8.4 h
Terpinolene	37 min	21 s	13 min
b-Caryophyllene	42 min	2 min	2 min
a-Humulene	28 min	1 min	2 min
Methanol	6 days	178 days	> 4.5 year
<i>Atmospheric lifetimes based on the following concentrations (molec cm<sup>-3</sup>):</i>			
[OH] = 2.0 × 10 <sup>6</sup>	[NO <sub>3</sub> ] = 5.0 × 10 <sup>8</sup>	[O <sub>3</sub> ] = 7.0 × 10 <sup>11</sup>	

large-eddy simulation modeling. Typically, winds decrease toward the surface in a vegetated forest canopy, slowing turbulent motions. However, processes in the canopy air space can drive secondary circulations that can be important for overall fluxes out of the top of the canopy. For example, in forests with little to no understory, winds can develop that may increase the movement of biogenic VOC within the canopy. In-canopy heating by incoming solar radiation can also generate additional in-canopy turbulence; therefore, the density and structure of the vegetation within the forest canopy can play a role in the turbulent transport of biogenic VOC. In general, sub-canopy flow and turbulence and its impact on biogenic VOC are very site-specific and depend on the overall forest canopy structure. As the wind flows around leaves, branches, and stems of plants, swirling currents of air occur as “wakes” on the downwind side. These local areas of turbulent wakes can potentially act as “reactor volumes,” increasing the time during which reactants can interact and thus enhancing the Damköhler number. Studies of within-canopy reactions and the various processes that affect the reaction and transport rates are still rudimentary and require further investigation.

As in-canopy turbulence can be important for understanding how biogenic emissions move from the plant and within the forest canopy air space, biogenic VOCs must also be transported from the forest canopy air space into the surface layer of the atmosphere. For this transport to occur, the VOC molecules must move through the lowest part of the atmosphere that interacts with the vegetation, which is frequently defined as the “roughness layer” or the “canopy sublayer” (Fig. 5). The interface between the canopy and the atmosphere represents a region of high wind



shear, where horizontal wind flows can be disrupted and create intermittent turbulent air motions that aid the transport of biogenic VOC. There is increasing evidence that much of this turbulent transport occurs through the mechanism of coherent wind structures (Finnigan 2000). Coherent wind structures are defined as distinct patterns of turbulence that occur at regular intervals and are described by two types of motion: (1) A “burst” or ejection of air from within the canopy to the atmosphere (representing upward motion) and (2) a “sweep” of air that brings air from the atmosphere into the forest canopy. These bursts and sweeps are due to instabilities in the air flow caused by the large differences in horizontal wind speeds near the top of the canopy. This can be visualized as a type of intermittent canopy “venting.”

Coherent structures, such as the sweeps and bursts, occur on time scales of seconds to minutes and are an important factor in the flux of biogenic emissions in and out of a forest canopy. While the role of coherent structures on the transport of biogenic VOC has yet to be quantified, results of studies on the transport of other trace gases suggest that biogenic VOCs are likely to be carried along with coherent structures and, depending on their chemical reactivity, vented to the atmosphere. Therefore, identifying these structures and quantifying their contribution relative to within-canopy reaction rate are key to understanding biosphere-atmosphere exchange.

## Top-of-the-Canopy Fluxes

The flux out of the top of the canopy into the planetary boundary layer represents the mass flux of biogenic VOC to the atmosphere, which is the most important emission metric for determining the role of biogenic VOC on atmospheric chemistry and climate. Biogenic VOC flux is defined as the mass of carbon (or compound) per area per time and can be measured in the field with several different techniques. Some studies have measured the fluxes of biogenic VOC at the leaf or branch level, where a leaf or branch is enclosed in a chamber and the flux can be quantified by measuring the flow and input and output concentrations (e.g., Fig. 3). These results must then be scaled with the biomass within the enclosure to represent the full canopy.

In addition to branch enclosure methods, micrometeorological methods are frequently employed to measure fluxes out of the canopy. Micrometeorological methods use high time resolution measurements of wind speed, including the turbulent and advective wind components, to estimate transport. The two most commonly used micrometeorological methods for biogenic VOC flux estimation are relaxed eddy accumulation (REA) and eddy covariance (EC). The REA method collects air samples at the top of the canopy in updrafts and downdrafts of the wind to determine a “top-of-the-canopy” flux. The EC method uses fast-response time measurements (e.g., 1–10 measurements per second) to derive fluxes as the statistical covariance between the turbulent wind speed and the time-dependent variance in VOC concentration. The EC approach is similar to



techniques implemented to measure surface energy fluxes (Foken 2008). Typically, the REA method is used when high-response chemical sampling of the fast fluctuations of biogenic VOC concentrations is unavailable. EC measurements of top-of-the-canopy fluxes of biogenic VOC are becoming more common in field sampling due to newer measurement techniques. The EC method is also advantageous to determine the role of coherent structure transport on the top-of-canopy fluxes of biogenic VOC, as the fast-response time measurements can indicate when coherent structures are present.

An additional metric often used to represent the fluxes of biogenic VOC out of the forest canopy is the escape efficiency (Stroud et al. 2005). The escape efficiency is defined as the fraction of the mass flux of biogenic VOC transported to the atmospheric boundary layer as compared to the mass flux emitted from vegetation. An escape efficiency of one therefore indicates that all biogenic VOC that is emitted is mixed into the atmosphere. Stroud et al. (2005) show that this escape efficiency is high (0.9) for less reactive species (e.g., isoprene and  $\alpha$ -pinene) but low (0.3) for  $\beta$ -caryophyllene (a sesquiterpene). This method has been employed in models to scale top-of-the-canopy flux estimates by removing the effect of in-canopy chemistry, which may reduce the source emissions of some very reactive biogenic VOCs.

## Transport in the Atmospheric Boundary Layer

If biogenic VOCs are transported out of the forest canopy and into the atmospheric surface layer without reacting, they will continue to be mixed upward and potentially into the free troposphere (Fig. 5). The fate of biogenic VOC is subject to the vertical mixing that occurs within the atmospheric boundary layer (typically 1–2 km under daytime conditions). Under sunny, daytime conditions, biogenic VOC will be transported with the large-scale atmospheric eddies that typically range in size from meters to kilometers and can be as large as the boundary layer height itself. Once into the boundary layer, the biogenic VOCs can impact atmospheric chemistry, air quality, and climate via various chemical pathways.

---

## Chemistry in the Troposphere

Once in the free troposphere, the chemistry of emitted VOCs is complex. Although 99.9 % of the atmosphere is composed of three compounds ( $N_2$ ,  $O_2$ , and water), it is the presence of the various trace components comprising the remaining 0.1 % that results in the changing chemical composition of the atmosphere. Depending on the emitted species and the background chemical composition of the air, the impact of biogenic VOC can act over different spatial scales (local, regional, or global) and different time scales from fractions of seconds to many centuries.

As shown in Table 1, the number of VOCs emitted from vegetation is large and the structure of these compounds is highly variable. While the ultimate fate of these emitted chemicals is to be deposited back to the terrestrial or marine land surface or broken

down into  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , the rate at which this occurs varies widely among compounds, and a diverse range of chemical by-products is also produced. Species with atmospheric lifetimes of hours to days can be transported to other parts of a region or continent, whereas VOCs with longer lifetimes become well-mixed in the free troposphere and can be transported across global scales. The implications of such atmospheric transport are discussed in the section “[Impacts on Air Quality and Climate](#)”.

While the precise reaction pathways of each emitted compound are determined by their chemical structure, as well as the atmospheric concentrations of their reactants, some generalizations can be made. The remainder of this section focuses on the chemistry governing the production and loss of ozone and the formation of secondary organic aerosols (trace components of the atmosphere that are both climatically active compounds and air pollutants) in which biogenic VOCs play a major role. Integration of the topic of VOC emissions from plants, as discussed above, with that of VOC reactions in atmosphere, as discussed in the next section, provides the true nexus required to understand how plants affect atmospheric chemistry.

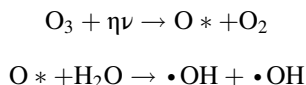
## Gas-Phase Chemistry

Generally, VOC emissions from plants are highly reactive, with atmospheric lifetimes on the order of seconds to hours. Once released into the atmosphere, biogenic VOCs react rapidly with atmospheric oxidants, primarily the hydroxyl (OH) and nitrate ( $\text{NO}_3$ ) radicals, and also ozone ( $\text{O}_3$ ) molecules. (It is important to note that hydroxyl and nitrate radicals are chemically different than hydroxide and nitrate ions. Free radicals contain one or more unpaired valence shell electrons and are thus highly reactive. It will be instructive to the student to explore the different chemical natures of radicals and ions. For example, see suggested reading by Seinfeld and Pandis (2006)). Table 3 shows the lifetimes of selected VOCs with typical atmospheric concentrations of OH,  $\text{NO}_3$ , and  $\text{O}_3$ . The reactions of biogenic VOC with these species produce secondary products that include  $\text{O}_3$ , stable organic nitrate compounds that can be transported for long distances, as well as low-volatility compounds that can condense to form particles in the atmosphere. These particles (also called aerosols) can remain suspended in the atmosphere for relatively long periods of time.

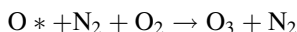
In the troposphere (Fig. 5), ozone ( $\text{O}_3$ ) is a pollutant and it can be harmful to human health, plants, and other man-made materials. (Tropospheric ozone is different in its ramifications for life on earth than stratospheric ozone. Stratospheric ozone protects the DNA in cells from mutagenic ultraviolet radiation, whereas tropospheric ozone damages cells by causing oxidation of membranes, proteins, and nucleic acids.) Tropospheric ozone is also a strong greenhouse gas. Tropospheric ozone is produced primarily through photochemically initiated reactions involving oxides of nitrogen ( $\text{NO}_x$ ) and VOC, including biogenic VOC species. The downward transport of ozone from the stratosphere to the troposphere is an additional source of tropospheric ozone, this source is small in comparison to the rate of chemical production in the troposphere itself. The main sink for tropospheric ozone is chemical loss, but there is also a significant flux to the surface where it is lost by the process of dry deposition (Royal Society 2008).

Biogenic VOCs therefore play an important role in the determination of ozone concentrations in the troposphere. The series of reactions leading to the formation or destruction of ozone in the troposphere can be broken down into three distinct phases: initiation reactions, free radical reactions, and termination reactions.

The initiation reactions involve the formation of OH radicals (the primary reactant), which occurs predominantly via the photolysis of ozone itself. During photolysis reactions, molecules absorb sufficient energy from sunlight to break down into their constituent atoms and smaller molecules as follows:

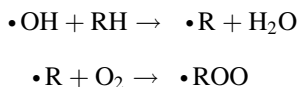


or

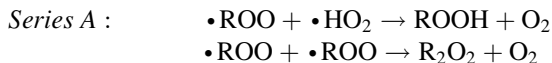


where  $\eta\nu$  denotes a photon of energy (i.e., from sunlight),  $\text{O}^*$  is an energetically excited oxygen atom,  $\bullet\text{OH}$  is a free radical, and  $+ \text{N}_2$  represents an energy-transferring collision with any inert molecule.

The reaction path proceeds with a series of **initiation reactions** mainly through reactions of organic compounds with the OH radical that produce peroxy radicals. While the dominant sources of such peroxy radicals are reactions involving methane ( $\text{CH}_4$ ) and carbon monoxide (CO), biogenic VOCs also undergo such initiation reactions:



where R denotes a hydrocarbon chain and  $\bullet\text{ROO}$  is a peroxy radical. The reaction chain is effectively ended in the **termination reactions** when free radicals mutually react to form relatively stable compounds, although these reaction products themselves can then go on to react with OH radicals to form their own peroxy radicals ( $\bullet\text{ROO}$ ). A general example of such termination reactions is shown in *Series A*:



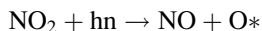
Alternatively, peroxy radicals can react with nitrogen oxide (NO) to produce stable molecules as shown in the general example *Series B*:



Thus, although the reaction chains are mostly initiated by the OH radical, the rate of chemical production and loss of ozone is governed by the termination

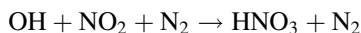
reactions that are followed by the peroxy radicals formed during the second reaction phase. This process is dependent on the concentration of  $\text{NO}_x$ . In very low  $\text{NO}_x$  environments, such as remote parts of the Southern Hemisphere and Pacific Ocean, the mutual termination reactions (*Series A*) predominate. As OH radicals are formed in the first instance through the photolysis of  $\text{O}_3$ , this sequence of reactions results in a net loss of tropospheric ozone.

At the moderate- $\text{NO}_x$  levels encountered over rural areas across much of the world, peroxy radical reactions with NO (*Series B*) predominate. Furthermore, the  $\text{NO}_2$  produced undergoes photolysis and breaks down into NO and  $\text{O}^*$ :



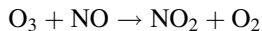
As the energetically excited oxygen atom can then react to form either new OH radicals or, more importantly,  $\text{O}_3$  molecules (as shown in the initiation reactions above), these regions are ozone producing. The rate of  $\text{O}_3$  production in such regions increases with increasing concentrations of  $\text{NO}_x$  but are relatively insensitive to changes in VOC emissions. Such regions are often described as “ $\text{NO}_x$ -limited” or “ $\text{NO}_x$ -sensitive” (Sillman 1999).

At even higher  $\text{NO}_x$  concentrations, for example in urban areas in industrialized or industrializing nations, the OH radical tends to react directly with  $\text{NO}_2$  to produce nitric acid ( $\text{HNO}_3$ ) as shown below:



When this reaction dominates the termination stage, insufficient  $\text{O}^*$  atoms are produced to outweigh the loss of  $\text{O}_3$  through photolysis, and the rate of  $\text{O}_3$  production declines. In such regions, an increase of hydrocarbons through VOC emissions increases the sink for OH, reducing the rate of  $\text{HNO}_3$  formation below the rate of  $\text{NO}_2$  photolysis. This results in an increased rate of  $\text{O}_3$  production, and these regions are often labeled “VOC-limited” or “VOC-sensitive” (Sillman 1999).

If  $\text{NO}_x$  concentrations rise further, a phenomenon known as “ $\text{NO}_x$  titration” occurs, and ozone concentrations fall as  $\text{O}_3$  reacts directly with NO to produce  $\text{NO}_2$  and  $\text{O}_2$  (Royal Society 2008):



As the  $\text{NO}_2$  produced from this reaction can subsequently reform NO and  $\text{O}_3$  through photolysis, there is a further consequence to such high levels of  $\text{NO}_x$ .  $\text{NO}_x$  titration leads to a rapid cycling of nitrogen and oxygen compounds and this effectively allows the  $\text{NO}_x$  to be transported away from the emission region (i.e., polluted urban environment) to regions of lower background  $\text{NO}_x$  levels, which may result in the enhanced formation of  $\text{O}_3$  downwind from the original  $\text{NO}_x$  emissions (Sillman 1999).

As well as governing the rate of production and loss of tropospheric ozone, the gas-phase reactions of biogenic VOCs play a key part in determining the atmospheric concentrations of a number of other gas-phase trace constituents of the atmosphere. Biogenic VOCs act as a major sink, particularly over land, for the OH radical, the atmosphere's most powerful oxidant. Emissions of biogenic VOCs thus mediate the oxidative capacity of the atmosphere, affecting the atmospheric lifetime of other chemical species, such as methane ( $\text{CH}_4$ ). Methane is oxidized in a similar set of reactions to those described above for non-methane VOCs. Thus, methane and other VOCs compete for hydroxyl radicals in the free troposphere. Simulations performed with atmospheric chemistry and transport computer models have demonstrated that including biogenic emissions of isoprene alone can increase the atmospheric lifetime of methane by up to 20 %, as compared to model simulations without isoprene (Forster et al. 2007). This is a result of direct competition for the OH radical; reactions with isoprene reduce the global OH budget by around 8 % in such simulations. Inclusion of biogenic methanol emissions results in similar impacts, though of lesser magnitude. Methanol is not only less reactive than isoprene, with an atmospheric lifetime ranging from a few days near the surface to a few weeks in the cold upper troposphere, but is also emitted in smaller quantities. Nevertheless, such emissions are sufficient to reduce the global average atmospheric concentration of the OH radical by around 2 %, thus further increasing the atmospheric lifetime of methane. As methane is a potent greenhouse gas, knowledge of the chemical reactions that affect its atmospheric lifetime and the ways in which the emissions of VOCs from plants can affect the lifetime are important issues to understand and to include in models of climate change.

Biogenic VOCs, and particularly isoprene, also play a key role in the distribution of reactive nitrogen (i.e., nitrogen that is available in a form that will readily react with other species rather than bound into long-lived stable virtually inert compounds such as  $\text{N}_2\text{O}$ ) in the atmosphere through the formation of organic nitrates, and in particular peroxyacetyl nitrate (PAN). PAN is a relatively long-lived compound, with an atmospheric lifetime of several months in the cold free troposphere. Vertical mixing lifts PAN from the boundary layer and lower troposphere, where it is formed from reactions involving peroxy radicals ( $\bullet\text{ROO}$ ) and  $\text{NO}_x$  to the free troposphere. Once there, its longevity allows it to be transported long distances before it is broken down by either thermal decomposition or photolysis, rereleasing reactive nitrogen. Thus, PAN acts to transport reactive nitrogen away from its source to other regions of the world. For some remote regions, the reactive nitrogen that is released from transported PAN (and other organic nitrates) can be the main source of  $\text{NO}_x$ . Atmospheric chemistry and transport model simulations show significant PAN increases in the remote tropics due to isoprene oxidation when biogenic isoprene emissions are included. The release of reactive nitrogen in such regions, where isoprene emissions are high and background levels of  $\text{NO}_x$  are low (i.e., “ $\text{NO}_x$ -sensitive regions”), can lead to enhanced ozone formation by shifting the region from a low- to a moderate- $\text{NO}_x$  regime, as outlined previously.

## Atmospheric Particles

In addition to the impact on ozone and other gas-phase constituents of the atmosphere described above, emissions of many VOCs from vegetation into the atmosphere affect the concentration of atmospheric particles or aerosols. The biosphere is a source of aerosols both directly through the release of particles such as pollen, plant detritus, bacteria, or spores, and indirectly as a result of the atmospheric reactions of gaseous compounds. The former are referred to as biogenic Primary Organic Aerosols (bPOA) and the latter as biogenic Secondary Organic Aerosols (bSOA). While bPOA are generally thought to be larger in size and, therefore, rapidly deposit back to the land or marine surface, bSOA are longer-lived, impacting the atmosphere via both chemical and physical pathways. Their respective atmospheric lifetimes are again reflected in the distances over which they can be transported and hence the impact they have on local and regional air quality and global climate (section “[Chemistry in the Troposphere](#)”).

Although the gas-phase reactions of biogenic VOCs are initiated through reactions with atmospheric oxidants to form peroxy radicals that go on to produce ozone, as outlined above, the products of these and subsequent reactions are often oxygenated species of lower volatility than the parent VOC. At sufficiently low volatility, these products can partition into the particle (or aerosol) phase, either through direct nucleation or by condensation onto existing particles (see, e.g., Hallquist et al. (2009) and references therein).

Detailed analyses of the composition of atmospheric aerosol have shown that the majority of their mass is biogenic in origin, even in highly polluted regions where urban anthropogenic emissions are dominant. However, the series of gas-phase reactions involved in SOA formation are complex and have not been fully elucidated for even the most common of VOCs. This is further complicated by the fact that VOCs and their products can also undergo reactions in the aerosol phase and participate in heterogeneous reactions (i.e., those that occur between compounds in the aerosol and gas phases). Knowledge of the processes of aerosol phase and heterogeneous chemistry and their controlling factors is even more limited than that of gas-phase atmospheric reactions (Hallquist et al. 2009). Our lack of understanding is clearly demonstrated by the mismatch between the magnitude and spatial distribution of SOA predicted by current theory and observations of aerosol concentration and composition (see, e.g., Spracklen et al. (2011)), although some of this lack of agreement is undoubtedly the result of the need to reduce and simplify the reactions included in most atmospheric chemistry models.

The biogenic VOCs that are emitted in the largest quantities, such as isoprene and methanol, as well as their reaction products, have very low yields of low-volatility condensable products and hence particles. In spite of their low yields, the magnitude of their emissions suggests they do contribute substantially to the total global SOA yield; but it is the longer-chained, and much more highly reactive (those with atmospheric lifetimes of seconds to minutes), biogenic VOCs, such as monoterpenes and sesquiterpenes, that are currently believed to have the highest yields of condensable products. Despite their low emission rate, the total

contribution of biogenic monoterpene and sesquiterpene emissions to the global SOA budget is of a similar order of magnitude as that of isoprene. While the spatial distribution of biogenic VOC emissions is highly heterogeneous, their reaction products and the SOA produced from biogenic compounds are much longer-lived (e.g., days to weeks) and can therefore become homogeneously mixed at the local to regional scale as they are transported long distances.

---

## Impacts on Air Quality and Climate

VOCs released from the terrestrial biosphere are for the most part emitted in such small quantities or have such short atmospheric lifetimes that they have virtually no direct impact on air quality or global climate. Most biogenic VOCs do not have absorption bands in the thermal parts of the electromagnetic spectrum and therefore do not contribute to the “greenhouse effect” by trapping earth-emitted radiation. However, as biogenic emissions play a key role in the regulation of tropospheric concentrations of ozone and particulate matter (see the section “[Chemistry in the Troposphere](#)”), their indirect impacts on both air quality and climate can be considerable. This section describes the effects of biogenic VOCs, ozone, and SOA, firstly on climate and then on regional or local air quality, and concludes with a reflection on the conflict between climate change drivers and air quality initiatives.

### Climate

Biogenic VOCs, in particular the terpenoids and other reactive species, have atmospheric lifetimes that are too short to directly affect global climate. Longer-lived species emitted from the terrestrial biosphere can be transported for long distances before reacting or decomposing and may survive long enough in the free troposphere to become well-mixed and ubiquitous in the atmosphere. However, their radiative forcing or global warming potentials and therefore climate impact are, as stated above, extremely low. The same is true of the organic gas-phase reaction products from biogenic VOCs.

By contrast, tropospheric ozone ( $O_3$ ) is a potent greenhouse gas. Estimates of its accumulated radiative forcing since preindustrial times place it third, behind only carbon dioxide and methane, in terms of contribution to anthropogenic global warming (see Fig. 1.1, Forster et al. (2007)). However, compared to both  $CO_2$  and  $CH_4$ ,  $O_3$  is short-lived, with an atmospheric lifetime ranging from a few days to several weeks in the upper troposphere. Ozone is therefore less well-mixed through the troposphere, and its climate impacts are regionally heterogeneous. As  $NO_x$  emissions in industrializing nations rise, it is to be expected that large areas of the tropics will be transformed from low- to moderate- $NO_x$  regimes. This will result in a considerable increase in  $O_3$  production from biogenic VOC reactions, likely to be sufficient to affect the climate in these regions. Furthermore, the gas-phase

atmospheric reactions of biogenic VOCs decrease the global budget of the OH radical (see section “[Moving VOC from the Leaf into the Atmosphere](#)”), resulting in higher atmospheric concentrations of other possible reactants, such as CH<sub>4</sub>. CH<sub>4</sub> is an important greenhouse gas. With fewer OH radicals available for reaction, the atmospheric lifetime of CH<sub>4</sub> increases and its radiative forcing (global warming potential) is similarly increased.

Aerosols influence climate both directly, by scattering and absorbing incoming solar and outgoing long-wave radiation, and indirectly, by inducing changes in cloud properties (Penner et al. 2001). Overall, aerosols exert a strong negative radiative forcing (i.e., a cooling effect), although there is considerable uncertainty in estimates of the magnitude of this effect (see Fig. 1.1, Forster et al. (2007)). Aerosols are relatively short-lived with an atmospheric lifetime of a few days. Hence, they cannot be considered to be well-mixed in the atmosphere and their impacts on climate vary from region to region. The picture is further complicated by the fact that the climate effects of aerosols are size-dependent (Penner et al. 2001). SOA tend to be relatively small, with diameters less than 2.5 μm (PM<sub>2.5</sub>), and therefore are longer-lived than larger particles (having lower Stokes numbers and therefore lower deposition velocities). The formation of SOA typically results in a higher number of smaller particles, which not only promotes cloud formation (e.g., Van Reken (2005)) but also increases the longevity of clouds and reduces the frequency of rain events. The clouds formed are also whiter and brighter, i.e., they have higher albedo and therefore reflect more incoming and outgoing radiation. Overall, these various effects combine to result in a negative climate forcing.

## Air Quality

As highlighted previously in this section, biogenic VOCs play an important role in the chemistry that produces tropospheric ozone. Ozone was first identified as a primary component of smog, and therefore a key atmospheric pollutant, in the 1950s (Haagen-Smit 1950, 1952). “Background” levels (i.e., annual average concentrations at rural sites) of ground-level ozone have now reached around 30–40 ppbv in the Northern Hemisphere and about 20 ppbv in the less-polluted Southern Hemisphere. Peak hourly concentrations of ozone of over 100 ppbv are regularly experienced during episodes of “photochemical smog,” with instantaneous concentrations over 400 ppbv recorded, caused by high temperatures and strong sunlight accelerating the production of O<sub>3</sub> from its precursors as well as promoting emissions of biogenic VOCs (Royal Society 2008).

Exposure to high levels of ozone has been shown to reduce lung function and cause inflammation of the airways (WHO 2005), and epidemiological studies from around the world have linked high ozone concentrations to increased cardiopulmonary mortality. For example, it has been estimated that around 22,000 deaths each year are attributable to ozone in Europe alone. Current air quality guidelines suggest a maximum daily ozone exposure limit of 50 ppbv (WHO 2005), although legal limits vary between regions, with Europe, for example, setting an exposure



limit of 60 ppbv (EC 2002). Although high concentrations of ozone usually occur with high temperatures, and often with high concentrations of other pollutants, e.g., NO<sub>2</sub> and PM<sub>10</sub> and PM<sub>2.5</sub> (themselves subject to air quality control regulations), meta-analyses of cardiopulmonary mortality data from epidemiological studies around the world have shown that it is possible to eliminate the effects of these confounders and deduce a concentration-response curve for the effects of ozone alone. Such analyses indicate that there is an increase of 0.6–1.0 % in daily mortality for every 10 ppbv increase in daily maximum ozone concentration above a threshold of 35 ppbv, and this response is significant to at least the 95th percentile. There is also growing evidence that long-term exposure to much lower levels of ozone causes chronic damage to respiratory function (WHO 2005).

As well as human health effects, ozone causes oxidative damage to vegetation. Ozone deposition onto vegetation surfaces leads to uptake through the stomata and subsequent oxidative damage to plant cells and functions. Such damage reduces photosynthesis, decreasing a plant's ability to assimilate carbon and therefore reducing productivity and crop yield (Sitch et al. 2007). Seed production and setting are also affected, propagating the impact through successive generations. Field studies of vegetation, particularly cash crops, have shown clear evidence of a strong link between reduced yields and accumulated damage due to high ozone concentrations. In Europe, this damage is measured using a cumulative metric known as "AOT40," defined as the sum of hourly ozone concentrations (during daylight hours when the stomata are open) above a threshold of 40 ppbv over the growing season of the crop, usually a 3-month period that varies according to latitude and crop type (CLRTAP 2004). More recently, it has been demonstrated that cellular damage can occur at air concentrations below the threshold of 40 ppbv in some instances, but that vegetation can conversely remain unaffected by concentrations above this level. As oxidative damage is governed by the rate of uptake of atmospheric ozone through the stomata (regulated by climate, soil moisture, atmospheric ozone concentrations, and plant growth stage), work is ongoing to develop flux-based criteria for measuring likely damage and identify critical levels for these metrics (see, e.g., CLRTAP (2010)).

It has been demonstrated that such concentration-based measures may not be the best way to identify areas at high risk of ozone damage to vegetation. Within Europe, for example, parts of Spain experience high ground-level ozone concentrations during the growing season; however, ozone fluxes into plant cells are relatively low as the stomata tend to be closed due to water stress during episodes of high ozone. Conversely, the East of England has much lower atmospheric ozone concentrations, but plant cells there have high ozone uptake as the water stress is lower and the stomata tend to remain open. Hence, although ozone damage to vegetation has been widely observed, robust methods to quantify such damage lag behind those developed for health impacts. This is in spite of the clear recognition of the economic and societal implications of the loss of food production due to such damage.

Aerosols have a very obvious impact on air quality, reducing visibility and creating visible haze (Went 1960). Particulate matter is also the biggest single cause of air quality-related health effects, with over two million deaths worldwide

attributable to particles each year (WHO 2005). While the majority of these occur in the developing world and are linked to indoor air pollution and cooking practices (WHO 2005), it is an issue that affects all regions. For example, around 280,000 deaths in Europe are thought to be caused by atmospheric particulate matter, an order of magnitude higher than those attributed to ground-level ozone. Air quality guidelines (WHO 2005) set limits for daily and annual exposure to aerosol particles with diameters of less than 10  $\mu\text{m}$  (of 50  $\mu\text{g m}^{-3}$  and 20  $\mu\text{g m}^{-3}$ , respectively) and 2.5  $\mu\text{m}$  (of 25  $\mu\text{g m}^{-3}$  and 10  $\mu\text{g m}^{-3}$ , respectively). In general, the smaller the particle, the more dangerous it is to the respiratory system as it is able to penetrate further, with particles below around 1  $\mu\text{g}$  able to reach the lung surfaces.

Unlike ozone, there are no recommended exposure limits for vegetation. Indeed, it has been speculated that the production of SOA is beneficial to vegetation as the increase in particle concentrations and possibly cloud cover results in a higher fraction of diffuse radiation relative to direct sunlight. "Diffuse" sunlight occurs as the aerosols reflect and refract incoming radiation resulting in radiation reaching the surface from all directions rather than solely from above. Shading of the lower canopy by leaves in the upper canopy is reduced, and lower leaves receive more radiation and are able to assimilate more  $\text{CO}_2$  through photosynthesis, resulting in a higher overall productivity.

## The Climate-Air Quality Conflict

Climate change and poor air quality are both major challenges to society. Identifying and implementing mitigation strategies are global priorities. Current policies focus on the reduction of emissions of greenhouse gases, primary pollutants, and precursor compounds (such as  $\text{NO}_x$  and VOCs in the case of ozone and other secondary pollutants).

While not simple to implement, for ozone the strategy is relatively straightforward to devise. In VOC-sensitive regions, VOC emission reduction measures are required; in  $\text{NO}_x$ -sensitive regions,  $\text{NO}_x$  emissions must be limited. Furthermore, reducing ozone concentrations in the troposphere both improves air quality and reduces future climate change.

The situation is more complex in the case of aerosols. A lack of understanding of the reactions and processes leading to SOA formation makes it hard for policy-makers to formulate successful strategies to tackle particulate pollution. While the majority of the global budget of SOA is believed to be biogenic in origin, the distribution of atmospheric aerosols reflects the distribution of anthropogenic pollutants, such as nitrate or sulfate compounds. It is thought that the reaction products of biogenic VOCs generally remain in the gas phase, even when theoretically of sufficiently low volatility to condense into the particle phase, until the presence of a so-called "seed" particle provides a surface on which they can condense. Hence, although the pollutant is biogenic, it is the anthropogenic emissions of the "seed" compounds that must be reduced in order to control SOA concentrations (Carlton et al. 2010).

However, in the case of aerosols, tackling air quality by reducing emissions of precursor compounds, and therefore the production of particles, creates a conflict with climate change mitigation, as aerosols exert an overall cooling effect. Currently, priority is being given to improving air quality, as this is an immediate issue and one in which both the problem and solution can be quantified, whereas the effect of aerosols on climate is poorly constrained and therefore highly uncertain, as well as being a problem for the future. The uncertainties surrounding the climate impacts of aerosol particles are a key area of research in the immediate future (Forster et al. 2007).

---

## Future Directions

- Constraining the quantity and environmental controls on biogenic emissions
- Developing improved models to simulate biogenic emissions based on climatic conditions
- Understanding the role of biogenic emissions in the formation of aerosols in the atmosphere
- Understanding the interactions between urban and anthropogenic emissions with biogenic emissions
- Understanding the interaction of biogenic VOCs, atmospheric chemistry, and climate in a changing world

---

## References

- Arya SP. Introduction to micrometeorology. San Diego: Academic; 2001. 420 pp.
- Carlton AG, Pinder RW, Bhawe PV, Pouliot GA. To what extent can biogenic SOA be controlled? *Environ Sci Technol.* 2010;44(9):3376–80.
- Chameides WL, Lindsay RW, Richardson J, Kiang CS. The role of biogenic hydrocarbons in urban photochemical smog – Atlanta as a case-study. *Science.* 1988;241(4872):1473–5. doi:10.1126/science.3420404.
- CLRTAP. Manual on methodologies and criteria for modelling and mapping critical loads and levels and air pollution effects, risks and trends. Convention on Long-Range Transboundary Air Pollution (CLRTAP). 2004. Available on-line from <http://www.icp-mapping.org>
- CLRTAP. Manual of methodologies for modelling and mapping effects of air pollution. Convention on Long-Range Transboundary Air Pollution (CLRTAP). 2010. Available on-line from <http://icpvegetation.ceh.ac.uk>
- Dlugi R, Berger M, Zelger M, Hofzumahaus A, Siese M, Holland F, Wisthaler A, Grabmer W, Hansel A, Woppmann R, Kramm G, Mollmann-Coers M, Knaps A. Turbulent exchange and segregation of HOx radicals and volatile organic compounds above a deciduous forest. *Atmos Chem Phys.* 2010;10(13):6215–35. doi:10.5194/acp-10-6215-2010.
- EC. Directive 2002/3/EC – relating to ozone in ambient air. Brussels: Commission of the European Communities; 2002. Available on-line from <http://ec.europa.eu/environment/air/legis.htm>
- Finnigan J. Turbulence in plant canopies. *Annu Rev Fluid Mech.* 2000;32:519–71. doi:10.1146/annurev.fluid.32.1.519.
- Foken T. *Micrometeorology.* Berlin: Springer; 2008. 328 pp.
- Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R. Changes in atmospheric

- constituents and in radiative forcing. In climate change 2007: the physical science basis. In: Solomon SD et al., editors. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press; 2007.
- Fuentes JD, Lerdau M, Atkinson R, Baldocchi D, Bottenheim JW, Ciccioli P, Lamb B, Geron C, Gu L, Guenther A, Sharkey TD, Stockwell W. Biogenic hydrocarbons in the atmospheric boundary layer: a review. *Bull Am Meteorol Soc.* 2000;81(7):1537–75. doi:10.1175/1520-0477(2000)081<1537:bhitab>2.3.co;2.
- Guenther AB, Monson RK, Fall R. Isoprene and monoterpene emission rate variability – observations with Eucalyptus and emission rate algorithm development. *J Geophys Res Atmos.* 1991;96(D6):10799–808. doi:10.1029/91jd00960.
- Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, Wang X. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci Model Dev.* 2012;5(6):1471–92. doi:10.5194/gmd-5-1471-2012.
- Haagen-Smit AJ. The air pollution problem in Los Angeles. *Eng Sci.* 1950;14(3):7–13.
- Haagen-Smit AJ. Chemistry and physiology of Los Angeles smog. *Ind Eng Chem Res.* 1952;44:1342–6.
- Hallquist M, Wenger JC, Baltensperger U, Rudich Y, Simpson D, Claeys M, Dommen J, Donahue NM, George C, Goldstein AH, Hamilton JF, Herrmann H, Hoffmann T, Iinuma Y, Jang M, Jenkin ME, Jimenez JL, Kiendler-Scharr A, Maenhaut W, McFiggans G, Mentel TF, Monod A, Prevot ASH, Seinfeld JH, Surratt JD, Szmigielski R, Wildt J. The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmos Chem Phys.* 2009;9(14):5155–236.
- Penner JE, Hegg D, Leaitch R. Unraveling the role of aerosols in climate change. *Environ Sci Technol.* 2001;35(15):332A–40. doi:10.1021/es0124414.
- Rasmussen R. Isoprene: identified as a forest-type emissions to the atmosphere. *Environ Sci Technol.* 1970;4:667–71.
- Rasmussen R. What do hydrocarbons from trees contribute to air pollution? *J Air Pollut Control Assoc.* 1972;22(7):537–43.
- Royal Society. Ground-level ozone in the 21st century: future trends, impacts and policy implications. Fowler D, editor. Science policy report 15/08. London: The Royal Society; 2008.
- Seinfeld JH, Pandis SN. Atmospheric chemistry and physics – from air pollution to climate change. 2nd ed. Wiley, New York; 2006.
- Sharkey TD, Wiberley AE, Donohue AR. Isoprene emission from plants: why and how. *Ann Bot.* 2008;101(1):5–18. doi:10.1093/aob/mcm240.
- Sillman S. The relation between ozone, NO<sub>x</sub> and hydrocarbons in urban and polluted rural environments. *Atmos Environ.* 1999;33(12):1821–45. doi:10.1016/s1352-2310(98)00345-8.
- Sitch S, Cox PM, Collins WJ, Huntingford C. Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature.* 2007;448(7155):791–4. doi:10.1038/nature06059.
- Spracklen DV, Jimenez JL, Carslaw KS, Worsnop DR, Evans MJ, Mann GW, Zhang Q, Canagaratna MR, Allan J, Coe H, McFiggans G, Rap A, Forster P. Aerosol mass spectrometer constraint on the global secondary organic aerosol budget. *Atmos Chem Phys.* 2011;11(23):12109–36. doi:10.5194/acp-11-12109-2011.
- Stroud C, Makar P, Karl T, Guenther A, Geron C, Turnipseed A, Nemitz E, Baker B, Potosnak M, Fuentes JD. Role of canopy-scale photochemistry in modifying biogenic-atmosphere exchange of reactive terpene species: results from the CELTIC field study. *J Geophys Res Atmos.* 2005;110(D17). doi:10.1029/2005jd005775
- VanReken TM, Ng NL, Flagan RC, Seinfeld JH. Cloud condensation nucleus activation properties of biogenic secondary organic aerosol. *J Geophys Res Atmos.* 2005;110(D7):D07206.
- Warneck P. Chemistry of the natural atmosphere. 2nd ed. San Diego: Academic; 2000.

- Warneck P, Williams J. The atmospheric chemist's companion. New York: Springer; 2012. doi:10.1007/978-94-007-2275-0. 436 pp.
- Went FW. Blue hazes in the atmosphere. *Nature*. 1960;187(4738):641–3.
- WHO. Air quality guidelines – global update 2005. Geneva: World Health Organisation; 2005.

## Further Reading

- Bender J, Weigel HJ. Changes in atmospheric chemistry and crop health: a review. *Agron Sustain Dev*. 2011;31(1):81–9. doi:10.1051/agro/2010013.
- Online version available at [http://www.knovel.com/web/portal/browse/display?\\_EXT\\_KNOVEL\\_DISPLAY\\_bookid=2126&VerticalID=0](http://www.knovel.com/web/portal/browse/display?_EXT_KNOVEL_DISPLAY_bookid=2126&VerticalID=0)
- Penuelas J, Staudt M. BVOCs and global change. *Trends Plant Sci*. 2010;15(3):133–44. doi:10.1016/j.tplants.2009.12.005.

Kimberly O’Keefe, Clint J. Springer, Jonathan Grennell, and Sarah C. Davis

## Contents

Introduction .....	602
Biomass Energy .....	605
Biomass Conversion Technologies .....	605
Bioenergy Feedstocks .....	611
The Case for Liquid Biofuels in the World Energy Market .....	614
Ecological Considerations Associated with Cellulosic Biofuel Production .....	615
Management Decisions .....	615
Greenhouse Gas Emissions .....	617
Soil and Nutrient Management .....	619
Water .....	622
Impacts on Wildlife and Biodiversity .....	623
Invasive Species Potential .....	624
Pests and Pathogens .....	626
Future Directions .....	627
References .....	627

## Abstract

- Renewable energy sources such as solar power, wind power, geothermal power, and bioenergy will improve energy sustainability and reduce environmental impacts associated with human energy use.

---

K. O’Keefe (✉)

Division of Biology, Kansas State University, Manhattan, KS, USA

e-mail: [kokeefe@k-state.edu](mailto:kokeefe@k-state.edu)

C.J. Springer

Department of Biology, Saint Joseph’s University, Philadelphia, PA, USA

e-mail: [cspringe@sju.edu](mailto:cspringe@sju.edu)

J. Grennell • S.C. Davis

Voinovich School of Leadership and Public Affairs, Ohio University, Athens, OH, USA

e-mail: [jg509012@ohio.edu](mailto:jg509012@ohio.edu); [daviss6@ohio.edu](mailto:daviss6@ohio.edu)

- Two types of bioenergy feedstocks exist: first-generation feedstocks that are derived from food crops and advanced feedstocks that are derived from nonfood plants. Advanced feedstocks include cellulosic bioenergy crops such as herbaceous perennial grasses, short-rotation woody crops, and annual crop residues.
- Due to the complex structure of lignocellulosic plant material, cellulosic bioenergy feedstocks are generally more difficult to process into liquid fuels than food crops. However, a variety of both thermochemical and biochemical conversion technologies exist or are being developed to improve the transformation of cellulosic biomass into alternative energy sources.
- Although cellulosic bioenergy crops are thought to have fewer adverse effects on natural ecosystems than first-generation bioenergy crops, the extent of their impact is determined by the bioenergy species grown, how the crop is managed, and the type of land-use changes associated with the cultivation of the bioenergy crop.
- Land-use changes associated with cellulosic bioenergy crop production can be direct (land-use change occurs directly for cultivating bioenergy feedstocks) or indirect (land-use change occurs on land not used for bioenergy production due to the displacement of land used for food crop production), and each can have different impacts on the environment.
- The cultivation of cellulosic bioenergy crops produces fewer greenhouse gas emissions than first-generation bioenergy crops. Highly productive cellulosic bioenergy crops may also sequester more atmospheric carbon dioxide, which can reduce greenhouse gas emissions associated with bioenergy land-use changes.
- Cellulosic bioenergy crops have the potential to reduce soil erosion, rehabilitate degraded soil, increase soil organic carbon (SOC), and counteract SOC losses due to food crop and first-generation bioenergy feedstock cultivation.
- Cellulosic bioenergy feedstocks generally use water and nutrients more efficiently than first-generation bioenergy crops, which may decrease irrigation and fertilization requirements for bioenergy feedstock production. This can benefit aquatic systems by reducing water-use and nutrient runoff.
- Land-use changes resulting in habitat loss and habitat fragmentation can impact native wildlife species. However, cellulosic biomass feedstocks have the potential to provide habitat for insects, small birds, and mammals if landscape heterogeneity is maintained.
- Some perennial biomass feedstocks have the potential to become invasive in ecosystems and also accelerate the spread of pathogens and other invasive species when grown in monocultures.

---

## Introduction

Nonrenewable natural resources such as coal, petroleum, and natural gas have long been exploited for energy consumption due to their historic relative abundance, versatility, transportability, and low cost. However, global reserves of these raw materials are finite and are rapidly decreasing as global demand for energy increases.

Extracting and using these energy sources can also have many negative environmental consequences. For example, fossil fuel combustion releases geologically stored carbon and other pollutants into the atmosphere, including greenhouse gases that cause climate change, indirectly damage ozone, contribute to acid deposition, and cause ocean acidification (Schlesinger and Bernhardt 2013). The physical exploitation of these fuels also damages the earth's surface layers, contaminates watersheds, and occasionally results in accidental marine contamination. Overall, the depletion of fossil fuel reserves, increasing global demand for energy, and the adverse environmental impacts associated with liquid and solid fossil fuel exploitation have highlighted the need to decrease dependence on nonrenewable fuel sources and have stimulated global interest in replacing fossil energy with alternative, sustainable solutions for future energy consumption.

Renewable energy technologies such as solar power, wind power, geothermal power, and bioenergy have the potential to improve energy sustainability and reduce the environmental consequences associated with human energy consumption. Bioenergy, in particular, is a renewable energy source that is primarily derived from plant material and is used to produce various energy products via direct combustion or chemical processing. Bioenergy feedstocks (i.e., biomass) include dedicated energy crops, agricultural food crops and residues, oil products, and other organic waste materials. Depending on the raw material and conversion pathway used, these feedstocks can produce an array of energy products ranging from liquid biofuels (e.g., biodiesel and bioethanol) to heat and electricity. Bioenergy is widely regarded as a viable alternative energy source because it has the potential to offer a broad range of socioeconomic and ecological benefits. In addition to reducing reliance on traditional fossil fuels, biomass production and biofuel processing can create employment opportunities, particularly in rural areas, and provide energy independence in both developing and industrialized countries. Bioenergy may also reduce carbon emissions because bioenergy crops sequester atmospheric carbon dioxide as they grow and because biomass combustion only releases as much carbon dioxide into the atmosphere as plant growth has sequestered. Therefore, bioenergy has the potential to become an economically beneficial and environmentally sustainable solution to the present energy crisis.

Although bioenergy is versatile and can provide various solutions to current energy concerns, biomass is predominantly used in the developed world for biodiesel or bioethanol to replace petroleum transportation fuels. These fuels are of particular interest because they do not require major modifications to current transportation systems and can be easily mixed with fossil petroleum as fuel additives. Presently, biofuels are produced from "first-generation" (i.e., conventional) sources that are also used commercially as food crops. For instance, bioethanol is fermented from sugar sources such as corn grain (*Zea mays* L.) or sugar cane (*Saccharum officinarum* L.), while biodiesel is processed from oil crops such as soybean (*Glycine max*, L.) and rapeseed (*Brassica napus* L.). The technologies used to produce first-generation biofuels are currently well established, and although biofuel additives/substitutes are not yet major energy sources in the transportation sector, their production is now commercial. Biofuels do have the



potential to contribute significantly to the transportation sector in the future; however, first-generation biofuels also raise several economic and environmental concerns (Williams et al. 2009). For example, first-generation bioenergy crops place increased pressure on the food industry because they compete with land used for food production and/or directly reduce the availability of feedstocks used commercially for food. These crops also require extensive water, fertilizer, and pesticide inputs, and their cultivation is associated with soil erosion, air and water pollution, and biodiversity losses as marginal grasslands and pastures are put into cultivation. Finally, the energy use associated with crop production and biofuel conversion processes may produce carbon emissions that do not result in a beneficial carbon balance. These disadvantages suggest that first-generation bioenergy crops may have long-term environmental costs and have thus generated an interest in developing bioenergy from alternative sources.

Biofuels produced from nonfood materials (i.e., “advanced” biofuels) have the potential to mitigate many of these concerns. Advanced biofuels are typically produced from cellulosic feedstocks, including dedicated herbaceous bioenergy crops (e.g., perennial C<sub>4</sub> grasses such as switchgrass or *Miscanthus x giganteus*), short rotation woody crops (e.g., hybrid poplar, willow), annual crop residues (e.g., corn stover), forest residues (e.g., commercial logging residues), and municipal solid waste (e.g., tree trimmings and paper products). These materials are generally cheap and abundant and have less potential to strain the food industry because they are derived from nonfood sources. Dedicated bioenergy crops, in particular, can produce high yields with relatively little water and nutrient inputs. When managed correctly, these crops can also have fewer adverse effects on the environment than first-generation crops (Howarth and Bringezu 2009). Like first-generation biomass feedstocks, cellulosic feedstocks can be burned directly for heat or can be chemically converted to liquid biofuels. However, cellulosic materials are more difficult to process than traditional biomass feedstocks and the additional conversion steps associated with breaking down lignocellulose into fermentable sugars render advanced biofuels cost-ineffective at the present (Carroll and Somerville 2009). If cellulosic biofuels were cost-competitive with first-generation biofuels, though, they could potentially become a commercially viable alternative energy source in the future. The costs and benefits of bioenergy production, as well as the environmental impacts associated with bioenergy production, will therefore be important to consider when evaluating the future sustainability of cellulosic biofuels.

The goal of this chapter is to provide an overview of biofuel production from cellulosic materials and to explore the environmental impacts associated with these processes. First, the various feedstocks, conversion technologies, and fuel products associated with cellulosic bioenergy production will be described in detail. The challenges of producing these biofuels will also be highlighted. Second, the potential impacts of cellulosic biofuel production on natural ecosystems will be explored. This section will provide an in-depth discussion of how different land-use changes and management practices associated with cellulosic biofuel production can affect greenhouse gas emissions, habitat fragmentation, biodiversity, soil properties, and water quality. Ultimately, this chapter will explore the advantages and

disadvantages of advanced cellulosic biofuels as an alternative fuel source, particularly with respect to production efficacy and environmental impacts.

---

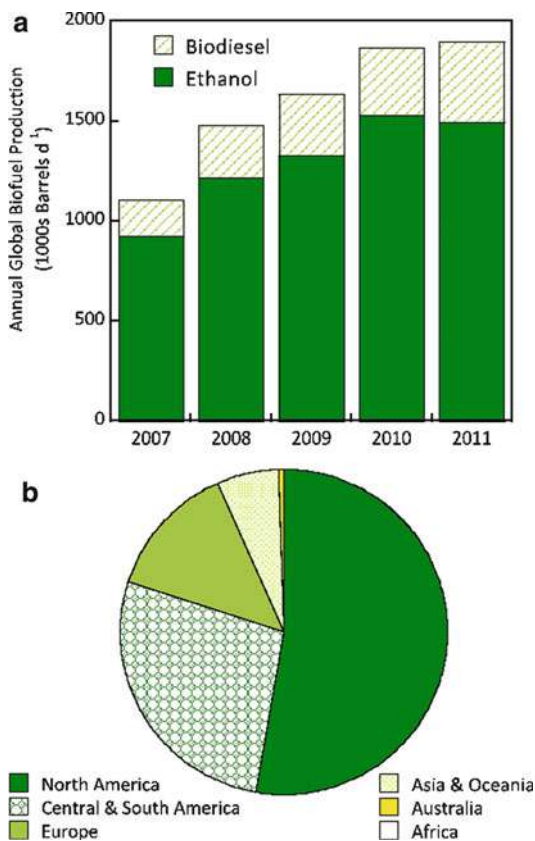
## Biomass Energy

Plants use solar radiation, carbon dioxide, water, and mineral nutrients to convert solar energy into chemical energy through the process of photosynthesis. In plants, this energy is stored primarily in the form of soluble and non-soluble carbohydrates that drive metabolic activity and form tissue structures, respectively. When carbohydrate bonds are broken via industrial conversion technologies, the energy released can be captured and used as a source of fuel for human consumption. If combusted, the energy takes the form of heat and can be used to produce electricity. If chemically converted to liquid hydrocarbons, the energy remains in a chemical form that can fuel combustion engines. In these processes, the plant tissue is referred to as *biomass*. Biomass has been a steady and reliable source of heat throughout human history and remains so in some developing nations where biomass-generated heat comprises up to 90 % of energy consumption. In developed nations, biomass energy delivers a significantly lower (~3–4 %) proportion of the total energy consumed (Demirbas 2009). However, there has been a concerted effort in recent years to increase the contribution of biomass energy to national energy budgets, particularly from advanced cellulosic sources. Currently, the United States leads the world in bioenergy production mainly due to the use of ethanol in blended fuels (Fig. 1). In the next few sections, the technology necessary for the conversion of biomass to liquid cellulosic fuels, the fuel products generated from these processes, and the most common plant species used for liquid cellulosic fuel production will be summarized, as will the potential for biomass energy to contribute to the global energy supply in the future.

## Biomass Conversion Technologies

Biomass conversion to useful energy forms can be accomplished using a variety of processes. Currently, biochemical conversion and thermochemical conversion techniques are the two main approaches to produce liquid fuels from cellulosic sources (Fig. 2). Many biochemical conversion processes ferment biomass carbohydrates into an alcohol product (bioethanol), while thermochemical conversion heats the raw biomass feedstock in the presence of varying oxygen concentrations to produce thermal energy or a variety of organic molecules. Generally, the method chosen to produce bioenergy depends on the type of biomass feedstock that is used, the requirements for end use, economic conditions, and environmental regulations associated with the energy source. The major fuels currently derived from biomass are biodiesel, methanol, dimethyl ether (DME), syngas, methyl tertiary-butyl ether (MTBE), biomethane from biogas, cellulosic ethanol, and hydrogen. This review will focus primarily on the production procedures used to generate the most widely used of these fuels.

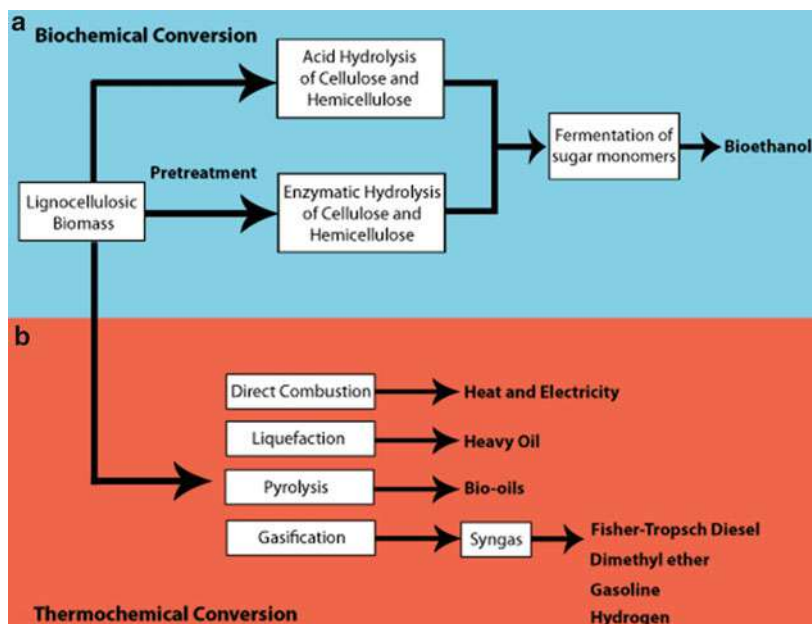
**Fig. 1** (a) Annual global liquid cellulosic biofuel production from 2007 to 2011. Total annual production increased 71 % globally across this period. (b) Proportion of total biofuel production by continent from 2007 to 2011. North America produces the most biofuel mostly due to the use of maize for ethanol production to be used in blended fuels (Source: *International Energy Agency*)



### Generation of Liquid Cellulosic Fuels Through Biochemical Conversion

Biochemical conversion of cellulosic biomass to liquid fuel has become an area of intense focus for the scientific and engineering communities in recent years. The primary reasons for this focus relate to the nearly scale-neutral production of fuels as well as the lower costs compared to thermal conversion technologies used in biofuel production. Similar to grain ethanol produced from first-generation sugar and starch crops, bioethanol derived from advanced cellulosic material is produced via fermentation reactions. However, advanced biomass is difficult to process directly into liquid fuel due to the properties of its structural components. Therefore, additional processing steps are required to produce ethanol from cellulosic sources.

Lignin, cellulose, and hemicellulose are the three primary constituents of plant cell walls in cellulosic (i.e., “lignocellulosic”) biomass. These molecules are also found in the greatest abundance in plant tissue and are therefore the main sources of energy derived from cellulosic bioenergy feedstocks. Cellulose microfibrils are large  $\beta$ -1,4 glucan chains that provide the structural framework for the plant cell wall.



**Fig. 2** Primary approaches used to convert lignocellulosic biomass into bioenergy via (a) biochemical and (b) thermochemical conversion processes. Also indicated are the most common fuel derivatives from each process

These macromolecules can be chemically deconstructed into predictable 6-carbon sugars, but due to their complex semicrystalline structure, as well as the insoluble nature of their  $\beta$ -1,4 glucan chains, they are difficult to degrade via hydrolysis (a processing step used in the biochemical conversion of cellulosic biomass to bioethanol). Hemicelluloses are polysaccharides comprised of both pentoses and hexoses. In a plant cell wall, hemicelluloses form hydrogen bonds with cellulose, binding the cellulose microfibrils together and creating a flexible network of macromolecules. Like cellulose, hemicellulose is also difficult to process during the biochemical conversion of biomass to liquid fuel, primarily because the bacterial and yeast species most commonly used to ferment plant sugars do not metabolize the five carbon sugars efficiently. Finally, lignins are large aromatic polymers that form a strong matrix around the cellulose and hemicellulose complex (Taiz and Zeiger 2010). This provides structural support to the plant cell, as well as protection from pests and pathogens. Due to its strength and durability, lignin is resistant to degradation and therefore exists as a by-product during the production of liquid cellulosic biofuels. The removal of lignin, as well as the conversion of complex cellulose and hemicelluloses to simple sugars, is therefore required prior to fermenting cellulosic biomass into ethanol. These initial steps make the processing of advanced feedstocks far more energetically expensive than those required to process first-generation feedstocks.

Four basic steps are required for the biochemical production of liquid cellulosic ethanol: size reduction, pretreatment, hydrolysis, and fermentation. The first, size reduction, is the mechanical reduction of biomass to a smaller size; this facilitates the access of reagents used in the fuel generation process to the array of carbohydrates found in cellulosic biomass. Once the initial feedstock is mechanically processed, the complex lignocellulosic molecules must be broken down into fermentable material via hydrolysis and, depending on the type of hydrolysis used, pretreatment prior to hydrolysis.

Defined generally, hydrolysis is the process of splitting a molecule into smaller fragments with the addition of water. In the biochemical conversion of biomass to liquid biofuel, hydrolysis is used to cleave (depolymerize) complex lignocellulosic polymers (cellulose and hemicellulose) into simple constituent sugar monomers (glucose, pentose, hexose, and xylose) that can subsequently be fermented into ethanol fuel. Hydrolysis of cellulosic biomass can occur chemically or enzymatically. Chemical, or acid hydrolysis, uses an acid (typically sulfuric acid or hydrochloric acid) in the presence of water to break down cellulose into glucose and hemicellulose into pentoses and hexoses. Xyloses can also be produced from hardwood and crop residue feedstocks. Lignin, however, is very resistant to degradation and therefore remains as a by-product in these reactions. Additional toxic by-products, such as hydrolyzates, may also form during this process.

Although acid hydrolysis does not require pretreatment beyond the mechanical processing of the raw biomass feedstock, this process is costly and energetically expensive because the sugar products must be conditioned before fermentation to remove toxic by-products and also because the acid used for hydrolysis must be recovered. Therefore, enzymatic hydrolysis is the more common process used to depolymerize cellulosic biomass into simple sugars.

If carried out enzymatically (enzymatic hydrolysis), the biomass feedstock must first undergo pretreatment. Pretreatment is the physical, chemical, or enzymatic degradation of biomass that is used to increase enzyme access to cellulose and other polysaccharide components of the biomass feedstock. This typically involves the breakdown of the biomass with the same chemicals used during acid hydrolysis (sulfuric acid or hydrochloric acid), which results in the partial hydrolysis of the biomass. After the biomass has been pretreated, the remaining material that has not been degraded by the acid can then undergo hydrolysis catalyzed by a mixture of enzymatic compounds. This process requires the use of cellulases, a class of enzymes derived from bacterial or fungal sources.

Following acid or enzymatic hydrolysis, the simple sugars, mostly xylose (derived from woody species) and glucose (derived from non-woody species), are converted to liquid bioethanol fuels through fermentation. Fermentation reactions can occur under aerobic or anaerobic conditions and are driven by many different kinds of microorganisms in nature. In the processing of liquid biofuel from cellulosic biomass, microorganisms use sugar monomers to produce ethanol. Once the enzymatic fermentation steps are complete, distillation and dehydration processes

can be used to yield anhydrous bioethanol (90–95 % purity by distillation and >99 % purity by distillation and dehydration) that can then be blended with gasoline for use as a fuel in the transport sector (Saxena et al. 2009).

### **Generation of Liquid Cellulosic Fuels Through Thermochemical Conversion**

The burning of biomass (direct combustion) under aerobic conditions to produce heat is the most basic type of energy derived from plant material and can be used to drive mechanical power or generate electricity. Direct combustion can derive energy from both first-generation and advanced biomass sources. However, more complicated thermochemical conversion processes have also been developed to produce liquid fuels from advanced cellulosic sources. If cellulosic biomass is heated under low oxygen conditions, hydrogen and organic gases are produced that can be further processed into liquid fuels such as Fischer-Tropsch biodiesel, dimethyl ether, or synthetic natural gas. The most common of these processes include gasification, pyrolysis, torrefaction, and liquefaction.

Gasification is the thermal conversion of biomass into a combustible gas mixture known commonly as synthesis gas or syngas. Syngas generally contains  $\text{CO}_2$ ,  $\text{CO}$ ,  $\text{CH}_4$ ,  $\text{N}_2$ , and  $\text{H}_2$  in varying proportions, depending on the feedstock used in the process. The conversion of biomass to syngas begins when biomass feedstocks are combusted at temperatures ranging from 800 °C to 1,000 °C. At these high temperatures, biomass decomposes quickly into the syngas components, as well as solid char and tar residues. Syngas, with the addition of different catalysts, can then be used to produce various fuel products including hydrogen, methanol, ethanol, and dimethyl ether (DME). For example, hydrogen gas can be produced using the water-gas-shift reaction (WGS). During this process,  $\text{CO}$  from syngas reacts with oxygen from water to produce  $\text{H}_2$  and  $\text{CO}_2$ . The  $\text{H}_2$  product can then be used to process other liquid fuels or it can be burned directly to produce electricity. Also produced from syngas are a number of hydrocarbons that can be altered further into waxes or liquid fuels that function similarly to traditional gasoline and diesel fuel. The Fisher-Tropsch process, for instance, is the reaction of  $\text{CO}$  and  $\text{H}_2$  in the presence of a metal catalyst to produce a mixture of liquid hydrocarbons that can be further processed into diesel fuel. Another pathway to generate liquid fuel is the methanol-to-gasoline (MTG) process, where methane industrially converted into methanol via specialized catalysts is then further converted to gasoline. Gasification is a useful conversion technology because it allows diverse feedstocks to be processed similarly despite differences in the chemical composition of the biomass feedstock. The versatility of syngas also makes it an attractive option to process liquid cellulosic fuels. For example, DME may be added directly to diesel fuel with no additional steps required, unlike methanol and ethanol.

Pyrolysis is another thermochemical conversion process used to convert biomass feedstocks into liquid cellulosic fuel. In general, pyrolysis is the decomposition of organic material in an anaerobic environment that leads to the production

of gas, solid carbon-based char, and liquid bio-oil fuel. The ratios of these pyrolysis products largely depend on a number of factors including the reaction temperature, pressure, rate of heating, the length of the reaction time, and the biomass feedstocks utilized at the onset. The process of pyrolysis begins by heating the biomass feedstock to a high temperature (ranging from 200 °C to >1,000 °C, depending on the method). At this point volatile organic compounds, or VOCs, form and leave behind a carbonaceous char, a process known as primary pyrolysis. The release of VOCs results in a transfer of heat from the hot VOCs to the feedstock that has yet to be pyrolysed. As the VOCs cool, they form tar. Finally, autocatalytic secondary pyrolysis occurs while primary pyrolysis occurs simultaneously, leading to the production of liquid biofuel. Currently, three types of pyrolysis reactions exist to produce char, tar, and liquid cellulosic fuels. The first is known as *slow pyrolysis*. The slow rate of heating in slow pyrolysis produces a higher ratio of char than liquid and gas products. The second, *fast pyrolysis*, involves fast heating rates and results in a much higher ratio of liquid cellulosic fuels. Finally, *flash pyrolysis* is a more efficient mechanism similar to fast pyrolysis except that very high reaction temperatures and very high heating rates of the reactions are used. Due to the extremely fast heating in flash pyrolysis, the conversion of biomass feedstocks to fuel is much more efficient and leads to a fuel product that does not require further refinement after the initial pyrolytic reactions have occurred.

An additional thermochemical conversion process of biomass feedstocks to liquid cellulosic fuels is known as liquefaction. Liquefaction is the process of heating biomass feedstocks to low temperatures under high pressure with the addition of a catalyst, solvent, and/or reducing gas such as hydrogen to produce a highly viscous insoluble oil that can be used for a variety of purposes. At this time, there is low interest in liquefaction as a viable thermoconversion process because of the complexity and expense associated with building reactors when compared to other thermoconversion processes like gasification and pyrolysis.

Finally, it should be noted that the physical properties of cellulosic plant material can often complicate thermochemical conversion processes. For instance, the high water and oxygen content of the plant material can produce large quantities of smoke during combustion, while the fibrous nature of its lignocellulosic cell walls can make the biomass physically difficult to process. Recent advancements therefore recommend pretreating the biomass to increase the quality of the biomass and to reduce undesirable side effects associated with fuel production. Torrefaction is a pretreatment method that is similar to pyrolysis but occurs at much lower temperatures (200–300 °C). This process removes oxygen from the plant tissue and decreases the volume of the tissue by as much as 62–69 %. In doing so, the energy content of the biomass is maintained because the material dries and partially de-volatilizes. This reduction in biomass and concomitant energy preservation can increase the energetic density of the material by approximately 20–30 %, which not only makes the material easier to process but also aids in transportation (Bhagwan Goyal et al. 2009).

**Table 1** Major environmental impacts and considerations for first-generation and advanced cellulosic bioenergy feedstocks

Feedstock type	Example feedstocks	Potential environmental impacts
<b>First generation</b>	<b>Sugar crops</b>	Increased land-use change
	<i>Zea mays</i> (corn)	Increased greenhouse gas emissions
	<i>Saccharum officinarum</i> (sugar cane)	Increased nutrient and chemical usage
		Increased soil erosion and runoff
	<b>Oil crops</b>	Decreased storage of soil organic carbon
	<i>Glycine max</i> (corn)	Increased water-use and impact on water quality
	<i>Brassica napus</i> (rapeseed)	Decreased wildlife diversity
<b>Advanced</b>	<b>Perennial C<sub>4</sub>grasses</b>	Increased land-use change
	<i>Panicum virgatum</i> (switchgrass)	Decreased greenhouse gas emissions
	<i>Miscanthus x giganteus</i>	Decreased nutrient and chemical usage
		Decreased soil erosion and runoff
	<b>Short rotation woody crops</b>	Increased storage of soil organic carbon
	<i>Populus</i> spp. (poplar)	Decreased water-use and impact on water quality
	<i>Salix</i> spp. (willow)	Decreased or increased wildlife diversity
		Increased invasiveness
	<b>Wastes and residues</b>	
	Corn stover	
	Forest residues	
	Municipal solid wastes	

## Bioenergy Feedstocks

The plant species that can be grown as cellulosic bioenergy crops are even more diverse than the processing technologies used to produce biofuels (Table 1). Many crops and wild plant species are currently being used as bioenergy feedstocks, are in development to be used to produce biofuel, or are excellent candidates for biofuel production in the future. Examples of such feedstock plants are agricultural wastes, trees wastes and residues, food crops, and perennial grasses. Generally, these plant species are classified into two categories: food crops and bioenergy crops. Plant species associated with each of these groups present unique challenges in the production of suitable biomass for liquid cellulosic fuel manufacturing and also have varying environmental concerns linked to their growth. A number of characteristics need to be considered when deciding which species to use as a feedstock, including mineral content, moisture content, nutrient and water requirements, dry matter production per unit land area, and the chemical composition of the tissue, especially lignin, hemicellulose, and cellulose content. Furthermore, the geographical distribution of the plant species, the effects of the species on the environment



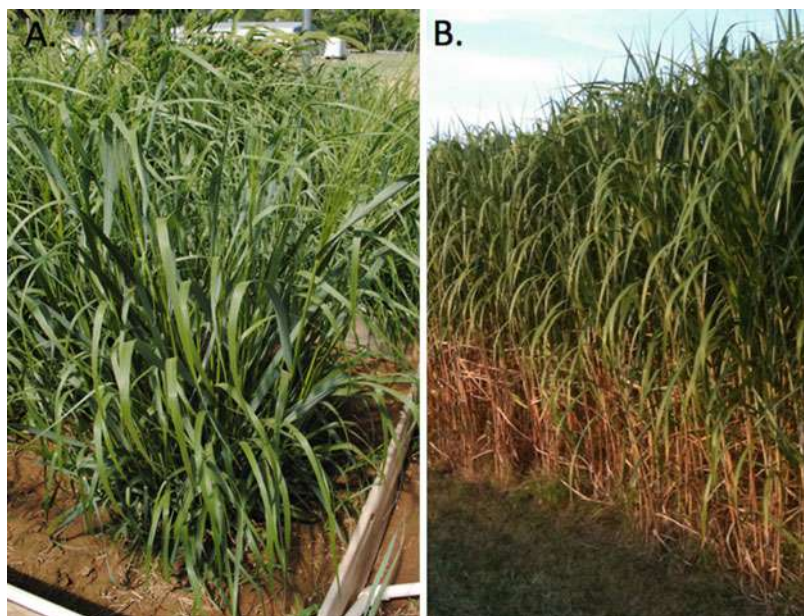
and ecology of the local ecosystem, their response to environmental conditions, their genetic diversity, and other agronomic considerations must be taken into account.

### **First-Generation Bioenergy Feedstocks**

Food crops have long been viewed as the least desirable for use as biofuel feedstock and are often termed first-generation biofuel feedstocks. Among the biggest concerns regarding the use of food crops for biofuels is the tremendous pressure that alternative uses place on an already marginal food supply. Another negative consideration is that most of these crops have an annual lifecycle that require them to be replanted each year, which leads to increased uses of energy for planting as well as pesticide and fertilizer use. For example, of all the plant species used for biofuel production, corn cultivation produces the greatest greenhouse gas emissions. Despite this, corn recently became the largest contributor to bioethanol production. Although corn is now the leading biomass feedstock globally, sugarcane is likely to be a significant contributor in the foreseeable future. Of the food crops used for biofuel production, sugarcane produces yields of up to 100 t/ha with little fertilizer input, thus resulting in a significantly higher energy transfer efficiency than corn. Sugarcane is already produced in significant quantities in South America and remains one of the most important crops globally. In sugarcane, sucrose accounts for 20 % of the dry matter produced and can be quickly converted to bioethanol. After the initial processing of soluble sugars to bioethanol, a by-product known as bagasse is produced. As conversion technologies of lignocellulose to ethanol continue to improve, sugarcane bagasse is likely to increase in use as biofuel feedstock thus making sugarcane an even stronger candidate for bioethanol conversion (Perlack et al. 2005).

### **Advanced Bioenergy Feedstocks**

Dedicated herbaceous bioenergy (nonfood) crops represent the next wave of biomass feedstocks for biofuel production and are also known as advanced biomass feedstocks. These biomass crops present major advantages over first-generation biomass feedstocks because of their long-term environmental sustainability. Of all of the potential energy crops, the two with the most promise in the future for temperate regions are the perennial rhizomatous grasses, switchgrass and miscanthus (*Miscanthus x giganteus* Greef et Deuter). Switchgrass (*Panicum virgatum* L.) is a warm-season, C<sub>4</sub> perennial grass that is native to North America and adapted to thrive in a wide range of environmental conditions (Fig. 3). Chosen as a model bioenergy species by the United States Herbaceous Energy Crops Program in the early 1990s, switchgrass has been the focus of research as a biomass feedstock for bioethanol production for a number of years. Switchgrass has many cultivars already developed for use as a forage stock, bioenergy crop, and as a restoration species. Switchgrass yields an average of 12 t/ha with recent increases in productivity of ~50 % over the last two decades because of cultivar improvements and agronomic technologies (Field et al. 2008). This total is still far below the productivity of the first-generation biomass feedstock sugarcane, but not always



**Fig. 3** *Panicum virgatum* L. (a) and *Miscanthus gigantea* (b), two species of grass with a high potential to become important feedstocks for second-generation biofuels (Photo credit: Kimberly O’Keefe (a) and Sarah Davis (b))

below the yields generated by corn. However, its extensive root system, low-nutrient and water requirements, and perennial lifecycle make switchgrass an attractive option as a biomass feedstock for conversion to bioethanol. In addition, the loss of biomass during harvest is low for switchgrass. Also increasing the attractiveness of switchgrass as a biomass feedstock are the genetic tools, such as linkage maps, that have been developed in recent years for use in breeding programs, an effort that results from the intense focus on switchgrass by the United States Department of Energy (Perlack et al. 2005).

*Miscanthus x giganteus* is another intensely researched option for biofuel production. *Miscanthus x giganteus* is also a rhizomatous, C<sub>4</sub> perennial grass that is a sterile hybrid of *Miscanthus saccharifloris* and *Miscanthus sinensis*, both native to Asia. Currently a single sterile hybrid, *Miscanthus x giganteus* Greef et Deuter, is the cultivar primarily studied for biofuel production. Like switchgrass, *Miscanthus x giganteus* produces a large aboveground component and stores much of its nutrients belowground in rhizomes prior to harvesting, thus reducing the nutrient requirements for the species. In fact, a number of studies have found *Miscanthus x giganteus* productivity to largely be unresponsive to nitrogen additions. Also like switchgrass, *Miscanthus x giganteus* has been successfully grown in a number of locations globally and therefore represents a single product for conversion technologies to use as a biomass feedstock. *Miscanthus x giganteus* is generally more productive than switchgrass, yielding 25–30 t/ha annually, but can also have greater loss of biomass at harvest

than switchgrass. One major unknown for *Miscanthus x giganteus* productivity that requires additional research is the effect of environmental conditions on the productivity of the species. Other challenges presented by *Miscanthus x giganteus* as a biomass feedstock are related to the self-incompatible nature of the species. This self-incompatibility has made genetic research on the species challenging and has hampered genetic improvements to date (but this also what eliminates the invasion risk of *Miscanthus x giganteus*). Recently, a low-density genetic map has been produced for *Miscanthus x giganteus* and has resulted in some genetic studies aimed at the heritability of select agronomic characteristics.

Other advanced dedicated bioenergy feedstocks that are currently under development are woody plant species such as *Populus* spp. (poplar) and *Salix* spp. (willow), also known as short-rotation woody crops (SRWCs). Of the woody plant species being considered for use as a biofuel feedstock, hybrid poplar received the most attention because it has the potential to yield biomass for conversion at a high rate. For example, yields of poplar are estimated between 12.4 t/ha on nonirrigated and unfertilized land to 22.5 t/ha on land that has been irrigated and fertilized. Poplar is also attractive because a number of genomic and genetic tools such as a fully sequenced genome are available for use by the existing research community. However, major setbacks associated with the use of poplar as a biofuel feedstock is the long generation time of the plant as well as the long-term sustainability of yields under low nitrogen inputs to the soil. Engineering the species for increased yields and decreased nitrogen requirements will therefore be important steps in developing poplar as a sustainable alternative energy solution. These improvements have the potential to also enhance production for the timber and paper industries as well (Field et al. 2008).

Finally, a readily available cellulosic feedstock can be gathered from agricultural waste products. Agricultural wastes include corn stover (the leaves and stalks of the corn crop), sugarcane residues, and rice hulls, as well as a number of other agricultural residues. Because of the inefficiencies that exist in current conventional agriculture production, these waste products remain among the most important feedstocks for biofuel production. In addition, the use of agricultural wastes is generally thought to be a better option for biofuel production than food crops such as corn because these waste products are the by-products of existing agricultural activity. This reduces the energy requirement needed for production as well as the use of pesticides, fertilizers, and water. In many cases, if agricultural wastes are not used for biofuel production, they are either burned or disposed in a landfill, both of which can have a higher environmental impact than the production of biofuel. Additionally, forest residues from logging, as well as municipal wastes, also have the potential to be used as cellulosic bioenergy feedstocks.

## **The Case for Liquid Biofuels in the World Energy Market**

As world energy demands, as well as the negative impacts of fossil fuel combustion on the natural environment, human health, and global economies continue to

increase, the case for liquid fuels produced from sustainable sources becomes stronger. Many factors contribute to this reality. First, because of the wide geographic ranges of the biomass feedstocks described above, biofuels present an opportunity for increased domestic energy production. This new area for domestic energy production also creates a unique opportunity for economies to further develop agricultural industries and build up rural communities. In addition to providing energy security, biofuels have the ability to decrease greenhouse gas emissions by reducing the use of fossil fuels while simultaneously increasing the potential for long-term sequestration of atmospheric carbon dioxide in plant tissues and soils.

Another attractive feature of liquid biofuels is the amount of energy, especially in the transportation sector, that they can generate currently and in the future. The earth receives annually  $\sim 3.8 \times 10^6$  exajoules ( $1 \text{ EJ} = 10^{18} \text{ J}$ ) of solar energy. This is such an extraordinary amount of energy that it meets the total global annual energy demand ( $450 \text{ EJ year}^{-1}$ ) in 1 h of daylight. Globally, plants fix many times more ( $\sim 2,900 \text{ EJ year}^{-1}$ ) than the total annual global energy demand by converting this solar energy into standing biomass through photosynthesis, a term known as net primary productivity (NPP). Unfortunately, not all of this energy is available for use as bioenergy feedstocks on a sustainable basis. Of this total, a number of estimates have placed the total energy potential from sustainable biomass feedstocks at  $100\text{--}300 \text{ EJ year}^{-1}$ . Currently, only  $40\text{--}55 \text{ EJ year}^{-1}$  of energy is produced from biomass globally. Because of the significant potential for much higher proportions, a number of developed nations have committed to significantly increasing the amount of bioenergy used by the year 2050. The Intergovernmental Panel on Climate Change estimates that by the year 2050 global energy demand will increase to  $\sim 560 \text{ EJ year}^{-1}$ . Currently, the energy generated from biomass has the potential to meet 32 % ( $180 \text{ EJ year}^{-1}$ ) of the future global energy demand laid out by the IPCC. This proportion is projected to increase to 46 % ( $325 \text{ EJ year}^{-1}$ ) by 2100. While the use of energy derived from biomass has the potential to change the global energy portfolio, there are a number of pressing environmental and sustainability considerations that must be accounted for now and in the future (Field et al. 2008).

---

## **Ecological Considerations Associated with Cellulosic Biofuel Production**

### **Management Decisions**

Cellulosic bioenergy crops are generally associated with fewer negative ecological consequences than first-generation bioenergy feedstocks and may even provide many benefits to the environment. Because these crops are highly productive, have extensive root systems, require few water and nutrient inputs, and can be grown for many years without requiring replanting, cellulosic bioenergy feedstocks may reduce greenhouse gas emissions, sequester soil organic carbon, improve soil and water quality, and create wildlife habitat. However, the extent to which

cellulosic bioenergy production will impact the environment will depend on the crops used, the land-use changes associated with them, and how these crops are managed. If managed poorly, their production may not provide any ecological benefits over first-generation bioenergy crops and may even adversely impact the environment. When evaluating the ecological consequences associated with cellulosic bioenergy production, the following factors should be considered.

### **Land-Use Change**

Bioenergy production on a scale large enough to meet current energy needs will require substantial areas of land provided via some form of land-use change (LUC) (Davis et al. 2011a, b). Land-use changes can directly alter existing land (e.g., agricultural land used previously for the production of other crops, natural ecosystems converted to agricultural land, or marginal land that is degraded from extensive agricultural practice and is unsuitable for further food crop production). Bioenergy cultivation can also indirectly induce land-use changes (i.e., indirect land-use change, iLUC), when uncultivated land is altered to produce a crop that was displaced by bioenergy feedstock cultivation on current agricultural land. For instance, in a hypothetical scenario where biofuel production replaces wheat production on a farm, the farmer may convert a native ecosystem in another area to compensate for the lost wheat production in the original location.

When addressing iLUC, it is important to note that tracking and predicting the many variables associated with iLUC is extremely difficult and associated with large uncertainty. Some agencies, such as the United States Environmental Protection Agency (EPA), have tried to evaluate effects of iLUC using models, but the models produce results with large variance because accounting for all associated variables can be difficult (Davis et al. 2011a). Although estimating the effects of iLUC is difficult, most models indicate that there are unintended consequences of biofuel policies for land use that should be addressed. Because iLUC has the potential to be counterproductive to mitigating climate change (through the development of uncultivated land), awareness of this potential consequence is key for policy makers to place regionally appropriate constraints on biofuel development.

Evidence is mounting that integrated land management policies might reduce unintended consequences for LUC and iLUC (Davis et al. 2011a). As discussed previously, many opportunities exist to coproduce biofuels and other resources, reducing the need for additional land. Wastes from many industries could serve as biofuel feedstocks, including residues from the timbering, paper, wood products, building, and agricultural industries. However, opportunities for integrated land management are often regionally specific and cannot be generalized globally for similar industries.

### **Crop Management**

Various management decisions associated with annual first-generation biofuel crop production, such as planting, harvesting, and tilling methods, can also determine the environmental impact of a bioenergy crop (Howarth and Bringezu 2009). For example, bioenergy crops can be planted as monocultures (single-species crop

stands) or in mixed assemblages (multiple species are planted in crop stands). They can also be planted in crop rotations, where different crops are planted and harvested on a rotational schedule (e.g., crop A is grown and harvested for 5 years and then crop B is grown and harvested for 5 years). The timing and pattern of crop harvest can also vary. Crops can be harvested annually, more than once per growing season, or less than once per growing season. The entire crop stand can be harvested at once, or alternatively, only sections of the stand can be harvested at once (i.e., strip harvesting). Crops can be harvested during different seasons. Finally, agricultural soil can be tilled, or mechanically disturbed to facilitate crop planting, in a variety of ways. Intensive tillage methods leave few crop residues, whereas less intensive, or reduced, tillage methods leave greater amounts of crop residues. No-till strategies do not till agricultural soil prior to planting, which leaves crop residues undisturbed. Strip-till methods only disturb the soil directly where the crop is planted, leaving strips of untilled soil between. Rotational tillage only tills the soil at particular intervals (i.e., every other year). Variation in any of these factors will ultimately influence the degree to which cellulosic bioenergy crops affect greenhouse gas emissions, soil properties, soil and water quality, and wildlife.

## Greenhouse Gas Emissions

Greenhouse gases are chemical compounds that absorb infrared radiation and trap heat in the atmosphere. Although this “greenhouse effect” is a naturally occurring process and is responsible for warming the planet by about 33 °C, human activities such as deforestation and fossil fuel combustion have increased the concentrations of many greenhouse gases in the atmosphere, which has further increased the temperature of the earth in recent years and driven other phenomena associated with global climate change (Schlesinger and Bernhardt 2013). Common greenhouse gases include carbon dioxide (CO<sub>2</sub>), water vapor, and other trace gases such as methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), tropospheric ozone (O<sub>3</sub>), and chlorofluorocarbons (CFCs). These gases are relatively inert so they remain in the troposphere (the lower atmosphere) for a long time and have greater potential to absorb more radiation over time compared to more reactive, short-lived gases. Thus, greenhouse gases can have long-lasting consequences on atmospheric chemistry and global climate once released into the troposphere by human activities such as fossil fuel combustion and land-use change.

Bioenergy feedstocks have the potential to mitigate global climate change phenomena because they act as carbon sinks by sequestering atmospheric CO<sub>2</sub> as they grow and because they can offset anthropogenic greenhouse gas emissions by slowing fossil fuel exploitation (Williams et al. 2009). However, land-use conversion and management practices associated with crop cultivation can also release greenhouse gases that may reduce or in some cases completely eliminate bioenergy potential to offset anthropogenic greenhouse gas emissions. For instance, bioenergy feedstock cultivation requires land, which usually involves a land-use change that



reduces soil organic carbon (SOC) and releases CO<sub>2</sub> into the atmosphere (see section “[Soil Organic Carbon](#)”). The magnitude of CO<sub>2</sub> release, however, depends on the type of land-use change used to cultivate the bioenergy crop. Land-use changes release more CO<sub>2</sub> if highly productive land, such as a forest ecosystem, is converted to a bioenergy crop field than if the conversion occurs on agricultural land or marginal land (a low productivity ecosystem) (Davis et al. 2011b). The bioenergy feedstock chosen for cultivation can also determine the impact of bioenergy production on greenhouse gas emissions over time. Perennial grasses, for instance, can grow for many years without the need to till and replant, resulting in greater accumulation of SOC relative to an annual cropping system (Blanco-Canqui 2010). Increased SOC sequestration in dedicated herbaceous bioenergy crops relative to first-generation bioenergy crops can therefore create a net greenhouse gas sink if this perennial system replaces annual corn agriculture. A popular metric that is used to determine if land-use change results in positive or negative consequences for ecosystem services is the payback time needed to neutralize the carbon debt incurred through soil disturbance and the removal of vegetation from the landscape (Davis et al. 2011a). The payback time is dependent on the original condition of the land (e.g., soil, aboveground plant community, management history), climate, and the rate at which the biofuel agroecosystem sequesters carbon.

Bioenergy production can also release N<sub>2</sub>O emissions if substantial fertilizer inputs are used to grow the crop. Fertilizers add nitrogen to the soil in the form of ammonium (NH<sub>4</sub><sup>+</sup>), which can increase rates of microbial nitrification and denitrification in the soil and subsequently produce gaseous nitric oxide (NO) and N<sub>2</sub>O as by-products (Schlesinger and Bernhardt 2013). N<sub>2</sub>O is highly inert and has a long residence time in the atmosphere, so it has great potential to warm the atmosphere over time (about 300× greater than atmospheric CO<sub>2</sub>). N<sub>2</sub>O also breaks down into NO when exposed to ultraviolet radiation in the stratosphere, which can promote stratospheric ozone destruction and subsequently increase the amount of harmful solar radiation that reaches the surface of the planet. Therefore, N<sub>2</sub>O production associated with agricultural activities can have wide-ranging consequences for atmospheric chemistry and climate. Cellulosic feedstocks are generally less likely to produce N<sub>2</sub>O emissions than first-generation bioenergy crops because dedicated herbaceous bioenergy crops are often characterized by high nutrient-use efficiency and can sometimes be grown without the addition of nitrogen fertilizer (see section “[Nutrient and Chemical Inputs](#)”) (Williams et al. 2009). Low nitrogen inputs reduce rates of nitrification and denitrification in the soil, which can ultimately reduce N<sub>2</sub>O emissions. Dedicated herbaceous bioenergy crops can also be grown under drier conditions than traditional row crop monocultures, which may reduce rates of denitrification and reduce N<sub>2</sub>O emissions compared to first-generation bioenergy crops. However, soil disturbance associated with land conversion can also accelerate nitrogen cycling processes, which may increase N<sub>2</sub>O emissions associated with the establishment of a dedicated herbaceous bioenergy crop despite their low-nutrient requirements. Therefore, bioenergy feedstock, land-use, and crop management must all be considered when assessing the impact of biofuel production on terrestrial N<sub>2</sub>O emissions.

## Soil and Nutrient Management

Many of the land-use changes and management strategies used to cultivate first-generation bioenergy crops can impact soil properties such as soil hydraulics, soil chemistry, and soil biodiversity. These crops are tilled often and require vast water and nutrient inputs, which reduces soil porosity, nutrient quality, water-holding capacity, and microbial activity, ultimately reducing soil productivity and exacerbating erosional processes. The biological characteristics and management requirements of cellulosic bioenergy feedstocks, however, have the potential to improve the biological, chemical, and physical properties of soils. These crops are highly productive, have extensive root systems that penetrate deep into the soil profile, and require few water and nutrient inputs, which can improve soil aggregation, soil hydraulic conductivity, soil water infiltration, water retention, soil organic matter, and nutrient retention. Perennial bioenergy crops therefore have great potential to improve degraded soils, although the degree to which these feedstocks can improve soil properties depends on the crop used, where the crop is grown, and how the crop is managed.

### Soil Erosion and Runoff

Surface runoff is the movement of water across a land surface (typically soil) that occurs when the soil is saturated or when the rate of precipitation is greater than the rate of water infiltration in the soil. Runoff can result in soil erosion, the transport of soil materials (i.e., nutrients, organic material, or contaminants) by some natural process (such as water or wind movement) to a different location. Runoff and soil erosion typically occur in agricultural systems when soil is harvested or disturbed so that biomass cover is reduced and/or the soil is compacted (U.S. Congress Office of Technology Assessment 1993). When biomass cover is reduced, a greater proportion of rainfall hits exposed soil, which dislodges particulate matter and washes nutrients and organic matter away from the upper soil layers. Runoff also occurs when soil becomes compacted because soil porosity (the amount of “empty” spaces in the soil) and water infiltration are reduced, increasing the rate of soil saturation. This can negatively impact agricultural systems because the loss of soil nutrients and organic matter associated with erosion reduces soil productivity and plant growth.

Runoff and erosion are often associated with the cultivation of annual row crops because these crops do not produce dense stands and also because they are managed extensively with large equipment during planting and harvesting each year. However, dedicated herbaceous bioenergy crops have the potential to reduce runoff and soil erosion rates (Lemus and Lal 2005). Perennial C<sub>4</sub> grasses, in particular, are high yielding and produce dense stands that intercept large quantities of rainfall, reducing the amount of water that directly hits the soil. Additionally, dense stands and the litter layers associated with them can reduce wind erosion. These species also have extensive root systems that decrease soil compaction, promote soil aggregation, and increase soil porosity, which can increase the amount of water that permeates deep soil layers. Finally, many perennial crops are replanted infrequently with some



species only replanted every 15–20 years, which reduces the degree of management by heavy equipment and thus reduces the risk of soil compaction. Reduced erosion can benefit agricultural and natural systems by maintaining soil structure, retaining soil organic matter and nutrients, and reducing the transport of undesirable nutrients and/or contaminants to other natural systems (e.g., nutrient deposition and in aquatic systems).

The degree to which cellulosic bioenergy crops reduce soil erosion depends on the crop used and how the crop is managed (Williams et al. 2009). Runoff and erosion are generally reduced by perennial  $C_4$  grasses and short-rotation woody crops. Conversely, harvesting annual crop residues such as corn stover may actually exacerbate the rate of surface runoff and soil erosion in an agricultural system because residue removal exposes soil to wind and rainfall and the heavy equipment used to remove the residues can compact the soil. Harvesting the crop during the winter or when the soil is dry can however reduce soil compaction. Minimum or no-till farming, as well as contour plowing (plowing along the landscape's elevation contour to form furrows that capture water), can also reduce surface runoff and erosion. Cellulosic bioenergy crops that are managed more intensely (i.e., are harvested multiple times throughout the year or are extensively tilled) can also counteract the benefits of perennial grasses on soil structure. The degree to which soil erosion is reduced by cellulosic bioenergy crops can depend on the type of soil in which the crop is growing, as well as on the length of time following establishment. Perennial  $C_4$  grasses, for instance, may not reduce erosion in the first year they are planted. In fact, these crops may not improve soil structure or soil hydraulic properties for many years after they are established (Howarth and Bringezu 2009). Therefore, cellulosic bioenergy crops do have the potential to reduce surface runoff and soil erosion, although this depends on crop management and may take decades.

### **Nutrient and Chemical Inputs**

Cellulosic bioenergy crops, particularly dedicated herbaceous bioenergy crops, can potentially benefit soil nutrients and nutrient cycling processes. Some perennial  $C_4$  grasses have low-nutrient requirements and high nutrient-use efficiency (i.e., they produce more biomass per fewer units of essential nutrients such as nitrogen or phosphorous); thus, they require little fertilizer inputs and can be grown on degraded, marginal soils (Carroll and Somerville 2009). These crops also require less herbicide and pesticide inputs than annual row crops, particularly because these chemicals are only applied in the first year of establishment and because these perennial crops are grown for many years (U.S. Congress Office of Technology Assessment 1993). Nutrients and chemicals are also better retained in the soil by dedicated herbaceous bioenergy crops because the organic material added to the soil by highly productive perennial  $C_4$  grasses provides a surface to which nutrients can adhere and because perennial roots retain nutrients between growing seasons. This has the potential to enhance crop productivity, as well as the productivity and diversity of soil microorganisms. However, these benefits are primarily associated with perennial  $C_4$  grasses and short-rotation woody crops; annual crop residue removal actually reduces essential plant nutrients from the soil and degrades soil quality.

The low chemical inputs required for cellulosic bioenergy crop cultivation can provide several benefits to the environment. First, low fertilizer inputs can greatly reduce energy consumption because the production of industrial nitrogen fertilizer (i.e., industrial nitrogen fixation via the Haber-Bosch process) is an energetically expensive process. Second, low fertilizer, herbicide, and pesticide inputs can improve soil quality and reduce the amount of chemicals that are present in surface runoff, thus reducing rates of nitrification and denitrification (see section “[Greenhouse Gas Emissions](#)”) and harmful ecological processes such as nitrogen leaching and eutrophication (see section “[Water Quality](#)”). Low fertilizer inputs, for example, can reduce nitrogen leaching in the soil by reducing rates of nitrification. Nitrification is the two-step process by which aerobic chemoautotrophs oxidize ammonium ( $\text{NH}_4^+$ ) to nitrite ( $\text{NO}_2^-$ ) and then nitrate ( $\text{NO}_3^-$ ), a highly soluble form of nitrogen (Schlesinger and Bernhardt 2013). Increasing  $\text{NH}_4^+$  inputs to a system via fertilization increases rates of nitrification and ultimately increases the concentration of soluble  $\text{NO}_3^-$  that can leach through the soil and contaminate groundwater. Thus, bioenergy crops that require low nitrogen inputs will reduce  $\text{NO}_3^-$  leaching associated with agricultural practices. Proper management regimes have the potential to enhance these environmental benefits. For instance, more nutrients can be retained in the soil by harvesting biomass after plant senescence, when nutrients have been translocated belowground to roots. Planting crop stands in mixed assemblages with nitrogen-fixing plant species interspersed among the biomass crop may also reduce the need for additional nitrogen input.

### Soil Organic Carbon

Soil organic carbon (SOC) is ecologically important in the global carbon cycle. This soil reservoir of organic residues contains approximately 1,500 Pg carbon, almost twice the amount of carbon contained in the atmosphere (approximately 780 Pg) and three times the amount stored in terrestrial biota (approximately 500 Pg) (Schlesinger and Bernhardt 2013). Thus, changes in the amount of carbon stored in soil, particularly reductions in SOC, can greatly impact other carbon cycling processes. Carbon lost from the soil primarily returns back to the atmosphere through heterotrophic respiration, which can have cascading effects on carbon fluxes between other carbon pools (e.g., atmosphere–ocean  $\text{CO}_2$  exchange). Reductions of SOC can also impact terrestrial systems by decreasing plant productivity, degrading soil quality, and decreasing water retention. SOC loss is caused by a variety of factors including soil erosion, root biomass reduction, or soil disturbances that increase decomposition rates and microbial respiration via increases in soil aeration and temperature (Lemus and Lal 2005). Although this is a naturally occurring process, intense agricultural management and land-use changes that convert natural ecosystems to agricultural land greatly increase the amount of carbon that is lost from the soil.

Perennial feedstocks have the potential to mitigate SOC losses associated with land-use changes by sequestering atmospheric  $\text{CO}_2$  and adding substantial amounts of organic material back to the soil carbon pool (Lemus and Lal 2005). For instance, the high yields associated with dedicated herbaceous bioenergy crops return

organic carbon back to the soil in the form of aboveground residues and root dieback. The extensive root systems produced by these crops also grow deep into the soil profile, which transfers organic carbon to deep soil layers where SOC decomposition rates are low. Thus, carbon inputs to the soil may be larger than carbon outputs, increasing SOC over time. Increasing SOC is highly beneficial in an agricultural system; higher SOC levels can improve soil structure, buffer soil acidity, increase crop quality and productivity, increase the abundance of soil microorganisms, reduce runoff, and improve water quality (U.S. Congress Office of Technology Assessment 1993). However, the amount of SOC that bioenergy crops can add to a system depends on a variety of factors, including soil type, climate, and land management. The amount of carbon that feedstocks can sequester and add to the soil also depends on the amount of carbon already present in the soil because soil can eventually become saturated with carbon. Although absolute limits are debated, greater amounts of carbon can be added to degraded soil that is carbon-depleted than highly productive soil that is closer to its carbon saturation point (Blanco-Canqui 2010). These crops therefore have greater potential to improve marginal lands compared to more productive lands. The amount of organic carbon that is added to the soil by bioenergy crops depends on the crop used and the way the crop is managed. Perennial  $C_4$  grasses and short-rotation woody crops tend to increase SOC, but removing annual crop or forest residues actually decreases SOC by directly removing organic material from the soil and by exposing the soil to higher air temperatures that increase rates of organic material decomposition (Lemus and Lal 2005; Williams et al. 2009). Greater amounts of SOC are also retained in the soil when crops are harvested less frequently and minimum or no-till farming regimens are used.

## Water

### Water Requirements

Agricultural crops, including food crops and first-generation bioenergy crops, can be characterized by low water-use efficiency (they assimilate less carbon per unit water transpired) and are sometimes irrigated with water collected from lakes, rivers, and groundwater to produce higher yields. This can have negative socioeconomic and environmental consequences because irrigation aggravates water shortages and reduces surface water flow necessary for wetland ecosystems and aquatic biota. Many dedicated herbaceous bioenergy crops can produce high yields without irrigation because these perennial grasses utilize the  $C_4$  photosynthetic pathway and use water more efficiently than plants that utilize the  $C_3$  photosynthetic pathway (Carroll and Somerville 2009; Williams et al. 2009). In addition, many perennial bioenergy feedstocks have extensive deep root systems that aid in retaining water in the soil more than the small root systems associated with annual row crops, further reducing the need for irrigation (Howarth and Bringezu 2009). Because they do not require as much irrigation, cellulosic bioenergy feedstocks compete less with food crops for water and are also less likely to impact aquatic systems than

first-generation bioenergy crops. There are some other considerations to be accounted for in water-use of biofuel species, including the length of growing season that may substantially increase the water needs across the growing season. Finally, these crops do require some additional water for chemical processing; however, they do not require greater amounts than processing first-generation bioenergy crops.

### **Water Quality**

Cellulosic bioenergy feedstocks can also improve water quality relative to annual row crops. These crops do not require substantial chemical inputs, and their extensive root systems, as well as their ample SOC inputs, reduce surface runoff and soil erosion. This can decrease chemical contamination of aquatic habitats and can subsequently reduce nitrogen leaching (see section “[Nutrient and Chemical Inputs](#)”) and aquatic eutrophication (i.e., aquatic ecosystem responses to nutrient additions). Thus, these feedstocks are less associated with negative aquatic processes such as phytoplankton or algal blooms and hypoxic conditions (oxygen depletion) than annual crops (Blanco-Canqui [2010](#)).

### **Impacts on Wildlife and Biodiversity**

Land-use changes associated with bioenergy production will likely affect various aspects of biodiversity including the number of species in a given habitat (species richness) and/or the relative abundance of each species in a given habitat (species evenness), which can potentially have cascading consequences on other biological processes at the community and ecosystem scales. Generally, land-use changes that convert natural ecosystems to agricultural land result in habitat loss and habitat fragmentation, which can ultimately reduce species richness and alter species evenness (Dauber et al. [2010](#)). Cellulosic bioenergy crops that directly or indirectly displace natural habitat can therefore negatively impact wildlife and biodiversity. If planted on marginal lands, these crops may have neutral or even positive impacts on wildlife. Perennial grasses such as switchgrass and *Miscanthus x giganteus* can improve the quality of degraded habitats and create an environment that structurally resembles a natural grassland ecosystem, which can provide nesting and foraging habitat for many birds and small mammals (Williams et al. [2009](#)). These high-yielding grasses also produce large amounts of litter and are seldom tilled, which provides substantial, undisturbed cover for ground-dwelling species.

However, wildlife benefits from cellulosic bioenergy cultivation will only occur if the feedstock is managed correctly. Perennial grasses planted in monoculture may actually reduce wildlife biodiversity if the crop system replaces a high productivity ecosystem because monoculture fields decrease environmental heterogeneity and reduce the number of species that can occupy an area (U.S. Congress Office of Technology Assessment [1993](#)). Switchgrass monocultures, for instance, primarily provide habitat for grassland birds that favor tallgrasses (although birds that prefer less cover may become more abundant following harvesting). Conversely, crops grown in mixed assemblages (i.e., two to three species) can enhance landscape

heterogeneity and create a more diverse environmental mosaic that can support more species in a given area. Harvesting strategies may also affect the degree to which bioenergy crops impact biodiversity (Fargione et al. 2009). Frequent harvests (>1 harvest per year) may favor species that prefer a short-grass habitat, while infrequent harvests may favor species that prefer tallgrasses. Rotational or strip harvesting can improve environmental heterogeneity and support the coexistence of multiple species that prefer different habitats. Crop harvests can also interfere with avian breeding seasons, so harvesting in the autumn or winter, after the breeding season of many bird species has ended, may benefit a variety of bird species (Dauber et al. 2010). However, autumn or winter harvests can reduce ground cover and consequently increase winter mortality for many ground-dwelling birds and mammals. Crop management strategies can therefore have wide-ranging impacts on many wildlife species, and these consequences must be carefully considered when making land management decisions to cultivate cellulosic bioenergy crops.

Other cellulosic bioenergy crops may also impact wildlife and biodiversity. For example, short-rotation woody crops can provide habitat for birds and small mammals, although these habitats are often less suitable than natural forests because crop stands are less complex than naturally occurring forest ecosystems (Dauber et al. 2010). Woody crops may also reduce habitat fragmentation if planted as a corridor to connect separated forest patches. Reduced habitat fragmentation can facilitate the movement of individuals and populations between habitats and is ultimately associated with high biodiversity. Finally, annual crop residues, as well as forest residues, tend to have fewer impacts on biodiversity than short-rotation woody crops or perennial grasses because their collection is not associated with land-use changes that reduce viable habitat or environmental heterogeneity. Residue removal, however, does reduce ground cover for wildlife and also decreases soil nutrients, which may impact the biodiversity of ground-dwelling animals or soil microorganisms. Therefore, the type of bioenergy feedstock used, as well as the strategy used to plant and maintain the crop, can strongly influence the degree to which cellulosic bioenergy cultivation impacts wildlife and biodiversity.

## **Invasive Species Potential**

### **Bioenergy Crops as Invasive Species**

An invasive species is one that occurs in location that is not part of its original (i.e., native or endemic) range. In order to successively invade a new range, a non-native plant must have certain characteristics that enable it to overcome multiple barriers (i.e., physical dispersal barriers, novel environmental conditions, competition with new species, predation by new enemies) (Hierro et al. 2005). Therefore, invasive plants typically have high relative growth rates, high competitive abilities, high fecundity under optimal conditions, and morphological and/or physical similarity to the native species in its new range.

Interestingly, these characteristics are also those associated with many cellulosic bioenergy crops. Dedicated herbaceous bioenergy crops, for example, are perennial, have rapid growth rates, produce high yields, utilize the C<sub>4</sub> photosynthetic pathway, have high resource-use efficiency, and propagate both vegetatively and by producing a seed crop. These species are also broadly adapted across a wide geographic range and are tolerant of various environmental conditions. For instance, crops such as switchgrass and *Miscanthus x giganteus* are tolerant of drought, as well as flooding, and can grow on low-nutrient, degraded soils. These traits promote efficient seedling establishment and quick production of high yields with relatively little water and nutrient inputs. However, these traits may also promote the undesirable invasion of bioenergy crops into nonagricultural areas, particularly if the crop is cultivated outside of its native range or if the crop is genetically modified to enhance qualities that concomitantly increase invasiveness (Raghu et al. 2006; Williams et al. 2009). Unintentional introduction can occur locally or on larger scales as a result of direct spread from the agricultural land source or by propagule release during harvesting and processing (Fargione et al. 2009). Biomass feedstocks are typically harvested following plant senescence, when seeds have been produced and are still attached to the plant, which can result in seed rain onto roadsides during transportation to biofuel production facilities. These seeds may also contaminate the equipment used to plant or harvest the crop, which may subsequently taint other agricultural crops if the equipment is not properly sterilized.

A non-native bioenergy crop may survive and form persistent populations because it will likely experience a decrease in pressure from specialist enemies (i.e., specialist pathogens and herbivores) when introduced to a new region (Hierro et al. 2005). The non-native species is not typically susceptible to the specialist enemies of the native species in its new range (assuming that these specialist enemies do not switch host preference to the invader) and should therefore experience a decrease in regulation by enemies relative to the native species in the new region. The risk of invasion, however, may decline if native crops or sterile cultivars (such as *Miscanthus x giganteus*) are cultivated, although other traits associated with these species may promote their invasiveness despite their lack of a viable seed crop. Invasiveness is not typically associated with other cellulosic bioenergy sources such as annual crop residues and short-rotation woody crops.

### **Risk of Invasion by Other Species**

Depending on how the crop is planted and maintained, cellulosic bioenergy crops also have the potential to increase the risk of invasion by other species in bioenergy agricultural lands. Dedicated herbaceous bioenergy crops can particularly promote the invasion of other non-native species if the crop is planted as a monoculture. Generally, habitat homogeneity can increase the susceptibility of a location to invasion by non-native species because less diverse communities (communities with fewer species) have more available resource niches compared to more heterogeneous communities, which can be utilized by an introduced species (Hierro et al. 2005). Cultivating bioenergy crops in mixed assemblages, however, may reduce the number of available niches in a community and subsequently reduce this

risk of invasion. Growing multiple genotypes of a single species may also increase landscape heterogeneity and reduce invasions by other species. Similarly, other species may become invasive in bioenergy agricultural lands if substantial amounts of water and/or nutrients are added to the crop, which may create more available resource niches that can potentially be utilized. Most cellulosic bioenergy crops, however, do not require substantial water or nutrient inputs, so this risk may actually be lower compared to traditional row crops.

## **Pests and Pathogens**

Cellulosic bioenergy crops can become infected by a variety of pests and pathogens including viruses, bacteria, fungi, insects, molds, and nematodes. Depending on the host and the type of disease, these infections have the potential to reduce photosynthetic rates, impair plant-water relations, decrease reproductive output, and ultimately reduce whole-plant yield and survival. This can significantly reduce the productivity of a crop stand and even impact other agricultural and natural ecosystems if the pathogen is transmitted via insects that can travel long distances. Thus, the interaction between cellulosic bioenergy crops and their pathogens can have significant consequences on both local and larger spatial scales.

The risk of infection by pests and pathogens may be a significant concern for dedicated herbaceous bioenergy crops because bioenergy cultivars can be genetically homogenous and are usually planted in monoculture. Generally, the probability of pathogen transmission between hosts increases with host abundance and distribution (Gonzalez-Hernandez et al. 2009). In natural ecosystems, susceptible hosts often co-occur with other species in a nonuniform distribution, decreasing the likelihood that pathogens will physically transfer from host to another. This probability is considerably higher when many individuals of the same species co-occur in a given area and are spaced uniformly, so herbaceous bioenergy monocultures are particularly vulnerable to the spread of pathogens and pests. Bioenergy cultivars may also be more susceptible to pests and pathogens because breeding programs have selected for certain traits that improve their yield and resistance to adverse environmental conditions (e.g., rapid growth rates, high resource-use efficiency, etc.); in doing so, bioenergy cultivars are somewhat genetically homogenous. This can increase the rate at which a pest or pathogen can adapt to a particular host genotype and will ultimately increase the probability of pathogen spread, as well as pathogen virulence (Gonzalez-Hernandez et al. 2009). Additionally, selecting for high yields may increase a cultivar's susceptibility to infection because plants that allocate more resources to growth typically invest fewer resources to defensive mechanisms (Schrotenboer et al. 2011). Quick-growing perennial grasses, therefore, have the potential to become highly susceptible to detrimental pests and pathogens. Planting bioenergy crops in mixed assemblages to enhance genotypic or species diversity, or even using crop rotations to disrupt the life cycles of many pests and pathogens, may reduce this risk and prevent the spread of disease within a crop stand and between other ecosystems.

## Future Directions

Biofuels produced from cellulosic sources have the potential to reduce the need for fossil fuel energy in the future. As mentioned throughout the previous sections, advanced cellulosic crops generally possess many ecological advantages over first-generation feedstocks. Cellulosic bioenergy crops typically have extensive rooting systems and produce high yields without requiring large water or nutrient inputs. These characteristics can increase SOC and reduce rates of greenhouse gas emissions, runoff, and eutrophication. Additionally, perennial cellulosic crops can be cultivated on land considered marginal for agricultural production and improve wildlife habitat quality. However, cellulosic feedstocks are also difficult to process into liquid fuel, and depending on how they are managed, their impact on natural ecosystems may not always be positive. Therefore, the development of cellulosic biofuels for widespread future production requires continued research in areas of feedstock propagation and conversion technologies. Specifically, future work in the development of biofuels from cellulosic sources should aim to improve the conversion of cellulosic biomass to liquid fuel. Genetically engineering bioenergy crop species to make lignocellulosic material easier to hydrolyze, either by reducing or modifying lignin content, may increase the cost efficiency of liquid biofuel production. Work in the future should also focus on the development of new enzymes that are better able to break down lignocellulosic biomass. Ultimately, these developments will require more research to better understand plant cell wall chemistry. A better understanding of the environmental impacts associated with bioenergy feedstock production is also needed. Bioenergy crops can have various impacts on the environment, depending on the crop used and how the crop is managed, so predicting how bioenergy production will impact various ecosystems in the future can be difficult. This will be especially important in the face of global climate change, as different crops will likely respond differently to changes in atmospheric chemistry and climate. Therefore, if the full benefits of cellulosic bioenergy production are to be realized, a dedication must be made to the production and management of bioenergy feedstocks that not only have few adverse impacts on the environment but that are also more efficient in generating liquid fuels from lignocellulosic material.

---

## References

- Bhagwan Goyal H, Saxena RC, Seal D. Thermochemical conversion of biomass to liquid and gaseous fuels. In: Pandey A, editor. Handbook of plant-based biofuels. Boca Raton: CRC Press; 2009. p. 29–43.
- Blanco-Canqui H. Energy crops and their implications on soil and environment. *Agronomy J*. 2010;102:403–19.
- Carroll A, Somerville C. Cellulosic biofuels. *Annu Rev Plant Biol*. 2009;60:165–82.
- Dauber J, Jones MB, Stout JC. The impact of biomass crop cultivation on temperate biodiversity. *Glob Change Biol Bioenerg*. 2010;2:289–309.
- Davis SC, House JI, Diaz-Chavez RA, Molnar A, Valin H, DeLucia EH. How can land-use modeling tools inform bioenergy policies? *J R Soc Interf Focus*. 2011a;1:212–23.



- Davis SC, Parton WJ, Del Grosso SJ, Keough C, Marx E, Adler P, DeLucia EH. Impacts of second-generation biofuel agriculture on greenhouse gas emissions in the corn-growing regions of the US. *Front Ecol Environ*. 2011b;10:69–74.
- Demirbas MF. World biofuel scenario. In: Pandey A, editor. *Handbook of plant-based biofuels*. Boca Raton: CRC Press; 2009. p. 13–28.
- Fargione JE, Cooper TR, Flaspohler DJ, Hill J, Lehman C, McCoy T, Nelson EJ, Oberhauser KS, Tilman D. Bioenergy and wildlife: threats and opportunities for grassland conservation. *BioScience*. 2009;59:767–77.
- Gonzalez-Hernandez JL, Sarath G, Stein JM, Owens V, Gedye K, Boe A. A multiple species approach to biomass production from native herbaceous perennial feedstocks. *In Vitro Cell Dev Biol Plant*. 2009;45:267–81.
- Hierro JL, Maron JL, Callaway RM. A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. *J Ecol*. 2005;93:5–15.
- Howarth RW, Bringezu S. Biofuels: environmental consequences and interactions with changing land use. Proceedings of the Scientific Committee on Problems of the Environment (SCOPE) International Biofuels Project Rapid Assessment; 2008 Sept 22–25; Gummertsbach, Germany. Ithaca/New York: Cornell University; 2009.
- Lemus R, Lal R. Bioenergy crops and carbon sequestration. *Crit Rev Plant Sci*. 2005;24:1–21.
- Perlack RD, Wright LL, Turhollow AF, Graham RL, U.S. Department of Agriculture and U.S. Department of Energy. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. GO-102005-2135. Washington, DC: Government Printing Office; 2005.
- Raghu S, Anderson RC, Daehler CC, Davis AS, Wiedenmann RN, Simberloff D, Mack RN. Adding biofuels to the invasive species fire? *Science*. 2006;313:1742.
- Saxena RC, Adhikari DD, Goyal HB. Biomass-based energy fuel through biochemical routes: a review. *Renew Sustain Energy Rev*. 2009;13:167–78.
- Schlesinger WH, Bernhardt ES. *Biogeochemistry: an analysis of global change*. San Diego: Academic; 2013.
- Schrotenboer AC, Allen MS, Malmstrom CM. Modification of native grasses for biofuel production may increase virus susceptibility. *GCB Bioenerg*. 2011;3:360–74.
- Taiz L, Zeiger E. *Plant physiology*. 5th ed. New York: Sinauer; 2010.
- U.S. Congress Office of Technology Assessment. Potential environmental impacts of bioenergy crop production-background paper, OTA-BP-E-118. Washington, DC: U.S. Government Printing Office; 1993.
- Williams PRD, Inman D, Aden A, Heath GA. Environmental and sustainability factors associated with next-generation biofuels in the U.S.: what do we really know? *Environ Sci Technol*. 2009;43:4763–75.

## Further Reading

- Buckeridge MS, Goldman GH. *Routes to cellulosic ethanol*. New York: Springer; 2011.
- Burkheisser EV. *Biological barriers to cellulosic ethanol*. Hauppauge: Noval Science; 2011.
- Canfield D, Glazer AN, Falkowski PG. The evolution and future of Earth's nitrogen cycle. *Science*. 2010;330:192–6.
- Cheng J. *Biomass to renewable energy processes*. Boca Raton: CRC Press; 2009.
- Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Hogberg P, Linder S, Mackenzie FT, Moore III B, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W. The global carbon cycle: a test of our knowledge of earth as a system. *Science*. 2000;290:291–6.
- Field CB, Campbell JE, Lobell DB. Biomass energy: the scale of the potential resource. *Trends Ecol Evol*. 2008;23:65–72.

- Gomez LD, Steele-King CG, McQueen-Mason SJ. Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist*. 2008;178:473–85.
- Horne R, Grant T, Verghese K. Life cycle assessment: principles, practice and prospects. Collingwood: CSIRO; 2009.
- Pimentel D. Global economic and environmental aspects of biofuels. Boca Raton: CRC Press; 2012.
- Rosenberg NJ. A biomass future for the North American great plains: toward sustainable land use and mitigation of greenhouse warming, *Advances in global change research*. New York: Springer; 2007.
- Tilman D, Socolow R, Foley JA, Hill J, Larson E, Lynd L, Pacala S, Reilly J, Searchinger T, Somerville C, Williams R. Beneficial biofuels – the food, energy, and environment trilemma. *Science*. 2009;325:270–1.

Jason G. Hamilton

## Contents

Introduction .....	632
Sustainability: From Word to Concept .....	633
Developing the Concept of Sustainability: What Is Being Sustained? .....	633
Defining the Concept of Sustainability: Focusing on Positive Change for All .....	635
The Foundational Premises of Sustainability .....	635
Operationalizing Sustainability .....	638
The Development of Sustainability Science .....	643
Focusing Where Knowledge Is Most Needed .....	644
Sustainability Science Represents a New Conceptual Model for the World .....	647
Future Directions in Sustainability Science .....	650
References .....	653

---

## Abstract

- Sustainability is concerned with meeting the essential needs of the large numbers of people on this planet whose needs are not being met.
- The novel insight provided by the concept of sustainability is that humans and their local and global environments exist as complex social-ecological systems.
- Sustainability science is a new field of research that deals with the interactions between natural and social systems and with how those interactions affect the challenges of sustainability.
- In sustainability science, human/nonhuman and basic/applied dichotomies are abandoned for a new way of viewing the natural world – one in which human demands on global ecosystems is integrated into the capacity of those ecosystems to persist.

---

J.G. Hamilton

Department of Environmental Studies and Sciences, Ithaca College, Ithaca, NY, USA

e-mail: [jhamilton@ithaca.edu](mailto:jhamilton@ithaca.edu); [jasonghamilton@gmail.com](mailto:jasonghamilton@gmail.com)

- Developing the science of sustainability forces a deep questioning of what the appropriate role of science in society is.
- The focus of sustainability science and most modern socio-ecological studies is to work for improvements in human health, ecosystem health, societal health, and economic health.
- One of the best examples of how the new conceptual model of sustainability science has been put into practice is in the Millennium Ecosystem Assessment. This is a paradigmatic example of how sustainability science, working at scales from local to global and studying processes occurring from short to long time scales, fully integrates existing knowledge into a framework useful for supporting sustainability.
- All systems consist of three component categories: parts, interconnections, and functions. The concept of sustainability as operationalized in sustainability science reminds us to have a conversation: How should the current social-ecological system be replaced with one that has, as its purpose, human well-being? Further, sustainability reminds us that there is no human well-being that is separate from the well-being of the social-ecological system as a whole.

---

## Introduction

The biological sciences are traditionally organized by scale (cells, tissues, organisms, populations, communities, ecosystems, etc.), taxonomic grouping (plants, animals, fungi, etc.), or process (competition, mutualism, evolution, etc.). In all of these organizational schemes, there has been a tendency to view the natural world as divided into two fundamentally different parts: humans and everything else. Even integrative fields such as ecology have often focused on “pristine” systems in the sense of trying to understand how ecosystems operate in the absence of human influence. Although the technique of using simplified systems to understand the fundamental properties of more complex systems has a long tradition in science (e.g., the idea of the frictionless plane developed by Galileo), the very act of simplifying the system alters the balancing and reinforcing feedback loops that have the potential to amplify or mute fundamental interactions between human actions and ecosystem processes. Thus, in simplifying our view down to a system of humans and everything else, the capacity to understand, predict, and manage the emergent properties of our social-ecological system is lost.

In the plant sciences, the human/nonhuman dichotomy has tended to manifest along the lines of basic versus applied perspectives. However, new fields such as agroecology and sustainable ecosystem management are bridging this conceptual divide. Sustainability science offers another approach that not only bridges this divide but also explicitly connects the study of ecology (at all scales and of all taxonomic groups) with other fields of study, especially those in the social sciences. In sustainability science, the human/nonhuman, basic/applied dichotomies are abandoned for a

new way of viewing the natural world – one in which human demands on global ecosystems is integrated into the capacity of those ecosystems to persist.

While terms such as “sustainability,” “sustainable development,” “sustainability science,” and “ecological sustainability” are increasingly being used in both lay and scientific vernacular, there is still much confusion regarding the meaning and ultimately the application of these concepts. The goal of this chapter is to provide the background and context to understand the concept of sustainability and the relationship among sustainability, sustainability science, and ecology. In addition, it will explore the historical development of sustainability science, provide illustrative examples of the application in sustainability science, and explore future directions in the development of this new field.

---

## **Sustainability: From Word to Concept**

Originally a noun meaning nothing more than “the property of being sustainable,” the word “sustainability” has become the concept of sustainability. This concept is responsible for spawning global movements, informing worldwide political discourse, and sparking new areas of scientific inquiry. First articulated in 1987, the concept of sustainability was evocative enough for the United Nations General Assembly (Resolution 42/187 1987) to call for it to become the “central guiding principle of the United Nations, Governments and private institutions, organizations and enterprises. . .” In the subsequent 25 years, sustainability has become a household word, and sustainability science has developed into a new field of research in its own right. For example, the National Academy of Sciences has established the National Academies’ Roundtable on Science and Technology for Sustainability with the goal to “mobilize, encourage, and use scientific knowledge and technology to help achieve sustainability goals and to support the implementation of sustainability practices.” The *Proceedings of the National Academy of Sciences of the United States of America* has launched a new section devoted to sustainability science. The British Royal Society has adopted sustainability as one of its four organizing themes. The American Association for the Advancement of Science has established a center for Science, Technology, and Sustainability in support of this new scientific field. And the national science academies of the world’s largest economies (the G-8 nations plus Brazil, China, India, Mexico, and South Africa) have issued joint statements on sustainability.

---

## **Developing the Concept of Sustainability: What Is Being Sustained?**

Languages are dynamic entities, and as languages change over time, confusion has often arisen when words already in common usage (e.g., “fitness” as in “physically fit”) develop an additional specialized technical meaning

(e.g., “fitness” in the Darwinian sense). There is generally no confusion for the practitioner using the word in the new sense because the meaning is gathered from its context. This means that knowledge about the context is required for understanding meaning. The concept of sustainability certainly carries with it the original meaning of “the property of being sustainable” with “sustainable” and “sustain” being used in the sense of being maintained or prolonged. But, clearly, there must be *something* that is being sustained or has the property of being sustainable. Thus, the defining question becomes, what exactly is being sustained? To answer this question, it is necessary to go back to the context under which the concept was originally developed.

The groundwork for the concept of sustainability was developed over the two decades spanning the late 1960s to the late 1980s in a series of United Nations reports, resolutions, conferences, and commissions. This work culminated in the first definition and description of sustainability, articulated in the 1987 *Report of the World Commission on Environment and Development* (entitled “*Our Common Future*” and often referred to as the “*Brundtland Report*” after the chairman of the commission, Gro Harlem Brundtland). The Brundtland Commission was formed at a time when there was increasing recognition of and concern over the linkages among accelerating environmental degradation, loss of natural resources, and deterioration of people’s economic and social conditions. The members of the commission were given a very clear charge: to propose ways to deal with environmental concerns that took into account the interrelationships between people, resources, environment, and development. The task of the commission was ambitious – it was charged with nothing less than formulating a “global agenda for change” (UNWCED 1987).

The Brundtland Commission, in formulating the new paradigm for improving overall human well-being by considering the coupled social-ecological system, used the concept of sustainable development to create the integrating framework of their approach. The term “development” was used in the broad sense of meeting the basic needs of all people and extending the opportunity to satisfy aspirations for a better life to everybody, with change being required in all countries, rich and poor alike. Thus, the “what” of sustainability, the thing that is being maintained, is *improvement in the human condition*. The report emphasizes that the sustainability of development or sustainable development is never a fixed endpoint. Rather, it is *a process of change* in which natural resource use, monetary investment, the orientation of technological development, and institutional change are consistent with future and present needs.

While intellectually revolutionary and forward-looking in most respects, the Brundtland Report didn’t yet take the full step of recognizing the inherent systems problem in maintaining a dichotomy between humans and the rest of the natural world. It argued very persuasively that human well-being depends on the delivery of goods and services supplied by well-functioning ecosystems and that ecosystem function relies on interactions among all the component species. However, it still described ecosystem health as a means for supporting improvements human well-being instead of recognizing that these two are inherently exactly the same thing. Humans are just one component of the social-ecological systems that is the life

support system of the planet. And while the focus on human well-being as a metric of particular concern may be chosen, it is not a distinct element from the functioning of the whole system.

---

## **Defining the Concept of Sustainability: Focusing on Positive Change for All**

In general, current definitions of sustainability (in the specialized sense) stem directly from the original definition of sustainable development as defined in the Brundtland Report: Sustainability is “meeting the needs of the present without compromising the ability of future generations to meet their own needs.” It is not surprising, given the value-laden nature of the concept, that there are now hundreds of definitions of sustainability and related terms (see, e.g., <http://www.sustainablemeasures.com/node/36> or <http://www.emrgnc.com.au/SustainabilityDefinitions.pdf>). Unfortunately, in popular parlance, sustainability has come to mean everything from buying recycled products, to “green” business practices, to another term for environmentalism. Despite this, the best definitions still attempt to focus on the whole of the social-ecological system. For example, sustainability is:

A vision of development that encompasses populations, animal and plant species, ecosystems, natural resources – water, air, energy – and that integrates concerns such as the fight against poverty, gender equality, human rights, education for all, health, human security, intercultural dialogue, etc. (UNESCO 2005)

Because sustainability is concerned with meeting the essential needs of the large numbers of people on this planet whose needs are not being met, it is therefore about creating the conditions for all people to have the opportunity to satisfy their aspirations for a better life. Creation of these conditions involves consideration of the functioning whole social-ecological system in which the relationships among economy, environment, politics, and social factors are linked into a complex, coupled system, in which no part can be viewed in isolation from the rest.

Disruption in the flows of matter and energy in natural ecosystems inevitably leads to disruption in the flows of goods and services to humanity, thus degrading the mean human condition. However, the negative effects of our actions are not shared equally, thereby enabling opportunities for certain portions of humanity to move further above the mean while others drop further below the mean. In other words, the mean human condition has been progressively degraded and, at the same time, that the variance around that mean has increased.

---

## **The Foundational Premises of Sustainability**

I doubt if there ever has been a period in history when a greater proportion of people have found themselves frankly puzzled by the way the world reacts to their best efforts to change it, if possible for the better. . . . recently things seem to have been going wrong so often, and in

so many different contexts, that many people are beginning to feel that they must be thinking in some wrong way about how the world works. I believe this suspicion is probably correct. C.H. Waddington, *Tools for Thought*, 1977, p. xi

Why is it that problems such as global climate change, long-lived organic toxins in our food chains, pernicious extreme poverty and hunger, lack of access to primary education, gaps in gender equality, childhood mortality, and deadly diseases such as HIV/AIDS and malaria are proving so remarkably resistant to our best efforts to understand and solve them? When traditional tools and approaches are not working, progress requires a new intellectual context. The novel insight provided by the concept of sustainability is that humans and their local and global environments exist as complex social-ecological systems. To really understand this statement, it is important to contrast the concept of “complexity” with the concept of “complicated.” Complicated systems are just simple systems with many parts. In simple systems, whether the parts are many or few, interactions among parts are well defined and predictable, and thus the system is, at least in theory, well defined and predictable. This does not mean that understanding complicated systems or the problems arising from them is easy. For example, cars, photocopiers, and spacecraft are complicated systems and most of us have only a tenuous grasp of how they actually work (or how to use them)!

Complex systems consist of few or many parts, but the source and essence of complexity arises from the richness, intensity, and character of the interactions among constituent parts. Typically, these interactions lead to nonlinear and/or emergent behavior (behavior that can't be predicted by studying the parts of the system individually). Furthermore, the interactions (as well as the specific connections over which these interactions occur) constantly change, compounding the difficulty of thorough analysis by the formation/dissolution of amplifying/stabilizing feedback loops. For example, human-induced climate change is a result of perturbing a complex system, and finding solutions is difficult because predicting the result of any decision strongly hinges on a thorough understanding of the countless interactions between ecological and human social and political factors.

While definitions of sustainability can be instructive and beneficial for communication, it is impossible for any definition to convey the richness and nuance that is being implied. In order to apply and further develop the concept of sustainability, a deeper understanding than just a definition is required. This necessitates an understanding of the *mental model* on which the concept of sustainability is based. Making the model explicit allows clear analysis of the strengths and weaknesses of the concept and allows for implementation and improvement. One way to succinctly describe the conceptual model of sustainability is to state it as a series of four foundational premises. Explicit articulation of the premises can then serve as a basis for developing research agendas, funding priorities, and mutually agreed-upon courses of action.

**Premise #1: The current state of human existence is not an acceptable endpoint of societal development.** Not designed to be inflammatory or accusatory, this statement is a simple recognition that regarding the state of humanity as a whole, we can always do better. It is not an indictment of the decisions we have



made in the past or of our current lifestyles. It is instead the fundamental driving force that keeps us working to improve the lives of all people worldwide. While humans have made phenomenal advances in medicine, food production and distribution, resource extraction, etc., the benefits of these advances are not enjoyed by large portions of humanity.

**Premise #2: Humans have reached a state where we are negatively impacting the ability of future generations to meet their needs and aspirations.**

The data are unequivocal that issues such as global climate change, ozone destruction, degradation of ecosystem services, depleted and limited fossil fuel resources, accumulation of persistent toxins in the environment, new and emerging diseases, and trends in food production all point to the same conclusion: Human impacts on global ecosystems are accumulating at a rate that endangers our present and future well-being. The most extensive scientific review of the data to date, contributed to by more than 2,000 authors and reviewers worldwide, concludes:

Human activity is putting such strain on the natural functions of Earth that the ability of the planet's ecosystems to sustain future generations can no longer be taken for granted. The provision of food, fresh water, energy, and materials to a growing population has come at considerable cost to the complex systems of plants, animals, and biological processes that make the planet habitable. ... Nearly two thirds of the services provided by nature to humankind are found to be in decline worldwide. In effect, the benefits reaped from our engineering of the planet have been achieved by running down natural capital assets. In many cases, it is literally a matter of living on borrowed time. (MEA 2005a)

There is no longer any doubt that we have to include issues of intergenerational equity in access to the earth's natural resources into our planning and decision-making.

**Premise #3: The major types of problems facing humanity have to be addressed simultaneously: There is no ranking of importance among social, environmental, and economic issues.** We exist as a part of a complex coupled social-ecological system that produces a set of nonlinear, time-dependent, multi-scalar outcomes replete with time lags and feedback loops. The parts cannot be viewed or studied in isolation. For example, global climate change is driven by changes in atmospheric composition of greenhouse gases, which in turn is affected by feedbacks in the physical and biological systems of the planet. Further, human social, political, and economic systems affect and are affected by these changes (water availability, food production, energy usage, land use change, etc.). In essence, our perspectives on the present and future states of global ecosystems must shift from one focused on human influences as "natural," not "unnatural."

**Premise #4: The complex, coupled social-ecological system of humans and the earth requires fundamental restructuring.** That is, we can't "fix" it; we have to fundamentally change it. Premises #1 and #2 establish the moral and physical imperative for change, and Premise #3 lays out the character of the problems and why conventional approaches have not worked. Essentially Premise #4 states that our current systems must be radically altered or replaced, not simply tweaked. Notice that this Premise does not say *what* particular changes need to happen; it states only *that* change needs to happen if a reversal in degradation of the human

condition is desired. Premise #4 is basically a restatement of the systems idea called “the central law of improvement” (Berwick 1996): Every system is perfectly designed to achieve the results it achieves. In any system of interacting parts, both the results that we desire and those we don’t are all products of the same system. Whether we like it or not, our system produces all the results we want and don’t want as a result of its inherent structure. This is not to say that the performance of systems doesn’t change over time. However, it does mean that average performance and the degree of variation around the mean are a function of the system itself. If we want fundamental change in results, i.e., if we want to improve things, we must institute fundamental structural changes in the system.

---

## Operationalizing Sustainability

The concept of sustainability is meant to be applied to the real world; it is a framework for making decisions to solve problems. Sustainability has been operationalized in a number of ways, the most common of which is to divide the problems facing humanity into three groups: Environmental, Social, and Economic. This division leads to the commonly used Venn-type diagrams where sustainability is viewed as the overlap of these three “realms,” “lenses,” “pillars,” “dimensions,” or “legs” (Fig. 1). This formulation has been quite attractive as it mirrors much of our past thinking and the societal structures that have already emerged from that thinking.

For example, this model maps easily onto existing academic disciplines and university departmental/school structures (natural sciences, social sciences, and economics/business), with the associated funding streams and research programs. It also mirrors the way governmental agencies are set up in many countries. In the United States, for instance, there is the Department of the Interior, the Department of Commerce, and the Department of Health and Human Services. This model also maps well onto existing NGOs and special interest groups, such as environmental groups, social justice groups, and free trade advocates.

While relatively easy to apply and useful in some respects, this model of viewing and operationalizing sustainability has resulted in much of the controversy, confusion, and misapplication surrounding the concept. The reductionist approach of dividing sustainability into parts is in direct opposition to the interdisciplinary systems approach that has propelled the concept of sustainability to its current positions as a fundamental organizing principle for a modern form of ecology and paradigm for future global development. One of the benefits of focusing on the integrated premises of sustainability is that the deficiencies of the “realms” approach are immediately illuminated: (1) In the real world, there aren’t different realms of problems facing humans (e.g., climate change cannot be confronted through isolated social, economic, or environmental approaches). (2) Defining a set of realms invites focus on the artificial boundaries that differentiate the realms instead of the whole system. This is the intellectual equivalent of the well-known mistake of dividing a complex system into an arbitrary set of subsystems and

**Fig. 1** The standard way of depicting sustainability as three realms of problems

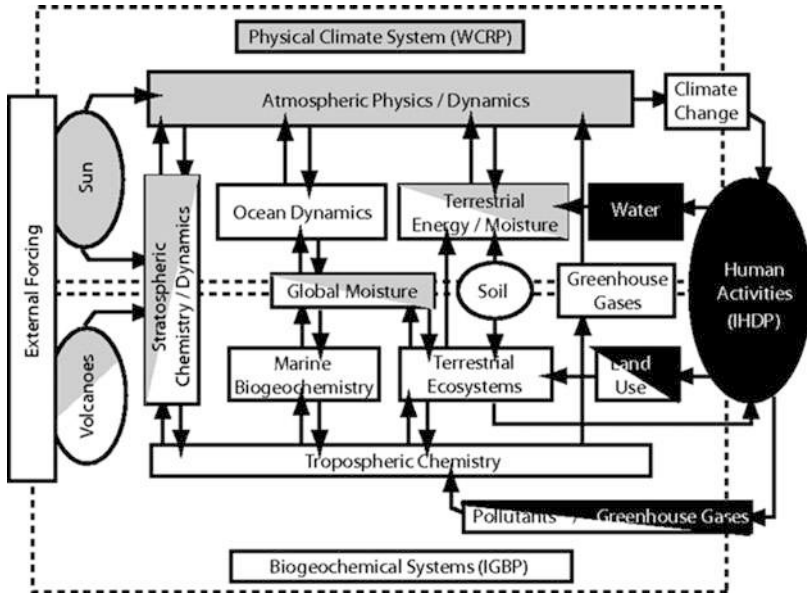


attempting to understand the whole system by just studying the subsystems. The complex coupled social-ecological system is just that – a system that must be studied as a single articulated system. (3) Existing institutional (and thought) structures must be fundamentally altered – the structure of the existing systems themselves has introduced unintended reinforcing feedbacks that allow the systems to persist and resist adaptive change in human–ecosystem interactions. The concept of sustainability is truly transdisciplinary in nature and requires a rethinking of all traditional boundaries. The Venn-type diagram encourages a mental disaggregation of sustainability into parts that can be assimilated into existing intellectual frameworks, but that further entrench existing social-ecological paradigms. The result of this can be seen in common use of terms such as “ecological sustainability,” “social sustainability,” and “economic sustainability.” Because the word “sustainability” refers to sustainable (as an ongoing) improvement in the state of the social-ecological system, these terms are either nonsensical or congruent.

Another way to operationalize sustainability is to formulate a true systems model that captures much more of the complexity of social-ecological systems. For example, a model of the earth system (see Fig. 2) can be coupled with a model of social processes (see Figs. 2 and 3).

Use of this type of model provides focus to whole-system processes and is helpful in studying fundamental system structure and process, but they are often too complicated to serve as tools for accurate prognosis or problem-solving. Also, it is very difficult to use a process model to make decisions regarding time-dependent resource allocation without extensive computer simulation. What is needed is a simple model that captures enough of the complexity of the real situation, can be used as a “dashboard” to measure progress toward goals, is evocative for thinking about connections and possibilities, and is analytical in its mathematical structure.

One promising alternative that is being used increasingly by both scientists and policy makers is the *rose diagram* (sometimes called an orientor star). A rose diagram is a pie chart variant with each sector of the circle having the same size.

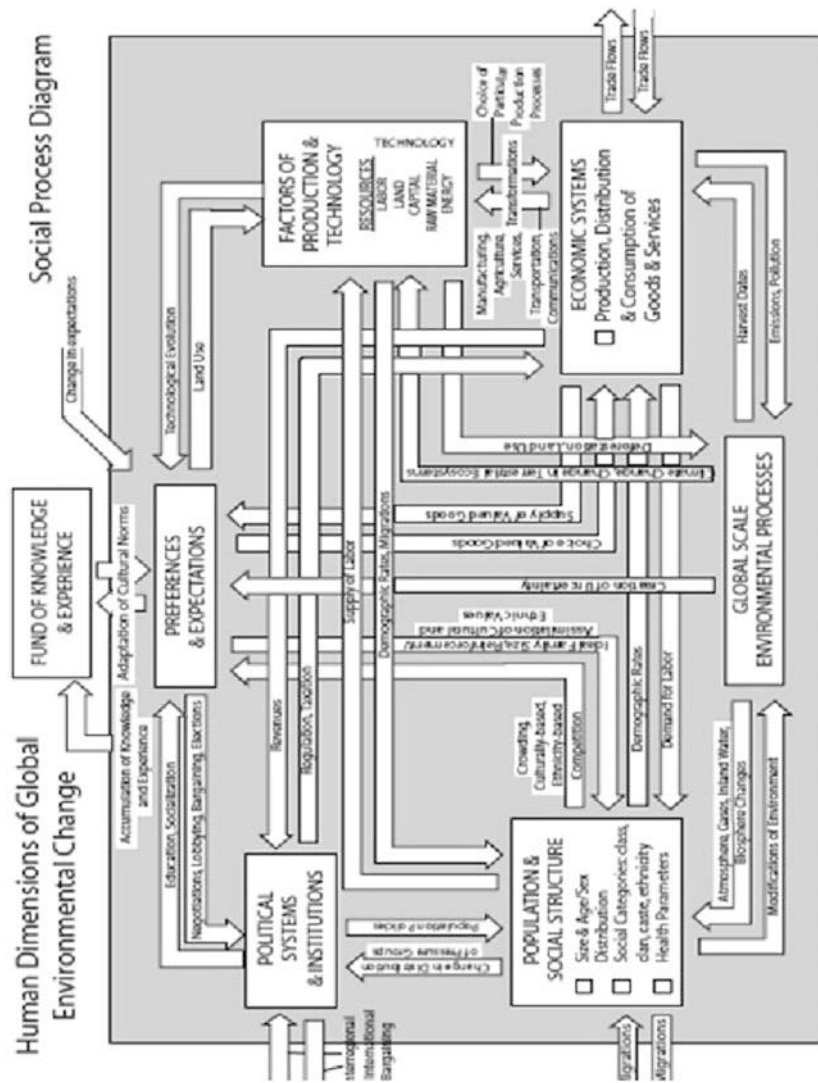


**Fig. 2** System model of ecological part of the coupled social-ecological system (From Mooney et al. (2013))

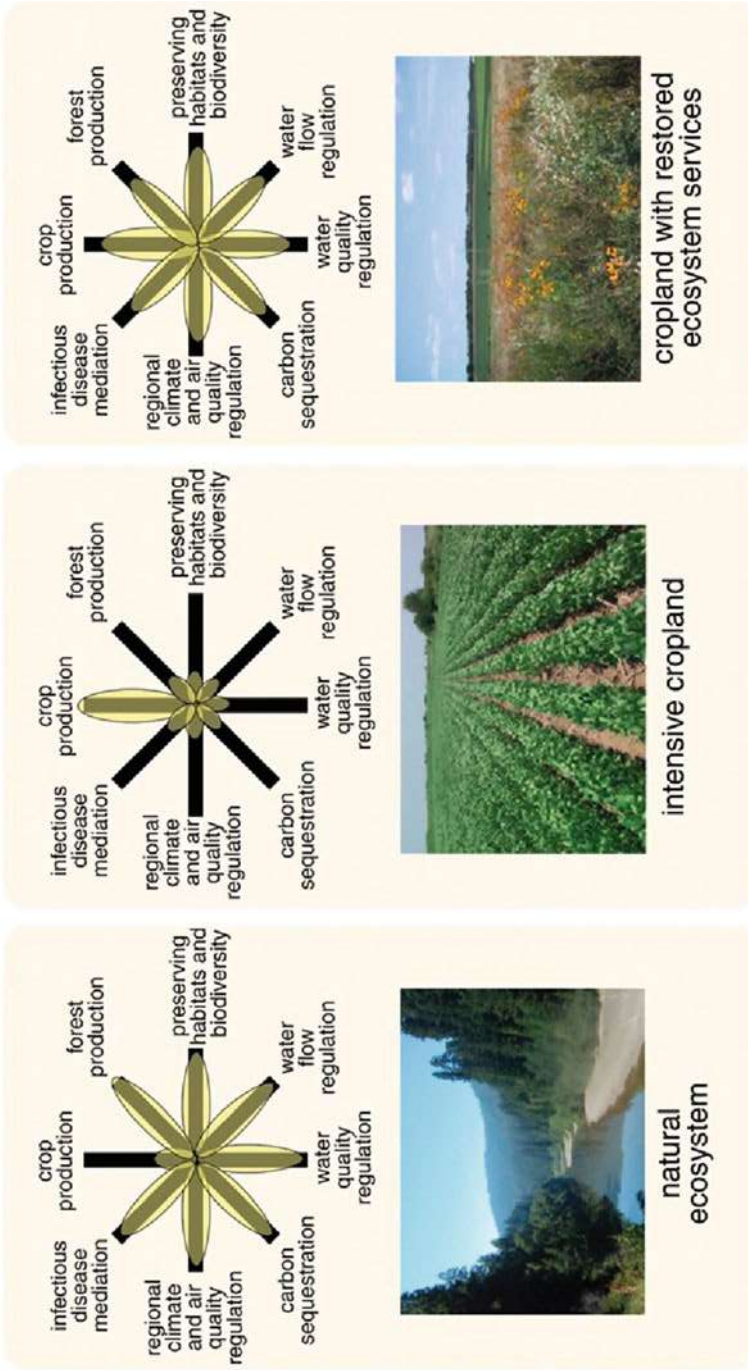
Variables are plotted as distance from the center either on the radial lines of the sectors or by filling in the sectors themselves. Concentric circles radiating from the center can be used for semiquantitative or even quantitative rendering. Rose diagrams are excellent conceptual ways to generate immediate visual representation of a large number of variables (see Fig. 4).

An excellent example of how these types of diagrams are being used to organize thinking, inform decisions, and operationalize sustainability is the United Nations’ Global Compact Cities Program (see Fig. 5). In this case it is possible to consider 28 variables simultaneously. By grouping variables in appropriate ways, this “dashboard” can show where efforts are having their greatest successes and where more effort and resources should be applied. In this example, the variables representing the social part of the social-ecological system (economics, politics, and culture) are roughly in similar states of acceptability, while those representing the ecological part (ecology) are, in some cases, reaching critical levels.

A weakness of the rose diagram conceptual approach is that it lacks a focus on social-ecological *processes*; it doesn’t show relationships among parts. It can’t be used to predict how various feedbacks among dynamic variables operate, and it doesn’t predict how leakage from one area to another might occur (i.e., how improvement in one area might cause decline in another). At the same time, it is extremely useful for getting a quick snapshot of a large number of important considerations. It allows for easy expansion in numbers of variables and in numbers of groupings of variables. Further, it avoids the Siren call of reductionism inherent

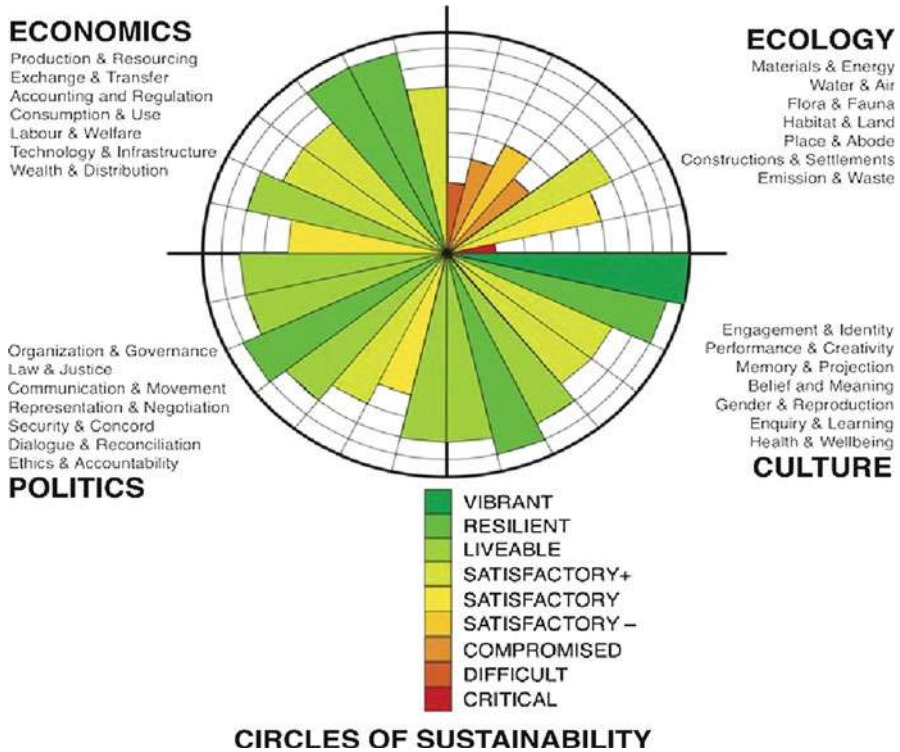


**Fig. 3** System model of the social-ecological system (From Mooney et al. (2013))



**Fig. 4** Example of rose diagrams used to compare trade-offs in land use. In this case, the state of each variable of interest is shown along each axis (Reproduced from Foley et al. (2005))





**Fig. 5** Example of a rose diagram that allows planners to view 28 variables simultaneously as a “sustainability dashboard” of the current state of a system. (From <http://citiesprogramme.com/aboutus/our-approach/circles-of-sustainability>)

in the Venn-type diagrams; for practical reasons, scientists and planners may still focus primarily on one quadrant of the whole, but it is always clear that each quadrant is an inseparable part of the whole. If the system starts to heavily favor one set of factors over another, it is immediately apparent.

### The Development of Sustainability Science

At first glance, it might appear counterintuitive that sustainability, a concept with such an ethically based, values-driven focus, can be associated with, and in fact spawn, a new science. In fact, developing the science of sustainability forces a head-on confrontation with two of the essential questions that underlie much of the scientific enterprise: (1) What is the optimal balance between pure and applied research? (2) What is the appropriate role of science in society? Because science is expensive and must be supported by society, these two questions are really the same question: How much support should society give science and scientists and what is appropriate to expect in return?

In a groundbreaking essay published in *Science*, Jane Lubchenco (1998) laid the foundation for the broad acceptance of sustainability science in her call for a new social contract between science and society. In this work, Lubchenco did two important things. First, she expanded the realm of the natural sciences by redefining what was meant by the environment. In the mainstream view, natural scientists study the natural world and other disciplines study everything else. Lubchenco presaged the systems approach now in common use in sustainability science by expanding the definition of “the environment” (the natural world) to include such things as human health, the economy, social justice, and national security; she had started defining the social-ecological system of the planet. Second, she suggested a new way to think about the age-old debate of the relative merits of pure versus applied science. Lubchenco clearly articulated in the most public and prestigious of forums the sense among a growing number of scientists that human-driven environmental change and environment-driven human suffering had become pressing enough that business as usual in the scientific community was no longer an option. She stated that it is incumbent upon science to “pay back” society for its support by prioritizing the problems facing the global society. At the same time, she acknowledged that pure research is the basis from which the tools for solving problems arise. She suggested that scientists and society make a new pact whereby a strategic framework is created to conduct research where knowledge is most needed (sometimes called use-inspired basic research; Fig. 6).

Sustainability science now defined as:

An emerging field of research dealing with the interactions between natural and social systems, and with how those interactions affect the challenge of sustainability: meeting the needs of present and future generations while substantially reducing poverty and conserving the planet’s life support systems. (<http://sustainability.pnas.org/page/about>)

has at its core the same philosophy as Lubchenco in her call for use-inspired basic research. The field explicitly emphasizes research on the fundamental character of interactions of the social-ecological system, as well as application of this knowledge to advance sustainability goals relevant to water, food, energy, health, ecosystem services, etc. It takes the traditional focus of ecology into a new realm of research – away from the study of pristine ecosystems isolated from anthropogenic influences and toward the study of humans as a dominating force causing change in ecosystem states and processes. Since its first description in NRC (1999), *sustainability science* has become an accepted discipline in its own right, with several specialized journals and an approximately exponential increase in numbers of publications worldwide (Fig. 7).

---

## Focusing Where Knowledge Is Most Needed

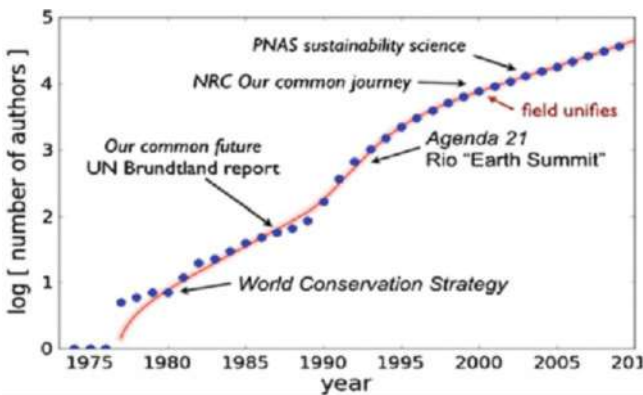
Given all the possible problems (local to global, specific to general) on which to focus, it is important for use-inspired basic research to focus on society’s most urgent challenges. Interestingly, determining which challenges to confront is an extra-scientific endeavor that must necessarily involve the human factors of hope,



**Beyond basic vs applied research:  
Science in Stoke's Quadrants**

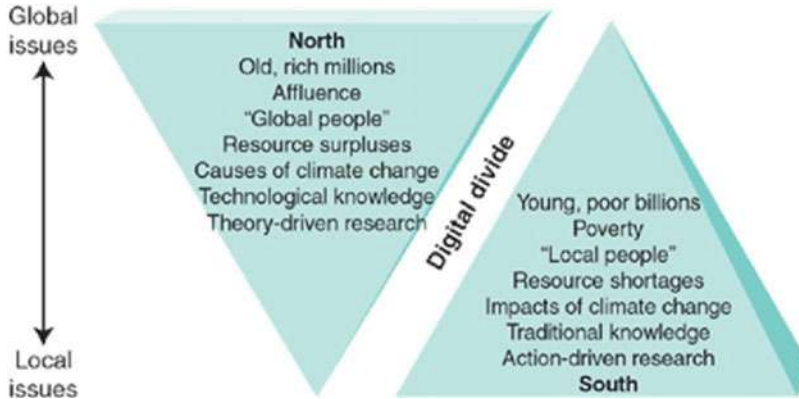
		Considerations of use?	
		No	Yes
Quest for fundamental understanding?	No		Applied research (Edison)
	Yes	Basic research (Bohr)	<i>Use-inspired basic research (Pasteur)</i>

**Fig. 6** The relationship among traditional basic and applied research and use-inspired research based upon the work of Donald Stokes (Clark 2007). Basic research is epitomized by the work of Bohr to determine atomic structure; applied research is epitomized by Edison in his work to commercialize electric lighting; use-inspired basic research is epitomized by the work of Pasteur. Work that explores particular phenomena without consideration for generality or application is represented in the *upper left* quadrant (Pasteur's Quadrant, Donald Stokes 1997)



**Fig. 7** The temporal evolution of sustainability science as depicted by the number of publications per year (From Bettencourt and Kaur (2011))

desire, politics, morality, etc. Probably, it is no coincidence that at the same time that sustainability science was germinating as a discipline, the global community was unifying to determine the world's most urgent sustainability challenges. In September 2000 the UN General Assembly adopted the United Nations Millennium Declaration laying out what has now become known as the Millennium Development Goals. These goals have since served as a framework for designing the organizing principles of sustainability science. While the terminology surrounding what is now known as the social-ecological system of sustainability science had not yet been developed, it is clear that the Millennium Development Goals were oriented toward focusing policy makers and scientists toward a systems approach to sustainability. These primary goals are generally listed as follows:



**Fig. 8** Organizing research agenda of sustainability science around spatial scale and economic status (From Kates et al. (2001))

1. Eradicating extreme poverty and hunger
2. Achieving universal primary education
3. Promoting gender equality and empowering women
4. Reducing child mortality
5. Improving maternal health
6. Combating HIV/AIDS, malaria, and other diseases
7. Ensuring environmental sustainability (in the sense of maintaining ecosystem function and continued production of natural resources)
8. Establishing a global partnership for development

The call is for improvements in human health, ecosystem health, societal health, and economic health. This has become the de facto identity of sustainability science and most modern socio-ecological studies. The recognition that all of these issues are linked, and that progress in any requires progress in all, is the type of integrative framework that is required to forge sustainability science as a discipline beyond the natural sciences. Another, complimentary, way to organize these goals that is, perhaps, more conducive for developing systemic research programs is to consider the different sorts of problems and pressures as they occur at different scales and in countries at different stages of economic development (Fig. 8).

Based on this early thinking, the sustainability science community has continued to define the new field by working to establish a coherent research agenda (Kates 2011):

- (i) What shapes the long-term trends and transitions that provide the major directions for this century?
- (ii) What determines the adaptability, vulnerability, and resilience of human–environment systems?
- (iii) How can theory and models be formulated that better account for the variation in human–ecosystem interactions?
- (iv) What are the principal trade-offs between human well-being and ecosystem states and processes?

- (v) Can scientifically meaningful “limits” be defined that would provide effective warning for instabilities or tipping points in human–ecosystem interactions?
- (vi) How can society most effectively guide or manage human–ecosystem interactions toward a sustainability transition, reversing degradation in the condition of both human societies and natural ecosystems?
- (vii) How can the “sustainability” of alternative pathways of environment and development be evaluated?

---

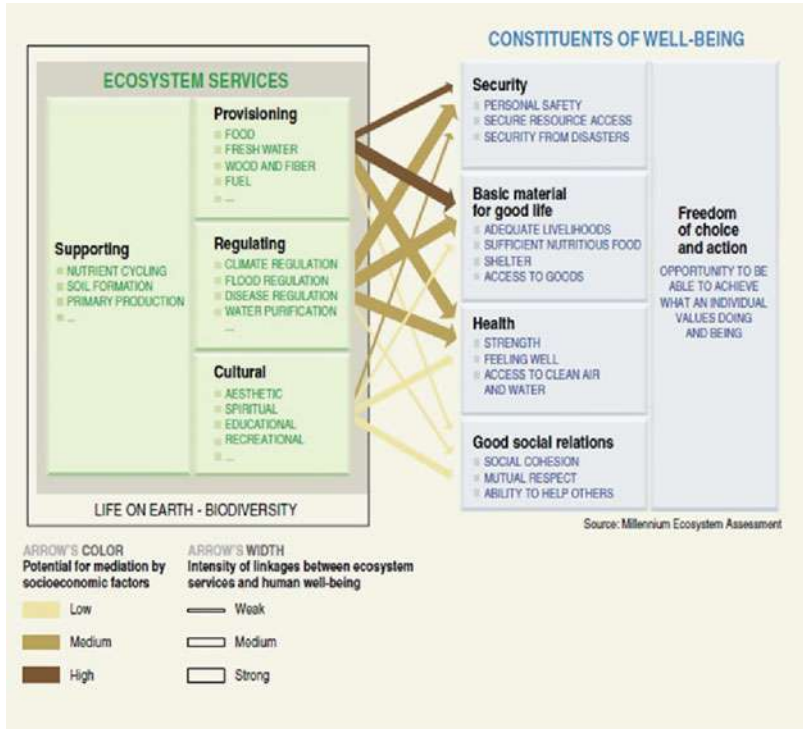
## **Sustainability Science Represents a New Conceptual Model for the World**

Much more than just a realignment of research priorities toward use-inspired basic research or an articulation of a coherent research agenda, sustainability science represents a new mental model for understanding and living in the human/nature system. The “new” insight is that the very concept of a human/nature system (a dichotomy between humans and the rest of the natural world) is, in fact, the problem. Models aren’t right or wrong, only more or less useful, and a human/nature dichotomy is no longer useful. Our current understanding of the biosphere is painting an ever more focused picture that the combined processes of all living things, including humans, create the very conditions that allow for living things; *life* creates its own life support system. The detailed composition of the atmosphere and ocean, the characteristics of soil, and the recycling of water and nutrients are all the result of life simply living. This is fundamentally a Gaian way of viewing social-ecological systems, without recognition of “intent” in the coevolution of humans and ecosystems, but with recognition of interconnected feedbacks. Thus, in order to understand and attempt to live in this system, we can’t separate out part of life: us.

One of the best examples of how the new conceptual model of sustainability science has been put into practice is in the Millennium Ecosystem Assessment (MA). The MA, launched by the United Nations in 2001, was designed to provide decision-makers with the scientific information necessary to understand the connections between ecosystem change and human well-being. This is a paradigmatic example of how sustainability science, working at scales from local to global and studying processes occurring from short to long time scales, fully integrates existing knowledge into a framework useful for supporting sustainability.

Central to sustainability science, and organized into a particularly useful structure in the MA, is the concept of *ecosystem services as measurable quantities*. Ecosystem services are the conditions and processes through which natural ecosystems and the species that make them up sustain and fulfill human life (Daily 1997). Divided into the categories of supporting services, provisioning services, regulating services, and cultural services, these ecosystem services are explicitly linked to the constituents of human well-being (Fig. 9)

The approach of the MA demonstrates how sustainability can be operationalized and advanced by sustainability science: The MA places human well-being and ecosystem services as the central focus and recognizes that humans, being but one

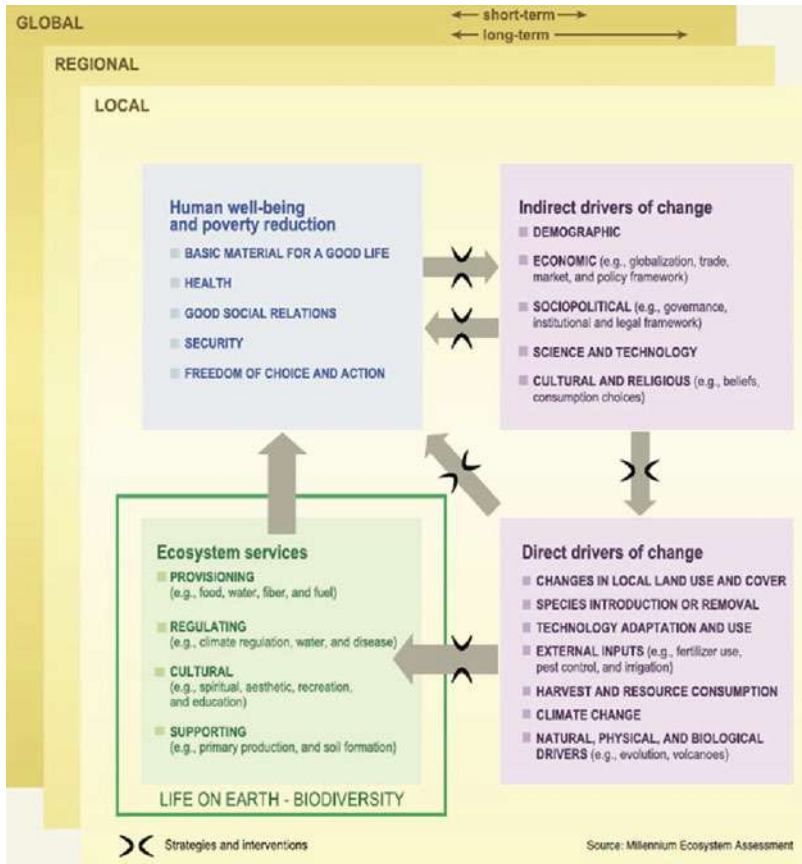


**Fig. 9** Ecosystem services and their linkages to human well-being (From MEA (2005b))

part of the social-ecological system, can't be separated from everything else. It makes explicit the idea that human well-being and ecosystem well-being are inextricably connected. Further, the MA explicitly rejects the idea that the planet is simply a "thing" to be used and manipulated by humans. It recognizes that ecosystems and biodiversity itself have *intrinsic value*. Human decisions must take into account not only human well-being, but also the intrinsic value of the rest of the social-ecological system (Fig. 10).

The assumption is that appropriate policy and management decisions can reverse ecosystem degradation, thereby enhancing ecosystem services and ultimately human well-being. The challenge for sustainability science is to provide decision-makers with sufficient understanding of the social-ecological system, coupled with the appropriate metrics, to allow for appropriate intervention. Of course, awareness and knowledge do not assure improved decision-making, but they are usually prerequisites for such.

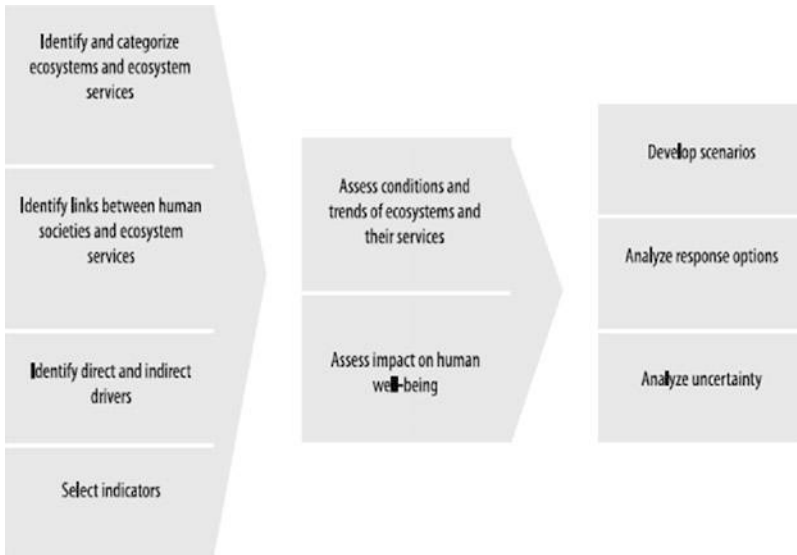
In addition to making human well-being the central focus, the MA describes and develops an explicit analytical approach. This provides a way for sustainability science to quantify ecosystem services, to make quantitative predictions regarding the effects of particular decisions on ecosystem services and human well-being, and



**Fig. 10** Conceptual model showing interaction among biodiversity, ecosystem services, human well-being, and drivers of change. Changes in indirect drivers that affect ecosystem function can lead to changes in direct drivers of ecosystem function. These changes affect ecosystem services, which, in turn, affect human well-being (From the MEA (2003))

to predict future trajectories of trends in human well-being for communities, countries, regions, and the planet (Fig. 11).

Many of the approaches used in the MA are well known in the scientific community, but the MA also highlights *scenarios* as a useful tool for converting scientific findings into knowledge that can be used by policy makers. While the use of scenarios (Fig. 11, top right) as an analytical tool predates the MA, the MA introduced it as a standard tool of sustainability science. Even with sophisticated process models of ecosystems, human social systems, or both, prediction of future conditions of ecosystem services and human well-being has too much uncertainty for policy decisions. Scenarios are tools that do not replace process models and other forms of forecasting, but serve as an important compliment to understanding the potential long-term effects of particular decisions. The ultimate goal of scenario



**Fig. 11** The analytical approach of the MA that serves as a model for sustainability science in general (From MEA (2003))

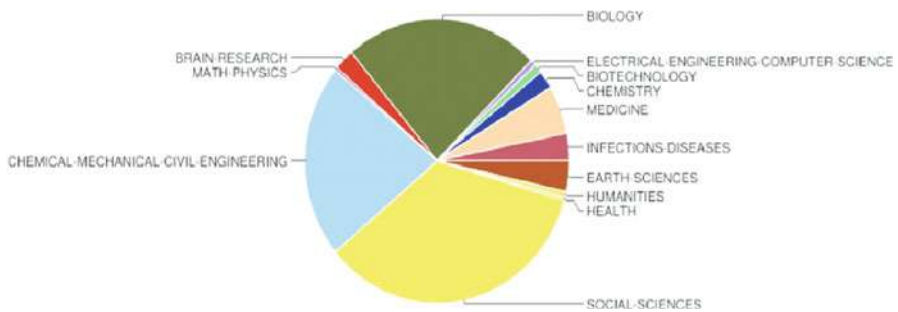
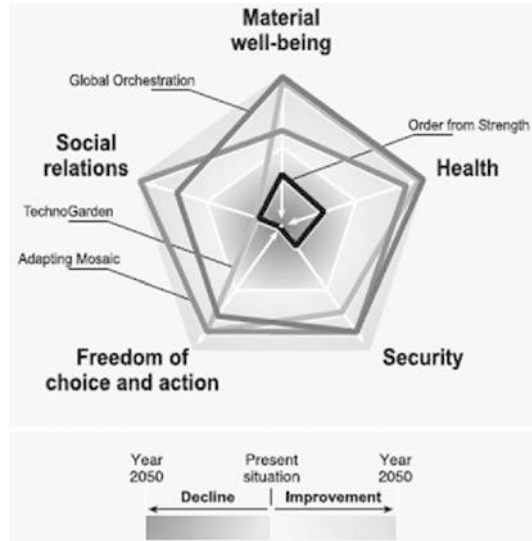
building is to ask: What are possible developments of the coupled social-ecological system? For example, the MA used four scenarios to try to understand potential trajectories of human well-being and ecosystem services. These scenarios made different assumptions regarding human decisions around issues such as freedom, social relations, material wealth, and security and resulted in quantitative predictions of trends in important ecosystem services (Fig. 12).

## Future Directions in Sustainability Science

Sustainability science has made significant advances in an extremely short period of time. One reason for this is that it is supported by contributions from a wide variety of existing disciplines (the proverbial “standing on the shoulders of giants”; Fig. 13).

While contributions from a range of existing disciplines is, and will continue to be, vitally important to the development of sustainability science, it is also problematic. Synthesis of knowledge collected from disparate sources making different assumptions and using different mental models and initial conditions is extremely difficult. The transdisciplinary nature of sustainability science necessitates breaking down barriers among other disciplines and integrating approaches and information. “Discipline-bound” approaches that focus on only one aspect of the social-ecological system eliminate much of the possibility for seeing emergent properties and lead to incomplete, or worse, incorrect results (Carpenter et al. 2009). One of the significant challenges is to develop systems for data management and

**Fig. 12** Example of changes in aspects of human well-being by the year 2050 in the different scenarios of the MA. The light pentagon in the *middle* represents the starting point (year 2000 in this case). *Lines* moving outward indicate improvement; moving inward indicates decline. *Bold words* are variable names; *small words* are names of different scenarios (From MEA (2005b))



**Fig. 13** Contribution to sustainability science from traditional scientific disciplines based on Institute for Scientific Information (ISI)-defined disciplines. Fractional contributions based on the classification of journals where publications appeared (From Bettencourt and Kaurc (2011))

interpretation and a scientific vocabulary that facilitates convergent research aims. Perhaps the very existence of sustainability science will break down some of the walls among existing disciplines for greater communication among scientists and between scientists and nonscientists.

Much current work in sustainability science is focusing on improvement to our mental models. For example, social-ecological systems evolve through time and it is challenging to incorporate this complexity into our mental models and analytical approaches. The endogenous restructuring in social-ecological systems arises in two different and connected ways: feedbacks creating complex adaptive systems (CAS) and temporal changes within hierarchical levels of the system itself (panarchy). These two concepts are important components of what has become known as *resilience theory* or *resilience thinking*.



There is growing recognition that social-ecological systems are, in fact, CAS. In CAS, system components are related such that the system as a whole has the ability to adjust or even fundamentally alter connections and interactions among components, and even the components themselves, based on experience with, pressure from, or even anticipation of external forces. A well-known example of the capacity for internal restructuring is in the redistribution of living biomass and changes in species composition across the surface of the earth in response to changes in climate driven by anthropogenic alterations in Earth's atmospheric composition. This reactive behavior, which amounts to "learning," is simply a consequence of the structure of the reinforcing (positive) and stabilizing (negative) feedback loops in the system, coupled with the capacity for internal reorganization. The ability of CAS to fundamentally alter internal structure in response to external forces is one of the primary reasons it is fundamentally impossible to control these systems for a constant performance despite humans' pressing social and economic interest in doing so.

Another topic of study in sustainability science is the development of techniques for integrating information drawn from multiple knowledge systems. A *knowledge system* is a set of propositions used to claim truth. Western science is one such knowledge system. Experience from the MA taught us that sustainability science cannot be successful if it draws only on information and models produced by the practice of traditional western science. Knowledge from other sources including local, traditional, and practitioner's knowledge must also be used because these are often the only source of information for local, site-specific resource use (Reid et al. 2006). The use of scenarios (see above) is one method that was employed by the MA in an effort to address this concern. While the MA did not achieve knowledge sharing to the extent that was hoped for, there were important lessons learned, and it laid the groundwork for further improvements (MA Multiscale Assessments MEA 2005c Vol. 4).

All systems consist of three component categories: parts, interconnections, and functions (Meadows 2008). That systems have *functions* does not imply sentience or intention. The function of a system is simply the result of the interconnections among the parts. Nonhuman systems have functions, but when referring to human systems, the word *purpose* is usually used instead. The concept of sustainability is ultimately just a reminder to have a conversation, perhaps the most important conversation we as scientists can have: What is the *purpose* of our social-ecological system? As we humans work through all the details and nuance of this issue, the concept of sustainability keeps us focused the fundamental question: How do we replace our current social-ecological system with one that has, as its purpose, human well-being (Beddoe et al. 2009)? Further, sustainability reminds us that there is no human well-being that is separate from the well-being of the social-ecological system as a whole. When we ask new questions, we often must augment our traditional approaches with new technologies, tools, or mental models to answer these questions. The new field of sustainability science, supported by its elder brothers and sisters the more established disciplines, is this tool. With its emphasis on use-inspired basic research, systems approaches, integrative mental



models of the social-ecological system that reject the human/nature dichotomy, analytical techniques, and methods of forecasting, sustainability science is providing the means to achieve our purpose.

---

## References

- Beddoe R, Costanza R, Farley J, Garza E, Kent J, Kubiszewski I, Martinez L, McCowen T, Murphy K, Myers N, Ogden Z, Stapleton K, Woodward J. Overcoming systemic roadblocks to sustainability: the evolutionary redesign of worldviews, institutions, and technologies. *Proc Natl Acad Sci U S A*. 2009;106(8):2483–9.
- Berwick DM. A primer on leading the improvement of systems. *Br Med J*. 1996;312(7031):619–22. Available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2350403/>
- Bettencourt LMA, Kaure J. Evolution and structure of sustainability science. *Proc Natl Acad Sci U S A*. 2011;108(49):19540–5.
- Carpenter SR, Mooney HA, Agard J, Capistrano D, DeFries RS, Díaz S, Dietz T, Duraipapp AK, Oteng-Yeboah A, Pereira HM, Perrings C, Reid WV, Sarukhan J, Scholes RJ, Whyte A. Science for managing ecosystem services: beyond the Millennium Ecosystem Assessment. *Proc Natl Acad Sci U S A*. 2009;106:1305–12.
- Clark WC. Sustainability science: a room of its own. *Proc Natl Acad Sci U S A*. 2007;104:1737–8.
- Daily GC, editor. *Nature's services: societal dependence on natural ecosystems*. Washington, DC: Island Press; 1997.
- Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK. Global consequences of land use. *Science*. 2005;309(5734):570–4.
- Kates RW. What kind of science is sustainability science? *Proc Natl Acad Sci U S A*. 2011;108(49):19449–50.
- Kates RW, Clark WC, Corell R, Hall JM, Jaeger CC, Lowe I, McCarthy JJ, Schellnhuber HJ, Bolin B, Dickson NM, Faucheux S, Gallopin GC, Grübler A, Huntley B, Jäger J, Jodha NS, Kasperson RE, Mabogunje A, Matson P, Mooney H, Moore III B, O'Riordan T, Svedin U. Sustainability science. *Science*. 2001;292(5517):641–42.
- Lubchenco J. Entering the century of the environment: a new social contract for science. *Science*. 1998;279:491–7.
- MEA: Millennium Ecosystem Assessment. *Ecosystems and human well-being: a framework for assessment*. Available at <http://www.unep.org/maweb/en/Framework.aspx> (2003).
- MEA: Millennium Ecosystem Assessment. *Living beyond our means, natural assets and human well-being, statement from the board*. Available at <http://www.maweb.org/documents/document.429.aspx.pdf> (2005a).
- MEA: Millennium Ecosystem Assessment. *Ecosystems and human well-being: scenarios, vol 2*. Available at <http://www.unep.org/maweb/en/Scenarios.aspx> (2005b).
- MEA: Millennium Ecosystem Assessment. *Ecosystems and human well-being: multiscale assessments: findings of the Sub-global Assessments Working Group*. Island Press. Available at <http://www.unep.org/maweb/en/Multiscale.aspx> (2005c).
- Meadows DH. *Thinking in systems*. White River Junction: Chelsea Green Publishing; 2008.
- Mooney HA, Duraipapp A, Larigauderie A. Evolution of natural and social science interactions in global change research programs. *Proc Natl Acad Sci U S A*. 2013;110:3665–72.
- NRC: National Research Council Board on Sustainable Development. *Our common journey: a transition toward sustainability*. Washington, DC: National Academy Press; 1999.
- Reid W, Berkes F, Wilbanks T, Capistrano D. *Bridging scales and knowledge systems: concepts and applications in ecosystem assessment*. Washington, DC: World Resources Institute, Millennium Ecosystem Assessment, Island Press; 2006.

- Stokes D. Pasteur's quadrant: basic science and technological innovation. Washington, DC: Brookings Institution; 1997.
- UNESCO. UNESCO and sustainable development. <http://unesdoc.unesco.org/images/0013/001393/139369e.pdf> (2005).
- United Nations General Assembly Resolution 42/187. <http://www.un.org/documents/ga/res/42/ares42-187.htm> (1987). Accessed 11 Dec 1987.
- UNWCED: United Nations World Commission on Environment and Development. Our common future (Brundtland report). Oxford: Oxford University Press. <http://www.un-documents.net/our-common-future.pdf> (1987).
- Waddington CH. Tools for thought: how to understand and apply the latest scientific techniques of problem solving. New York: Basic Books; 1977.

## Further Reading

- Blewitt J, editor. Understanding sustainable development. London: Routledge; 2008.
- Gunderson LH, Holling CS, editors. Panarchy: understanding transformations in human and natural systems. Washington, DC: Island Press; 2001.
- Handl G. Declaration of the United Nations conference on the human environment. Audiovisual Library of International Law. <http://untreaty.un.org/cod/avl/ha/dunche/dunche.html>
- IPCC. Climate change 2007: synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report of the Intergovernmental Panel on Climate Change. In: Core Writing Team, Pachauri RK, Reisinger A, editors. Geneva: IPCC; 2007. 104 pp. Available at [http://www.ipcc.ch/publications\\_and\\_data/ar4/syr/en/contents.html](http://www.ipcc.ch/publications_and_data/ar4/syr/en/contents.html) (2007).
- Lynam T, Brown K. Mental models in human–environment interactions: theory, policy implications, and methodological explorations. *Ecol Soc.* 2012;17(3):24.
- National Institute of Allergy and Infectious Diseases. Emerging and re-emerging infectious diseases. Available at <http://www.niaid.nih.gov/topics/emerging/Pages/Default.aspx>
- Norberg J, Cumming GS, editors. Complexity theory for a sustainable future. New York: Columbia University Press; 2008.
- Staudinger MD, Grimm NB, Staudt A, Carter SL, Chapin FS III, Kareiva P, et al. Impacts of climate change on biodiversity, ecosystems, and ecosystem services: technical input to the 2013 National Climate Assessment. Cooperative Report to the 2013 National Climate Assessment. 2012. 296 p. Available at <http://assessment.globalchange.gov>
- United Nations. We can end poverty 2015, millennium development goals. Background. Available at <http://www.un.org/millenniumgoals/bkgd.shtml>
- Volk T. Gaia's body: toward a physiology of earth. Cambridge: MIT Press; 2003.
- Walker B, Salt D. Resilience thinking: sustaining ecosystems and people in a changing world. Washington, DC: Island Press; 2006.
- Waltner-Toews D, Kay JJ, Lister NE. The ecosystem approach: complexity, uncertainty, and managing for sustainability. New York: Columbia University Press; 2008.

---

# Index

## A

Abiotic filtering, 73–74  
Aboveground net primary production (ANPP), 212  
Abscisic acid (ABA), 12  
Absorbed photosynthetically-active radiation (APAR), 209–210, 215, 225  
Actinorhizal N-fixing, 182  
Actinorhizal symbioses, 183  
Actual evapotranspiration (AET), 218  
Aerenchyma, 429–430  
Alkaloids, 155  
Allometry, 495  
Alpine plants  
    carbon and nitrogen storage, 345–348  
    convection, 343  
    CO<sub>2</sub> availability, 349–352  
    description, 330  
    latent heat exchange and water, 342  
    microclimate and energy balance, 335–338  
    physiological and ecological constraints, 353  
    radiation stress, 351–353  
    radiation, 339–342  
    soils, 334–335  
    stress response and growth strategies, 345  
    temperature stress, 348–349  
    temperature, 338–339  
Analytical flow cytometry (AFC), 488  
Anemophily, 92  
Anoxia, 429  
Anthropocene, 534  
Ants transport, 104  
AOT40 595  
*Araucaria cunninghamii*, 34  
Arbuscular mycorrhizae (AM), 185  
Arbutoid mycorrhizae, 187  
Arms race, 146

Atmospheric lifetimes, 585  
Autogamous, 91

## B

Belowground net primary production (BNPP), 213  
Beringia, 369  
Biochemical arms race hypothesis, 147  
Biodiversity, 257, 624  
    effects, 232  
Bioenergy, 603  
    crops, 612  
    feedstocks, 611, 617  
Biofilm formation, 191  
Biofuels, 603–604  
Biogenic Primary Organic Aerosols (bPOA), 592  
Biogenic Secondary Organic Aerosols (bSOA), 592  
Biogenic VOC emission, 575, 578, 583, 591–592  
Biogenic VOC flux, 586  
Biogeochemical cycling, 15, 294  
Biological soil crusts, 318  
Biomass, 209, 211  
    energy, 605  
Biomes, 250  
Biotic interactions, 75–76  
Boundaries of population, 61  
Brundtland Commission, 634

## C

Canopies, 286–287  
Canopy sublayer, 585  
Carbon cycling, 281–282  
Carbon sequestration, 448

- Cavitation, 313  
 Censusing populations, 61–63  
 Chiropterophily, 92  
 Climate change, 408  
 Climate, 593–594  
 Coastal squeeze, 451  
 Coherent wind structures, 586  
 Community assembly, 68–71, 84  
 Complex adaptive systems (CAS), 651  
 Convection, 343  
 Corpse flower, 90  
 CO<sub>2</sub> fertilization effect, 8  
 C<sub>4</sub> photosynthesis, 402  
 Cyanogenic glycosides, 154
- D**
- Damköhler number, 584  
 Demographic parameters, 33  
 Demographic stochasticity, 47  
 Denitrification, 447  
 Density-independent limiting forces, 41  
 Desert, 367  
 Detritus, 229  
 Developmental noise, 122  
 Diaheliotropism, 307  
 Dioecious plants, 96  
 Direct inducible defenses, 169  
 Dispersal, 71–73  
 Drought, 302, 407  
 Dwarf mangrove habitat, 443
- E**
- Ecological succession, 278–280  
 Ecosystem respiration (ER), 211  
 Ecosystem services, 543  
 Ectendomycorrhizae, 187  
 Ectomycorrhizae (EM), 185–187  
 Edaphic factors, 221  
 Emission factor, 578, 581  
 Enemy release hypothesis, 46  
 Energy loss mechanisms, 286  
 Environmental stochasticity, 46  
 Environmental variability, 55  
 Enzymatic hydrolysis, 608  
 Epiphyte grazers, 469–470  
 Ericoid mycorrhizae, 187  
 Erosion control, 448–449  
 Escape efficiency, 587  
 Estuary, 427  
 Eutrophication, 452  
 Evolutionary arms race, 147
- F**
- Feeding deterrents, 162–163  
 Fertilizer, 621  
 Fire, 289–290, 408–412  
 Flora, Arctic, 369  
 Flux measurements, 19  
 Fog deserts, 301  
 Food crops, 612  
 Forest gap models, 279  
 Forest, 275  
 Functional equilibrium hypothesis, 237  
 Functional traits, 70
- G**
- Gasification, 609  
 Genetic diversity, 477–479  
 Geometric model, 40  
 Geomorphic processes, 382–385  
*Geukensia demissa*, 435  
 Glucosinolates, 154–155  
 Grazing, 412–416  
 Greenhouse effect, 535–536  
 Greenhouse gas emissions, 537–539,  
 617–618  
 Gross primary production (GPP), 209, 210,  
 505–508
- H**
- Habitat filtering, 73, 79  
 Haleakala silversword, 30  
 Halophytes, 316, 427  
 Harmful algal blooms (HABs), 524  
 Heat shock proteins, 555  
 Heliotropism, 306  
 Hermaphroditic flowers, 96  
 Heterocystous cyanobacteria, 184  
 Human activities, 112–114  
 Humanity, 636  
 Human/non-human dichotomy, 632  
 Hydraulic lift, 312  
 Hydric soils, 426  
 Hydrodynamics process, 467–468  
 Hydrolysis, 608  
 Hypoxia, 427
- I**
- Indirect inducible defenses, 170–171  
 Indirect land-use change (iLUC), 616  
 Initiation reactions, 589  
 Insect enzymatic detoxification system, 163

Interspecific competition, 322  
Islands of fertility, 317

**K**

Krummholz, 333, 357

**L**

Land-use change (LUC), 616  
Latent heat exchange, 337  
Leaf area index (LAI), 6, 18, 210, 215, 225, 229  
Leaf endophytes, 188–189  
Leaf energy balance, 305–310  
Life table, 50–51  
Life-cycle diagram, 52  
Light-use efficiency (LUE), 215  
*Linanthus parryae*, 42  
Liquefaction, 610  
Liquid biofuels, 615  
Logging, 293–294  
Long Term Ecological Research (LTER), 216  
Long wave radiation, 340–341

**M**

Macroevolutionary hypotheses, 146–147  
Mainland-island model, 61  
Mangrove swamps, 437  
Marine protected areas (MPA), 480  
Matrix model, 54–55  
Mean annual temperature (MAT), 240  
Melittophily, 92  
Meristem limitation, 228  
Meta-analysis, 240  
Metapopulation, 60  
Michaelis-Menten model, 350  
Microbial diversity  
  phyllosphere, 196  
  rhizosphere, 198  
  small-subunit ribosomal RNA, 196  
Microclimate, 12  
Microflagellates, 491  
Microphytoplankton, 491–492  
Mid-continent deserts, 300  
Millennium development goals, 645  
Millennium Ecosystem Assessment (MA), 647  
Mitigation, 476  
Model-data fusion (MDF), 23  
Monoecious species, 96  
Monotropoid mycorrhizae, 187–188  
Mycorrhizal diversity, 199

**N**

N-fixing mutualisms, 180  
Nanoflagellates, 491  
Nanophytoplankton, 491  
Natural communities, 234  
Net community production (NCP), 508–509  
Net ecosystem production (NEP), 211, 216  
Net primary production (NPP)  
  abiotic controls, 218–226  
  ANPP, 212  
  biodiversity effects, 232–235  
  biotic controls, 230–231  
  BNPP, 213–214  
  community change, 235  
  disturbance, 229–230  
  herbivory, 236–237  
  legacy effects, 227  
  remote sensing and modeling approaches, 214–216  
  sequential limitation, 219  
  vegetation structure, 231–232  
Niche conservatism, 82  
Nitrogen fixation, 447  
Non-leaf photosynthetic structures, 310–311  
Non-native bioenergy crop, 625  
Non-photochemical quenching, 352  
Non-protein amino acids, 153–154  
Nurse plant/nurse-protégé association, 322  
Nutrient limitation, 518–519  
Nutrients, 380–381

**O**

Ocean acidification (OA), 503, 523, 526  
Ocean warming, 525  
Oil palm, 267–268  
Ontogenetic drift, 127–128  
Optimal partitioning models, 124  
Optimal partitioning theory (OPT), 130  
Orchids, 187  
Ornithophily, 92  
Ozone (O<sub>3</sub>), 9, 563

**P**

Park Grass Experiment, 391  
Pelagic environment, 496–498  
Permafrost, 377–378  
Pests and pathogens, 626  
Phenolics, 156–157  
Phenotypic plasticity  
  adaptive response, 124  
  definition, 121

- Phenotypic plasticity (*cont.*)  
 developmental noise, 122  
 importance in plants, 123  
 methodological approaches, 136–138  
 nonadaptive/maladaptive, 126  
 role in evolution, 126–127  
 techniques for evaluation, 127–136  
 vs. developmentally programmed changes,  
 127
- Photochemical smog, 594
- Photoinhibition, 352
- Photosynthesis, 304
- Photosynthetically active radiation (PAR),  
 5, 206
- Phreatophytes, 311
- Phylogenetics, 78
- Phytoliths, 400
- Phytoplankton ecology, 513–514
- Phytoremediation, 447
- Picophytoplankton, 489–490
- Plant functional traits (PFT), 287–288
- Plant functional types (PFTs), 493
- Plant growth promoting rhizobacteria (PGPR),  
 190–191
- Plant-microbe interactions (PMI), 178
- Plants  
 adaptation to environmental change,  
 566–567  
 and ecosystem services, 543  
 in global carbon cycle, 541–542  
 response to CO<sub>2</sub> 550–551  
 response to drought, 558–563  
 response to ozone, 563–566  
 response to temperature, 551–558
- Poa annua*, 42
- Pollination syndromes, 93
- Pollination  
 animals benefit, 91–92  
 delivery system, 92–93  
 dispersal agents, 92  
 evolutionary dynamics, 100–101  
 genetic and evolutionary consequences,  
 98–100  
 plant mating systems, 96  
 plants benefit, 91
- Population dynamics, 32, 38, 59, 401
- Posidonia australis*, 462
- Precipitation, 7, 302, 381, 539
- Proteinase inhibitors, 155
- Pseudovivipary, 437
- Pycnocline, 501
- Pyrolysis, 609
- R**
- Radiative heat exchange, 337
- Rain-shadow deserts, 300
- Ramsar Convention, 452
- Relative humidity (rH), 7
- Relaxed eddy accumulation (REA), 586
- Remote sensing (RS), 19, 509
- Residence time, 211
- Resilience theory, 651
- Respiratory needs (Rr), 213
- Restoration  
 grassland, 418–420  
 and recovery, 476–477  
 wetland, 453–454
- Rhizobia-legume mutualism, 180–182
- Rhizosphere C flux (RCF), 237
- Rhizosphere effect, 191–192
- Root endophytes, 188
- Root systems, grassland, 404
- Rose diagram, 639
- Rubisco, 553
- Runoff, 619–620
- S**
- Salt marshes, 431
- Seagrass ecosystems  
 abiotic factors, 465  
 average vs. marine and terrestrial  
 ecosystems, 465  
 economic goods and services, 466–467  
 epiphyte grazers, 469–470  
 food webs, 470–472  
 future threats, 472–476  
 generas of, 464  
 genetic diversity, 477–479  
 grazers, 469  
 hydrodynamics and resilience, 467–468  
 nurseries for juvenile fish, 464  
 restoration and recovery, 476  
 species in, 459
- Secondary metabolites, 144
- Seed dispersal  
 agents, 103–105  
 animals benefit, 102  
 evolutionary dynamics, 109  
 fruit characteristics, 105–106  
 packaging, 103  
 patterns, 106–109  
 plants benefit, 102
- Sensible heat exchange, 337
- Shortwave, solar radiation, 339

- Sky islands, 320, 329  
Snowbed species, 359  
Snow glades, 344  
Socio-ecological system, 632  
Soil organic carbon (SOC), 621–622  
Solar radiation (SR), 5–6  
Source-sink model, 61  
*Spartina alterniflora*, 430, 432, 434, 435  
Species pool, 68, 83  
Species richness, 232  
Standing crop, 209  
Stomata, 559  
Structured models, 40  
Suberin, 437  
Subsidence, 450  
Succulent plants, 314–315  
Superorganismic, 279  
Sustainability, 632  
Synthetic communities, 234
- T**  
Temperate forests, 275–276  
Termination reactions, 589  
Terpenoids, 155–156  
Thermal acclimation, 310  
Thermokarst, 384  
Tillers, 398  
Total root allocation (TRA), 213–214, 237–238  
Total soil respiration (TSR), 214, 238  
Traditional growth analysis techniques, 134  
Trapline, 98  
*Trichodesmium*, 494  
Tropical rain forests  
  ant-plant symbioses, 256  
  biogeography, 250–252  
  cambial dormancy, 254  
  climate, 250–252  
  factors, 259  
  future aspects, 268  
  lianas, 254  
  mycorrhizal associations, 257  
  physiognomic properties, 252  
  plant-pest interactions, 262  
  productivity and nutrient cycling, 263–265  
  threats, 265–267  
  tree fall gaps, 262  
Tropospheric ozone, 588  
Tundra, 366  
Turnover coefficient (TC), 214, 237–238  
Turnover rate, 211
- U**  
Unstructured models, 39
- V**  
Vapor pressure deficit (VPD), 7  
Vegetation structure constraint, 232  
Vertebrate disperser, 104  
Vivipary, 437  
Volatile organic compounds (VOCs), 17
- W**  
Water cycling, 285–287  
Water requirements and quality, 622–623  
Water-mediated dispersal, 104  
Wetland hydrology, 426  
Wildlife, 623–624  
Wind dispersal, 103  
Wind speed, 7
- Z**  
Zeldovich reaction, 8