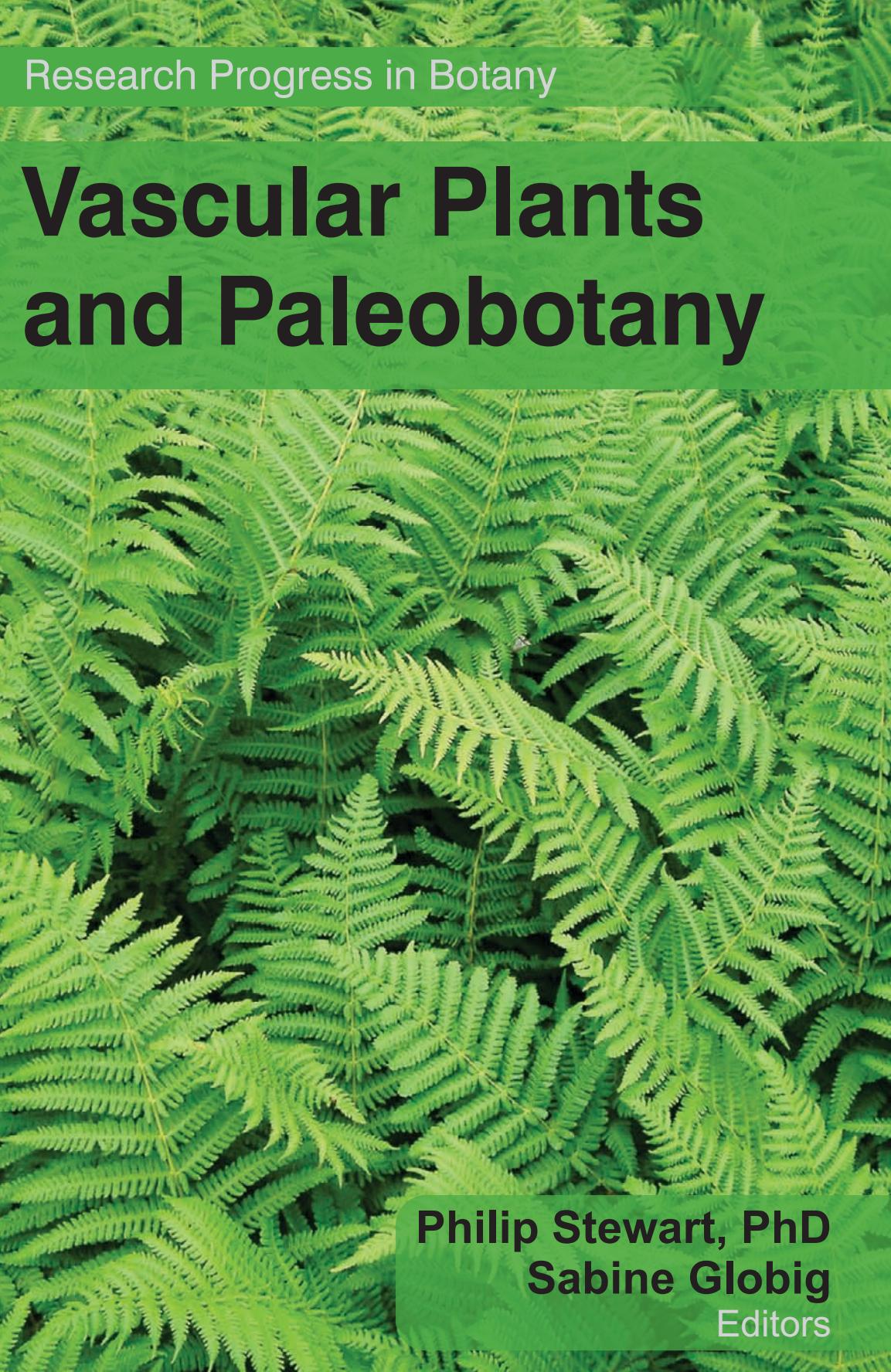


Research Progress in Botany

Vascular Plants and Paleobotany

The background of the entire image is a close-up photograph of many green fern fronds, creating a textured, overlapping pattern.

Philip Stewart, PhD
Sabine Globig
Editors

VASCULAR PLANTS AND PALEOBOTANY

This page intentionally left blank

Research Progress in Botany

VASCULAR PLANTS AND PALEOBOTANY

Edited By

Philip Stewart, PhD

Head, Multinational Plant Breeding Program;
Author; Member, US Rosaceae Genomics,
Genetics and Breeding Executive Committee;
North Central Regional Association of State Agricultural
Experiment Station Directors, U.S.A.

Sabine Globig

Associate Professor of Biology, Hazard Community
and Technical College, Kentucky, U.S.A.



Apple Academic Press

TORONTO NEW YORK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Apple Academic Press, Inc
3333 Mistwell Crescent
Oakville, ON L6L 0A2
Canada

© 2012 by Apple Academic Press, Inc.
Exclusive worldwide distribution by CRC Press an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20120530

International Standard Book Number-13: 978-1-4665-5863-2 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

For information about Apple Academic Press product
<http://www.appleacademicpress.com>

Preface

The chapters within this book, focusing as they do on the paleobotany of vascular plants—in other words, the recovery and identification of plant remains from geological contexts and their use for the biological reconstruction of past environments—provide the reader with a fascinating perspective on the evolutionary history of plants, which has bearing upon the evolution of life in general.

Some plants have remained remarkably unchanged throughout Earth's history; early ferns, for example, had developed by the Mississippian Period and conifers by the Pennsylvanian. Some plants of prehistory are the same ones around today and are thus living fossils, such as *Ginkgo biloba* and *Sciadopitys verticillata*. Other plants have changed radically, or have gone extinct entirely.

The research described within this volume is important to the reconstruction of ancient ecological systems and climate, and it provides a fundamental basis to the study of green plant development and evolution. What is more, it offers the reader a unique window on the history of early terrestrial life.

—Philip Stewart, PhD

This page intentionally left blank

List of Contributors

Patricia M. Anderson

Department of Earth and Space Sciences and Quaternary Research Center, University of Washington, Seattle, Washington, USA.

Jo Ann Banks

Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA.

Phillip N. Benfey

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

Jörg Bohlmann

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Eric D. Brenner

The New York Botanical Garden, 200th St. and Kazimiroff, Bronx, NY 10458-5126, USA.

Thomas A. Brown

Lawrence Livermore National Laboratory, Center for Accelerator Mass Spectrometry, Livermore, California, USA.

Linda B. Brubaker

College of Forest Resources, University of Washington, Seattle, Washington, USA.

Gloria M. Coruzzi

New York University, Department of Biology. 1009 Main Building, New York, NY 10003, USA.

Guangwei Dai

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

Andrew W. Douglas

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

Songlin Fei

Assistant Professor, Department of Forestry, University of Kentucky, Kentucky.

F. Fernandoy

Earth Sciences Department, University of Concepcion., Casilla 160-C, Concepción, Chile.

Lei Gao

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, PR China.
Graduate School, Chinese Academy of Sciences, Beijing 100039, PR China.

Peter J. Gould

Research Forester, Pacific Northwest Research Station, USDA Forest Service.

J. Guiot

CEREGE, CNRS/University and Paul C'ezanne UMR 6635, BP 80, 13545 Aix-en-Provence cedex, France.

Dawn Hall

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Björn Hamberger

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Britta Hamberger

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Mitsuyasu Hasebe

Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA.

Masami Hasegawa

School of Life Sciences, Fudan University, Shanghai, China.

Maria Herrero

Department of Pomología, Experimental Station “Aula Dei”, CSIC, Apdo. 202/50080 Zaragoza, Spain.

Philip E. Higuera

College of Forest Resources, University of Washington, Seattle, Washington, USA.

Jose I. Hormaza

Experimental Station “The Mayor”, CSIC, 29760 Algarrobo-Costa, Málaga, Spain.

Feng Sheng Hu

Department of Plant Biology, University of Illinois, Urbana, Illinois, USA.

Department of Geology, University of Illinois, Urbana, Illinois, USA.

Wen Ying Jiang

Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, 100029 Beijing, China.

Melanie J. Kaeser

Lead Research Technician, Joseph W. Jones Ecological Research Center, Ichauway, 3988 Jones Center Dr. Newton, GA.

Manpreet S. Katari

New York University, Department of Biology 1009 Main Building, New York, NY 10003, USA.

Makoto Kato

Graduate School of Human and Environmental Studies, Kyoto University, Sakyo, Kyoto, 606-8501, Japan.

Masahiro Kato

Department of Botany, National Museum of Nature and Science, Tsukuba 305-0005, Japan.

Christopher I. Keeling

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Alison T. Kennedy

Department of Earth Sciences, Montana State University, Bozeman, Montana, USA.

M. Leppe

Scientific Department, Chilean Antarctic Institute, INACH, Plaza Muñoz Gamero 1055, Punta Arenas, 000 billion, Chile.

Damon P. Little

Cullman Program for Molecular Systematics Studies, The New York Botanical Garden, Bronx, New York, USA.

C. Loehle

National Council for Air and Stream Improvement, Inc., 552 S Washington St., #224, Naperville, IL 60540, USA.

Jorge Lora

Experimental Station "The Mayor", CSIC, 29760 Algarrobo-Costa, Málaga, Spain.

Yun Li Luo

Institute of Botany, Chinese Academy of Sciences, 100093, Beijing, China.

Rob A. Martienssen

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

Klaus F. X. Mayer

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

Richard W. McCombie

Genome Research Center, Cold Spring Harbor Laboratory, 500 Sunnyside Blvd, Woodbury, NY 11797, USA.

W. Michea

Department of Geology, University of Chile, Plaza Ercilla 803, Casilla 13518, Correo 21, Santiago, Chile.

Walter N. Moss

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

C. Muñoz

Department of Earth Sciences, University of Concepcion., Casilla 160-C, Concepción, Chile.

Harufumi Nishida

Faculty of Science and Engineering, Chuo University, Tokyo 112-8551, Japan.

Claire Oddy

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Yudai Okuyama

Graduate School of Human and Environmental Studies, Kyoto University, Sakyo, Kyoto, 606-8501, Japan.

Wayne Owen

Southern Region, USDA Forest Service.

Peter M. Palenchar

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

S. Palma-Heldt

Department of Earth Sciences, University of Concepcion., Casilla 160-C, Concepción, Chile.

Richard D. Periman

Rocky Mountain Research Station, USDA Forest Service, Suite 115, 333, Broadway SE, Albuquerque, New Mexico, USA.

Maria C. Risueño

Center for Biological Research, CSIC, Ramiro de Maeztu 9, 28040, Madrid, Spain.

Carol Ritland

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Kermit Ritland

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Stephen A. Rudd

Centre for Biotechnology, Tykistökatu 6, FIN-20521 Turku, Finland.

Suzan J. Runko

The New York Botanical Garden, 200th Street and Kazimiroff, Bronx, NY 10458-5126, USA.

George Rutherford

Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA.

P. E. Ryberg

Department of Ecology and Evolutionary Biology, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045-7534, USA.

Chodon Sass

Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, California, USA.

Richard M. K. Saunders

Division of Ecology and Biodiversity, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, PR China.

Chelsea D. Specht

Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, California, USA.

Kim C. Steiner

Professor, School of Forest Resources, The Pennsylvania State University, Pennsylvania, USA.

Giulia M. Stellaris

Department of Plant Biology, Cornell University, Ithaca NY 14850, USA.

Dennis W. Stevenson

Cullman Program for Molecular Systematics Studies, The New York Botanical Garden, Bronx, New York, USA.

Ying-Juan Su

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, PR China.

Yvonne C. F. Su

Division of Ecology and Biodiversity, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, PR China.

Milos Tanurdzic

Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907, USA.

E. L. Taylor

Department of Ecology and Evolutionary Biology, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045-7534, USA.

Natural History Museum and Biodiversity Research Center, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045-7534, USA.

T. N. Taylor

Department of Ecology and Evolutionary Biology, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045-7534, USA.

Natural History Museum and Biodiversity Research Center, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045-7534, USA.

Pilar S. Testillano

Center for Biological Research, CSIC, Ramiro de Maeztu 9, 28040, Madrid, Spain.

Richard W. Twigg

Department of Biology, Duke University, Box 91000, Durham, North Carolina, 27708, USA.

Kazuhiko Uemura

Department of Geology and Paleontology, National Museum of Nature and Science, 3-23-1 Hyakunincho, Tokyo 169-0073, Japan.

Masayoshi Umebayashi

Division of Life Sciences, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 920-1192, Japan.

Ting Wang

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, PR China.

Hai Bin Wu

SKLL, Institute of Earth Environment, Chinese Academy of Sciences, Xian 710075, China.

Institute of Environmental Sciences, UQAM, Montreal PQ, Canada H3C 3P8.

Toshihiro Yamada

Division of Life Sciences, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 920-1192, Japan.

Yong-Xia Yang

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, PR China.

Graduate School, Chinese Academy of Sciences, Beijing 100039, PR China.

Xuan Yi

Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, PR China.

Graduate School, Chinese Academy of Sciences, Beijing 100039, PR China.

Takahiro Yonezawa

School of Life Sciences, Fudan University, Shanghai, China.

Mack Yuen

Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T1Z4, Canada.

Bojian Zhong

School of Life Sciences, Fudan University, Shanghai, China.

Yang Zhong

School of Life Sciences, Fudan University, Shanghai, China.

This page intentionally left blank

List of Abbreviations

ABI	Association for biodiversity information
AGCM	Atmospheric general circulation models
AIC	Akaike information criterion
AMS	Accelerator mass spectrometry
APG	Angiosperm phylogeny group
BAC	Bacterial artificial chromosomes
BI	Bayesian inference
bp	Base pairs
BP	Before present
BSA	Bovine serum albumin
CBoL	Consortium for the barcode of life
CHARs	Charcoal accumulation rates
cp	Chloroplast
CR model	Correlated-rates model
CSRE	Conifer specific repeat element
CTM	Central Transantarctic Mountains
DEM	Digital elevation model
DIVA	Dispersal-vicariance analysis
DOGMA	Dual organellar genome annotator
dsRNA	Double-stranded RNA
EAM	East asian monsoon
ELF5	Early Flowering 5
ESA	Endangered Species Act of 1973
ESS	Effective sample size
ESTs	Expressed sequence tags
ETS	External transcribed spacer
FL	Full-length
FRI	Fire return intervals
FT	Flowering Locus T
GISs	Geographical information systems
GLRs	Glutamate receptors
GPS	Global positioning system
GSW	Weighted gain steps
HPD	Highest posterior density
HPT	Hashed position tree

IR model	Independent-rates model
IR	Inverted repeat
IS	Insertion sequence
ITS	Internal transcribed spacer
LBC	Long branch clade
LGM	Last glacial maximum
LOB	Lateral Organ Boundaries
LRT	Likelihood ratio test
LSC	Large single copy
LSW	Weighted loss steps
LTR	Long terminal repeat
MCMC	Markov Chain Monte Carlo
MeJA	Methyl jasmonate
MIPS	Munich Information Center for Protein Sequences
MP	Maximum parsimony
NJ	Neighbor-joining
NPRS	Non-parametric rate smoothing
nrITS	Nuclear ribosomal internal transcribed spacer
ORFs	Open reading frames
PAS	Periodic acid-Schiff's reagent
PBS	Phosphate buffered saline
PcGs	Polycomb-group proteins
PCR	Polymerase chain reaction
PFT	Plant functional types
PL	Penalized likelihood
Plant GDB	Plant Genome DataBase
PMC	Pollen mother cells
PP	Posterior probabilities
PTGS	Post-transcriptional gene silencing
PWG	Plant working group
RSCU	Relative synonymous codon usage
SAdM	S-Adenosylmethionine
SAM	Shoot apical meristem
SBC	Short branch clade
SEM	Scanning electron microscopy
SSC	Small single copy
TAIR	The <i>Arabidopsis</i> information resource
TBR	Tree bisection-reconnection
TEs	Transposable elements

TNC's	The Nature Conservancy's
TPSs	Terpenoid synthases
tRNA-Sec	tRNA-selenocysteine
UCLD	Uncorrelated lognormal
UPA	Universal plastid aplicon
WAAA	Weighted ancestral area analysis

This page intentionally left blank

Contents

1. Paleobotany of Livingston Island.....	1
M. Leppe, W. Michea, C. Muñoz, S. Palma-Heldt, and F. Fernandoy	
2. Targeted Isolation Sequence Assembly and Characterization of White Spruce	5
Björn Hamberger, Dawn Hall, Mack Yuen, Claire Oddy, Britta Hamberger, Christopher I. Keeling, Carol Ritland, Kermit Ritland, and Jörg Bohlmann	
3. Expressed Sequence Tag Analysis in <i>Cycas</i>	21
Eric D. Brenner, Dennis W. Stevenson, Richard W. McCombie, Manpreet S. Katari, Stephen A. Rudd, Klaus F.X. Mayer, Peter M. Palenchar, Suzan J. Runko, Richard W. Twigg, Guangwei Dai, Rob A. Martienssen, Phillip N. Benfey, and Gloria M. Coruzzi	
4. Complete Chloroplast Genome Sequence of a Tree Fern <i>Alsophila Spinulosa</i>.....	36
Lei Gao, Xuan Yi, Yong-Xia Yang, Ying-Juan Su, and Ting Wang	
5. Reverse Genetic Analyses of Gene Function in Fern Gametophytes.....	54
George Rutherford, Milos Tanurdzic, Mitsuyasu Hasebe, and Jo Ann Banks	
6. Distribution and Dynamics of Hayscented Fern.....	67
Songlin Fei, Peter J. Gould, Melanie J. Kaeser, and Kim C. Steiner	
7. Glossopterid Seed Ferns from the Late Permian, Antarctica.....	74
E. L. Taylor, T. N. Taylor, and P. E. Ryberg	
8. Unveiling Cryptic Species Diversity of Flowering Plants	82
Yudai Okuyama and Makoto Kato	
9. Pollen Development in <i>Annona Cherimola</i> Mill. (Annonaceae)	102
Jorge Lora, Pilar S. Testillano, Maria C. Risueño, Jose I. Hormaza, and Maria Herrero	
10. DNA Barcoding in the Cycadales	117
Chodon Sass, Damon P. Little, Dennis Wm. Stevenson, and Chelsea D. Specht	
11. Chloroplast Genomes in Ancestral Grasses	131
Bojian Zhong, Takahiro Yonezawa, Yang Zhong, and Masami Hasegawa	
12. EST Analysis in <i>Ginkgo biloba</i>.....	144
Eric D. Brenner, Manpreet S. Katari, Dennis W. Stevenson, Stephen A. Rudd, Andrew W. Douglas, Walter N. Moss, Richard W. Twigg, Suzan J. Runko, Giulia M. Stellari, W.R. McCombie, and Gloria M. Coruzzi	
13. Frequent Fires in Ancient Shrub Tundra	158
Philip E. Higuera, Linda B. Brubaker, Patricia M. Anderson, Thomas A. Brown, Alison T. Kennedy, and Feng Sheng Hu	
14. History of Native Plant Communities in the South	169
Wayne Owen	

15. Predicting Pleistocene Climate from Vegetation in North America.....	193
C. Loehle	
16. Past Vegetation Patterns of New Mexico's Rio Del Oso Valley	205
Richard D. Periman	
17. East Asian Monsoon and Paleoclimatic Data Analysis	220
J. Guiot, Hai Bin Wu, Wen Ying Jiang, and Yun Li Luo	
18. Evolutionary Divergence Times in the Annonaceae.....	231
Yvonne C.F. Su and Richard M.K. Saunders	
19. Early Cretaceous of Northern Japan.....	256
Toshihiro Yamada, Harufumi Nishida, Masayoshi Umebayashi, Kazuhiko Uemura, and Masahiro Kato	
Permissions.....	264
References.....	266
Index	311

Chapter 1

Paleobotany of Livingston Island

M. Leppe, W. Michea, C. Muñoz, S. Palma-Heldt, and F. Fernandoy

INTRODUCTION

This is the first report of a fossil flora from Hannah Point, Livingston Island, South Shetland Islands, Antarctica. The fossiliferous content of an outcrop, located between two igneous rock units of Cretaceous age are mainly composed of leaf imprints and some fossil trunks. The leaf assemblage consists of 18 taxa of Pteridophyta, Pinophyta, and one angiosperm. The plant assemblage can be compared to other Early Cretaceous floras from the South Shetland Islands, but several taxa have an evidently Late Cretaceous affinity. A Coniacian-Santonian age is the most probable age for the outcrops, supported by previous K/Ar isotopic studies of the basalts over and underlying the fossiliferous sequence.

The Cretaceous was a crucial time of vegetation change, largely due to the evolutionary and geographic radiation of angiosperms. Throughout the Late Cretaceous angiosperms progressively infiltrated the pre-existing vegetation, but gymnosperms, ferns, and sphenophytes dominated land-plant biomass until the Cenozoic (Cantrill and Poole, 2002; Spicer et al., 1993).

In the Turonian-Coniacian of the Antarctic, the abundance and diversity of angiosperms remained low (Dettmann, 1989), while cryptogams were still an important part of the vegetation. This was, however, an important transitional time when the established southern podocarp conifer forest vegetation was diversifying, and new angiosperm families that henceforth typified southern vegetation started to appear (Askin, 1989; Crame, 1992, Cantrill and Poole 2002, 2004). Palynomorph assemblages indicate the predominance of temperate, podocarp-araucarian-fern forest vegetation, with lycopod and fern moorland vegetation in some areas (Cantrill and Poole, 2005; Dettmann, 1986; Dettmann and Thomson, 1987; Dettmann et al., 1992; Douglas and Williams, 1982; Truswell, 1990). The podocarp-araucarian-fern forest association continued through this interval in southern high latitudes (Askin and Spicer, 1995). The southern high latitude region was a locus of evolutionary innovation from the Turonian to the end of the Cretaceous (Askin and Spicer, 1995). This important forest dominated southern high latitude forests from the Santonian through the Paleocene, and produced pollen. It is now restricted to wet, cool temperate, maritime western margins of Tasmania (Dettmann, 1989).

A rich Campanian leaf assemblage from King George Island, South Shetland Islands, suggests a broad-leaved forest community, including evergreen types (with thick, coriaceous leaves) and deciduous *Nothofagus* growing in a subhumid mesothermal climate (Birkenmajer and Zastawniak, 1989). Abundant ferns, especially gleicheniaceous types, grew on moist lowland areas at the end of the Cretaceous.

The fossiliferous leaf content of Hannah Point (Figure 1) exhibits the dominance of a Podocarp-Araucarian-fern forest association, but with evidence of angiosperms, restricted to the Late Cretaceous in the South Shetland Islands.

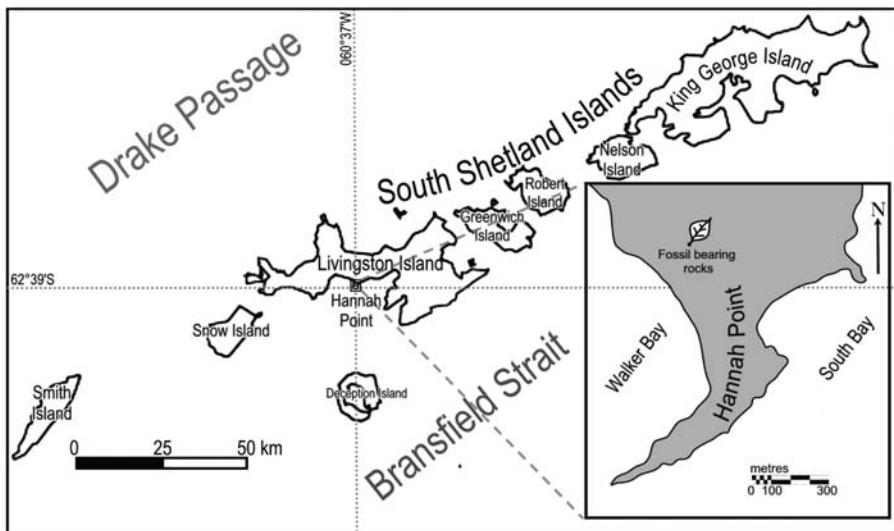


Figure 1. Podocarp-Araucarian-fern forest association.

GEOLOGICAL SETTING

Lithologies that belong to the main tectonic units of the Antarctic Peninsula region (such as forearc basin, magmatic arc, and extension-related volcanics) crop out on Livingston Island (Hathway, 2000; Pallàs et al., 1999). Smellie et al. (1996) ruled out any cogenetic relationship between the volcanic rocks of Hannah Point and the volcanics included in the Byers Formation, but suggested a possible cogenetic relationship with the dolerites cropping out at Siddons Point. Later isotopic analyses (Sr, Nd, and Pb) and the study of La/Sm-La and Ce/Yb-Ce pairs indicate that the rocks of Hannah Point display crustal contamination and correspond to a different magma source from the rocks of Siddons Point, Byers Peninsula, and Cape Shirreff (Xiangshen et al., 1996).

Hannah Point consists of a homoclinal, north-, northwest-dipping sequence with a thickness of about 500 m, comprising basaltic-andesite to dacite lavas interbedded with a variety of pyroclastic and thin sedimentary beds. Two lava samples were obtained, both are basaltic andesites, from the top and just above the center of the sequence (Smellie et al., 1996). According to Hobbs (1968), the basal 43 m of the Hannah Point succession consists of massive andesites interbedded with green agglomerates and amygdaloidal lava sand tuffs. The overlying 110 m are composed of friable agglomerates, fissile ashes, and amygdaloidal lavas, while the top 195 m are composed of massive andesite layers (much thicker than the basal ones) interbedded with amygdaloidal lavas.

Pallàs et al. (1999) defined five units for the volcanic succession cropping out at Hannah Point. The upper part corresponds to an approximately 80 m thick coherent andesitic lava layer that shows columnar jointing. Smellie et al. (1996) analyzed, with K-Ar isotopic method, two igneous rock samples from Hannah Point that yielded conflicting ages of 87.9 ± 2.6 My and 67.5 ± 2.5 My from the center and top of the sequence, respectively. The younger sample belongs to a 150 m thick upper part of the E unit of Pallàs et al. (1999). However, in 1999, when the paper of Pallàs et al. was published, the ice and snow covered the lower part of the unit E. During the 2006–2007 field season of the XLIII Chilean Antarctic Scientific Expedition, the exposure of the lower layers of the Pallàs E unit was revisited, and a sequence was discovered consisting of 25 m thick fine-grained green sandstone with abundant leaf imprints.

PALEOBOTANY

The leaf megaflora of the Hannah Point outcrops consists of: *Adiantites* sp., *Elatocladus* sp., *Matonia jeffersonii*, Angiosperm form 1 and 2, *Gleichenites* sp., *Podozamites cf. binatus*, *Brachyphyllum* sp., *Microphyllopteris unisoria*, *Ptilophyllum* sp., *Thrysopteris elegans*, *Coniopteris* sp., *Sphenopteris* sp., *Cladophlebis Antarctica*, *Pachypterus* sp., *Cladophlebis oblonga* (Figure 2).

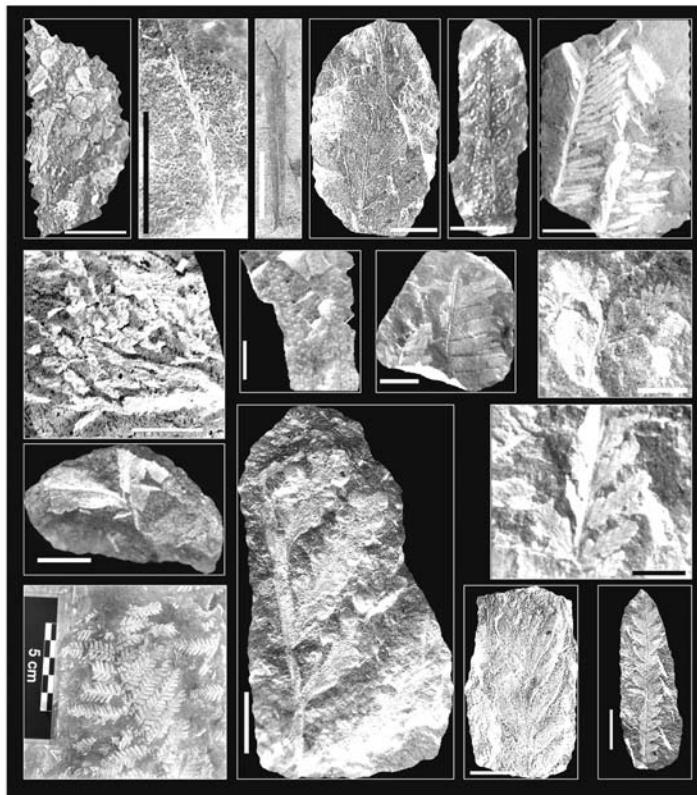


Figure 2. Leaf megaflora of the Hannah Point outcrops.

DISCUSSION

Several leaf types have been described from the Albian-Aptian floras of Byers Peninsula and Snow Island, and less from Alexander Island and James Ross Island, reflecting a strong connection with the lower Cretaceous Antarctic floras, but the presence of other common elements to younger outcrops, puts a reasonable doubt on an Early Cretaceous age for this flora.

The fossil-bearing rocks belong to a previously dated volcanic-sedimentary succession at Hannah Point. According to this previous isotopic analysis (Pallàs et al., 1999; Smellie et al., 1996) the age could be Coniacian and probably also Santonian age, but it is necessary to confirm that with other geochemical analyses. The sequences vary in composition locally, reflecting lateral facies changes, taphonomic processes, and climatic differences. The assemblages are commonly dominated by ferns, mainly of the families Osmundaceae, Dicksoniaceae and Gleicheniaceae, and foliage of conifers (mostly referable to the Podocarpaceae). The occurrence of angiosperm leaves is rare, but recently angiosperm pollen has been found in pollen and spore-bearing rocks from Hannah Point (Palma-Heldt, personal communication). The age proposed here is similar to the isotopic dating (Pallàs et al., 1999; Smellie et al., 1996) and with the phases described by Cantrill and Poole (2005) for the vegetational development in Antarctica during the Cenomanian to Santonian.

It is not easy to understand how angiosperms migrated rapidly into the conifer-fern dominated floras of Phase 1 and how they subsequently rose to dominance with-in Phase 2 due to the lack of knowledge of Cenomanian to Turonian floras (Cantrill and Poole, 2005). By Phase 3 of the Antarctic ecosystem evolution fossil leaf floras suggest that angiosperms were diverse (Hayes, 1999) and an array of angiosperms had evolved to form a significant tree-forming element of the vegetation during the Coniacian (Chapman and Smellie, 1992; Poole and Cantrill, 2001). Hannah Point's flora can help to complete the lack of knowledge of the evolution of the plant communities of the early upper Cretaceous.

KEYWORDS

- **Albian-Aptian floras**
- **Campanian**
- **Cretaceous**
- **Early Cretaceous**
- **Fossiliferous**

ACKNOWLEDGMENTS

This chapter was prepared as a collaborative work between the Chilean Antarctic Institute (INACH) project 05-06 and the Chilean Bicentennial Ring of Science and Technology Program (PBCT) “Geological Connection between West Antarctica and Patagonia since the Late Paleozoic”, both under the IPY Full Proposal GeoPaSAA (Nº 115), project under the Plates & Gates IPY Cluster Project. Thanks to photographer and graphic designer Jeniffer Muñoz.

Chapter 2

Targeted Isolation Sequence Assembly and Characterization of White Spruce

Björn Hamberger, Dawn Hall, Mack Yuen, Claire Oddy,
Britta Hamberger, Christopher I. Keeling, Carol Ritland,
Kermit Ritland, and Jörg Bohlmann

INTRODUCTION

Conifers are a large group of gymnosperm trees which are separated from the angiosperms by more than 300 million years of independent evolution. Conifer genomes are extremely large and contain considerable amounts of repetitive DNA. Currently, conifer sequence resources exist predominantly as expressed sequence tags (ESTs) and full-length (FL) cDNAs. There is no genome sequence available for a conifer or any other gymnosperm. Conifer defense-related genes often group into large families with closely related members. The goals of this study are to assess the feasibility of targeted isolation and sequence assembly of conifer bacterial artificial chromosomes (BAC) clones containing specific genes from two large gene families, and to characterize large segments of genomic DNA sequence for the first time from a conifer.

We used a PCR-based approach to identify BAC clones for two target genes, a terpene synthase (3-carene synthase; 3CAR) and a cytochrome P450 (CYP720B4) from a non-arrayed genomic BAC library of white spruce (*Picea glauca*). Shotgun genomic fragments isolated from the BAC clones were sequenced to a depth of 15.6- and 16.0-fold coverage, respectively. Assembly and manual curation yielded sequence scaffolds of 172 kbp (3CAR) and 94 kbp (CYP720B4) long. Inspection of the genomic sequences revealed the intron-exon structures, the putative promoter regions, and putative *cis*-regulatory elements of these genes. Sequences related to transposable elements (TEs), high-complexity repeats and simple repeats were prevalent and comprised approximately 40% of the sequenced genomic DNA. An *in silico* simulation of the effect of sequencing depth on the quality of the sequence assembly provides direction for future efforts of conifer genome sequencing.

We report the first targeted cloning, sequencing, assembly, and annotation of large segments of genomic DNA from a conifer. We demonstrate that genomic BAC clones for individual members of multi-member gene families can be isolated in a gene-specific fashion. The results of the present work provide important new information about the structure and content of conifer genomic DNA that will guide future efforts to sequence and assemble conifer genomes.

Conifers (Coniferales) are a large group of gymnosperm trees which are separated from the angiosperms by more than 300 million years of independent evolution. The conifers include the economically and ecologically important species of spruce

(*Picea*) and pine (*Pinus*), which dominate many of the world's natural and planted forests (Ralph et al., 2006). The development of genomic resources for conifers has focused on the discovery and characterization of expressed genes in the form of ESTs and full-length (FL) cDNAs. The available conifer cDNA sequence resources are extensive (1,158,419 ESTs as of December 3, 2008), representing almost 9% of all ESTs in the plant genome database (<http://plantgdb.org/>, http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). The EST and FLcDNA resources developed for white spruce (*Picea glauca*), Sitka spruce (*P. sitchensis*), and a hybrid white spruce (*P. glauca* × *P. engelmannii*) (Pavy et al., 2005; Ralph et al., 2008), have enabled transcriptome profiling (Friedmann et al., 2007; Holliday et al., 2008; Pavy et al., 2008; Ralph et al., 2006), proteome analysis (Lippert et al., 2005, 2007, 2009), marker development (Bérubé et al., 2007; Namroud et al., 2008; Pelgas et al., 2006; Rungis et al., 2004), and the functional characterization of gene products (Keeling et al., 2008; Martin et al., 2004; Phillips et al., 2007). These functional genomics studies have provided considerable insights into conifer defense against insects and pathogens, adaptation to the environment, and development (Ralph et al., 2006; Friedmann et al., 2007).

Beyond the characterization of cDNAs and their encoded proteins, the lack of a gymnosperm reference genome sequence limits our knowledge of the organization, structure, and gene space of conifer genomes. Sequencing a conifer genome has not yet been attempted and will remain a daunting task, given that conifer genomes range in size from 20 to 40 Gbp, which is 200–400-fold larger than the genome of *Arabidopsis* and larger than any other genome sequenced to date. The sequencing of a conifer genome may also be challenging due to a very high content of repetitive DNA (Ahuja, 2005) and the tendency of conifers to out-cross, preventing the development of inbred strains. An important step in assessing the feasibility of conifer genome sequencing will be the isolation, in random or targeted fashion, of genomic gDNA in the form of BAC clones, followed by the sequencing and assembly of large segments of gDNA. However, to the best of our knowledge, sequencing of a complete BAC clone or any large segment of nuclear gDNA has not yet been reported in the literature for a conifer or any other gymnosperm species. Recently, a loblolly pine (*Pinus taeda*) gDNA BAC library was used to assess the contribution of a novel pine-specific retrotransposon family (Gymny) to conifer genome size (Morse et al., 2009).

Unlike in angiosperms, conifers are not thought to have undergone recent genome duplication events (Ahuja, 2005; Cui et al., 2006). However, two features of conifer genomes pose untested challenges for the targeted isolation and sequence assembly of BACs containing genes of interest involved in conifer defense. First, many conifer defense genes exist as closely related members of large families. For example, genes encoding the oleoresin producing terpenoid synthases (TPSs) (Keeling et al., 2008; Martin et al., 2004), cytochrome P450 monooxygenases (P450s) involved in diterpene resin acid formation (CYP720B) (Hamberger and Bohlmann, 2006; Ro et al., 2005), TIR-NBS-LRR disease resistance proteins (Liu and Ekramoddoullah, 2003), pathogenesis-related (PR)-10 proteins (Liu and Ekramoddoullah, 2004), and dirigent proteins (Ralph et al., 2006, 2007a) are members of such multigene families. Against the background of large gene families it may be difficult to isolate BACs for a specific target gene. Second, the abundance of TEs, specifically those of the Copia and Gypsy

classes, which have been demonstrated by *in situ* hybridizations as diverse families of retroelements across conifer chromosomes (Friesen et al., 2001; L'Homme et al., 2000), may cause additional problems with genome sequence assemblies.

In this chapter we report a successful strategy for the targeted BAC identification and isolation of TPS and P450 genes using PCR-based screening of a non-arrayed white spruce BAC library of 3X genome coverage, and the subsequent gDNA insert sequencing, sequence assembly, and sequence characterization. When extended to other conifers, our strategy will enable a comparative analysis of synteny of specific target regions of conifer genomes.

MATERIALS AND METHODS

White Spruce BAC Library

Genomic gDNA was isolated from 200 g fresh weight of apical shoot tissue collected in April 2006 from a single white spruce (*Picea glauca*, genotype PG29) tree at the Kalamalka Research Station (British Columbia Ministry of Forests and Ranges, Vernon, British Columbia, Canada). A BAC library cloned into the HindIII site of pIndigoBAC-5 was made by BioS&T (<http://www.biost.com/>, Montreal). The non-arrayed library consisted of approximately 1.1 million BAC clones with an average insert size of 140 kbp, representing approximately 3× coverage of the white spruce genome.

BAC Library Screening and Shotgun Subcloning into Plasmid Libraries

The BAC library was screened by BioS&T for two target genes, a *TPS* gene encoding 3-carene synthase (3CAR), and a P450 gene encoding a diterpene oxidase (CYP720B4) using the procedures detailed in Isodore et al. (2005). In brief, the entire BAC library was plated (977 plates; approximately 1,200 colonies per plate) and colonies were transferred into ten 96-well plates with approximately 1,000 BAC clones per well (pool). Twenty super-pools of BAC clones were generated for each of the ten 96-well plates by combining the wells from 12 vertical rows and eight horizontal columns. These super-pools were screened by PCR for the two target genes. We used all available spruce EST and FLcDNA sequence information to design PCR primers that are, to the best of current knowledge, specific for the two target genes, while suppressing amplification of other known members of the spruce *TPS* and *P450* gene families. Primers were designed to amplify fragments of approximately 500 bp, were evaluated with white spruce PG29 gDNA. The primer sequences (shown in 5'-3' orientation) are CTTCAAGCCAATACCAAAGGCCTG and GGAAATGGCAATCACTG-CATTGGTATAG for CYP720B4; and GGAGAATTAGTGAGTCATGTCGATG and CTCTGTCTGATTGGTGGAACAGGC for 3CAR. The PCR products from super-pools were sequenced to confirm the identity of the target DNA. The individual pool (well containing the target gDNA clone) was identified, confirmed by PCR, and individual BAC clones isolated as described in Isodore et al. (2005).

Isolated BAC clones PGB02 (3CAR) and PGB04 (CYP720B4) were digested with NotI to release the insert, and insert DNA size was determined by pulse field gel electrophoresis. The gDNA inserts of PGB02 and PGB04 were isolated by gel purification

and sheared using a nebulizer (Invitrogen). After blunt-end repair, gDNA fragments were size fractionated on SeaPlaque agarose gels (CBM Intellectual Properties, Inc.). Fragments of 700–2,000 bp were recovered and ligated into the SmaI site of pUC18. Plasmids were transformed in *E. coli* DH10B.

Sequencing and Automated Sequence Assembly

Shotgun subcloned plasmid libraries for PGB02 and PGB04 were arrayed in 384-well plates and gDNA inserts were Sanger-sequenced from both ends. Sequences were scanned and masked for vector sequences and contaminating bacterial sequences, eliminating 21.4% (PGB02) and 27.9% (PGB04) of the total sequences. This high level of contaminating DNA resulted from prolonged growth of bacterial cultures prior to BAC isolation. We have subsequently found that the use of Plasmid-Safe ATP-dependent DNase (Epicenter) reduces the amount of contaminating bacterial DNA.

Sequences were processed using PHRED software (version 0.020425.c) (Ewing and Green, 1998), quality-trimmed according to the high-quality contiguous region determined by PHRED, and vector-trimmed using CROSS_MATCH software <http://phrap.org/>. Vector and bacterial contaminated DNA sequences were identified by sequence alignments using megaBLAST to all UniVec and non-redundant bacterial sequences from NCBI respectively, and hits with 95% identity were subsequently masked with N's. Processed sequences were assembled with PHRAP <http://www.phrap.org/> using the base quality files and with the bi-directional reads generated for each clone pre-assembled by PHRAP to match paired reads. The two commonly used assembling routines CAP3 and PHRAP were tested for their capability of assembling the BAC sequences. Despite CAP3 employing a higher stringency as compared to PHRAP (Huang and Madan, 1999), The PHRAP assemblies of both BAC clones resulted in fewer but higher quality contigs which included more total sequences (PGB02: CAP3 49 contigs, PHRAP 14 contigs; PGB04: CAP3 19 contigs, PHRAP 14 contigs). The gDNA sequences identified in this work were submitted to NCBI GenBank under accession numbers FJ609174 (PGB02) and FJ609175 (PGB04).

Manual Curation of Sequence Assemblies

The contigs for PGB02 (15 contigs) and PGB04 (14 contigs) obtained by automated sequence assembly were manually curated. Sequences that prevented correct assembly such as sequences from chimeric DNA were removed and the remaining contigs were re-aligned. PGB02 was manually assembled into two contigs. Assembly of PGB04 into a single contig required the re-introduction of several sequences which had been previously identified as contaminating *E. coli* sequence. Examination of this *E. coli* sequence identified it as the insertion sequence (EcIS10) of the plasmid-associated bacterial transposon Tn10, which was presumably inserted into the BAC during proliferation. The left and right arms of the BAC vector (pIndigoBAC-5) were used to orient the remaining contigs, resulting in the final builds of PGB02 and PGB04.

Oligonucleotide primers were designed to bridge gaps in automated and manually curated sequence assemblies of PGB02. The PCR using PGB02 BAC DNA and primers placed 1,112 bp and 993 bp on either side of the gap generated a single band

of approximately 2.2 kbp. Sequencing of this PCR product verified up to 900 bp of sequence on either side of the gap but no additional sequences for the gap region were obtained, possibly due to low sequence complexity. For sequence finishing, oligonucleotide primers (shown in 5'-3' orientation) were designed based on the sequence scaffolds of PGB02 (AATTGGTCAATTCTAAAACACCATG, AAATTATGGGTT-TAAGGGCTAGAGTTC) and PGB04 (AACAAATTACTCATTACCCGTGA, CCCATCAAATCCATGCCAAG, TTCCAAGTTCTGTGGGAGGAG, GACT-GATTTCTCTCCACCAAGCAAG).

Sequence Analysis

Repetitive DNA was identified with the RepeatMasker software (A.F.A. Smit, R. Hubley and P. Green, unpublished data. Current Version: open-3.2.6 (RMLib: 20080801)), using the viridiplantae section of the RepBase Update (Jurka et al., 2005) as a database. Gene models were predicted using the *ab initio* gene finder FGENESH (dicot matrix; (Salamov and Solovyev, 2000)), Genscan and GeneMark.hmm with default parameters. Regions with similarity to DNA transposons were identified with RepeatMasker (Chen, 2004; Jurka et al., 2005) with a threshold score over 200 and a length over 100 bp.

Cloning and Sequencing of Up-Stream Regions of 3CAR and CYP720B4

The regions upstream of the start codon including the 5'UTR and promoter regions for 3CAR and CYP720B4 were amplified by PCR using white spruce PG29 gDNA as a template. Gene-specific oligonucleotide primers (shown in 5'-3' orientation) were based on the BAC scaffolds of PGB02 (3CAR) (ACCCATCTTCACAAAATTAC, GTAGTCCATAACGAGCAGAA) and PGB04 (CYP720B4) (TGATATTGGTCT-GCCATGGCG, CATTCCCTGCATGTATTCAATGCC, CCACCACATAGTTA-GACCGTGATGC).

DISCUSSION

Sequencing and Assembly of BAC Clones as a Test for Conifer Genome Sequencing

To date there is no sequence report for large segments of conifer gDNA, and researchers have avoided sequencing a conifer genome due to the large size and high content of repetitive elements. Several approaches are currently being considered for future efforts to sequence a conifer genome including the high-throughput sequencing of BAC libraries. To assess the feasibility of sequencing and assembling long, continuous segments of conifer gDNA, we targeted two white spruce defense genes, 3CAR and CYP720B4, for BAC clone isolation, sequencing, and assembly. These genes were chosen because they are known to be members of large gene families with key functions in terpenoid biosynthesis.

Pre-assembled bi-directional reads of shotgun plasmid libraries for each BAC clone were assembled using PHRAP software resulting in a large number of contigs (15 for PGB02 and 14 for PGB04). Both BAC clones had areas of reduced quality reads with low or no sequence coverage bordered by regions of low complexity sequence

repeats, which necessitated manual curation of the sequence assembly resulting in substantially improved sequence assemblies of two (PGB02) and one (PGB04) contigs. High-complexity and simple repeats did not interfere with the automated PHRAP assembly and manual inspection of the contigs did not reveal falsely matched reads within the repeat regions. The use of pre-assembled paired reads and quality scores produced by PHRED balanced between tolerating discrepancies and complete mis-assembly of the data sets (de la Bastide and McCombie, 2007). We found that most problems for automated sequence assembly resulted from chimeric clones in the plasmid libraries, bacterial DNA contamination, low-quality sequences, and low-complexity repeats.

Targeted BAC Isolation of Members of Large Conifer Defense Gene Families Provides Insights into Gene Content of a Conifer Genome

The two genes targeted for BAC sequencing are members of large defense-related *TPS* and *P450* gene families in spruce (Hamberger and Bohlmann, 2006; Keeling and Bohlmann, 2006). In the TPS gene family, members with more than 90% sequence identity can have distinct biochemical functions with non-overlapping product profiles (Keeling et al., 2008; Martin et al., 2004). In this study we demonstrate for the first time that it is possible to isolate, in an efficient and targeted fashion, BAC clones for specific members of the large conifer *TPS* and *CYP720* defense gene families, thus providing new opportunities to characterize members of these important defense gene families at the genome level.

The *3CAR* gene contains 10 exons and nine introns, identical to the exon-intron structure of the grand fir (*Abies grandis*) monoterpane synthase genes (–)-limonene synthase and (–)- α / β -pinene synthase, previously cloned by PCR amplification of the gDNAs between the start and stop codons identified in the corresponding FL-cDNAs (Figure 2C) (Trapp and Croteau, 2001). The identity of the deduced amino acid sequence to the previously functionally characterized Norway spruce *3CAR* (Fäldt et al., 2003) is 84%. The *CYP720B4* gene has nine exons and eight introns, and is the first genomic structure reported for a gymnosperm P450 gene. A comparison of the *CYP720B4* gDNA structure with the gDNA structures of *Arabidopsis* P450s shows highly conserved intron-exon boundaries between *CYP720B4* and *Arabidopsis* *CYP88*, which is involved in the primary metabolism of giberellin biosynthesis. Both families of P450s share a similar reaction mechanism and catalyze consecutive oxidation steps of structurally similar substrates (Ro et al., 2005). These findings suggest a common ancestor of *CYP88* (primary metabolism) and *CYP720B4* (secondary metabolism).

Despite the large size of conifer genomes (estimated 20–40 Gbp; 200–400-fold larger than the genome of *Arabidopsis*), it is not likely that the spruce genome contains a proportionally larger number of protein coding genes than *Arabidopsis* as estimated from EST and FLcDNA discovery (Ralph et al., 2008). In contrast to previously sequenced angiosperm genomes, the spruce gDNA sequences of PGB02 and PGB04 reveal a low gene density, with a single gene per 172 kbp and 94 kbp respectively, which is at least 10-fold lower than the overall gene density of the genomes of *Arabidopsis*, rice, poplar, and grapevine (Table 1). This observation of low gene density has also

been confirmed by additional sequencing of several randomly selected spruce BAC clones (K. Ritland et al., unpublished results).

In angiosperms, several mechanisms contribute to the expansion of gene families, including whole genome and chromosome segmental duplications (De Bodt et al., 2005), and tandem duplication of closely related genes (Rizzon et al., 2006). For the gene family members targeted in this work, we did not find evidence for local tandem duplication.

The Upstream Regions of 3CAR and CYP720B4 Contain Putative *cis*-Acting Elements Consistent with the Roles of these Genes in Induced Defense

A large volume of previous research on the regulation and coordination of defense responses in spruce has targeted processes at the anatomical and molecular levels of induced metabolite accumulation, enzyme activities, and transcript abundance of genes involved in the biosynthetic pathways of terpenoid and phenolic defenses (Bohlmann, 2008; Hudgins et al., 2006; Keeling and Bohlmann, 2006; Martin et al., 2002, 2003; McKay et al., 2003; Miller et al., 2005; Phillips et al., 2007; Ralph et al., 2007a, 2007b). In particular, 3CAR transcripts were up-regulated by real and simulated insect attack in Sitka spruce (Miller et al., 2005) and in Norway spruce (Fäldt et al., 2003). In loblolly pine transcripts of the *CYP720B4* related *CYP720B1* were up-regulated in response to MeJA treatment (Ro et al., 2005). In addition, large-scale proteome and gene expression profiling has identified putative transcription factors in spruce that were up-regulated in response to real or simulated insect attack (Lippert et al., 2005, 2009; Ralph et al., 2006). This is the first report of the upstream sequences of conifer defense-related genes and the putative *cis*-acting elements located in those regions.

The upstream sequences of 3CAR and *CYP720B4* each have more than five elements with sequence identity to *cis*-acting elements putatively involved in wound, stress, and defense responses in angiosperms. The promoter region of the *CYP720B4* gene is 95–99% identical with the corresponding PCR-amplified regions across several genotypes of Sitka spruce, hybrid interior spruce, and white spruce (data not shown). The conserved W-box motif present upstream of *CYP720B4* is recognized and bound by transcription factors of the plant specific WRKY class which mediate pathogen defense responses in angiosperms (Eulgem et al., 2000). More than 80 members of the WRKY family have been reported in pine (Liu and Ekramoddoullah, 2009; Zhang and Wang, 2005) and more than 10 different sequences with 60–80% identity to the *Arabidopsis* WRKY proteins AtWRKY6, AtWRKY3, and AtWRKY4, involved in defense, stress, and pathogen responses (Lai et al., 2008; Robatzek and Somssich, 2001) were found in the white spruce EST databases. These putative promoter regions and *cis*-acting elements represent valuable tools for future studies of the transcriptional regulation of conifer defense genes. Transformation of white spruce for characterization of promoters has been reported (Bedon et al., 2009; Godard et al., 2007). In future work we will use this transformation system, in parallel with transformation in heterologous plant systems, for functional testing of spruce TPS and P450 promoter constructs linked to reporter genes.

The finding of a novel 44 bp sequence element which is detected four times in the 5'UTR of the white spruce *3CAR* gene on PGB02 was also found 19 times in the 5'UTR of the orthologous gene isolated as a cDNA in Sitka spruce. The conservation of this short sequence across spruce species suggests that this element has an important functional role in the regulation of the *3CAR* gene.

Genomic Regions Surrounding the *3CAR* and *CYP720B4* Genes Contain DNA and RNA_B Transposable Elements

The genomic regions surrounding the *3CAR* and *CYP720B4* genes contain retrotransposons, DNA transposons, and simple repeat sequences. With the exception of a fully preserved IS10 element present in the genomic sequence of PGB04 (likely the result of transposition from the bacterial host *E. coli* genome), all repetitive sequences appear to have accumulated a large number of mutations, deletions, and rearrangements suggesting that these elements are no longer functional. The repeat regions in the gDNA of PGB02 (15%) and PGB04 (17%) have up to 89% similarity to white spruce TE-related ESTs. The presence of ESTs for these TEs indicates that members of these retrotransposon families may actively proliferate in conifers, potentially increasing genetic variability.

Remnants of DNA transposons of the cut-and-paste and copy-and-paste classes were found within 4 kbp and 500 bp of *3CAR* and *CYP720B4*, respectively. In maize, the DNA-transposon helitron is associated with the duplication of CYP72A (Jameson et al., 2008), and DNA-based transposons have been implicated in the capture and transduplication of host genes in rice, *Lotus japonicas*, and *Arabidopsis* (Hoen et al., 2006; Holligan et al., 2006; Juretic et al., 2005). The proximity of DNA transposons to the protein coding *3CAR* and *CYP720B4* genes is consistent with the possibility that a DNA transposon-mediated translocation mechanism may contribute to the diversification of the conifer TPS and P450 gene families.

RESULTS

Targeted Isolation of BAC Clones Containing *TPS* (*3CAR*) and *P450* (*CYP720B4*) Genes

Our first objective was to test if individual BAC clones containing conifer genes of large gene families could be isolated in a gene-specific manner. A white spruce (genotype PG29) gDNA BAC library of approximately 3X genome coverage was constructed, aliquoted into pools in ten 96-well plates, and screened in a hierarchical fashion by PCR as described previously (Isidore et al., 2005). The primers used to screen pooled BAC clones for a specific *TPS* gene were based on the functionally characterized Norway spruce (*Picea abies*) and Sitka spruce 3-carene synthase FLcDNAs (*3CAR*) (Fäldt et al., 2003), D. Hall, J. Robert, C.I. Keeling, J. Bohlmann, unpublished results). Primers used to screen for a specific target *P450* gene were based on the functionally characterized diterpene oxidase *CYP720B4* from Sitka spruce and its white spruce orthologue (B. Hamberger, T. Ohnishi, J. Bohlmann, unpublished results). The function of the spruce *CYP720B4* gene is similar to that of loblolly pine *CYP720B1* in diterpene resin acid formation (Hamberger and Bohlmann, 2006; Ro et al., 2005).

Primers used for gene-specific screening for *TPS* (*3CAR*)- or *P450* (*CYP720B4*)-containing BAC clones were assessed in silico against other known members of the large conifer TPS-d family (Martin et al., 2004) and other members of the conifer-specific CYP720B family (Hamberger and Bohlmann, 2006), respectively, to minimize the chance of isolating non-target members of these gene families.

From a total of 960 BAC pools (ten 96-well plates), which were initially screened as 200 super-pools (20 super-pools per 96-well plate) we identified 23 and 18 pools that yielded PCR products with the *3CAR* and *CYP720B4* primers, respectively. The 23 independent PCR products obtained with *3CAR* primers represented four unique *3CAR*-like sequences with at least 95% identity (in the open reading frame) amongst each other and to the Sitka spruce *3CAR* FLcDNA Q09. We also sequenced five independent PCR products obtained by screening the BAC pools with *CYP720B4* primers. All five sequences were 100% identical with the target *CYP720B4* sequence. For each of the two target genes, a single individual BAC clone was isolated, verified by sequencing the PCR product, and the gDNA inserts were excised and their size estimated based upon their mobility in pulsed field gel electrophoresis. The BAC clone PGB02 (*3CAR*) contained a gDNA insert of approximately 185 kbp and BAC clone PGB04 (*CYP720B4*) contained an insert of approximately 110 kbp. These gDNA inserts were sheared into fragments of 700–2,000 bp and shotgun-subcloned into plasmid libraries for sequencing.

Automated Sequence Assemblies of PGB02 and PGB04

The shotgun plasmid libraries for PGB02 and PGB04 were arrayed in 384-well plates. Plasmid inserts from 10 and five 384-well plates were Sanger-sequenced for PGB02 and PGB04, respectively, resulting in 6,954 and 3,677 paired sequence reads. The average plasmid insert length was 1,102 bp for the PGB02 library and 1,056 bp for the PGB04 library. Sequences were scanned and masked for vector sequences and contaminating bacterial sequences, eliminating 21.4% (PGB02), and 27.9% (PGB04) of the total sequences. Using PHRAP, we assembled the sequences into 15 contigs for PGB02 and 14 contigs for PGB04. For PGB02, the two largest contigs assembled in this automated fashion covered a total length of 172,403 bp (91.2% of the sequence reads); the three largest contigs for PGB04 covered over 93,905 bp (94.4% of the sequence reads).

Manual Curation of the Sequence Assemblies of PGB02 and PGB04

To improve the assembly of PGB02 and PGB04, we inspected each contig generated with the PHRAP software. We found that chimeric sequences, resulting from the ligation of independent gDNA fragments during the production of shotgun plasmid libraries, were included in some of the plasmid insert sequences, which together with low-quality sequences and low-complexity repeats, prevented the automated assembly into continuous sequence. In addition, we manually aligned shorter contigs with low sequence representation to the larger contigs. The left and right arms of the pIndigoBAC-5 vector, which were subcloned together with the gDNA inserts into the plasmid shotgun libraries, provided orientation for the scaffolds of PGB02 and PGB04 (Figure 1).

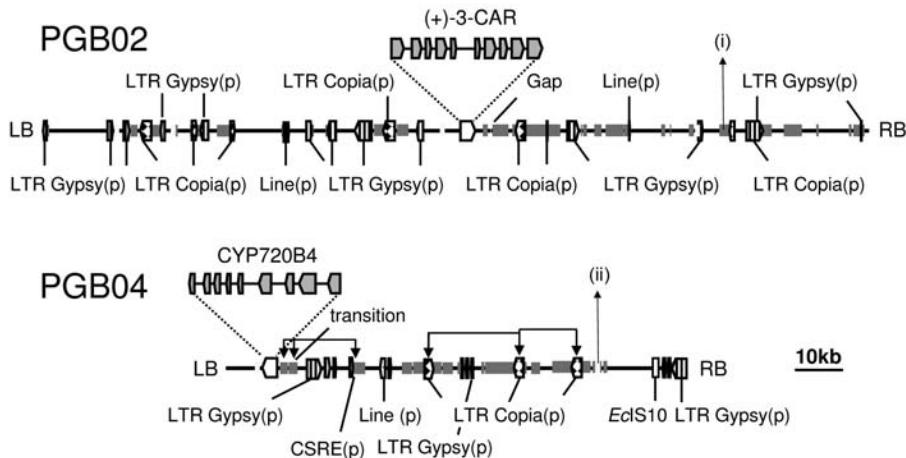


Figure 1. Structure of white spruce genomic DNA of BAC clones PGB02 and PGB04. The position of the target genes *3CAR* and *CYP720B4* is indicated. i and ii bars represent repeated segments and segments with similarity to DNA transposons, respectively. Transposable elements were identified with the RepeatMasker using the viridiplantae section of the RepBase Update database. EclS10, *E. coli* individual insertion sequence (IS) of the bacterial transposon Tn10; CSRE, conifer specific repeat element; LB/RB left and right border of pINDIGO; arrows in PGB04 indicate local putative segment duplications. The scale bar represents 10 kbp. (p) pseudogene, based on the accumulation of deleterious mutations and the absence of transcript with >90% identity.

The final assembly of PGB02 contained two contigs separated by a short gap (approximately 25–50 bp based on PCR amplification of the gap region) without sequence coverage. The gap is flanked by long stretches of low-complexity repeat sequence. It is likely that the sequence gap resulted from physical repeat structures (e.g., hairpins) which interfered with sequencing this region. Manual curation resulted in a single complete contig for the PGB04 gDNA. In PGB04 two high-complexity repeats and several low-complexity repeats extend for over 1 kbp on either side of a region of approximately 200 bp with low sequence coverage (transition) (Figure 1).

In summary, the combined automated and manual sequence assemblies resulted in two contigs for PGB02 with a combined sequence length of 172,056 bp and 15.6 \times sequence coverage, and into a single contigs for PGB04 with a sequence length of 93,592 bp and 16.0 \times sequence coverage. The size of the assembled sequence contigs for PGB02 and PGB04 agree well with the size of BAC inserts as estimated by PFGE (185 kbp and 110 kbp, respectively).

In Silico Analysis of the Effect of Sequencing Depth on Assembly Quality

Using the high sequence coverage (16 \times) and high-quality manually curated sequence assembly (93,592 bp) for PGB04 we analyzed the effect of plasmid shotgun library sequencing depth on the quality of the automated assembly. This assessment can guide cost-effective sequencing of BAC clones for future efforts of conifer genome sequencing. The sequences obtained from the plasmids of five 384-well plates for PGB04 were assembled into independent builds in all permutations of two, three, four, or

five plates. With sequences obtained from one plate, an average coverage of 3.2 \times was obtained and the number of nucleotides assembled into contigs (average contig number of 22.2) was less than 90 kbp (representing 93.0% coverage). By assembling sequences from two plates, the coverage doubled to an average of 6.4 \times , the number of contigs (average 9.9) was reduced, the assembly included over 95 kbp in contigs, and the full length scaffold had over 98% coverage relative to the reference PGB04 assembly. When sequences from three, four, or five plates were used in the assembly, coverage increased to 9.6 \times , 12.8 \times , and 16 \times , respectively, with a further increase in the number of nucleotides assembled. The assembly of sequences from three, four, or five plates also resulted in an increase of the number of contigs. Even with five plates, the coverage obtained by automated assembly never reached 100% relative to the PGB04 reference assembly, which involved manual curation.

Gene Content of PGB02 and PGB04

Results from the overall sequence analyses of the BAC clones PGB02 and PGB04, visualized using gbrowse, are available as online information at <http://treenomix3.msl.ubc.ca/cgi-bin/gbrowse/PGB02/> (username: treenomix; password: conifer). These descriptions include BLAST annotations (against NCBI NR, MIPS coniferales repeats, spruce ESTs), GC content and gene predictions (Genemark Prediction (Eukaryotic HMM), FGENESH Prediction, Genescan Prediction). PGB02 and PGB04 each contained a single functional gene identified by BLAST searches, which match the target genes *3CAR* (*PGB02*) and *CYP720B4* (*PGB04*) (Figure 1). Relative to the complete gDNA sequence length of PGB02 and PGB04, the gene density with a single gene per 172 kbp and 94 kbp, respectively, is at least 10-fold lower than the overall gene density of the sequenced genomes of *Arabidopsis*, rice, poplar, and grapevine (Table 1). The GC content (37%) of the two white spruce gDNAs was lower than the GC content of the rice genome (43.6%) and higher than those of the *Arabidopsis* (36%), poplar (33.7%), and grapevine (34.6%) genomes (Table 1) (AGI, 2000; IRGSP, 2005; Jaillon et al., 2007; Tuskan et al., 2006).

Table 1. General features of the gDNA sequences of the white spruce BAC clones PGB02 and PGB04 as compared to the genome sequence features of *Arabidopsis*, rice, poplar and grapevine.

	Genome Size (Mbp)	Predicted genes	Avg Gene length (bp)	Gene density (kbp per gene)	% TE	GC content (%)
<i>Arabidops isithaliana</i> ¹	115	25,498	1,992	4.5	14.0	36.0
<i>Orzya sativa</i> ²	389	37,544	2,699	9.9	34.8	43.6
<i>Populus trichocarpa</i> ³	485	45,555	2,392	10.6	42.0	33.7
<i>Vitis vinifera</i> ⁴	487	30,434	3,399	16.0	41.4	34.6
PGB02 ⁵	0.172	I	3,138	172	36.0	38.0
PGB04 ⁵	0.094	I	3,131	93.6	41.6	37.0

¹ [30–33]

²BACinsertsize

Analyzes of the gDNA Sequences for 3CAR and CYP720B4

The genomic region of the 3CAR gene on PGB02 covers 3,541 bp, including a 198 bp 5'-UTR and 205 bp 3'-UTR which are part of the corresponding transcript isolated from cDNA (Figure 2A). The gene contains 10 exons and nine introns, with intronic regions accounting for 35.4% of the gene sequence between the start and stop codon of this *TPS* gene. The genomic region of the CYP720B4 gene on PGB04 covers 3,131 bp over nine exons (1,452 bp) and eight introns and includes transcribed 5'- and 3'-UTRs of 38 bp and 134 bp, respectively (Figure 2B). The intronic region covers 50% of the gene sequence between the start and stop codon. The introns of 3CAR and CYP720B4 are of much lower GC content than the exons (% GC content exons/introns: 3CAR, 42.3/27.8; CYP720B4, 41.4/25.5).

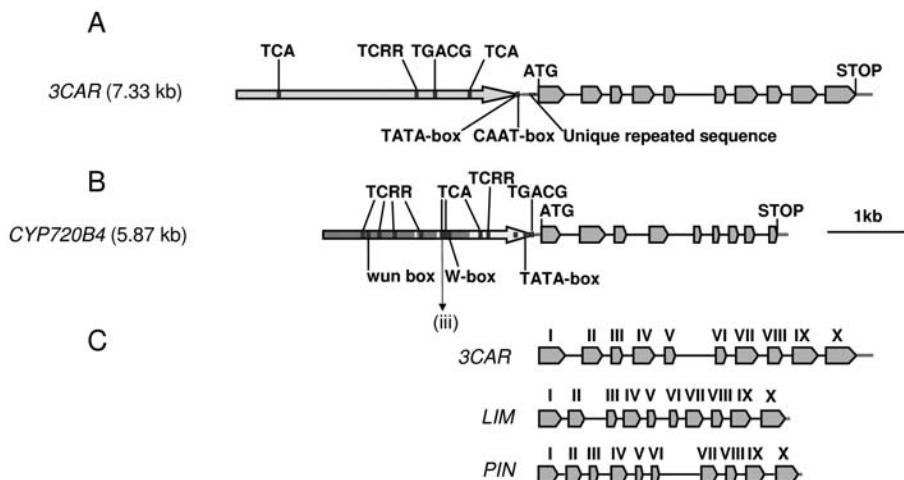


Figure 2. Gene structure of white spruce 3CAR (A) and CYP720B4 (B) and comparison of 3CAR with the grand fir (*Abies grandis*) limonene synthase (LIM) and pinene synthase (PIN) genes (C). Exons of the 3CAR and CYP720B4 genes matching the cDNA sequences are shown with grey arrows separated by introns. The UTRs are shown with grey lines. ATG, start codon. Putative *cis*-acting elements were identified using the PlantCARE database and positions are highlighted in iii (not to scale): wun-box, wound-responsive element (*Brassica oleracea*); W-box, fungal elicitor responsive element (*Petroselinum crispum*); TCRR, TC-rich repeats, *cis*-acting element involved in defense and stress responsiveness (*Nicotiana tabacum*); TCA, *cis*-acting element involved in salicylic acid responsiveness (*Brassica oleracea*); TGACG, *cis*-acting regulatory element involved in the MeJA-responsiveness (*Hordeum vulgare*). LIM, AF326518; PIN, AF326517; roman numbers in part C indicate conserved exons in 3CAR, LIM, and PIN; the scale bar represents 1 kbp.

Analyses of Upstream Promoter Regions of 3CAR and CYP720B4

Our analysis of upstream sequences for *cis*-regulatory elements covered 3,793 bp upstream of the ATG start codon for 3CAR and 2,500 bp upstream of the ATG start codon for CYP720B4. Putative *cis*-regulatory elements were identified by a similarity search of the PlantCARE database (Lescot et al., 2002). The region upstream of the ATG in 3CAR is unique until -3,973 bp which marks the location of a DNA transposon (Figure 1). In contrast, only the region from -1 bp to -749 bp upstream of the start codon

of *CYP720B4* is unique, followed by repetitive sequence (Figures 1 and 2). Several promoter enhancing sequences (TATA and CAAT boxes) were identified in the region immediately upstream of the start codon of the *3CAR* and *CYP720B4* genes (Figure 2).

Since the spruce *TPS* and *CYP720B* genes are involved in the biosynthesis of defense-related terpenoids induced by insects, pathogens, wounding, or methyl jasmonate (MeJA) (Bohlmann, 2008; Byun-McKay et al., 2006; Miller et al., 2005; Phillips and Croteau, 1999; Ro et al., 2005), we analyzed the upstream genomic regions of *3CAR* and *CYP720B4* for putative *cis*-acting elements associated with plant defense responses (Figure 2). In *CYP720B4*, a conserved W box motif (TTGACC), which interacts in *Arabidopsis* with members of the WRKY transcription factor family to mediate responses to wounding or pathogen responses (Eulgem et al., 2000), is located at position -1,129 relative to the ATG of *CYP720B4* on PGB04. A similar element (TGACG), involved in the MeJA-responsive gene expression in barley (*Hordeum vulgare*) (Rouster et al., 1997), is detected at -1,266 relative to the start codon of *3CAR* and at -79 relative to the start codon of *CYP720B4*. The upstream regions of *3CAR* and *CYP720B4* also contain a TCA-element at positions -815 and -3,291 in PGB02 and at positions -1,227, -676, and -1,162 (TCAGAAGAGGG, GAGAAAGAATA, and CAGAAAAGGA) in PGB04, respectively. This element was first characterized as a *cis*-acting element involved in salicylic acid responsiveness and systemic acquired resistance in wild cabbage (*Brassica oleracea*) (Goldsborough et al., 1993). In addition, we identified several TC-rich repeats (ATTTTCTCCA) in the up-stream regions of *3CAR* (one on PGB02) and *CYP720B4* (six on PGB04). These sequences were previously described in tobacco (*Nicotiana tabacum*) as *cis*-acting elements involved in defense and stress responsiveness (Klotz and Lagrimini, 1996).

The upstream regions of the *3CAR* and *CYP720B4* genes also include a large number of putative transcription factor binding sites (37 for *3CAR*; 19 for *CYP720B4*), implicated in light responsiveness in several other plant species. Interestingly, the promoter sequence including the transcribed 5'-UTR of the *3CAR* gene on PGB02 contains a unique and conserved repeated sequence of 44 bp (TCAGGTTCTGCCATTGCCTTTTAGTCATTATCTTGAGCTGCC) which is located four times (with no more than two nucleotide changes) between -21 and -199 bp upstream of the start codon. Seventeen of the 44 bp in this repeated sequence have high levels (94–100%) of sequence identity to plant I-box transcription factor binding sites, which are involved in light responsiveness (Giuliano et al., 1988). The actual role of this sequence in gene regulation is unknown, however, the prevalence of this sequence in the transcribed 5'-UTR of the *3CAR* gene on PGB02, as well as in the 5'-UTR of two white spruce *3CAR*-like ESTs (GQ03804.B7_I10 and GQ03313.B7_P23) and one Sitka spruce *3CAR*-like EST (WS02910_I02) would make this sequence a relevant target for future transcription factor binding site analysis. In addition, several *cis*-acting elements previously identified in other plant species to be involved in responses to gibberellin (GARE, TAACAGA; P-box; GCCTTTGAGT), auxin (ARF, TGTCTC; TGA-element, AACGAC; AUX28, ATTTATATAAAT), ethylene (ERE, AWTTCAA), and abiotic stresses (HSE, AAAAAATTC; MBS, TAACTG; LTR, CCGAAA) were found in the upstream regions of *3CAR* and *CYP720B4*.

Identification and Distribution of High- and Low-Complexity Repeats in PGB02 and PGB04

Since repeat regions may offer a particular challenge for genome sequence assembly in conifers, it is important to accurately detect and mask high- and low-complexity repeats. A comparison of the PGB02 and PGB04 sequences with the genome sequences of *Arabidopsis*, rice, poplar, and grapevine (AGI, 2000; IRGSP, 2005; Jaillon et al., 2007; Tuskan et al., 2006) identified 3.7% of PGB02 and 3.0% of PGB04 with similarity (E-value < 10⁻⁵) to repetitive regions found in these angiosperms <http://www.phytozome.net> (Table 2). Using RepeatMasker (Chen, 2004) we found that high-complexity repeats contribute to 21.9% and 17.6% of the sequence of PGB02 and PGB04, respectively (Table 2). We identified regions with similarity to RNA-based retroelements, predominantly Ty1/Copia, and Gypsy/DIRS1 (long terminal repeat (LTR) element class) and a few segments of L1/CIN4 (long interspersed element (LINE) class) (Figure 1). In contrast to the large number of retroelement-based TEs, we found few regions (0.7% of total sequence of PGB02 and PGB04) with similarity to DNA-based transposons (EnSpm, Helitron, MuDR and hAT). Although PGB02 and PGB04 represent only a small fraction of the spruce gDNA, the identification of these DNA-based TEs is important as this is the first report of these elements in a gymnosperm.

Table 2. High-complexity repeats in the white spruce gDNA of PGB02 and PGB04.

BAC	Repetitive sequences with similarities in angiosperms ¹	TEs detected with RepeatMasker ²	Total repeat content ³	Similarity to ESP (%)
PGB02	3.7%	21.9%	36.0%	14.7%
PGB04	3.0%	17.6%	41.6%	17.1%

¹Portion of the white spruce gDNA sequences of PGB02 and PGB04 with similarity to repeat regions identified in the genomes of *Arabidopsis*, rice, poplar and grapevine (cut-off E-value <10⁻⁵); this excludes the coding regions of 3CAR and CYP720B4.

²Percentage of PGB02 and PGB04 sequences consisting of TEs as detected by the RepeatMasker using the viridiplante section of the RepBase Update.

³Percentage of PGB02 and PGB04 sequences consisting of high complexity repeats as detected by pairwise comparisons of the two gDNA sequences.

⁴Fraction of the PGB02 and PGB04 sequences with similarity (at least 80–90% nucleotide sequence identity) to white spruce ESTs; this excludes the coding regions of 3CAR and CYP720B4; no EST hits were detected outside repeat regions.

While LTR retrotransposons have been reported in spruce with a high copy number, it is not known if members of the Ty1/Copia or Gypsy/DIRS1 families are active in spruce (L'Homme et al., 2000). Presence of retrotransposons in the transcriptome and sequence conservation indicates that they are active. A BLAST search of the repetitive regions of PGB02 and PGB04 against EST databases (plant genome database, <http://plantgdb.org/>) yielded significant hits with ESTs from white spruce, Sitka spruce, interior spruce, and Norway spruce as well as with pine species (Table 2). Pairwise comparison of the gDNA sequences of PGB02 and PGB04 revealed substantial sequence conservation within the repeat regions (Table 2). All regions with similarity to TEs reside in large, often continuous sections with high homology (average identity 86% over up to 3,000 bp) on PGB02 and PGB04 (Figure 1).

Screening for homologous regions between and within PGB02 and PGB04 also identified several previously undetected repeated elements, one of which represents a putative conifer specific repeat element (CSRE), which appears to have locally multiplied in PGB04 (Figure 1). A white spruce transcript with 91% identity to this CSRE is also present in the EST database (accession number WS0339.C21_N21). The occurrence of high-complexity repeats in the BAC clones is estimated at 36.0% in PGB02 and 41.6% in PGB04, values which are substantially higher than those found in the fully sequenced genomes of *Arabidopsis* (10%) and poplar (12.6%), and similar to the genomes of rice (35%) and grapevine (38.8%) (AGI, 2000; IRGSP, 2005; Jaillon et al., 2007; Tuskan et al., 2006) (Table 2).

CONCLUSION

We report the first sequence assembly and annotation of large segments of gDNA from a conifer. We also demonstrate that genomic BAC clones for specific members of large conifer defense gene families can be isolated in a very efficient and targeted fashion. This work provides important new information about the structure and content of conifer genome regions associated with the *3CAR* and *CYP720B4* genes in white spruce. Features of low gene density, high content of repetitive sequence regions, and richness of TEs identified in this work are likely characteristic of conifer genomes in general.

This work also provides relevant information for future efforts to sequence a conifer genome. Cost-efficiency is a critical factor in genome sequencing and is a function of sequencing chemistry, the complexity of the region being sequenced, and the quality of the assembly. Our simulation of the effect of BAC sequencing depth on assembly coverage showed that increasing the sequencing depth beyond 5–7 × coverage results in only a marginal improvement of the sequence assembly. The future sequencing of a conifer genome will likely use a combination of ultra-high throughput methods in combination with sequencing of BAC clones to anchor the high throughput reads. The bi-directional Sanger sequencing used in this study generated high quality sequences of more than 1,000 bp average length which were critical for the assembly of full-length BAC clones. Low-quality reads resulting in poor sequence coverage occurred in regions of complex and simple repeats, which may also provide challenges for ultra high-throughput sequencing.

KEYWORDS

- **Angiosperms**
- **BAC clones**
- **Conifers**
- **White spruce**

AUTHORS' CONTRIBUTIONS

Björn Hamberger, Dawn Hall, Mack Yuen, Christopher I. Keeling, and Jörg Bohlmann designed experiments, conducted the data analysis, and interpretation of data and

results. Björn Hamberger, Dawn Hall, Claire Oddy, and Britta Hamberger carried out experiments. Jörg Bohlmann and Kermit Ritland conceived of the overall study. Carol Ritland participated in the design of the study and coordination. Björn Hamberger, Dawn Hall, Mack Yuen, and Jörg Bohlmann wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Dr. Alvin Yanchuk, Dr. John King, and Mr. Barry Jaquish from the British Columbia Ministry of Forests and Range for access to white spruce trees and generous support of this project. We thank Ms. Karen Reid (UBC) for excellent support with project and laboratory management, BioS&T (Montreal) for producing the PG29 BAC library and clone identification, and the Michael Smith Genome Sciences Centre (Vancouver) for sequencing and bioinformatics support. The work described in this chapter was supported with funding from Genome British Columbia and Genome Canada for the Treenomix Conifer Forest Health project <http://www.treenomix.ca> (to Jörg Bohlmann and Kermit Ritland) and the Natural Sciences and Engineering Research Council of Canada (EWR Steacie Memorial Fellowship and Discovery Grant to Jörg Bohlmann, and a Postdoctoral Fellowship to Dawn Hall). Jörg Bohlmann is supported in part by the University of British Columbia Distinguished University Scholars program.

Chapter 3

Expressed Sequence Tag Analysis in *Cycas*

Eric D. Brenner, Dennis W. Stevenson, Richard W. McCombie,
Manpreet S. Katari, Stephen A. Rudd, Klaus F.X. Mayer,
Peter M. Palenchar, Suzan J. Runko, Richard W. Twigg, Guangwei Dai,
Rob A. Martienssen, Phillip N. Benfey, and Gloria M. Coruzzi

INTRODUCTION

Cycads are ancient seed plants (living fossils) with origins in the Paleozoic. Cycads are sometimes considered a “missing link” as they exhibit characteristics intermediate between vascular non-seed plants and the more derived seed plants. Cycads have also been implicated as the source of “Guam’s dementia”, possibly due to the production of S(+)-beta-methyl-alpha, beta-diaminopropionic acid (BMAA), which is an agonist of animal glutamate receptors (GLRs).

A total of 4,200 expressed sequence tags (ESTs) were created from *Cycas rumphii* and clustered into 2,458 contigs, of which 1,764 had low-stringency BLAST similarity to other plant genes. Among those cycad contigs with similarity to plant genes, 1,718 cycad “hits” are to angiosperms, 1,310 match genes in gymnosperms and 734 match lower (non-seed) plants. Forty-six contigs were found that matched only genes in lower plants and gymnosperms. Upon obtaining the complete sequence from the clones of 37/46 contigs, 14 still matched only gymnosperms. Among those cycad contigs common to higher plants, ESTs were discovered that correspond to those involved in development and signaling in present-day flowering plants. We purified a cycad EST for a *GLR*-like gene, as well as ESTs potentially involved in the synthesis of the GLR agonist BMAA.

Analysis of cycad ESTs has uncovered conserved and potentially novel genes. Furthermore, the presence of a GLR agonist, as well as a *GLR*-like gene in cycads, supports the hypothesis that such neuroactive plant products are not merely herbivore deterrents but may also serve a role in plant signaling.

The Cycadales (cycads) are the most primitive living seed plants and have endured over 270–280 million years since their origins in the Lower Permian (Gao and Thomas, 1989; Mamay, 1969). Cycads have a fern or palm-like appearance, largely due to their pinnately compound leaves (Figure 1a, b). Unlike ferns or palms, however, cycads belong to the gymnosperms, or non-flowering seed plants. Of the four orders that comprise the gymnosperms, the Cycadales are considered to be the most ancestral compared to Ginkgoales, Gnetales, and Coniferales (Figure 2) (Nixon et al., 1994; Soltis et al., 2002). Cycads (non-flowering seed plants) exhibit a number of characteristics that reflect their evolutionary position between ferns (non-seed plants) and angiosperms (flowering seed plants). Such characteristics include pollen tubes, which release motile

sperm before fertilization; dichotomous branching (versus axillary branching in higher plants); and ovules, which contain a large, free-nuclear megagametophytic stage, that are borne on the margins of leaf-like megasporophylls (Chamberlain, 1919; Loconte and Stevenson, 1990; Norstog and Nicholls, 1997). These characteristics, among others, place cycads at a key node in plant evolution.

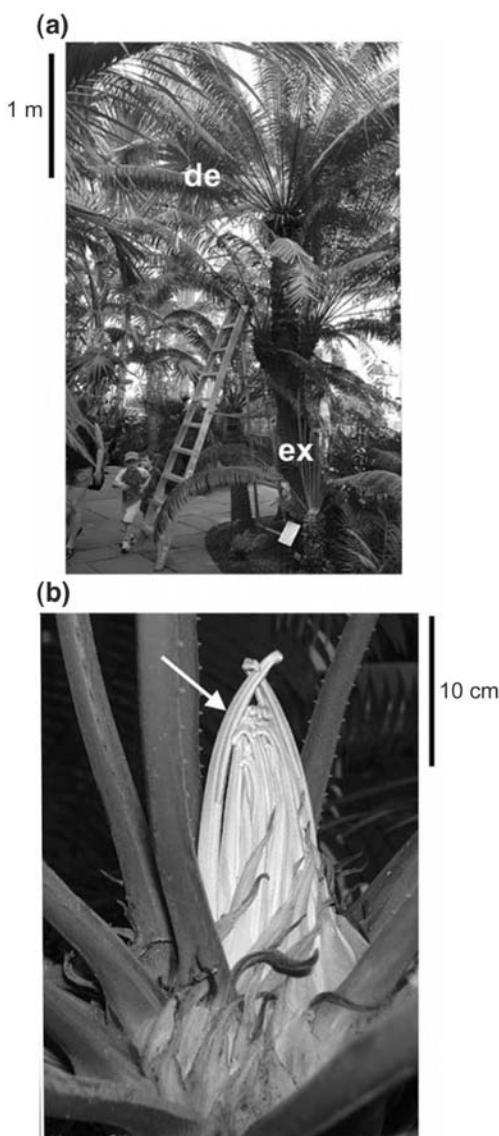


Figure 1. *Cycas rumphii* used for cDNA library construction. (a) Mature cycad trunk with developed (de) leaves and young, expanding (ex) leaves. (b) Young emergent leaves (arrow) at the crown, which were used to generate a cDNA library database.

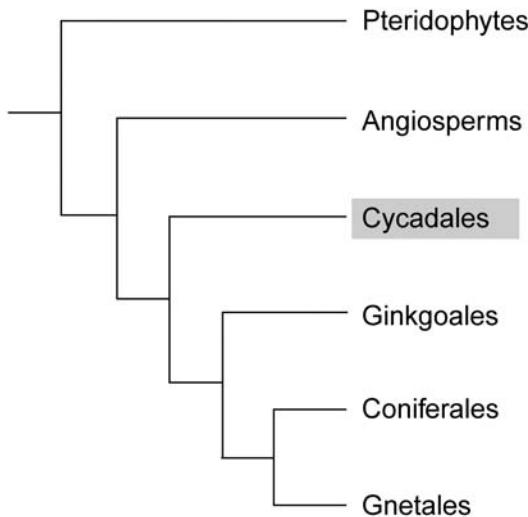


Figure 2. Cycads are the sister group to the seed plants. A phylogenetic tree shows that cycads (highlighted) are the least derived of the seed plants. Cycads are believed to be the oldest extant seed plants.

In addition to their evolutionary importance, cycads have also been studied in the field of medicine, because they produce neurotoxic compounds. In particular, cycads produce a secondary compound, BMAA (S(+)-beta-methyl-alpha, beta-diaminopropionic acid), which has been implicated as the possible cause of Guam's dementia (Vega and Bell, 1967). This disorder occurs among the indigenous Chomorro people, who ate cycads as food, and now suffer from Alzheimer's and Parkinson's dementia (Kurland, 1972; Spencer et al., 1987; Whiting, 1963). BMAA production is unique to cycads, where it has been used as a monophyletic character in plant classification (Locoste and Stevenson, 1990). It is present in both seeds and leaves of all genera of the Cycadaceae (Charlton et al., 1992). BMAA is neurotoxic in mammals (Seawright et al., 1999; Spencer et al., 1987) because of its excitotoxic action as an agonist of GLRs (Brownson et al., 2002). The discovery of GLR-like genes in *Arabidopsis* suggests that plant-derived GLR agonists, as well as acting as potential deterrents to herbivores, might also operate in signaling during plant growth and development, by interacting with native plant GLRs (Lam et al., 1998). In partial support of this hypothesis, BMAA was shown to affect the development of *Arabidopsis* and consequently was used in a pharmacologically-based genetic screen to isolate mutants in a putative GLR pathway in *Arabidopsis* (Brenner, 2000).

Despite the importance of cycads in the study of plant evolution, and their role in neurological disorders in humans, nothing is known about the genes responsible for these traits—primarily because cycads are recalcitrant to genetic analysis. Unlike, genetically tractable plants such as tomato, maize, and *Arabidopsis*, cycads are dioecious (male and female organs on separate plants), produce a limited number of seeds and take up to 30 years to become reproductive. Furthermore, cycad genomes are large

(20,000–30,000 million base-pairs (Mbp)) (Murray, 1998; Ohri and Khoshoo, 1986) compared to *Arabidopsis* (125 Mbp) (The *Arabidopsis* Genome Initiative, 2000). Consequently, cycads have remained outside the realm of both traditional genetic studies and modern genome-sequencing initiatives. Fortunately, recent advances in plant genomics (Daly et al., 2001; Mayer and Mewes, 2002), provide new tools to study genetically complex species such as cycads. In particular, the availability of the complete, annotated sequence of two angiosperm genomes—the dicot *Arabidopsis thaliana* (Martienssen and McCombie, 2001; The *Arabidopsis* Genome Initiative, 2000) and the monocot rice (*Oryza sativa*) (Goff et al., 2002; Yu et al., 2002)—now makes it possible to study the genomes of evolutionarily important plants by comparing the expressed genes of cycads (ESTs) to the complete genomes of higher plants.

To begin a survey of expressed genes of cycads, the genus *Cycas* was chosen for EST analysis because *Cycas* is at the basal node—that is, the sister taxon to the rest of the Cycadales (Crane, 1985; Stevenson, 1990; Treutlein and Wink, 2002). Furthermore, the species *Cycas rumphii* Miq. was selected for this analysis as it is suspected to be the dietary cause of Guam's dementia. It has been established that in *C. rumphii*, from which the EST library was made, BMAA levels are nearly 0.1 mg/g tissue (Duncan et al., 1989). Because of its evolutionary position as a key node within the plant kingdom, as well as its medicinal significance to humans, *Cycas* is ideally suited for genomic prospecting (Brenner et al., 2003).

Here, we describe the construction of a cycad EST database from RNA of young *C. rumphii* leaves. Using this database, our comparison revealed conserved genes, including those involved in development and signaling in present-day flowering plants. Our analysis defined a set of cycad clones that have no similarity to any known angiosperm genes, but possess similarity only to genes of other gymnosperms. Furthermore, as a first step to understanding the function of neurotoxins produced in cycads, we defined a number of candidate genes that encode putative enzymes involved in the biosynthesis of BMAA, as well as a cycad GLR-like gene, the suspected target of BMAA action in animal brains. These cDNA tools will be useful to test whether BMAA, which has been postulated to serve as an herbivore deterrent (Norstog and Nicholls, 1997), also acts to regulate GLR function in plants.

MATERIALS AND METHODS

Tissue Collection and Library Construction and DNA Purification

Newly emerged immature leaves from the crown of a *C. rumphii* tree, accession 808/59 A, were collected from the New York Botanical Garden Conservatory. Leaves collected ranged from 5 to 30 cm in length. Tissue was frozen in liquid nitrogen. RNA was extracted from pulverized, frozen tissue in a mortar and pestle with the RNeasy maxi kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Purified *Cycas* RNA was precipitated in 2 M LiCl, washed twice with 70% ethanol, and resuspended in 50 µl water. Poly(A) RNA was subsequently purified from total RNA with the Oligotex Maxi kit (Qiagen). A cDNA library was constructed using the Lambda ZAP-CMV cDNA synthesis kit (Stratagene, La Jolla, CA) using 10 µg poly(A) RNA. Before cloning, cDNA was size fractionated over a Sepharose CL-6b column. The

first five fractions containing a total of around 100 ng cDNA were collected, pooled and precipitated in 70% ethanol/0.3 M sodium acetate and resuspended in 3.5 µl water. cDNA (0.5 µl) was then directionally sub-cloned into the vector at the EcoRI and XhoI sites.

DNA was collected from unemerged *C. rumphii* sporophylls using the DNeasy purification kit (Qiagen).

EST Sequencing

Plasmid DNA was collected as described in the manual (Stratagene) catalog number 200450 in the *in vivo* mass excision section. Sequence analysis was performed at Cold Spring Harbor Laboratory using an ABI 3700 capillary sequencer (Applied Biosystems, Foster City, CA) for separation and nucleotide detection. Reactions were performed using a 1/16 Big Dye Terminator. Sequencing was performed with either the -21 M13 forward and/or reverse primer.

EST Clustering and Assignment into Functional Categories

The EST sequences were clustered and assembled using the HarvESTer application (Biomax informatics, Martinsried, Germany). The default HarvESTer settings were optimized to screen for vector against the UniVec nonredundant database of vector and polylinker sequences (VecScreen, 2010). The Hashed Position Tree (HPT) clustering used a similarity link threshold of 0.7 and a maximum distance of six steps was required to define a cluster from the similarity network, thus encouraging the separation of likely paralogs. Cluster consensus sequences and concomitant alignments were derived from the HPT clusters using the CAP3 application with default settings. The HarvESTer assemblies and coordinate alignments were imported into the Sputnik EST and cluster analysis application (Rudd et al., 2003).

Peptide Extraction

The BLASTX (Altschul et al., 1990) was performed against a nonredundant protein database for each of the cluster consensus sequences. Likely coding sequences were derived for each cluster consensus sequence by parsing the best BLASTX match and filtering the results using the arbitrary expect value <1e-10. Dicodon usage frequencies and probabilities were extracted using tools from the ESTate package (Slater, 2000). A peptide sequence was predicted for each of the cluster consensus sequences using the Framefinder application from the ESTate package with the cycad-specific codon usage statistics. Framefinder was run using the default parameters. The derived peptide sequences were used as the basic scaffold for peptide-based annotation in Sputnik.

Sequence Annotation

Sequence annotation on each of the cycad cluster consensus sequences and derived peptides were performed within the Sputnik application. Results were assessed for possible contamination by searching for homology to the *Escherichia coli* and human genomes and were scored for homology to a wide range of non-coding RNAs and plant chloroplast and mitochondrial genomes. Similarity searches were performed

using the BLAST application (Altschul, 1990) and results were filtered using the expectation value < 1e-10. Functional assignment was performed on both cluster consensus sequence and the peptide sequence. Assignments were made using BLASTX and BLASTP respectively against the MIPS catalog of functionally assigned proteins (funcat) (Ewing and Green, 1998; Ewing et al., 1998); tentative functional assignments were filtered using the expectation value < 1e-10.

Categorization of Cycad Contig

All cycad contigs sequences were aligned against the PlantEST database using TblastX (Altschul et al., 1990) and BlastX against the NR(aa) database. The PlantEST database was created by downloading all plant ESTs in GenBank and assembling them using Phrap (Ewing and Green, 1998; Ewing et al., 1998). Todd Wood from Clemson University provided the PERL script that creates the PlantEST databases as described above. The NR(aa) database is a nonredundant database of protein sequences from GenBank.

Determination of Gymnosperm-Specific Genes

All available plant ESTs were downloaded from GenBank and separated into three datasets consisting of angiosperms (monocots and dicots), gymnosperms, or lower plants (ferns, mosses and algae). Downloaded ESTs were assembled using Phrap (Ewing and Green, 1998; Ewing et al., 1998). All matches with an expect value < 1e-5 were considered significant.

DISCUSSION

Cycads can be Regarded as Living Fossils

Extant genera, such as *Cycas*, have changed little in morphology from their extinct relatives, such as *Crossozamia*, which existed during the Permian (Gao and Thomas, 1989; Mamay, 1969). The study of cycads has proved to be useful in reconstructing plant evolution, in particular in understanding the rise of important plant structural innovations such as the evolution of seeds (Foster and Gifford, 1974). Cycads also produce a variety of neuroactive compounds, some of which are suspected to be the source of Guam's dementia (Kurland, 1972; Khabazian et al., 2002). However, despite their scientific importance in plant biology and medicine, virtually nothing is known regarding gene expression, development and signaling in the Cycadales. As a first step in this direction, a cDNA library was made from young, developing *C. rumphii* leaves to produce a cycad EST database.

A Cycad EST Database: A Foundation to Study the Evolution of Early Seed Plants

One advantage of a genomics approach is that it provides rapid access to genes important for evolutionary studies. The more traditional homology-based gene-cloning approach is limited by tedious gene-by-gene purification. It is also limited in that it may miss related genes if the degeneracy is too great or if nonconserved regions of the protein are chosen during primer design. Finally, the targeted gene approach can never be used to discover new genes.

Sequence Analysis of Contigs with BLAST Similarity to Gymnosperms but not Angiosperms

An EST project in *Pinus taeda* (loblolly pine) sampled 59,797 transcripts from wood-forming tissues (Kirst et al., 2003). In this analysis, 66 *P. taeda* contigs showed BLAST similarity at low stringency only to other gymnosperms. Similarly, in our analysis, we found 46 cycad contigs that only matched gymnosperms (including *P. taeda*) and/or lower plant ESTs, but were not found in the genomes of higher plants or non-plants. Complete sequencing of 37 of these cycad cDNA clones showed that 14 clones, ranging in length from 586 to 1,899 bp, were still found only in other gymnosperms. Having no homology to the completely sequenced genomes of two different angiosperm species—*Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000) (a dicot) and rice (Goff et al., 2002; Yu et al., 2002) (a monocot)—suggests that these 14 genes are found only in gymnosperms or lower plants, in which genomic studies have only just begun. However, because ESTs as well as contigs usually represent only a portion of the full-length gene sequence, these results are preliminary. For instance, in *P. taeda*, larger contigs have a higher BLAST match rate to other plant genes than do shorter contigs (Kirst et al., 2003). Thus, these preliminary results of clade specificity are tenuous and presumably will change as more ESTs, as well as full-length gene sequences, from cycads and other species are generated in the future.

Genes with Potential Developmental Roles in Cycads

As in higher plants, cycad leaves are derived from the shoot apical meristem (SAM) (Stevenson, 1981). In *Cycas* leaflet primordia, meristematic growth ceases at the apex, while proceeding basipetally where it becomes localized to the leaflet margins (Stevenson, 1981). The presence of these marginal meristems may explain why a surprising number of developmental genes were identified in a relatively small number of ESTs from young cycad leaves (Table 2).

A gene with identity to the *YABBY* gene family was among the cycad ESTs. The *YABBY* genes encode transcription factors expressed on the abaxial side of all lateral organs that promote abaxial cell fate (Eshed et al., 1999). In *Arabidopsis*, mutations in the *YABBY* gene INO (INNER-NO-OUTER), lead to the loss of the outer integument (Villanueva et al., 1999) reminiscent of gymnosperm (and cycad) unitegmy (the presence of a single integument). Unitegmy is considered to be the ancestral condition in seed plants (Foster and Gifford, 1974; Norstog and Nicholls, 1997). An analysis of *YABBY* gene expression in cycads may help to explain the origin of the integument in gymnosperms, and/or possibly the second integument in angiosperms. One cycad EST from the library has highest similarity to COP9. The COP9 encodes a subunit of the COP9 signalosome complex, which controls multiple signaling pathways that regulate development in all eukaryotes (Chamovitz and Glickman, 2002; Hellmann and Estelle, 2002). In *Arabidopsis*, the COP9 mutant is constitutively photomorphogenic in dark-grown seedlings (Wei and Deng, 1992). Some gymnosperms, (in particular the Coniferales) are constitutively photomorphogenic when grown in the dark (Bogdanovic, 1973; Peer et al., 1996). As yet, the phenotype of dark-grown cycad seedlings has not been fully evaluated. The discovery of a gene encoding a putative

subunit of the COP9 complex in cycads could be a first step to define the ancestral, developmental role of the signalosome in gymnosperms, particularly with regard to its role in photomorphogenesis.

Another gene potentially involved in cycad development has highest similarity to the *CONSTANS* gene family, which are regulators of flowering time that follow internal and external (environmental) inputs in *Arabidopsis* (Suarez-Lopez et al., 2001). Because cycads predate the evolution of flowers, it would be of interest to determine if *CONSTANS* genes in cycads temporally regulate sporophyll and cone induction, which typically follows a yearly cycle (Chamberlain, 1919; Norstog and Nicholls, 1997).

A Cycad *GLR*-Like Gene Expressed in Tissue Producing the *GLR* Agonist BMAA

An unexpected finding of the *Arabidopsis* EST genome project was the discovery of *GLR*-like genes, or “neural” receptor genes, in plants (Lam et al., 1998). In *Arabidopsis*, the *GLR*-like gene family comprises 20 members (Lacombe et al., 2001). Pharmacological evidence has linked *Arabidopsis GLRs* to light and/or growth signaling pathways (Brenner et al., 2000; Lam et al., 1998). Supplying exogenous BMAA to growing *Arabidopsis* seedlings was shown to block light-induced hypocotyl shortening and cotyledon expansion (Brenner et al., 2000). Because BMAA has such profound effects on *Arabidopsis* development, we have previously proposed that BMAA, or glutamate, the natural agonist of *GLRs* in humans, plays a physiological role in *Arabidopsis* (Brenner et al., 2000; Lam et al., 1998). Continuing genetic studies in *Arabidopsis* aim to identify the endogenous components of the BMAA-targeted pathway in plants (Brenner et al., 2000).

Cycads produce BMAA (Spencer et al., 1987; Vega and Bell, 1967). One EST uncovered in the *C. rumphii* leaf cDNA library has a high degree of similarity to plant *GLR* genes (Table 2). This discovery is intriguing, because it suggests that BMAA might be interacting with native *GLR* gene products in cycads. To further investigate the relationship between cycad *GLR* genes and BMAA, we sought to identify cycad genes potentially involved in BMAA synthesis.

From the structure of BMAA, we hypothesized that cycads produce BMAA in a simple two-step pathway, beginning with a β -substituted alanine. To enhance the probability of finding genes involved in BMAA synthesis, we made our cDNA library from tissues that produce relatively large quantities of BMAA (nearly 0.1 mg/g tissue) (Duncan et al., 1989). According to Ohlrogge and Benning, there is a 95% chance of finding the gene for a specified enzyme when it is expressed at 0.1% mRNA/protein by sampling only 3,000 ESTs from an unnormalized library (Ohlrogge and Benning, 2000). Considering the prevalence of BMAA in *Cycas*, it is not surprising that we discovered cognate genes for the predicted enzymes for this BMAA biosynthetic pathway in the cycad EST database (Figure 5, Table 2). Future biochemical and molecular studies will determine if these genes play a part in BMAA synthesis.

The discovery of *GLR*-like genes in *C. rumphii* raises the intriguing possibility that endogenous BMAA may interact with native cycad *GLRs* as a regulatory molecule. Future studies aim to understand the role of *GLRs* in plants, as well as the role of BMAA in herbivore defense versus endogenous signaling. The production of additional ESTs from cycads will increase the variety of genes available for study, so that a detailed expression profile can be evaluated during cycad development. Complementation studies of these genes in orthologous *Arabidopsis* mutations will help define their roles in cycads. This combined approach to studying cycad gene structure and function will help reveal molecular changes in genes involved in signaling, metabolic and developmental pathways that led to the rise of the seed plants.

RESULTS

Construction of a cDNA Library from *Cycas rumphii*

At maturity, *C. rumphii* leaves can reach up to 3 meters in length (Figure 1a). The tissue used in this study consisted of 10 to 40 cm of the immature leaf terminus protruding from the crown collected shortly after emergence (Figure 1b). Immature leaves consist of a petiole, a central rachis and circinate leaflets composed of both expanding and meristematic cells (Stevenson, 1981). RNA extracted from this tissue was used to construct a cDNA library from *C. rumphii*. Size fractionation was used to enrich for full-length cDNAs during library construction. It was determined that 53% of the cDNA clones were over 500 bp long. From this cDNA library, 4,210 sequences reads (ESTs) were generated. The majority of these reads (3,917) were generated from the 5' end of the cDNA; however, small subgroups (293) were sequenced from the 3' end. Cluster analysis performed at the Munich Information Center for Protein Sequences (MIPS) of the entire EST dataset produced a UniGene set of 2,458 contigs consisting of 1,917 singlettons and 541 assemblies. Of the clustered ESTs, the longest contig was 1,836 bp. The entire UniGene set can be viewed on the MIPS Sputnik website (*Sputnik Cycas rumphii*), which features sequence annotations and peptide sequence predictions. At the MIPS Sputnik site there are links to download the complete cycad sequences as an EST fasta file, a cluster fasta file or as the derived peptide fasta file.

Classification of *C. Rumphii* ESTs by Functional Categories

Each contig from the database was automatically assigned to a functional category on the basis of its top match against the complete genomic sequence of *Saccharomyces cerevisiae* and *A. thaliana* databases using BLASTP. A non-stringent expects value (E-value) of <1e-10 was chosen as the threshold. The pie chart in Figure 3 illustrates the relative fraction that each functional category comprises within the entire UniGene set. The four largest predominant categories of cycad ESTs according to this functional categorization are: “cellular organization” (22%), “metabolism” (10%), “unclassified proteins” (10%), and “cell growth, cell division/DNA synthesis” (9%).

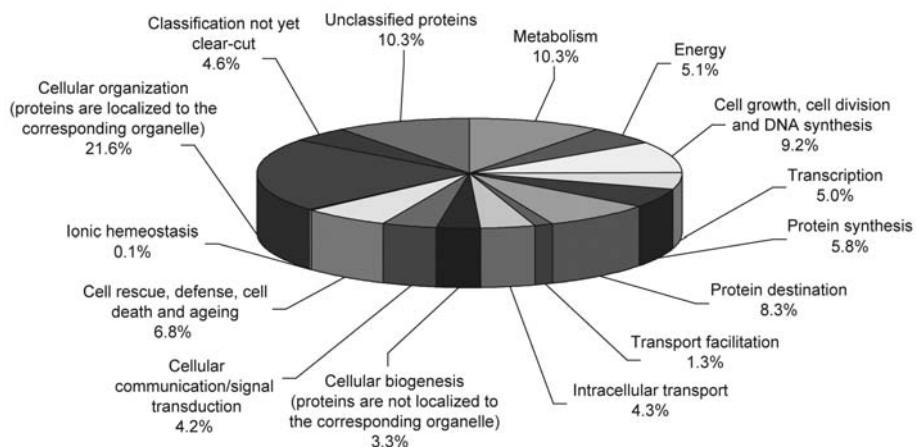


Figure 3. Functional gene categories of cycad ESTs. Clustered cycad ESTs were assigned to a functional category based on top BLASTP similarity scores. An expect value (E-value) of $> 1e-10$ was chosen as the cut-off threshold. The analysis was performed at the Munich Information Center for Protein Sequences.

Cycad Contig Matches to Genes in Angiosperms, Gymnosperms, and Lower Plants

Using TBLASTX, a comparison was made between the *C. rumphii* UniGene set versus all available ESTs from GenBank and predicted *Arabidopsis* genes from The *Arabidopsis* Information Resource (TAIR). Both EST and predicted genes were grouped into three subcategories: angiosperms, gymnosperms, and lower plants. The angiosperm database encompasses all annotated rice and *Arabidopsis* genes identified from their respective genomic sequences, as well as all higher plant ESTs. The gymnosperm database contains ESTs from all gymnosperms, the majority of which came from the *Pinus taeda* EST sequencing project (Kirst et al., 2001; Whetten et al., 2001). The lower plant databases included genes from all remaining plant ESTs including ferns, fern allies, bryophytes and algae available in GenBank. The angiosperm subgroup consisted of 84.5%, the gymnosperms 6.5% and lower plants 9.0% of the total genes used in this analysis.

The Venn diagram shown in Figure 4 displays the total number of cycad contigs shared between one or more of the plant gene datasets at very low BLAST stringency values (expect $< 1e-5$). The majority of cycad contigs (1,764/2,458) have counterparts in other plants, leaving 694 with no match to other plant genes. As one would expect, most *Cycas* hits (1,718) are to angiosperms, because of the predominance of angiosperm accessions in GenBank. Many of the cycad matches to angiosperms also match gymnosperms and/or lower plants (1,416). There are 1,310 cycad contigs that match gymnosperm genes and 734 that match genes from lower plants.

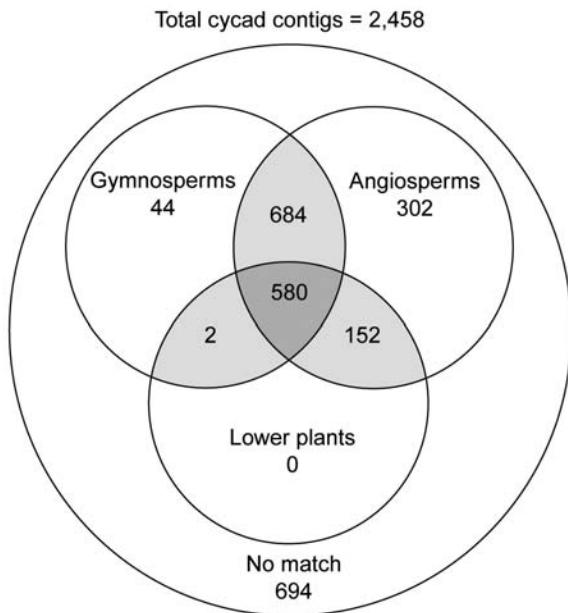


Figure 4. A Venn diagram reveals shared gene sets between cycad contigs versus lower plants, gymnosperms and/or angiosperms. BLASTX (cut-off E value > 1e-5) was used to compare the cycad contigs against all angiosperm ESTs and annotated genes from the full *Arabidopsis* and rice genome sequence from GenBank. Genes that do not have a match to angiosperm genes were then compared to available ESTs from all gymnosperms or lower-plant ESTs available in GenBank. Genes that are common to cycads and more than one group are shown in the intersecting (shaded) regions.

Full-Length Sequencing of Cycad Clones that Match Only Gymnosperm Genes

As shown in Figure 4, 44 *Cycas* ESTs specifically match only genes in the gymnosperm subgroup. Two additional *Cycas* ESTs match genes from gymnosperms and lower plants, but not angiosperms. To further analyze these 46 contigs that match only gymnosperms and/or lower plants, we next sequenced these *Cycas* cDNAs in their entirety to determine whether this “gymnosperm/lower plant” specific grouping held up when the remaining portions of the cDNA were sequenced. Because ESTs, even when clustered into contigs, usually represent only a portion of the actual gene (particularly for genes poorly represented in the library) 37 of the 46 *Cycas* cDNAs were sequenced in their entirety (the remaining nine clones were not successfully recovered for sequencing), and this sequence can be downloaded from the Internet (Index of full-length sequences). Of these 37 fully sequenced cDNAs, 14 clones still showed no similarity to any known angiosperm genes, even at this low stringency cut-off. The insert size for each clone ranges from 586 bp to 1,899 bp, with predicted open reading frames (ORFs) varying from 69 to 527 residues (Table 1). None of these 14 *Cycas* cDNA clones is homologous to any known genes outside the plant kingdom, although Interpro analysis identified a small number of conserved motifs, which are listed in Table 1. To confirm that these genes were indeed derived from *C. rumphii*,

gene-specific primers designed to each of the 14 genes were able to amplify a fragment from genomic DNA isolated from a different *C. rumphii* specimen and different tissue (sporophyll) from the source tissue of the cDNA library (data not shown). This distinct *C. rumphii* specimen was cultivated in a geographically separate location (Florida) from the cDNA source *C. rumphii* specimen used for cDNA library construction (New York).

Table 1. Fully sequenced cycad clones from contigs that match only genes in gymnosperms.

Contig	GenBank accession number	Transcript length (bp)	Peptide (residues)	InterPro result
gy79c04_704	CB090702	586	72	No matches found
gy78g12_244	CB090673	627	84	No matches found
te82h02_741	CB093328	647	107	No matches found
he95e08_721	CB091708	651	114	No matches found
hf04g07_288	CB092366	684	141	ASP_RICH (unintegrated)
hk42a07_743	CB093061	790	142	No matches found
gp23c01_369	CB089407	791	189	No matches found
gp26f08_297	CB089628	827	69	No matches found
gy82g05_181	CB090964	840	118	No matches found
he92f06_688	CB091462	935	170	No matches found
gy81ell_544	CB090877	948	211	ASP_RICH (unintegrated)
he97c12_740	CB091858	965	140	No matches found
gp32b03_724	CB089926	1311	335	Peptidoglycan-binding LysM
te83a03_729	CB093338	1899	527	No matches found
Average		893	173	

All available ESTs and annotated genes from GenBank were divided into three categories (angiosperms, gymnosperms and lower plants) and compared with the *C. rumphi* UniGene set. Forty-six cycad ESTs that had no similarity to angiosperm genes, but matched gymnosperm and lower plant genes, were fully sequenced, of which 14 clones (listed) still have no similarity to angiosperms. To confirm that these genes were of cycad origin, all 14 were successfully amplified from the DNA of a *C. rumphii* specimen other than the one used to construct the cDNA library. The transcript length, as well as the predicted translation product size, is shown. Interpro analysis identified conserved motifs in three of these cycad ESTs as shown.

Cycad Genes Similar to Developmental Regulators

A survey of the cycad EST dataset reveals a surprisingly large number of genes with highest similarity (BLASTP score < e-5) to genes with defined roles in growth and development in angiosperms (Table 2). Some of these *Cycas* genes have similarity to *Arabidopsis* transcription factors, including *CONSTANS* (Putterill et al., 1995; Suarez-Lopez et al., 2001), two distinct *homeobox* genes (Chan et al., 1995) and a *YABBY* gene (Eshed et al., 1999; 2001). Other cycad ESTs have similarity to other regulators of *Arabidopsis* development, including *ARGONAUT* (Bohmert et al., 1998) and *COP9* (Chamovitz and Glickman, 2002; Schwechheimer and Deng, 2001).

Table 2. Genes in *Cycas rumphii* with potential roles in signaling, development and biosynthesis of BMAA.

	GenBank accession number	Subject description	E-value	%ID	Score
Development	CB092871	Argonaute-like protein I (AGO I) <i>Arabidopsis thaliana</i>	8.00e-10	0.85	34
	CB090033	YABBY2 <i>A. thaliana</i>	2.00e-36	0.58	151
	CB089539	Multisubunit regulator protein COP9 - spinach COP9 <i>Spinacia oleracea</i>	2.00e-31	0.62	98
	CB092157	CONSTANS B-box zinc finger family protein <i>A. thaliana</i>	1.00e-47	0.48	221
	CB092462	CRHB3 homeoprotein <i>Ceratopteris richardii</i>	3.00e-44	0.70	131
	CB089344	Homeodomain protein HB2 <i>Picea abies</i>	3.00e-29	0.62	117
	CB089945	Photolyase/blue-light receptor PHR2	8.00e-76	0.69	197
	CB091652	Putative glutamate receptor protein GLR3.4b	2.00e-45	0.54	161
	CB093220	Calmodulin-like protein; protein ids At5g44460.1 <i>A. thaliana</i>	3.00e-07	0.58	45
	CB089469	14-3-3 protein <i>Fritillaria cirrhosa</i>	8.00e-38	0.80	94
Signaling	CB091066	Ser/Thr protein kinase isolog; protein ids. supported by cDNAs <i>Arabidopsis</i>	1.00e-10	0.28	185
	CB090652	Ser/Thr specific protein phosphatase 2A B regulatory subunit beta <i>Medicago</i>	4.00e-86	0.94	162
	CB093099	Auxin regulated protein (1M 13) <i>A. thaliana</i>	1.00e-34	0.63	125
	CB089385	Auxin-induced protein IAA9 <i>A. thaliana</i>	8.00e-29	0.55	III
	Biosynthetic enzymes of cycad specific phytochemicals (BMAA)				
	Cysteine synthase	CB089577 Cysteine synthase (O-acetylserine sulfhydrylase)	3.00e-50	0.75	128
	Methyl transferases	CB092214 Plastid cysteine synthase 2 <i>Solanum tuberosum</i>	5.00e-27	0.64	83
		CB091906 Caffeic acid O-methyltransferase II <i>Nicotiana tabacum</i>	3.00e-35	0.56	122
		CB090738 Caffeoyl-CoA 3-O-methyltransferase <i>Oryza sativa</i>	1.00e-37	0.47	188
SAdM metabolism					
Adenosylhomocysteine (S-adenosyl-L-homocysteine hydrolase)	CB091477	Adenosylhomocysteinase <i>Phalaenopsis</i>	1.00e-87	0.84	185
	CB091821	Adenosylhomocysteinase <i>Triticum aestivum</i>	3.00e-78	0.90	156
	CB090818	Adenosylhomocysteinase <i>Medicago sativa</i>	2.00e-18	0.68	66
	CB091682	S-adenosylmethionine synthetase <i>Brassica juncea</i>	4.00e-90	0.94	167
	CB090997	S-adenosylmethionine synthetase (methionine adenosyltransferase) <i>Petunia</i>	1.00e-69	0.94	133
Homocysteine methyltransferase	CB090407	S-adenosyl-L-methionine synthetase <i>Elaeagnus umbellata</i>	1.00e-93	0.88	191
	CB092344	Methionine synthase protein <i>Sorghum bicolor</i>	4.00e-94	0.90	190
	CB091647	5-methyltetrahydropteroylglutamate - homocysteine S-methyltransferase	3.00e-79	0.76	205

C. rumphii ESTs were compared to GenBank with a BLASTP score < e-5. The top match produced from the BLAST search to the cycad EST is listed under subject description.

Cycas Genes with Similarity to *Arabidopsis* Genes Involved in Signaling

A number of genes in our cycad EST library showed similarity to components of signaling pathways found in higher plants (Table 2). These genes include a photolyase blue-light receptor, genes involved in secondary signaling (including those for calmodulin, kinases, and phosphatases), a 14-3-3 protein, and genes involved in phytohormonal responses, including auxin (IAA-9 and IAA-13) pathways as reviewed in Chory and Wu (2001). Surprisingly, a Cycas EST with high similarity to plant *GLR*-like genes was also found (Table 2) (Chiu et al., 2002; Lam et al., 1998). The presence of a *GLR*-like gene in cycads is of particular interest as it relates to BMAA, as described below.

A Predicted Pathway for BMAA Synthesis in Cycas is Supported by EST Analysis

The BMAA, an agonist of mammalian *GLRs*, is a suspect causative agent of neurological disorders (Seawright et al., 1999; Spencer et al., 1987). However, nothing is known about the genes and enzymes involved in the biosynthesis of BMAA. Because the structure of BMAA is similar to other beta-substituted alanines (Warrilow and Hawkesford, 2000, 2002), it is likely that BMAA biosynthesis utilizes phosphoserine, cysteine, o-acetylserine, or cyanoalanine as a beginning substrate. On this basis, a likely BMAA biosynthetic pathway is shown in Figure 5. This would require a two-step reaction initiated with the transfer of NH₃⁺ at the beta-carbon of the substituted alanine (Figure 5a), followed by an addition of CH₃ (Figure 5b) to produce BMAA (Figure 5c). The NH₃⁺ transfer would require a nucleophilic reaction catalyzed by a cysteine synthase-like protein. A preliminary survey of genes in the cycad EST library identified candidate genes for both of these enzymatic steps (Table 2). The cycad leaf EST library contains two ESTs, which each encode a cysteine synthase.

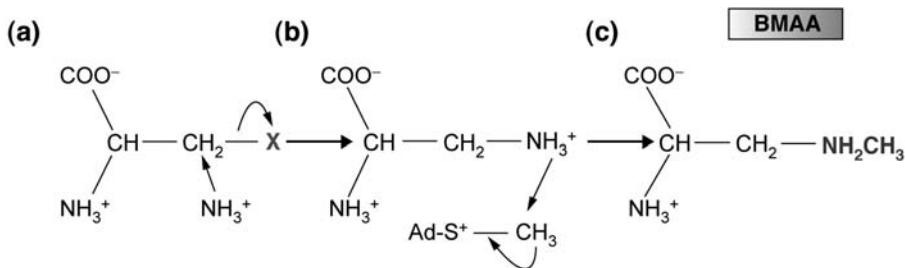


Figure 5. Predicted two-step pathway for the biosynthesis for BMAA in cycads. A postulated route for BMAA biosynthesis supported by cycad EST analysis is shown. In this simple, two-step scheme, BMAA synthesis begins with (a) the transfer of NH₃⁺ to β-substituted alanine, where X = phosphoserine, cysteine, o-acetylserine or cyanoalanine, to form (b) an intermediate. The reaction is catalyzed by a cysteine synthase-like enzyme. This step is followed by transfer of a methyl group from S-adenosylmethionine (Ad-S-CH₃) to the new amine group by a methyltransferase, which would lead to the formation of (c) BMAA. Candidate cycad genes encoding probable cysteine synthase-like enzymes and methyltransferase, as well as S-adenosylmethionine-regenerating enzymes that were identified in the cycad EST collection are listed in Table 2.

To catalyze the second step of BMAA synthesis, the EST library contains two potential methyltransferases (caffeic acid *O*-methyltransferase II and caffeoyl-CoA 3-*O*-methyltransferase). The second step would require a methyl donor, the most likely candidate being S-adenosylmethionine (SAdM). Consumption of SAdM would require the presence of enzymes to regenerate SAdM. A number of cycad ESTs can be implicated in SAdM recycling including: adenosylhomocysteinate, S-adenosylmethionine synthetase and homocysteine methyltransferase. Taken together, the cycad EST library contains candidate genes for all of the enzymes predicted to be present during the biosynthesis of BMAA.

KEYWORDS

- **Contigs**
- **Cycadales**
- **Megasporophylls**
- **Meristematic cells**

ACKNOWLEDGMENTS

We thank Francesco Coelho, Javier Francisco Ortega, and the Montgomery Botanical Center, Florida for providing plant tissue; Dan Chamovitz, and Trevor Stokes for reviewing the manuscript; Vivekanand Balija and Neilay Dedhia for sequence generation and curation; Eduardo de la Torre and Eugene Mueller for helpful discussions; and Alex Clark and Ayelet Levy for technical help. Funding for this work comes from the Plant Genomics Consortium. The Plant Genomics Consortium is made possible by the generosity of the Altria Group, The Mary Flagler Cary Charitable Trust, The Eppley Foundation for Research, The Leon Lowenstein Foundation, The Ambrose Monell Foundation, The Wallace Genetic Foundation, and the National Institutes of Health, grant number GM-32877 to G.C. and an NIH postdoctoral fellowship to E.B.

Chapter 4

Complete Chloroplast Genome Sequence of a Tree Fern *Alsophila Spinulosa*

Lei Gao, Xuan Yi, Yong-Xia Yang, Ying-Juan Su, and Ting Wang

INTRODUCTION

Ferns have generally been neglected in studies of chloroplast genomics. Before this study, only one polypod and two basal ferns had their complete chloroplast (cp) genome reported. Tree ferns represent an ancient fern lineage that first occurred in the Late Triassic. In recent phylogenetic analyses, tree ferns were shown to be the sister group of polypods, the most diverse group of living ferns. Availability of chloroplast genome sequence from a tree fern will facilitate interpretation of the evolutionary changes of fern chloroplast genomes. Here we have sequenced the complete chloroplast genome of a scaly tree fern *Alsophila spinulosa* (Cyatheaceae).

The *Alsophila* chloroplast genome is 156,661 base pairs (bp) in size, and has a typical quadripartite structure with the large (LSC, 86,308 bp) and small single copy (SSC, 21,623 bp) regions separated by two copies of an inverted repeat (IRs, 24,365 bp each). This genome contains 117 different genes encoding 85 proteins, 4 rRNAs, and 28 tRNAs. Pseudogenes of *ycf66* and *trnT-UGU* are also detected in this genome. A unique *trnR-UCG* gene (derived from *trnR-CCG*) is found between *rbcL* and *accD*. The *Alsophila* chloroplast genome shares some unusual characteristics with the previously sequenced chloroplast genome of the polypod fern *Adiantum capillus-veneris*, including the absence of 5 tRNA genes that exist in most other chloroplast genomes. The genome shows a high degree of synteny with that of *Adiantum*, but differs considerably from two basal ferns (*Angiopteris evecta* and *Psilotum nudum*). At one endpoint of an ancient inversion we detected a highly repeated 565-bp-region that is absent from the *Adiantum* chloroplast genome. An additional minor inversion of the *trnD-GUC*, which is possibly shared by all ferns, was identified by comparison between the fern and other land plant chloroplast genomes.

By comparing four fern chloroplast genome sequences, it was confirmed that two major rearrangements distinguish higher leptosporangiate ferns from basal fern lineages. The *Alsophila* chloroplast genome is very similar to that of the polypod fern *Adiantum* in terms of gene content, gene order, and GC content. However, there exist some striking differences between them: the *trnR-UCG* gene represents a putative molecular apomorphy of tree ferns; and the repeats observed at one inversion endpoint may be a vestige of some unknown rearrangement(s). This work provided fresh insights into the fern chloroplast genome evolution as well as useful data for future phylogenetic studies.

The chloroplast genome has long been a focus of research in plant molecular evolution and systematics due to its small size, high copy number, conservation, and extensive characterization at the molecular level (Raubeson and Jansen, 2005). More recently, with technical advances in DNA sequencing, the number of completely sequenced chloroplast genomes has grown rapidly. Aside from providing information on genome structure, gene content, gene order, and nucleotide composition, complete chloroplast genome sequences also offer a unique opportunity to explore the evolutionary changes of the genome itself. In general, chloroplast genomes are structurally highly conserved across land plants. However, structural rearrangements, for example gene loss, inverted repeat (IR) loss, or expansion and inversion, do occur in certain lineages and have been shown to be extremely informative in resolving deep phylogenetic relationships because they may exhibit less homoplasy than sequence data (Raubeson and Jansen, 2005). For example, a 30-kb inversion shared by all vascular plants except lycopsids identifies the lycopsids as the basal lineage in the vascular plants (Raubeson and Jansen, 1992). Two inversions and an IR expansion can be used to clarify basal nodes in the leptosporangiate ferns (Raubeson and Stein, 1995; Stein et.al., 1992).

Currently, one limiting factor in comparative chloroplast genomics is the sparse taxon sampling in spore-bearing land plants. The representation of genome sequencing almost always favors plants of economic interest (Pryer, 2002). Complete chloroplast genomes have been sequenced for more than 100 seed plants. Among these, more than 10 completed sequences each are from cereals, crucifers, and conifers respectively. But for other land plants, excluding seed plants, only 10 chloroplast genome sequences have been achieved in total, of which only three are from ferns prior this study. For further insights into the evolutionary dynamics of chloroplast genome organization, more data from plant species representative of other crucial evolutionary nodes is needed (Pryer, 2002)..

Ferns (*monilophytes*), with more than 10,000 living species, are the most diverse group of seed-free vascular plants (Schneider et al., 2004; Smith et al., 2006). Previous studies have uncovered considerable genomic rearrangements in fern chloroplast genomes, but the details and exact series of these events have not yet been fully characterized (Raubeson and Stein, 1995; Roper et al., 2007; Stein et al., 1992). The completed chloroplast genome sequence of the polypod fern *Adiantum capillus-veneris* shows some unusual features not seen in vascular plants before, including tRNA gene losses, which had only been observed in chloroplast genomes of non-photosynthetic plants (Wolf and Roper, 2008; Wolf et al., 2003). For example, a putative tRNA-selenocysteine (*tRNA-Sec*) gene in *Adiantum* (Wolf et al., 2003) replaces the typical *trnR-CCG* gene. Unfortunately, because *Adiantum* is the only sequenced representative of leptosporangiates, the most diverse fern lineage, it is difficult to tell which characteristics are unique to *Adiantum* or diagnostic of a much larger clade. Therefore, complete chloroplast genome data from more fern clades are necessary to better resolve these issues.

As part of an effort to shed more light on the chloroplast genome evolution in ferns, we have sequenced the complete plastid genome of a scaly tree fern *Alsophila*

spinulosa (*ab.* *Alsophila*) (Cyatheaceae). This taxon was chosen because it is an easily available representative of an ancient lineage tree ferns, for which no chloroplast genome has been sequenced before. In addition to tree ferns, heterosporous, and polypod ferns are the other two main lineages within the “core leptosporangiates” (Pryer et al., 2004). The three major lineages of “core leptosporangiates” were thought to have originated from a Late Triassic diversification (Pryer et al., 2004). Recent phylogenetic studies further demonstrated a sister relationship between tree ferns and polypods (Pryer et al., 2001, 2004; Schneider et al., 2004). After the Late Triassic diversification, polypods remarkably re-diversified along with angiosperms in the Cretaceous (Pryer et al., 2004; Schneider et al., 2004). Similarly, the scaly tree ferns (Cyatheaceae) also radiated very recently and diversified at an exceptionally high rate (Janssen et al., 2008). A comparison of the complete chloroplast genome sequences between *Alsophila* and the polypod fern *Adiantum* will aid interpretation of unusual characters observed in *Adiantum*, such as some missing and novel genes (Wolf and Roper, 2008; Wolf et al., 2003).

Moreover, sequences of all four published fern chloroplast genomes (including that of *Alsophila*) will enable more detailed comparisons of the organization and evolution of the chloroplast genomes in ferns. Our comparative analyses corroborate that fern chloroplast genomes have undergone substantial changes in gene orders during evolution: two main rearrangements contribute to major differences between “higher” and basal ferns. In addition, the comparisons also identify some unique characteristics in the *Alsophila* chloroplast genome including a novel tRNA, interesting pseudogenes, and a highly repeated 565-bp-region spanning one endpoint of an ancient inversion.

MATERIALS AND METHODS

Genome Sequencing and Assembly

Young leaves of *Alsophila spinulosa* were collected from a plant growing in the greenhouse in Wuhan Botanical Garden, Chinese Academy of Sciences. A voucher specimen was deposited at Wuhan Botanical Garden. Total DNA was extracted using the CTAB-based method (Gawel and Jarret, 1991). The chloroplast genome was amplified using polymerase chain reaction (PCR). In brief, the coding sequences were extracted from known chloroplast genomic sequences of three ferns (GenBank: NC_003386, NC_008829, and NC_004766), three bryophytes (GenBank: NC_001319, NC_005087, and NC_004543), and one lycophyte (GenBank: NC_006861) according to their annotations in GenBank. The PCR primers were developed from alignments of the above coding sequences. Overlapping regions of each pair of adjacent PCR fragments exceeded 150 bp. We did not clone two IRs separately, but designed primers to amplify the regions spanning the junctions of LSC/IRA, LSC/IRB, SSC/IRA, and SSC/IRB. Using these primers, we covered the entire chloroplast genome of *Alsophila* with PCR products ranging in size from 500 bp to 5 kb. All PCR reactions were performed using TaKaRa LA taq (TaKaRa Bio Inc, Shiga, Japan). Amplified chloroplast genome fragments were cloned into TaKaRa pMD19-T plasmids (TaKaRa Bio Inc, Shiga, Japan), which were then used to transform *E. coli* DH5 α . Multiple (≥ 6) clones were randomly selected and commercially sequenced using ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). For long fragments (>1.4 kb), walking primers were

designed based on acquired sequences and used for sequencing remaining sequences step-by-step. Gap regions (caused by unsuccessful PCR amplification or failed primer walking sequencing) were amplified using primers that flank the gaps, then cloned and sequenced as above. From the individual reads we excluded vector, primer, and low-quality sequences, then we assembled the reads using Phrap (Ewing and Green, 1998). Since automated assembly methods cannot distinguish two IRs, we input the reads as two parts and acquired two large contigs, with each contig including one IR and its adjacent partial large and small single copy (LSC and SSC) regions. Then the two large contigs were manually assembled into the complete circular genome sequence. The IRs were identified through alignment of the final complete genome sequence against itself via BLAST 2 sequences at the National Center for Biotechnology Information (Tatusova and Madden, 1999). We accumulated 1,415,559 bp sequences, which is about 9-fold coverage.

Annotation and Related Study

Annotation of the *Alsophila* chloroplast genome was performed using DOGMA (Dual Organellar GenoMe Annotator) (Wyman et al., 2004). Genes that were undetected by DOGMA, such as *ycf1*, *ycf2*, *rps16*, *ndhF*, *ndhG*, and *matK*, were identified by BLASTX <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. From this initial annotation, putative starts, stops, and intron positions were determined by comparisons with homologous genes in other chloroplast genomes and by considering the possibility of RNA editing, which can modify the start and stop positions. The *tRNA* genes were annotated using DOGMA and ARAGORN v1.2 <http://130.235.46.10/ARAGORN/> (Laslett et al., 2004), and then confirmed by ERPIN <http://tagc.univ-mrs.fr/erpin/> (Gautheret and Lambert, 2001) and TFAM Webserver v1.3 (Taquist et al., 2007). The circular gene map of the *Alsophila* chloroplast genome was drawn by GenomeVx (Conant and Wolfe, 2008) followed by manual modification.

Synteny among fern chloroplast genomes was analyzed and visualized by using online zPicture software <http://zpicture.dcode.org/> (Ovcharenko et al., 2004).

Examination of GC Content

Overall GC content was calculated for 118 land plant plastid genomes. For the *Alsophila* chloroplast genome, GC content was further determined for three groups of genes, protein-coding genes (85), *rRNA* genes (4), and *tRNA* genes (28), respectively. For protein-coding genes, GC content was calculated for the entire gene and the first, second, and third codon positions, respectively. Protein-coding genes were partitioned into three main functional groups: photosynthetic genes, genetic system genes, and *NADH* genes. The GC content of the three groups of genes was then determined. The genes included in each of these three groups were: (1) photosynthetic genes (*rbcL*, *atp**, *pet**, *psa**, and *psb**); (2) genetic system genes (*rpl**, *rps**, *rpo**, *clpP*, *infA*, and *matK*); and (3) NADH genes (*ndh**).

Dispersed Repeats

Direct and IRs in the *Alsophila* and *Adiantum* (GenBank: NC_004766) chloroplast genomes were determined by using REPuter (Kurtz et al., 2001) at a repeat length

≥ 30 bp with a Hamming distance of 3. The entire genome was used to detect repeats in order to map them in both copies of the IR, but numbers of repeats were based on results from only one IR copy.

RESULTS AND DISCUSSION

General Features

The chloroplast genome of *Alsophila spinulosa* (GenBank: FJ556581) is 156,661 base pairs with a LSC region of 86,308 bp separated from a 21,623-bp SSC region by two IRs, each of 24,365 bp (Figure 1). The genome is the largest among the four sequenced fern chloroplast genomes (Table 1), but is smaller than previous estimates of other Cyatheaceae species, for example, *Alsophila bryophila* (165 kb), *Cyathea furfuracea* (179.2 kb), and *Sphaeropteris cooperi* (164.3 kb), using the mapping method (Conant et al., 1994). When the IR is considered only once, the *Alsophila* chloroplast genome contains 117 genes, encoding 85 proteins, 4 rRNAs, and 28 tRNAs (Table 1). Pseudo-genes of ycf66 and trnT-UGU were also detected in this genome (Figure 1). More than half of the *Alsophila* chloroplast genome is composed of coding regions (92,691 bp, 59.17%) with the protein-coding regions accounting for the major portion (81,111 bp, 51.77%) followed by rRNA genes (9,086 bp, 5.80%), and tRNA genes (2,494 bp, 1.59%) (counting both IRs).

Table 1. Comparison of general features of fern chloroplast genomes.

	<i>Alsophila spinulosa</i>	<i>Adiantum capillus-veneris</i>	<i>Psiilotum nudum</i>	<i>Angiopteris evecta</i>
Total length (bp)	156661	150568	138829	153901
GC content (%)	40.43	42.01	36.03	35.48
LSC length (bp)	86308	82282	84617	89709
GC content (%)	39.62	40.82	33.63	33.66
SSC length (bp)	21623	21392	16304	22086
GC content (%)	37.85	37.10	29.97	33.05
IR length (bp)	24365	23447	18954	21053
GC content (%)	43.02	46.33	44.00	40.65
Number of gene ^a	117	117	118	121
Protein gene	85	84	81	85
rRNA gene	4	4	4	4
tRNA gene	28	29 ^b	33	32

^a The genes in IRs were considered only once

^b *trnSeC* was counted here according to the annotation in GenBank and the reference [10]

The *Alsophila* chloroplast genome has an overall GC content of 40.43%, which is lower only than *Adiantum capillus-veneris* among the four sequenced fern chloroplast genomes (Table 1) and is the fourth highest among sequenced land plant chloroplast genomes. Like other land plants (Cai et al., 2006; Shimada and Sugiura, 1991), The GC content is unevenly distributed across the *Alsophila* chloroplast genome by location, functional group, and codon position. The GC content in rRNA genes (55.18%)

and tRNA genes (54.55%) is much higher than in protein-coding regions (40.87%). The GC percentage in IRs is the highest (Table 1), reflecting the high GC content of rRNA genes. Among the protein genes, photosynthetic genes possess the highest GC content (43.85%), followed by genetic system genes (40.80%), while NADH genes have the least (39.54%). The GC content also varies by codon position with the first (47.75%) > second (40.94%) > third (33.91%) position in turn.

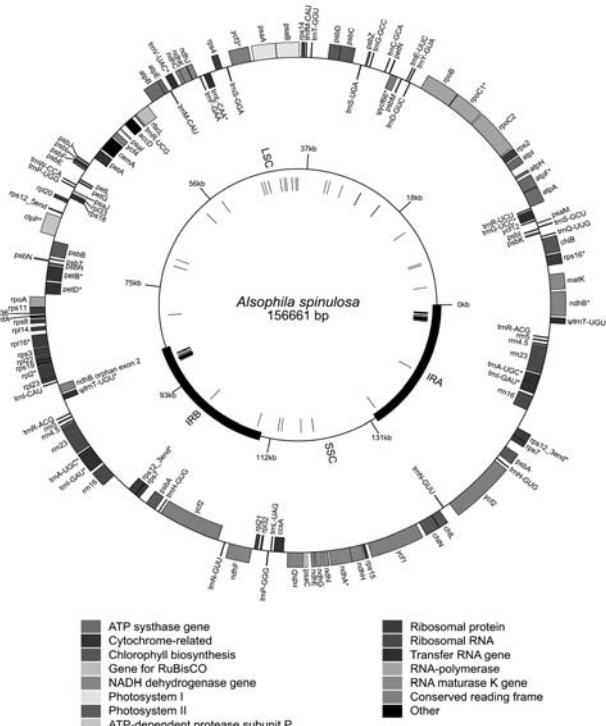


Figure 1. Gene map of the *Alsophila spinulosa* chloroplast genome. Thick black lines on inner cycle indicate the inverted repeats (IRA and IRB) which separate the genome into the large (LSC) and small (SSC) single copy regions. Genes shown on the inside of the circle are transcribed counterclockwise and those on the outside clockwise. Gene boxes are color coded by functional group as shown in the key. Asterisks denote genes with introns. Ψ represents pseudogene. Nucleotide positions are numbered starting at the boundary of IRA and LSC, with position 1 in the intron of *ndhB*. The circle of hashmarks indicates the location of direct and inverted repeats detected by REPuter (Kurtz et al., 2001).

The start codons of 85 protein genes were inferred by comparisons with previously annotated land plant chloroplast genomes. Sixty-three of these genes start with AUG, 20 with ACG and 2 with GUG (*psbC* and *rps12*). An ACG codon may be restored to a canonical start codon (AUG) by RNA editing, whereas a GUG initiation codon has been reported in other chloroplast genomes (Kuroda et al., 2007; Sugiura et al., 1998). Inferring translation start positions based only on genome sequences is merely hypothetical (Wolf et al., 2003). Future determination of sequences from complementary DNA (cDNA) and/or proteins will help to substantiate the putative translation start positions as well as RNA editing sites.

There are in total 27,046 codons in all protein-coding regions (including coding regions in both IRs) (Table 2), representing the total coding capacity of the *Alsophila* chloroplast genome; of these, 2,771 (10.25%) are for leucine, 2,365 (8.74%) for serine, 2154 (7.96%) for isoleucine, and 1847 (6.83%) for glycine. One-third of the total codons are represented by these four amino acids. The codon usage of the *Alsophila* chloroplast genome reflects an apparent AT bias. Most codons end in A or U (66.13%). As shown in Figure 2, both codon numbers and RSCU (Relative Synonymous Codon Usage) values are negatively correlated with codon GC content (represented by the number of G+C in a given codon). It appears that nucleotide composition bias has a significant influence on codon usage.

Table 2. Total numbers of each codon detected in all putative protein-coding regions in the *Alsophila spinulosa* chloroplast genome, indicated with tRNAs for which genes have been identified.

AA	Codon	Number	tRNA	AA	Codon	Number	tRNA
Phe	UUU	740		Ser	UCU	612	
Phe	UUC	539	trnF-GAA	Ser	UCC	390	trnS-GGA
Leu	UUA	787		Ser	UCA	499	trnS-UGA
Leu	UUG	877	trnL-CAA	Ser	UCG	268	
Tyr	UAU	669		Cys	UGU	204	
Tyr	UAC	267	trnY-GUA	Cys	UGC	91	trnC-GCA
Ter	UAA	82		ter	UGA	52	
Ter	UAG	27	Trp		UGG	435	trnW-CCA
Leu	CUU	471		Pro	CCU	389	
Leu	CUC	237		Pro	CCC	312	trnP-GGG
Leu	CUA	479	trnL-UAG	Pro	CCA	344	trnP-UGG
Leu	CUG	220		Pro	CCG	185	
His	CAU	407		Arg	CGU	397	trnR-ACG(x2)
His	CAC	192	trnH-GUG(x2)	Arg	CGC	144	
Gin	CAA	580	trnQ-UUG	Arg	CGA	314	trnR-UCG
Gin	CAG	247		Arg	CGG	170	
lie	AUU	1051		Thr	ACU	530	
lie	AUC	488	trnl-GAU(x2)	Thr	ACC	294	trnT-GGU
lie	AUA	615	trnl-CAU	Thr	ACA	394	
Met	AUG	543	trnfM-CAU	Thr	ACG	208	trnM-CAU
Asn	AAU	938		Ser	AGU	431	
Asn	AAC	332	trnN-GUU(x2)	Ser	AGC	165	trnS-GCU
Lys	AAA	883		Arg	AGA	497	trnR-UCU
Lys	AAG	472		Arg	AGG	232	
Val	GUU	563		Ala	GCU	734	
Val	GUC	216		Ala	GCC	262	
Val	GUA	564	trnV-UAC	Ala	GCA	428	trnA-UGC(x2)
Val	GUG	236		Ala	GCG	204	
Asp	GAU	885		Gly	GGU	683	
Asp	GAC	241	trnD-GUC	Gly	GGC	191	trnG-GCC
Glu	GAA	1,006	trnE-UUC	Gly	GGA	658	trnG-UCC
Glu	GAG	460		Gly	GGG	315	

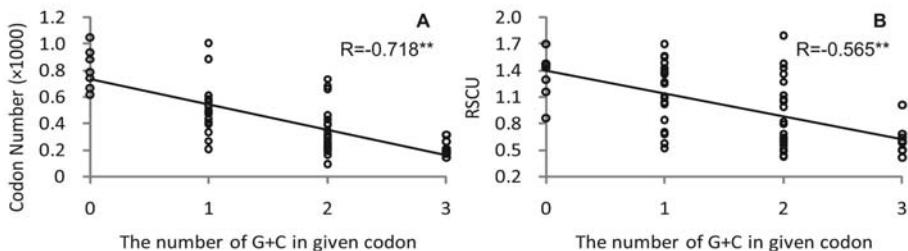


Figure 2. Correlations between codon usage and codon GC content in the *Alsophila spinulosa* chloroplast genome. Codon usage is represented by total number (A) and RSCU (Relative Synonymous Codon Usage) (B) of each codon; codon GC content is indicated by the number of G+C in given codon. Each point represents one of the 59 degenerate codons. Pearson correlations shown in the figure are all significant at $p < 0.001$. (A) Correlation between total number and GC content of each codon; (B) correlation between RSCU value and GC content of each codon.

Gene Order

The *Alsophila* chloroplast genome shares three key inversions with other ferns relative to bryophytes (Figure 3): (1) a 30-kb inversion at the beginning of LSC (close to IRA) (Raubeson and Jansen, 1992); (2) an approximately 3 kb inversion involving trnT, psbD, psbC, trnS, psbZ, and trnG (Roper et al., 2007; Wakasugi et al., 1998; Wolf et al., 2003); and (3) a minor inversion containing a single gene trnD-GUC. The first of these inversions is also shared by all vascular plants except lycophytes, whereas the latter two are restricted to ferns.

To our knowledge, the trnD-GUC inversion has not been previously documented. Three conserved and consecutive tRNA genes, trnD-GUC, trnY-GUA, and trnE-UUC, have been identified in all land plant chloroplast genomes. Excluding ferns, the three genes have the same directions of transcription. However, in ferns trnD is inverted relative to trnY and trnE (Figure 3). The simplest interpretation of this change is a single minor inversion involving only trnD. Based on current data, it remains unknown whether the 3-kb inversion or the trnD inversion occurred first in ferns.

Overall, the *Alsophila* chloroplast genome shows a high degree of synteny with the previously sequenced chloroplast genome of *Adiantum* (Figure 4A). In contrast, there exist striking differences between *Alsophila* and *Angiopteris* (Figure 4B) as well as between *Alsophila* and *Psilotum* (Figure 4C). A set of complex rearrangements in the IRs, involving a rare duplication of *psbA* gene, was found in “higher” ferns using physical mapping (Raubeson and Stein, 1995; Stein et al., 1992). The IR gene orders of “higher” ferns, such as *Adiantum*, *Cyathea*, and *Polystichum*, are highly rearranged in comparison to that of basal leptosporangiate *Osmunda* (Hasebe et al., 1992; Raubeson and Stein, 1995; Stein et al., 1992). Complete chloroplast genome data from *Angiopteris*, *Psilotum*, *Adiantum*, and *Alsophila* detail these rearrangements. The IR gene order in *Alsophila* appears to be the same as that in *Adiantum*, while *Angiopteris* and *Psilotum* have the *Osmunda* gene order. To explain the complex rearrangements, a “two inversions” hypothesis was proposed (Hasebe et al.,

1992). Figure 5 illustrates the great gene order changes within these rearrangements and the updated version of the “two inversions” model incorporating gene order data from the *Alsophila* and *Angiopteris* chloroplast genomes. Recently, Wolf and Roper (2008) indicated that the two major inversions did occur in turn and the second inversion (Figure 5, Inversion II) took place on the branch leading to the common ancestor of the heterosporous fern clade and its sister group. Thus, it seems reasonable to hypothesize that the *Adiantum* gene order represents a common feature of the three lineages within core leptosporangiates (including heterosporous ferns, tree ferns, and polypod ferns).

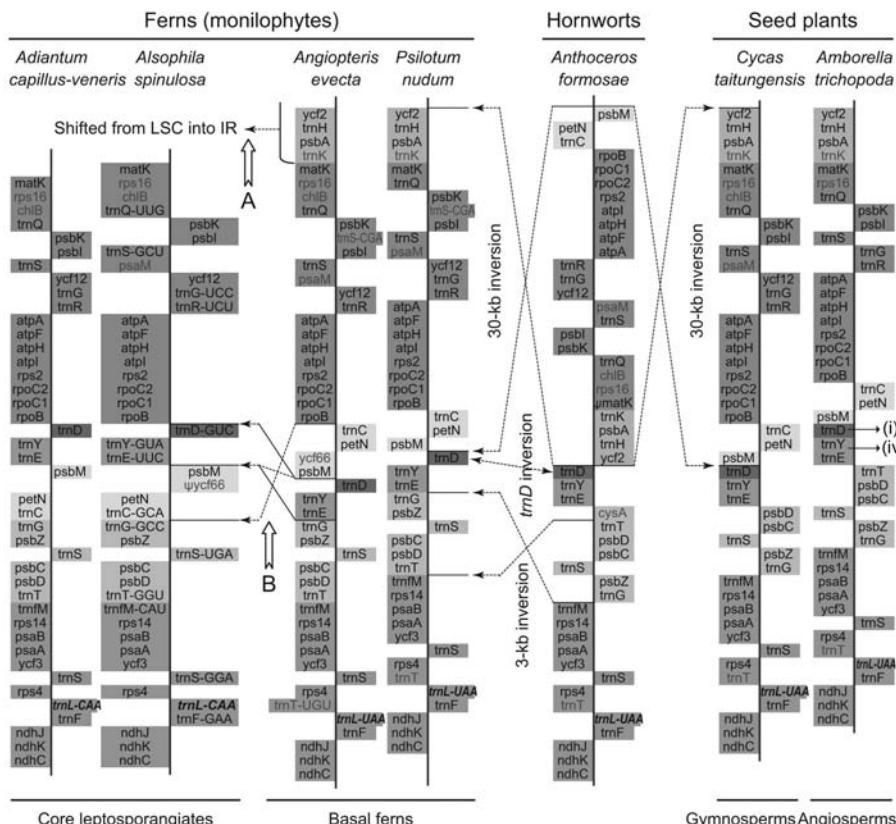


Figure 3. Expected rearrangements in the evolution of fern chloroplast genomes. Genes are represented by boxes extending right or left of the base-line according to the direction of transcription. Each colored gene segment shows the same gene order region among the seven land plants chloroplast genomes. The boxes highlighted in i denote the inversion of trnD-GUC. Excluding *Alsophila spinulosa*, the unchanged tRNA anticodon is abbreviated in the other six chloroplast genomes. The genes that are missing in one or more chloroplast genomes are shown in iv. The tRNA-leu (CAA/UAA) gene between rps4 and ndhJ is indicated in bold italic type. The pseudogene is denoted by ψ . A, details are shown in Figure 5; B, hypothetical pathways to explain this rearrangement are illustrated in Figure 6.

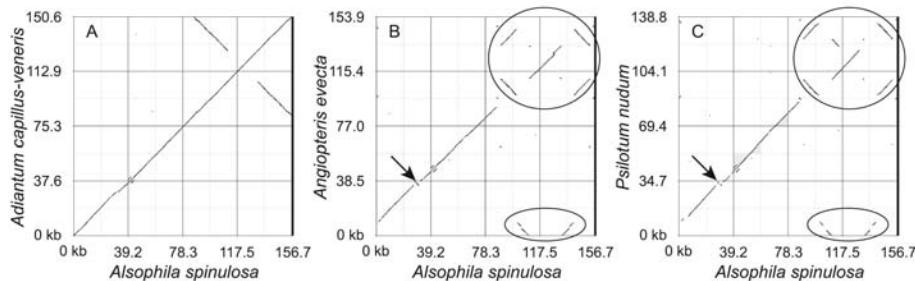


Figure 4. Comparisons of the gene order of *Alsophila spinulosa* chloroplast genome with other ferns. Comparison by using online zPicture software <http://zpicture.dcode.org/> (Ovcharenko et al., 2004). Points along the positive slope are in the same orientation in both genomes, whereas points along the negative slope indicate sequences that are oriented in opposite directions. (A) The two groups of points that fall along the negative slope in the upper right corner represent the sequences of IRs. (B) and (C) ellipses denote the complex rearrangements in IRs and arrows indicate the gene order change between *rpoB* and *psbZ*.

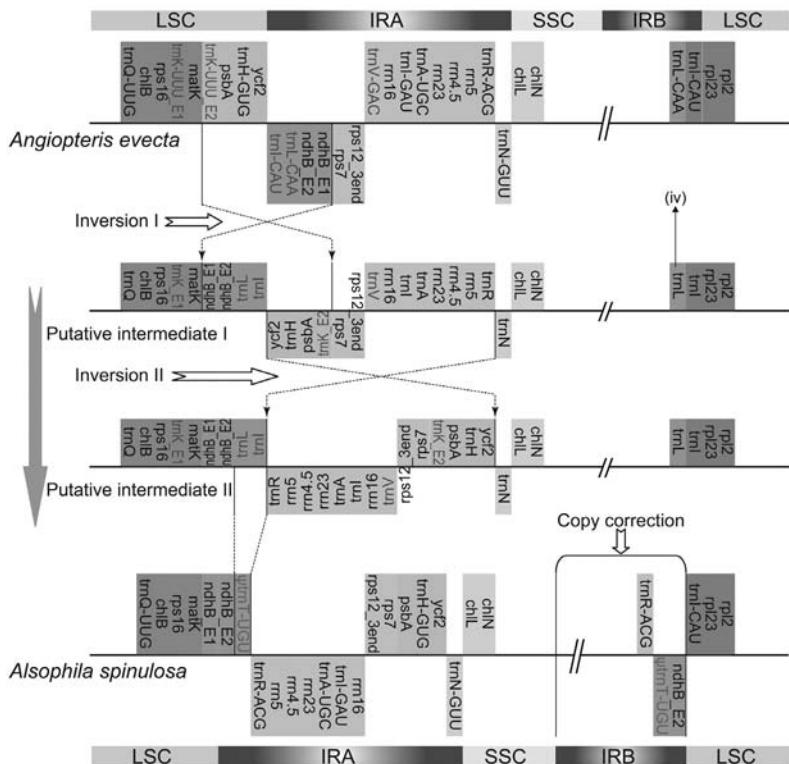


Figure 5. The “two inversions” model for IR rearrangements in fern chloroplast genomes. Genes are represented by boxes extending above or below the base-line according to the direction of transcription. The genes that are absent in the *Alsophila* chloroplast genomes are shown in iv; these loci are merely hypothetical in putative intermediates since the course of their loss is unclear. The pseudogene of *trnT-UGU* is represented indicated by ψ . The tRNA anticodon is abbreviated in putative intermediates.

Interestingly, in the *Adiantum* chloroplast genome, an intron-containing trnT-UGU was identified between *trnR-ACG* and *ndhB* (Table 3) (Wolf et al., 2003). The *Alsophila* chloroplast genome possesses no intact intron-containing trnT. However, two fragments that are similar to the two exons within the *Adiantum* trnT were annotated as a Ψ trnT-UGU in this study (Table 3). This new trnT or Ψ trnT is just at one endpoint of the Inversion II (Figure 5). Therefore, the generation of intron-containing trnT-UGU may be associated with the IR rearrangements.

Table 3. Comparison of gene contents of fern chloroplast genomes.

Gene ^a	<i>Alsophila spinulosa</i>	<i>Adiantum capillus-veneris</i>	<i>Psilotum nudum</i>	<i>Angiapteris evecta</i>
tRNA gene				
<i>trnR-CCG</i> (<i>rbcL-accD</i>)	<i>trnR-UCG</i>	<i>trnSeC</i>	●	●
<i>trnL-CAA</i> (<i>ndhB</i> Exon2 3')	○	○	●	●
<i>trnL-UAA</i> (<i>rps4-ndhJ</i>)	<i>trnL-CAA</i>	<i>trnL-CAA</i>	●	●
<i>trnK-UUU</i> (<i>MatK</i>)	○	○	●	●
<i>trnS-CGA</i> (<i>psbK-psbI</i>)	○	○	●	●
<i>trnT-UGU</i> (<i>rps4-ndhJ</i>) (without intron)	○	○	●	●
<i>trnT-UGU</i> (<i>ndhB</i> Exon2 3') (with one intron)	▲	●	○	○
<i>trnV-GAC</i> (<i>rrn</i> 16- <i>rps</i> 12)	○	○	●	●
<i>trnG-UCC</i> (<i>ycf12-atpA</i>)	●	●	●	○
Protein gene				
<i>psaM</i>	●	○	●	●
<i>ycf66</i>	▲	○	○	●
<i>chlB</i>	●	●	○	●
<i>chlL</i>	●	●	○	●
<i>chiN</i>	●	●	○	●
<i>rps/6</i>	●	●	○	●
<i>ycf1</i>	●	●	●	▲

^aOnly the genes that are missing in one or more cp genomes are shown. The loci of tRNA genes are denoted by their neighbor protein genes. A filled/open circle denotes the presence/absence of a gene. A filled triangle indicates a pseudogene.

Alsophila and *Adiantum* share another rearranged region between *rpoB* and *psbZ* in LSC relative to *Angiopteris* and *Psilotum* (Figure 3). For the latter two, gene order in this region is “*rpoB-trnC-petN-psbM-trnD-trnY-trnE-trnG-psbZ*”, whereas in *Alsophila* and *Adiantum* it is “*rpoB-trnD-trnY-trnE-psbM-petN-trnC-trnG-psbZ*” (Genes with boldface are unchanged) (Figure 3). Roper et al. (2007) noted that this gene order change is not caused by a single inversion. Two alternative pathways may account for this rearrangement (Figure 6), but more data are needed to determine the order of the two inversions.

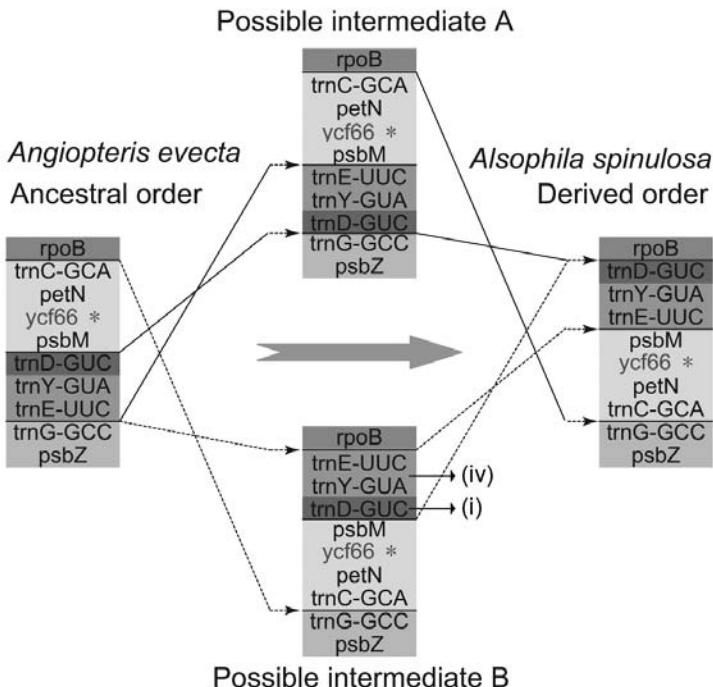


Figure 6. Two hypothetical pathways to explain rearrangement between rpoB and psbZ. Genes are represented by boxes, the colors of which are consistent with Figure 3. *In *Angiopteris evecta* ycf66 has an intact ORF, but in *Alsophila spinulosa* it is a pseudogene.

Gene Content

A total of 117 different genes are present in the *Alsophila* chloroplast genome (Table 1). This gene content is similar to that of most land plants (Wakasugi et al., 2001). However, there are some interesting differences among the four sequenced fern chloroplast genomes (Table 3). The *Alsophila* chloroplast genome possesses the least number of tRNA genes due to five missing tRNA genes in comparison to basal ferns (*Psilotum nudum* and *Angiopteris evecta*). Its protein gene number is equal to that of *Angiopteris*, but higher than that of both *Adiantum* and *Psilotum*. Details of these differences are discussed below.

Novel tRNA Gene

A unique *trnR-UCG* gene, encoding tRNA-Arg, is found between *rbcL* and *accD* in the *Alsophila* chloroplast genome (Figure 1; Table 3). Another type of tRNA-Arg gene *trnR-CCG* resides in the same locus in non-flowering land plants including *Angiopteris* (Roper et al., 2007) and *Psilotum* (Wakasugi et al., 1998). In *Adiantum*, an apparent tRNA gene is annotated as *trnSec* (Wolf et al., 2003). It is uncertain whether the occurrence of *trnR-UCG* in the *Alsophila* chloroplast genome represents a unique feature for this species or is an apomorphy for a larger clade such as Cyatheaceae or tree ferns. To address this question, we collected all fern *rbcL-accD* intergenic

sequences deposited in GenBank and examined the *tRNA* genes within them using ARAGORN (Laslett et al., 2004). The results indicate that *trnR-UCG* is restricted to tree ferns, whereas *trnR-CCG* is widespread in non-core leptosporangiates and basal ferns (Table 4). However, neither *trnR-UCG* nor the *trnR-CCG* gene is identified at this locus in polypod ferns. Therefore, the existence of *trnR-UCG* may reflect a putative molecular apomorphy of tree ferns.

Table 4. tRNA genes in fern rbcL-accD intergenic spacer sequences.

Order ^b	Family	Number of sequences	tRNA gene ^c
Cyatheales (tree ferns)	Cyatheaceae	140	trnR-UCG
	Dicksoniaceae	6	trnR-UCG
	Lophosoriaceae	1	trnR-UCG
	Hymenophyllopsidaceae	1	trnR-CCG
Gleicheniales	Dipteridaceae	1	trnR-CCG
	Gleicheniaceae	2	trnR-CCG
Hymenophyllales	Hymenophyllaceae	84	trnR-CCG
Osmundales	Osmundaceae	18	trnR-CCG
Marattiales	Marattiaceae	1	trnR-CCG
Psilotales	Psilotaceae	1	trnR-CCG
Ophioglossales	Ophioglossaceae	1	trnR-CCG

^a All sequence data were obtained from GenBank at March 30, 2009;

^b Group names at ordinal level follow Smith et al. [7];

^c tRNA genes were identified by using ARAGORN v 1.2 [http://130.235.46.10/ARAGORN/\[22\]](http://130.235.46.10/ARAGORN/[22])

Sequence alignment indicates that *trnR-UCG* and *trnR-CCGs* have quite similar primary sequences with 44 of 74 nucleotides unchanged across seven representative land plants (Figure 7A). In addition, the *Adiantum trnSeC* shares 51, 41, and 40 identical nucleotides with the *Alsophila trnR-UCG*, the *Psilotum trnR-CCG*, and the *Angiopteris trnR-CCG* respectively (Figure 7A). Due to their similarities and conserved loci, we propose that *Alsophila trnR-UCG*, *Adiantum trnSeC* as well as *trnR-CCGs* in other land plants are orthologous. Tree fern *trnR-UCG* can transfer arginine even though its anticodon alters from CCG to UCG. However, *Adiantum trnSeC* has undergone major changes: (1) its anticodon is UCA (unmatchable for an Arg codon), and (2) it contains up to 18 nucleotide differences relative to all other land plant *trnR* genes (Figure 7A). Our findings imply that the *trnR-UCG* is derived from the *trnR-CCG* by the alteration of one anticodon base; then the *Adiantum trnSeC* evolves from the *trnR-UCG* by altering one anticodon base further, becoming a *trnR-UCG* pseudogene (Figure 7B). If this is the case, the *Adiantum trnSeC* should be annotated as Ψ *trnR*. Sugiura and Sugita (2004) argued that the *trnR-CCG* is not essential for plastid function although it is conserved in non-flowering plants. The evolutionary scenario of *trnR-CCG* in ferns (Figure 7B) tends to support this view.

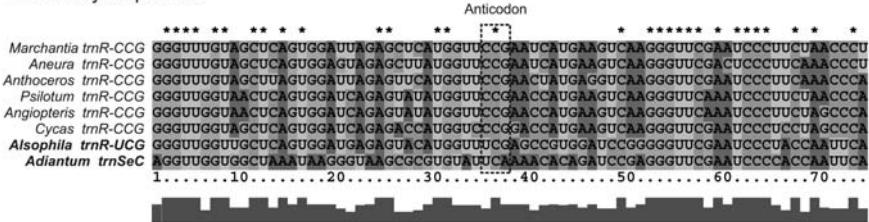
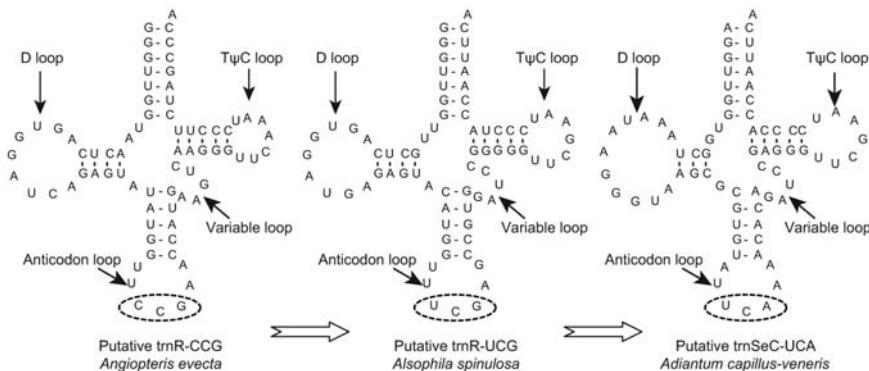
A: Primary sequences**B: Secondary structures**

Figure 7. Comparisons of the sequences and putative secondary structures of *trnR-CCG*, *trnR-UCG*, and *trnSeC* (tRNA-selenocysteine). (A) Primary sequences of *Alsophila spinulosa* *trnR-UCG*, *Adiantum capillus-veneris* *trnSeC*, and *trnR-CCGs* from other six land plants. Dashed rectangle indicates anticodons. (B) The putative secondary structures of *Angiopteris evecta* *trnR-CCG*, *Alsophila* *trnR-UCG*, and *Adiantum* *trnSeC*. Dashed ellipses indicate anticodons.

At the locus between *rps4* and *ndhJ*, the *Alsophila* and *Adiantum* chloroplast genomes encode a *trnL-CAA* (tRNA-Leu) rather than a *trnL-UAA* gene (Table 3). However, they lose another *trnL-CAA* gene (Table 3), which is found at the 3' downstream of *ndhB* in almost all other land plant plastid genomes. Consequently, *Alsophila* and *Adiantum* only possess the *trnL-CAA*, whereas the *Angiopteris* and *Psilotum* chloroplast genomes contain both the *trnL-UAA* and the *trnL-CAA*. In the *Adiantum* chloroplast, the missing *trnL-UAA* could be provided for the heavily used UUA codon by a partial C-to-U edit in the *trnL-CAA* anticodon (Wolf et al., 2004). Since the UUA is also a preferred leucine codon for the *Alsophila* chloroplast genome (*RSCU* = 1.70), the same editing event might occur in the *Alsophila* chloroplast as well.

Missing tRNA Gene

Only 28 tRNA genes are encoded in the *Alsophila* chloroplast genome, whereas 29, 32, and 33 are annotated in *Adiantum*, *Angiopteris*, and *Psilotum*, respectively (Table 1). For chloroplast genomes, it is believed that a set of 30 tRNA species is sufficient for the translation of chloroplast mRNAs (Shinozaki et al., 1986). In the *Angiopteris* and *Psilotum* chloroplasts, tRNAs can read all codons by using two-out-of-three and wobble mechanisms (Pfitzinger et al., 1990). However, in *Alsophila* and *Adiantum* chloroplasts, both lysine codons lack a corresponding tRNA-Lys (encoded by *trnK*)

(Tables 2 and 3). The loss of *trnK* suggests cytosolic tRNAs may be imported into chloroplasts, despite a lack of experimental evidence (Lung et al., 2006). As an incidental consequence of the *trnK* loss, the matk open reading frame (ORF) is not nested in the *trnK* intron (Figure 1).

Apart from the *trnK* and the *trnL-CAA*, the *Alsophila* chloroplast genome also shares other three *tRNA* gene losses, including the *trnS-CGA*, the *trnV-GAC*, and the *trnT-UGU* (intron-free), with *Adiantum* relative to basal ferns *Angiopteris* and *Psilotum* (Table 3). The shared absence of *tRNA* genes between *Alsophila* and *Adiantum* suggests that they may derive from a common ancestor.

Protein Genes

The *Alsophila* chloroplast genome contains a *psaM* gene encoding photosystem I reaction center subunit M. This gene has been detected in *Psilotum* (Wakasugi et al., 1998) and *Angiopteris* (Roper et al., 2007), but not in *Adiantum* (Wolf et al., 2003) (Table 3). Besides ferns, *psaM* also exists in bryophytes, lycophytes, and gymnosperms, but not in angiosperms, implying its independent loss in ferns and angiosperms (Wolf et al., 2003). *Alsophila* and *Adiantum* represent tree ferns and polypods, respectively. Due to their sister relationship, we speculate that the loss of *psaM* in ferns occurred after the split of polypods and tree ferns.

A putative pseudogene of *ycf66* is identified in the *Alsophila* chloroplast genome (Figure 1; Table 3). The 5' ends of its two exons are both destroyed. In the four sequenced fern chloroplast genomes, only *Angiopteris* contains an intact *ycf66* gene (Roper et al., 2007). For other land plants, this gene only occurs in *Marchantia polymorpha* (liverworts), *Physcomitrella patens* subsp. *patens* (mosses), *Syntrichia ruralis* (mosses), and *Huperzia lucidula* (lycophytes). The findings suggest that *ycf66* is lost independently in multiple clades of land plants including hornworts, ferns and seed plants.

Inversion Endpoint as Hotspot for Repeats

A total of 133 pairs of repeats (≥ 30 bp) were identified in the *Alsophila* chloroplast genome by using REPuter (Kurtz et al., 2001), of which 106 are direct and 27 are IRs. This number of repeats is less than are found in some highly rearranged chloroplast genomes (e.g., *Trachelium caeruleum*) but more than are present in unrearranged ones (e.g., *Nicotiana tabacum*) (Chumley et al., 2006; Haberle et al., 2008). Up to 66 direct repeats (no IR) are restricted to a region spanning only 565 bp (153,682–154,246 bp in IRA or 88,724–89,288 bp in IRB) between *trnR-ACG* and ψ *trnT-UGU* in the IRs (Figure 1). The GC content of this 565-bp-region (35.93%) is lower than that of IRs and the overall GC content of the whole genome. Detailed sequence analyses revealed that this region is composed of tandem iterations of 11 similar segments ranging from 40 to 58 bp (Figure 8). The core repeated motif is AAAATCCTAGTAGTTAgAGCTT-TATCcaGGGtaTaGgACT (the lowercase letters denote variable bases) with variant lengths of heads and/or tails (Figure 8).

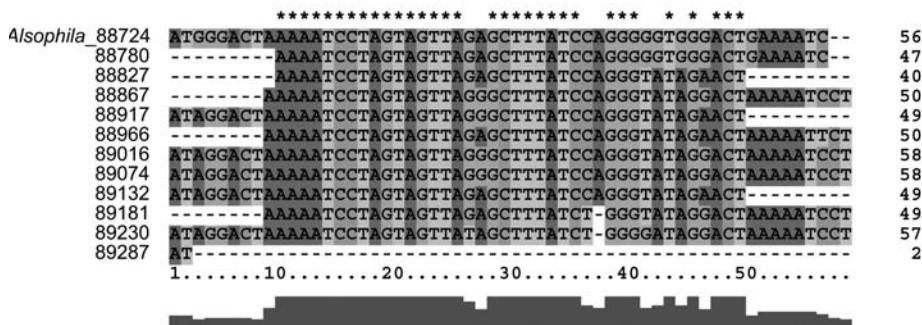


Figure 8. Repetitive units within the highly repeated 565-bp sequence. The sequence from 88,724 to 89,287 bp in IRB is shown in this figure. The numbers on the right hand side indicate the lengths of each segment.

In contrast to *Alsophila*, dispersed repeats (≥ 30 bp) are rare in the *Adiantum* chloroplast genome, with only five short IRs (30–36 bp); and none of these occurs between the trnR-ACG and the trnT-UGU. In the *Alsophila* chloroplast genome, the length of the intergenic region between trnR-ACG and ψ trnT-UGU is 1,467 bp, whereas in *Adiantum* it is 913 bp, the difference being 554 bp. We noted that this length is very similar to that of the highly repeated 565-bp-region, and speculate that the difference is caused by the presence of the highly repeated region. To test this hypothesis, we extracted the sequence from trnR-ACG to ψ trnT-UGU in *Alsophila* and from trnR-ACG to trnT-UGU in *Adiantum*. The sequence alignment indicates that the highly repeated 565-bp-region is indeed lost in the *Adiantum* chloroplast genome (Figure 9).

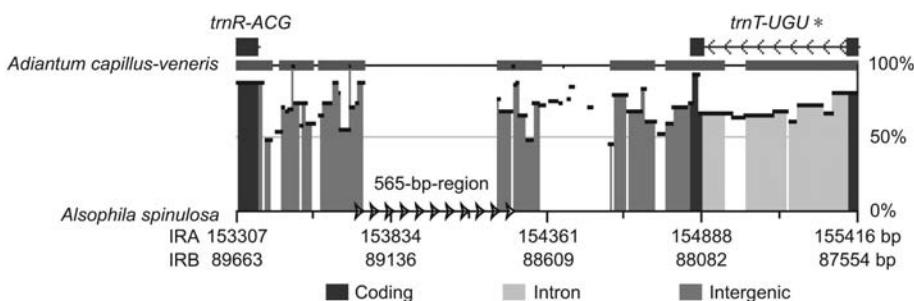


Figure 9. Alignment of the sequence from trnR-ACG to ψ trnT-UGU in *Alsophila* and corresponding region in *Adiantum*. Comparison by using online zPicture software with ECR criteria of ≥ 100 bp and $\geq 70\%$ identity <http://zpicture.dcode.org/> (Ovcharenko et al., 2004). Similarities between aligned regions are shown as average percent identity. The cluster of arrows denotes the location of the highly repeated 565-bp-region in the *Alsophila* chloroplast genome. *, an intact trnT-UGU was identified in *Adiantum*, while a trnT-UGU pseudogene was found in *Alsophila*.

In the *Alsophila* chloroplast genome, the location of the highly repeated 565-bp-region is exactly at the endpoint of the second inversion of the IR rearrangements (Figure 5, Inversion II). Dispersed repeated sequences have been reported from several chloroplast genomes. These are associated with numerous DNA rearrangements, particularly inversions (Ogihara et al., 1988; Pombert et al., 2005, 2006). In extensively rearranged chloroplast genomes, the endpoints of rearranged gene clusters are usually flanked by repeated sequences (Cai et al., 2008; Chumley et al., 2006; Haberle et al., 2008). If repeat-mediated recombination is the major mechanism generating inversions in chloroplast genomes (Palmer, 1985; Palmer et al., 1987), the preservation of repeats would destabilize the genome structure. After inversions, the repeats should be deleted to guarantee genome stability (like the situation in *Adiantum*). The repeats observed at the endpoint of the ancient inversion (Figure 5, Inversion II) may be a vestige of recent rearrangement(s) that are undiscovered. The existence of these repeats implies that the region is a potential hotspot for genomic reconfiguration.

CONCLUSION

In this study, we present the first complete chloroplast genome sequence from a tree fern and provide a comprehensive comparative analysis of chloroplast genomes in ferns. The chloroplast genome size of *Alsophila* is larger than that of *Adiantum*, *Psilotum*, and *Angiopteris*. Besides 117 genes, two pseudogenes $\Psi ycf66$ and $\Psi trnT-UGU$ are also detected in the *Alsophila* chloroplast genome. An intact *ycf66* is identified in *Angiopteris*, while an intron-containing *trnT-UGU* is found in *Adiantum*. Based on the findings, we speculate that $\Psi ycf66$ reflects an intermediate during *ycf66* gene loss, and the genesis of *trnT-UGU* may be associated with the IR rearrangements. A *trnR-UCG* gene was detected between *rbcL* and *accD* in *Alsophila*, and this seems to be a molecular apomorphy of tree ferns. In the *Adiantum* chloroplast genome, the *trnR-UCG* gene degenerates to a pseudogene. The *Alsophila* chloroplast genome shares several unusual characteristics with the previously sequenced *Adiantum* (a polypod fern) chloroplast genome, such as five missing *tRNA* genes and two major rearranged regions. These common characters probably derive from their common ancestor. In the *Alsophila* chloroplast genome, a highly repeated 565-bp-region, which is composed of tandem iterations of 11 similar segments, occurs at one endpoint of an ancient inversion, but it is not detected in the genome of *Adiantum*. Nonetheless, the origin and function of these repeats remain to be characterized in future studies.

KEYWORDS

- ***Alsophila* chloroplast**
- **Chloroplast genome**
- **GC content**
- **Inverted repeat**
- **PCR primers**

AUTHORS' CONTRIBUTIONS

Lei Gao participated in the conception of this study, carried out part of the genome sequencing, performed all sequence analyses, annotated the genome, generated tables and figures, and drafted the manuscript. Xuan Yi and Yong-Xia Yang participated in genome sequencing. Ying-Juan Su and Ting Wang conceived and supervised the project, contributed to the interpretation of the data, and prepared the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We give special thanks to Paul G. Wolf for his constructive comments and suggestions. We are grateful to Bobby A. Brown and George Littlejohn for significantly improving English, to Ling-Ling Gao, Cynthia Gleason, and Su-Min Guo for helpful comments on the manuscript, and to Chen Lu and Gong Xiao for providing valuable assistance on the use of computer hardware and software. This work was supported by the “100 Talent Project” of the Chinese Academy of Sciences (Grant No.: 0729281F02), the “Outstanding Young Scientist Project” of the Natural Science Foundation of Hubei Province, China (Grant No.: O631061H01) and the Open Project of the State Key Laboratory of Biocontrol, China (Grant No.: 2007–01). We also gratefully acknowledge the valuable comments by three anonymous reviewers.

Chapter 5

Reverse Genetic Analyses of Gene Function in Fern Gametophytes

George Rutherford, Milos Tanurdzic, Mitsuyasu Hasebe,
and Jo Ann Banks

INTRODUCTION

Ceratopteris richardii is a useful experimental system for studying gametophyte development and sexual reproduction in plants. However, few tools for cloning mutant genes or disrupting gene function exist for this species. The feasibility of systemic gene silencing as a reverse genetics tool was examined in this study.

Several DNA constructs targeting a *Ceratopteris* protoporphyrin IX magnesium chelatase (*CrChlI*) gene that is required for chlorophyll biosynthesis were each introduced into young gametophytes by biolistic delivery. Their transient expression in individual cells resulted in a colorless cell phenotype that affected most cells of the mature gametophyte, including the meristem and gametangia. The colorless phenotype was associated with a 7-fold decrease in the abundance of the endogenous transcript. While, a construct designed to promote the transient expression of a *CrChlI* double stranded, potentially hairpin-forming RNA was found to be the most efficient in systemically silencing the endogenous gene, a plasmid containing the *CrChlI* cDNA insert alone was sufficient to induce silencing. Bombarded, colorless hermaphroditic gametophytes produced colorless embryos following self-fertilization, demonstrating that the silencing signal could be transmitted through gametogenesis and fertilization. Bombardment of young gametophytes with constructs targeting the *Ceratopteris* filamentous temperature sensitive (*CrFtsZ*) and uroporphyrin dehydrogenase (*CrUrod*) genes also produced the expected mutant phenotypes.

A method that induces the systemic silencing of target genes in the *Ceratopteris* gametophyte is described. It provides a simple, inexpensive, and rapid means to test the functions of genes involved in gametophyte development, especially those involved in cellular processes common to all plants.

Plants differ from animals by incorporating into their life cycle a multicellular, haploid phase, the gametophyte, which alternates with a diploid, sporophyte phase. Although, the gametophyte is extremely reduced and inconspicuous in flowering plants, it is essential for sexual reproduction in all land plants as it produces gametes, facilitates fertilization, and, for at least a brief time, nurtures the young embryo. We have used the homosporous fern *Ceratopteris richardii* as a model system for studying gametophyte development because *Ceratopteris* gametophytes are autotrophic, small (~1 mm) and develop rapidly (Chatterjee and Roux, 2000). They can also be manipulated to develop as males or hermaphrodites by the pheromone antheridiogen

(Banks et al., 1993) and are easily crossed. Because all gametophytes are haploid, mutations affecting the gametophyte development are easily selected within days of spore mutagenesis and growth on selective medium. While *Ceratopteris* has proven to be a useful genetic system for dissecting its sex determination pathway (Banks, 1997, 1999; Strain et al., 2001), it has yet to be stably transformed, which makes it difficult to clone genes known only for their mutant phenotype or to test the functions of gametophytically expressed genes.

Recent advances in epigenetic gene silencing have led to its use as a reverse genetics tool for examining gene function in plants and animals (Ashrafi et al., 2003; Chuang and Meyerowitz, 2000; Fraser et al., 2000; Gonczy et al., 2000; Maeda et al., 2001; Misquitta and Paterson, 1999; Stoutjesdijk et al., 2002; Wessely et al., 2001). Referred to as post-transcriptional gene silencing (PTGS), co- or sense-suppression, antisense suppression, quelling, or RNA interference (RNAi) depending on the organism or the method employed, these processes result in post-transcriptional and sequence-specific gene silencing upon introduction of a transgene or double-stranded RNA (dsRNA) (reviewed in (Hannon, 2002; Pickford et al., 2002; Vaucheret et al., 2001)). What links these processes together is the presence of small 21–23 nt RNA molecules that mediate the degradation of complementary homologous RNA. Genetic screens in *Arabidopsis thaliana*, *Caenorhabditis elegans*, and *Neurospora crassa* have identified homologous genes required for gene silencing (Cogoni and Macino, 1997, 1999; Ketting and Plasterk, 2000; Mourrain et al., 2000), indicating that they share a common and evolutionarily conserved mechanism that is likely to be present in all plants, including ferns. One striking feature of gene silencing is that the silencing effects are non-cell autonomous and spread to neighboring cells. For this reason, transient expression rather than stable integration of a transgene is sufficient to induce phenotypes resulting from gene silencing.

This study investigates the feasibility of DNA vector based gene silencing as a reverse genetics technique for studying gene function in gametophytes. Our selection of *Ceratopteris* genes to target was based upon three criteria: they result in visible phenotypes when mutated in other flowering plants; they are expressed in the gametophyte; and they are present in a *Ceratopteris* EST library generated from germinating spores (Stout et al., 2003). The three genes selected included protoporphyrin IX magnesium chelatase (*CrChl*), filamentous temperature sensitive Z (*CrFtsZ*), and uroporphyrin dehydrogenase (*CrUrod*), which are necessary for chlorophyll biosynthesis, or chloroplast development. Here we show that, systemic gene silencing occurs in *Ceratopteris* gametophytes when appropriate transgene constructs are introduced into young gametophytes by particle bombardment, that the silencing effects are non-cell autonomous, and that the phenotype resulting from gene silencing can be transmitted from the gametophyte to the sporophyte generation, although at low frequencies. By comparing the efficiency of various gene-silencing constructs, we also show that cDNA constructs without any recognizable promoter sequence are sufficient to induce silencing in gametophytes at high frequencies.

MATERIAL AND METHODS

Gametophyte Culture and Imaging

The hermaphroditic her19 mutant used in this study is described in Eberle and Banks (1996). Hermaphrodite cultures were generated from sterilized spores plated on fern medium, or FM, consisting of 0.5× MS salts, pH 6.5 as previously described (Eberle and Banks, 1996). In preparation for bombardment, spores suspended in water were plated on 60 mm Petri dishes at a density of 2,500–3,000 spores per dish. Cultures were then placed in plastic bags to maintain high humidity and incubated at 30°C. To self-fertilize gametophytes, individual virgin hermaphrodites were placed into the wells of 96-well microtiter plates containing FM plus agar; enough water to submerge each gametophyte was then added to each well. Electron microscopy was done as previously described (Banks et al., 1993). All light photography was done using a Spot II camera attached to a Leica dissecting microscope; images were processed with Adobe Photoshop.

Gametophyte Bombardment

Several parameters were optimized before using the following conditions for bombardment, including the age of the gametophyte at time of bombardment, the amount of DNA delivered, gold size, rupture disc type, and shelf placement in the biolistic apparatus. 1.6 µm gold was prepared and coated with unmodified plasmid DNA purified using a Qiagen kit (Qiagen, CA) prior to bombardment using a PDS 1000 Helium System, all according to manufacturer's instructions (BioRad, CA). Gametophyte cultures were placed on a shelf 9 cm below the stopping screen of the apparatus; 1100 psi rupture discs were used. Gametophytes were bombarded 6 days after spore inoculation, and the amount of DNA delivered was 0.7 µg per shot. Unless indicated otherwise, gametophytes were scored for phenotype and histologically stained for beta-glucuronidase (GUS) activity (Jefferson et al., 1987) 6 days after bombardment. In comparing the efficiencies of different plasmid constructs in inducing gene silencing, the null hypothesis (i.e., the efficiency of gene inactivation between two different constructs is the same) was tested by applying the z test for two proportions.

DNA Constructs

A skeleton silencing DNA vector, named 35S:irint, into which a targeted gene could be easily cloned in opposing orientations separated by an intron was made by removing the GUS coding region from pFF19G (Timmermans et al., 1990) by digestion with NruI and PstI and replacing the GUS fragment with another containing the castor bean catalase intron 1 (Suzuki et al., 1994). This intron was PCR amplified using pIGI121Hm (Hiei et al., 1994) as template and the two adapter/primer sequences: 5'CGACGGACCGATCTAGAACATGGATCCCTACAGGGTA and 5' TCAGCTGCAGACTAGTTACAGGACGGACGAGTCGACGGTTC. The PCR product was digested with PstI and NruI then ligated to the pFF19G-minus GUS vector.

The genes targeted for silencing were chosen from a *Ceratopteris* EST database of ~3700 cDNAs generated from germinating, 20-hours-old spores (Stout et al., 2003). The cloning vector for the cDNA library from which ESTs were derived was pCMVSPORT6

(Invitrogen, CA). The inverted repeat-intron CrChII (Cr referring to *Ceratopteris richardii*) gene silencing construct (35S:irintCrChII) was made by PCR amplifying two 393 bp fragments from a *Ceratopteris* cDNA clone (GenBank accession number BE642494) using the primer pairs 5'GATACGGACCGGTTCTGGCAATCCAGAG-GAA and 5'ATCGGGATCCAAGGCAATTGGGAATCACTG for cloning CrChII in the sense orientation, and 5' GACGGTCGACAAGGCAATTGGGAATCACTG and 5'CGTAACTAGTGTCTGGCAATCCAGAGGAA for cloning CrChII in the anti-sense orientation. Constructs for targeting the *CrFtsZ* gene (35S:irintCrFtsZ) were made by PCR amplifying two 435 bp *CrFtsZ* fragments from a *Ceratopteris* cDNA clone (GenBank accession number BE64351) using the primer pairs 5'CATA CG-GACCGGCTCTTGAGGCCATTGAAAG and 5'ATCGGGATCCGGATCAGCCA-AGCTGGTAAC for cloning in the sense orientation, and 5'CGTAACTAGTGTCT-TGAGGCCATTGAAAG and 5'GACGGTCGACGGATCAGCCAAGCTGGTAAC for cloning in the antisense orientation. All sense orientation PCR fragments were digested with RsrII plus BamHI and cloned into the same sites of the skeleton silencing vector. All antisense orientation PCR fragments were digested with SalI and SpeI and cloned into the same sites of silencing vector. The promoterless construct irintCrChII was made by removing a 791 bp HindIII- RsrII fragment that contains the enhanced 35S promoter. Additional *CrChII* constructs were generated by deleting various sequences from 35S:irintCrChII by digestion with appropriate restriction enzymes followed by religation of the plasmid, or were generated as intermediates during the cloning of the final 35S:irintCrChII construct. The antisense construct for silencing the endogenous *CrUrod* gene (GenBank accession number BE642240) was made by PCR amplifying a 413 bp *CrUrod* fragment from the appropriate *Ceratopteris* cDNA clone using the primers 5'CGTAACTAGTGTGCTGAGAAGCACCCCAGTTTC3' and 5'GACGGTCGACAAAGACTTGGGTGCCTGATG3'. The SalI and SpeI digested PCR fragments were then cloned into the SalI and SpeI sites of 35S:irint.

Quantitative PCR

RNA was isolated from pools of 80 green and 80 colorless gametophytes 1 week after bombardment with the 35S:irintCrChII plasmid using an RNeasy Plant Mini kit (Qiagen, CA). Total RNA was reverse transcribed using 200 U SuperscriptII reverse transcriptase (Invitrogen, CA) in the presence of d(T)15. Single-stranded cDNA was then used as a template in a real time PCR reaction using SYBR green PCR Master Mix from Applied Biosystems. Approximately 2 ng of cDNA, corresponding to the amount of RNA isolated from three gametophytes, was used as a template. The reactions were performed in an ABI Prism 7700 machine with real-time SYBR Green I detection using default parameters and the primers 5'AACGAGCAGGATGTGAAATG3' and 5' AACGAGCAGGATGTGAAATCG3'. Reactions were performed in quadruplicate for each template to assess standard deviation of threshold cycle (Ct) measurements of the amount of *CrChII* transcripts in the green and colorless samples. Each measurement was normalized to the amount of CrEF1 α (\bullet Ct) (GenBank accession number BE642078) transcript using the same amplification conditions and the primers 5' CAGACCAGTCGGAGCAAAGT3' and 5'TCCTGTGGGAAGGGTGGAA3'. The

fold decrease in abundance of CrChlI in green and colorless gametophytes is equal to $2 \bullet (\bullet Ct)$.

RESULTS AND DISCUSSION

Biostatic Introduction of a CrChlI Potential Hairpin-Forming Construct Suppresses the Endogenous *CrChlI* Gene

A *Ceratopteris* EST that encodes a putative protein that is >75% identical in amino acid sequence to the barley ChlI protein was initially selected to test silencing of endogenous genes in *Ceratopteris*. This gene, which is required for chlorophyll biosynthesis, was chosen because its inactivation results in an easily scorable phenotype. In barley, plants homozygous for ChlI mutations are yellow-seedling lethal, whereas heterozygotes are yellow-green (Hansson et al., 1999). Because the introduction of silencing transgenes by biostatic bombardment of tissues has been shown to trigger systemic gene silencing in angiosperms (Crete et al., 2001; Klahre et al., 2002; Palauqui and Balzergue, 1999; Voinnet et al., 1998), the same method was applied here. Hermaphroditic *Ceratopteris* gametophytes were initially bombarded with the 35S:irintCrChlI plasmid (Figure 1), which potentially drives the expression of a dsRNA hairpin with an intron loop that, if spliced, forms a 393 bp double-stranded CrChlI RNA molecule. Such an inverted repeat construct design is known to be very efficient in inducing gene silencing in flowering plants (Wesley et al., 2001). At the time of bombardment, 6d-old hermaphrodites are very small (~0.3 mm) and beginning to initiate a lateral, multicellular meristem (Figure 2A). The simple morphology of the hermaphroditic gametophyte, which consists of a single layer of cells dotted with egg-forming archegonia and sperm-forming antheridia, makes it simple to detect gene silencing in all cells of the gametophyte while its gender makes it possible to self-fertilize gametophytes and assess the transmission of a silenced phenotype to the sporophyte generation following fertilization. If silencing of the endogenous *CrChlI* gene spreads from the bombarded to neighboring cells, gametophytes co-bombarded with 35S:irintCrChlI and the GUS reporter pFF19G plasmid (Figure 1) were expected to develop a sectored or completely yellow-to-colorless prothallus but express GUS in only a single cell of the gametophyte that was present at the time of bombardment. As shown in Figure 2, hermaphrodites co-bombarded with both plasmids generated colorless cells throughout most of their prothalli, including the meristem, antheridia, and archegonia, yet displayed GUS activity only in older cells of the gametophyte. While GUS activity could be detected in several adjacent cells (Figure 2C), this is most likely due to diffusion of either the GUS protein or reaction product from an individual cell transiently expressing the *GUS* gene. Only one GUS positive sector was observed in >99% of all GUS positive gametophytes and more than two GUS sectors never observed. Of the gametophytes co-bombarded with the 35S:irintCrChlI and pFF19G and having a colorless or GUS positive phenotype, ~89% were both colorless and GUS positive 7 days after bombardment (Table 1), indicating a high efficiency of introduction of both plasmids into the same cell. These results demonstrate that the presence of this CrChlI expression construct in one or a small number of cells is sufficient to inactivate the endogenous *CrChlI* gene in almost all cells of the growing prothallus leading to a colorless phenotype, especially in cells formed after bombardment.

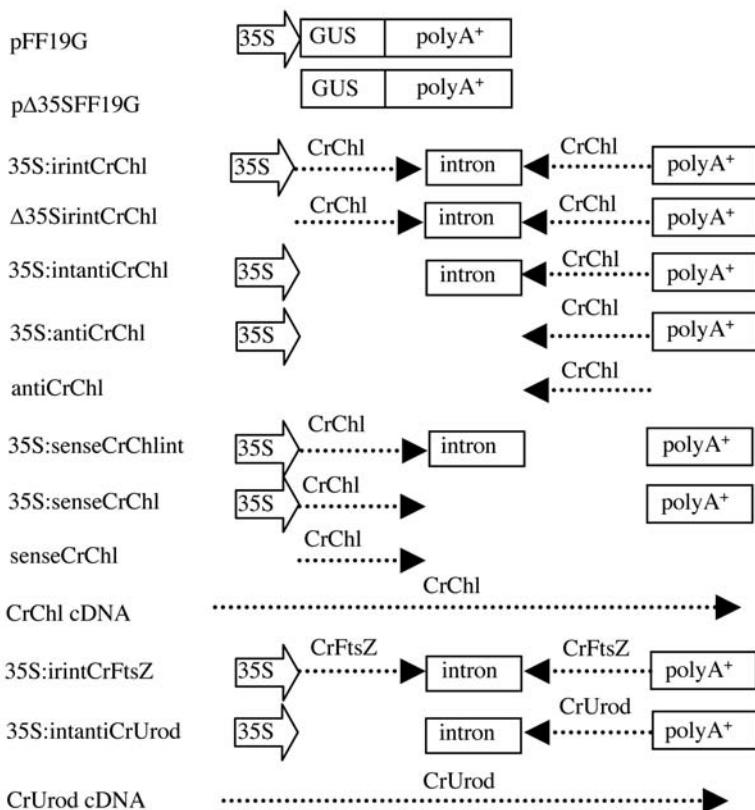


Figure 1. DNA constructs used to bombard gametophytes. Plasmid sequences are not shown.

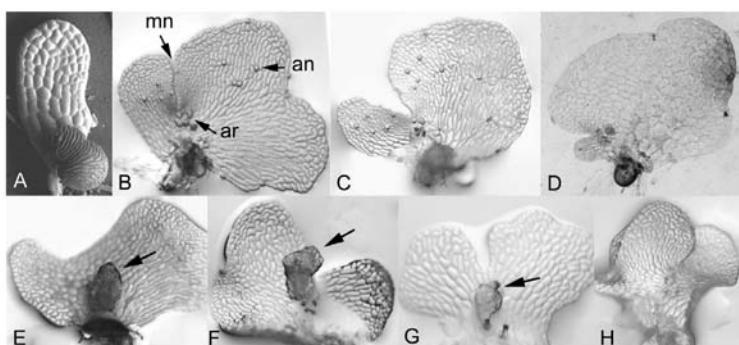


Figure 2. Gametophyte phenotypes after bombardment with 35S:irintCrChl. (A) Scanning electron microscopy of 6-day-old hermaphrodite initiating a lateral meristem. (B) Non-bombarded 15-day-old hermaphrodite showing meristem notch (mn), archegonia (ar) and antheridia (an). (C) A 15-day-old hermaphrodite 8d after bombardment with a large colorless sector including the meristem and gametangia. (D) Bombarded colorless gametophyte stained for GUS activity. (E)-(H) Phenotypes of self-fertilized colorless gametophytes ~3 weeks after bombardment. Embryos are indicated by arrows.

Table 1. Frequencies of gametophyte phenotypes after bombardment with various plasmids.

Plasmids introduced	% colorless & GUS positive ^{a,b}	% colorless & GUS negative ^a	% green & GUS positive ^a	n ^c
35S:irintCrChl+pFF 19G	88.8 (+/-1.36)d	6.8 (+/-0.46)d	4(+/-1.41)d	1263
irintCrChl+pFF 19G	81.7 (+/-0.87)	8.3 (+/-1.20)	10 (+/-1.53)	775
35S:intantiCrChl+pFF 19G	61.0 (+/-1.73)	11.0 (+/-1.73)	28.0 (+/-3.21)	564
35S:antiCrChl+pFF 19G	12.0 (+/-1.53)	0	88.0 (+/-1.53)	724
antiCrChl+pFF 19G	0.7 (+/-0.33)	0	99.3 (+/-0.33)	479
35S:senseCrChl+pFF 19G	37.7 (+/-1.12)	6.3 (+/-0.19)	56 (+/-1.00)	559
35S:senseCrChl+pFF 19G	2.3 (+/-0.88)	0	97.7 (+/-0.88)	761
senseCrChl+pFF 19G	1.7 (+/-0.33)	0	98.7 (+/-0.42)	677
CrChl cDNA+pFF 19G	13.7 (+/-0.88)	0	86.0 (+/-1.16)	563
Δ35SpFF 19G+ 35SinintCrChi	48.5 (+/-6.36)	51.0 (+/-5.66)	0.5 (+/-0.71)	470
pFFI9G	0 % with lesions & GUS positive ^{a,b}	0 % with lesions & GUS negative ^a	100 % without lesions & GUS positive	600 n ^c
CrUrod cDNA+pFF 19G	61.0 (+/-0.71)	20.2 (+/-0.69)	18.9 (+/-0.43)	774
35S:intantiCrUrod+pFF 19G	55.7 (+/-1.40)	13.5 (+/-1.61)	31.1 (+/- 2.25)	472

^aAll gametophytes were stained 7 days after shooting; each percent represents the average of three replicates; standard error is given in parentheses. The differences between treatments for all pairwise comparisons are statistically significant (≥ 3.4 . P < 0.0003) with the exception of 35S:senseCrChl+pFF 19G and senseCrChl+pFF 19G. ^ctotal number of gametophytes displaying a phenotype and scored. ^dthe average of 5 replicates.

The relative abundance of endogenous *CrChlII* transcripts in silenced gametophytes was assessed by quantitative real-time RT-PCR using cDNA generated from green and colorless gametophytes 6d after bombardment with the 35S:irintCrChlII plasmid. The abundance of *CrChlII* was normalized to the abundance of the *Ceratopteris* EF1 α gene in both gametophyte populations; the latter did not vary between the two populations (Figure 3A). A ~7-fold decrease in the abundance of *CrChlII* transcripts was observed in colorless gametophytes compared to green gametophytes (Figure 3B) indicating that the silencing mechanism interferes either with *CrChlII* transcription or the stability of the endogenous *CrChlII* transcripts.

Endogenous Gene Silencing in the Gametophyte is Both Reversible and Heritable

The heritability of the colorless phenotype of gametophytes bombarded with 35S:irintCrChlII was assessed by placing 195 colorless hermaphrodites individually into microtiter wells and adding water, allowing sperm to swim to, and fertilize the egg. After 3 weeks, four classes of gametophytes, illustrated in Figures 2E,2F,2G,2H, were observed, including those that turned green and produced a green embryo (32%; Figure 2E); those that remained colorless and produced a green embryo (26%; Figure 2F); those that remained colorless and produced a colorless embryo (7%) that did not develop beyond the stage illustrated in Figure 2G; and those that remained colorless and produced no embryo (35%; Figure 2H). The inability of many colorless gametophytes

to form sporophytes was not due to the lack of antheridia, motile sperm or archegonia as these structures appeared to develop normally, but may be due to the lack of sufficient photosynthate to support embryonic growth after fertilization. The relative proportions of each gametophyte class indicate that the silenced *CrChlII* gene is almost equally likely to reactivate or remain silenced in the gametophyte, but tends to reactivate upon or after fertilization. Reactivation of the endogenous *CrChlII* gene, reflected by the greening of colorless gametophytes, also indicates that *CrChlII* silencing is reversible and, therefore, epigenetic. Although, the percentage of colorless embryos that developed from colorless gametophytes was low, their presence demonstrates that the silenced state can be maintained through gametogenesis and fertilization and can be transmitted to the next sporophyte generation. Because colorless sporophytes did not develop beyond the embryonic stage, it was not possible to assess whether the colorless phenotype could be maintained through meiosis and the subsequent gametophyte generation.

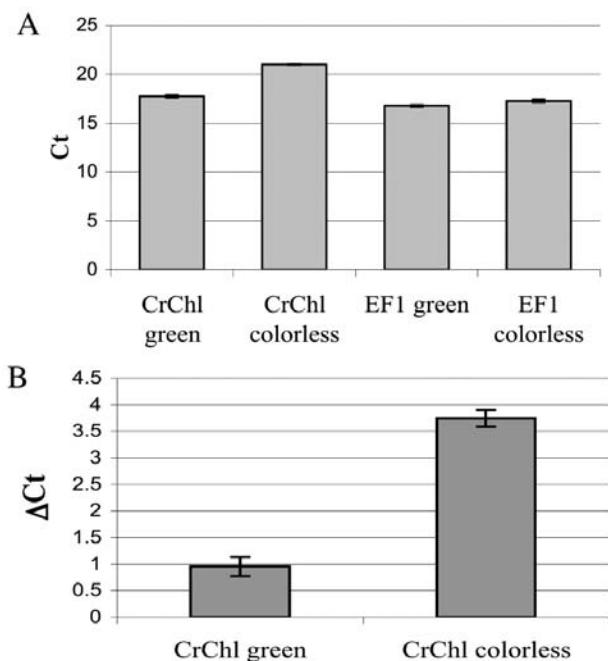


Figure 3. Results of quantitative PCR comparing *CrChlII* message abundance in green and colorless gametophytes. (A) Ct values using PCR templates derived from green and colorless gametophytes and primers specific for either the *CrChlII* or the *Ceratopteris EF1* gene. (B) ΔCt values of *CrChlII* abundance, normalized to *EF1α*, in green and colorless gametophytes.

Defining Transgene Elements Required for Gene Silencing

While a potential hair-pin-forming construct was found to be effective in promoting the systemic silencing of the endogenous *CrChlII* gene, the assembly of this construct involves a two-step cloning procedure, which is a limitation for high throughput

screening of gene function. For this reason, a variety of elements were removed from the inverted repeat-intron construct (illustrated in Figure 1) and their ability to induce gene silencing assessed.

To determine if silencing of the endogenous *CrChII* gene requires a promoter, the 35S promoter was deleted to give Δ 35S:irintCrChII; gametophytes were co-bombarded with this plasmid plus pFF19G. Surprisingly, the absence of the 35S promoter resulted in a high percentage of colorless gametophytes, with ~82% of the colorless and/or GUS-positive gametophytes displaying both phenotypes (Table 1), indicating that either the CrChII DNA sequences alone are sufficient for gene silencing, or that transcription of irintCrChII can occur and promote gene silencing in the absence of a plant promoter. To test the latter possibility, gametophytes were co-bombarded with two constructs: the 35S:irintCrChII plasmid as a marker for transformation; and the pFF19G construct from which the 35S promoter had been removed (to give Δ 35S:GUS). The latter construct was used as a marker for *in vivo* gene expression of the GUS gene in the absence of a plant promoter. About one-half of the colorless gametophytes that developed after bombardment were also GUS positive (Table 1), although the intensity of GUS staining was less than that observed in gametophytes bombarded with pFF19G (data not shown). No gametophytes bombarded with 35S:irint, a plasmid lacking the GUS reading frame (see materials and methods) stained positive for GUS activity (data not shown). Since, a promoterless GUS construct was found to promote GUS activity in the gametophyte, transcription of the *CrChII* silencing transgene is also likely to occur in the absence of a plant promoter. The systemic silencing of genes using promoterless silencer constructs similar to that described here has also been observed in angiosperms (Voinnet et al., 1998).

To address the contribution of the intron, the 35S promoter and a second copy of the transgene to *CrChII* silencing in the gametophyte, each was deleted from 35S:irintCrChII then co-bombarded with pFF19G into gametophytes. In comparing constructs with the *CrChII* fragment cloned in an antisense orientation, the proportion of colorless, GUS positive gametophytes decreased ~28% with the deletion of the sense copy of CrChII from 35S:irintCrChII (giving 35S:intantiCrChII) and decreased a further 49% when the intron was also deleted (giving 35S:antiCrChII) (Table 1). Almost no colorless gametophytes were observed when bombarded with the antiCrChII plasmid, which lacks the 35S promoter, intron, poly(A)+ signal and the second (sense) copy of *CrChII*. The same trend was observed when gametophytes were bombarded with the *CrChII* fragment cloned in the sense orientation (Table 1). These results indicate that the intron plays an important but unknown role in gene silencing and that the intron and the 35S promoter together enhance the efficiency of endogenous gene silencing. In angiosperms, an intron used to separate the two copies of a gene in an inverted orientation (as in the 35S:irintCrChII construct) also increases the efficiency of gene silencing, although how this occurs is unclear (Wesley et al., 2001).

To determine the efficiency of *CrChII* silencing by the unmodified CrChII cDNA from which the CrChII EST was originally derived, gametophytes were co-bombarded with pFF19G plus the CrChII cDNA plasmid (Figure 1). About 14% of the GUS positive gametophytes were colorless (Table 1), which is significantly greater than the

~1–2% observed following bombardment with the antiCrChlI or senseCrChlI plasmids that also lack a 35S promoter and intron sequences (Table 1). The differences in frequency observed are likely due to differences in the size of the *CrChlII* insert, as the senseCrChlII and antiCrChlII plasmids harbor a 393bp cDNA insert while the cDNA plasmid harbors a 1.2kb cDNA insert. A similar correlation between the length of the silencer transgene and silencing efficiency has been observed in angiosperms (Klahre et al., 2002).

***CrFtsZ* and *CrUrod* Gene Silencing in the Gametophyte**

Gametophytes were bombarded with two additional genes selected from a *Ceratopteris* EST library to further test the application of gene silencing as a reverse genetics tool in *Ceratopteris* gametophytes. The two chosen, *CrFtsZ* and *CrUrod*, encode putative proteins that are at least 70% identical in amino acid sequence to the *FtsZ* and *Urod* proteins in angiosperms. The *FtsZ* gene encodes a tubulin-like component of the filamentous plastoskeleton and the plastid division ring, which is essential for chloroplast division in plants (Osteryoung et al., 1998; Reski, 2002; Strepp et al., 1998). While *Arabidopsis* has four *FtsZ* genes representing two gene families that differ in the cellular targeting of the *FtsZ* protein (Osteryoung and McAndrew, 2001; Reski et al., 2002; Strepp et al., 1998), inactivation of a single member of either family results in cells with fewer and larger chloroplasts, a phenotype that should be easy to visualize in *Ceratopteris* gametophytes. Following bombardment with the inverted repeat-intron-forming *CrFtsZ* construct 35S:intCrFtsZ (Figure 1), hermaphrodites developed a prothallus with cells having larger and as few as three chloroplasts per cell compared to non-bombarded gametophytes with >50 chloroplasts per cell (Figures 4A,4B). The larger-and-fewer chloroplast phenotype was not evident until 3 weeks after bombardment, indicating that the phenotype requires new cell divisions to occur. Because GUS activity fades in time and is usually undetectable 3 weeks after bombardment, the frequencies of GUS positive and *CrFtsZ* silenced phenotypes were impossible to determine for this gene. Gametophytes with altered chloroplast morphology also formed shallow meristems and no functional archegonia or antheridia, preventing self-fertilization of the affected hermaphrodites (Figure 4C). Whether a functional *CrFtsZ* gene product is directly or indirectly involved in the development of meristem and gametangia is unknown, but this result suggests that *CrFtsZ* inactivation also can affect the differentiation of the gametangia and the organization of the multicellular meristem.

The *Urod* gene encodes an enzyme that catalyzes the decarboxylation of uroporphyrinogen III to coproporphyrinogen III, a precursor of chlorophyll and heme production in plants (Elder and Roberts, 1995; von Wettstein et al., 1995). Maize plants heterozygous for Les22, a mutation of *Urod* gene, develop discrete, tiny, colorless, or necrotic spots on their leaf blades, while homozygous plants are seedling lethal (Hu et al., 1998). The lesions in Les22 heterozygotes result from the accumulation of uroporphyrin, which is toxic to cells upon exposure to light. Inactivation of the endogenous *Ceratopteris CrUrod* gene was expected either to inhibit growth of the gametophyte or lead to gametophytes that developed necrotic lesions of unknown size. About 60% of the hermaphrodites co-bombarded with the construct 35S:intantiCrUrod (Figure 1)

and pFF19G developed one or two discrete necrotic lesions within 1 week of bombardment (Figures 4D,4E) and were also GUS positive (Table 1). Each lesion consisted of several adjacent colorless cells that eventually died. GUS staining could be detected only in cells of the mature gametophyte that were likely present at the time of bombardment. The GUS-positive cells may have survived following bombardment with the 35S:intantiCrUrod construct because each had produced sufficient *CrUrod* protein prior to bombardment and were thus able to catalyze the decarboxylation of uroporphyrin. Of the gametophytes that developed lesions or were GUS positive following bombardment with the original *CrUrod* cDNA plasmid plus pFF19G, 61% were both GUS positive and developed lesions (Table 1), a percentage similar to that obtained when gametophytes were bombarded with the intron-antisense construct 35S:intantiCrUrod plus pFF19G.

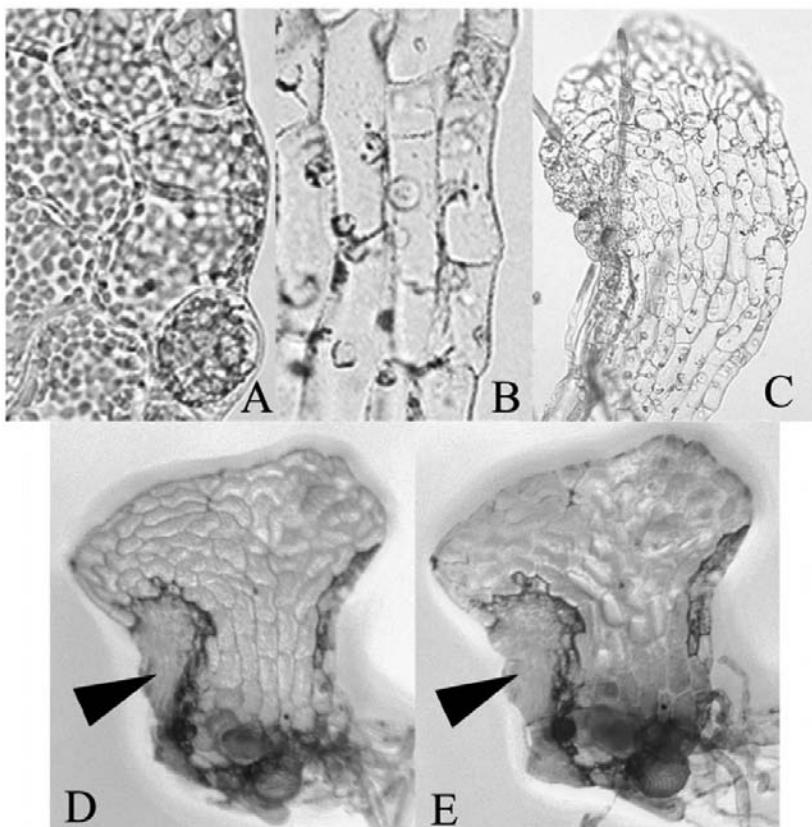


Figure 4. Phenotypes of sporophytes and gametophytes bombarded with various plasmids. (A) Chloroplasts of non-bombarded gametophyte. (B) Chloroplasts of gametophyte ~3 weeks after bombardment with the 35S:intCrFtsZ plasmid. (C) Hermaphrodite prothallus ~3 weeks after bombardment with 35S:intCrFtsZ. (D), (E) Gametophyte 7d after bombardment with 35S:intantiCrUrod before (D) and after (E) staining for GUS activity. The lesion associated with the inactivation of the *CrUrod* gene is indicated by the arrow.

Application of Gene Silencing for High Throughput Analyses of Gene Function

Our results show that DNA vector-based gene silencing in the fern gametophyte is effective in systemically inactivating targeted genes and results in a gene-specific mutant phenotype that is apparent throughout most of the gametophyte. While, PTGS and RNAi have been shown to silence genes in flowering plants, the fern gametophyte offers many technical advantages as a system for studying gene function. One is that the method for generating a silenced phenotype is rapid and robust, requiring only a cDNA plasmid clone, a biolistic apparatus and a plate of 6-day-old *Ceratopteris* gametophytes. Since each gametophyte matures quickly and is not much larger than a yeast colony, large numbers (>100) of transformed and gene-silenced gametophytes can be generated with each bombardment and identified 1 to 2 weeks following bombardment. For this reason, it is feasible to bombard and test a minimum of 500 different cDNA clones within a 2-week period, making this method suitable for high throughput testing of gene function. Like an onion epidermal peel, each hermaphroditic gametophyte is a single cell layer thick, making phenotypes easy to observe without interference from adjacent cell layers. Yet, another technical advantage of the fern gametophyte is that lethal phenotypes associated with the inactivation of essential genes, such as the *CrUrod* gene, can be easily recognized because the gametophyte is not bombarded until after it has germinated from the spore. By varying the age of the gametophyte at the time of bombardment, one can control when during development gene silencing occurs.

The method described will be useful for identifying the biological functions of genes that are involved in post-germination processes in the gametophyte, including meristem development, archegonia, antheridia, sperm and egg differentiation, sperm chemotaxis, fertilization, and early embryo development. Most of these processes are hallmarks of the plant gametophyte that are difficult to study in flowering plants because their gametophytes develop embedded within sporophytic tissues of the flower. Because the gold particles cannot penetrate the spore coat, genes that control processes that only occur prior to the emergence of the prothallus cannot be silenced using the method described in this study. Germination, the initial cell division, establishment of polarity as well as sex determinations are among the processes that occur during these early days of growth (Banks, 1999). Recently, an RNAi method first described in *Marsilea vestita* (Klink and Wolniak, 2000, 2001; Tsai and Wolniak, 2001) was applied to *Ceratopteris* spores (Stout et al., 2003). In this study, spores were incubated in the presence of *in vitro* synthesized dsRNA corresponding to each of five genes known to be expressed in germinating spores. Although phenotypes associated with each treatment were not reported, mixing spores with dsRNA was shown to reduce steady state mRNA levels after a 24 hr incubation period. Should this method prove fruitful in generating informative phenotypes, the combination of the two techniques will allow the examination of gene function throughout all stages of gametophyte development.

CONCLUSION

The expression of endogenous genes in the fern gametophyte can be systemically suppressed by introducing a transgene construct into single cells of the gametophyte by

particle bombardment. The DNA constructs that promote the expression of potential hairpin-intron loop or antisense transcripts are the most effective, although constructs having a promoterless gene are sufficient in inducing systemic gene silencing in the gametophyte. The silencing of the three genes tested (*CrChlI*, *CrFtsZ*, and *CrUrod*) resulted in phenotypes that mimic mutant phenotypes in other plants. The method developed is a useful reverse genetic tool to quickly and efficiently screen the functions of gametophytically expressed genes.

KEYWORDS

- **Feasibility**
- **Homosporous fern**
- **Meristems**
- **Sporophytic tissues**

AUTHORS' CONTRIBUTIONS

George Rutherford made the constructs. Milos Tanurdzic carried out the plant transformations and molecular analysis of the transgenic plants. George Rutherford, Milos Tanurdzic, and Jo Ann Banks participated in the design of this study. Jo Ann Banks conceived the study, participated in its design and drafted the manuscript together with Mitsuyasu Hasebe and Milos Tanurdzic. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Luke Gumaelius, Drew Schultz, Amber Hopf, and Erlinda Embuscado for technical assistance, insightful discussions and help with harvesting plants and spores, and thank Kenzo Nakamura for graciously providing the pIG121Hm plasmid. We also thank Dr. Stan Roux for providing the cDNA library from which ESTs were generated at the Purdue University Genomics Facility. This work was supported by the National Science Foundation and the Purdue Agricultural Research Programs.

Chapter 6

Distribution and Dynamics of Hayscented Fern

Songlin Fei, Peter J. Gould, Melanie J. Kaeser, and Kim C. Steiner

INTRODUCTION

The distribution and dynamics of hayscented fern were examined as part of a large-scale study of oak regeneration in Pennsylvania. The study included 69 stands covering 3,333 acres in two physiographic provinces. Hayscented fern was more widely distributed and occurred at higher densities in the Allegheny Plateau physiographic provinces versus the Ridge and Valley region. After partial overstory removal, the density and distribution of fern increased in stands that were not treated with herbicide. Herbicide treatments successfully reduced fern densities and created a “window of opportunity” about 4 years post-harvest for the establishment of regeneration. In the mixed hardwood forests of Pennsylvania, dense groundcover of hayscented fern (*Dennstaedtia punctilobula* (Michx.) Moore) and other herbaceous species interfere with the development of advance regeneration of oaks (Horsley et al., 1992; Steiner and Joyce 1999). Regeneration of tree seedlings may be adversely affected due to hayscented fern’s ability to influence the availability of light, nutrients, and water (Engelman and Nyland, 2006; George and Bazzaz, 1999; Lyons and Sharpe, 1996; Messier et al., 1989). Hayscented fern has been classified as a competitor species because of its ability to respond aggressively to sudden resource availability by way of vegetation expansion of rhizomes and sexual reproduction through spore dispersal (Groninger and McCormick, 1991; Hughes and Fahey, 1991).

Hayscented fern’s ability to occupy a variety of habitats makes it a dominant component of the understory in various regions throughout Pennsylvania. Small-scale studies have shown that hayscented fern is a large component of the groundcover in relatively productive oak stands (Allen and Bowersox, 1989) and on poorly drained soils (Marquis, 1979). There is limited information about the distribution of hayscented fern among other forest community types, however, previous studies have focused on only one or several stands and failed to look at landscape-level patterns of fern colonization. Yet in light of the oak regeneration problem facing forest managers in Pennsylvania, the distribution and dynamics of this competing species need to better understood (McWilliams et al., 1995). This chapter focuses on the distribution of hayscented fern at a landscape scale in Pennsylvania, and articulates its response to overstory removal and herbicide treatments.

STUDY AREA

The study area crosses the Allegheny Plateau and Ridge and Valley physiographic provinces of Pennsylvania (Figure 1). Soils in both provinces in stands represented

in this study are derived from sandstone, siltstone, and shale and are typically well drained and support moderately productive forests. Stand elevations range from 250 m above sea level in the Ridge and Valley province to 700 m on the Allegheny Plateau. Precipitation, temperature, and length of growing season vary with latitude and topography. Mean annual precipitation ranges from 960 to 1,070 mm and mean annual temperature ranges from 8 to 11°C. The growing season ranges from 120 to 140 days in the northwest, and 140 to 180 days in the southeast.

MATERIALS AND METHODS

Data Collection

Field measurements in 69 stands in Pennsylvania were performed on a total area of 3,333 acres of state land during 1996–2006. Thirty-one stands were located on the Allegheny Plateau and 38 stands in the Ridge and Valley physiographic provinces. All the stands were measured 1 year prior to silvicultural treatment, 52 stands were re-measured 1 year after treatment, 43 stands were re-measured 4 years after treatment, and 17 stands were re-measured 7 years after treatment. Depending on stand area, 15–30 acre permanent plots (26.3 ft radius) were systematically installed in a square grid to represent the whole stand. Four permanent milacre subplots (3.72 ft radius), were also set up within each plot. On each subplot, percentage cover of hayscented fern was estimated. In total, we measured 7,481 subplots before treatment, and re-measured 5,163 subplots 1 year after treatment, 4,240 subplots 4 years after treatment, and 1,715 subplots 7 years after treatment.

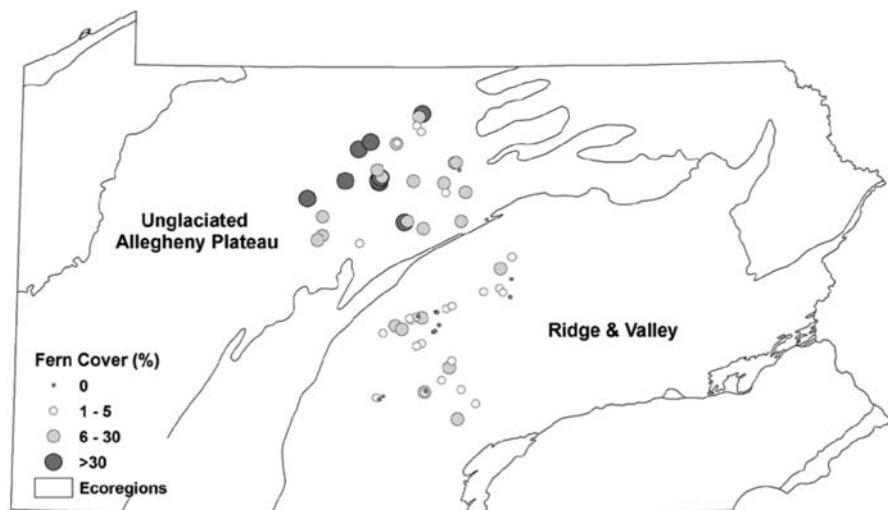


Figure 1. Distribution of average hayscented fern cover in 69 Pennsylvania stands before harvest.

Treatments

Treatments were based upon the forester's management objectives for each stand and were not experimentally controlled. The primary objective of the treatments was to

establish and/or release desirable regeneration. Overstory treatments ranged from 20 to 80% basal area reduction at the stand level. Stands were treated with herbicide, if hayscented fern densities appeared likely to inhibit regeneration as judged by the staff forester in charge. OUST® (sulfometuron methyl) or an ACCORD®/OUST® mix was applied to dense hayscented fern at various rates depending on fern density. Herbicide was applied before the first re-measurement in 22 of the 31 stands in the Allegheny Plateau and four of the 38 stands in the Ridge and Valley physiographic provinces.

Data Analysis

For each stand, average fern cover, percentage of subplots with fern, and percentage of subplots with over 30% fern cover were calculated. Percentage cover by hayscented fern was divided into four classes—none, low (1–5% cover), moderate (6–30%), and heavy (over 30%). The heavy class reflects the level of competing vegetation that was considered problematic by Marquis (1994). Fern cover class transition rates from before harvest to 1, 4, and 7 years after harvest were calculated. Transition rates were calculated as the percentage of plots in a given pre-treatment cover class that fell into each cover class 1, 4, or 7 years after harvest. Consequently, transition rates sum to 100% for each pre-treatment cover class. Subplots in the Ridge and Valley and Allegheny Plateau physiographic provinces were combined, but transition rates were calculated separately for herbicided and non-herbicided subplots.

DISCUSSION

Competition between tree seedlings and hayscented fern has been identified as a factor contributing to the oak regeneration problem in Pennsylvania (McWilliams et al., 1995). Our results indicate that hayscented fern is more prevalent on the Allegheny Plateau than in the Ridge and Valley.

Without herbicide treatments, hayscented fern became more widely distributed and increased in density following partial overstory removal. Most of the stands in the Ridge and Valley were not treated with herbicide because hayscented fern was either absent or did not occur at problematic levels before harvest. After harvest, both the distribution and density of hayscented fern increased as a result of stand disturbance associated with overstory removal. However, hayscented fern did not reach problematic densities after 7 years in the Ridge and Valley because initial densities were low. In contrast, the distribution of fern in this region increased dramatically. It is uncertain whether this increase in distribution is a precursor to a greater fern problem in the future, or a short-lived phenomenon associated with overstory disturbance. The expansion of fern has the greatest potential to become a problem in stands that fail to regenerate quickly.

Herbicide treatments created a “window of opportunity” for the establishment of seedlings following partial overstory removals. One year after treatment, fern cover was reduced to non-problematic levels on most of the subplots with initially heavy fern cover. After 7 years, fern cover had partially recovered from the herbicide treatments, but less than one-third of the initially heavy fern cover plots had returned to that condition. Given that oaks are expected to produce large acorn crops every 4–6 years,

the window of opportunity provides a chance to capture a regeneration cohort. A heavy mast year shortly after the herbicide treatment, or initially high levels of advance regeneration, should result in the successful establishment of oak regeneration (Johnson et al., 1989). The rate of fern recovery between two consecutive surveys accelerated after treatments and a complete recovery is likely (Horsley et al., 1992). If a mast year does not occur in the several years following the herbicide treatment, the window of opportunity will likely close as fern recovery accelerates (Figure 2).

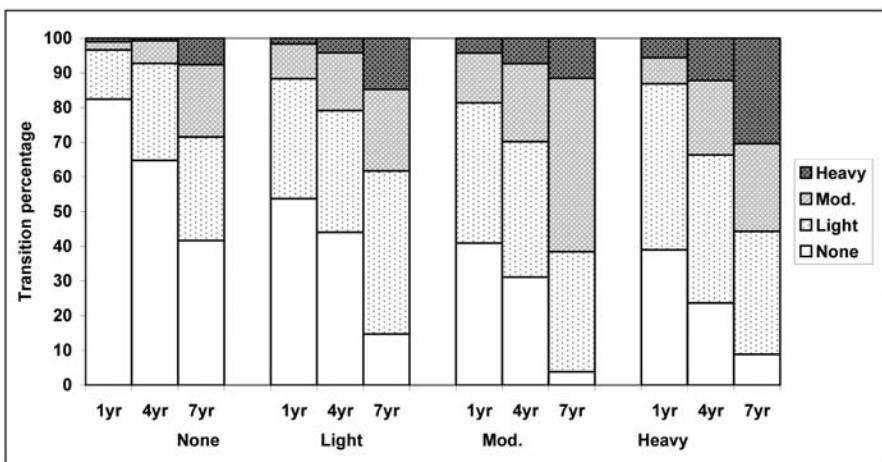


Figure 2. Hayscented fern cover class transition percentages from 1 year before harvest to 1, 4, and 7 year after harvest without herbicide treatment (e.g., 78% of subplots classified as none cover class before harvest retained in the same cover class and 20% transited to the light cover class 1 year after harvest).

Caution is needed when applying the findings to forest management practices. In our stands, we have different intensities of overstory removal, ranging from 20 to 80%. Although over 70% of the stands had at least 50% of their overstory removed, there might be a relationship between the rate or extent of fern spread after cutting, and the cutting intensity. Other factors such as forest community types might also influence the findings. Additional analyzes of those factors may be needed in future studies.

RESULTS

Average hayscented fern cover percentage ranged before treatment for all 69 stands were plotted on the Pennsylvania physiographic provinces map (Figure 1). Generally, stands on the Allegheny Plateau had a higher percentage of hayscented fern cover than stands in Ridge and Valley. However, there were two stands on the Allegheny Plateau that had no fern cover, and seven stands in the Ridge and Valley had moderate fern cover.

Before overstory removal, Ridge and Valley stands had lower fern cover, frequency of occurrence, and frequency of heavy hayscented fern than the Allegheny

Plateau stands (Table 1). Average percentage of hayscented fern cover for the Ridge and Valley stands remained nearly unchanged from 1 year before treatment (3.0%) to 1 year after treatment (2.8%), but increased 4 and 7 years after treatment (4.8 and 9.6%, respectively). Average frequency of subplots with the presence of hayscented fern in the Ridge and Valley showed a pronounced monotonic increase during the survey period. One year before treatment, 11.6% of subplots had fern cover, increasing to 27.1% 1 year after treatment, to 46.1% 4 years after treatment, and to 55.5% 7 years after treatment. Changes in the percentage of plots with heavy fern cover occurred in a pattern similar to the average percentage of hayscented fern cover.

The Allegheny Plateau stands had an average of 19.6% fern cover 1 year before treatment. Fern cover dropped to 7.8% 1 year after treatment and 7.4% 4 years after treatment (Table 1). The decrease in fern cover was likely due to herbicide application as 58% of the stands were treated with herbicide after the first measurement. Fern cover recovered to 11.2% 7 years after treatments. Percentage of plots with fern cover decreased slightly from 52.0% 1 year before treatment to 46.1% 1 year after treatment, but recovered and surpassed 4 and 7 years after treatment (57.1 and 62.4%, respectively). Changes in percentage of plots with heavy fern cover had a similar pattern to fern cover percentage.

Table 1. Stand-level average of hayscented fern by physiographic provinces and time of assessment.

	Time	Fern Cover (%)	Fern Freq. ¹ (%)	Dense Fern Freq. ² (%)
Ridge and Valley	1yr-pre (n = 38)	3.0	11.6	4.2
	1yr-post (n = 29)	2.8	27.1	3.7
	4yr-post (n = 25)	4.8	46.1	5.5
	7yr-post (n = 13)	9.6	55.5	13.6
Allegheny Plateau	1yr-pre (n = 31)	19.6	52.0	28.3
	1yr-post (n = 23)	7.8	46.1	9.5
	4yr-post (n = 18)	7.4	57.1	9.7
	7yr-post (n = 4)	11.2	62.4	14.0

¹percentage of subplots with hayscented fern cover.

²percentage of subplots with heavy hayscented fern cover (>30%).

Overstory removals without herbicide treatment resulted in an increase of fern after harvest throughout the survey period (Figure 3). About 21.6% of subplots that had no fern cover before harvest moved into other fern cover classes 1 year after harvest. Four and 7 years after harvest, the percentage increased to 39.2 and 44.9, respectively. A considerable proportion of subplots that started in the light cover class moved into the moderate and heavy cover classes. Some reduction of fern cover density was observed on subplots that were classified as moderate or heavy fern cover classes. For subplots that had moderate fern cover before harvest, 23.9% fell into the light or no fern cover classes 1 year after harvest, and the percentage increased to 39.3 and 43.9 in

4 and 7 years after harvest, respectively. However, more than one-seventh of moderate fern cover subplots transitioned into heavy fern cover class throughout the post-harvest survey period. Less than 5% of the subplots that were classified in the heavy fern class fell into the no cover class after harvest. The overall trend is that stands without herbicide treatment after harvest had a spread of fern cover and increased density for subplots that had low or no fern cover before overstory removal.

Herbicide treatments in conjunction with overstory removals successfully reduced the cover of hayscented fern to non-problematic levels across all density classes (Figure 2). One year after treatment, fern cover was absent or light on 86.9% of the subplots that started with heavy fern cover. Among all of the pre-treatment cover classes, fern cover was light or absent in over 80% of the subplots. Four years after treatment, the recovery of hayscented fern from the herbicide treatment was evident. Furthermore, the pre-treatment cover class influenced the level to which fern abundance recovered. More than one-tenth of the subplots that had heavy fern cover before harvest returned to that class 4 years after harvest.

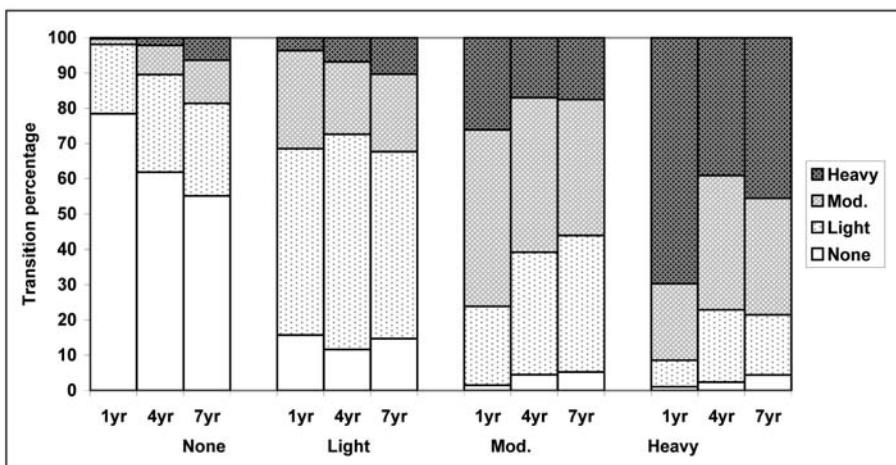


Figure 3. Hayscented fern cover class transition percentages from 1 year before harvest to 1, 4, and 7 year after harvest with herbicide treatment (e.g., of all subplots classified as heavy cover class before harvest, 39%, 24%, and 9% transited to the none cover class 1, 4, and 7 year after harvest, respectively).

An additional 7.3% of the subplots that had moderate pretreatment fern cover reached heavy fern cover by 4 years after harvest. The recovery of hayscented fern from herbicide treatment was even more pronounced 7 years after harvest. Over three-fifths of the subplots that had moderate fern cover before harvest returned to the moderate class or increased to the heavy class, and over one-half of the subplots that had heavy fern cover before harvest recovered to the moderate or heavy class 7 years after harvest.

CONCLUSION

Hayscented fern is more prevalent on the Allegheny Plateau than in the Ridge and Valley physiographic provinces of Pennsylvania. It became more widely distributed and increased in density after overstory removal in stands that were not treated with herbicide. Herbicide treatments created a window of opportunity for the establishment of regeneration. After 4 years, fern recovery was accelerated and the opportunity to establish regeneration appeared to decline.

KEYWORDS

- **Allegheny Plateau stands**
- **Hayscented fern**
- **Overstory removal**
- **Overstory treatments**

Chapter 7

Glossopterid Seed Ferns from the Late Permian, Antarctica

E. L. Taylor, T. N. Taylor, and P. E. Ryberg

INTRODUCTION

The *Glossopteridales* are an extinct group of seed ferns that dominated Gondwana during the Permian. The strap-shaped leaves of *Glossopteris* are widespread and provided early evidence of continental drift, but reproductive organs of this group have remained enigmatic since they were first described in the 1950s. Based on compression-impression fossils, there are at least two basic types of ovulate (seed-bearing) organs. In one form, numerous ovules are borne on the surface of a leaf-like megasporophyll (e.g., *Plumsteadia*, *Dictyopteridium*), while the second type consists of stalked, apparently uniovulate cupules borne on a branching system (e.g., *Lidgettonia*). More than 30 genera of seed-bearing organs have been described, many based on compressions or, more commonly, impression fossils, in which it is often impossible to discern the three-dimensional morphology or attachment of these structures. Both morphological types have now been found in permineralized peat from Skaar Ridge, Central Transantarctic Mountains (CTM), and provide some of the first anatomical detail of the diversity of reproductive organs of *Glossopteris*. The multiovulate organ is small (6 mm wide), with ovules borne on the upper (adaxial) surface; the uniovulate structure consists of four stalked cupules, each containing a single ovule about 2 mm long. These two forms provide important information about the diversity of the glossopterids and their status as a natural group of seed plants.

The glossopterid seed ferns (pteridosperms) have occupied both a prominent and perplexing position in the history of paleobotany, prominent because the spathulate *Glossopteris* leaves with complex venation represented some of the earliest evidence for the existence of the supercontinent Gondwana and subsequent movement of continental plates, and perplexing because the phylogenetic position of the group has continued to remain equivocal. While *Glossopteris* has been used as an index fossil for the Permian, it has also been described from the Early Triassic of India (Pant and Pant, 1987; Shah, 2000) and from the Fremouw Formation in Antarctica (McManus et al., 2002), which has previously been considered to be Triassic (Collinson et al., 2006). McLoughlin (1993) noted that while some glossopterid reproductive organs may be used in biostratigraphy, leaves, and other vegetative organs are not useful for correlation due to problems in delimiting species, especially among impression floras. Permineralized leaves have been described from Antarctica (Pigg, 1990) and Australia (Gould and Delevoryas, 1977; Pigg and McLoughlin, 1997), along with evidence of their attachment to the parent plant (Pigg and Taylor, 1993). The distinctive roots of

the *Glossopteris* plant, Vertebraria (e.g., Neish et al., 1993), can be identified in both compression-impression and permineralized specimens. *Glossopteris* remains have been found in a number of types of depositional environments in Antarctica (Cúneo et al., 1993).

The ovule-bearing structures of the glossopterids continue to be poorly understood; many are known only from impression fossils, making interpretation of three-dimensional morphology difficult. Numerous taxa have been erected based on size, shape, number of ovules (often only ovule scars), and various other features (see, e.g., McLoughlin, 1990a, 1990b; Pant, 1977; Rigby, 1978; Surange and Chandra, 1975). Attachment of reproductive organs to the parent plant is based on impression specimens only and is controversial.

All glossopterid ovulate reproductive organs suggest a basic morphological organization in which a vegetative leaf subtends a structure that bears ovules or seeds. Historically, this seed-bearing unit has been termed a megasporophyll, cupule, or other, less specific morphological entities (e.g., fertilizer, capitulum, cladode); the latter terms are no longer in use. The reproductive structure is generally regarded as an axillary shoot that has become reduced and flattened evolutionarily (e.g., Doyle, 2006), much like the secondary shoot of the cordaites (Schopf, 1976). In some cases, this structure appears to be adnate to a typical *Glossopteris* leaf, although this is difficult to confirm based only on impressions. Gould and Delevoryas (1977) provided the first description of a structurally preserved, seed-bearing organ from the Bowen Basin, Australia. It showed definitively that the ovules were borne on a leaflike megasporophyll, thus confirming the assignment of these plants to the pteridosperms. Another permineralized megasporophyll (Figures 1 and 2) was described from the Permian of Antarctica (Taylor and Taylor, 1992); it provided additional evidence that the ovulate organ was a megasporophyll, and showed that the seeds were borne on the adaxial (upper) surface of the megasporophyll (Figure 2), based on the orientation of the vascular bundles. The presence of bisaccate, striate pollen 10th International Symposium on Antarctic Earth Sciences grains in the pollen chambers (Figure 3) confirmed the glossopterid affinities. Nishida et al. (2007) have recently confirmed that the ovules in the Bowen Basin specimens were borne on the anatomically adaxial surface of the megasporophyll. The exact relationship between the megasporophyll and the subtending leaf, however, has yet to be determined, and is an important piece of the puzzle that is needed to accurately homologize the ovule-bearing structures with other seed plants.

An examination of the more than 30 ovulate organ morphogenera that have been named to date suggests that there are two basic types, as previously suggested by Surange and Chandra (1975). The multiovulate type consists of a flattened, leaf-like megasporophyll, e.g., *Plumsteadia* (McLoughlin, 1990a; Rigby, 1962) and the cupulate type consists of a branching structure, e.g., *Denkania* (Surange and Chandra, 1973) or *Lidgettonia* (Thomas, 1958). Both structures are subtended by a vegetative leaf and presumed to represent ovulate organs, although ovules are often not preserved. The multiovulate types often show only scars where the ovules would be attached (e.g., McLoughlin, 1990b). In impression fossils with no cuticle preserved (e.g., Anderson and Anderson, 1985), it is often difficult to determine whether structures represent

ovules or ovule scars. The preliminary description of an ovule-bearing structure in this chapter represents the first to provide anatomical details of the cupulate type of reproductive organ.

MATERIAL AND METHODS

The specimen (#13677) is from a silicified peat in the Upper Buckley Formation, Skaar Ridge, Beardmore Glacier region, CTM (84° 49' 15.8" S, 163° 20' 18.9" E, Buckley Island Quadrangle, Barrett and Elliot, 1973; Taylor et al., 1989) and is considered Late Permian, based on floral content and palynology (Farabee et al., 1991). Acetate peels were made of the surface after etching in hydrofluoric acid (Galtier and Phillips, 1999); peels were mounted on slides for light microscopy and digital photography. All materials are housed in the KU Natural History Museum, Division of Paleobotany (<http://paleobotany.bio.ku.edu/PaleoCollections.htm>).

DISCUSSION

The permineralized specimens from Antarctica confirm the existence of two morphological types of ovule-bearing reproductive organs in the glossopterid seed ferns, an idea which was initially proposed based on compression-impression remains only (see, e.g., Surange and Chandra, 1975). They also provide the first definitive evidence of ovules borne within cupules, since the impression taxa described to date have no cellular details preserved. One type consists of a flattened leaf bearing a few too many seeds on the adaxial surface, the multiovulate type (e.g., Gould and Delevoryas, 1977; Nishida et al., 2007; Taylor and Taylor, 1992), and the second type encompasses uniovulate cupules borne terminally on elongate stalks, which themselves are part of a branching system (cupulate type). Both morphotypes contain orthotropous ovules and integumentary structures vary among taxa.

Compression-impression material suggests that both types of ovulate structures were borne in some relation to a vegetative leaf, either along the petiole on the upper (adaxial) surface of the leaf or in an axillary position. Only a very few specimens have been described that show glossopterid ovulate organs compressed laterally, so that the relation of the megasporophyll and the subtending vegetative leaf can be seen clearly. One of these, described by Pant and Singh (1974), shows a stalk terminated by an enlarged, flattened megasporophyll bearing numerous ovules on the surface. The stalk is attached in the axil of a typical vegetative leaf of *G. taenioides* with the ovules facing the stem (i.e., apparently on the adaxial surface of the megasporophyll). Based on this and other observations, Pant hypothesized in several publications (e.g., Pant, 1977; Pant and Singh, 1974) that all the multi-ovulate types bore their ovules on the adaxial (upper) surface of the megasporophyll, which was then attached to the stem in the axil of a vegetative leaf. Holmes (1973) also illustrated a specimen of *Austroglossa walkomii* with an attached axillary stalk bearing a megasporophyll and ovules. The specimen is somewhat crushed and may be twisted, but he hypothesized, based on this fossil, that the ovules were borne on the abaxial (lower) surface of the megasporophyll. In this scenario, the ovules would face the vegetative leaf and be appressed to it. Anatomical studies have now clearly demonstrated (Nishida et al., 2007; Taylor and

Taylor, 1992) that the ovules were attached to the adaxial (upper) surface of the megasporophyll, based on the orientation of the vascular bundles in the structure.

Several authors have reconstructed the glossopterid megasporophyll with the ovules attached to the abaxial (lower) surface of the megasporophyll (e.g., Retallack and Dilcher, 1981; Schopf, 1976) and these reconstructions have been widely copied. These interpretations, however, have relied heavily on theoretical considerations necessary to evolve the angiosperm carpel from the glossopterid megasporophyll (e.g., Doyle, 2006). Information from compressed fossils does suggest that the ovules probably faced the surface of the subtending *Glossopteris* leaf (e.g., McLoughlin, 1990a), that is, they faced abaxially, but the permineralized structures from Antarctica and Australia illustrate that anatomically, the ovules were borne on the adaxial surface of the megasporophyll. This result would imply that either the peduncle of the megasporophyll was twisted at the base, or perhaps the vascular bundles twisted, either in the stem or in the peduncle, a situation known to occur in cycad leaves (Hermsen et al., 2007; Le Goc, 1914).

Glossopterid Reconstructions and Phylogenetic Relationships

In attempting to elucidate the phylogenetic position of the Glossopteridales, many authors have considered the group as a single terminal clade in phylogenetic analyses (e.g., Doyle, 2006; Hilton and Bateman, 2006; Nixon et al., 1994). To date, there is no whole-plant reconstruction of a *Glossopteris* plant based on organic attachment or anatomical evidence, so the “*Glossopteris*” plant used in phylogenetic analyses has been a composite of morphotaxa based on dispersed organs, which masks the potential diversity in the group. A number of impression-compression leaf specimens have been described attached to axes (see references in Pigg and Taylor, 1993). Taylor and Taylor (1992) attributed the small megasporophyll from Antarctica to the same plant that bore *G. schopfii* (Pigg, 1990) leaves, based on similar anatomy of the vascular bundles. Pigg and Taylor (1993) described leaves of *G. skaarensis*, also from the Antarctic permineralized peat, attached to woody stems assigned to the morphotaxon *Araucarioxylon*. Pigg and Nishida (2006) reconstructed three “whole plants” of *Glossopteris* based on permineralized specimens, one with *G. homevalensis* leaves from Australia, and two based on *G. schopfii* and *G. skaarensis* leaves from Antarctica; these leaf species also occur in Australia. Immature ovulate organs have been found attached to axes with *G. homevalensis* leaves, but these cannot be related to the mature ovulate structure from the Homevale locality (Nishida et al., 2007; Pigg and Nishida, 2006). With the exception of ovulate organ of Taylor and Taylor (1992) and the immature structure, all reconstructions to date consist of vegetative parts only. Any attribution of reproductive organs has been based on common occurrence at the same locality, and pollen organs have not been described from these sites. Despite a leaf morphotype that is relatively uniform throughout Gondwana in the Permian, the reproductive organs, in particular the ovule-bearing parts of the plant that bore *Glossopteris* leaves, are both morphologically and anatomically diverse. It should come as no surprise, therefore, that the glossopterids represent a heterogeneous group of Late Paleozoic seed plants, and, like the so-called seed ferns in the Mesozoic, constitute a major challenge in understanding gymnosperm phylogeny.

RESULTS

The new cupulate organ consists of four small, uniovulate cupules that appear to be arranged in a C-shaped pattern. Each cupule is approximately 3 mm long and 1 mm in diameter (Figures 4, 5) and contains a single ovule; the cupule extends distally beyond the level of the micropyle. Ovules are sessile, orthotropous, and approximately 2 mm long. The multilayered integument has short wings at the apex, and bilaterally symmetrical ovules (1.2–2.0 mm wide in the primary plane). These features distinguish this ovule from the similar-sized, obovoid ovules of *Chaonostoma* (Klavins et al., 2001) previously described from this same locality. The cupules are attached to the distal end of an axial system; each axis undergoes two dichotomies resulting in a total of four uniovulate cupules in each unit (Figure 5). One dichotomy, resulting in two terminal cupules, is shown in Figure 5. The most proximal end of the cupule axis has a C-shaped vascular strand. This cupulate structure, like the megasporophyll previously described from Skaar Ridge, is not preserved attached to the parent plant or in relationship to a vegetative leaf, and thus the homologies of the fertile organ continue to remain imprecise.

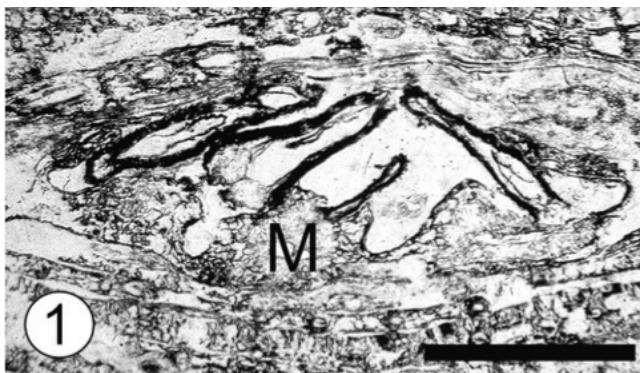


Figure 1. Cross section of multiovulate megasporophyll (M) with several ovules attached to adaxial surface. One ovule (just left of M) has broken off and is reversed, with its micropyle pointing back towards the megasporophyll. Two ovules (above M and, right) are still attached (see Taylor and Taylor, 1992 for further explanation); Scale bar = 2 mm.

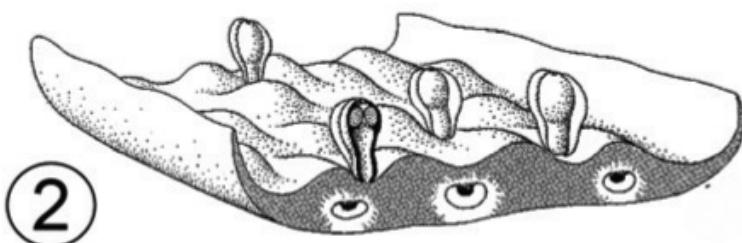


Figure 2. Suggested reconstruction of megasporophyll in Figure 1 (redrawn from Taylor and Taylor, 1992).

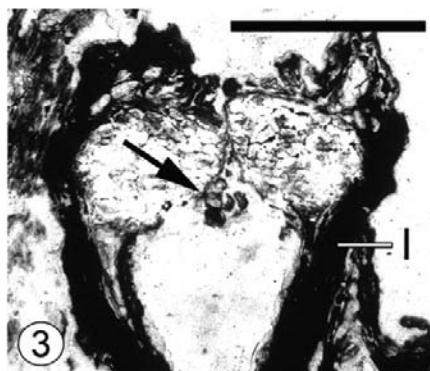


Figure 3. Longitudinal section through the micropylar end of a dispersed ovule of the same type as Fig. 1, showing pollen grain (arrow) and poorly preserved integument (I); Scale bar = 1 mm.

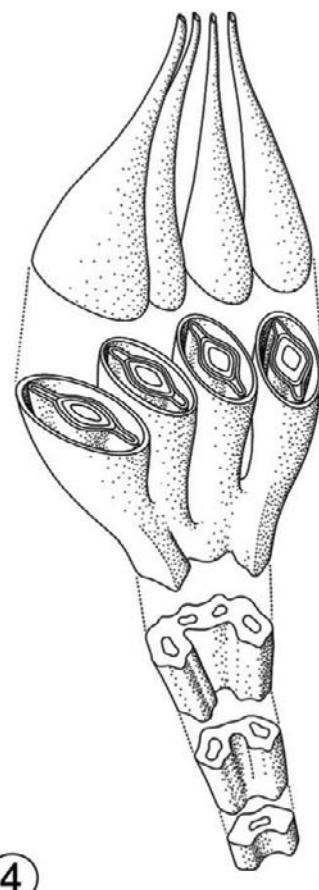


Figure 4. Suggested reconstruction of cupulate organ in Figure 5.

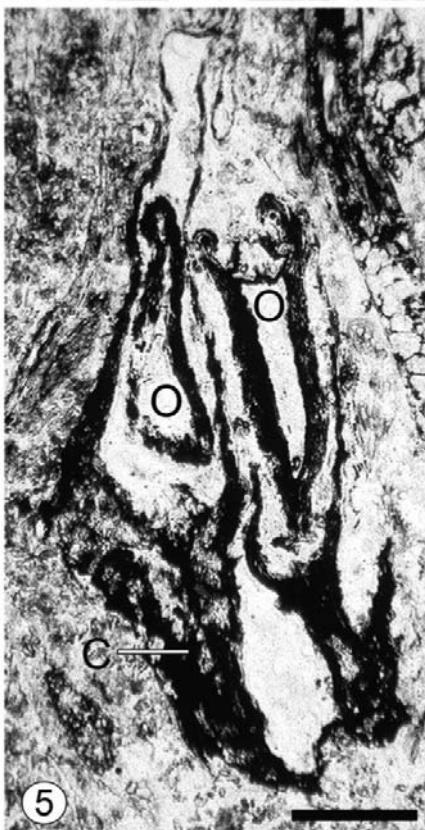


Figure 5. Section of cupulate reproductive organ with two terminal branches, each with a single, uniovulate cupule (one on left marked C). O = ovule. The cupule on the left extends above the seed (see two thin lines above). Note connection of two cupules at the base (= point of dichotomy). Scale bar = 1 mm.

CONCLUSION

The glossopterid seed ferns dominated Gondwana during the Permian, which was an important period of transition in seed plant evolution from more archaic Paleozoic forms to those appearing in the Mesozoic. Two types of anatomically preserved ovule-bearing organs of the Glossopteridales occur in a permineralized peat deposit on Skaar Ridge in the CTM. Although, multiovulate reproductive structures have been found previously in peat from the Bowen Basin of Australia and from Skaar Ridge, this report represents the first anatomical detail of the cupulate type of reproductive organ, as well as the first definitive evidence of ovules found within the cupules. The confirmation of two distinct types of seed-bearing organs in the glossopterids indicates that the group is more diverse than the single leaf morphotype, *Glossopteris*, would suggest. Anatomically preserved glossopterid reproductive organs are crucial in furthering our knowledge of the relationships within this group, and in understanding their phylogenetic position in relation to the other seed plants.

KEYWORDS

- **Compression-impression fossils**
- **Glossopteridales**
- **Ovules**
- **Paleozoic seed plants**

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (OPP-0229877). We thank current and former researchers in our laboratory and former field party members for their assistance in collecting and processing the permineralized peat. Thank you to Dr. K.B. Pigg for bringing our attention to their “in press” paper and to two anonymous reviewers for their helpful suggestions on an earlier version of the manuscript.

Chapter 8

Unveiling Cryptic Species Diversity of Flowering Plants

Yudai Okuyama and Makoto Kato

INTRODUCTION

Although DNA sequence analysis is becoming a powerful tool for identifying species, it is not easy to assess whether the observed genetic disparity corresponds to reproductive isolation. Here, we compared the efficiency of biological species identification between nuclear ribosomal and chloroplast DNA sequences, focusing on an Asian endemic perennial lineage of *Mitella* (*Asimitellaria*; Saxifragaceae). We performed artificial cross experiments for 43 pairs of 10 taxonomic species, and examined their F1 hybrid pollen fertility *in vitro* as a quantitative measure of postzygotic reproductive isolation.

A nonlinear, multiple regression analysis indicated that the nuclear ribosomal DNA distances are sufficient to explain the observed pattern of F1 hybrid pollen fertility, and supplementation with chloroplast DNA distance data does not improve the explanatory power. Overall, with the exception of a recently diverged species complex with more than three biological species, nuclear ribosomal DNA sequences successfully circumscribed 10 distinct biological species, of which two have not been described (and an additional one has not been regarded as a distinct taxonomic species) to date.

We propose that nuclear ribosomal DNA sequences contribute to reliable identification of reproductively isolated and cryptic species of *Mitella*. More comparable studies for other plant groups are needed to generalize our findings to flowering plants.

Plant systematics is one of the most active areas of biology because of marked progress in molecular phylogenetics during recent decades (Savolainen and Chase, 2003). Many of the long-standing enigmas regarding systematic positions of various taxonomic groups, for example, the relationships among gymnosperms, basal angiosperms, monocots, and dicots, have been resolved, and overall agreement has now been reached regarding circumscription of the major orders and families, with current practical taxonomic systems now following the Angiosperm Phylogeny Group (APG) system (The Angiosperm Phylogeny Group, 1998, 2003) with little controversy. The number of research articles on plant molecular phylogenetics has increased markedly in recent years, focusing mainly on extending studies to lower taxonomic groups. At the same time, the rapid accumulation of DNA sequence data for phylogenetic studies has prompted recent endeavors to use them for precise and efficient delineation of biodiversity (DNA taxonomy (Monaghan et al., 2003; Vogler et al., 2007)).

Many studies attempting to resolve plant evolutionary relationships and/or to identify plant species using DNA sequences have assumed that intraspecific genetic diversity

is usually lower than interspecific genetic diversity and that sequences derived from a species usually form a monophyletic group. Consequently, sampling of a few individuals (or even only one) is considered sufficient to represent the genetic characteristic of the species. However, these assumptions are not thoroughly supported by empirical data. For example, a recent survey of DNA sequences of a nuclear-encoded gene in the *Pinus* subgenus *Strobus* indicated that 58% of the taxonomic species studied did not form a monophyletic group (Syring et al., 2007). These authors also reported that many published studies that include multiple accessions per taxonomic species failed to reconstruct species monophyly for up to 100% of the species examined. If such species non-monophyly is common among plants, any attempt at DNA-based approaches for taxonomy would lose their relevance. Moreover, the frequency of allelic non-monophyly among plant biological species is not only methodologically but also conceptually crucial for our understanding of plant speciation. Assuming that long-term maintenance of reproductively distinct species results in allelic uniqueness of some, if not all, gene loci for each species, this should directly lead to a classic debate on the nature of plant species (Diamond, 1992; Levin, 1979; Mayr, 1992; Rieseberg et al., 2006) because a species that cannot be recognized genetically may not be a real entity (but see (Rieseberg and Brouillet, 1994)). Nevertheless, very little information is available regarding whether a plant species can indeed be regularly recognized as a genetically distinct group (e.g., only 17 studies are available (Syring et al., 2007)). More specifically, most of these studies examining the correspondence between supported clades in a phylogenetic tree and species rely heavily on traditional taxonomic species circumscriptions, obscuring whether such patterns of species non-monophyly, if observed, can be attributed to true non-monophyly or only to poor resolution of the present taxonomic system.

A more constructive approach to establish methodologies for plant DNA taxonomy would be to find genetic markers that are most likely to achieve species monophyly of the group under study, because the probability of supporting species monophyly should vary across markers and lineages in response to the marker-specific coalescence time and lineage-specific life history traits. Importantly, although many recent papers on plant DNA barcoding have placed strong emphasis on the use of markers on the chloroplast genome (Chase et al., 2007; Kress and Erickson, 2007; Lahaye et al., 2008; Taberlet et al., 2007) the chloroplast genome constitutes a non-recombining, single linkage group so that the differences among markers on the chloroplast genome might be limited to differences in the amount of information or its resolution, but not to their accuracy.

Assuming that biological species are the entities that have some, if not a complete, degree of reproductive isolation from each other, such an ideal marker for species delimitation should also have the capacity to estimate the degree of reproductive isolation among the plant individuals, from which sequence data are available but species identities are unknown. Nevertheless, few studies have compared the relationship between genetic divergence and reproductive isolation in plants. To our knowledge, only three empirical studies (genus *Glycine* (Fabaceae), *Silene* (Caryophyllaceae), and *Streptanthus* (Brassicaceae)) have been published in which a general trend of correlation between prezygotic/postzygotic reproductive isolation and genetic distance was

observed, and each of these studies used only one measure of genetic distance (nuclear ribosomal ITS DNA sequences for the former two, and allozyme distance for the last (Moyle et al., 2004)). In fact, no study has compared the relationship of different gene loci to the degree of reproductive isolation.

Here, we report that species within the Asian *Mitella* section *Asimitellaria* can mostly be recognized as a distinct, monophyletic clade that exhibits reproductive isolation (measured by sterility of pollen from F1 hybrids) based on nuclear ribosomal external and internal transcribed spacer (ETS and ITS) DNA sequences. In contrast, we found that the relatively long sequence reads (>1.5 kbp) of the chloroplast psbA-trnH interspecific spacer plus the *matK* gene, which are the most frequently used markers for plant DNA barcoding, were much less effective for recognizing the biological species boundaries likely due to natural hybridizations in *Asimitellaria*.

Asimitellaria is a monophyletic group of perennials that diversified into more than 10 species exclusively within Japan and Taiwan, which enables comprehensive sampling of genetic diversity that presumably derived from a single ancestor. All *Asimitellaria* species and varieties have the same chromosome number ($2n = 28$), with very few exceptions of intraspecific variations in chromosome number, that is, some triploid plants of *M. pauciflora* in the northernmost populations (Wakabayashi, 1973) implying that a complex polyploid formation has not been responsible for speciation. By analyzing a comprehensive collection of nuclear ribosomal ETS and ITS DNA sequences and the chloroplast psbA-trnH spacer and *matK* gene DNA sequences from samples of *Asimitellaria* plants throughout their distribution range, we first examined if distinct genotypic clusters reflect species circumscription. Furthermore, we examined pollen fertility of 43 lines of artificially crossed F1 hybrids to determine whether reproductive isolation occurs between species and how the parental genetic distances are related to the observed degrees of reproductive isolation. Furthermore, we determined whether the distinct cluster recognized by the nucleotide sequence data corresponds to a distinct taxonomic or biological species. Finally, we discuss the utility and limitations of these DNA sequences as identification tools for plant species.

MATERIALS AND METHODS

Study Organisms

The genus *Mitella* section *Asimitellaria* (Saxifragales; Saxifragaceae in the APG system) is a monophyletic group of perennials endemic to Japan and Taiwan. Nine species and an additional two varieties endemic to Japan and one species endemic to Taiwan have been described to date (Wakabayashi, 2001). We sampled 158 individuals of all 10 *Asimitellaria* species and two varieties throughout their distribution range (Figure 1) for DNA sequencing, of which 17 individuals were only sequenced for nuclear ribosomal DNA but not for chloroplast DNA because of sample loss. For each species and variety, sequences from 2 to 20 populations encompassed the entire distribution range (Figure 2). Nine other species of the genus *Mitella* were used as outgroups, as they are clearly not included in *Asimitellaria* (Okuyama et al., 2008). Overall, 105, 150, and 116 individuals were newly sequenced for nuclear ribosomal DNA, chloroplast psbA-trnH, and *matK*, respectively, whereas the remaining sequences were obtained from previous studies (Okuyama et al., 2005, 2008).

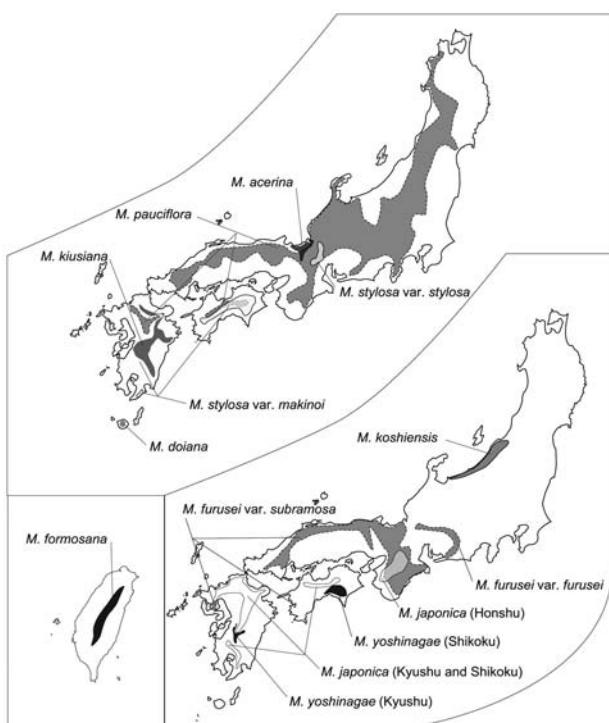


Figure 1. Geographic distribution ranges of 10 *Asimitellaria* taxonomic species (and an additional two taxonomic varieties) drawn from the records of Wakabayashi (1973) and our own studies. Note that the taxonomic species are arbitrarily separated onto two maps of Japan to minimize overlap.

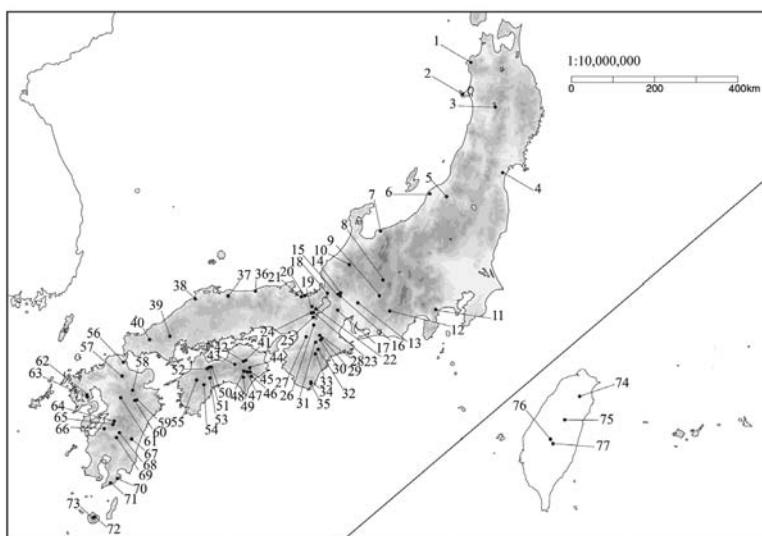


Figure 2. Locations of 77 populations from which *Asimitellaria* plants were collected.

Analysis of Nucleotide Sequences

The DNA sequences newly generated in this study were deposited in DDBJ (National Institute of Genetics, Mishima, Japan) under accession numbers AB492287–AB492762. The obtained sequences of nuclear ribosomal ETS and ITS regions and the chloroplast *matK* gene were easily aligned manually, with very few insertions/deletions (indels). In contrast, sequence alignment of the chloroplast psbA-trnH spacer was less straightforward. For example, 8- to 26-base stretch of poly-T sequences with very few other bases on the aligned site 170–195 of the chloroplast psbA-trnH spacer was impossible to align and therefore excluded from the data. Moreover, a careful inspection revealed a 37-base inversion on the aligned site 100–136 of the chloroplast psbA-trnH spacer. We therefore excluded this region and instead coded it as a single binary character with character states inverted or non-inverted. The indels in the aligned matrix were unambiguously coded as separate characters using the methods described in Simmons and Ochoterena (2000). The 5.8S region that is flanked by the ITS-1 and ITS-2 regions was removed from the dataset, as this region is missing in some of the sequences from previous studies. Genetic distances between all pairs of plant individuals were calculated separately with PAUP*4.0b10 (Swofford, 2002) for each of the nuclear and chloroplast DNA datasets, using the Tamura-Nei + I + Γ model of nucleotide substitutions (gamma shape = 0.8578, proportion of invariable sites = 0.3018) for the nuclear dataset, and the K81uf + Γ model (gamma shape = 0.2834) for the chloroplast DNA dataset, both of which were selected using ModelTest 3.7 (Posada and Crandall, 1998). The relationship between all pairwise genetic distances calculated from each of nuclear and chloroplast dataset (hereafter referred to as nuclear genetic distance and chloroplast genetic distance, respectively) were also examined. All statistical analyses were performed using the R package version 2.7.0 (R Development Core Team, 2008) unless otherwise mentioned.

For cladistic grouping of collected nucleotide sequences, maximum parsimony (MP), neighbor-joining (NJ), and Bayesian (Bayesian inference: BI) tree searches were performed using PAUP*4.0b10 for MP and NJ analyses and MrBayes 3.1 (Huelsenbeck and Ronquist, 2001) for BI. A heuristic search with tree bisection-reconnection (TBR) branch swapping and 100 additional sequence replicates, saving a maximum of 100 trees per replicate, was used for the MP tree search, and a distance measure under the maximum likelihood settings was used for the NJ tree search. To assess topological uncertainty, bootstrapping (1,000 and 10,000 replicates for MP and NJ, respectively) was also performed, using the same settings as in the original tree searches, except that we reduced both the maximum tree number and additional sequence replicates to 10 in the MP analysis. For BI, the GTR + I + Γ base substitution model, F81 base substitution model, and GTR+ Γ base substitution model were used with an uninformative prior for combined nuclear ribosomal ETS and ITS, chloroplast psbA-trnH, and chloroplast *matK*, respectively. The standard model with equal rate variation among sites was used for indel and inversion data. The nucleotide substitution models for BI were selected using MrModelTest 2.2 (Nylander, 2004). Two independent runs of Markov chain Monte Carlo (MCMC) simulation were allowed for 1.2 or 2.2 million generations each (for nuclear and chloroplast datasets, respectively), with trees sampled every 1,000 generations, to achieve independence among samples. The likelihood

scores of the obtained trees were plotted to confirm that the two independent runs reached virtually identical stationarity well before the first 201 or 1,201 trees of each run, which were discarded as burn-in. As a result, 2,000 trees were retained, and a majority-rule consensus tree (hereafter referred to as the Bayesian consensus tree) was constructed using these trees.

Examination of F1 Hybrid Fertility via Artificial Cross Experiments

To examine the presence of reproductive isolation among and within the genotypic clusters identified from the ETS and ITS sequence data, we performed artificial cross experiments using 55 individuals of 10 *Asimitellaria* species and two varieties collected from 26 populations. Specifically, because we observed genotypic clusters within each of *M. stylosa*, *M. japonica*, and *M. yoshinagae*, we suspected that these clusters may form distinct species, and therefore examined the presence of hybrid sterility among these clusters. We also checked for the presence of hybrid sterility among *M. furusei*, *M. pauciflora*, and *M. koshiensis*, as this species complex could not be separated by genotypic clustering, implying the need for assessment of concordance between taxonomic species and boundaries of biological species within the complex.

We used the pollen germination ratio of F1 hybrids for 43 combinations of crosses (corresponding to cross strain nos. 10–52 listed in Table 1) as a measure of F1 hybrid fertility, because this measure was highly quantitative and highly variable among the cross designs. The other measures, such as fruit/seed set of crossed plants and F1 plant growth were generally high, with a few exceptions in crosses between very distantly related species.

Table 1. The parental genetic distances measured with chloroplast and nuclear DNA, and average pollen fertility of nine wild-collected species (strain ID numbers 1–9) and 43 F1 hybrids (strain ID numbers 10–52).

Strain ID	Cross design ^a	Maternal (Population ^b)	Paternal (Population ^b)	N ^c	Nuclear genetic distance	Chloroplast genetic distance	Average fertility (\pm s.d.)
1	wild, Clade A	<i>M. pauciflora</i>		5	0	0	0.872 \pm 0.123
2	wild, Clade A	<i>M. koshiensis</i>		4	0	0	0.798 \pm 0.169
3	wild, Clade A	<i>M. (furusei) var. subramosa</i>		5	0	0	0.838 \pm 0.107
4	wild, Clade B	<i>M. kiusiana</i>		6	0	0	0.685 \pm 0.277
5	wild, Clade B	<i>M. stylosa</i> var. <i>stylosa</i>		7	0	0	0.813 \pm 0.101
6	wild, Clade B	<i>M. stylosa</i> var. <i>makinoi</i>		8	0	0	0.860 \pm 0.122
7	wild, Clade C	<i>M. japonica</i> (Shikoku & Kyushu)		6	0	0	0.923 \pm 0.096
8	wild, Clade C	<i>M. japonica</i> (Honshu)		4	0	0	0.808 \pm 0.123
9	wild, Clade C	<i>M. yoshinagae</i> (Kyushu)		2	0	0	0.995 \pm 0.007
10	WS, Clade A	<i>M. furusei</i> var. <i>furusei</i>	<i>M. furusei</i> var. <i>subramosa</i> (16)	6	0.0087	0.0033	0.170 \pm 0.104
11	WS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (29)	<i>M. furusei</i> var. <i>furusei</i>	4	0.0087	0.0033	0.205 \pm 0.205

Table 1. (Continued)

Strain ID	Cross design ^a	Maternal (Population ^b)	Paternal (Population ^b)	N ^c	Nuclear genetic distance	Chloroplast genetic distance	Average fertility (\pm s.d.)
12	WS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (18)	<i>M. furusei</i> var. <i>subramosa</i> (38)	6	0.0121	0.0061	0.103 \pm 0.024
13	WS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (38)	<i>M. furusei</i> var. <i>subramosa</i> (18)	5	0.0121	0.0061	0.028 \pm 0.013
14	WS, Clade B	<i>M. stylosa</i> var. <i>stylosa</i> (15)	<i>M. stylosa</i> var. <i>makinoi</i> (55)	12	0.0133	0.0033	0.417 \pm 0.217
15	WS, Clade C	<i>M. japonica</i> (60)	<i>M. japonica</i> (60)	3	0	0	0.837 \pm 0.015
16	WS, Clade C	<i>M. japonica</i> (70)	<i>M. japonica</i> (56)	9	0.0023	0.004	0.867 \pm 0.065
17	WS, Clade C	<i>M. japonica</i> (70)	<i>M. japonica</i> (51)	19	0.0023	0.0027	0.788 \pm 0.114
18	WS, Clade C	<i>M. yoshinagaе</i> (44)	<i>M. yoshinagaе</i> (64)	6	0.0217	0.0007	0.019 \pm 0.029
19	WS, Clade C	<i>M. japonica</i> (60)	<i>M. japonica</i> (29)	9	0.0364	0.0007	0.348 \pm 0.348
20	WS, Clade C	<i>M. japonica</i> (29)	<i>M. japonica</i> (60)	8	0.0364	0.0007	0.260 \pm 0.178
21	WS, Clade C	<i>M. japonica</i> (70)	<i>M. japonica</i> (29)	5	0.0378	0.0027	0.038 \pm 0.064
22	WS, Clade C	<i>M. japonica</i> (29)	<i>M. japonica</i> (56)	9	0.0385	0.0013	0.034 \pm 0.023
23	BS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (29)	<i>M. pauciflora</i> (29)	6	0.0022	0.0014	0.303 \pm 0.073
24	BS, Clade A	<i>M. koshienensis</i> (6)	<i>M. pauciflora</i> (52)	1	0.0087	0.002	0.140 \pm 0
25	BS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (24)	<i>M. pauciflora</i> (24)	6	0.0099	0.0034	0.343 \pm 0.077
26	BS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (19)	<i>M. koshienensis</i> (5)	7	0.011	0.0027	0.271 \pm 0.109
27	BS, Clade A	<i>M. acerina</i> (19)	<i>M. furusei</i> var. <i>subramosa</i> (19)	7	0.0122	0.0013	0.266 \pm 0.124
28	BS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (19)	<i>M. acerina</i> (19)	7	0.0122	0.0013	0.254 \pm 0.110
29	BS, Clade A	<i>M. furusei</i> var. <i>subramosa</i> (19)	<i>M. acerina</i> (19)	4	0.0122	0.0013	0.188 \pm 0.021
30	BS, Clade A	<i>M. acerina</i> (19)	<i>M. furusei</i> var. <i>subramosa</i> (29)	3	0.0156	0.004	0.270 \pm 0.173
31	BS, Clade B	<i>M. kiusiana</i> (60)	<i>M. stylosa</i> var. <i>makinoi</i> (71)	8	0.018	0.0068	0.274 \pm 0.128

Table 1. (Continued)

Strain ID	Cross design ^a	Maternal (Population ^b)	Paternal (Population ^b)	N ^c	Nuclear genetic distance	Chloroplast genetic distance	Average fertility (\pm s.d.)
32	BS, Clade B	<i>M. kiusiana</i> (68)	<i>M. doiana</i> (72)	10	0.0203	0.0138	0.114 \pm 0.097
33	BS, Clade B	<i>M. stylosa</i> var. <i>makinoi</i> (55)	<i>M. doiana</i> (72)	12	0.025	0.0054	0.068 \pm 0.068
34	BS, Clade C	<i>M. yoshinaga</i> (69)	<i>M. japonica</i> (29)	11	0.0214	0	0.015 \pm 0.019
35	BS, Clade C	<i>M. japonica</i> (60)	<i>M. yoshinaga</i> (47)	9	0.0301	0.0013	0.320 \pm 0.238
36	BS, Clade C	<i>M. formosana</i> (77)	<i>M. japonica</i> (29)	3	0.0316	0.0075	0
37	BS, Clade C	<i>M. yoshinaga</i> (44)	<i>M. formosana</i> (77)	2	0.0316	0.0082	0.005 \pm 0.007
38	BS, Clade C	<i>M. japonica</i> (60)	<i>M. yoshinaga</i> (64)	13	0.0354	0.0007	0.262 \pm 0.136
39	BC	<i>M. acerina</i> (19)	<i>M. stylosa</i> var. <i>makinoi</i> (55)	2	0.0227	0.004	0.007 \pm 0.007
40	BC	<i>M. furusei</i> var. <i>subramosa</i> (38)	<i>M. kiusiana</i> (67)	3	0.0288	0.0096	0.023 \pm 0.040
41	BC	<i>M. koshienensis</i> (5)	<i>M. kiusiana</i> (60)	2	0.032	0.0075	0.030 \pm 0.028
42	BC	<i>M. kiusiana</i> (68)	<i>M. pauciflora</i> (52)	6	0.0346	0.0089	0.028 \pm 0.026
43	BC	<i>M. furusei</i> var. <i>subramosa</i> (24)	<i>M. stylosa</i> var. <i>makinoi</i> (55)	3	0.0347	0.0033	0.02
44	BC	<i>M. pauciflora</i> (52)	<i>M. stylosa</i> var. <i>makinoi</i> (45)	7	0.0371	0	0.103 \pm 0.216
45	BC	<i>M. japonica</i> (60)	<i>M. acerina</i> (19)	4	0.0549	0.011	0
46	BC	<i>M. kiusiana</i> (60)	<i>M. japonica</i> (60)	3	0.0646	0.0007	0
47	BC	<i>M. japonica</i> (60)	<i>M. pauciflora</i> (39)	2	0.0672	0.0083	0
48	BC	<i>M. japonica</i> (60)	<i>M. furusei</i> var. <i>subramosa</i> (38)	3	0.0672	0.0103	0
49	BC	<i>M. formosana</i> (77)	<i>M. acerina</i> (19)	2	0.0691	0.0054	0
50	BC	<i>M. formosana</i> (77)	<i>M. stylosa</i> var. <i>makinoi</i> (55)	3	0.072	0.0027	0
51	BC	<i>M. yoshinaga</i> (64)	<i>M. kiusiana</i> (64)	3	0.0747	0.0013	0.003 \pm 0.006
52	BC	<i>M. furusei</i> var. <i>subramosa</i> (29)	<i>M. japonica</i> (29)	2	0.0747	0.0082	0

a. Abbreviations; wild: wild collected, cultivated plants examined for pollen fertility, WS: cross within biological species, BS: cross between species within the clade (A, B, C) defined in Figure 2, BC: cross across the clades.

b. Their geographic origins are shown as a number in parentheses, each of which corresponds to the population ID no. in Additional file I: Table S1.

c. The number of plant individuals examined for their fertility.

Just prior to flowering, potted plants used for the crosses were transferred from the garden to growth chambers (NK System, Osaka, Japan) from which potential pollinator insects were excluded. Hermaphrodite flowers of maternal plants were emasculated before anthesis (this procedure was omitted for female flowers of several sexually dimorphic species), and subsequently used for the cross. A sufficient amount of pollen was applied onto the stigma of an emasculated flower using a toothpick. Three to 5 weeks after the cross, mature seeds were collected from these crossed fruits. The collected seeds were surface-sterilized for ~15 min with 0.04% TritonX-100 and sodium hypochlorite solution (~0.05% chlorine) and then plated onto sterilized nutrient agar in plastic Petri dishes. The resultant seeds germinated normally within 3 weeks of sowing. Two to 5 months after germination, the seedlings were transplanted to pots filled with well-fertilized soil and grown in the garden to examine their pollen viability in the next flowering season (March–May). Pollen grains were collected from each individual plant from just dehiscent anthers using a toothpick, and the pollen was scattered on a spot of liquid culture dropped onto a 1.5 cm × 1.5 cm square 1% agar culture, optimized for *Asimitellaria* pollen germination with 5% sucrose and 5.0 × 10–3% boric acid. After a 24 hr incubation at 25°C in a humid plastic case, the agar culture with pollen grains was fixed with 3:1 ethanol:acetic acid, stained with 0.1% aniline blue for >5 h, washed with 1% acetic acid, and dried to prepare microscope slides. For each slide specimen, 100 pollen grains were chosen at random to assess their germination ability under a binocular microscope, and the count was used as the measure of fertility for each individual plant. Accordingly, 1–19 individual measures of fertility were obtained for each of 43 intra- and interspecific cross strains. In addition, two to eight wild-collected individuals from each of nine *Asimitellaria* species were used to confirm their high pollen germination ability under our experimental conditions. To maximize the number of cross designs within limited time and space, we conducted the cross within a genotypic cluster as a control only in clade C, and instead used the measures of fertility for the wild-collected individuals originating presumably from spontaneous crosses within each genotypic cluster. The average pollen fertility was calculated for each of the 43 crossed strains and the nine wild-collected species. The artificially crossed strains within each of clades A–C observed in nuclear DNA phylogeny (Figure 3) with more than five individual plants were examined statistically for any reduction in hybrid fertility, compared to pollen fertility of corresponding wild-collected species using the Mann–Whitney U test with Bonferroni correction. The crossed strains across the clades were always nearly or completely sterile (Table 1); thus we did not assess these statistically for fertility reduction.

We further examined the correlations between genetic distance of parental plant individuals and their F1 hybrid fertility by plotting the results of the artificial cross experiments with parental nuclear or chloroplast genetic distances of the crossed lines. Genetic distance was simply calculated using the DNA sequences of parental individuals when available, whereas in several cases, a different plant individual of the same geographic origin was used as the representative of the accidentally lost plant used for the cross, as the intraspecific sequence divergence within a single population was always negligible (<0.001 for chloroplast DNA and <0.002 for nuclear DNA). For wild-collected plants and the intraspecific cross of plants from the same population (strain

ID numbers. 1–9, 15 in Table 1), their parental genetic distances were regarded as zero to discriminate them from the intraspecific cross of plants from different populations. A multiple regression analysis using a generalized additive model with smoothing splines was performed using each of the nuclear and chloroplast genetic distances as the independent variable and F1 hybrid pollen fertility as the dependent variable, applying the gam and smooth spline function of the R package. An optimal regression model was selected by comparing the Akaike information criterion (AIC) values for all possible models, whereby nuclear and chloroplast genetic distance terms as well as their interaction term were incorporated in the full model.

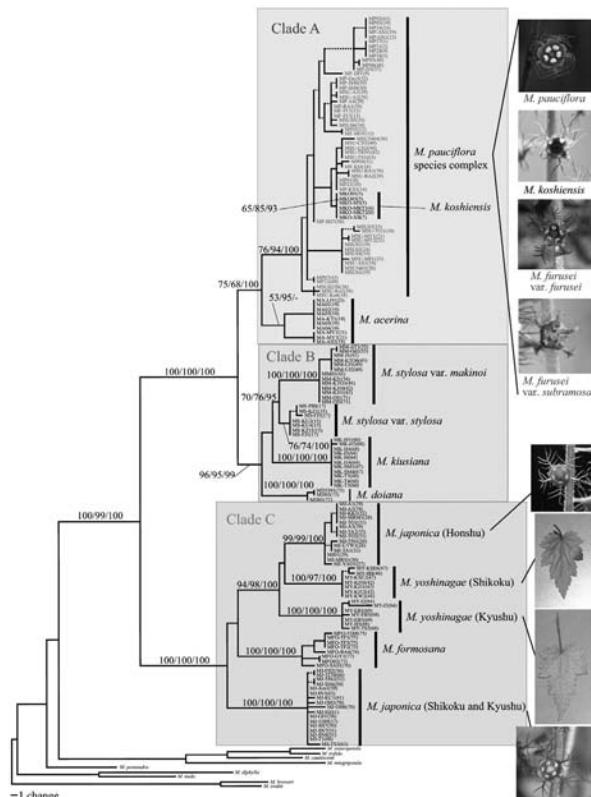


Figure 3. One of the 8,100 most parsimonious trees ($L = 476$, $CI = 0.7542$, $RI = 0.9722$) obtained by cladistic genotypic clustering of combined ETS and ITS sequences from 158 *Asimitellaria* plants and nine outgroups. Branches that collapse in the strict consensus tree are shown by dashed lines. Nodal support values by bootstrapping or posterior probability are indicated near branches (MP/NJ/BI) where needed. Vertical bars on the right, with the exception of the *M. pauciflora* complex, represent distinct biological species proposed in the present study, which have substantial reproductive barriers to each other (>39% of fertility reduction if crossed). For the *M. pauciflora* species complex, the labels each represent the taxonomic species *M. pauciflora*, *M. furusei* var. *furusei*, and *M. furusei* var. *subramosa*, respectively, at least among those that have a substantial level of reproductive isolation (>57% of fertility reduction if crossed). Pictures on the right indicate some diagnostic characters (i.e., flower or abaxial side of the leaf) for (cryptic) biological species in several species complexes. The population ID numbers from which the individual accessions were collected are indicated in parentheses.

DISCUSSION

The Relationships Among Genetic Divergence, Postzygotic Isolation, and Taxonomic Species Boundaries in Asian *Mitella*, and their Implications for DNA Taxonomy

A rapidly evolving endeavor in recent taxonomy is to utilize DNA sequences for precise and efficient delineation of biodiversity (Vogler et al., 2007) but the information regarding how observed genetic disparity corresponds to reproductive isolation has been critically lacking. In the present study, we have comprehensively illustrated the relationships among genetic divergence, postzygotic reproductive isolation, and taxonomic species boundaries using *Asimitellaria* as a model group. In *Asimitellaria*, we found that the degree of postzygotic reproductive isolation correlates consistently with genetic distance measured by nuclear ribosomal DNA only (Figure 7).

Consequently, the distinct subclades observed in the phylogeny (Figure 3) each corresponded to a distinct biological species, with the cross among which always result in at least 39% fertility reduction compared to the cross within the subclades (Table 1, Figure 4). Furthermore, we found that the genotypic clustering based on nuclear ribosomal DNA distance was mostly concordant with a morphology-based system (Figure 3). These findings have several significant implications for application of DNA taxonomy in flowering plants. Together with the previous finding in other three genera (*Glycine*, *Silene*, and *Strephanthus* (Moyle et al., 2004) now we have a strong evidence to assume that in general the degree of postzygotic reproductive isolation well correlates with genetic distance (note here we assume no polyploidy, although it is undoubtedly a major factor generating reproductive isolation in plants). Nevertheless, the goodness of correlation can vary largely among genetic markers used for the distance measure, and thus careful examination is necessary to determine which marker should be chosen. In the marker choice for DNA taxonomy, comparing the relative goodness of fit to a morphology-based system might be very helpful; in the case of *Asimitellaria*, the marker that fit better to the pattern of reproductive isolation also fit better to the morphology-based system. Accordingly, in the case when artificial cross experiments are impractical, it would be a good practice for researchers to compare multiple, unlinked markers such as chloroplast DNA and nuclear ribosomal DNA for goodness of fit to the morphology-based system of the plant group under study.

We would note, however, that the postzygotic reproductive isolation measured by pollen fertility in the present study is only a very small fraction of reproductive isolation that exists in nature. It is suggested that the prezygotic isolation has more important role in keeping different species genetically distinct in both plants and animals (Lowry et al., 2008). Whether the degree of prezygotic reproductive isolation correlates consistently with genetic distance is not clear because prezygotic isolation involves many adaptive traits such as flowering time, pollinator difference, and floral morphology, which can be direct targets of natural selection (Coyne and Orr, 2004). Nevertheless, the general trend of strong correlations between genetic distance measured with a specific genetic marker and postzygotic reproductive isolation can be used (after the choice of appropriate markers) as a strong basis for regarding that

phylogenetically supported distinct clusters can be used for a minimum assessment of biological species diversity.

Surprisingly, until now, limited examples of cryptic species within higher plants, including angiosperms, have been reported (Bickford et al., 2007; Grundt et al., 2006; Nicolè et al., 2007; Whittall et al., 2004). This may suggest that cryptic species are less frequent in flowering plants because the morphology-based taxonomic system is highly reliable; however, there have been too few empirical studies to make conclusive inferences regarding the prevalence of cryptic species within flowering plants. Nevertheless, considering the present findings of at least three cryptic species present within *Asimitellaria*, a relatively taxonomically well examined lineage with regard to comparative morphology, embryology, and cytology (e.g., Wakabayashi, 1973, 1977, 1992), it is likely that many angiosperm lineages contain at least some cryptic species (see also (Grundt et al., 2006) who reported cryptic species diversity of the genus *Draba* in the Arctic region). Thus, the establishment of a concrete framework for DNA taxonomy would be valuable to unravel cryptic diversity of flowering plants.

Utility and Limitations of Nuclear Ribosomal DNA Sequences for Delimiting Species

In the present study, genotypic clustering using nuclear ribosomal ETS and ITS sequences was shown to be fairly successful for delimiting biological species of *Asimitellaria* (10 of 11 clusters corresponded to distinct biological species), except for the *M. pauciflora* complex. This good correspondence was achieved from low intraspecific sequence diversity and consistent sequence monophyly for each biological species, in addition to the high overall variability of the sequences, an essential prerequisite for DNA taxonomy. It is noteworthy that recent plant DNA barcoding studies have suggested the potential utility of ITS regions for identifying plant taxonomic species, partly because it is the most frequently sequenced locus in plant phylogenetic studies. Importantly, Kress and colleagues found the highest interspecific sequence divergence of ITS among 10 genetic markers tested (Kress et al., 2005), and Chase and colleagues reported a high probability (93.21%) of assigning taxonomic species using the ITS-1 region as a BLAST query sequence against GenBank (Chase et al., 2005). However, in addition to these previously suggested advantages of ITS, it is especially important to achieve consistent species monophyly (each clade or genotypic cluster corresponding to one species) to be useful for species recognition and identification. Accordingly, the identifier loci are required to have a relatively short coalescence time. This may be the case in ETS and ITS regions because, unlike other nuclear loci, the sequence homogeneity within a genome is strongly maintained by concerted evolution (Okuyama et al., 2005; Soltis et al., 2004; Wendel et al., 1995). Note that *Asimitellaria* is a paleotetraploid lineage ($2n = 28$), which has double chromosome number compared to most of the remaining species of *Mitella* and its allies (*Heuchera* group; mostly $2n = 14$). Therefore, even the case in which ribosomal DNAs are likely located in multiple chromosome blocks, each of the species can be recognized consistently as a monophyletic group because of concerted evolution process comparable to those of diploids (see also Soltis et al., 2004; Wendel et al., 1995 in which clear evidence for rapid concerted evolution among different chromosomal locations is reported).

Another ideal property of potential loci for plant species delimitation is their robustness for genetic introgression via interspecific hybridization, as plant species with morphological and/or ecological distinctiveness are often reported to have intensive gene exchange (e.g., Yatabe et al., 2007). It is theoretically predicted that the loci under divergent selection and those linked to hybrid incompatibility are unlikely to introgress across species (Barton, 1979) but it is usually very difficult to find such loci for many non-model plant lineages. The nuclear ribosomal gene is an exception, as in most eukaryotic genomes, including those of flowering plants, whereby the physical locations of highly repetitive, nuclear ribosomal gene clusters are confined within telomeric regions (e.g., (Shishido et al., 2000; Zhang and Sang, 1999) the physical locations of ribosomal RNA genes for various model organisms are also available in the MapViewer at <http://www.ncbi.nlm.nih.gov/> website) where recombination is presumably suppressed (Gerton et al., 2000; Wu et al., 2003). Importantly, in a recent genome-wide survey of genetic differentiation between frequently hybridizing, sympatric sunflower species, *Helianthus annuus* and *H. petiolaris*, Yatabe and colleagues found that the chromosomal segments that differentiated these species are usually very small (even undetectable in sunflowers) except for the regions neighboring chromosomal breakpoints (Yatabe et al., 2007). Consequently, it is possible that the ribosomal ETS and ITS sequences, which are located near the chromosomal breakpoints, remain distinct under a substantial degree of interspecific genetic introgression, as in the case of *Asimitellaria* (Okuyama et al., 2005).

Taken together, we propose that the ETS and ITS regions are the most promising currently available candidate markers for DNA taxonomy of flowering plants, with presumably short coalescent times and robustness against introgressive hybridization (Okuyama et al., 2005); note, however, that there have been several reports of interspecific gene flow in the ITS regions (Aguilar et al., 1999a, 1999b). Even in *Asimitellaria*, there is a clear example of interspecific gene flow in ITS region (but not in ETS) between *M. acerina* and *M. furuei* var. *subramosa* (Okuyama et al., 2005). Therefore, it is worth testing in *Asimitellaria* and other plant lineages whether nuclear ribosomal DNA is indeed more robust against interspecific gene flow compared to other loci such as protein-coding nuclear genes. Of course, in either case, it would be better to keep in mind that there is unlikely to be any genetic markers free of interspecific gene flow, which is always a major challenge in plant taxonomy.

Moreover, in some cases, another caution is needed for use of the ETS and ITS because there are non-negligible numbers of reports for the presence of multiple divergent paralogs of ribosomal DNA in a single genome (Alvarez and Wendel, 2003; Bailey et al., 2003; Buckler-IV et al., 1997) which makes it impossible to compare orthologous sequences among individuals, an essential step for DNA taxonomy. Such a problematic nature of ribosomal DNA for plant DNA taxonomy might be more general phenomenon, considering even reporting bias might be present for the plant group in which ribosomal DNAs cannot be sequenced directly in a simple manner as in *Asimitellaria*.

Also note that even our genotypic clustering approach failed to recognize the three taxonomic species within the *M. pauciflora* complex (Figure 3), especially to

discriminate between *M. furusei* and *M. pauciflora* (*M. koshiensis* could be recognized with a species-specific sequence nested within the complex in the phylogenetic tree). This does not mean that the complex should be grouped as a single biological species, as the phenotypic discontinuity among the three taxonomic species is obvious ((Wakabayashi, 1973, 2001) Figure 3), and *M. furusei* and *M. pauciflora* co-occur in many populations and have very distinct life history traits, including flowering phenology, pollinator species, and modes of reproduction (Okuyama et al., 2004, 2008). In addition, the artificial cross experiments within the *M. pauciflora* complex (Figure 4) also supported the conclusion that the three taxonomic species are reproductively isolated. Even the cross between genetically divergent populations of *M. furusei* var. *subramosa* was found to result in strong hybrid sterility (<10% pollen fertility; strain ID numbers 12 and 13 in Figure 4). Therefore, it is clear that the species complex comprises more than three biological species, although clarifying a comprehensive pattern of reproductive isolation within the *M. pauciflora* complex is beyond the scope of the present study. The divergence between *M. furusei* and *M. pauciflora* appears to have occurred very recently compared to other speciation events outside the *M. pauciflora* complex, and this observation probably indicates limitation to the sole use of ETS and ITS sequences for recognizing plant biological species that have undergone recent speciation. Accordingly, it is expected that a recently radiated plant lineage would be most difficult for biological species recognition based on ETS and ITS sequences, even if sequence variations are present, as in the case of the *M. pauciflora* complex. At present, there is no conventional, DNA-based methodology for recognizing these recently diverged biological species (amplified fragment length polymorphism-PCR may be a candidate, although this method is fairly labor-intensive; e.g., (Kardolus et al., 2008; Pellmyr et al., 2007)). A recent simulation-based study suggested that sampling of a moderate number (up to 10) of loci could correctly delimit recently diverged biological species with a coalescent theory-based approach, even without relying on their reciprocal monophyly (Knowles and Carstens, 2007) Thus, there is no doubt that adding the data from different loci would result in more precise delimitation of biological species, including those that have differentiated recently, although the cost and effort would also increase substantially.

RESULTS

The numbers of sites in the aligned data matrices of nuclear ribosomal ETS, ITS-1, ITS-2, and indels were 453, 276, 224, and 30, of which 84, 68, 47, and 24 were parsimony-informative, respectively, whereas those of psbA-trnH, matK, and indels (and an inversion) were 500, 1164, and 15, of which 28, 48, and 11 were parsimony-informative, respectively. Statistical summaries of genetic diversity observed within taxonomic species and varieties of *Asimitellaria* are listed in Table 2. Because chloroplast genetic distances in *Asimitellaria* were found to have very different information from that of the nuclear genetic distance, we regarded the nuclear and chloroplast DNA sequence data as different sources of information, and analyzed each separately.

Table 2. Statistical summaries of intraspecific genetic diversity of nuclear and chloroplast DNAs observed in *Asimitellaria*

Species	No. of populations	No. of individuals	Nuclear distance mean (\pm S.D.)	Chloroplast distance mean (\pm S.D.)
<i>M. formosana</i>	4	8	0.0037 \pm 0.0026	0.0005 \pm 0.0004
<i>M. japonica</i> (all)	20	31	0.0185 \pm 0.0171	0.0013 \pm 0.0015
<i>M. japonica</i> (Honshu)	8	14	0.0010 \pm 0.0013	0.0000 \pm 0.0000
<i>M. japonica</i> (Shikoku and Kyushu)	12	17	0.0012 \pm 0.0013	0.0021 \pm 0.0014
<i>M. yoshinagae</i> (all)	7	14	0.0130 \pm 0.0107	0.0011 \pm 0.0011
<i>M. yoshinagae</i> (Shikoku)	4	7	0.0010 \pm 0.0011	0.0021 \pm 0.0013
<i>M. yoshinagae</i> (Kyushu)	3	7	0.0024 \pm 0.0024	0.0000 \pm 0.0000
<i>M. doiana</i>	2	3	0.0007 \pm 0.0006	0.0004 \pm 0.0004
<i>M. kiusiana</i>	4	11	0.0001 \pm 0.0004	0.0012 \pm 0.0010
<i>M. stylosa</i> (all)	11	20	0.0060 \pm 0.0053	0.0016 \pm 0.0014
<i>M. stylosa</i> var. <i>stylosa</i>	2	7	0.0005 \pm 0.0007	0.0014 \pm 0.0009
<i>M. stylosa</i> var. <i>makinoi</i>	9	13	0.0011 \pm 0.0011	0.0001 \pm 0.0002
<i>M. acerina</i>	3	10	0.0019 \pm 0.0031	0.0007 \pm 0.0007
<i>M. furusei</i> (all)	20	31	0.0079 \pm 0.0044	0.0030 \pm 0.0016
<i>M. furusei</i> var. <i>furusei</i>	4	6	0.0051 \pm 0.0030	0.0025 \pm 0.0013
<i>M. furusei</i> var. <i>subramosa</i>	16	25	0.0076 \pm 0.0046	0.0030 \pm 0.0017
<i>M. koshiensis</i>	3	6	0.0000 \pm 0.0000	0.0016 \pm 0.0014
<i>M. pauciflora</i>	20	24	0.0039 \pm 0.0038	0.0015 \pm 0.0012

Cladistic grouping of the nuclear ribosomal DNA sequences using MP, NJ, and BI all resulted in nearly identical topology, subdividing the entire ingroup into three subclades, A–C, in which two to five distinct genotypic clusters were consistently recognized (Figure 3). In total, at least 11 distinct subclades were found within *Asimitellaria* (Figure 3), and the identified clades were not largely incongruent with the present taxonomic system. *Mitella koshiensis*, *M. stylosa*, *M. kiusiana*, *M. doiana*, and *M. formosana* were always supported to be monophyletic, with moderate to high MP (65–100%) and NJ (76–100%) bootstrap values and high Bayesian posterior probability (>93%). The monophyly of *M. acerina* was not supported by BI, although this was apparently due to genetic introgression of ITS sequences with sympatric *M. furusei* var. *subramosa*, which has been reported previously (Okuyama et al., 2005). Intriguingly, as expected from the excess of their intraspecific genetic diversity (Table 2), none of the four taxonomic species, *M. pauciflora*, *M. furusei*, *M. japonica*, or *M. yoshinagae*, formed a monophyletic group, with the former two species forming an inseparable clade together with *M. koshiensis*, whereas *M. japonica* and *M. yoshinagae* were each composed of two non-sister clades with very high nodal support (>97%; Figure 3). In addition, *M. stylosa* was further subdivided into two strongly supported (100%/100%/100% and 76%/74%/100% in MP, NJ, and BI, respectively) monophyletic

clades, with each corresponding to the two taxonomic varieties, *M. stylosa* var. *stylosa* and *M. stylosa* var. *makinoi*.

The pairs of two distinct clades found within both *M. japonica* and *M. yoshinagae* have not been recognized previously, but careful reexamination of morphological characters found some support for these clusters from the morphology of petal and leaf surface structure (Figure 3). The measure of F1 hybrid sterility by artificial crossing experiments revealed the presence of clear reproductive barriers between the clusters. The cross between *M. japonica* individuals of the same genotypic cluster (Kyushu and Shikoku populations), each collected from geographically well isolated (279–335 km) populations, showed little hybrid sterility (Figure 4, Table 1, strain numbers 16 and 17). Therefore, the significant reduction in pollen fertility in the F1 hybrids resulting from the inter-genotypic crosses cannot be explained by ordinary isolation by distance within a single species, whereby a gradual decrease in hybrid fertility is expected, but by an incompatibility between reproductively isolated, discontinuous species. Similarly, the two genotypic clusters observed in *M. stylosa*, corresponding to the taxonomic varieties *M. stylosa* var. *stylosa* and *M. stylosa* var. *makinoi*, would be two distinct species, as a significant reduction in hybrid fertility between them was observed. An exception is the species complex of *M. koshiensis*, *M. furusei*, and *M. pauciflora* (henceforth referred to as *M. pauciflora* complex), in which fairly large genetic variations were observed. Nevertheless, each of the three taxonomic species cannot be recognized as a distinct genotypic cluster.

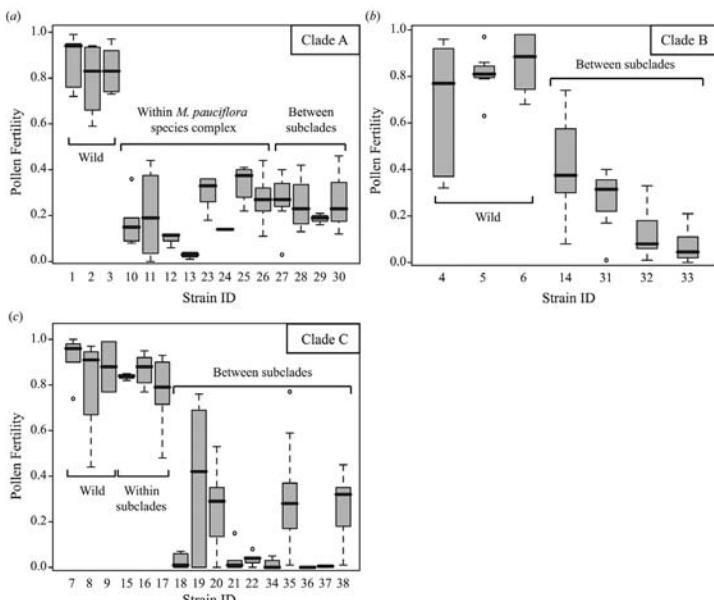


Figure 4. Reduction in pollen fertility in artificially crossed interspecific hybrid strains of *Asimitellaria*. Boxes and error bars represent the range distribution for individual strains, with thick horizontal bars representing the median. (a) Crosses within clade A. (b) Crosses within clade B. (c) Crosses within clade C. The strain ID numbers are as in Table 1. Note that the strain IDs 1–9 are the wild-collected individuals used as controls.

Surprisingly, we found that the genetic information from chloroplast DNA sequences was very different from nuclear ETS and ITS DNAs in *Asimitellaria*. As shown in Figure 5, less nucleotide divergence in chloroplast DNA did not necessarily coincide with less divergence in ETS and ITS, and *vice versa*. In addition, unlike the nuclear ETS and ITS data, very few genotypic clusters that potentially correspond to species were found in chloroplast DNA. Actually, eight out of 10 taxonomic species in *Asimitellaria* were recovered to be paraphyletic or polyphyletic in the chloroplast DNA data (e.g., subclades D1, E1, and E2 in Figure 6), although none of these groupings was supported morphologically. Only *M. formosana* and *M. doiana*, both of which are strictly allopatric with other *Asimitellaria* species, each formed an exclusively monophyletic group, although the nodal support for the *M. formosana* clade was weak (Figure 6). This pattern is consistent with our previous finding that chloroplast DNA in *Asimitellaria* is highly sensitive to rare interspecific gene flow (Okuyama et al., 2005). This finding is further supported by the fact that the pattern of F1 hybrid pollen sterility expected from chloroplast genetic distances (Figure 7a; deviance explained = 52.8%), fitted the data much less compared to that from nuclear genetic distances (Figure 7b; deviance explained = 88.5%), and multiple regression analysis indicated that only the nuclear genetic distances are necessary and sufficient to explain the observed pattern of F1 hybrid pollen fertility (the AIC value for the optimal model was -64.04, whereas that for the full model was -63.62).

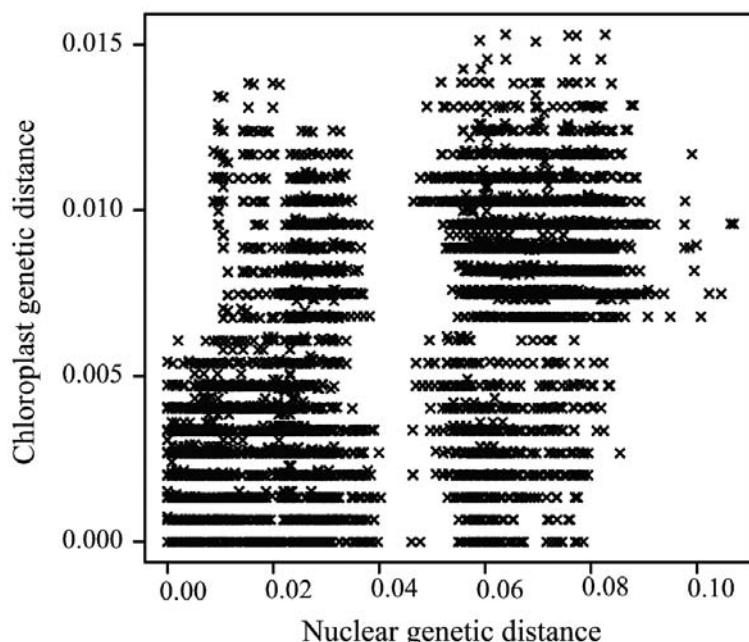


Figure 5. Strong incongruence between nuclear and chloroplast genetic distance for all possible pairs of 141 *Asimitellaria* plant individuals. The large gap along the x-axis (nuclear genetic distance) corresponds to the large genetic gap between clade A+B and clade C.

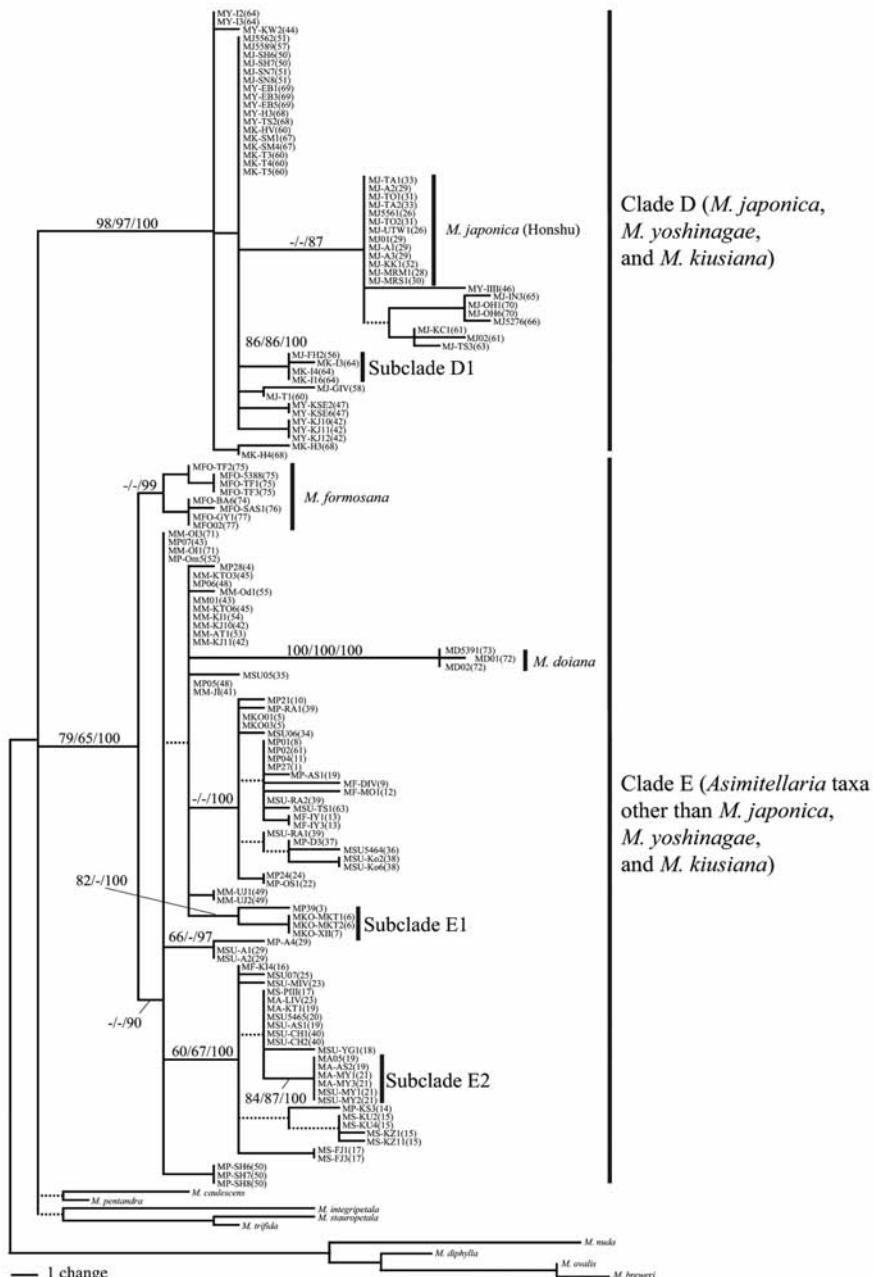


Figure 6. One of the 7,700 most parsimonious trees ($L = 189$, $CI = 0.8042$, $RI = 0.9671$) obtained via cladistic genotypic clustering of combined chloroplast DNA sequences from 141 *Asimitellaria* plants and nine outgroups. D1, E1, and E2 are the strongly supported subclades, each consisting of two reproductively isolated, distinct biological species. Other descriptions are as in Fig. 3.

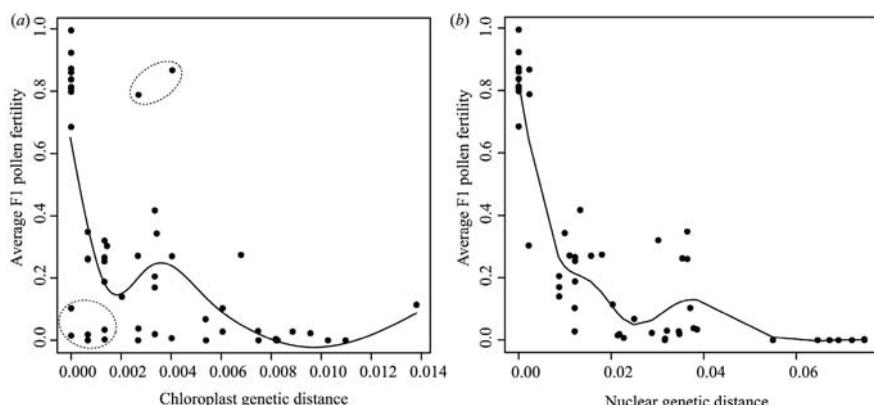


Figure 7. Nonlinear, single regression of average F1 pollen fertility (a) against average parental chloroplast genetic distance (K81uf + G model of nucleotide substitutions with gamma shape = 0.2834) and (b) against the average parental nuclear genetic distance (Tamura–Nei + I + Γ model of nucleotide substitutions with gamma shape = 0.8578, proportion of invariable sites = 0.3018). Dashed circles indicate data points showing strong discordance between chloroplast genetic distance and average F1 pollen fertility.

CONCLUSION

To our knowledge, this is the first comprehensive study that links quantitative measures of postzygotic reproductive isolation to genetic distinctness observed in an angiosperm lineage. We showed that with appropriate selection of genetic markers, most reproductively isolated species of *Asimitellaria* could be recognized as distinct genotypic clusters. With only the present results being available, however, we could not conclude whether the low efficiency of biological species recognition using chloroplast DNA is a general trend in flowering plants. It is widely appreciated that chloroplast DNA has a general advantage of the availability of nearly universal primers that are applicable to entire flowering plants, and less risk of sampling multiple copies from one plant individual, which in turn is one of the major obstacles in using nuclear ribosomal DNA sequences (Alvarez and Wendel, 2003; Bailey et al., 2003; Buckler-IV et al., 1997). Nevertheless, extensive introgression of chloroplast DNA via interspecific hybridization is a well-known and frequently reported phenomenon not restricted to *Asimitellaria* (e.g., (Bänfer et al., 2006; Fehrer et al., 2007; Kron et al., 1993; Petit et al., 1997) older examples are reviewed in (Rieseberg and Soltis, 1991). Therefore, it would be necessary to reassess how common chloroplast DNA introgression is among flowering plants, especially after sampling a sufficient number of individuals for each taxonomic or biological species. Also, it might be helpful to compare the use of nuclear ribosomal and chloroplast DNA in plant lineages without chloroplast DNA introgression. Further attempts at DNA taxonomy in plant lineages with various life history traits (annuals, perennials, trees, aquatics) and evolutionary backgrounds (recent and old radiations, oceanic island endemics) are required to generalize the utility of ETS and ITS for accurate and efficient delimitation of plant biological species.

KEYWORDS

- Akaike information criterion (AIC) values
- *Asimitellaria*
- Bootstrapping
- Cladistic genotypic clustering
- Delimiting species
- DNA taxonomy
- Genetic distance
- Postzygotic reproductive isolation

AUTHORS' CONTRIBUTIONS

Yudai Okuyama designed and performed the research, analyzed the data, and wrote the manuscript. Makoto Kato designed research and wrote the manuscript. Both authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors are grateful to M. Wakabayashi, N. Fujii, K. Nishikawa for plant materials, Y. Kameda and T. Okamoto for technical assistance, and A. Kawakita, T. Sota, V. Savolainen, and anonymous reviewers for critical and constructive comments on manuscript. Special thanks to T. W. Hsu and H. M. Chang (Institute of Endemic Species Research) for their generous help on field trip in Taiwan. This study was supported by Grant-in-Aids No. 15370012 from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (to Makoto Kato), and by JSPS Research Fellowships for Young Scientists No. 02323 (to Yudai Okuyama).

Chapter 9

Pollen Development in *Annona cherimola* Mill. (Annonaceae)

Jorge Lora, Pilar S. Testillano, Maria C. Risueño, Jose I. Hormaza, and
Maria Herrero

INTRODUCTION

In most flowering plants, pollen is dispersed as monads. However, aggregated pollen shedding in groups of four or more pollen grains has arisen independently several times during angiosperm evolution. The reasons behind this phenomenon are largely unknown. In this study, we followed pollen development in *Annona cherimola*, a basal angiosperm species that releases pollen in group of four, to investigate how pollen ontogeny may explain the rise and establishment of this character. We followed pollen development using immunolocalization and cytochemical characterization of changes occurring from anther differentiation to pollen dehiscence.

Our results show that, following tetrad formation, a delay in the dissolution of the pollen mother cell wall and tapetal chamber is a key event that holds the four microspores together in a confined tapetal chamber, allowing them to rotate and then bind through the aperture sites through small pectin bridges, followed by joint sporopollenin deposition.

Pollen grouping could be the result of relatively minor ontogenetic changes beneficial for pollen transfer or/and protection from desiccation. Comparison of these events with those recorded in the recent pollen developmental mutants in *Arabidopsis* indicates that several failures during tetrad dissolution may convert to a common recurring phenotype that has evolved independently several times, whenever this grouping conferred advantages for pollen transfer.

Pollen development is a well characterized and highly conserved process in flowering plants (Blackmore et al., 2007; McCormick, 2004; Scott et al., 2006). Typically, following anther differentiation, a sporogenous tissue develops within the anthers producing microsporocytes or pollen mother cells. Prior to meiosis, pollen mother cells (PMC) become isolated by a wall with the deposition of a callose layer. Each pollen mother cell, as the result of the two meiotic divisions, generates four haploid cells forming a tetrad and, for a short time, these four sibling microspores are held together in a persistent pollen mother cell wall that is surrounded by callose. The tapetum then produces an enzyme cocktail that dissolves the pollen mother cell wall and the microspores are shed free and become independent (Scott et al., 2006). The unicellular microspores go through an asymmetric mitotic division (pollen mitosis I) to produce a pollen grain with two cells, a larger vegetative cell that hosts a smaller generative cell; the latter will divide once more to produce two sperm cells

(pollen mitosis II). Pollen mitosis II can take place before or after pollen release and, depending on when it occurs, the pollen will be bicellular or tricellular at the time of anther dehiscence. Throughout the manuscript we will use the term “pollen tetrads” for mature pollen to avoid confusion with the tetrads of early developmental stages (“microspore tetrads”).

Angiosperms pollen is most commonly released as single pollen grains or monads (Pacini and Franchi, 1999) which represent the basic angiosperm pollen-unit. Dehiscence of aggregated pollen (mostly in groups of four) is considered a recent apomorphic characteristic (Harder and Johnson, 2008; Walker and Doyle, 1975) that has arisen independently several times during evolution primarily in animal-pollinated taxa although, in some cases, monads may have evolved secondarily from groups of four grains (Harder and Johnson, 2008). Pollen release as tetrads has been reported in some or all members of 55 different angiosperm families and also in some pteridophytes (Pacini et al., 1985). Blackmore and Crane (1988) put forward that the maintenance of pollen tetrads could be the result of relatively minor ontogenetic changes and, consequently, this could be an excellent example of convergence in situations where the release of pollen as tetrads is an effective reproductive strategy. Interestingly, the dissemination of pollen as tetrads has also been reported in the quartet mutants of *Arabidopsis* (Preuss et al., 1994; Rhee et al., 2003).

Annonaceae, included in the order Magnoliales, is the largest family within the basal angiosperm Magnoliid clade (APG II, 2003; Soltis et al., 2005). Due to its phylogenetic position among the basal angiosperms, the family has been the object of considerable interest from a taxonomic and phylogenetic point of view (Doyle and Le Thomas, 1994, 1997a; Pirie et al., 2007) and a number of studies have focused on pollen morphology (Doyle and Le Thomas, 1997b; Le Thomas, 1980, 1981; Walker, 1971a, 1971b). Although, most genera of the Annonaceae produce solitary pollen at maturity, in several species of the family pollen is released aggregated in groups of four or in polyads (Walker, 1971b). Recent studies on the mechanism of pollen cohesion in this family have been performed in species of the genera *Pseuduvaria* (Su and Saunders, 2003), *Annona* and *Cymbopetalum* (Tsou and Fu, 2002, 2007). Pollen cohesion in these species is generally acalymmate (four pollen grains are grouped only by partial fusion) with simple cohesion (Su and Saunders, 2003). But these studies show differences in cohesion mechanisms; thus, while pollen grains in *Pseuduvaria* are connected by wall bridges (crosswall cohesion), involving both the exine and the intine, in *A. glabra*, *A. montana*, and *Cymbopetalum* cohesion is achieved through a mass of callose-cellulose. Evolutionary transitions in flowering plant reproduction are proving to have a clear potential in plant evolutionary biology (Barrett, 2008), and the need for more detailed ontogenetic studies in the family has been put forward (Tsou and Fu, 2002). Indeed the fact of being the largest family among basal angiosperms, together with the puzzling connection mechanisms so far described in the different species examined, provide an excellent opportunity to investigate the ontogeny of pollen development and its evolutionary implications.

In this work, pollen development is characterized in *A. cherimola*, one of the species in the Annonaceae where pollen is shed aggregated in groups of four, paying

special attention to the events close to pollen formation and retention of the individual pollen grains together, observed by immunolocalization of different wall components. Results are discussed in relation to the shedding of pollen in groups of four in other species and how this event may have occurred and settled during evolution.

MATERIAL AND METHODS

Plant Material

The research was performed on adult *A. cherimola*, cv. Campas trees of located in a field cultivar collection at the EE la Mayora CSIC, Málaga, Spain. To study the relationship between flower bud length and developmental stages, tagged flower buds were measured sequentially on the trees. Buds were measured twice a week for 8 weeks from leaf unfolding, when the buds were visible but buried under the leaf petiole until anthesis. *A. cherimola*, as other members of the Annonaceae, presents protogynous dichogamy (Gottsberger, 1999). The flower opens in the female stage and remains in this stage until the following day in the afternoon when at a precise time, around 6 pm. under our environmental conditions, it changes to the male stage: the anthers dehisce, the petals open more widely and the stigmas shrivel (Baker and Baker, 1979).

Light Microscope Preparations

To follow pollen development, anthers were collected from flower buds of a range of stages, with petals 6, 9, 12, 16, 22, 24, and 30 mm long. Anthers were also collected from flowers one day prior to anthesis and at the female (F) and male (M) stages of mature flowers. The anthers from three flowers of each stage were fixed in glutaraldehyde at 2.5% in 0.03 M phosphate buffer (Sabatini et al., 1963), dehydrated in an ethanol series, embedded in Technovit 7100 (Kulzer & Co, Wehrheim, Germany), and sectioned at 2 µm.

Sections were stained with periodic acid-Schiff's reagent (PAS) for insoluble carbohydrates and with PAS/Toluidine Blue for general histological observations (Feder and O'Brien, 1968). Sections were also stained for cutine and exine with 0.01% auramine in 0.05 M phosphate buffer (Heslop-Harrison, 1977) and for cellulose with 0.007% calcofluor in water (Hughes and McCully, 1975). Intine and exine were observed with a 3:1 mixture of 0.01% auramine in water and 0.007% calcofluor in water.

To observe nuclei during pollen development, anthers collected from flowers at the same developmental stages ranging from 9 mm long to anthesis were also fixed in 3:1 (V1/V2) ethanol-acetic acid, embedded as described above, sectioned at 5 µm and stained with a solution of 0.25 mg/ml of 4',6-diamidino-2-phenylindole (DAPI) and 0.1 mg/ml p-phenylenediamine (added to reduce fading) in 0.05 M Tris (pH 7.2) for 1 hr at room temperature in a light-free environment (Williams et al., 1999). Preparations were observed under an epifluorescent Leica DM LB2 microscope with 340–380 and LP 425 filters for auramine, calcofluor, and DAPI.

For the study of pollen morphology and pollen size, dehisced anthers were sieved through a 0.26 mm mesh sieve and the pollen was placed in glacial acetic acid for acetolysis. Pollen grains were transferred to a mixture of 9:1 acetic anhydride:concentrated sulphuric acid at 65°C for 10 minutes, then washed with glacial acetic acid and washed

again three times with water following a modification of the method by Erdtman (1960).

Scanning Electron Microscopy

Pollen for scanning electron microscopy (SEM) was fresh dried with silica gel and directly attached to SEM stubs using adhesive carbon tabs and observed with a JSM-840 scanning electron microscope (JEOL) operated at 10 kV.

Immunocytochemistry

Immunocytochemistry was performed on Technovit 8100 (Kulzer & Co, Wehrheim, Germany) embedded semithin sections and revealed by fluorochromes, as described previously (Satpute et al., 2005; Solís et al., 2008). Anthers from three flowers per developmental stage with petals 6, 9, 12, 16, 22, 24, and 30 mm long and at anthesis were fixed in 4% paraformaldehyde in phosphate buffered saline (PBS) at pH 7.3 overnight at 4°C, dehydrated in an acetone series, embedded in Technovit 8100 (Kulzer), polymerized at 4°C and sectioned at 2 µm. Sections were placed in a drop of water on a slide covered with 3-Aminopropyltriethoxy-silane 2% and dried at room temperature.

Different antibodies were used to localize specific cell components: an anti-RNA mouse monoclonal antibody, D44 (Eilat and Fischel, 1991; Mena et al., 1994), for total RNA detection; JIM5 and JIM7 rat monoclonal antibodies (Professor Keith Roberts, John Innes Centre, Norwich, UK) which respectively recognize low and high-methyl-esterified pectins (Knox, 1997) for localization of pectins; and an anti-callose mouse monoclonal antibody (Biosupplies, Parkville, Australia) for callose.

Sections were incubated with PBS for 5 min and later with 5% bovine serum albumin (BSA) in PBS for 5 min. Then, different sections were incubated for 1 hr with the primary antibodies: JIM5, JIM7, and anti-RNA undiluted and anti-callose diluted 1/20 in PBS. After three washes in PBS, the sections were incubated for 45 min in the dark with the corresponding secondary antibodies (anti-rat, for JIM5 and JIM7, and anti-mouse, for anti-RNA and anti-callose) conjugated with Alexa 488 fluorochrome (Molecular Probes, Eugene, Oregon, USA) and diluted 1/25 in PBS. After three washes in PBS and water, the sections were mounted in Mowiol 4–88 (Polysciences), examined with a Zeiss Axioplan epifluorescent microscope, and photographed with a CCD Digital Leica DFC 350 FX camera.

DISCUSSION

Pollen Development

Pollen in *A. cherimola* is shed in groups of four, originating from the same meiotic division and, hence, the same tetrad. Pollen development, however, continues beyond tetrad formation and, although held together, pollen grains are fully mature upon anther dehiscence. Meiosis cytokinesis occurred through the formation of ingrowths of callose that are also found in genera of some primitive angiosperms (Furness et al., 2002; Nadot et al., 2008) including species of the Magnoliid clade as *Magnolia tripetala* (Farr, 1918) and *Degeneria vitiensis* (Dahl and Rowley, 1965) in the Magnoliales, *Laurelia novae-zelandiae* (Sampson, 1969) and *Liriodendron tulipifera* (Guzzo, 1995)

in the Laurales or *Asarum* in the Piperales (Furness et al., 2002) as well as in monocots as *Sisyrinchium* (Farr, 1922).

Starch accumulated prior to meiosis and the first pollen mitosis and vanished with the onset of these two divisions; this also occurred 6 days before anther dehiscence, preceding the shedding of starchless pollen. The accumulation of starch in PMC has also been reported in other primitive angiosperms, such as *Anaxagorea brevipes* (Gabareva, 1995) or *Austrobaileya maculata* (Zavada, 1984), and in other evolutionarily more recent angiosperms (Pacini and Franchi, 1988; Xiang-Yuan and Demason, 1984). Vacuolization also follows a conserved pattern (Bedinger, 1992; Maheshwari, 1950). The cytoplasm enlarges through a first vacuolization and, later on, following the first pollen mitosis, small vacuoles appear as starch builds up.

In most species a dehydration process takes place prior to pollen shedding and starch is hydrolyzed to form sucrose that protects pollen against desiccation (Franchi et al., 1996). Starchless pollen is the most common pollen type in the angiosperms (Grayum, 1985), being more frequent in bicellular than in tricellular pollen (Baker and Baker, 1979; Brewbaker, 1967). In *A. cherimola*, pollen is shed in a highly hydrated stage (Lora et al., 2006) and this lack of dehydration may explain why the second mitotic division continues in free pollen after pollen shedding, producing a mixed population of bi and tricellular pollen (Lora et al., 2009). However, both types of pollen are starchless at anther dehiscence.

Pollen Cohesion

Several reasons could account for the release of pollen in groups of four. In *Arabidopsis*, failure of different enzymes during the dissolution of the pectic layer that surrounds the PMC wall has been reported in two quartet mutants (Francis et al., 2006; Rhee et al., 2003). In our work the immunolocalization of esterified and non-esterified pectins showed that, although they were clearly present in the PMC wall, the pectins disappeared following tetrad formation. A closer examination of the photographs reveals that the PMC wall, which stains for cellulose, remained beyond the tetrad stage. Interestingly, quartet mutants of *Arabidopsis* also show a defect in the degradation of the PMC wall (Rhee et al., 2003). Cellulase has also been shown to be involved in the breakdown process of the PMC wall (Neelam and Sexton, 1995) and a delay in its action could lead to this phenomenon. However, this failure does not seem permanent, since 25 days later this wall is completely dissolved. Thus, the permanence of the PMC wall appears as a key factor contributing to pollen grouping as pollen tetrads in *A. cherimola*, similar to the observations in *Arabidopsis* mutants. A mixture of enzymes is required to break down the complex PMC wall (Scott et al., 2006), and a failure of one or more of these enzymes could result in a similar final result.

Different events that take place during the retention of this wall may explain pollen adherence once this wall disappears. The confining of pollen within the tapetal chamber keeping the young microspores in close proximity may contribute to this wall maintenance. Surprisingly, the young microspores are apparently separated from their sibling cells allowing some free movement indicated by the strange 180° rotation

of the pollen aperture sites. Thus, those aperture sites that originally looked outwards rotate inward to face each other. A similar rotation has been reported previously in other Annonaceae (*A. glabra* and *A. montana* (Tsou and Fu, 2002)) and *Cymbopetalum* (Tsou and Fu, 2007), and also in species of the Poaceae (Rowley, 1964). This distal-proximal microspore polarity transition in development contrasts with the evolutionary shift from a proximal to a distal aperture that has been long regarded as one of the major evolutionary innovations in seed plants (Rudall and Bateman, 2007). Proximal apertures predominate in the spores of mosses, lycophytes, and ferns while distal apertures are more common in extant seed plants including gymnosperms, cycads and early-divergent angiosperms (Rudall and Bateman, 2007). In fact, species in the Annonaceae with monad pollen are reported to have distal apertures (see Le Thomas, 1980 for review). However, a complete study of 25 Annonaceae genera with species that release aggregated pollen showed proximal apertures (Walker, 1971a) and, consequently, the distal-proximal transition observed in pollen development of *A. cherimola* and other Annonaceae (Tsou and Fu, 2002, 2007) could represent a widespread situation in this basal family.

Another reason proposed for this permanent binding of pollen in groups of four could be a failure in the synthesis of the callose layer during microspore separation in the tetrad (Blackmore and Crane, 1988). However, the results shown in this work in *A. cherimola* indicate that callose is layered following the standard pattern and vanishes later, after meiosis is completed, similar to the way it occurs in *Arabidopsis* quartet mutants, in which callose dissolution proceeds normally (Rhee and Somerville, 1998). However, the use of antibodies against callose showed that callose remains for a while in the area where pollen apertures will form hampering the layering of sporopollenin. Callose remnants in this area have also been reported in other Annonaceae and it has been suggested that these remnants pull the pollen grains to undergo the 180° turning (Tsou and Fu, 2002, 2007). In the formation of the pollen wall, callose dissolution occurs concomitantly with the layering of the exine (Testillano et al., 1995) and the formation of the pollen aperture is related to endoplasmic reticulum blocking the deposition of primexine (Blackmore et al., 2007). The callose remnant at the pollen aperture sites has not been investigated in detail in other species and, given the high conservation of pollen ontogeny in angiosperms, this is a topic worthy of a detailed study. Interestingly, in an *Arabidopsis* mutant lacking the gene responsible for callose synthesis, pollen develops unusual pore structures (Enns et al., 2005).

Further binding at the aperture sites could follow this initial adhesion process through the observed joint deposition of pollenkit that has also been reported in other species (Pacini and Franchi, 1999). Thus, two key processes could contribute to holding together the four pollen grains in *A. cherimola*, the confined space that permits the delay in the dissolution of the PMC wall and the tapetal chamber and pollen rotation that allows the adhesion of the sticky proximal faces by the formation of small pectin bridges. Later, the joint deposition of sporopollenin would further strengthen this initial binding.

Biological Significance of the Pollen Dispersal Unit

A failure or delay in the dissolution of the PMC wall and tapetal chamber appears to be a critical step, resulting in the continued proximity of the four microspores produced by meiosis of a single PMC. However, this phenotype could also result from failure in the different enzymes that dissolve the PMC wall. The distribution of this character, together with the information provided by *Arabidopsis* mutants, shows that this has occurred independently several times during evolution, suggesting that it must provide some evolutionary advantages (Blackmore and Crane, 1988).

The adaptive advantages derived from aggregated pollen have been reviewed recently (Harder and Johnson, 2008). The release of aggregated pollen in insect pollinated species could increase pollination efficiency, since more pollen grains could be transferred in a single pollinator visit and, in this sense, a correlation between pollen tetrads and polyads with a high number of ovules per flower has been shown in a survey of the Annonaceae (Walker, 1971b). The release of aggregated pollen is more advantageous in situations where pollinators are infrequent (Harder and Johnson, 2008) and in situations of short pollen viability and pollen transport periods. A short pollen viability period has been reported in *A. cherimola*, (Lora et al., 2006; Rosell et al., 1999) and a short pollen transport episode is common in several Annonaceae (Ratnayake et al., 2006) and in other beetle pollinated species of early divergent angiosperm lineages (Davis et al., 2008).

An additional possible benefit of aggregated pollen is protection against desiccation and entry of pathogens through the thin walls of the pollen aperture sites. Pollen grouped in dyads, tetrads or polyads show a strong proximal reduction of the exine in Annonaceae (Le Thomas et al., 1986). *A. cherimola* pollen is inaperturate and germinates in the proximal face, showing a large area of unprotected intine (Lora et al., 2006; Rosell et al., 2006). More evolutionarily recent species present a colpus that, in dehydrated pollen, is just a narrow slit protected by loose pollenkit. Only upon hydration, when the pollen faces a wet surface on the stigma, this slit swells developing a wider colpus through which the pollen tube protrudes (Heslop-Harrison and Heslop-Harrison, 1992). Inaperturate pollen does not have this protection from desiccation and the development of inward facing intines may play a role in protecting pollen against desiccation.

RESULTS

The mature *A. cherimola* flower is a syncarpous gynoecium with a conic shape composed of about 100 fused carpels surrounded at its base by several rows of anthers with up to 200 stamens, encircled by two whorls of three petals. The flower cycle from opening to anther dehiscence lasts 2 days: the flower opens on the morning of the first day in the female stage and remains in this stage until the afternoon of the following day when the flower enters the male stage. Anther dehiscence occurs concomitantly in all stamens of a flower and, as the anthers dehisce, they detach from the flower and fall over the open petals.

Flower buds of *A. cherimola* develop in the leaf axes following leaf expansion; the basal nodes are differentiated in the year preceding anthesis. The uppermost distal buds differentiate in synchrony with shoot growth (Higuchi and Utsunomiya, 1999). Flower bud growth begins 39 days prior to anthesis. Anther differentiation proceeded centripetally, with the most developmentally advanced anthers placed in the outermost rows and the different stages of anther and pollen development present within the same flower. This fact was helpful for establishing successive stages of anther development. The anther becomes septate with PMC positioned between rows of interstitial tapetum similar to the anthers described in a sister species, *Annona squamosa* (Periasamy and Kandasamy, 1981).

To determine if pollen development followed a standard pattern and whether pollen tetrads at anther dehiscence corresponds with the cytological and morphological features of mature pollen, anther and pollen development were examined from microsporogenesis to maturity. Special attention was given to the events responsible for pollen cohesion. Microgametogenesis and tapetum degeneration were also examined sequentially.

Microsporogenesis

Initial hypodermal archesporial cells were apparent 24 days before anthesis (Figure 1A). From them, anther septa initials and PMC developed in 9 cm long flower buds 19 days before anthesis (Figure 1B). Each anther contained a uniserial row of PMC with a conspicuous common wall. The PMC began to accumulate starch grains (Figure 1C) and increased in size (Figures 1C and 1D). Starch grains vanished concomitantly with the beginning of meiosis, some 15 days before anthesis, as a translucent cell wall layer was apparent surrounding the PMC (Figure 1D). Meiosis proceeded rapidly and was followed by a new accumulation of starch grains in the young microspores (Figure 1E) 14 days before anthesis. The microspore tetrads remained together in isolated tapetal chambers surrounded by the PMC wall that stained positively with PAS for carbohydrates (Figures 1F and 1G).

Immunocytochemical essays revealed the localization of various cell wall components (Figure 2). Callose surrounded the PMC wall and, following meiosis I, an additional furrow of callose developed inwards (Figure 2A) forming a callose positive band between the dyad cells (Figures 2B, 2C). Successive cytokinesis followed (Figure 2C), resulting in a tetrad (2D), each separated by callose. Dyad and tetrad stages coexist in the flower as centripetal maturation progresses. The PMC wall also reacted positively to JIM7 (Figure 2E) and JIM5 (Figure 2F) staining, indicating the presence of methyl-esterified and unesterified pectins respectively. However, while the walls of the anther somatic and tapetal cells also reacted positively to the JIM7 for methyl-esterified pectins (Figure 2E), they showed only a faint staining for the presence of unesterified pectins (Figure 2F).

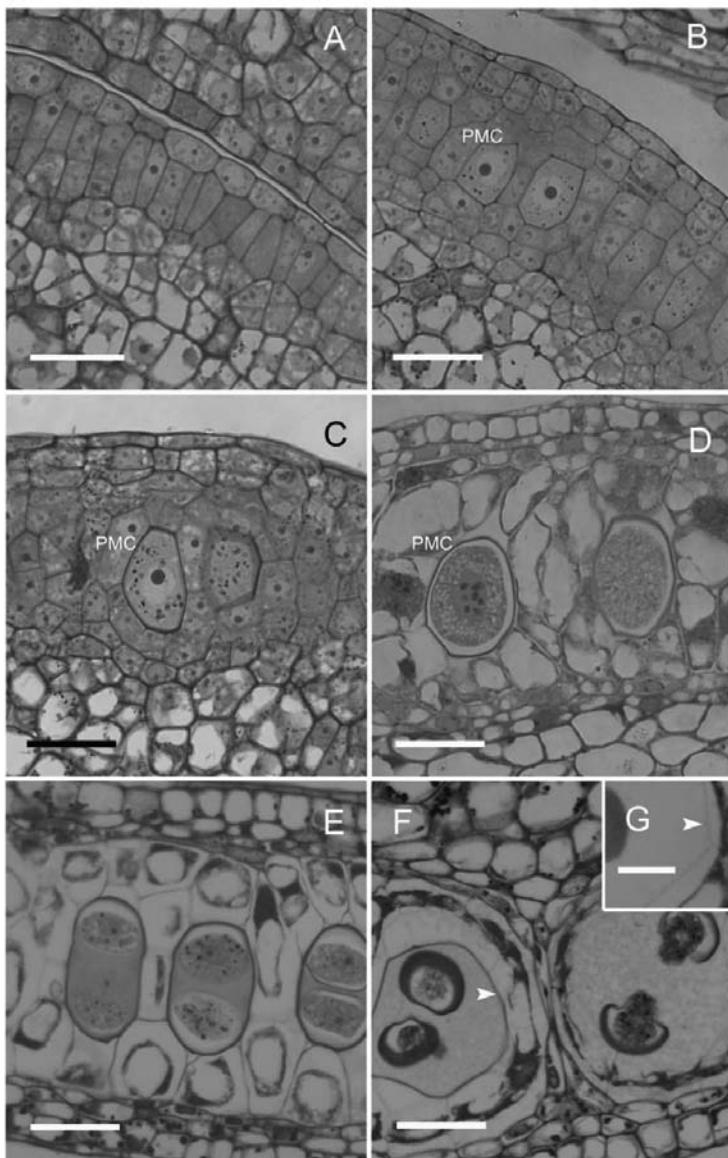


Figure 1. Microsporogenesis of *Annona cherimola*. (A) Uniseriate row of arquesporial cells. (B) Septal and pollen mother cells (PMC), showing a conspicuous wall, alternate in the sporogenous tissue. (C) PMC increase in size and starch grains are visible. (D) Starch grains vanish, a translucent layer appears in the PMC wall, and PMC starts meiosis. The tapetum vacuolates and the tapetal chambers are apparent. (E) Following meiosis, starch grains accumulate again in the young microspores, which are surrounded by callose. (F) The young microspores, with an incipient exine, appear to float and turn within the still remaining PMC wall (arrow) that holds the four microspores together. (G) Detail of PMC wall (arrow). Longitudinal sections of the anthers stained with PAS and Toluidine. Bar = 20 μ m.

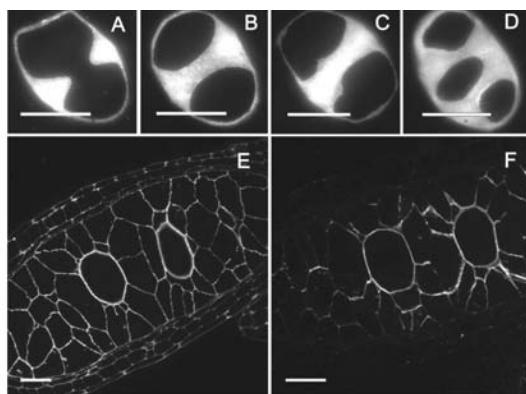


Figure 2. Callose and pectins during microsporogenesis in *Annona cherimola*. (A–D) Anticallose in dyad/tetrad phases. (A) A furrow of callose developed inwards, forming a wall between the dyad cells. (B) Dyad phase, showing (B) an incipient, and (C) A well developed callose wall. (D) Tetrad microspore showing in the section plane three of the microspores separated by callose walls. (E) PMC and other anther tissue walls showing methyl-esterified pectins. (F) PMC wall also shows unesterified pectins. Specific cell components were localized using antibodies against callose (A, B, C, D), methyl-esterified pectin (JIM7) (E), and unesterified pectin (JIM5) (F). A-D Bar = 10 µm. E-F: Bar = 20 µm.

Pollen Cohesion

Following tetrad formation, callose disappeared but the microspores remained within the PMC calcofluor-positive cellulosic wall (Figure 3A). At this developmental stage, an interesting event was detected: the microspores within each tapetal chamber, which initially had their pollen aperture sites facing outward towards the PMC wall (Figure 3A), appeared to float and rotate within their individual chambers (Figure 3B). This movement was not random, but the microspores turned 180° until the pollen aperture sites faced each other (Figure 3B). The remaining PMC cellulosic wall, which persisted for some time, together with the confined space provided by the tapetal chamber, appear to contribute towards keeping the microspore tetrad together. Subsequently, the PMC cellulosic wall disappeared completely (Figure 3C) and the tapetum degenerated as the microspores increased in size. They remained in their new orientation attached by their apparently sticky aperture sites that now faced each other (Figure 3D).

At this stage, both the cell walls of the somatic cells of the anther and the inner wall of microspores (intine) reacted similarly for methyl-esterified pectins (Figure 4A), while unesterified pectins were present just in the microspore intine (Figure 4B). The exine showed a low unspecific autofluorescence but in a different fluorescence wavelength (yellowish color) than the fluorescent marker of the antibodies, AlexaFluor 488, which emitted green fluorescence. As a consequence, exine autofluorescence was clearly differentiated from the immunofluorescence signals. Anti-callose immunofluorescence revealed remnants of callose at the pollen aperture sites where the thick external layer of the microspore wall, the exine, was extremely thin or absent. These callose remnants were apparent in all microspores at this stage (Figure 4C). The four microspores showed crosswall cohesion bridges that stained with antibodies against unesterified and methyl-esterified pectins in the microspore wall (Figures 4D–H).

Following this inter-intine cohesion, additional deposition of sporopollenin with a joint layering of the four microspores further strengthened this connection (Figure 4I).

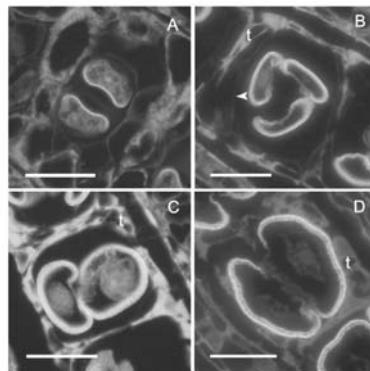


Figure 3. Pollen development within the tapetal chamber in *Annona cherimola*. (A) Two young microspores in a tetrad which still keeps the pollen mother cell wall. Aperture sites are located towards the outside facing the pollen mother cell wall. (B) Pollen is shed free, within the PMC wall, in the tapetal chamber. Within this confined space the young microspores turn (C) with their aperture sites facing now each other as the PMC cellulosic wall is digested. (D) The pollen grains regroup sticking through the aperture sites, and enlarge as the tapetum degenerates. Longitudinal anther 2 µm resin sections stained with calcofluor and auramine. Bar = 20 µm.

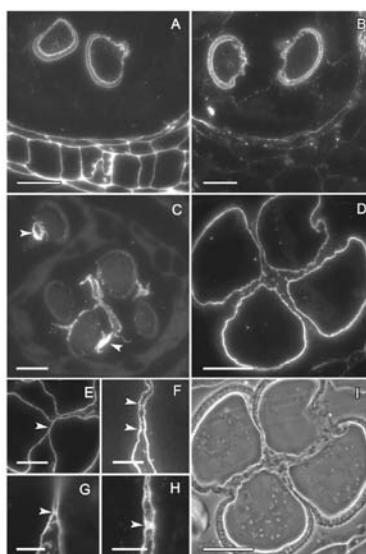


Figure 4. Establishment of pollen cohesion in *Annona cherimola*. (A) Microspore walls show methyl-esterified pectins, and also (B) unesterified pectins. (C) As callose is digested, remnants of callose (white arrow) are observed layering the pollen aperture sites. (D–F) Microspores show crosswall cohesion bridges showing the presence of unesterified pectins. (G–H) Details of crosswall cohesion bridges, showing the presence of methyl-esterified pectins. (I) Phase contrast of a mature pollen grain showing internal cohesion and a joint sporopollenin layering. Specific cell components were localized using antibodies against: methyl-esterified pectin (JIM7) (A, –H), unesterified pectin (JIM5) (B, D, E, F) Callose (C). A–E, I: Bar = 10 µm. F–H: Bar = 3 µm.

Microgametogenesis

As the microspores increased in size, their cytoplasm became vacuolated (Figure 5A) and starch grains were absent (Figure 5B). During this vacuolization, nuclear migration preceded the first mitosis to form bicellular pollen grains. Following the first mitosis, 4–6 days before anthesis, the vacuoles decreased in size (Figure 5C) and starch was again stored (Figure 5D). Young pollen grains had no vacuoles (Figure 5E) and numerous starch grains were present (Figure 5F). The second mitotic division producing the first tricellular pollen grains started some 4 hr prior to anther dehiscence (Figure 5G) and one day prior to anther dehiscence starch grains began to hydrolyze (Figure 5H). Mitotic divisions were not synchronized within a pollen tetrad and single pollen grains with different numbers of nuclei could be observed in the same tetrad resulting in the coexistence of bicellular and tricellular pollen upon anther dehiscence.

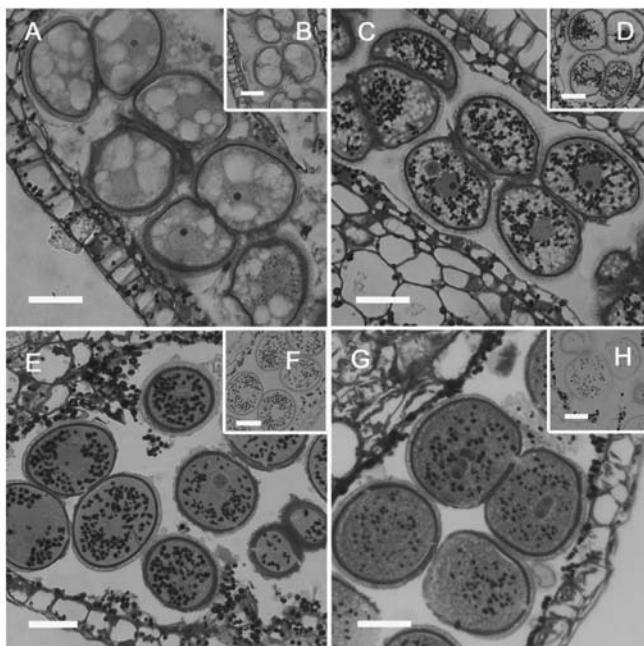


Figure 5. Microgametogenesis in *Annona cherimola*. (A) Microspores increase in size as vacuoles appear in the cytoplasm, (B) Microspores at this stage do not have starch grains. (C) Microspores following mitosis I; (D) as vacuoles decrease in size, starch grains are stored. (E) Young pollen grains without vacuoles (F), which accumulated starch grains. (G) Close to the time of anther dehiscence, the second mitosis occurs, the tapetum is completely degenerated and (H) starch is digested. Longitudinal sections of anthers stained with PAS and Toluidine (A, C, E, G), and with PAS (B, D, F, H) to show starch grains. Bar = 20 μ m.

In mature pollen, while intine thickness was similar around the pollen grain, the exine was thinner or absent at the pollen aperture sites where contact points between siblings pollen grains were established (Figure 6A). At these areas unesterified and methyl-esterified pectin bridges were maintained throughout pollen development although

these connections seemed to be less strong in mature pollen (Figure 6B). However, mature pollen tetrads resisted separation during acetolysis, showing the permanence of joint sporopollenin (Figure 6C).

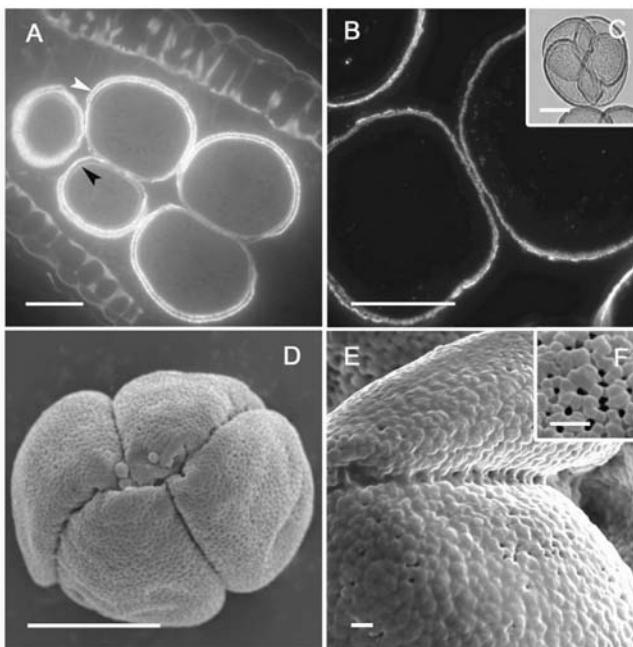


Figure 6. Mature pollen of *Annona cherimola*. (A) Intine (black arrow) is similar all around the pollen grain, but exine (white arrow) is thinner in the pollen aperture site. Longitudinal section stained with a 3:1 mixture of Auramine and Calcofluor. (B) Sibling pollen grains have a faint cohesion that showed with JIM 5 antibody the presence of unesterified pectins. (C) Mature pollen tetrad following acetolysis. (D, E, F) Mature pollen observed with scanning electron microscopy (SEM). (D) Mature pollen grains with a globose shape and a radiosymmetric disposition. (E) Exine cohesion helps keeping sibling pollen grains together. (F) Pollen exine shows a tectate perforate appearance. A, B, D: Bar = 20 µm; C: Bar = 10 µm; E, F: Bar = 2 µm.

Scanning electron micrographs revealed that mature pollen had a radiosymmetric globose shape, was inaperturate, tectate perforate, and with a diameter of 40 µm (Figures 6D–6F). Mature pollen was shed in groups of four sibling pollen grains that stick together having an exine cohesion, clearly visible with high magnification SEM images (Figure 6E).

Tapetum Degeneration

A. cherimola has a secretory tapetum with tapetal-type septa similar to those described in other species of the genus *Annona* such as *A. squamosa* (Periasamy and Kandasamy, 1981) and *A. glabra* (Tsou and David, 2003). Prior to meiosis, septal initials formed tapetal chambers that host the PMC (Figure 7A). After meiosis, the tapetum showed a vacuolization and a progressive degeneration as the tapetal chamber enlarged

(Figure 7B). The nuclei of the tapetal cells displayed elongated and lobular shapes together with a extremely high chromatin condensation, revealed by an intense DAPI fluorescence (Figure 7C), typical features of programmed cell death (Wu and Cheung, 2000), which have also been found in the tapetal nuclei of other species (Testillano et al., 2007). At the same time, tapetal cells released their cellular contents that coated the pollen grains to form the pollenkit. At anther dehiscence the tapetum was completely degenerated and had disappeared (Figure 7D).

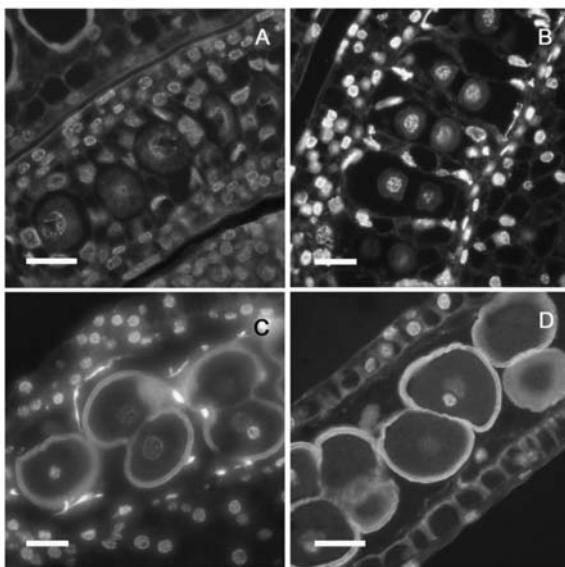


Figure 7. Tapetum degeneration in *Annona cherimola*. (A) Pollen mother cells in Prophase I and an active tapetum. (B) Dyad phase in enlarged tapetal chambers. (C) Anther, 4 days before anthesis, showing bicellular pollen and degenerated tapetum, with nuclei displaying elongated shapes and chromatin condensation. (D) Tapetum has disappeared in anthers of flowers at the female stage showing mature pollen. Longitudinal 5 μ m resin sections stained with DAPI. Bar = 20 μ m.

CONCLUSION

The results obtained in this work support the hypothesis that aggregated pollen could be the result of relatively minor ontogenetic changes beneficial for pollen transfer or/and protection from pollen desiccation. Comparison of the events reported here with those recorded in recent pollen development mutants in *Arabidopsis* suggests that a simple event along development, the delay in the dissolution of the pollen mother cell wall and tapetal chamber, results in conspicuous morphological changes that lead to the release of pollen in tetrads. A variety of different mutations within the enzymes required to breakdown this wall, may contribute to this common morphology. These changes have occurred and recur in nature and, due to their adaptive advantages for pollen transfer, have been selected during evolution several independent times, representing an example of convergent evolution.

KEYWORDS

- **Anthesis**
- **Microsporocytes**
- **Protogynous dichogamy**
- **Scanning electron microscopy (SEM)**
- **Vacuolization**

AUTHORS' CONTRIBUTIONS

Jorge Lora performed most of the experimental analyses, Pilar S. Testillano had an active contribution to the immunocytochemistry assays, María C. Risueño designed and discussed the immunocytochemistry essays, José I. Hormaza participated in the design of the experiments, María Herrero coordinated the study. All authors contributed to the draft and read and approved the final manuscript.

ACKNOWLEDGMENTS

Financial support for this work was provided by the Spanish Ministry of Science and Innovation (Project Grants AGL2004-02290/AGR, AGL2006-13529-C01, AGL2007-60130/AGR, AGL2008-04255 and BFU2008-00203), GIC-Aragón 43, Junta de Andalucía (AGR2742) and the European Union under the INCO-DEV program (Contract 015100). Jorge Lora was supported by a grant of the Junta de Andalucía. The authors thank K. Pinney for helpful suggestions to improve the chapter.

Chapter 10

DNA Barcoding in the Cycadales

Chodon Sass, Damon P. Little, Dennis Wm. Stevenson,
and Chelsea D. Specht

INTRODUCTION

Barcodes are short segments of DNA that can be used to uniquely identify an unknown specimen to species, particularly when diagnostic morphological features are absent. These sequences could offer a new forensic tool in plant and animal conservation—especially for endangered species such as members of the Cycadales. Ideally, barcodes could be used to positively identify illegally obtained material even in cases where diagnostic features have been purposefully removed or to release confiscated organisms into the proper breeding population. In order to be useful, a DNA barcode sequence must not only easily PCR amplify with universal or near-universal reaction conditions and primers, but also contain enough variation to generate unique identifiers at either the species or population levels. Chloroplast regions suggested by the Plant Working Group (PWG) of the Consortium for the Barcode of Life (CBoL), and two alternatives, the chloroplast *psbA-trnH* intergenic spacer and the nuclear ribosomal internal transcribed spacer (nrITS), were tested for their utility in generating unique identifiers for members of the Cycadales. Ease of amplification and sequence generation with universal primers and reaction conditions was determined for each of the seven proposed markers. While none of the proposed markers provided unique identifiers for all species tested, nrITS showed the most promise in terms of variability, although sequencing difficulties remain a drawback. We suggest a workflow for DNA barcoding, including database generation and management, which will ultimately be necessary if we are to succeed in establishing a universal DNA barcode for plants.

Barcode all described species is an enormous task with large sums being spent annually toward this end (Ebach and Holdrege, 2005). The proposed utility of the Barcode of Life project has been debated (Chase et al., 2005; Hebert et al., 2003, 2004; Kress et al., 2005; Meier et al., 2006; Rubinoff, 2006; Will and Rubinoff, 2004) and fundamental challenges have been acknowledged that focus on (a) The identification of DNA regions useful at the appropriate taxonomic level, (b) The development of universal primers for these regions, and (c) The proper use of DNA barcodes as taxonomic identifiers. Proponents argue that molecular barcodes can be used to identify new species and eliminate the need for the complex taxonomic training that is currently required for species description and identification (Blaxter, 2004) helping to ease the taxonomic crisis, especially in countries with high biodiversity and small numbers of practicing taxonomists. However, the patterns of sequence variation make it logically impossible to use DNA barcodes for species circumscription as originally proposed (see Meier et al., 2006 for an empirical example, see Little and Stevenson, 2007 for a theoretical

example). Although barcodes are appealing as a powerful tool to identify already described species, the cautious among us argue that the use of a single locus for identification may produce misleading results especially considering the different evolutionary histories of organellar and nuclear genomes within a single species (Moritz and Cicero, 2004). Moreover, there is limited intraspecific sequence variation data for the proposed barcoding loci in plants. Others reject the use of barcodes for taxonomic purposes on the grounds that species description and identification requires full taxonomic revisions and that “phylogenies” produced by barcoding genes do not necessarily represent evolutionary history (Will and Rubinoff, 2004; Will et al., 2005).

Ultimately the ability to identify a sample to species could be useful in cases where specimens are not of adequate quality to make accurate identifications (e.g., adult forms vs. larval forms, sterile vouchers of plant specimens) and for ecologists and conservation biologists to rapidly assess biological diversity. In this sense, barcoding acts as a “forensic” tool for the accurate identification of a sample to species. The species, in this case, needs to be both described as unique (i.e., monographed) with a known range of morphological and sequence variation and be represented in a DNA barcoding database. This is an enormous task, requiring active participation of taxonomists, DNA sequencing facilities, database managers, and funding agencies to support monography, DNA sequencing, continuous specimen, and database management, and potentially, the recircumscription of species as new data become available.

In order for a region of DNA to be operative as a barcode, it must simultaneously contain enough variability to be informative for identification (i.e., contain unique identifiers), be short enough to sequence in a single reaction, and contain invariant regions that can be used to develop universal primers (Stoeckle, 2003). Unfortunately, it is difficult to find a single region of DNA that has all three of these properties. For animals, the mitochondrial cytochrome oxidase I (*coxI*) gene has been successfully used for identification (Hebert et al., 2003, 2004; Ward et al., 2005) although there are exceptions (e.g., see Gompert et al., 2006; Meier et al., 2006; Meyer and Paulay, 2005). For land plants, the *coxI* gene, and the mitochondrial genome in general, is not useful for identification at the species level because of low levels of primary sequence variability (Cho et al., 2004; Palmer and Herbon, 1988) Other regions often used for phylogenetic analysis across large groups of plants (e.g., *rbcl*) do not usually contain enough variability to identify individual species (but see Driscoll and Barrington, 2007; Kress and Erickson, 2007). Developing a barcoding region for plants is further complicated by extensive genome-wide horizontal gene transfer, hybridization, and homoplasy (Chase et al., 2005).

Despite these obstacles, several gene regions have recently been proposed for use in land plants (Chase et al., 2007; Kress et al., 2005; Presting, 2006) One set of loci includes a nuclear region, the ribosomal internal transcribed spacer with embedded 5.8S (nrITS), and a chloroplast region, the *psbA-trnH* intergenic spacer (Kress et al., 2005) The combination of these two regions to positively and accurately identify taxa to species was tested on a subset of plants in the published analysis, but the combination is predicted to yield difficulties at the species level because nrITS is extremely variable in length—making analysis potentially more difficult—and *psbA-trnH* is likely to

provide insufficient variation to reliably identify an organism to species, especially in groups with low divergence.

A portion of the chloroplast encoded large subunit ribosomal DNA, that is potentially “universally” amplifiable (Universal Plastid Aplicon; UPA), has also been proposed as a barcode for photosynthetic organisms. Available data suggest that although UPA may be variable at the species level in some algal lineages, it is not particularly variable among land plants (Presting, 2006).

A consortium of institutions operating under the auspices of the PWG of the CBoL initially suggested five chloroplast gene regions for evaluation as potential barcodes: *matK*, *rpoC1*, *rpoB*, *accD*, and *YCF5*, with *ndhJ* as a potential sixth region (<http://www.kew.org/barcoding/>). These markers were proposed because of their potential for amplification with universal primers and because they may harbor sufficient sequence diversity, individually or in combination, to distinguish among species. In order for either of these criteria to be demonstrated, members of the community must devote time and effort to evaluate the proposed regions in the plant group they study with the goal of developing a defined “barcoding workflow” for the taxonomic group in question (Figure 1).

The last step of any barcoding workflow is to use newly generated sequence data in combination with a well-maintained database to positively identify the species in question. The BOLD identification system developed as part of the ongoing barcoding initiative at the University of Guelph (www.barcodinglife.org: (Ratnasingham and Hebert, 2007)) uses a Hidden-Markov Model to align a query sequence to a reference database of *coxI* sequences generated for animal barcodes and then select the most similar sequence(s) as the identification. Unfortunately this algorithm is only applicable to sequences that can be globally aligned (Ratnasingham and Hebert, 2007). Some of the proposed plant barcodes are non-coding regions that cannot be sensibly aligned across land plants and thus could not benefit from BOLD-ID. Little and Stevenson (Little and Stevenson, 2007) demonstrated that search algorithms can be successfully used on unaligned nucleotide sequence data, the most accurate and precise algorithms were, respectively, the commonly used local alignment search tool, BLAST (Altschul et al., 1990) and a diagnostic method, DNA-BAR (DasGupta, 2005a, 2005b). The DNA-BAR was originally intended as an algorithmic tool to select oligonucleotides for identification of microorganisms by Southern hybridization, but DNA-BAR’s output file can be queried by a PERL script (DEGENBAR) that uses a simple matching algorithm to pick the most similar sequences(s) in the reference database. Provided that DNA-BAR is run on an input file containing each sequence and its reverse complement, both forward and reverse query sequences can be used to search the reference database.

The Cycadales are unique in their evolutionary position and importance for conservation, and as such are important to include in tests of proposed barcoding regions. Cycads are often thought of as “living fossils” and the extant taxonomic assemblage represents only a sampling of the ancient diversity. Most extant genera have representative fossils that date to the tertiary with some dating to the early Permian—indicating a minimum of 50–60 million years of morphological evolution that might enable us to observe greater nucleotide divergence than one would expect in more recently

derived species (Hermsen et al., 2006; Jones, 1993; Whitelock, 2002). Because of the relictual nature of the genera and their high value in illegal horticultural trade, cycads are an important focus for conservation efforts (Donaldson, 2003). Most cycad genera are listed in CITES (2007). An easy-to-use and inexpensive identification system would enable non-experts to identify illegally harvested individuals and help prevent the illegal trade of these species. Ideally it would be possible to identify an individual to species and perhaps even identify the population from whence the specimen was removed, allowing for proper repatriation of illegally harvested individuals.

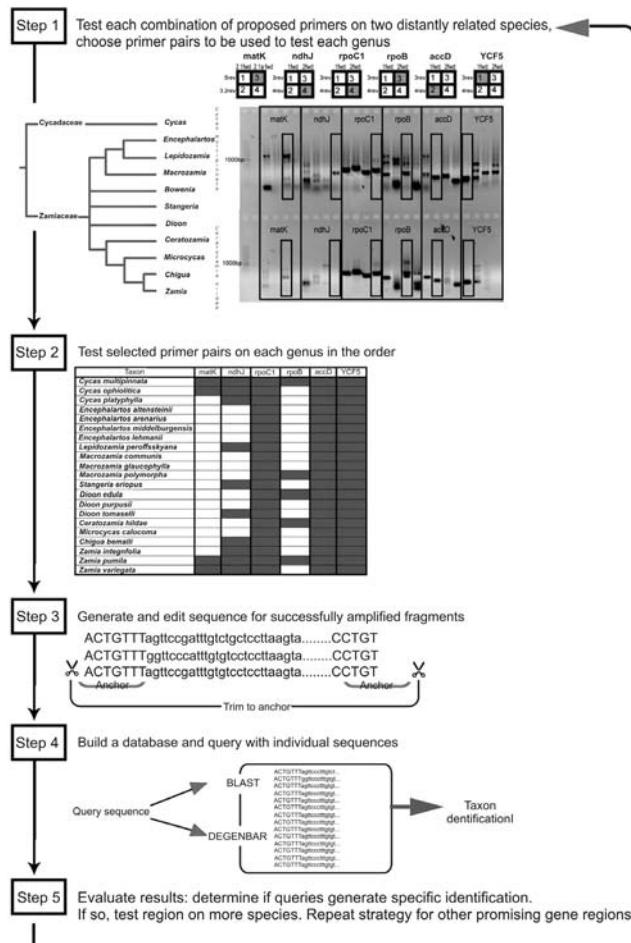


Figure 1. Barcoding optimization workflow. Step 1: genera used for testing all primer pairs, amplification products of each combination of primer pairs, and primer pair combination chosen for testing on more genera. Step 2: taxa subject to further testing and success of amplification with chosen primer pairs (highlighted in gray). Step 3: trimming all sequences to highly similar anchor regions. When possible, anchor regions were actually the primer binding sites. Step 4: each sequence entered into a database and used as a query sequence. The process is repeated for more species, with promising regions, or with new markers. The PWG suggested primer regions (www.kew.org/barcoding) are used as the example.

The only way to determine if it is possible to use DNA barcodes across a wide variety of plant life is to test the proposed loci and search algorithms. In this study, we test the proposed barcoding regions in the members of the ancient gymnosperm order Cycadales in an effort to develop a functional barcoding workflow for this order.

MATERIALS AND METHODS

Taxon Sampling and Primer Testing Strategy

For each region, the PWG designed four primers (two forward and two reverse) in their Phase I trials (www.kew.org/barcoding) in an attempt to increase the likelihood of finding a working combination. The primer pairs were first tested in all combinations on two species—*Cycas ophiolitica* and *Ceratozamia hildae*—chosen based on their distant placement in the Cycadales phylogeny. Primers were considered successful if they amplified a single product. If a single band was obtained by more than one primer pair, the pair that generated the largest and brightest (highest PCR yield) of the bands was chosen. If amplification was successful in only one of the two species, the pair generating the brightest band for that species was selected. The best working primers were then tested for a set of 21 species representing 10 of 11 cycad genera. Gene regions with universal or near universal success in amplification were sequenced. Gene regions with variability that enabled specific positive identification were tested on additional species within each genus to further test the region's ability to provide identification at the species level. This workflow is outlined in Figure 1.

Plant Collection, DNA Extraction, and Amplification

Leaflets were clipped from live plants, dried in silica gel, and then stored at -80°C. Whole genomic DNA was extracted using DNeasy Plant Mini Kits (Qiagen, Valencia, CA) or a modified CTAB method (Doyle and Doyle, 1987) from fresh or frozen tissue. The PCR amplification was performed from genomic DNA according to instructions on Kew's website (www.kew.org/barcoding) for the six chloroplast regions or following Kress et al. (2005) (Table 1). Some modifications were made to accommodate the use of iProof™ High-Fidelity DNA polymerase (Bio-Rad, Hercules, CA). Amplified products were inspected on 1% agarose/TAE gels. Amplicon was cleaned by digestion with Exonuclease and Shrimp alkaline phosphatase or through gel extraction using the QiaQuick™ Gel Extraction Kit (Qiagen, Valencia, CA). Cycle sequencing was performed using AmpliTaq™ (Amersham, Piscataway, NJ) or BigDye® v3.1 (Applied Biosystems, Foster City, CA) sequencing chemistry and an ABI PRISM® 3100 sequencer (Applied Biosystems, Foster City, CA).

Sequence Alignment and Determination of Barcoding Regions

Sequence editing and contig generation were performed using Sequencher (Gene Codes Corp., Ann Arbor, MI). Additional sequences for nrITS were downloaded from GenBank. If sequences did not include primer regions, all sequences were trimmed to an area with highly similar (>98% identity) sequence regions at the ends of the sequence reads—anchor regions (Figure 1). This was only necessary for *nrITS* and *psbA* sequences. Sequences from these loci were longer and more variable than other regions

and as a result primer regions were not always sequenced. Sequences were used for further analysis only if they contained the anchor regions—ensuring that identification success was due to internal variability and not arbitrary factors such as sequence read length. In order to identify and trim sequences to the anchor regions, *nrITS* and *psbA-trnH* regions were aligned using CLUSTAL W (Thompson et al., 1994) and then manually adjusted using MacClade (Sinauer Associates, Inc., Sunderland, MA). After elimination of sequences with ambiguous nucleotides and non-anchor containing sequences, databases were created and individual sequences were queried against the databases with BLAST and DNA-BAR/DEGENBAR. Sequences excluded from the ordinal level database due to the absence of anchor regions were included in secondary databases that contained only a single genus provided that genus-specific anchors could be identified from sequences that were not long enough to be included in the ordinal-level database (this was only the case for *nrITS* sequences from *Encephalartos*, *Cycas*, and *Macrozamia*; Table 3).

Comparability of Results with Different Algorithms: BLAST and DNA-BAR/DEGENBAR

The same sets of sequences were used both to generate databases and as query sequences for both BLAST and DNA-BAR/DEGENBAR (Little and Stevenson, 2007). The BLAST queries were run without filtering. Before generating the database with DNA-BAR the sequences were run through a PERL script that added a reverse complement for each sequence in order to ensure that query sequences would match the database in either the forward or the reverse orientation. To test for unique species-specific barcodes that could be used for a species level identification, the sequence belonging to each species was copied from the database and used as a query sequence. If the query sequence returned an exact match only to itself, this was scored as a positive identification at the species level. If the query sequence returned an exact match to itself and other members of the same genus, this was scored as a negative identification at the species level, but a positive identification to the genus level. The DNA-BAR/DEGENBAR returns only the highest scoring match(es), so the cutoff for genus and species identification is straightforward. For BLAST, an additional constraint was added: to positively score identification at the genus level the best match as well as the next most similar sequence had to match the genus of the query sequence. If any other genus was included in the top two hits, the result was not considered genus specific.

DISCUSSION

Proposed Barcoding Loci

Three of the six regions proposed by the PWG did not easily amplify across all Cycadales. A recently posted Phase II update based on research from the PWG (<http://www.kew.org/barcoding/update.html>) indicates a new primer pair for *matK* that was successful in *Encephalartos* and may be successful across all genera of cycads, while YCF5 was determined to not be suitable as a barcode region for all land plants due to its apparent absence in bryophytes. In addition, the Phase II update suggests two options for a combined approach to DNA barcoding, involving the use of three gene

regions to accurately identify a sample to species. Option one uses *rpoC1*, *rboB*, and *matK* while option two utilizes *rpoC1*, *matK*, and *trnH-psbA*. However, neither of these options is likely to provide the resolution necessary to identify a cycad sample to species based on the results presented above. The three regions that did easily amplify and were tested (including *rpoC1*, a member of both proposed three-region barcode options) did not provide enough variation to specifically identify the cycads tested in this study. It is possible that *matK* may provide more variation than other regions tested in this analysis, and continued studies with the newly proposed primers are necessary to evaluate its utility as a barcoding region for cycads. Our results are emblematic of the challenge faced in plant DNA barcoding. Additional search is required to find regions that both amplify easily and contain variation if the goal of a universal primer set (or sets) is to be reached.

nrITS

The *nrITS* shows promise as a barcoding region because it contains enough variability to identify many samples to species. However, intraspecific differences are greater than interspecific differences in some cases. For the 14 species for which multiple individuals were sampled, eight of the BLAST queries resulted in correct species identification, but the next highest BLAST hit (based on e-value) was not the correct species (data not shown). Many differences between species can be attributed to only one base pair difference. Six of the 26 unique identifications within *Encephalartos* were only a one base pair different from the next highest BLAST hit, and 17 of the 26 had less than five base pair differences. This suggests that once all species of *Encephalartos* have been sampled there may be a complete lack of informative variation. In addition, because sequences are generated directly from amplifications of whole genomic DNA, rare alleles (less than 10% of the amplicon) may not be evident and potential variation within a species will be missed. If *nrITS* is to be used as barcoding locus, further sequencing using cloned PCR products will be necessary to ensure that all alleles of each species are captured in the database. Allelic variation could result in false identification if all alleles for each species are not included in the reference database (Jakob and Blattner, 2006). Finally, because some identifications are based on single nucleotide positions, sequencing errors could cause further false identifications.

psbA-trnH

The placement of *psbA* has shifted in and out of the chloroplast inverted repeat in various lineages making the *psbA-trnH* intergenic spacer difficult to work with. For example, in ferns *psbA* is located inside the inverted repeat (Wolf et al., 2003); in eudicots it is located outside of the repeated region; in *Pinus contorta*, *psbA* has undergone a tandem duplication, with one truncated copy (Lidholm and Gustafsson, 1991; Lidholm et al., 1991). In cycads, PCR with the *psbA-trnH* primers suggested by Kress et al. (2005) generated two products in all genera tested except *Cycas* (Figure 3). Additional primers designed to include more of the *psbA* coding region and a portion of *psbA-trnH* intergenic spacer specific to either the small or large fragment (Table 1) were tested on *Encephalartos nubimontanus*. Sequences from these amplifications had two distinct protein coding regions, indicating that *psbA* is present in two copies

Table 1. nr/TS identification success for each genus.

Gene	Primer	Sequence 5'-3'	Rxn conditions
<i>matK</i>	2.1 forward	CCTATCCATCTGGAAATCTTAG	94°C-4 min
	2.1 a forward	ATCCATCTGGAAATCTTAGTTC	94°C-30 sec
	5 reverse	GTTCTAGCACAAAGAAAGTCG	53°C-40 sec } 40 ×
	3.2 reverse	CTTCTCTGTAAAGAATTG	72°C-40 sec 72°C-7 min
<i>ndhJ</i> (1 st reserve)	1 forward	CATAGATCTTGGGCTTYGA	94°C-4 min
	2 forward	TTGGGCTTCGATTACCAAGG	94°C-30 sec
	3 reverse	ATAATCCTTACGTAAAGGCC	53°C-40 sec } 40 ×
	4 reverse	TCAATGAGCATCTGTATTTC	72°C-40 sec 72°C-7 min
<i>rpaC1</i>	1 forward	GTGGATACACTTCTTGATAATGG	94°C-4 min
	2 forward	GGCAAAGAGGGAAAGATTTCG	94°C-30 sec
	3 reverse	TGAGAAAACATAAGTAAACGGGC	53°C-40 sec } 40 ×
	4 reverse	CCATAAGCATATCTTGAGTTGG	72°C-40 sec 72°C-7 min
<i>rpoB</i>	1 forward	AAGTGCATTGTTGGAAGTGG	94°C-4 min
	2 forward	ATGCAACGTCAAGCAGTTCC	94°C-30 sec
	3 reverse	CCGTATGTGAAAAGAAGTATA	53°C-40 sec } 40 ×
	4 reverse	GATCCCAGCATCACAAATTCC	72°C-40 sec 72°C-7 min
<i>accD</i>	1 forward	AGTATGGATCCGTAGTAGG	94°C-4 min
	2 forward	GGRGCACGTATGCAAGAAGG	94°C-30 sec
	3 reverse	TTTAAAGGATTACGTGGTAC	53°C-40 sec } 40 ×
	4 reverse	TCTTTAACCGCAAATGCAAT	72°C-40 sec 72°C-7 min
<i>YCF5</i>	1 forward	GGATTATTAGTCACTCGTTGG	94°C-4 min
	2 forward	ACTTAGAGCATATATTAACTC	94°C-30 sec
	3 reverse	ACTTACGTGCATCTTAACCA	53°C-40 sec } 40 ×
	4 reverse	CCCAATACCATCATACTTAC	72°C-40 sec 72°C-7 min
<i>psbA-trnH</i> [4]	fwd	GTTATGCATGAACGTAATGCTC	94°C-5 min
	rev	CGCGCATGGTGGATTACAATCC	94°C-1 min
			55°C-1 min } 30 ×
			72°C-1.5 min 72°C-7 min
<i>psbA-trnH</i> (including protein coding region)	fwd	CGAGCCTGTTCTGGTTCTC	98°C-3 min
	Rev (short-fragment)	GGGGTGTGGTAGAGCAGT	98°C-10 sec
			60°C [-0.5%/cycle] - 20 sec } 10 ×
	Rev (long-fragment)	CCGACGACGAACTAACATTG	72°C-1 min } 10 ×

Table 1. (Continued)

Gene	Primer	Sequence 5'-3'	Rxn conditions
			98°C-10sec }
			55°C-20 sec } 25 ×
			72°C-1 min }
			72°C-7 min
<i>nrlTS</i>	5a fwd	CCTTATCAITTAGAGGAAGGAG	94°C-5 min
	4 rev	TCCTCCGCTTATTGATATGC	94°C-1 min }
			50°C-1 min } 30 ×
			72°C-1.5 min +3 sec /cycle }
			72°C-7 min
	2c rev (sequencing only)	GCTACGTTCTCATCGTGGC	N/A

Conditions for chloroplast markers from the Plant Working Group (www.kew.org/barcoding/protocols.html); conditions for *psbA-trnH* adopted from Kress et al., 2005.

in some cycads or is present as a pseudogene. The longer fragment in *Encephalartos nubimontanus* corresponded to the protein sequence that is most similar to *psbA* protein sequence of other gymnosperms (*Ginkgo biloba* and *Pinus korianus*). These analyses were performed on whole genomic DNA, so it remains unclear whether both genes are being amplified from the chloroplast genome or if the second fragment could be nuclear DNA that was transferred from the chloroplast, a well documented phenomenon (Timmis et al., 2004). Problems with amplification aside, *psbA-trnH* does not show promise as a barcoding locus for cycads because of its inability to provide specific identification for taxa that could not be distinguished with *nrITS*.

Algorithm Comparisons: BLAST and DNA-BAR/DEGENBAR

For our data sets, there was no difference between BLAST and DNA-BAR/DEGENBAR. For optimization, BLAST offers several advantages: It generates a more detailed output and is readily available and downloadable from NCBI. For use in barcoding in practice, either method seems to be similarly successful (Little and Stevenson, 2007). Standardization of an algorithm used for database searches as part of the DNA barcoding workflow should be promoted in order to provide maximally consistent results.

RESULTS

Proposed Regions

The primer pairs chosen using *Ceratozamia hildae* and *Cycas ophiolitica* (Figure 1) for *ndhJ*, *rpoB*, and *matK* did not work well for the remaining taxa (Table 2): non-specific primer binding resulted in multiple bands or complete lack of amplification. Because the purpose of these experiments was to test the functionality and utility of the proposed barcoding conditions and primers (as per www.kew.org/barcoding) on cycads we did not develop novel cycad-specific primers or reaction conditions. Further

analyses were performed only on those primers that successfully generated single products under universal conditions: *accD*, *YCF5*, and *rpoC1*.

Table 2. Primers and reaction conditions used in this study.

Marker	Successful amplification (single bands)	Non-specific amplification (multiple bands)	No amplification	Used for identification
<i>accD</i>	26/27 = 96%	1/27 = 4%	0/27 = 0%	yes
<i>YCF5</i>	66/66 = 100%	0/66 = 0%	0/66 = 0%	yes
<i>rpoC1</i>	29/29 = 100%	0/29 = 0%	0/29 = 0%	yes
<i>ndhJ</i>	12/21 = 57% *	6/21 = 29%	3/21 = 15%	no
<i>rpoB</i>	7/21 = 33%	14/21 = 67%	0/21 = 0%	no
<i>matK</i>	5/21 = 24%	11/21 = 52%	5/21 = 24%	no

Only markers with near universal amplification success were sequenced and tested for identification. An * indicates very weak bands.

Sequences generated from these three regions were tested for their ability to provide unique species identifications using both BLAST and DNA-BAR/DEGENBAR. Neither algorithm was able to positively identify individuals to species due to a lack of unique species-specific sequence for all species tested. Both algorithms had some success with identification of individuals to genus with 63–93% of query sequences correctly identified depending on the marker used (Figure 2). Inspection of the alignment revealed that there were very few variable positions. Over the three tested DNA regions, approximately 10% of the bases were variable (93 of 917 total bases): for *accD*, 28 of 242 base pairs were variable; for *rpoC1*, 41 of 476 base pairs were variable; and for *YCF5*, 24 of 199 base pairs were variable.

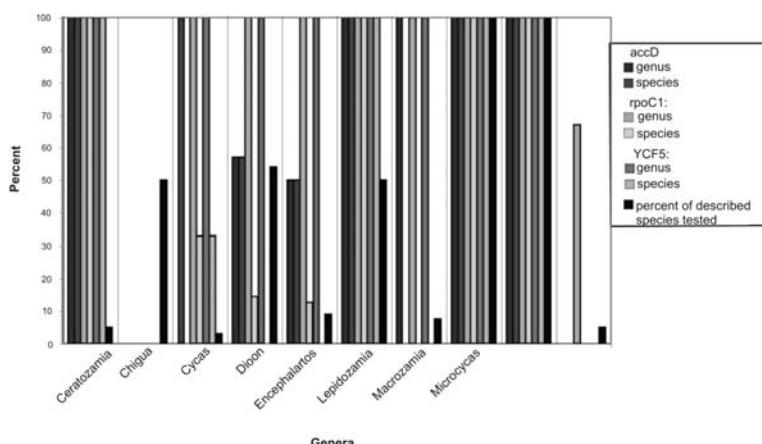


Figure 2. Success of species and genus level identification using CBOL proposed gene regions. Performance (% correct identification) at genus and species levels is noted for each marker for each of the 10 genera tested. No bar indicates failure of identification (0% success). Values are identical for BLAST and DNA-BAR/DEGENBAR.

Secondary Regions (*nrITS* and *psbA-trnH*)

Because the chloroplast gene regions initially suggested by the PWG did not promise to distinguish among species even with our rather incomplete sampling, the alternative regions suggested by Kress et al. (2005) -*psbA-trnH* intergenic spacer coupled with *nrITS*-were tested on the original 27-species set. The *nrITS* repeat (nrITS 1, 5.8S, and nrITS2) amplified cleanly in most species, but sequencing was difficult because *nrITS* in cycads (and other gymnosperms) is variable in length—approximately 1100 bp in most species, but around 1400 bp in *Stangeria eriopus*. In many species, the use of internal primers was required to generate contigs of the full sequence—making *nrITS* less desirable as a DNA barcode for cycads. A second potentially negative factor is the presence of long poly-G, poly-C, and poly-A repeats that are difficult to sequence through. Despite these issues, *nrITS* had sufficient variation to correctly identify individuals to species for the 27 individuals initially tested plus four additional species represented by sequences downloaded from GenBank (due to sequencing difficulties for *Zamia* and lack of fresh tissue samples for *Bowenia*).

Additional species were sampled from *Dioon* and *Encephalartos* to further test the functionality of *nrITS* for species-level identification. These genera were chosen because tissue samples were available that maximized the total percent coverage of species within each genus (seven out of 13 *Dioon*, 44 out of 65 *Encephalartos*). To further increase the number of species represented, available sequences from GenBank were included in the reference databases. For *nrITS*, a total of 96 sequences comprising 74 taxa were included in the ordinal-level analysis. Each genus was included in the ordinal analysis and where possible more than one species from each genus was included (Table 3). In the ordinal-level database, all species were correctly identified to genus and 90.5% of queries correctly and uniquely identified the query sequence in the reference database. The success of self-identification is broken down by genus in Table 3. Genus-specific databases were made for *Encephalartos*, *Cycas*, and *Macrozamia* because some species could not be included in the ordinal-level database as the sequences did not contain the necessary anchor regions. In the generic-level databases the percent identification decreased: For *Encephalartos*, 26 of 44 (59.1%) species identified uniquely; for *Cycas*, 11 of 12 (91.7%) species identified uniquely; and for *Macrozamia*, eight of eight (100%) species identified uniquely. The *nrITS* locus had the highest success rate of the any of the markers tested; even though not all species could be correctly identified. Because of variation in length and sequence, complete alignments were not generated and the number of variable characters was not counted.

The *psbA-trnH* spacer primers and reaction conditions specified by Kress et al. (2005) yielded distinct double bands in all but *Cycas* species (Figure 3A). Even with greatly increased annealing temperature, double bands were still present (Figure 3B). The utility of this region for barcoding was tested by sequencing the larger of the two fragments (after gel excision) from species that could not be uniquely identified in the *nrITS* database. The addition of *psbA-trnH* sequence data did not further resolve the non-specific identifications made by *nrITS* for the species tested. Of 322 total

characters, including gaps, in the cursory alignment used to ensure the presence of anchor regions, 83 characters were parsimony informative; this variability was mostly due to differences between *Cycas* and the remaining genera and is not directly translated into sequence variation that is useful for barcoding.

Table 3. Amplification success of suggested primer pairs (<http://www.kew.org/barcoding/protocols.html>) with broad sampling of cycads.

Genera	Number of species analyzed		Total number of named species [40]	Percent of species that are represented and success rate of unique identification in ordinal level and generic level analyses				
	Order level	Genus level		Ordinal level		Generic level		
				Percent represented	Success	Percent represented	Success	
<i>Cyeas</i>	11	12	99	11.1	11/11 = 100%	12.1	11/12 = 91.7%	
<i>Zamia</i>	8	—	59	13.6	8/8 = 100%	—	—	
<i>Chigua</i>	2	—	2	100	2/2 = 100%	—	—	
<i>Ceratozamia</i>	8	—	21	38.1	8/8 = 100%	—	—	
<i>Macrozamia</i>	7	8	40	17.5	7/7 = 100%	20	8/8 = 100%	
<i>Stangeria</i>	1	—	1	100	1/1 = 100%	—	—	
<i>Encephafartos</i>	25	44	65	38.5	18/25 = 72%	67.7	26/44 = 59.1 %	
<i>Lepidozamia</i>	2	—	2	100	2/2 = 100%	—	—	
<i>Microcyeas</i>	1	—	1	100	1/1 = 100%	—	—	
<i>Bowenia</i>	2	—	2	100	2/2 = 100%	—	—	
<i>Olooon</i>	7	—	13	53.8	7/7 = 100%	—	—	
TOTAL	74	94	305	24.2	67/74 = 90.1%	32.5	76/94 = 80.9%	

Success indicates results for both BLAST and DNA-BAR/DEGENBAR, which were identical.

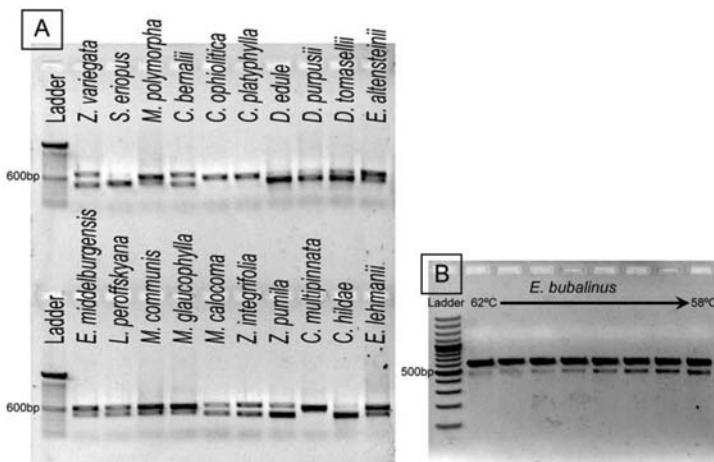


Figure 3. Amplification of PCR producing using published *psbA-trnH* primers. (A) All genera except *Cycas* showed double bands, some genera had more prominent smaller fragments (e.g., *Dioon*), while others had more prominent larger fragments (e.g., *Macrozamia*). (B) When more stringent reaction conditions were applied by running the amplification with the primer annealing temperature at 62°C, double bands were still evident.

Algorithm Differences: BLAST and DNA-BAR/DEGENBAR

As tested on our cycad database, there were no differences in the ability of BLAST and DNA-BAR/DEGENBAR to correctly identify species.

CONCLUSION

The goal of finding universal primer pairs and reaction conditions with unique internal sequence for all land plants remains elusive—not surprising given the complex history of land plant genomes. At least in cycads, the chloroplast regions tested do not have sufficient variability to provide the unique sequences (characters or combinations of characters) necessary to identify an individual to species. Nuclear regions may provide more usable variability, but such regions have not yet been identified. Perhaps a set of primers designed for each of the major clades of land plants (such as gymnosperms, pteridophytes, angiosperms, mosses, etc.) could be used simultaneously if universal tails were added to the primers so that although only one set of primers would amplify an unknown sample, the amplicon could be sequenced using a primer that matched the tail sequence. This approach would be especially useful in situations where little morphological information is available from the sample (e.g., determination of diet based on scat collections, identification of degraded, fragmented, or sterile tissue). Alternatively improved technology such as sequencing long regions of DNA (e.g., whole or partial chloroplast genomes) may enable identification based on both genome architecture and additional variation captured by simply increasing the total amount of sequence.

KEYWORDS

- **Barcode of Life**
- **Cycadales**
- **DNA barcodes**
- ***psbA-trnH***

AUTHORS' CONTRIBUTIONS

Conceived and designed the experiments: Chelsea D. Specht, Chodon Sass, and Dennis Wm. Stevenson. Performed the experiments: Chodon Sass. Analyzed the data: Chelsea D. Specht, Chodon Sass and Damon P. Little. Contributed reagents/materials/analysis tools: Chelsea D. Specht, and Damon P. Little. Wrote the chapter: Chelsea D. Specht, Chodon Sass, and Damon P. Little.

ACKNOWLEDGMENTS

The authors thank the University of California Botanical Garden at Berkeley for access to cycad specimens and especially C. Carmichael, H. Forbes, and J. Finn (UCBG) for aiding in sampling efforts. We also thank H. Driscoll for discussions, laboratory, and computational support, H. Driscoll and M. Bartlett for assistance in

collecting tissue, and T. Gregory for ongoing support of cycad research at the UCBG and UCB.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

Chapter 11

Chloroplast Genomes in Ancestral Grasses

Bojian Zhong, Takahiro Yonezawa, Yang Zhong,
and Masami Hasegawa

INTRODUCTION

It has been suggested that the chloroplast genomes of the grass family, Poaceae, have undergone an elevated evolutionary rate compared to most other angiosperms, yet the details of this phenomenon have remained obscure. To know how the rate change occurred during evolution, estimation of the time-scale with reliable calibrations is needed. The recent finding of 65 Ma grass phytoliths in Cretaceous dinosaur coprolites places the diversification of the grasses to the Cretaceous period, and provides a reliable calibration in studying the tempo and mode of grass chloroplast evolution.

By using chloroplast genome data from angiosperms and by taking account of new paleontological evidence, we now show that episodic rate acceleration both in terms of non-synonymous and synonymous substitutions occurred in the common ancestral branch of the core Poaceae (a group formed by rice, wheat, maize, and their allies) accompanied by adaptive evolution in several chloroplast proteins, while the rate reverted to the slow rate typical of most monocot species in the terminal branches.

Our finding of episodic rate acceleration in the ancestral grasses accompanied by adaptive molecular evolution has a profound bearing on the evolution of grasses, which form a highly successful group of plants. The widely used model for estimating divergence times was based on the assumption of correlated rates between ancestral and descendant lineages. However, the assumption is proved to be inadequate in approximating the episodic rate acceleration in the ancestral grasses, and the assumption of independent rates is more appropriate. This finding has implications for studies of molecular evolutionary rates and time-scale of evolution in other groups of organisms.

The grass family, Poaceae, is one of the largest plant families, comprising about 10,000 species including the most important agricultural plants, rice, wheat, and maize, and grass-dominated ecosystems comprise about one-third of the Earth's vegetative cover and support a vast number of animals (Jacobs et al., 1999). It has long been suggested that the chloroplast genomes of the grass family have undergone an elevated evolutionary rate compared to other angiosperms (Bousquet et al., 1992; Chaw et al., 2004; Gaut et al., 1992; Muse and Gaut, 1997), yet little is known when, why and how this rate change occurred.

To examine how the rate change occurred during evolution, it is prerequisite to know the time-scale of evolution. It has become increasingly feasible to estimate the phylogenetic tree and the time-scale of angiosperm evolution by using chloroplast

genome sequences (Chaw et al., 2004; Jansen et al., 2007; Leebens-Mack et al., 2005; Martin et al., 2005; Moore et al., 2007). A reliable calibration is necessary to obtain reliable time estimates, but lack of good fossil evidence of the ancestral grasses has prevented us from addressing this issue. The recent finding of 65 Ma grass phytoliths in Cretaceous dinosaur coprolites (Prasad et al., 2005; Piperno and Sues; 2005) places the diversification of the grasses to the Cretaceous period, and provides a reliable calibration in studying the tempo and mode of grass chloroplast evolution. By using this calibration, we here find that episodic rate acceleration occurred in the common ancestral branch of the core Poaceae (a clade formed by rice *Oryza*, wheat *Triticum*, maize *Zea*, and their allies) accompanied by adaptive evolution in several chloroplast proteins, while the rate reverted to the slow rate typical of most monocot species in the terminal branches. We also find that the widely used method for estimating divergence times based on the assumption of correlated rates between ancestral and descendant lineages (Kishino et al., 2001; Sanderson, 1997; Thorne et al., 1998) proved to be inadequate in approximating the process of grass chloroplast evolution, and the assumption of independent rates (Rannala and Yang, 2007) is more appropriate to studies of rate change over time. These results have implications for studies of molecular evolutionary rates and time-scale of evolution in other groups of organisms.

MATERIALS AND METHODS

Since our main interest was on grass evolution, we used all the monocot genera (13 species) and selected 18 species from outside monocots (31 species in total) among the 64 species in ref. (Jansen et al., 2007). We used 75 chloroplast genes among 77 protein-encoding genes in ref. (Jansen et al., 2007), excluding *infA* and *ycf2* because of missing data.

Estimation of Divergence Times

The concatenated 75 gene sequences of chloroplast from 31 species (from ref. (Jansen et al., 2007)) and the tree topology in ref. (Jansen et al., 2007) were used. To estimate divergence times and molecular evolutionary rates, a Bayesian method implemented in MCMCTREE (in PAML (Yang, 2007)) was applied either with the correlated-rates (CR) model (Kishino et al., 2001; Thorne et al., 1998) or with the independent-rates (IR) model (Rannala and Yang, 2007) (using the GTR model with a discrete gamma distribution with five rate categories (Γ_5) for nucleotide substitutions), and multiple calibrations were incorporated through the time prior. The gymnosperm/angiosperm divergence time was set at 280–310 Ma (Chaw et al., 2004; Moore et al., 2007). Three nodes were constrained with minimum ages as follows; (1) the minimum age of the *Zea/Oryza* divergence was set either to 65 Ma (Prasad et al., 2005; Piperno and Sues; 2005) or without this constraint, (2) >115 Ma constraint to the divergence of Poales from other monocots based on the earliest fossils of Poales (Herendeen and Crane, 1995; Linder and Rudall, 2005), and (3) >125 Ma for the most basal divergence in eudicots (Moore et al., 2007). In order to check the robustness of the time estimation on the choice of the substitution model, the codon-substitution (Goldman and Yang, 1994; Muse and Gaut, 1994)+ Γ_5 model was also used. The program adopts soft bounds, so that the probability that the true divergence time is outside the bounds is

small but not zero (Yang and Rannala, 2006). In the Bayesian framework, priors are assigned not only on times, but also on the overall substitution rate parameter μ and on the rate-drift parameter σ^2 . So we roughly estimated the prior mean of the overall rate μ using the strict molecular clock with 295 Ma constraint to gymnosperm/angiosperm divergence time, and assigned the gamma prior G(4, 80) and G(4, 22) for this prior parameter in applying the nucleotide and codon-substitution models, respectively. We next examined the impact of the rate-drift parameter σ^2 by giving various priors for σ^2 in applying the nucleotide-substitution model. Posterior distributions of parameters were approximated using two independent MCMC analyses of 107 steps each, following a discarded burn-in of 106 steps. All the analyses were repeated with different inseed values to check for convergence of the MCMC chain.

Non-Synonymous/Synonymous Rate Ratio

To the concatenated sequences of 75 protein-encoding genes of chloroplast from six Poaceae species (*Oryza*, *Triticum*, *Hordeum*, *Zea*, *Saccharum*, and *Sorghum*), *Typha*, and *Musa*, we applied the codon-based likelihood models that allow for variable ω ratios among different lineages (Yang and Bielawski, 2000). We used the likelihood ratio test (LRT) to compare the likelihood of one- ω ratio model, which assumes the same ω for all branch in the tree, with the two- ω ratio models, which assumes two different ω ratios. One of the two-ratio models (named “Simple 2 ω -model”) assumes that Poaceae (including the common ancestral branch) has different ω from other parts of the tree as is represented by

$$(Musa\#\omega_o, Typha\#\omega_o, (\text{crown Poaceae } \#\omega_i)\#\omega_i),$$

while the other (named “Reverted 2 ω -model”) assumes that only the ancestral branch of Poaceae has a different ω ratio than all the other branches in the tree as is represented by

$$(Musa\#\omega_o, Typha\#\omega_o, (\text{crown Poaceae } \#\omega_o)\#\omega_i).$$

All the analyses were carried out with the CODEML program in PAML (Yang, 2007) using the codon-substitution model with the F61 codon frequency.

Branch-Site Test of Positive Selection

The branch-site test was applied to the dataset of 11 monocot species in our dataset excluding the two most basal monocots; that is, *Dioscorea* and *Acorus*. The branch preceding the common ancestor of the core Poaceae was specified as a foreground branch, and all the others as background branches. The LRT is constructed to compare an alternative model that allows for some codons under positive selection on the foreground branch with a null model that does not. The null model restricts codons on the foreground lineage to be undergoing neutral evolution ($\omega = 1$). The specific codons which evolved under positive selection were identified on the foreground branch using a Bayes empirical Bayes procedure (Yang et al., 2005; Zhang et al., 2005).

DISCUSSION

In our study, the IR model gives more consistent results than the CR model, which has been widely used in estimating divergence times (Aris-Brosou and Yang, 2003;

Bremer, 2002; Hasegawa et al., 2003; Kishino et al., 2001; Leebens-Mack et al., 2005; Lepage et al., 2007; Nikaido et al., 2001; Sanderson, 1997; Sanderson et al., 2004; Thorne et al., 1998; Yoder and Yang, 2004). A basic assumption of the CR model is that rates change gradually over the tree. Our results suggest that the magnitude of the rate of acceleration is underestimated by the CR model and that the IR model is more appropriate in approximating the rate change in the grass chloroplast evolution. Although there exists a case in which the CR model outperforms the IR model (Lepage et al., 2007), a number of authors have recently begun to notice that the IR model is superior to the CR model in approximating the evolution of evolutionary rates in several cases (Brown et al., 2008; Drummond et al., 2006; Kitazoe et al., 2007; Renner et al., 2008).

Our analysis has revealed an episodic acceleration of the evolutionary rate of chloroplast genomes during the emergence of core Poaceae, accompanied by adaptive evolution in several protein-encoding genes. Because the elevation of the rate occurred not only in non-synonymous substitutions but also in synonymous substitutions and because the elevated substitution rates were accompanied also by an elevated rate of insertions/deletions of nucleotides (Leebens-Mack et al., 2005), the elevation of the mutation rate of chloroplast genomes might have acted as a trigger of the adaptive evolution in the ancestral grasses, which might have facilitated the successful radiation and diversification of their descendants.

Suggested positive selection of *clpP* in *Oenothera* and Sileneae accompanied by elevated synonymous rate (Erixon and Oxelman, 2008) might be related to our finding of rate acceleration in ancestral grasses both in terms of synonymous and non-synonymous substitutions. A more extensive study of chloroplast genomes showed highly accelerated non-synonymous rates of ribosomal protein and RNA polymerase genes in Geraniaceae accompanied with the elevation of the ω ratio (Guisinger et al., 2008). Interestingly, the four genes (*atpE*, *cemA*, *rpoB*, and *rpsII*) detected to have positively selected sites in our analysis (Table 6) are included in the gene group with significantly high ω ratio in Geraniaceae relative to other angiosperms (*clpP* was not analyzed in ref. (Guisinger et al., 2008)).

Recently, Smith and Donoghue (2008) tested evolutionary rates across five groups of angiosperms, and found that the rates are generally low in trees/shrubs compared to related herbs. This is an interesting finding which links life history of plants to their rates of molecular evolution, and their conclusion generally holds in five different groups of Angiosperm. What we have shown in this work, however, is that the pattern of rate change during evolution is more complicated than has previously been anticipated. Our finding highlights the need for paying attention to rates of internal branches rather than averaging along a lineage in addressing the rate heterogeneity problem.

RESULTS

Estimation of Time-Scale and Pattern of Rate Change

Figure 1 shows the ML phylogenetic tree of angiosperm chloroplast with gymnosperm as an outgroup. The elongated branches of Poaceae are in accord with their widely accepted rate acceleration (Bousquet et al., 1992; Chaw et al., 2004; Gaut et al., 1992;

Muse and Gaut, 1997). The global clock model in Poales (including Poaceae and *Typha*) +*Musa* was rejected when compared with the 2-local-clocks model (Poaceae lineages have a different rate from basal lineages such as *Typha* and *Musa*) by the LRT ($\chi^2 = 2903.89$, $P < 10^{-300}$ with the codon-substitution model). Moreover, longer distances of the Poaceae species from *Musa* than the *Typha/Musa* distance both in terms of non-synonymous and synonymous substitutions (Figure 1) indicate that both types of substitutions have undergone rate acceleration along the line leading to Poaceae.

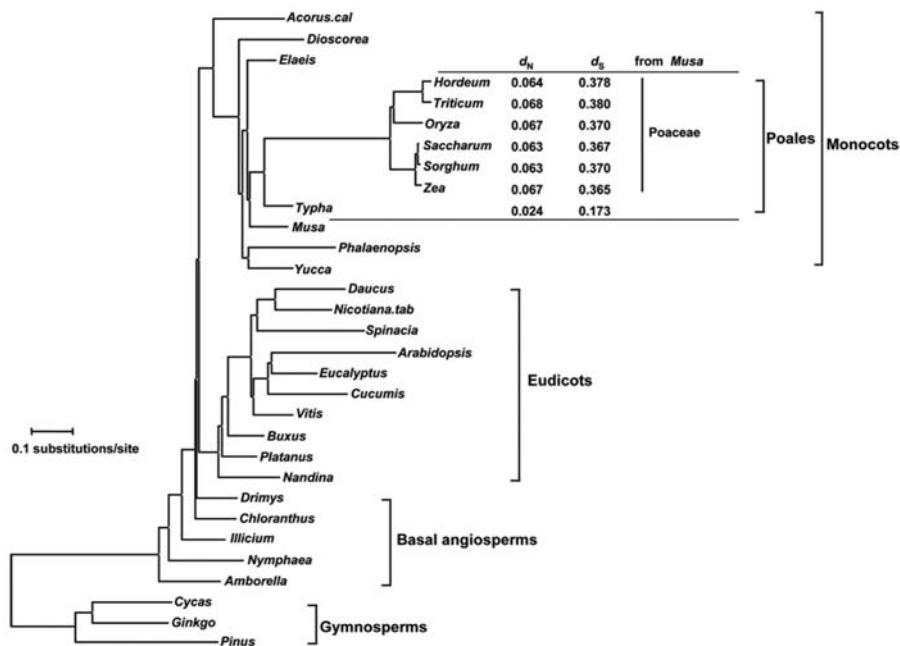


Figure 1. The phylogenetic tree of chloroplast genomes for the 31 species. The tree topology in Figure 3 of ref. (Jansen et al., 2007) was used, and the branch lengths were estimated by the ML with the codon-substitution model (Goldman and Yang, 1994; Muse and Gaut, 1994) (CODEML in PAML (Yang, 2007)). The root was arbitrarily placed between Gymnosperm and Angiosperm. Non-synonymous (d_N) and synonymous (d_S) distances of Poales from *Musa* were estimated by CODEML.

To explore the pattern of rate change during the course of grass evolution in more detail, we estimated the time-scale of Angiosperm phylogeny, particularly focusing on monocots. Although several powerful methods have been developed for molecular time estimation allowing the rate change (a relaxed clock) (Drummond et al., 2006; Huelsenbeck et al., 2000; Kishino et al., 2001; Rannala and Yang, 2007; Sanderson, 1997; Thorne et al., 1998), the poor quality of the fossil record for early grasses has prevented us from addressing this issue. Previously, the divergence among major groups of Poaceae was thought to have occurred in early Cenozoic, and the 60 Ma (Wolfe et al., 1989) and 50–60 Ma (Chaw et al., 2004) date calibrations for the maize/wheat divergence were used in estimating the monocots/eudicots divergence time with

chloroplast DNA sequences. However, recent findings of grass phytoliths in Cretaceous dinosaur coprolites (Piperno and Sues, 2005; Prasad et al., 2005) provided evidence that the major groups of core Poaceae had already diversified before Cretaceous/Tertiary (K/T) boundary of 65 Ma. Figure 2A gives time estimates of the monocots evolution (Figure 3 and Table 1 for the whole angiosperms) by a relaxed clock based on the Bayesian method implemented in MCMCTREE (in PAML (Yang, 2007)) with a constraint of >65 Ma for the *Zea/Oryza* divergence and with the IR model for the rate change along lineages (Drummond et al., 2006; Rannala and Yang, 2007). In order to illustrate the rate change during evolution, a single instance of estimated rates along the lineage from the root to *Oryza* is also shown in Figure 2A, where elevation of the rate only occurred on the common ancestral branch of Poaceae after they diverged from *Typha*.

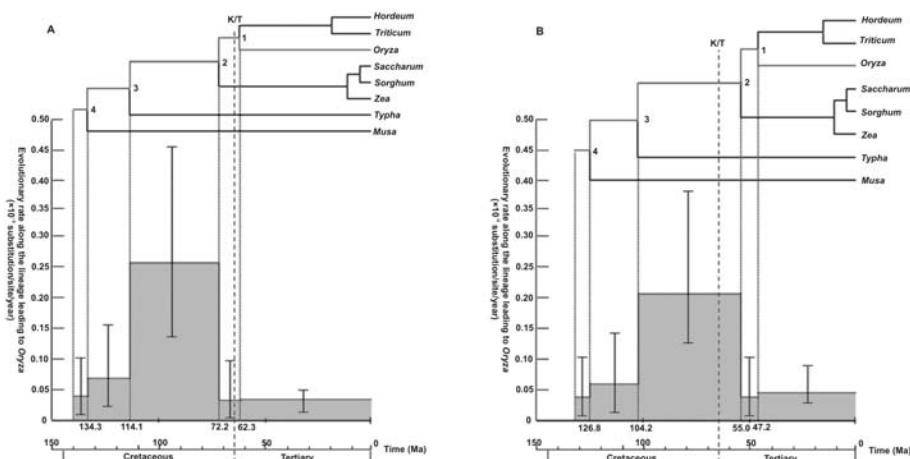


Figure 2. Posterior estimates of divergence times and rate change during evolution. Estimations were done by using MCMCTREE in PAML (Yang, 2007) with the IR model (Rannala and Yang, 2007) for the rate change along lineages. Shape and scale parameters, α and β , in the gamma prior for parameter σ^2 were 1.0 and 10.0, respectively. Only Poales + *Musa* part of the whole tree is shown, and a numbering of a node follows that of the whole tree in Figure 3. The upper lines of the colored area trace the estimated rates along the lineage from the root to *Oryza* (the lineage indicated by colored lines in the tree) where 95% highest posterior density (HPD) is shown by a vertical line segment with two short horizontal line segments at boundaries. (A) >65 Ma constraint and (B) no constraint to the *Zea/Oryza* separation (for other calibrations, see Materials and Methods).

Although the fossil evidence for the >65 Ma constraint of the *Zea/Oryza* divergence is important in demonstrating the rate acceleration in ancestral grasses with subsequent slow-down, it is not a prerequisite. Even when the constraint was removed, almost the same pattern of rate change as that with the >65 Ma constraint was obtained when the IR model was used (Figure 2B and Table 2), although the time estimate of the *Zea/Oryza* separation became younger (55.0 Ma). This time estimate is consistent with a conservative date of >50 Ma presented in refs. (Chaw et al., 2004; Vicentini et al., 2008), and our conclusion of the reverted slow rate in contemporary Poaceae can be regarded as robust to the calibration points used.

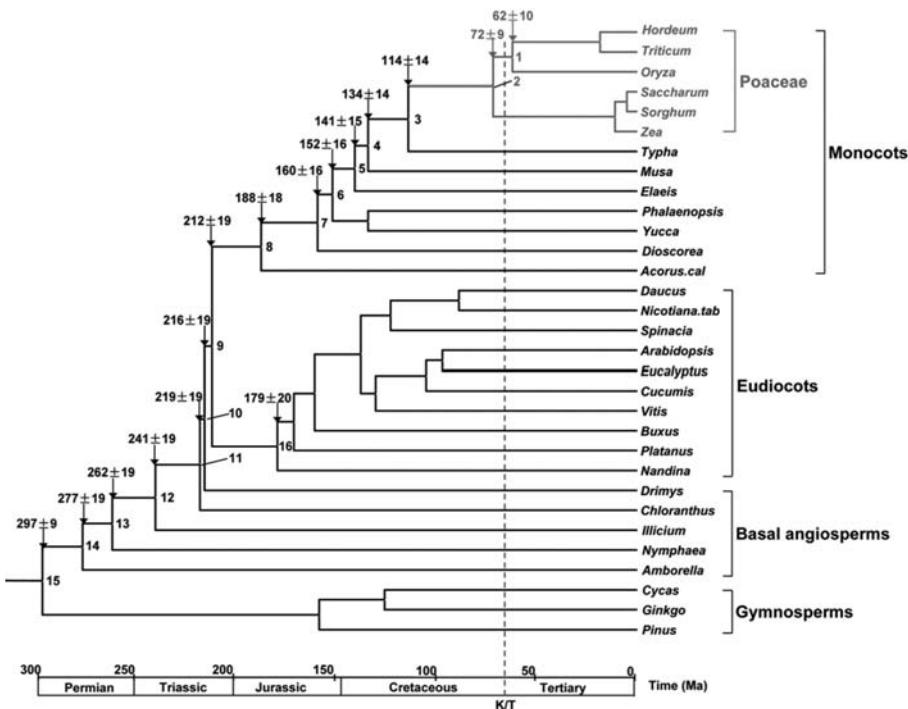


Figure 3. Posterior estimates of divergence times of a whole Angiosperm tree. Estimations were done by using MCMCTREE (Yang, 2007) with the IR model (Rannala and Yang, 2007). The >65 Ma constraint to the *Zea/Oryza* separation was applied.

In the first model of the relaxed clock implemented by Thorne and colleagues (Kishino et al., 2001; Thorne et al., 1998), rates are auto-correlated between ancestral and descendant lineages on the tree, and the model is called the CR model. Sander-son's method of non-parametric rate-smoothing (Sanderson, 1997) was also based on the same idea. Later, an alternative model named the IR model with no auto-correlation was developed (Drummond et al., 2006; Rannala and Yang, 2007). In Table 1, estimates of divergence times with the >65 Ma constraint for the *Zea/Oryza* separation are compared between the IR and CR models. The CR model tends to give older estimates for the nodes preceding the *Zea/Oryza* separation than the IR model. For example, the monocots/eudicots divergence time estimate was 239.1 Ma with the CR model, while the estimate with the IR model was 212.5 Ma which is more in accord with the recently published estimate of 140–150 Ma (Chaw et al., 2004) even though it is still older. Without the >65 Ma constraint, the time estimate of the *Zea/Oryza* separation became too young (36.9 Ma) from the CR model to be compatible with the suggestion of >50 Ma from the previous works (Chaw et al., 2004; Vicentini et al., 2008), while the IR model gave compatible estimate of 55.0 Ma as mentioned before.

Table 1. Posterior estimates of divergence times with the >65 Ma constraint to the *Zea/Oryza* separation

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
1 (<i>Triticum</i>)	47.2 (28.4, 71.5)	0.040 (0.012, 0.105)	32.4 (21.6, 49.6)	0.059 (0.028, 0.103)
2 (<i>Zea</i>)	55.0 (35.2, 80.4)	0.212 (0.124, 0.377)	36.9 (24.6, 56.7)	0.119 (0.087, 0.159)
3 (<i>Typha</i>)	104.2 (78.7, 135.5)	0.063 (0.026, 0.144)	119.8 (90.9, 156.5)	0.059 (0.034, 0.096)
4 (<i>Musa</i>)	126.8 (105.7, 158.6)	0.040 (0.012, 0.105)	140.8 (113.0, 176.2)	0.043 (0.025, 0.069)
5 (<i>flaeis</i>)	133.6 (111.1, 166.1)	0.044 (0.014, 0.112)	145.8 (117.4, 181.1)	0.057 (0.033, 0.091)
6 (<i>Phalaenopsis/Yucca</i>)	145.3 (120.2, 179.3)	0.041 (0.012, 0.108)	153.2 (124.5, 187.7)	0.043 (0.026, 0.066)
7 (<i>Dioscorea</i>)	152.6 (125.8, 187.8)	0.080 (0.034, 0.179)	158.7 (129.4, 193.2)	0.059 (0.035, 0.093)
8 (<i>Acorus</i>)	181.7 (147.1, 220.3)	0.054 (0.022, 0.128)	193.7 (159.4, 226.9)	0.046 (0.027, 0.073)
9 (<i>Eudicots</i>)	207.7 (169.2, 245.1)	0.036 (0.010, 0.098)	220.4 (184.1, 251.4)	0.047 (0.029, 0.072)
10 (<i>Drimys</i>)	211.3 (172.3, 249.1)	0.034 (0.009, 0.094)	222.6 (186.2, 253.2)	0.044 (0.026, 0.070)
11 (<i>Chloranthus</i>)	213.6 (174.3, 251.5)	0.060 (0.023, 0.141)	223.9 (187.4, 254.5)	0.047 (0.029, 0.072)
12 (<i>Illicium</i>)	236.2 (193.9, 272.9)	0.053 (0.020, 0.128)	248.6 (210.3, 277.3)	0.048 (0.027, 0.079)
13 (<i>Nymphaea</i>)	258.2 (213.6, 291.8)	0.042 (0.014, 0.107)	269.2 (229.4, 296.1)	0.048 (0.027, 0.078)
14 (<i>Amborella</i>)	272.8 (227.0, 304.2)	0.056 (0.012, 0.177)	279.7 (239.6, 305.8)	0.053 (0.031, 0.086)
15 (<i>Gymnosperm</i>)	296.6 (280.3, 310.2)	–	297.2 (280.4, 310.3)	–
16	176.9 (138.5, 217.4)	0.061 (0.026, 0.139)	178.5 (141.1, 215.1)	0.040 (0.024, 0.068)
Terminal branch to <i>Oryza</i>	0	0.051 (0.032, 0.080)	0	0.074 (0.046, 0.106)

MCMCTREE in PAMI. [19] was used with the GTR+ Γ_3 model. Shape and scale parameters, α and β , in the gamma prior for parameter σ^2 were 1.0 and 10.0, respectively.

95% HPD is shown in parentheses. Rate ($\times 10^8$ substitutions/nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

In order to examine the impact of including rapidly evolving Poaceae in the analysis, a comparison between the two models was carried out excluding Poaceae (Table 3). The time estimates were similar between the two models, and were similar to those from the IR model including Poaceae. For the monocots/eudicots separation, the IR model gave almost consistent estimates of 212.5, 207.7, and 216.5 Ma, respectively, with the >65 Ma constraint for the *Zea/Oryza* separation, without the constraint, and excluding Poaceae, while the CR model gave more diverged estimates of 239.1, 220.4,

and 223.2 Ma, respectively. Interestingly, the estimates for this separation were very close between the two models when Poaceae species were excluded. This suggests that the episodic rate acceleration in ancestral Poaceae causes biased estimates, which the CR model cannot accommodate.

Table 2. Posterior estimates of divergence times without constraint to the *Zea/Oryza* separation

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
1 (<i>Triticum</i>)	62.3 (44.4, 84.4)	0.033 (0.009, 0.090)	58.4 (45.5, 76.2)	0.030 (0.014, 0.058)
2 (<i>Zea</i>)	72.2 (59.1, 93.7)	0.253 (0.137, 0.4601)	67.7 (54.3, 86.9)	0.119 (0.087, 0.1671)
3 (<i>Typha</i>)	114.1 (91.4, 144.5)	0.069 (0.029, 0.157)	150.7 (123.2, 179.5)	0.062 (0.033, 0.104)
4 (<i>Musa</i>)	134.3 (111.9, 165.8)	0.040 (0.012, 0.106)	171.0 (145.1, 197.7)	0.048 (0.027, 0.079)
5 (<i>Elaeis</i>)	141.0 (117.0, 173.4)	0.046 (0.015, 0.116)	175.6 (149.3, 202.1)	0.065 (0.038, 0.104)
6 (<i>Phalaenopsis/Yucca</i>)	152.2 (125.7, 185.4)	0.041 (0.012, 0.109)	182.0 (156.0, 208.0)	0.047 (0.028, 0.076)
7 (<i>Dioscorea</i>)	159.6 (131.8, 193.6)	0.083 (0.035, 0.185)	187.0 (160.7, 212.9)	0.070 (0.043, 0.107)
8 (<i>Acorus</i>)	187.8 (153.1, 224.4)	0.057 (0.023, 0.136)	216.3 (188.8, 241.5)	0.053 (0.032, 0.085)
9 (<i>Eudicots</i>)	212.5 (175.0, 247.5)	0.035 (0.009, 0.096)	239.1 (211.5, 262.3)	0.055 (0.033, 0.086)
10 (<i>Drimys</i>)	216.2 (178.0, 251.4)	0.033 (0.008, 0.091)	240.9 (213.4, 263.8)	0.050 (0.029, 0.079)
11 (<i>Chloranthus</i>)	218.6 (180.1, 254.0)	0.061 (0.024, 0.145)	242.1 (214.6, 264.9)	0.055 (0.034, 0.086)
12 (<i>Illicium</i>)	240.8 (199.7, 274.7)	0.053 (0.020, 0.125)	263.0 (235.3, 283.9)	0.056 (0.032, 0.091)
13 (<i>Nymphaea</i>)	262.4 (219.8, 293.3)	0.043 (0.014, 0.108)	280.8 (253.3, 299.5)	0.057 (0.032, 0.092)
14 (<i>Amborella</i>)	276.9 (233.3, 305.1)	0.054 (0.011, 0.174)	289.8 (262.2, 307.8)	0.059 (0.035, 0.096)
15 (<i>Gymnosperm</i>)	297.2 (280.5, 310.3)	–	299.5 (281.4, 310.6)	–
16	179.4 (142.1, 218.8)	0.057 (0.024, 0.135)	196.7 (166.5, 230.0)	0.039 (0.025, 0.071)
Terminal branch to <i>Oryza</i>	0	0.038 (0.027, 0.052)	0	0.040 (0.030, 0.051)

MCMCTREE in PAMI. [19] was used with the GTR+ Γ_5 model. Shape and scale parameters, α and β , in the gamma prior for parameter σ^2 were 1.0 and 10.0, respectively.

95% highest posterior density (HPD) is shown in parentheses. Rate ($\times 10^{-8}$ substitutions/nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

Table 3. Posterior estimates of divergence times excluding Poaceae

Node	Independent-rates (IR) model		Correlated-rates (CR) model	
	Time (Ma)	Rate	Time (Ma)	Rate
4 (<i>Musa/Typha</i>)	115.5 (96.3, 136.7)	0.033 (0.011, 0.083)	110.6 (91.6, 128.6)	0.031 (0.016, 0.052)
5 (<i>Elaeis</i>)	125.1 (103.9, 149.2)	0.033 (0.012, 0.081)	119.2 (99.0, 138.7)	0.034 (0.022, 0.053)
6 (<i>Yucca/Phalaenopsis</i>)	139.5 (115.5, 167.3)	0.034 (0.010, 0.086)	130.8 (109.9, 150.3)	0.034 (0.021, 0.052)
7 (<i>Dioscorea</i>)	147.7 (122.0, 177.2)	0.060 (0.029, 0.128)	137.3 (115.4, 158.1)	0.037 (0.026, 0.056)
8 (<i>Acorus</i>)	184.6 (151.4, 219.8)	0.041 (0.019, 0.090)	190.8 (162.0, 218.1)	0.037 (0.022, 0.060)
9 (<i>Eudicots</i>)	216.5 (181.5, 248.9)	0.033 (0.010, 0.085)	223.2 (194.0, 246.9)	0.041 (0.027, 0.060)
10 (<i>Drimys</i>)	220.1 (184.7, 252.6)	0.031 (0.009, 0.081)	225.5 (196.2, 248.9)	0.041 (0.026, 0.062)
11 (<i>Chloranthus</i>)	222.2 (186.6, 254.9)	0.057 (0.024, 0.128)	226.7 (197.4, 250.2)	0.041 (0.028, 0.060)
12 (<i>Illicium</i>)	244.7 (206.9, 275.8)	0.050 (0.020, 0.114)	253.8 (222.5, 276.8)	0.045 (0.028, 0.070)
13 (<i>Nymphaea</i>)	266.3 (227.3, 294.5)	0.040 (0.014, 0.096)	274.8 (241.9, 297.0)	0.044 (0.028, 0.068)
14 (<i>Amborella</i>)	280.2 (240.4, 305.9)	0.047 (0.011, 0.139)	285.2 (251.9, 307.0)	0.049 (0.031, 0.076)
15 (<i>Gymnosperms</i>)	298.3 (280.8, 310.4)	–	300.3 (282.0, 310.7)	–
16	190.4 (157.5, 223.6)	0.067 (0.031, 0.141)	188.6 (161.2, 213.2)	0.045 (0.030, 0.066)

MCMCTREE in PAMI. [19] was used with the GTR+ Γ_5 model. Shape and scale parameters, α and β , in the gamma prior for parameter σ^2 were 1.0 and 10.0, respectively.

95% HPD is shown in parentheses. Rate ($\times 10^8$ substitutions/nucleotide/year) refers to the rate of the branch preceding the node. Node numbers refer to those in Fig. 3, and taxa in parentheses refer to those branched off from the lineage leading to *Oryza*.

In the above mentioned analyses, before fixing the shape and scale parameters (α and β) in the gamma prior for parameter σ^2 , which specifies how variable the rates are across branches, impact of priors on these parameters to posterior time and rate estimates was examined in detail. Posterior time estimate for the *Zea/Oryza* separation depended less on the choice of the gamma prior for parameter σ^2 with the IR model than with the CR model, and therefore α and β in the gamma prior for parameter σ^2 were arbitrarily chosen to be 1.0 and 10.0, respectively, in the analyses of Tables 1–3 and Figures 2 and 3.

In order to further check the robustness of the time estimation on the choice of the substitution model, additional analyses based on a more realistic model of codon-substitution (Goldman and Yang, 1994; Muse and Gaut, 1994) were carried out.

Adaptive Evolution

Non-synonymous/synonymous rate ratio ($\omega = dN/dS$) is widely used as an indicator of adaptive evolution or positive selection (Yang, 2006). Table 4 compares ω ratios along the branches estimated by different models. The minimum AIC (Akaike, 1973) model shows that a pronounced increase of ω ratio occurred in the common ancestral lineage of Poaceae after they diverged from *Typha*, followed by reversion in the terminal branches to the lower level typical of basal lineages. The elevation of the ω ratio can be due either by adaptive evolution or by relaxation of selective constraints. A higher ω value than 1 is usually regarded as an evidence of adaptive evolution, but since the analysis shown in the table averages over the entire genomes, we would not get such a high value even if positive selection operated in some parts of some proteins. Therefore, the branch-site model (Yang et al., 2005; Zhang et al., 2005) was applied.

Table 4. Estimation of non-synonymous/synonymous rate ratio (ω) under different models by using CODEML in PAML (Yang, 2007).

Model	ω_0	ω_1	ω_2	Ln L	LRT with 1ω -model	AIC
1ω	0.1518	—	—	-115,741.5	—	231,485.1
Simple 2ω	0.1265	0.1617	—	-15,729.5	9.63×10^{-7}	231,463.0
3ω	0.1255	0.2189	0.1246	-15,669.4	4.95×10^{-32}	231,344.8
Reverted 2ω	0.1250	0.2189	—	-15,669.4	3.29×10^{-33}	231,342.8

1ω -model: (Musa# ω_0 , Typha# ω_0 , (crown Poaceae # ω_0) # ω_0).

Simple 2ω -model: (Musa# ω_0 , Typha# ω_0 , (crown Poaceae # ω_1) # ω_1).

3ω -model: (Musa# ω_0 , Typha# ω_0 , (crown Poaceae # ω_2) # ω_1).

Reverted 2ω -model: (Musa# ω_0 , Typha# ω_0 , (crown Poaceae # ω_0) # ω_1).

“Crown Poaceae” includes all Poaceae branches in our tree except for the common ancestral branch (stem Poaceae). The codon-substitution model with the F61 codon frequency was used. Minimum AIC (Reverted 2ω -model) is shown in bold italic.

Table 5. LRT of 1ω -model vs. reverted 2ω -model for individual genes

Gene	Ancestral Poaceae ω_1	Other branches ω_0	In L (reverted 2ω)	InL (1ω)	$X^2 = 2\Delta\text{ln L}$	P
<i>rpoB</i> *	0.3469	0.1013	-8546.12	-8574.82	57.41	3.53E-14
<i>rps11</i> *	0.7406	0.0538	-1126.87	-1144.90	36.06	1.91E-09
<i>clpP</i> *	0.5943	0.1374	-1767.99	-1778.73	21.47	3.95E-06
<i>atpE</i> *	0.9642	0.1374	-1132.80	-1143.33	21.05	4.48E-06
<i>rps3</i> *	0.5767	0.1273	-1958.45	-1968.07	19.24	1.15E-05
<i>cemA</i> *	1.4337	0.3047	-2214.75	-2224.29	19.09	1.25E-05
<i>rp122</i> *	0.6319	0.1139	-1067.47	-1075.53	16.11	5.97E-05
<i>rpoC1</i>	0.2385	0.1294	-5702.89	-5708.31	10.85	0.000989
<i>atpA</i>	0.2296	0.1102	-4201.36	-4206.20	9.68	0.001860
<i>P513</i> #	0.0540	0.1663	-3542.69	-3547.23	9.08	0.002578
<i>rps2</i>	0.3623	0.1288	-1886.14	-1890.56	8.83	0.002960
<i>rpoC2</i>	0.3364	0.2179	-10553.23	-10557.46	8.45	0.003642
<i>psaC</i> #	0.0001	0.0780	-603.22	-607.35	8.26	0.004047
<i>rbcL</i> #	0.0344	0.0937	-3805.48	-3809.48	8.00	0.004669
<i>ndhH</i>	0.1405	0.0592	-3144.67	-3148.58	7.81	0.005202
<i>rbcL</i> #	1.2673	0.2151	-865.87	-869.76	7.78	0.005297
<i>rps12</i>	0.6567	0.1220	-697.24	-700.69	6.91	0.008589
<i>rps19</i>	0.6328	0.1836	-727.58	-730.62	6.08	0.013646
<i>rp116</i> #	0.0268	0.1060	-1131.04	-1134.08	6.07	0.013739
<i>rpl14</i>	0.2858	0.0995	-972.76	-975.00	4.47	0.034457

ω and ω_0 refer to the estimates of non-synonymous/synonymous rate ratios based on the reverted 2ω -model. Only 20 genes rejecting the 1ω -model with $P < 0.05$ are listed. In this analysis, 11 monocot species (6 Poacea species, *Typha*, *Musa*, *Elaeis*, *Phalaenopsis*, and *Yucca*) were used. 75 genes were analyzed, but genes with related

functions were concatenated if the lengths are shorter than 180 nucleotides, and therefore the number of tests was 61.

PS13 refers to the concatenated sequences of petG+petL+petN+psaI+psaJ+psbF+psbL+psbJ+psbK+psbL+psbM+psbN+psbT.

“refers to a gene which remains significant after the Bonferroni correction.

#refers to a gene with $\omega_1 < \omega_0$.

To identify positively selected sites, among 61 protein-encoding “genes,” we at first selected 16 “genes,” for which the reverted 2ω -model (with the rate ratio ω_1 of the common ancestral branch of Poaceae larger than the rate ratio ω_0 of other branches) is significantly better than the 1ω -model ($P<0.05$) (Table 5), and by using the branch-site model (Yang et al., 2005; Zhang et al., 2005), we identified five genes (*atpE*, *cemA*, *clpP*, *rpoB*, and *rps11*) which have P value of the branch-site LRT less than 0.05 and contain positively selected sites (Table 6).

Table 6. Branch-site test of positive selection.

Gene	LRT	Positively selected sites
<i>atpE</i>	0.0484	<i>2T</i> →K, <i>17S</i> →C, <i>41A</i> →N, <i>64M</i> →W, <i>132V</i> →W
<i>cemA</i>	0.0021	<i>55N</i> →R, <i>76Y</i> →K, <i>161W</i> →F, <i>1901</i> →F, <i>2041</i> →C
<i>clpP</i>	0.0081	<i>26R</i> →V, <i>48V</i> →T, <i>86F</i> →T, <i>1121</i> →P, <i>134E</i> →R, <i>182T</i> →D
<i>rpoB</i>	0.0352	<i>90R</i> →F, <i>338G</i> →K, <i>1026G</i> →N
<i>rps11</i>	0.0082	<i>54V</i> →P, <i>62A</i> →S, <i>82A</i> →R, <i>105L</i> →S, <i>115R</i> →A, <i>120L</i> →R

The numberings of amino acids are those of *Zea mays* [45]. Positively selected sites were inferred at $P_b = 95\%$ with those reaching 99% shown in bold italic.

The analyses were carried out for the 16 genes selected in Table 5, and only genes with positively selected sites and with $P<0.05$ (LRD are listed.

Among the 16 genes with significantly higher ω_1 than ω_0 in Table 5 and among the five genes with positively selected sites in Table 6, only *atpE* is among the 14 genes with significant heterogeneity of nucleotide substitution rates for maize versus rice, maize versus wheat, or rice versus wheat comparisons listed in Table 5 of ref. (Matsuoka et al., 2002). Four “genes,” *psaC*, *rbcL*, *rpl6*, and *PS13*, have significantly lower ω_1 than ω_0 (stronger purifying selection in the ancestral branch of Poaceae than in other branches). In Table 5, we carried out multiple tests for 61 “genes.” The Bonferroni correction is a safeguard against multiple tests falsely giving the appearance of significance, since one out of every 20 hypotheses tests is expected to be significant at the 5% level purely by chance. After performing the Bonferroni correction, seven genes with * in Table 5 remained significant, that means, all the genes listed in Table 6 remained significant even by the conservative test of Bonferroni. On the other hand, the four “genes” with lower ω_1 than ω_0 in Table 5 were not significant after the Bonferroni correction.

KEYWORDS

- **Poaceae**
- **Phytoliths**
- **Musa**
- **Typha**
- **Zea/Oryza separation**

AUTHORS' CONTRIBUTIONS

Conceived and designed the experiments: Bojian Zhong and Masami Hasegawa. Analyzed the data: Bojian Zhong and Takahiro Yonezawa. Wrote the chapter: Bojian Zhong, Takahiro Yonezawa, Yang Zhong, and Masami Hasegawa.

ACKNOWLEDGMENTS

We thank James Crabbe for improving the manuscript. Simon Joly provided thoughtful comments that helped improve the manuscript.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

Chapter 12

EST Analysis in *Ginkgo biloba*

Eric D. Brenner, Manpreet S. Katari, Dennis W. Stevenson,
Stephen A. Rudd, Andrew W. Douglas, Walter N. Moss,
Richard W. Twigg, Suzan J. Runko, Giulia M. Stellari,
W.R. McCombie, and Gloria M. Coruzzi

INTRODUCTION

Ginkgo biloba L. is the only surviving member of one of the oldest living seed plant groups with medicinal, spiritual, and horticultural importance worldwide. As an evolutionary relic, it displays many characters found in the early, extinct seed plants, and extant cycads. To establish a molecular base to understand the evolution of seeds and pollen, we created a cDNA library and expressed sequence tag (EST) dataset from the reproductive structures of male (microsporangiate), female (megasporangiate), and vegetative organs (leaves) of *Ginkgo biloba*.

The RNA from newly emerged male and female reproductive organs and immature leaves was used to create three distinct cDNA libraries from which 6,434 ESTs were generated. These 6,434 ESTs from *Ginkgo biloba* were clustered into 3,830 unigenes. A comparison of our *Ginkgo* unigene set against the fully annotated genomes of rice and *Arabidopsis*, and all available ESTs in Genbank revealed that 256 *Ginkgo* unigenes match only genes among the gymnosperms and non-seed plants—many with multiple matches to genes in non-angiosperm plants. Conversely, another group of unigenes in *Ginkgo* had highly significant homology to transcription factors in angiosperms involved in development, including *MADS box* genes as well as post-transcriptional regulators. Several of the conserved developmental genes found in *Ginkgo* had top BLAST homology to cycad genes. We also note here the presence of ESTs in *G. biloba* similar to genes that to date have only been found in gymnosperms and an additional 22 *Ginkgo* genes common only to genes from cycads.

Our analysis of an EST dataset from *G. biloba* revealed genes potentially unique to gymnosperms. Many of these genes showed homology to fully sequenced clones from our cycad EST dataset found in common only with gymnosperms. Other *Ginkgo* ESTs are similar to developmental regulators in higher plants. This work sets the stage for future studies on *Ginkgo* to better understand seed and pollen evolution, and to resolve the ambiguous phylogenetic relationship of *G. biloba* among the gymnosperms.

Ginkgo biloba is a widely popular tree that is native to China and has been cultivated for well over a millennium. In Asia, *G. biloba* is used medicinally and its seeds are also a popular cuisine item. In the West, *Ginkgo* leaf extracts are commonly used for a variety of folk remedies (for review see: (Hori et al., 1997)) including as a treatment for improving cognitive function (Curtis-Prior et al., 1999; Gold et al., 2003).

Today's *Ginkgo biloba* is the sole surviving species of an ancient group (Ginkgophytes) of seed plants that may even date from the Permian (approximately 150–200 million years ago) (Rothwell et al., 1997). The genus *Ginkgo* itself goes back to the Jurassic period—approximately 170 million years ago (Zhou and Zheng, 2003). Although it is widely believed that the survival of *G. biloba* depended upon Buddhist monks, who venerated the tree cultivated in their temple grounds, molecular evidence suggests that some stands in China (Wuchuan, Guizhou) are of natural origin representing vestige populations (Fagard et al., 2000). As a living fossil, *Ginkgo biloba* has changed little in morphology from its extinct relatives (Zhou and Zheng, 2003). Along with the Cycadales, Coniferales, and Gnetales, the Ginkgoales is one of four orders of non-flowering seed plants (gymnosperms) that form a sister group to the angiosperms (Figure 1).

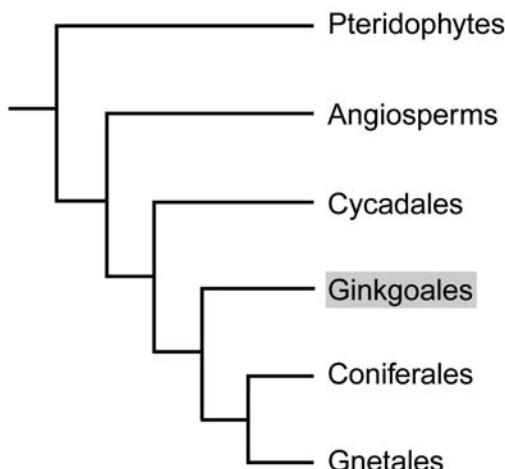


Figure 1. Gene tree of extant gymnosperms. *Ginkgo* displays characters suggesting it forms a basal subgroup among the gymnosperms with cycads. Alternately *Ginkgo* is a sister group of the conifers. Here the placement of *Ginkgo* is shown as ambivalent between these two scenarios.

Morphological (Donoghue and Doyle, 2000; Zhou and Zheng, 2003) and molecular analysis have not yet succeeded in defining the precise phylogenetic hierarchy of the four gymnosperm clades (Magollón and Sanderson, 2002). *Ginkgo* potentially forms a sister group with the Coniferales (partly due to similar characteristics such as axillary branching and simple leaves). Another model, based on molecular sequence data, places *Ginkgo* with the Cycadales (Bowe and Coat, 2000; Chaw et al., 2000; Hasebe, 1997). Interestingly cycads and *Ginkgo* both share certain plesiomorphic (ancestral) characters found in early fossil seed plants such as haustorial pollen (Friedman, 1987; Hasebe, 1935), which release motile male gametes (Ikeno and Hirase, 1897) as well as a large four celled opening in the neck of the archegonia (Foster and Gifford, 1974; Hasebe, 1935). Despite the presence of these and other early seed-plant characteristics, surprisingly little work has been performed on *Ginkgo* and cycads. Some recent molecular (Zhang et al., 2004) and genomic (Brenner et al., 2003) research on cycads

have been conducted and molecular studies of *Ginkgo* genes have been initiated as well (Chinn and Silverthorne, 1993; Chinn et al., 1995; Jager et al., 2003). However, no genomic work on *Ginkgo biloba* has been completed to date.

To begin our genomic treatment of *Ginkgo biloba*, we focused our initial efforts on developing reproductive and vegetative tissues (Figure 2). Separating *Ginkgo* male and female structures at an early stage is straightforward because *Ginkgo* is strictly a dioecious plant (male and female organs on separate individuals). Organ emergence can generally be pinpointed to a specific time of the year in that both reproductive and vegetative tissues are regularly produced in the beginning of May at our collection site in New York. The reproductive structures, megasporangia bearing ovules (Figure 2A–C) (from female trees) and microsporangia bearing pollen (Figure 2D–F) (from male trees), emerge at the apex of short, determinate (spur) shoots. A discreet flush of leaves are also produced in male and female short-shoots (Figure 2A and 2D) (Bierhorst, 1971; Foster and Gifford, 1974; Hasebe, 1935). Long-shoots (not shown) exhibit indeterminate growth and yield only vegetative organs. Long-shoots are identifiable by their obvious longer internodes, whereas short-shoots (Figure 2A and 2D) have telescoped internodes. Each season, short-shoots might exhibit extensive internode growth and be transformed into long-shoots and *vice versa*. Consequently, reproductive shoots can become vegetative or vegetative shoots can become reproductive.

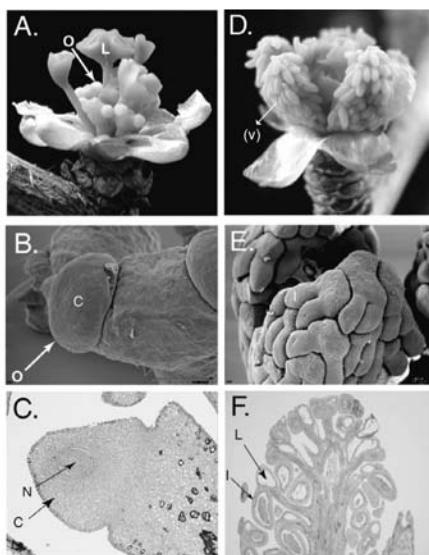


Figure 2. *Ginkgo* male and female short-shoots. (A) The fertile female structure (megasporangia) has just emerged from the bud. Two ovules set on a v stalk are visible. Young, unfurled leaves, which have also emerged have extended above the megasporangia. The bracts, which originally enclosed the bud, are now completely opened below the leaves and megasporangia. (B) The SEM of an ovule, which is completely enclosed by an integument. (C) A longitudinal cross section of the megasporangia reveals the integument enclosing the nucellus. (D) The male reproductive structure is a cluster of microsporangia. In the center of the bud are partly emerged leaves (E) The SEM shows two microsporangial lobes containing ripening pollen sacks attached to a stalk. (F) Longitudinal cross section shows a large mucilage containing cavity juxtaposed from the microsporangia filled with immature pollen. C, integument; N, nucellus; I, microsporangia; L, mucilage cavity. O, ovule.

Until now, little is known regarding the genetic regulation of development in the oldest living seed plants. In order to uncover the genetic controls directing growth and development in *Ginkgo biloba*, we generated ESTs from cDNA libraries of very young, recently emerged organs of fertile short-shoots where a large number of regulatory genes are expected to be present. Below is an analysis of these ESTs from *Ginkgo biloba*. In all three tissues examined, vegetative, microsporangia (male), and megasporangia (female), we found a large number of ESTs with similarity to angiosperm developmental genes. Conversely, a certain number of *Ginkgo biloba* ESTs were uncovered with homology to genes only found in gymnosperms and non-seed plants, including a set of *Ginkgo* ESTs that were only common to our cycad EST dataset, further strengthening their classification as gymnosperm specific.

MATERIALS AND METHODS

Tissue Collection and Library Construction and DNA Purification

Newly emerged microsporangia from accession 76163B, megasporangia from accession 76163D, and immature leaves from both accessions were collected from newly opened buds of *Ginkgo biloba* growing in the New York Botanical Garden outdoor collection on April 12, 2002. Organs were snap frozen in liquid nitrogen. RNA was collected from each organ and a cDNA library was constructed from fractionated cDNA according to (Brenner et al., 2003).

Microscopy

Ginkgo apices were collected on April 19. Bract tissues were removed from the apex leaving the leaves and reproductive tissue, which was fixed in FAA (50% ethanol, 5% glacial acetic Acid, 3.7% formaldehyde) under vacuum (20 In. Hg) at room temperature. Fresh FAA was vacuum infiltrated two additional times. Tissue was stored in 70% ethanol at 4°C.

For histology, tissue was prepared by sequential (overnight 4°C incubation at each alcohol grade) dehydration in 80, 90, 95, and finally 100% ethanol plus Eosin Y (National Medicinal Products) followed by two treatments in 100% ethanol for 2 hr at room temperature. The tissue was next placed in a 1:1 solution of ethanol and toluene, then twice in toluene alone, each time for 2 hr at room temperature. The tissue was then placed in toluene with a quarter-volume of paraffin (PARAPLAST X-TRA® (Fisher)) chips at 60°C overnight. The tissue was then embedded in melted paraffin with six wax changes over the course of three days at 55°C. Apices were sectioned on a MICROM HM 355 microtome. Eight µm thick sections were taken using a blade angle of 9°. The tissue was stained with Astra Blue and Safranin. After mounting on slides, sections were imaged using a Nikon DXM1200F digital microscope camera.

For SEM, fixed materials were dissected, dehydrated in ethanol and critical point dried in a Denton critical point dryer. Dried materials were affixed to aluminum EM stubs and coated with between 80–240 Å of palladium in a Hummer II Sputter Coater. Coated materials were then observed using a Jeol SEM 15 or 20 kV. Images were digitally recorded and evaluated using Adobe Photoshop 9.0.

EST Sequencing and Gene Analysis

Plasmid DNA was collected as described in the manual (Stratagene), catalog number 200450 in the *in vivo* mass excision section. Sequence analysis was performed at CSHL using an ABI 3,700 Capillary sequencer for separation and nucleotide detection. Reactions were performed using a 1/16 Big Dye Terminator. Sequencing was performed with either the -21 M13 forward and/or reverse primer. The ESTs were assembled using Phrap (Ewing and Green, 1998; Ewing et al., 1998) and clustered into contigs using the CAP3 program (Huang and Madan, 1999).

Peptide Extraction

Peptide sequences were derived for all unigenes using the ESTScan application (Iseli et al., 1999) run with the default parameters. Prior to the ESTScan predictions, a *Ginkgo* species-specific ESTScan model was created. The ESTScan was trained with *Ginkgo* ORFs identified from the best match of BLASTX analyses performed on the unigene sequence against the Swissprot protein database. All BLASTX matches were filtered using the arbitrary expectation value of 1e-10.

Sequence Annotation

Sequence annotation on each of the *Ginkgo* cluster consensus sequences and derived peptides were performed within the openSputnik application (Rudd, 2005). Results were assessed for possible contamination by searching for homology to the *E. coli* and human genomes and were scored for homology to a wide range of non-coding RNAs and plant chloroplast and mitochondrial genomes. Homology searches were performed using the BLAST application (Altschul et al., 1990) and results were filtered using the expectation value <1e-10. Functional assignment was performed on both cluster consensus sequence and the peptide sequence. Assignments were made using BLASTX and BLASTP respectively against the MIPS catalog of functionally assigned proteins (FunCat) (Ewing and Green, 1998; Ewing et al., 1998), tentative functional assignments were filtered using the expectation value <1e-10.

Categorization of *Ginkgo* Contigs

All *Ginkgo* contigs sequences were aligned against a PlantEST database using TBLASTX (Altschul et al., 1990) and BLASTX against the NR (aa) database. The PlantEST database was created by downloading all plant ESTs in GenBank and assembling them using Phrap (Ewing and Green, 1998; Ewing et al., 1998). Todd Wood from Clemson University provided the PERL script that creates the PlantEST databases as described above. The NR (aa) database is a non-redundant database of protein sequences from GenBank.

Determination of Gymnosperm Specific Genes

All available plant ESTs were downloaded from GenBank and separated into three datasets consisting of angiosperms (monocots and dicots), gymnosperms, or non-seed plants (ferns, mosses and algae). Downloaded ESTs were assembled using Phrap (Ewing and Green, 1998; Ewing et al., 1998). All matches with an expect value <1e 10-5 are considered significant.

DISCUSSION AND RESULTS

Construction of a cDNA Library from *Ginkgo biloba* Fertile and Vegetative Tissue

Young organs (Figures 2A and 2D) were collected during the spring from the opening buds of short-shoots immediately after their emergence. At this stage, the megasporangium consists of an axis typically bearing two ovules (Figure 2A–C). The ovule is composed of a single integument surrounding a developing nucellus (Figure 2C). The male structure consists of a main axis bearing two or more microsporangia (Figure 2D–F). The RNA was extracted from the following organs: megasporangia, microsporangia, and two sets of leaves collected from either male or female trees. The mRNA isolated from all four tissues was used to construct four separate cDNA libraries. (Both male and female leaf sequences were pooled during subsequent bioinformatic analysis). Size fractionation was used to enrich for full-length cDNAs during library construction. From this cDNA library, 6,434 sequence reads (ESTs) were generated. All *Ginkgo biloba* EST reads have been deposited in GenBank. It was determined that 3,739 (58%) of the cDNA clones were over 500 bp long. The 3,618 of the reads were generated from the 5' end of the cDNA, and 2,816 were sequenced from the 3' end. Cluster analysis on the EST sequence produced a unigene set of 3,830 contigs consisting of 2,851 singletons and 979 assemblies. Of the clustered ESTs, the longest contig was 2,172 bp.

The entire unigene set or complete *Ginkgo* BLAST files can be downloaded at the website (<http://nypgenomics.org>). Each *G. biloba* contig is given a numeric identifier. The constituent ESTs for each contig can be obtained at this website. Additional bioinformatic analysis of the *Ginkgo biloba* dataset can be accessed at the open Sputnik Comparative Genomics Platform at (<http://sputnik.btk.fi/>). This site features sequence annotations, peptide sequence predictions, protein domain architectures, and putative molecular markers (ISSRs) for the *Ginkgo* EST derived unigenes. The sequence can be downloaded either as a fasta file, a clustered fasta file or as the derived peptide fasta file. In addition, BLAST analysis can be performed with the clustered ESTs from a given *Ginkgo* organ against all genes in *Arabidopsis thaliana* or distinct plant clustered EST datasets using the ViCoGenTa program available at the New York Plant Genomics Consortium website (<http://nypgenomics.org>)

Ginkgo Contig Matches to Genes in Angiosperms, Gymnosperms, and Non-Seed Plants

The TBLASTX (expect <1eX10-5) was used to compare the *G. biloba* unigenes against all available plant ESTs from TIGR (The Institute for Genomic Research), and the Plant Genome DataBase (Plant GDB). The ESTs from these databases were downloaded and clustered into unigenes, which were used in the comparison. Next the *Ginkgo* unigene set was compared against the *Arabidopsis* and rice genome annotated protein sequences downloaded from TIGR. All genes used in this comparison against *Ginkgo* were divided into one of three taxonomically relevant categories: (1) angiosperms, (2) gymnosperms, and (3) non-seed plants. The angiosperm category encompasses all annotated rice and *Arabidopsis* genes identified from their respective

genomic sequences, as well as all higher plant ESTs. The majority of the gymnosperm ESTs came from the conifer groups pine and spruce but also include ESTs generated from the Plant Genomics Consortium containing ESTs from the two other gymnosperm clades, Cycadales, and Gnetales. The non-seed plant category consisted of genes from all remaining plant ESTs including ferns, fern allies, bryophytes, and algae available with the majority of the sequences originating from *Physcomitrella patens* and *Chlamydamonas reinhardtii*.

A Venn diagram shown in Figure 3A displays the number of *Ginkgo* contigs, which are shared between one or more of the plant EST datasets at low BLAST stringency value (expect $<1eX10^{-5}$). From the Venn diagram, it can be seen that a majority of *Ginkgo* unigenes (2749/3830) match genes in other plants, and 1,081 have no match to other plant genes. Of those 2,749 *Ginkgo biloba* unigenes with matches to other plant genes, a subgroup of 256 unigenes had no corresponding match to genes in the angiosperm dataset. Of these 256 *Ginkgo* genes that do not match angiosperms, four also match genes in non-seed plants.

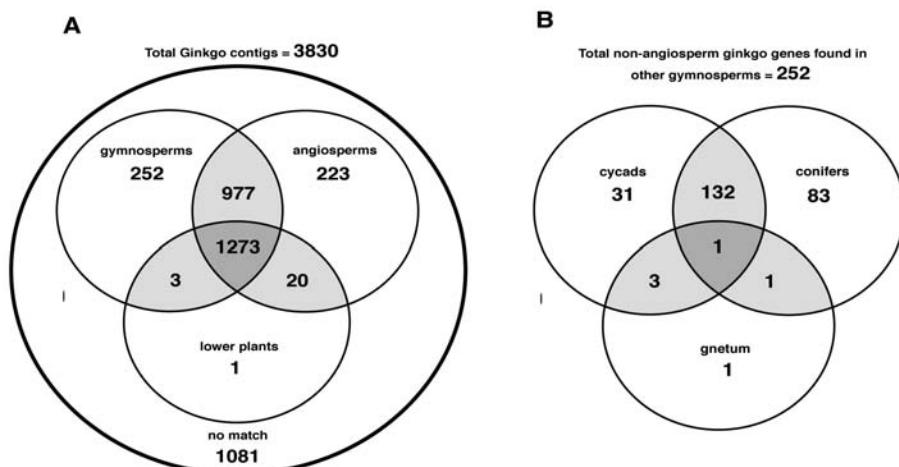


Figure 3. A Venn diagram illustrating the number of *Ginkgo* contigs with shared homology to genes found in non-seed plants, gymnosperms, and/or angiosperms (A). A BLASTX E value $> 10^{-5}$ was used as a cut-off. (B) The *Ginkgo* contigs with similarity to gymnosperms but (no match to angiosperm genes) were further subgrouped according to their BLAST score homology (E value $> 10^{-5}$) within gymnosperm taxa.

The 252 *Ginkgo biloba* unigenes that only match gymnosperm genes were next partitioned into matches between the three other gymnosperm orders: Cycadales, Coniferales, and Gnetales (Figure 3B). Since there are significantly more conifer unigenes ($>60,000$) than cycad unigenes (5,459) that were used in this comparison, one would expect that the number of matches between *Ginkgo* and conifers would be significantly greater than matches between *Ginkgo* and cycads. However, the actual

number of matches between *Ginkgo* and conifers (215) is only slightly more than for *Ginkgo* and cycads (163). In other words, despite the fact that there is over 10 times the number of conifer genes than cycad genes used in this comparison, *Ginkgo* matches to conifers are only 1.3 times greater than matches between *Ginkgo* and cycads. Of the matches between *Ginkgo* and gymnosperms, 31 match only cycads, (22 match *Cycas rumphii*, six match only unigenes from the cycad species), *Zamia furfuracea*, (753 contigs deposited in Genbank from Brenner et al., unpublished data); and three *Ginkgo* contigs match unigenes from both *Cycas rumphii* and *Zamia furfuracea*.

As one might expect, unigenes with matches to other plants are somewhat longer than those that have no match to other plants. Of all *Ginkgo* unigenes with matches to other plants, 89% are greater than 300 bp, whereas 72% are greater than 300 bp than those *Ginkgo* unigenes with no matches to other plants.

Common Genes between Cycads and *Ginkgo*

Our comparative analysis of the *Ginkgo* EST dataset builds upon our results from a previous genomic study on the cycad, *Cycas rumphii* (Brenner et al., 2003). In our current analysis, three (CB090673:GinkgoA3816, CycadCB089620:GinkgoA3730, and CycadCB089926:GinkgoA1532) of the 14 unigenes from cycads previously found only among gymnosperms (after the full-length clone was sequenced), were also homologues to *Ginkgo* genes found only in gymnosperms. Considering the relatively small number of unigenes from *Ginkgo* (3,830) and cycads (4706) available for our comparative studies, the detection of the same gene match in *Ginkgo* and cycads with homology to only gymnosperms strengthens the argument that these genes are gymnosperms specific.

***Ginkgo* Matches to Non-Seed Plants but Not Angiosperms**

An additional four *Ginkgo* unigenes that are not found in angiosperms were detected in non-seed plants. Three of these *Ginkgo* unigenes (*GinkgoA2411*, *GinkgoA3214*, and *GinkgoA325*) match non-seed plants and other gymnosperm genes whereas the forth *Ginkgo* gene (*GinkgoA2273*) only matches non-seed plants with similarity to a gene in *Chlamydomonas*.

Classification of *G. biloba* ESTs by Functional Categories

Each contig from the *Ginkgo* dataset was automatically assigned to a functional category (FunCat) based on its top match against the MIPS FunCat list of functionally annotated gene sequences from *S. cerevisiae* and *A. thaliana* databases using BLASTP. A non-stringent expect value (E-value) of <e-10 was chosen as the threshold. Table 1 illustrates the relative fraction that each functional category comprises within the entire unigene set compared to our previous study in *Cycas rumphii* (Brenner et al., 2003). The four largest categories of *Ginkgo* ESTs according to this functional categorization are: “cellular organization” (19%), “metabolism” (11%), “unclassified proteins” (14%), and “protein synthesis” (8%). In general, these same categories are also the highest in the cycad EST library from our previous work, except for “protein synthesis,”

which appears increased in *Ginkgo*, whereas interestingly, the category of cell growth, cell division, and DNA synthesis is reduced in *Ginkgo* compared to cycads.

Table 1. Placement of *Ginkgo* unigenes into functional categories (FunCats). *Ginkgo* genes with a BLASTP expect value (E-value) of >e-10 were assigned into FunCats based on their similarity score. The analysis was performed at the Centre for Biotechnology, Turku, Finland. Previous FunCat analysis data from *Cycas rumphii* is shown.

Functional category	Ginkgo %	Cycad %
Metabolism	11.0	10.3
Energy	5.0	5.1
Cell Growth, Cell Division and DNA Synthesis	5.4	9.2
Transcription	3.6	5.0
Protein Synthesis	8.4	5.8
Protein Destination	5.8	8.3
Transport Facilitation	2.0	1.3
Intracellular Transport	3.4	4.3
Cellular Biogenesis	3.5	3.3
Cellular Communication/Signal Transduction	3.8	4.2
Cell Rescue, Defense, Cell Death and Ageing	7.2	6.8
Ionic Homeostasis	0.5	0.1
Cellular Organization	18.8	21.6
Classification Not Yet Clear-Cut	7.3	4.6
Unclassified Proteins	14.1	10.3

***Ginkgo* Genes Involved in Development**

Analysis of the *Ginkgo biloba* dataset revealed a number of ESTs with highest BLAST similarity to genes with known roles in higher plant developmental processes. A sampling of some of these genes is shown in Table 2. These genes included the *Polycomb* gene CURLY LEAF (Goodrich et al., 1997; Kohler and Grossniklaus, 2002) as well as Lateral Organ Boundaries (LOB) (Shuai et al., 2002), Early Flowering 5 (ELF5) (Noh et al., 2004), Flowering Locus T (FT) (Weigel et al., 2000), and CONSTANS (for review see (Hayama and Coupland, 2004)) as well as five ESTs that match *MADS box* genes, some of which appear identical to previously cloned fragments of *MADS* genes from *Ginkgo biloba* including the *G. biloba* ortholog of AGAMOUS (Jager et al., 2003). Other genes in our EST library have homologies to proteins that regulate development through protein turnover including SPA-1 (Saijo et al., 2003), COP1 (Seo et al., 2003), and COP9 (Chamovitz and Glickman, 2002; Hellmann and Estelle, 2002). This sampling reveals that the EST dataset of *Ginkgo biloba* is a rich source of genes encoding proteins with known roles in development at the transcriptional as well as post-transcriptional stage.

Table 2. Similarity match of *Ginkgo* unigenes to genes involved in development. The *G. biloba* unigene set was compared to Genbank using a BLASTP cut-off score <e-5. The top match is listed under subject description. The organ(s) from which the listed ESTs were detected are: G = megagametophyte, I = microgametophyte, and L = leaf.

Genes in *Ginkgo biloba* with similarity to developmental genes in other plants

Contig id.	Organ	BLAST homology match	E-value
A2725	G	gi 1903019 emb CAA7 599.1 curly leaf [<i>A. thaliana</i>]	4.00E-37
A3095	I	gil 523 I388 ref NP_188001.1 LOB domain family protein [<i>A. thaliana</i>]	2.00E-21
A2815	I	gi 42541 156 gb AAS 19471.1 EARLY FLOWERING 5 [<i>A. thaliana</i>]	1.00E-33
AD51	I	gi 4903 I39 dbi BAA77836. 1 extensive homology to FT (FLOWERING LOCUS T, AB027504)	4.00E-10
A241	I, L	gi 41 323976 gb AAS00054. 1 CONSTANS-like protein CO I [<i>Populus deltoides</i>]	1.00E-30
AI591	L	gb AAG43405.1 AF 17293 1_1 homeobox 1 [<i>Picea abies</i>]	5.00E-23
AI591	G	gi 14715 I83 emb CAC44080. 1 putative MADS-domain transcription factor DEFH7 [<i>A. majus</i>]	2.00E-39
A2737	G	gi 30230270 gb AAM76208.1 AGAMOUS-like MADS-box transcription factor [<i>Ginkgo biloba</i>]	5.00E-28
A2850	I	gi 25307918 pir S5 1935 probable MADS-box protein dall - Norway spruce dall [<i>Picea abies</i>]	5.00E-72
A629	G, L	gi 7446554 pir T I075 I MADS-box protein MADS9 - Monterey pine	6.00E-17
A352	G, I	gi 7446559 pir T0957I MADS box protein MADS2 - Monterey pine	4.00E-75
A914	G, I	gi 1840 1293 ref NP_565632.1 COP9 / CSN signalosome complex subunit [A. thaliana]	1.00E-45
A2730	G	gi 15225760 ref NP_1 80854. 1 COPI regulatory protein [<i>A. thaliana</i>]	2.00E-60
A944	G, I	gi 30694320 ref NP_849784.1 argonaute protein(AGO I) [<i>A. thaliana</i>]	1.00E-56

CONCLUSION

The Importance of *Ginkgo* for the Study of Plant Evolution

As the sole remaining species of an ancient genus of plants which has survived nearly 170 million years from the Jurassic (Zhou and Zheng, 2003). *Ginkgo biloba* is a taxonomic and geographic relict that may be even older because fossils displaying a “ginkgophyte” vegetative morphology have been found as early as the Permian (Rothwell et al., 1997). *Ginkgo* has a number of plesiomorphic (unspecialized) as well as apomorphic (derived) traits that make it a valuable tool to study the evolution of seed plants. Here we used a genomic approach to investigate the genes involved in regulating development in *Ginkgo* by creating an EST library from both reproductive and vegetative tissues.

Similar to our previous analysis in *Cycas rumphii*, our *Ginkgo* EST study has found significant BLAST homology between *Ginkgo* ESTs with plant genes in gymnosperms and non-seed plants but not in angiosperms. Since ESTs, even when clustered in contiguous genes, may not represent the complete gene (Rudd, 2003), often one will find homology to angiosperm genes when the remaining *Ginkgo* gene sequence is revealed. For example in the gymnosperm, *Pinus taeda* EST collection, contigs of increasing length have a higher likelihood than shorter contigs matching a known gene

in the *Arabidopsis* genome (Kirst et al., 2003). However, in this same study a significant subcategory of very long contigs (>1,900 bp) have no homology to *Arabidopsis* (Kirst et al., 2003). It is likely that at least some of these long contigs with no match to angiosperm genes represent full-length genes that are specific to the gymnosperm and/or seedless plant clades. Our strategy to address this question involves screening for these same genes in additional taxa of gymnosperms, in the case of this study, *Ginkgo biloba*. In our analysis three *Ginkgo* genes that were only found in gymnosperms also matched the 14 ESTs from our previous study of gymnosperm common cycad genes.

Along these same lines, our results suggest the presence of genes common to non-seed plants and gymnosperms that are not present in angiosperms. This non-seed plant/gymnosperm grouping is not surprising considering the fact that gymnosperms have morphologically common characters that are not found in the angiosperms—particularly in their reproductive structures. For example, the megagametophyte is highly reduced both in cell number and in structural organization in angiosperms when compared to gymnosperms. Although these results cannot say for certain that these genes are specific to non-seed plants and gymnosperms, or more specifically that these genes are found in gymnosperm structures that are not found in seed plant, it nonetheless represents an important starting point to correlate the presence or absence of gymnosperm genes in angiosperms and/or lower plants.

Are Cycads and *Ginkgo* Sister Taxa?

One result from our study found that the number of *Ginkgo* contig matches to conifers are only 1.3 times greater than matches between *Ginkgo* and cycads despite the fact that there is over 10 times the number of conifer genes than cycad genes used in this comparison. Taken together these results might indicate a closer evolutionary association between *Ginkgo* and cycads than between *Ginkgo* and conifers. This bias towards cycad/*Ginkgo* similarity correlates with the fact that the majority of molecular phylogenetic studies place the cycads sister group to *Ginkgo*. Hopefully, this preliminary data will encourage further phylogenomic studies to fully resolve the hierarchy among extant gymnosperm orders. Until the full genome sequence becomes available for key gymnosperm taxa, EST sampling provides an important initial step for large scale identification of molecular markers to generate robust phylogenetic trees.

Developmental Regulators in *Ginkgo*

In *Ginkgo biloba* we note here a variety of genes with similarity to developmental regulators in angiosperms. We also note below that homologues to some of these developmental regulators are also present in our *Cycas rumphii* library as either orthologs to those found in higher plants or at least, belonging to the same gene family. An EST from *Ginkgo biloba* that was detected in the megagametophyte library has high similarity to the *Arabidopsis* CURLY LEAF (*CLC*) gene, which belongs to the Polycomb-group proteins (PcGs). The PcGs epigenetically regulate downstream target genes (Kohler and Grossniklaus, 2002). The PcGs modify chromatin-protein complexes that repress homeotic gene transcription and influence cell proliferation. In *Arabidopsis* *PcG* genes have been shown to regulate *MADS box* genes (Katz et al., 2004). The *CLC* protein product regulates the expression of AGAMOUS (Goodrich et al., 1997), a

gene controlling floral organ identity (Bowman et al., 1991). Interestingly, an ortholog for angiosperm AGAMOUS was also detected in the *Ginkgo* megagametophyte library (Table 2). *Ginkgo* AGAMOUS, (previously named GBM5) was identified in a study where the MADS domains were examined in *Ginkgo* (Jager et al., 2003). In this work *Ginkgo* AGAMOUS was shown via RT-PCR to be expressed in not only female but also in male and vegetative tissue. In our analysis, five total MADS box homologues were also detected in the *Ginkgo* EST dataset. Three of the *Ginkgo* ESTs from our library, GinkgoA2340, GinkgoA2730, and GinkgoA2850, are perfectly identical to the MADS domain gene fragments previously cloned by (Jager et al., 2003) as degenerate PCR products. The other two unigenes from our dataset have homologies to *MADS* genes (*GinkgoA629* and *GinkgoA352*), but do not specifically match any of the PCR fragments isolated in their study. These two *MADS box* unigenes either do not include the small region amplified in their degenerate PCR screen or could alternatively be unique *MADS* genes not isolated in their study. Unlike the degenerate primer approach used to isolate *MADS* genes, our EST approach offers the additional advantage of cloning entire genes or at least substantially large gene fragments. Among the few developmental genes examined in gymnosperms, most attention has focused on the expression of *MADS* homologs (Brenner and Stevenson, 2005; Theissen et al., 2000).

Other developmental genes found in the *Ginkgo* EST library include those with homology to regulators of flowering such as Early Flowering 5 (ELF5), which controls the levels of the gene FLC, which itself is a central regulator of flowering (Noh et al., 2004). Another *Ginkgo* EST includes Flowering Locus T (FT), which belongs to a small family of genes (FT/TFL1) that act to promote flowering as a downstream component from CONSTANS (Kardailsky et al., 1999). The CONSTANS is a transcription factor that has a critical role integrating circadian rhythms and light signals (for review see (Hayama and Coupland, 2004)). As one would expect an EST homolog for the CONSTANS gene family was found in *Ginkgo*. The CONSTANS belongs to a large gene family, which may have redundant roles in plants (Lagercrantz and Axelson, 2000). Not surprisingly, we also found homologs to CONSTANS in our previous study on cycad leaf ESTs (Brenner et al., 2003). In that flowering plants are believed to have evolved from gymnosperms, a survey of CONSTANS, ELF, and FT in gymnosperms, particularly in very young reproductive tissue might help define the origins of reproductive induction in non-flowering plants. Among the other genes related to developmental regulators includes a homologue to Laterol Organ Boundaries (LOB) domain gene family which in *Arabidopsis* has over 40 members (Shuai et al., 2002). The molecular mechanism of LOB domain containing genes is unknown, but one gene in *Arabidopsis*, ASYMMETRIC LEAVES2, is required for normal leaf development, by potentially acting as a regulatory repressor of KNOX genes (Lin et al., 2003). A KNOX homolog is also present in our EST library and was found in male reproductive tissues and HOX genes were also detected in our previous analysis in *C. rumphii*.

Another important component regulating development occurs at the level of protein degradation. A gene recognized in our EST library includes COP1. The COP1, serves as an E3 ubiquitin targeting photomorphogenic factors such as HY5 for degradation (Saijo et al., 2003). Another *Ginkgo* EST from the library has highest similarity to COP9. In our previous EST analysis in *Cycas rumphii* an EST was also isolated

with similarity to COP9 (Brenner et al., 2003). The COP9 is a subunit of the COP9 signalosome complex that controls multiple signaling pathways that regulate development in all eukaryotes (Chamovitz and Glickman, 2002; Hellmann and Estelle, 2002). In *Arabidopsis*, the cop9 and cop1 mutants are constitutively photomorphogenic in dark grown seedlings (Suzuki et al., 2002). Unlike angiosperms, seedlings from conifers are constitutively photomorphogenic when grown in the dark (Bogdanovic, 1973; Peer et al., 1996). In *Ginkgo*, chlorophyll and chloroplast development is completely dependent on light, however this process proceeds at a markedly slower pace than in flowering plants. That is, photomorphogenic development in *Ginkgo* seedlings is strongly delayed after transfer from dark grown conditions to light grown conditions when compared to seed plants (Chinn and Silverthorne, 1993; Chinn et al., 1995). The dark grown phenotype of cycads is unreported. Considering this variability in photomorphic development among and between the gymnosperms and the angiosperms, the discovery of genes encoding photomorphogenic regulators in gymnosperms will help understand the evolution of photomorphogenesis in seed plants.

Taken together, our genomics analysis of *Ginkgo biloba* is an important additional step to analyze the role of molecular development of early seed plants. Thus the stage is set to further determine the role of these genes during the development of ancillary structures found between *Ginkgo*, cycads, and other gymnosperms with higher plants as well as the role of those in structures that are unique to gymnosperms and/or the non-seed plants as a step to understand the evolution of the seed plant habit.

KEYWORDS

- *Cycas rumphii*
- Expressed sequence tags
- *Ginkgo biloba*
- *Megasporangia*
- Venn diagram

AUTHORS' CONTRIBUTIONS

Eric D. Brenner conceived of this project. He participated in its design, experiments, and drafted the manuscript. Dennis W. Stevenson and Gloria M. Coruzzi also conceived this project and participated in its design and coordination. Manpreet S. Katari played the major role in the bioinformatics analysis. Stephen A. Rudd performed the FunCat analysis and built the Sputnik website, Andrew W. Douglas performed the SEM work, Walter N. Moss and Giulia M. Stellari performed the histological sectioning, Richard W. Twigg and Suzan J. Runko performed the cDNA library construction, RM facilitated the EST sequencing. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank Vivekanand Balija for sequence generation. Funding for this work comes from the Plant Genomics Consortium. The Plant Genomics Consortium is made possible by the generosity of the Altria Group, Inc., The Mary Flagler Cary Charitable Trust, The Eppley Foundation for Research, Inc., The Ambrose Monell Foundation, The Wallace Genetic Foundation, Inc., and the National Science Foundation Plant Genome Group grant number DBI-0421604.

Chapter 13

Frequent Fires in Ancient Shrub Tundra

Philip E. Higuera, Linda B. Brubaker, Patricia M. Anderson,
Thomas A. Brown, Alison T. Kennedy, and Feng Sheng Hu

INTRODUCTION

Understanding feedbacks between terrestrial and atmospheric systems is vital for predicting the consequences of global change, particularly in the rapidly changing Arctic. Fire is a key process in this context, but the consequences of altered fire regimes in tundra ecosystems are rarely considered, largely because tundra fires occur infrequently on the modern landscape. We present paleoecological data that indicate frequent tundra fires in North Central Alaska between 14,000 and 10,000 years ago. Charcoal and pollen from lake sediments reveal that ancient birch-dominated shrub tundra burned as often as modern boreal forests in the region, every 144 years on average (± 90 s.d.; $n = 44$). Although paleoclimate interpretations and data from modern tundra fires suggest that increased burning was aided by low effective moisture, vegetation cover clearly played a critical role in facilitating the paleofires by creating an abundance of fine fuels. These records suggest that greater fire activity will likely accompany temperature-related increases in shrub-dominated tundra predicted for the 21st century and beyond. Increased tundra burning will have broad impacts on physical and biological systems as well as on land-atmosphere interactions in the Arctic, including the potential to release stored organic carbon to the atmosphere.

Tundra and boreal ecosystems store one third of the world's soil carbon (Post et al., 1982) The fate of this vast carbon stock has become a major concern to global-change scientists because its release to the atmosphere could exacerbate CO₂-related climate change (Chapin et al., 2000; Mack et al., 2004; Oechel et al., 2000; Shaver et al., 2006; Sitch et al., 2007). Unfortunately, uncertainty about a number of ecosystem processes hampers predictions of future tundra carbon cycling and the potential consequences to the climate system. One of the most important processes is how vegetation and climate change will alter fire regimes of tundra regions (Bond-Lamberty et al., 2007; Chapin et al., 2000; Sitch et al., 2007). Available evidence suggests that ongoing vegetation and climate change could significantly increase the rate of burning in northern tundra (ACIA, 2004), which is currently dominated by low-biomass communities (graminoids, herbs, and dwarf shrubs) that seldom burn (e.g., only 3% of Alaskan tundra burned between CE 1950 and 2005; Figures 1). In particular, a marked increase in shrub abundance and density, likely resulting from climate warming (Tape et al., 2006) is changing the physiognomic structure of arctic and subarctic regions. Shrubby growth forms increase the abundance of fine fuels available for burning, and in light of 3–5°C warming predicted over the next century (ACIA, 2004) such fuel changes could result in fire regimes vastly different from those in modern tundra. Unfortunately, short

observational fire records (e.g., 48 and 57 years in Canada and Alaska; 9,11), a lack of fire-history studies, and the possibility of novel future vegetation (Edwards et al., 2005) result in little information to evaluate how tundra fire regimes may respond to future climate and vegetation change. The paleoecological approach circumvents these limitations and offers the only way to obtain long-term empirical records of fire and vegetation change relevant for understanding tundra fire regimes under future climate and vegetation scenarios.

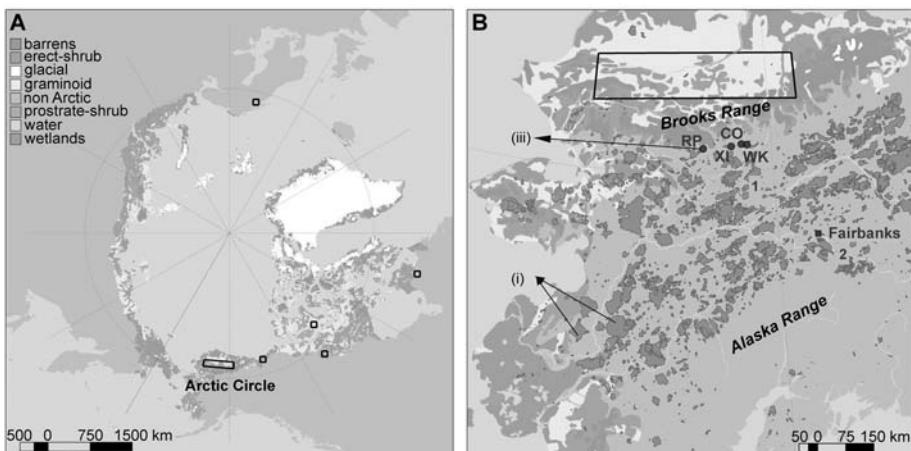


Figure 1. Distribution of modern circumpolar Arctic tundra (Walker et al., 2005), Alaskan fires from CE 1950–2005, and sites referred to in the text. (A) Black rectangles indicate circumpolar regions showing recent increases in shrub densities and/or extent (Tape et al., 2006). (B) Alaskan fires from CE 1950–2005 (i polygons) in tundra and boreal forest. Fires burned only 3% percent of Alaskan tundra, representing 6% of the total area burned in the state. iii dots identify lakes used in this study: Ruppert (RP) and Xindi (XI) lakes contain records of fire and vegetation from the Shrub Tundra Zone; Ruppert, Code (CO), and Wild Tussock (WK) lakes contain records from the Boreal Forest Zone (5.5–0 ka BP). Sediment-charcoal records from Sithylemenkat Lake (1) (Earle et al., 1996) and Lost Lake (2) (Tinner et al., 2006) also show qualitative evidence of increased fire activity within the Shrub Tundra Zone.

Here we present fire and vegetation reconstructions from northcentral Alaska that document frequent fires in shrub tundra during the late-glacial and early-Holocene periods (1410 ka BP [ka BP = thousand calendar years before present, CE 1950]). Vegetation and climate controls of these unusual fire regimes are inferred from paleovegetation records from each of two sites and from regional paleoclimate interpretations for this period. We also present an analysis of the climate space occupied by modern tundra vegetation and modern tundra fires in Alaska (CE 1950–2004). This analysis provides additional support for the climate-fire relationships inferred from the paleo-data.

MATERIALS AND METHODS

Lake Sediment Cores

We reconstructed fire and vegetation history from macroscopic charcoal and palynological data preserved in sediments from four lakes in the South Central Brooks Range

(Figure 1B). Ruppert Lake (3 ha; N 67°04'16", W 154°14'45"; 230 m asl) and Xindi Lake (7 ha; N 67°04'42", W 152°29'30"; 240 m asl) have records spanning late glaciation and the early Holocene (15–9 ka BP). Both sites are surrounded today by *Picea mariana* (black spruce) dominated boreal forest. Additionally, late-Holocene (last 5.5 ka BP) charcoal records from Ruppert, Code (2 ha; N 67°09'29", W 151°51'40"; 250 m asl), and Wild Tussock (15 ha; N 67°07'40", W 151°22'55"; 290 m asl) lakes provide information about fire regimes from the modern boreal forest [as defined by (Anderson and Brubaker, 1994)] for comparison with late-glacial and early-Holocene records.

Two parallel, overlapping sediment cores were collected from the center of each lake in summer 2001 (Code), 2002 (Ruppert), and 2003 (Xindi, Wild Tussock) using a modified Livingstone-type piston corer (Wright et al., 1984) and sliced at 0.25–0.5 cm intervals in the laboratory. Subsamples of 1 cm³ were prepared at varying intervals for pollen analysis according to PALE protocols (PALE, 1994) and pollen was counted to a terrestrial sum >300 grains at 400–1,000× magnification. For charcoal analysis, 3–5 cm³ subsamples were taken from contiguous core slices, soaked in sodium metaphosphate for 72 hr, washed through a 150 µm sieve, and bleached with 8% H₂O₂ for 8 hr. Charcoal was identified at 10–40× magnification based on color, morphology, and texture (Rhodes, 1998).

Chronologies

Chronologies are based on accelerator mass spectrometry (AMS) ¹⁴C dates of *Betula* macrofossils, concentrated *Picea* pollen grains, and/or concentrated charcoal particles, and all ages are expressed as calibrated ¹⁴C years before present (CE 1950). AMS ¹⁴C ages were calibrated using CALIB 5.0 and the IntCal 04 dataset (Reimer et al., 2004). Calibrated dates and corresponding confidence intervals represent the 50th, 2.5th, and 97.5th percentiles of the cumulative probability density function of calibrated ages, respectively (Telford et al., 2004). Chronologies were developed using a weighted cubic smoothing spline with the smoothing parameter determined by the average distance (cm) between dates, such that greater sampling resulted in a more flexible spline. The inverse of the 95% confidence interval of the calibrated ¹⁴C date was used for weighting.

Given the density of radiocarbon dates in and around the Shrub Tundra Zone, and that CHARs are sensitive to sedimentation rates, we evaluated whether general features of the CHAR series at Xindi and Ruppert lakes varied significantly when using 5–7 alternative age-depth models. In no case did high CHARs or the distinct peaks of the Shrub Tundra Zone disappear. Charcoal concentrations (pieces cm⁻³) are also high in this period, giving us confidence that the high CHARs reflect increased charcoal accumulation and are not chronological artifacts.

Statistical Treatment of Charcoal Data

Peaks in the charcoal accumulation rate (pieces cm⁻² yr⁻¹; CHAR) in lake sediment records have been shown both empirically (Lynch et al., 2004) and through mechanistic models (Higuera et al., 2007) to be associated with the local (0.5–1.0 km) occurrence of individual or multiple high-severity fires (“fire events”). Local fires

introduce charcoal to a lake via airborne fallout and create distinct CHAR peaks that exceed variability around a low-frequency trend. This characteristic can be taken advantage of to infer when local fires occurred in the past. We estimated the timing of fire events in our charcoal records by removing low-frequency trends (i.e., “background”); reflecting changes in the rates of charcoal production, secondary transport, sediment mixing, and sediment sampling (Rhodes, 1998) and applying a locally-defined threshold value that separates fire-related CHAR peaks (i.e., signal) from non-fire-related variability in CHARs (i.e., noise). Our approach accounts for changes in both the mean and variability of CHARs through time and the statistical nature of charcoal counts.

Prior to quantitative analysis, charcoal data were interpolated to constant 15 years time steps, approximating the median temporal resolution of each record. Low-frequency trends in CHARs, $C_{\text{background}}$, were modeled with a 500 years running median, smoothed with a locally-weighted regression (also with a 500 years window). We subtracted $C_{\text{background}}$ from the interpolated charcoal series to obtain a residual “peak” series, C_{peak} . For each sample in each record, we identified charcoal peaks when C_{peak} exceeded a sample-specific threshold value. Our threshold criterion assumes that fires create charcoal peaks that exceed C_{peak} variations related to sediment mixing, sediment sampling, and analytical noise, and that this variability changes on time scales ≥ 500 years. Thus, for each 500-year period, we assume that the distribution of C_{peak} values contains two sub-populations: C_{noise} and C_{fire} . C_{noise} is a normally-distributed population centered near 0 (i.e., $C_{\text{background}}$); C_{fire} samples are high CHARs exceeding variations in C_{noise} , presumably caused by local fires. We used a Gaussian mixture model to identify the mean and variance of the C_{noise} distribution (Gavin et al., 2006) and we used the 99th percentile of this distribution as the threshold value separating C_{fire} from C_{noise} . For all records, this procedure was done for each overlapping 500-year period, producing a unique threshold for each sample. Individual thresholds for each sample were smoothed with a locally-weighted regression (to 500 years). Finally, all peaks exceeding the locally-defined threshold were screened based on the original charcoal counts contributing to each peak. If the maximum count in a CHAR peak had a >5% chance of coming from the same Poisson-distributed population as the minimum charcoal count within the proceeding 75 years, then the “peak” was not identified (e.g., Charster user’s guide, accessed September 2007, <http://geography.uoregon.edu/gavin/charster/Analysis.html>; (Shiu and Bain, 1982). Our methods are contained within the program CharAnalysis, written by PEH and freely available at <http://CharAnalysis.googlepages.com>.

Quantifying Fire Regimes

We used dates of estimated fire events to calculate fire return intervals (FRIs) (years between fire events; FRIs), and we fit a two-parameter Weibull model to the distribution of FRIs within each vegetation zone using maximum likelihood techniques (Johnson and Gutsell, 1994). Each Weibull model passed a Kolmogorov–Smirnov goodness-of-fit test ($p > 0.10$), and we estimated 95% confidence intervals for the Weibull scale b , and shape c , parameters based on 1,000 bootstrapped samples from

each population. Confidence intervals for the mean FRI were calculated in the same manner. We used a likelihood-ratio test, based on likelihood values of the Weibull models, to test the null hypothesis that any two FRI distributions were similar (Johnson and Gutsell, 1994) (Thoman and Bain, 1969). The probability of Type I Error, p , was estimated using a permutation test, and the null hypothesis was rejected if $p < 0.05$.

Climate Space of Modern Tundra and Tundra Fires

The climate space occupied by modern tundra vegetation and tundra fires was quantified using tundra classification data from the circumpolar Arctic vegetation map (Walker et al., 2005) temperature and precipitation data from the Global Historical Climatology Network (W. Cramer, 2006 University of California-Berkeley, Integrative Biology, and U.S. Geological Survey, Alaska Geographic Science Office. Accessed on-line in January 2007: <http://agdc.usgs.gov/data/projects/hlct/hlct.html#A>), and area burned data from the Alaska Fire Service (accessed on-line in January 2007: <http://agdc.usgs.gov/data/blm/fire/index.html>). Climate data represent averages across variable periods, starting from 1888–1968 and generally ending in 1990. Each dataset was imported into a raster-based geographic information system with a 1 km² cell size. Climate space was determined based on the average June precipitation and average June temperature values from all cells with: (1) CAVM classification of tundra, and (2) burned cells with a CAVM classification of tundra.

DISCUSSION

High fire frequencies in the ancient shrub tundra prompt questions about the relative roles of vegetation (fuels) and climate (summer temperature and precipitation) in controlling fire regimes in the Shrub Tundra Zone and the implications of this natural experiment for understanding future Arctic environmental change. Climate is perhaps most often invoked to explain past changes in fire regimes. However, the influence of climate on the fire regime in the Shrub Tundra Zone is not straightforward. Near the end of *Betula* shrub dominance and afterwards (ca 11.59.0 ka BP), summer temperatures in northern Alaska may have approached or exceeded modern levels (Kaufman et al., 2004). However, such a temperature rise cannot explain the increase in fire frequencies at the beginning of the Shrub Tundra Zone, ca 14.012.0 ka BP. In contrast, paleoclimate proxies (Anderson et al., 2004) suggest that this period was characterized by cooler-than-present summers. Furthermore, lowered lake levels in interior Alaska indicate that effective moisture was lower than present throughout the Shrub Tundra Zone (Abbott et al., 2000). Because summer temperatures were cooler than modern, low effective moisture must have been a key factor facilitating the fuel drying necessary to maintain high fire activity within the ancient shrub tundra. The importance of low effective moisture for facilitating tundra burning is evident in the pattern of 232 tundra fires that burned in Alaska between CE 1950–2005. These fires were significantly skewed to tundra regions with relatively dry and/or warm summer climate conditions, that is, with mean June precipitation between 20–30 mm and mean June temperature between 6–10°C (Figure 2).

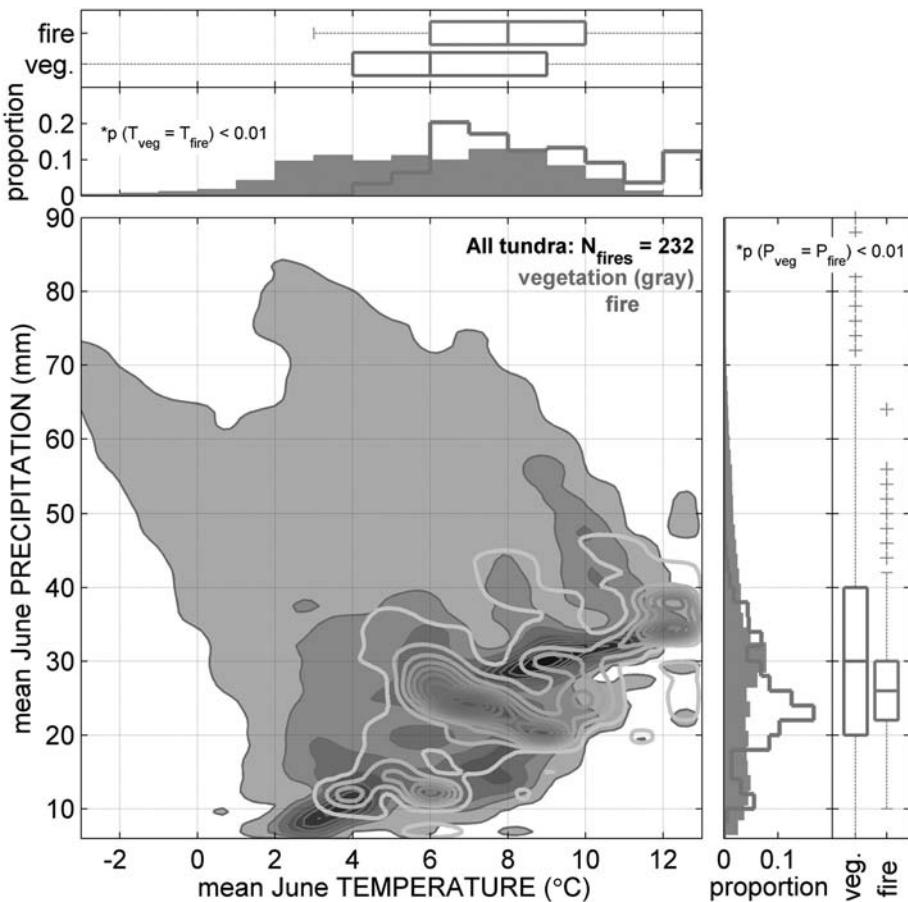


Figure 2. Fire return intervals (FRIs) from the Shrub Tundra Zone and the conifer-dominated Boreal Forest Zone (5.5–0 ka BP).

FRIs from the Shrub Tundra Zone at (A) Xindi Lake and (B) Ruppert Lake with fitted Weibull models. Weibull (Wbl) b (yr) and c (unitless) parameters, and the mean FRI (mFRI; yr) all include 95% confidence intervals. (C) Weibull models from the Shrub Tundra Zone (solid lines) and the Boreal Forest Zone (black dashed lines). All FRI distributions presented are statistically similar based on likelihood-ratio tests ($p>0.29$; see Results). The Weibull b and c parameters, and mFRI for Ruppert (boreal forest), Code, and Wild Tuussock lakes are 188 (147–239), 150 (123–178), and 149 (123–174); 1.53 (1.31–2.06), 1.85 (1.52–2.60), and 1.96 (1.61–2.75); 171 (135–216), 135 (113–160), and 135 (113–157), respectively.

Given our current understanding of late glaciation and the early Holocene, increased burning in the Shrub Tundra Zone was not a simple function of climate change. The distinct increase in CHARs and CHAR peaks at the onset of the Shrub Tundra Zone suggests that vegetation was a key element facilitating fires. The tall growth form, small stem diameters, and highly resinous twigs of *B. glandulosa* (Dugle, 1966) make it susceptible to fire on modern landscapes (de Groot and Wein, 1999) and a widespread cover of *B. glandulosa* in the past would have created the continuity of

flammable fuels necessary for fire spread. In addition, vigorous sprouting following fires (de Groot and Wein, 1999) would have provided the regeneration necessary to sustain fire frequencies similar to those of modern boreal forests (Figure 4). Based on paleo and modern relationships between tundra fire occurrence and corresponding climatic conditions, the role of fuels is central to understanding past and future shifts in tundra fire regimes. In the case of the Shrub Tundra Zone, the combination of abundant flammable fuels and low effective moisture overwhelmed the mitigating effects of low temperatures on landscape flammability.

Overall, paleorecords from northcentral Alaska imply that ongoing shrub expansion and climate warming will result in greater burning within northern tundra ecosystems. The geographic extent of fire-regime changes could be quite large, as shrubs are expected to expand over the next century in both herb and low shrub tundra ecosystems, which comprise 67% of circumpolar Arctic tundra (Tape et al., 2006; Walker et al., 2005) (Figure 1). Over this same period, annual temperatures in the Arctic are projected to increase between 3–5°C over land, lengthening the growing season and likely decreasing effective moisture (in spite of increased summer precipitation) (ACIA, 2004). How long might it take for the current shrub expansion to trigger a significant change in fire frequencies? Within the chronological limitations of our records, past shrub expansion, and fire-regime changes at each site occurred within a few centuries (Figure 3). The duration of this shift is consistent with the estimated rate of shrub expansion within a large area of northern Alaska [0.4% yr⁻¹ for ca 200,000 km²; 10]. Based on a simple logistic growth model and the assumption of a constant expansion rate, Tape et al. (2006) hypothesize that the ongoing shrub expansion in this region started roughly 125 years ago and should reach 100% of the region in another 125 years. Thus, if fuels and low effective moisture are major limiting factors for tundra fires, we predict that fire frequencies will increase across modern tundra over the next several centuries.

Although our fire-history records provide unique insights into the potential response of modern tundra ecosystems to climate and vegetation change, they are imperfect analogs for future fire regimes. First, ongoing vegetation changes differ from those of the late-glacial period: several shrub taxa (*Salix*, *Alnus*, and *Betula*) are currently expanding into tundra (Tape et al., 2006) whereas *Betula* was the primary constituent of the ancient shrub tundra. The lower flammability of *Alnus* and *Salix* compared to *Betula* could make future shrub tundra less flammable than the ancient shrub tundra. Second, mechanisms of past and future climate change also differ. In the late-glacial and early-Holocene periods, Alaskan climate was responding to shrinking continental ice volumes, sea-level changes, and amplified seasonality arising from changes in the seasonal cycle of insolation (Anderson et al., 2004) in the future, increased concentrations of atmospheric greenhouse gases are projected to cause year-round warming in the Arctic, but with a greater increase in winter months (ACIA, 2004). Finally, we know little about the potential effects of a variety of biological and physical processes on climate-vegetation-fire interactions. For example, permafrost melting as a result of future warming (ACIA, 2004) and/or increased burning (Liljedahl et al., 2007) could further facilitate fires by promoting shrub expansion (Tape et al., 2006) or inhibit fires by increasing soil moisture (Liljedahl et al., 2007).

Despite these uncertainties, Alaskan paleorecords provide clear precedence of shrub-dominated tundra sustaining higher fire frequencies than observed in present-day tundra. The future expansion of tundra shrubs (Tape et al., 2006; Walker et al., 2006) coupled with decreased effective moisture (ACIA, 2004) could thus enhance circumpolar Arctic burning and initiate feedbacks that are potentially important to the climate system. Feedbacks between increased tundra burning and climate are inherently complex (Chapin et al., 2000; Mack et al., 2004; Oechel et al., 2000; Shaver et al., 2006) but studies of modern tundra fires suggest the possibility for both short- and long-term impacts from (1) increased summer soil temperatures and moisture levels from altered surface albedo and roughness (Liljedahl et al., 2007) and (2) the release soil carbon through increased permafrost thaw depths and the consumption of the organic layer (Liljedahl et al., 2007; Racine et al., 2006). Given the importance of land-atmosphere feedbacks in the Arctic (Eugster et al., 2000; McGuire et al., 2006; Zimov et al., 2006) the precedence of a fire-prone tundra biome should motivate further research into the controls of tundra fire regimes and links between tundra burning and the climate system.

RESULTS

Trends in charcoal accumulation rates (pieces $\text{cm}^{-2} \text{yr}^{-1}$, CHARs) correspond markedly with shifts in pollen assemblages at Xindi and Ruppert lakes (Figure 3). Both records start in herb-dominated tundra (Herb Tundra Zone), indicated by high pollen percentages of Cyperaceae (sedge), Poaceae (grass), and minor herb taxa (e.g., *Artemisia* (wormwood), data not shown). Raw CHARs are low (medians = 0.01 and 0.00 pieces $\text{cm}^{-2} \text{yr}^{-1}$) with few identified peaks in the detrended series (Figure 3), suggesting little or no burning in the late-glacial herb tundra near these sites. Increases in CHARs (medians = 0.05 and 0.02 pieces $\text{cm}^{-2} \text{yr}^{-1}$) and the frequency of peaks in the detrended series coincide with a prominent rise in *Betula* (birch) pollen percentages (from <5 to 50–75%; 14.3 and 13.3 ka BP at Xindi and Ruppert lakes, respectively), which marks the expansion of *Betula* shrubs in the study area (Figure 3). These pollen assemblages (Shrub Tundra Zone) have higher *Betula* percentages than pollen assemblages from modern tundra in North America (Anderson et al., 2004) (e.g., 70% vs. 40%) and are thought to represent extensive thickets of tall (>1 m) *Betula glandulosa* (resin birch, inferred from measurements of pollen morphology) (Brubaker et al., 1983). The inferred vegetation of the Shrub Tundra Zone contrasts with the majority of modern circumpolar Arctic tundra, where only 12% of the area contains shrubs taller than 0.4 m (i.e., Low-shrub tundra; (Walker et al., 2005)). However, the vegetation structure of the Shrub Tundra Zone may be analogous to future Arctic tundra, which is predicted to have a major component of >0.5-m tall *Betula*, *Salix* (willow), and *Alnus* (alder) shrubs (Tape et al., 2006; Walker et al., 2006). Deciduous woodlands (Deciduous Woodland Zone), identified by samples with >10–20% *Populus* (poplar) pollen, characterized the vegetation from 10.59.0 ka BP (Figure 3). As in the Herb Tundra Zone, the low raw CHARs (medians = 0.02 and 0.01 pieces $\text{cm}^{-2} \text{yr}^{-1}$) and few peaks in the detrended series suggest less frequent fires as compared to the Shrub Tundra Zone.

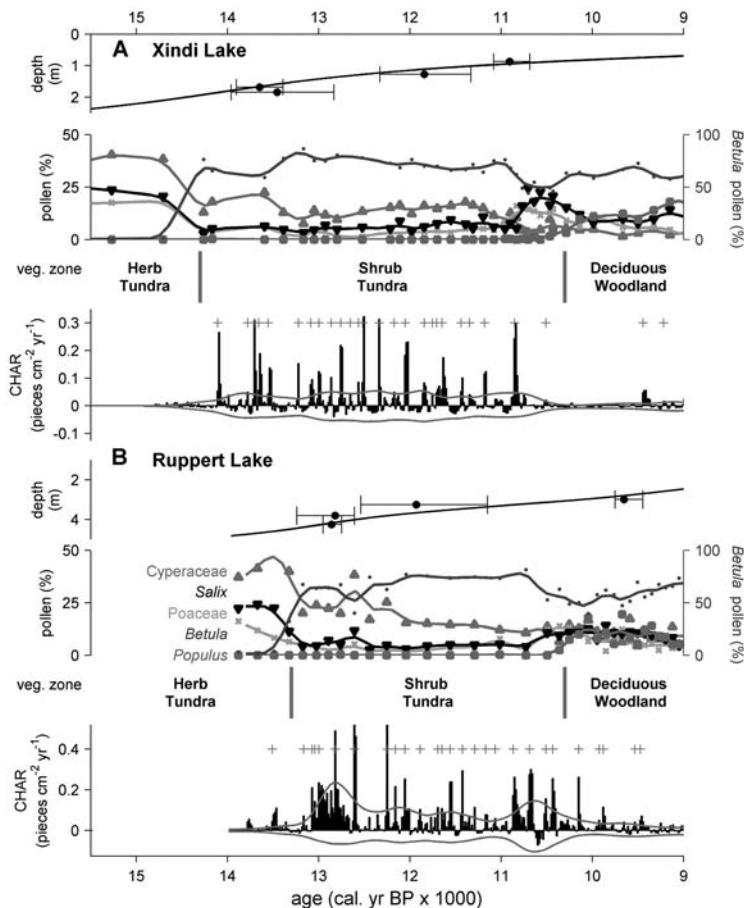


Figure 3. Climate space occupied by Alaskan tundra in the circumpolar Arctic vegetation map (Walker et al., 2005) and area burned within the same region from CE 1950–2005.

Darker shades represent a greater proportion of total tundra vegetation (gray) or total area burned within the climate space. Mean June temperature and precipitation distributions associated with tundra vegetation and area burned are shown as histograms and box plots. For both temperature and precipitation, distributions of vegetation, and area burned differ significantly based on a Kolmogorov–Smirnov test with $N_{\text{fires}} = 232$ degrees of freedom ($p < 0.01$). Most fires occurred in areas with a mean June temperature of 6–10°C and a mean June precipitation of 20–30 mm. The general bias towards warm and/or dry portions of the total climate space suggests an overriding importance of low effective moisture for facilitating tundra burning

Estimated fire frequencies within the Shrub Tundra Zone (Figures 3 and 4) were much higher than in modern tundra (Kasischke et al., 2002; Stocks et al., 2002) (Figure 1). Fire events (i.e., CHAR peaks) occurred on average (95% CI) every 150 (113–189) years at Xindi Lake and 137 (107–171) years at Ruppert Lake, with high variability around these means (FRI s range from 30–360 year; Figure 4). The FRI distributions at these two sites were statistically indistinguishable during this period ($p = 0.60$, $n = 24, 20$) and from FRI distributions in the late-Holocene boreal forests around Ruppert,

Code, and Wild Tussock lakes (p ranges from 0.29 to 0.99, n ranges from 20 to 39; see Materials and Methods; Figure 4). The fire-vegetation relationships observed at Ruppert and Xindi lakes during the Shrub Tundra Zone are likely regional in scale, as this tundra type is documented in a large network of pollen and macrofossil records in North Central Alaska (Anderson and Brubaker, 1994; Anderson et al., 2004; Edwards et al., 2005), and high fire activity has been qualitatively inferred from discontinuous charcoal records at other sites in interior Alaska (Earle et al., 1996; Tinner et al., 2006) (Figure 1B).

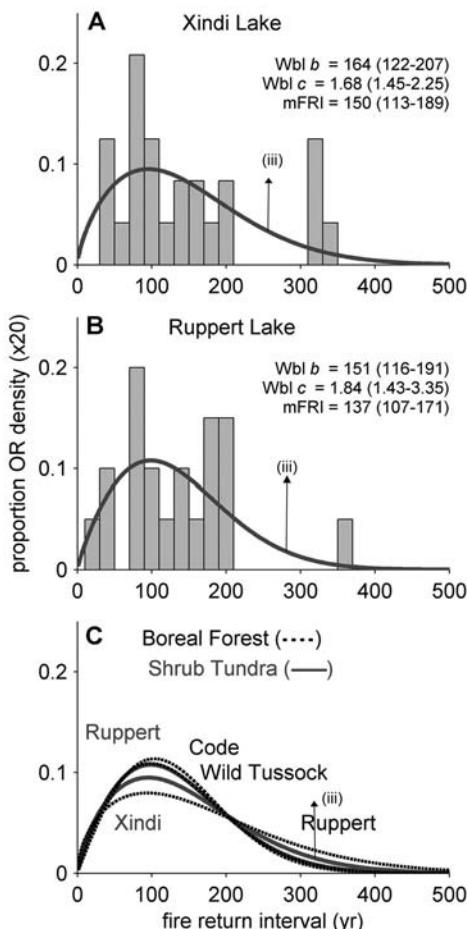


Figure 4. Fire and vegetation reconstructions from northcentral Alaska.

Chronology, pollen stratigraphy, inferred vegetation, and high-frequency variations in charcoal accumulation rates (CHARs) from (A) Xindi Lake and (B) Ruppert Lake. Pollen percentage curves are smoothed to 500 years and color coded. CHAR records represent residuals after removing 500-year trends, and i lines around $CHAR = 0$ are thresholds identifying noise-related variations. i plus marks identify CHAR peaks exceeding the positive threshold (and a minimum-count screening; see Materials and Methods) and are interpreted as local fire events. At both sites CHARs and CHAR peaks increase distinctly with the rise in *Betula* pollen percentages, marking the transition from the Herb Tundra Zone to the Shrub Tundra Zone.

KEYWORDS

- Accelerator mass spectrometry
- CHAR peaks
- High fire frequencies
- Shrub Tundra Zone

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Feng Sheng Hu, Philip E. Higuera, Linda B. Brubaker, and Patricia M. Anderson. Performed the experiments: Philip E. Higuera. Analyzed the data: Philip E. Higuera and Alison T. Kennedy. Wrote the chapter: Philip E. Higuera. Other provided significant input to the manuscript: Feng Sheng Hu. Gathered and summarized data on modern tundra fires: Alison T. Kennedy. Oversaw 14C dating and assisted in chronology development: Thomas A. Brown. Counted pollen from Xindi Lake: Patricia M. Anderson. Provided significant input to the manuscript: Patricia M. Anderson and Linda B. Brubaker. Counted pollen from Ruppert Lake: Linda B. Brubaker.

ACKNOWLEDGMENTS

Sampling was conducted under permit from Gates of the Arctic National Park and the Bureau of Land Management. We thank B. Clegg, J. Mauro, and K. Shick, for field assistance, Brooks Range Aviation, and VECO Polar Resources for field logistics, C. Adam, E. Cudaback, J. Leach, A. Lilienthal (Yambor), J. Smith, and E. Spaulding for laboratory assistance, and C. Whitlock, C. Carcaillet, and two anonymous reviewers for constructive comments on the manuscript.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

Chapter 14

History of Native Plant Communities in the South

Wayne Owen

What are the history, status and projected future of native plant communities in the South?

INTRODUCTION

Nowhere in America is there a greater variety of a native plant communities, native plant species, or rare and endemic native plants than in the forests of the Southeast. However, this exceptional bounty of diversity is under an increasing burden of impacts as a result of habitat conversion, alterations in community composition, and exotic pest and disease species. Human activities have impacted native plant communities since the first aborigines settled in the region, and humans are likely to remain a formative part of the southern landscape for the foreseeable future. As was the case at the beginning, the human use of native plants and their communities continues to mirror current societal needs. At the beginning of the 21st century the forested plant communities of the South are producing more than ever. Although the vast majority of the region's plant communities have been altered to a greater or lesser extent, an increasingly important societal need is the conservation of natural areas and the restoration of public lands. Rare vascular plant species are not evenly distributed throughout the South. Peaks of rare species diversity occur in the Southern Appalachians, the Florida Panhandle, and the Lake Wales Ridge region of Florida. Secondary peaks of rare species diversity are located in Arkansas' Ouachita Mountains and on the Cumberland Plateau.

Native plant communities in the South have been much studied and written about since the Bartrams explored the region in the 18th century (Bartram, 1791). Bartram noted that Native Americans, as well as European settlers altered native plant communities by intentional burning, land clearing for agriculture, clearcutting of timber, and introductions of exotic species from Europe and the Caribbean. The plant communities of the South were not pristine in Bartram's time, and they were not pristine when Europeans first arrived on these shores. The southern landscape had already seen 10,000 years of human history. The last 400 years, however, have brought more radical changes than any caused by Native Americans.

Today's landscape and vegetation are not only the result of a very long history of change; they are also the starting point of tomorrow's vegetation. To better understand the resource at hand, it is valuable to remind ourselves of how we got here so that, perhaps, we can do better in the future. For the purposes of this assessment, a native plant community is defined as:

A set of populations of plants naturally indigenous to an area that are interacting to the extent and degree that would have been observed prior to European settlement and share critical physiognomic and compositional traits.

It is somewhat arbitrary to confine the limits of what is natural to a pre-European timeframe, but it is impossible to separate the influences of native cultures from the “historical” landscape. And, even at the height of aboriginal culture in the South Eastern United States, Native Americans were not able to impact the native vegetation to the degree that the Europeans have.

Plant communities, both native and otherwise, are defined not only by their inter- and intra-specific interactions and composition—which species are present and in what numbers—but also by their structure. Major structural elements include seral stage; the relative abundance, age distribution, spatial arrangement of dominant species in each canopy layer; as well as physical metrics such as the height, size, and spatial arrangement of individuals. Natural disturbances such as hurricane blowdowns, ice storms, and drought are common events that markedly influence the structural condition of plant communities and have contributed to the perpetuation of a full spectrum of structural and seral conditions.

MATERIALS AND METHODS

The literature was reviewed for information about the history of southern vegetation. There are already several reviews of this material. The better treatments of the subject include Delcourt and Delcourt (1993), Mac et al. (1998), Ricketts et al. (1999), and Stein et al. (2000). An extensive and detailed primary literature exists on the paleobotany of the region based on palynology (the study of ancient pollen). Only a small portion of that information was used in this work, but anyone interested in further reading can consult the reviews of Watts (1980) and Delcourt and Delcourt (1998).

DISCUSSION

Plant communities of the South deserve many superlatives. They are exceptionally diverse, being rich in both the number of species and the number of endemic taxa. Forests of the South are also among the most heavily impacted in North America. They are heavily fragmented, have experienced greater levels of human habitation for longer than any other forests in North America, and have the greatest number of exotic species. The native plant communities of the South have a history of increasingly intensive use, but recent changes in social attitudes are a source of great hope to those that appreciate the very special qualities of the native southern landscape. There is no chance that the South will ever see the communities that Cabeza de Vaca or de Soto saw, or even the relatively more modified landscapes first described by Bartram or Nuttall. In fact, continuing urbanization and population pressures will almost certainly conspire to keep the majority of the South’s landscape working hard to support its people (Table 1). However, the remaining public land in the region is increasingly being managed for uses other than commodity production, and native plant community restoration and species protection activities on both public and private land, is at an all time high. Changes will continue into the future, most of them detrimental to

the overall health of native plant communities in the South. Increasing human populations and resource demands will further fragment the remaining forests and natural areas. Invasive species will occupy increasingly larger proportions of the Southern landscape. Global climate change will also impact the composition and distribution of plant communities in the South. However, increasing awareness of the value of forests and natural areas has slowed the pace of land conversion in the South, and recent efforts by State and Federal government landowners to improve forest conditions through restoration suggest that, at least in part, some of the inevitable changes coming to southern native plant communities are going to be improved. The native plant communities of the South will never be what they were, but if the future is able to bring forth increasing functionality to the remaining intact ecosystems of the South, then the conservation and restoration efforts of today will have been successful.

Table 1. Percentage of wetland acres lost in Southeast, 1780s through 1980s.

State	Hardwoods			Softwoods		
	All Owner-ships	National Forests	Industrial Forests	All Owner-ships	National Forests	Industrial Forests
Alabama	21931.9	605.4	5499.4	7447.1	237.2	2789.9
Arkansas	18392.1	2371.8	4514.6	5077.0	831.8	2450.3
Florida	14650.7	1029.5	4601.5	7437.8	725.5	2921.9
Georgia	23796.1	710.7	4890.5	10805.4	192.4	3154.3
Kentucky	12347.3	698.9	204.5	682.1	64.2	0.0
Louisiana	13783.0	568.5	4422.5	5006.7	327.9	2357.1
Mississippi	18587.4	1106.6	3314.1	5751.0	505.3	1579.7
North Carolina	18710.4	1082.4	2420.4	6261.9	168.0	1528.2
South Carolina	12454.9	560.0	2394.3	5561.5	311.2	1492.3
Tennessee	13965.0	556.8	1393.0	1468.9	93.3	336.6
Virginia	12094.9	1360.9	714.5	3352.8	137.2	840.3
Total	180713.7	10651.5	34369.3	58852.2	3594.0	19450.6

Data from Southern Region Forest Inventory and Analysis, <http://www.srsfia.usfs.msstate.edu/>

NEEDS FOR ADDITIONAL RESEARCH

The TNC national vegetation classification system is the most important development for the study of natural plant communities in the last decade. This uniform, standardized method for classifying plant communities will provide a reliable means for comparing where we are with where we have been. Alternatively, efforts to model the current and projected distributions of plant communities or forest trees can substantially aid our understanding of the distribution of plant diversity throughout the South. For example, Prasad and Iverson (1999) have developed multiple maps of the current and projected distributions of 80 eastern forest trees based on a variety of sets of projected condition.

Even though trained botanists have been exploring the Southern United States for over 300 years, the mapping of native plant communities has just begun. A full

accounting of the variation and geography of species and their communities is critical. This information is essential to make an accurate assessment of the conservation needs of the region.

The greatest challenges to natural plant communities throughout the nation, but particularly in the South, are conversion to agriculture, the creation of tree plantations, and urbanization. The fourth common source of degradation of natural plant communities is the incursion of exotic and invasive plant species. There is a great need to investigate more effective methods of control, whether chemical, biological, or physical. There are many safety concerns associated with chemical and biological control methods, but physical methods usually prove slow and expensive. It is impossible to eliminate exotic species from our region, but we can still take steps to reduce their impact on native plant communities and learn to better manage the impacts.

There is currently a management emphasis on the retention and development of old-growth forests, or forest stands with old-growth characteristics on public land. However, concerns over the habitat needs of wildlife, especially migratory birds, has recently highlighted the broader need for forests with a range of structural traits. Early successional forest stands in particular support a very different array of native plant communities than do mature forests. There is a significant opportunity for research to contribute to a better understanding of the historic abundance and distribution of open areas in the South.

A final future research priority for native plant communities should be restoration ecology. In the past, restoration has meant the establishment of any kind of vegetative cover on denuded landscape such as eroded farmland or strip mines. In the last decade, there has been a significant trend toward restoration of native communities using native plant material. However, the availability of native material is limited, and there is a growing concern about the source of the plant material used in restoration. We have much to learn about the distribution of genetic diversity in the native species commonly used for restoration, and even more to learn about the potential for use in restoration of the majority of plant species native to the South.

RESULTS

Prehistory of Southern Native Plant Communities

Through an understanding of the history of native plant communities in the South, this assessment hopes to put into context the background on which change has occurred. It is important to understand the roles that global climate change and indigenous human cultures played in shaping the plant communities that are considered “native” or “natural” today. In this assessment, only those works that address Quaternary, 2 million years before present (BP), and later floras are discussed. The primary focus is on the vegetation history of the Holocene, 10,000 years BP.

For the majority of the Quaternary, the climate of the Southeast has been colder than at present (Greller, 1988). During this period, there were multiple continental glaciation episodes that did not impact our region directly, but nonetheless had significant impacts on the composition of our native plant communities. These glaciations have been attributed by most to Milankovitch (1941) variations in the orbit of the Earth

about the sun. The components to the Milankovitch cycle are expressed at periods of approximately 100,000, 41,000, and 21,000 years (Delcourt and Delcourt, 1993). The interaction of these three cycle periods has been correlated with the relative severity of glacial periods and the rapidity with which glacial advances or retreats occurred.

The coastlines of the Southeastern United States achieved their present approximate position and shape during the early Quaternary (Christensen, 1988). Changes in sea level associated with Quaternary glaciations have profoundly affected the vegetation of the historic Coastal Plains, though due to normal coastal processes, most of the evidence of paleo-coastal plant communities has been obliterated. Likewise, the major Quaternary glaciations also profoundly impacted the depositional landscape, especially in the Mississippi Basin.

The composition of native plant communities of the Southeastern United States has changed less than those of any other region in the country during the last 20,000 years (Delcourt and Delcourt, 1993). This is not to suggest that plant communities in the South have been static over that period. About 18,000 years ago, at the peak of the last major glacial period, the influence of Arctic air masses and boreal vegetation extended to about 33° north latitude—the approximate latitude of Birmingham, Alabama, Atlanta, and Georgia (Delcourt and Delcourt, 1993).

These forests were dominated by various spruce species (*Picea* spp.) and Jack pine (*Pinus banksiana*); fir (*Abies* spp.) was abundant in some locations. The understories of these forests were generally typical of modern spruce-fir forests, with the exception of the absence of certain prairie elements (Wright, 1981). Today, Jack pine is essentially limited to boreal forest types and higher elevations in New England, Wisconsin, Minnesota, and northward. Modern boreal forests dominated by spruce and fir are similarly restricted to New England and Canada.

Temperate deciduous forests dominated the landscape South of 33° north latitude, to about 30° north latitude, including most of the then Gulf Coast from about 84° west longitude. The climate of this region was similar to, or slightly drier than modern conditions, based on the analysis of the species present in pollen profiles collected from lake sediments deposited during this time. Oak (*Quercus* spp.), hickory (*Carya* spp.), chestnut (*Castanea dentata*), and southern pine species were abundant. Walnuts (*Juglans* spp.), beech (*Fagus grandifolia*), sweetgum (*Liquidambar styraciflua*), alder (*Alnus* spp.), birch (*Betula* spp.), tulip tree (*Liriodendron tulipifera*), elms (*Ulmus* spp.), hornbeams (*Carpinus* spp. and *Ostrya* spp.), tilias (*Tilia* spp.), and others that are generally common in modern southern deciduous forests were also common then. Pollen of members of the grass, sedge, and sunflower plant families (Poaceae, Cyperaceae, and Asteraceae) were also common in samples from this time period (Delcourt and Delcourt, 1993; Greller, 1988; Watts, 1980).

The vegetation south of 30° north latitude, in peninsular Florida, was dominated by sand/scrub communities with xeric pine-oak forests in the uplands. Swamps and marshes occupied low-lying and coastal areas (Delcourt and Delcourt, 1993; Greller, 1988; Watts, 1980). The areas that were occupied by coastal marshes at that time are now submerged because sea levels during the time of peak glacial extent were significantly

lower than modern levels. The sand/scrub communities still occupy significant areas of upland central Florida (Ricketts et al., 1999).

During glacial periods, extensive mesophytic forest communities, similar in character and overall composition to modern lowland and bottomland forests, occurred along major river drainages, especially the Mississippi embayment, the Alabama-Coosa-Tallapoosa Basin, the Apalachicola-Chattahoochee-Flint Basin, and the Savannah River Basin (Delcourt, 1993 and Greller, 1988; Delcourt).

From approximately 15,000 years BP to approximately 10,000 years BP there was a gradual warming trend throughout the region, but the period of 14,000 years BP to about 12,000 years BP was marked by a high degree of climatic variability, including increased seasonality and other climatic extremes (Delcourt and Delcourt, 1993). By approximately 10,000 years BP, deciduous forests had expanded northward throughout the region, with pockets of boreal elements remaining only at high elevations in the Appalachian Mountains and in a few other refuges. Broadleaf evergreen and pine forests occupied an area similar in extent to which they occupy today, primarily in the Coastal Plains. Mesophytic and bottomland forest communities continued to occupy the major river drainages of the region (Delcourt and Delcourt, 1993).

Although the exact date is in question, this was also the period in which humans first colonized the Southeast. Archeologists date the earliest potential date of human habitation at approximately 12,500 years BP. Between 12,500 and 10,000 years BP, the human population of the region is thought to have been largely nomadic and very sparsely distributed. Their influence on the region's vegetation was almost certainly trivial and highly localized.

At about this time, many large herbivores that heretofore had been common in the region went extinct (Martin and Klein, 1984). Among these animals were the mastodon, ground sloth, and giant bison. In other parts of the world where large grazing animals still exist, they are known to exert a profound influence on the composition and condition of the native plant communities. Likewise, their extinction would lead to a variety of (largely unpredictable) changes. It is not clear why this guild of plant-eating animals disappeared from the region, but over-exploitation by aboriginal Americans and an inability to adjust to climatic changes are most often posited. It is certain that their disappearance altered regional patterns of vegetation (Martin and Klein, 1984).

At the beginning of the Holocene (10,000 years BP), the climatic conditions in the Southeast were comparable to conditions today (Delcourt and Delcourt, 1993). However, the existence of "modern" climatic conditions does not necessarily imply the existence of modern native plant communities. Although the major modern community types were flourishing in the Southeast by 10,000 years BP, the understory flora had not yet come to resemble modern herbaceous floras. Mixed hardwood forests dominated the majority of upper Coastal Plains, Piedmont, and lower Mountain regions. Southern pine communities dominated the middle and lower Coastal Plains, whereas evergreens and some remnant boreal elements occupied higher elevation sites. Canopy openings in the mixed hardwood and high-elevation forest regions are thought to have been infrequent and due either to local edaphic conditions or natural disturbance (Delcourt and Delcourt, 1993; Watts, 1980).

Evidence of human habitation in the region becomes common at about 10,000 years BP (the Paleo-Indian period), but there is little evidence that these cultures had significant, or large-scale impacts on the landscape (University of Illinois, 1997).

Around 8,700 years BP to approximately 5,000 years BP, a period of significant warming and drying, often called the hypsithermal period, began impacting the vegetation of the Southeast. During the hypsithermal period, extensive expansions of prairies and savannahs occurred throughout the region (Delcourt and Delcourt, 1993) and xeric oak and oak-hickory forest types proliferated. Many species with more northerly affinities migrated northward and, to the extent possible, upward in elevation. Given the limited heights of the Appalachian Mountains, many of these boreal elements were extirpated during this period. Others were relegated to isolated refuges (Delcourt, 1979; Delcourt and Delcourt, 1998). Further retraction of boreal forest elements caused a proportional increase in pine-dominated forests in the Appalachians. The hypsithermal was also responsible for the expansion of sand and scrub habitats in central Florida (Delcourt and Delcourt, 1993; Watts, 1971). The grasslands and savannas of the time expanded and were also linked to the great interior plains grasslands to the west of the region. As a result, elements of the prairie flora became established throughout the region, first by simple migration, but then also by invading disjunct openings (including glades and barrens) that were forming in the canopy of more mesic forests (Delcourt and Delcourt, 1993).

During most of the climatic shifts of the last hundred thousand years, most plant migration in Eastern North America occurred along a more or less north-south axis. The hypsithermal was significant because it made conditions favorable for the invasion and establishment of species from the center of the continent.

With the warming and drying of the climate throughout the region, species with more mesic proclivities retreated to shrinking riparian and riverine areas.

During this period, the population density of aboriginal peoples increased substantially. The hypsithermal also saw the transition from Paleo-Indian to Archaic Indian cultures. During this period, the Archaic Indians' settlements and populations tended to increase in size. Archaic Indians remained; like their Paleo-Indian ancestors, they were largely nomadic but were able to remain in some areas for extended seasons by practicing more concentrated resource usage. Increased resource use was made possible by technological advances that improved the efficiency of the harvest, collection, and processing, for example, native plant materials. More concentrated occupation had significant, but still local, impacts on the abundance and regeneration of tree species (University of Illinois, 1997).

At the end of the hypsithermal interval, about 5,000 years BP, all of the components of the modern southern forests were in place. As the climate cooled and precipitation increased, species migrated so that communities were reassembled in new form. The boreal elements of the early Quaternary enjoyed a modest expansion. Riparian, bottomland, and wetland plant communities expanded. Grasslands and savannahs contracted and retracted westward.

Within approximately 1,000 years of the end of the hypsithermal, the distribution of species within plant communities of the Southeast had more or less stabilized

and would see only minor changes until the colonization by Europeans (Delcourt and Delcourt, 1993).

At about, 4,000 years BP, the Archaic Indian cultures began practicing agriculture throughout the region. Technology had advanced to the point that pottery was becoming common and the small-scale felling of trees became feasible. Some of their crop plants, such as corn and squashes (*Zea mays* and *Cucurbita* spp.), were acquired through trading with cultures from the South that had a longer tradition of agriculture (Delcourt, 1987). Other crop plants were selected from local natives on the basis of desirable cultivation and harvesting traits. This period also saw increasing emphasis on some forms of passive agriculture, in which existing perennial plants were cared for to increase or improve their output of desired products such as beechnuts or cranberries. Concurrently, the Archaic Indians began using fire in a widespread manner in large portions of region. Intentional burning of vegetation was taken up to mimic the effects of natural fires that tended to clear forest understories, thereby making travel easier, and facilitating the growth of herbs and berry producing plants that were important for both food and medicines.

Approximately concurrent with the transition from the Archaic Indian culture to the Woodland Indian culture, around 2,800–2,500 years BP, aboriginal groups began to establish relatively large settlements. People from these settlements visited sites to exploit specialized resources such as fish, medicinal plants, and cherts. There was a trend, however, toward more permanent occupations to maintain local agricultural plots (University of Illinois, 1997). It was during this time that Mound cultures began to develop and flourish. Woodland Indian Culture evolved into the Mississippian Indian Culture in large portions of the region approximately 1,000 years BP (University of Illinois, 1997). Mississippian Culture agriculture became more highly developed and villages, both large and small, were able to support a more specialized citizenry (Delcourt 1987). Mounds became larger and more numerous, and the amount of land needed to support these populations increased. The majority of Mississippian Culture sites are associated with wetland, riparian, or riverine habitats, and these people became quite expert at altering local hydrological patterns to keep their villages dry and their fields irrigated, and to supply community water needs. In some places, soil erosion became locally significant.

Indian use of fire in land management continued from approximately 4,000 years BP to approximately 500 or 600 years BP (Adams, 1992; Cowell, 1998; Delcourt and Delcourt, 1997). This practice significantly affected the structure of forest stands and the relative abundance of species over large portions of the region. It is not clear to what extent fire influenced the composition or richness of regional floras.

For reasons that are unclear, approximately 500 years ago, aboriginal populations declined significantly throughout Eastern North America, and more broadly throughout the Americas. Most anthropologists attribute this depopulation to the transmission and spread of pathogens brought to North America by Europeans. Some communities are known to have lost 98% of their population, though in general it seems that approximately two-thirds of the Indian population of the Eastern United States was eliminated in a very short time. As a consequence, large areas that had been cleared,

burned, and farmed by native peoples were left fallow. Thus, by the time the first European observers were reporting the nature of the vegetation of the region, it is likely to have changed significantly since the regional peak of Indian influence.

A myth has developed that prior to European culture the New World was a pristine wilderness. In fact, the vegetation conditions that the European settlers observed were changing rapidly because of aboriginal depopulation. As a result, canopy closure and forest tree density were increasing throughout the region.

When Europeans started making regular visits to the New World approximately 500 years BP, and during subsequent colonization (specifically in Florida, but also shortly afterwards northward along the Atlantic coast), they also began introducing Eurasian and nonnative tropical plant species. Exotic plants first became prevalent around permanent settlements, especially along the coasts, and then spread inland along travel routes to other suitable locations.

The earliest exotic plants to become established in the region came originally as packing material (often rough hay) in shipping crates or animal bedding material. Later, food, forage, and medicinal plants were introduced in support of the settlements (Carrier, 1923). The introduction of exotic animals (especially hogs, cattle, and rats) also began at this time. These animals, also have had a significant and permanent impact on the vegetation of the region.

In June 1527, a group of Spaniards, including Cabeza de Vaca, began a 10-year expedition from Florida, along the Gulf Coast into Texas, and on into the American Southwest (Cabeza de Vaca, 1542). In his account of the journey, Cabeza de Vaca reported that: (1) the natives of Florida cultivated large quantities of corn; (2) palmetto was abundant and was used commonly for food, fiber, and fuel; (3) extensive areas of heavy timber (almost certainly longleaf pine) were present with considerable large woody debris on the ground. The chronicles of other early Spanish explorers, such as Hernando de Soto and Ponce de Leon, contain similarly superficial accounts of the existing native vegetation. The first really useful and widely available information on the natural vegetation of the Southeast was not published until more than 200 years after the Spanish exploration of the region.

Southern Native Plant Communities in Historical Times

Information about the historic native plant communities of the region can be difficult to interpret. Since the modern concept of a plant community did not evolve until the late 19th and early 20th century, earlier writers seldom included the kind of information we would like to have for this assessment. Also, most common paleobotany methods have limited value in the study of historic vegetation because they have poor resolving capabilities over the relatively short period of the last 500 years. These difficulties aside, there is currently a great deal of interest in the nature of native plant communities at the time of European settlement, largely motivated by the current trend toward the restoration of a variety of native southern plant communities.

Although Europeans began to explore and settle the Southeast by the mid- and late 16th century, their impact on the native plant communities of the region was limited largely to Coastal Plain, savannah, and bottomland forests. For the most part, the earliest

settlements were established in coastal areas and on broad river terraces accessible by boat and barge. Even the rare interior settlements, such as the Arkansas Post established in 1686, were built along major rivers to avail themselves of local patterns of commerce. These areas were often cleared to make way for agriculture. Some of the clearings were made for subsistence farming, but the largest were made for commercial farming and livestock production. The quantity of timber taken during this time was limited both by technology and local demand. Consequently, large areas of upland forest in the South went essentially untouched until the 19th century.

The exploitation of natural resources, such as timber and forage, increased as population increased and as an industrial base was built in North America. Improved agricultural efficiency, a growing population, and better access to European markets by the end of the 18th century provided both the motivation and the capital necessary to expand the conversion of native vegetation to agriculture (Carrier, 1923). People began to move westward into the interior of the region and began to clear increasingly large tracts of land. In this era of increased trade, additional exotic species were introduced to the South and exotic plants that had become well established moved with the expanding population.

Although the Native American population had declined significantly, these people were sufficiently common in the early 18th century to exert a continued impact on wide areas of the southern landscape through their agriculture and, more importantly, their use of fire as a means of manipulating vegetation. The aboriginal practice of burning the forests was adopted by European settlers soon after permanent settlements were established.

Like the Indians, the European settlers of the interior South tended to choose specific areas in which to build homes and farms. Relatively flat topography, access to water and timber, and proximity to trade routes via waterways or overland were important criteria for settlement sites. Such places are most typically found either along the terraces of large river systems or on the Coastal Plain. Consequently, riverine forest communities and longleaf pine communities were the first natural vegetation types in the interior South to be impacted by the expansion of European settlement. However, these native plant communities had long been inhabited by aboriginal people. In some cases the Europeans removed the Indians by force so that they could occupy their land. Europeans selected and exploited other areas on the basis of their strategic value for military outposts or their proximity to mineral resources. These areas were less common, but usually had equally significant impacts on the local vegetation.

Until the 20th century, the economy of the South was based largely on agriculture. Technology changed the kinds of crops grown, especially for the export market. From the late 18th century until the early 20th century, resin extraction from pines, especially longleaf pine, for use by American and European navies shaped the management of longleaf pine forests in the Coastal Plains. The naval stores industry began to decline with the development of metal hull ships at the end of the 19th century. Large, extensive farms became common in the region by the early 19th century, due in great part to technological improvements like the invention of the cotton gin in 1793. Until the beginning of the 19th century, tobacco accounted for the majority of southern exports;

thereafter and well into the 20th century, mechanized cotton production dominated the South. Large tracts of agricultural land were created out of the native plant communities of the Coastal Plain where cultivation was relatively easy. This form of land use also greatly affected longleaf pine communities, as well as a wide range of hardwood communities that existed on river terraces.

Increases in farm size had the effect of concentrating economic power in the hands of relatively few established families and companies. There was little incentive for these families to develop new centers of agriculture or diversify the crops being grown. The majority of new settlements in the interior South were based either on a subsistence economy or service to relatively small areas. Certain areas were completely converted to agriculture, with permanent and deleterious implications to the native plant communities. In areas dominated by subsistence farming, less obvious impacts to the native plant communities occurred, such as the disruption of population processes caused by fragmentation, the introduction of exotic species, impacts to rare communities such as mountain bogs and glades, and widespread alterations in forest community structure related to timber harvesting and fuel-wood gathering.

There was considerable curiosity in 17th and 18th century Europe about North American ornamental and medicinal plants. In fact, most of the “botanists” of this time were collectors for wealthy Europeans. These botanists, however, usually did not catalog the natural resources of the region. It was left to the early 18th century botanists from the Northeast to first explore and describe the vegetation of the Southeast. Most notable among these early explorers were John (1699–1777) and William Bartram (1739–1823).

The Bartrams made several journeys of botanical exploration and collection and published accounts of the natural history of the areas that they visited. William Bartram’s “Travels through North and South Carolina, Georgia, East and West Florida...” became an international best seller shortly after being published in 1791. This success was no doubt due in part to John Bartram’s reputation and to his and William’s extensive correspondence with European botanists. William Bartram states that the purpose of his trip through the South was the “discovery of rare and useful products of nature, chiefly in the vegetable kingdom,” and to “obtain specimens and seeds of some curious trees and shrubs (which were the principal objects of this excursion).”

Although “*Travels*” is full of details of soil conditions in various places, lists of species encountered, and in some cases detailed descriptions of particular species, Bartram did not generally offer useful accounts of the native plant communities. He did recount the occurrence of many of the broad community types we are familiar with, including various forests, savannas, glades, and swamps, but they are usually described in a fashion such as:

“...expansive green meadows or savannas, in which are to be seen glittering ponds of water, surrounded at a great distance, by high open pine forests and hommocks, and islets of oaks and bays projecting into the savannas...”

He noted large areas of clearcut longleaf pine (*Travels*, p. 312) and “expansive ancient Indian fields” (*Travels*, p. 458). Bartram was particularly interested in the agricultural potential of the South, noting not only the areas used by the aborigines for

cropping (*Travels*, p. 511), but also areas that would be suitable for the cultivation of European crops as diverse as olives and oranges (*Travels*, p. 337). He also documents the early trade in useful native plants such as ginseng (*Travels*, p. 327) and rosinweed (*Silphium*, *Travels*, p. 398). Bartram also offers accounts of introduced species such as barnyard grass (*Echinochloa*, *Travels*, p. 430) as well as a description of Franklin tree, (*Franklinia altamaha*, *Travels*, p. 467) a species that is now extinct in the wild. Perhaps most remarkable about the landscapes described by Bartram is that many of these places remained unchanged until the late 19th century.

Thomas Nuttall, traveling in the Arkansas Territory around 1819 (Nuttall, 1821), also described what he saw in general terms; thickets of dwarf oaks, hills of pine and oak, and scattered areas of prairie. He too noted the hand of man on the landscape, mentioning annual fires set by the white settlers and extensive areas of cutover pine. Nuttall cataloged many non-woody plants as well. As was customary at the time, he did not elaborate about the specific conditions in which these plants were growing, but simply stated this or that species was growing under oaks, along streams, or high upon a hill.

Bartram and Nuttall are the most important of the early botanical explorers of the South, but their work is of limited value in determining the nature of native plant communities in existence at the time. Their approach reflected the philosophy of natural history and botany at the time. At the cusp of the 19th century, ecology was not yet a word, much less a science. Linneaus had developed his natural classification system only a half century prior, there was not yet a concept of natural selection or evolution, and it was a time of global exploration and discovery. All of the major seafaring European nations were establishing colonies around the world. The purpose of this exploration was the acquisition of power and wealth, and because many plants were the source or great wealth there developed a need for botanists to travel to “unexplored” parts of the world to catalog the plant life. At the time, this was called phytogeography, a term that describes the endeavor well enough. The primary concern of phytogeographers was to identify the location and distribution of plant species. While phytogeography was a necessary step in the development of plant ecology, at the beginning of the 19th century little effort was being expended to describing the interrelations among the species that they were so faithfully cataloging.

After Bartram and Nuttall, a procession of botanists and naturalists, often physicians with an interest in botany, collected plants in the areas around their homes. For the most part, these collectors did not directly contribute to the understanding of the distribution of native plant communities. However, their work would become important later, in the late 19th and early 20th centuries, as regional floras for the South were developed.

In 1835, the first railroad system in the South began operating in North Carolina, in the heart of the longleaf pine forests of the Coastal Plain (Croker, 1987). The Industrial Revolution had brought to the South the means by which its abundant forest resources could be transported great distances and still turn a tidy profit. The longleaf pine forests of the Coastal Plains were not only a source of high-quality timber for a growing population, but also the nation’s most important source of “naval stores.” Naval stores,

the processed or unprocessed resin, or tar, made from the sap of longleaf pine was used to seal the hulls of ships and many other things. The naval stores industry began in North Carolina, but later spread throughout the Coastal Plains with the railroad (Croker, 1987). By 1854, the railways had reached the Mississippi River.

In the mid-19th century, clearcutting was the primary logging method employed. Modern forestry, as practiced in Europe at the time, would not become commonplace in North America until the early 20th century. In the first half of the 19th century, extensive areas of forest were leveled to create pastureland. In many places the native forest has never recovered. Forested areas surrounding major river ports were extensively cut to fuel steamboats. Vast acreages of wetlands and river terraces were drained or plowed by the mid-19th century, causing significant losses to local biodiversity in some areas. Strip mining, especially for coal to stoke hungry steamboats and railroad locomotives, became commonplace where deposits were sufficiently shallow to exploit (such as the Upper Cumberland Plateau). Strip mining eliminating forest cover and frequently altered or killed riparian and aquatic plant and animal communities downstream from the spoil piles. Although much of this activity in the region slowed during the 1860s, logging resurged quickly thereafter. By the 1880s, a broad sector of Americans, mostly in the Northeast and West, were becoming concerned about the unbridled exploitation of the nation's forest and wetland resources.

The evolution of forest protection laws, and the establishment of national forests in the South parallel the development of the modern conservation movement in the United States (Williams, 2000). Issues such as farmland erosion, forest clearcutting, and the hyper-exploitation of buffalo were on the national conscience. The first use of the word conservation in the context of the protection of natural resources was in 1875, by John Warder, president of the American Forestry Association. The leadership of America's conservation movement was borne by Gifford Pinchot, John Muir, Charles Sargent, and Theodore Roosevelt.

The Federal Government began setting aside tracts of land as "forest reserves" when Congress passed the Forest Reserve Act of 1891 (Williams, 2000). This legislation allowed the President to "from time to time, set apart and reserve, in any state or territory having public land bearing forests, in any part of the public lands, wholly or in part covered with timber or undergrowth, whether commercially valuable or not, as public reservations..." Federal forest administration was consolidated under the leadership of Gifford Pinchot in 1905 with the establishment of the U.S. Department of Agriculture's Forest Service (Williams, 2000). The first national forest established in the South was the Arkansas National Forest (1907). Two national forests in Florida were added to the growing system in 1908 (Ocala and Choctawhatchee). Most of the national forests throughout the South are a result of the Weeks Act of 1911. This Act broadened the mandate of the Forest Service and provided for the purchase of land, largely for watershed protection. From the time of their establishment until the beginning of the Second World War, the national forests of the South served primarily as conservation areas (Williams, 2000). National forest lands have since been critical refuges of functional native plant communities in the South.

At the turn of the 20th century, the logging industry in the South was producing lumber at its historic peak. So much forest land had been logged out that timber companies were finding it difficult to access merchantable trees and were beginning to close mills and move to the newly opened virgin timberlands of the Northwest. Although the First World War caused a short-lived resurgence in the demand for timber and naval stores, the conversion of the shipbuilding industry to steel by 1920 caused demand for southern timber and naval stores to fall drastically. By 1930 the majority of the Coastal Plains longleaf pine communities had been essentially cut over (Croker, 1987), as had the interior shortleaf pines (*Pinus echinatus*). Upland hardwood forests fared somewhat better, at least in some places.

After 300 years of land conversion and alien plant introduction, it is no surprise that in the early part of the 20th century exotic plant species were common throughout the region. Some had been planted purposefully as ornamentals, as forage for livestock, or increasingly as erosion control agents by State and Federal agencies. Others were simply accidental tourists that made their way across the region without the direct assistance of people, such as in stocks of hay or the coats of domestic animals. Palmer (1926) notes an abundance of “introduced species [and] adventive woody species” in the vicinity of Hot Springs, Arkansas. He specifically noted Japanese honeysuckle (*Lonicera japonica*), Princess tree (*Paulownia tomentosa*), and many other introduced species.

Vascular plants were not the only exotic species to be introduced to the United States during historic times. Among the most destructive exotics were fungal pathogens of trees. Chestnut blight (*Cryphonectria parasitica*) was introduced into this country in New York in 1904. It spread rapidly and was actively killing trees in Southern Appalachians by the 1920s. By the early 1950s, American chestnut (*Castanea dentata*) was ecologically extinct throughout its range in Eastern America. This tree was once a dominant member of Appalachian forests. In some areas, one tree of every four was a chestnut. Although the loss of the chestnut was significant in terms of the resultant change in forest composition, there is some disagreement as to the ecological impact of chestnut blight. Only one extinction is suspected as a result of the blight (American chestnut moth, *Ectodemia castaneae*) and the greatest impacts to the native plant communities seem to have been a change in tree density (a temporary result of the many canopy gaps created by the death of the chestnuts), and a realignment of dominant overstory tree species as a result of competition among species (Woods and Shanks, 1959; Stein et al., 2000). Different tree species have replaced the chestnut as the dominant canopy species in different portions of the chestnut’s former range.

Dutch elm disease (*Ophiostoma ulmi*, and *O. novaeulmi*) entered the United States in 1930 in logs imported from Europe. There is differential susceptibility among *Ulmus* species, but the American elm, a common street and landscaping tree, has been the hardest hit. By the late 1970’s Dutch elm disease was known to have impacted elm trees throughout the country (Schlarbaum, 1997).

Butternut canker (*Sirococcus calvigeni-juglanacearum*), which impacts *Juglans cinerea*, was first observed in the United States in 1967, but it is believed to have been infecting trees for many years by that time. By 1995, the USDA Forest Service

estimated that over three quarters of all butternut trees had perished from the disease (Schlarbaum, 1997).

There have been many other exotic disease-causing fungi and insects that have had significant impacts on the native plant communities of the South. Examples include white pine blister rust (*Cronartium ribicola*), the gypsy moth (*Lymantria dispar*), and the balsam wooly adelgid (*Adelges piceae*). Many introduced disease organisms are still impacting our native plant communities, and it is likely that new pests will be periodically introduced to our region. No one can tell what damage they might bring in the future.

The study of the flora of the South was in some respects dependent on the publication of local and regional floras. Improvements in the knowledge of the botany of the region required these tools. Several local floras had been published for portions of the South, including Walter's *Flora Caroliniana* (1788), Mohr's *Flora of Alabama* (1901), and Gattinger's *Flora of Tennessee* (1901). The first comprehensive flora of the Southeast was published in 1860 by Chapman. It was an important, though incomplete work. Unfortunately, it seemed to stifle further serious assessments of the local flora of the region until the early 20th century. It was not until 1903, with the publication of Small's *Manual of the Southern Flora*, that the region had a comprehensive, systematic flora. Revised in 1933, Small's *Manual* is a monumental work of 1,500 pages and was the standard of southern botany floras for over 50 years (Reveal and Pringle, 1993). The last 20 years have seen the development of several important new floras (Smith, 1994; Wunderlin and Hansen 2000).

The lack of specific information about native plant communities in the South from settlement times to the end of the 19th century is the product of two conspiring circumstances. First and foremost, the Southeast has been continuously occupied for longer than any other region of the United States so that by the early 19th century when the nation became interested in its natural resources, the focus was on the wild and unknown West rather than the familiar South.

Secondly, the development of plant ecology as a modern science took place largely in Europe beginning in the early and mid-19th century. There and then the concepts of succession and plant associations were first developed into forms recognizable today. However, at the time, the study of plant ecology was a subdiscipline of plant geography. Plant geography, the description of the distribution of plants, was the primary concern of European academics, capitalists, and naturalists. In the 19th century, naturalists from many nations were traveling around North America cataloging plants. The pinnacle of plant geography studies was reached in the early 20th century and coincided with the rise of the modern study of plant ecology. The earliest focus of the fledgling field of ecology was the study of plant community succession. That research was done in the Midwestern plains and Eastern forests.

Henry Cowles first described the dynamic (changing) nature of vegetation. Prior to Cowles, the plant geographers were content to map the current condition and extent of vegetation. Many of Cowles' students went on to make important contributions to the study of succession throughout North America. E. Lucy Braun became renowned

for her descriptions of virgin forests in the Eastern States, especially the Appalachian Mountains. Her work is still read and used as a reference.

Fredrick Clements was arguably the first “community” ecologist in America. Working largely with prairie and old-field communities in the Midwest, Clements described much of the vegetation of North America, named many plant associations and identified successional stages for his named communities. He described the plant community as a form of “super-organism” in order to highlight his perception of the interdependence of all of the parts of a community and described succession as of the development or life cycle of an organism.

Clements notion of the “super-organism” was not universally accepted. In 1926, Henry Gleason, who conducted his research in forested communities similar to those common throughout the South, wrote an influential paper that criticized Clements views and posited that the nature of plant associations is determined by the individualistic behavior of plant species. Gleason’s “individualistic” notion of plant communities eventually won out over Clement’s idea of the super-organism.

The complexity of Southern forest plant communities hampered the development of a comprehensive and consistent community classification system, such as those developed early in the history of land management in the Midwest and West.

Beginning with the study of plant succession in the first quarter of the 20th century, a practical science of plant and community ecology evolved. From this point forward meaningful data became available about the nature of native plant communities. However, because the South had been settled for centuries, by the early 20th century, vast tracts of native plant communities had been converted, planted, logged over, infested with weeds, or otherwise impacted so opportunities to study intact native communities were rare.

The Great Depression of the early 1930’s was exceptionally difficult for the people of the South, but it did a lot for the native plant communities of the region. The Federal Government purchased land and established many national forests. The Civilian Conservation Corps, established in 1933 during the Franklin Roosevelt administration, did extensive reforestation in the South. The formal teaching of forest sciences in the United States had finally matured by the 1920s and 1930s so that an abundance of well-trained foresters working for the USDA Forest Service, State forestry agencies, and the CCC itself, were available to supervise and direct the work (Williams, 2000). The fledgling Forest Service was working to control unauthorized timber cutting on Federal land. Unfortunately, this was also the time in which wide spread fire suppression activities began. Although this practice was well intentioned at the time, it eventually led to significant declines in native plant communities throughout most of the Southeast.

The timber industry in the South remained depressed until the outbreak of the Second World War. At about the same time, serious scientific research was started at government and university labs to increase the productivity of forest land. Much of this work focused on the development of “improved” tree selections and cultivation practices. One of the innovations that arose was the growing of pines in plantations.

Plantation cultivation of pines turned out to be an exceptionally productive way to produce trees. Newly developed tree selections thrived in the prepared conditions of the plantation. Large tracts of cutover land, especially in the Coastal Plain and Piedmont, would eventually be converted to pine plantations. This methodology focused timber production on developed sites. Although those sites were forever altered, this intensive form of silviculture saved many acres of native forest from more traditional timber harvesting.

The next large threat to native plant communities in the South came from another, unlikely, advancement in technology. From the time of settlement the South was largely rural, agrarian, and sparsely populated. The widespread availability of air conditioning in the 1950s and 1960s made living and conducting business much easier in the sweltering heat of southern summers. The South, therefore, began to see significant increases in immigration and urbanization. Land was developed and large tracts were fragmented. These trends led to rapid increases in demand for building materials, electricity, and additional agricultural production.

Improvements in technology and mechanization (especially in agriculture), and decreasing Federal commodity price supports led to significant consolidations in the timber and farm industries. Former farmers migrated to cities in the North and South. In the 1940s, 42% of the population in the South lived on farms. By the 1950s, only 15% of Southerners lived on farms. The majority of the population of the region became isolated from the landscape and forever changed the way Southerners viewed their forests.

After the end of the Second World War, pine forests in the South, including those on State and Federal land, were predominantly managed for timber production. The birth of the modern conservation movement in the 1960s came, in part, as a reaction to concerns about public land management priorities and the lax enforcement of environmental laws.

The Current Condition of Native Plant Communities in the South

Ecosystems—interacting aggregations of plant and animal communities and the abiotic factors affecting them—in the Southeastern United States are as diverse as any in the world. No place in North America has more diverse forests in terms of plants or animals, or more different types of forests. One very important source of this diversity in plant communities in the Southeast is the exceptionally high degree of endemism (occurrence restricted to a particular region or area) in the regional flora, especially in Coastal Plain conifer forests and in Appalachian forests.

In contrast, the South has the greatest absolute number of introduced plant species in North America. Florida alone reports 800 introduced species existing outside of cultivation (FLEPPC, 2001).

One of the most important tools in the study of any system, including plant communities, is a comprehensive means of classification of the observed diversity. Several large-scale vegetation classification methods are in current use; the most important are those described by Kuchler (1985), Bailey (1994, 1998), and The Nature Conservancy (1999). Each of these systems divides the region on the basis of either general

physiography or potential natural vegetation. Although many other methods exist, these methods illustrate the basic philosophies of large-scale vegetation classification. Although most vegetation classification systems are in agreement on the general distribution of regional plant communities, there is still much discussion and continuing research concerning how to define the transitions between vegetative communities.

Small-scale community classification can be generally useful in understanding the dynamics of local vegetation. Hierarchical and geographically comprehensive systems such as The Nature Conservancy's (TNC's) National Vegetation Classification System (Anderson et al., 1998; Grossman et al., 1998) define literally thousands of plant associations based on the presence of "dominant" and associated species. The utility of this system (and similar systems) is its inherent flexibility.

One of the most useful qualities of the TNC National Classification System is the assignment of rarity ranks to plant communities (ABI, 2001). A comprehensive system of rarity ranks across the nation allows for an assessment of the geography of community diversity.

According to TNC figures, the Southeastern United States has the highest number of endangered ecosystems of any region of the country (Stein et al., 2000). According to TNC, more than 30% of all natural plant communities throughout the Southeast are critically endangered and the Southeast has the highest proportion of imperiled plant communities in the United States, exclusive of Hawaii (Stein et al., 2000). A great number of the rare plant communities in the Southeast are inherently rare and their rarity is a function of the great plant diversity in the region. However, the majority of rare communities in the Southeast are rare because of habitat alteration or degradation.

The majority of inherently rare plant communities are relatively small patches of plants in unique combinations, often due to the presence of equally rare edaphic conditions. These "patch" communities can be seen as occurring within a matrix of more common, widespread community types. Most habitat conservation activities tend to focus on the "patch" habitats.

Because there has not been a single consistent convention for the identity of plant communities during the majority of the history of the Southeast, it is essentially impossible to discuss the specific changes to those plant communities over time. However, this is not to say that we cannot assess the overall trends in conditions of plant communities. On the basis of conversion, alteration, and impedance of function, more than 99% of all plant communities in the South are not in the condition they were in prior to European settlement. Some of these changes have been subtle, but most are readily distinguishable. It is impossible from the perspective of current times to precisely estimate what has been lost, but we can understand the degree of loss sustained by southern native plant communities.

Among the communities to have seen the greatest change in historic times are the region's forests. All of the forests of the South have been touched, directly or indirectly, at one time or another by the hand of humanity. Sometimes that hand has been gentle, but in most cases it has not.

By some estimates, all of the upland hardwood forests of the Appalachians have been altered. The hardwood forests have suffered from chestnut blight, Dutch elm

disease, and butternut canker. Even if the impact of disease is discounted, less than 10% of the original, native forest area of the region has not been eliminated or altered. Most was cleared prior to the 1930s. Estimates vary from state to state but, on average, approximately half of all pre-settlement hardwood forest has been eliminated (Walker and Oswald, 1999), and the majority (essentially all) of what remains is compromised by fragmentation, exotic pest, and disease organisms, and altered natural processes such as fire and livestock grazing (Mac et al., 1998; Noss et al., 1995).

Coastal Plains longleaf pine forests, renowned for their high levels of diversity, endemism, and species rarity, have been reduced by more than 98%, compared to pre-settlement conditions. Most have been converted to agriculture or pine plantations, (two plant communities renowned for their lack of diversity, endemism, and species rarity.) Most of the longleaf pine forests were cut by the 1920s, but longleaf pine habitat was still being clearcut and converted into plantations in the 1980s (Noss et al., 1995; Stein et al., 2000). They have been used as a source of timber since aboriginal times, but European settlers were clearcutting vast areas of longleaf pine by the mid 18th century. Longleaf that was not cut for lumber was commonly used as a source of naval stores beginning in the 17th century, a practice that continued into the early 20th century (Croker, 1987). The remaining large blocks of longleaf exist almost exclusively in public forests (notable privately owned large tracts of longleaf include the Moody tract in Southern Georgia and Green Swamp in North Carolina). Many areas of longleaf forests are being managed for the endangered red-cockaded woodpecker. Remaining blocks are, in some places, threatened by exotic plant species, such as *Imperata cylindrica* (Cogon grass), fire suppression, and some forestry (site preparation) practices that disturb the forest understory plants (in lieu of burning) to facilitate the growth of the trees. There is also much concern, but little that can actually be done, about the fragmentation of the original longleaf community (Croker, 1987). Only minor fragmentation agents, such as roads, can be managed to increase longleaf habitat continuity, whereas the major fragmentation factors—conversion to agricultural and urban land uses—are essentially intractable. Many public land management agencies are currently practicing longleaf forest restoration activities, and others are encouraging restoration on private land. These efforts, while very important, vary greatly in their success. While it is relatively simple to successfully grow longleaf pine, the reconstitution of the original plant community is very difficult.

Fewer than 50% of the pre-settlement spruce-fir forests still exist in the Appalachians (Noss et al., 1995). Of that quantity, more than 98% have been either altered or are under attack by introduced pests. Over 90% of the red spruce forests in Central Appalachian forests have been lost (Noss et al., 1995).

Approximately 90% of the forested habitats in Florida have been altered or eliminated, including 60–75% of the forested uplands of Lake Wales Ridge, an area of exceptionally high species rarity and endemism. Only on the Atlantic and Gulf coastal barrier islands does a majority of the natural forest cover remain. It has survived due to its isolation and unsuitability for agriculture or development (Noss et al., 1995; Stein et al., 2000).

More than 98% of the pre-settlement old-growth forests in the South have been altered or lost (Stein et al., 2000). The vast majority of the remaining old-growth forests in the South are on Federal land in national forests and national parks. Of the original 60–90 million acres of Coastal Plain pinelands, only 3% survive today as old growth (Croker 1987; Noss et al., 1995; Walker and Oswald, 1999). Less than 2% of the forests in Kentucky have old growth characteristics (Noss et al., 1995). In Tennessee, only about 5% of the pre-settlement old-growth forest on the Cumberland Plateau remains and no more than 20% of the forest of Tennessee's Blue Ridge Province can be classified as old growth (Noss et al., 1995). Those few tracts of old growth not on public land are mostly in fragments of 100 acres or less, which hamper their value (Stein et al., 2000). Most of the forest types classified as old growth today are actually second or third growth forests that have or are developing the structural characteristics of old growth.

Open habitats in the South such as glades, barrens, and prairies were common at the time of European settlement as noted by the earliest travelers to the region. There are, however, no good estimates of how much of the landscape was occupied by these open areas. The current best approximation suggests that as much as 10% of the plant communities of the South were historically open habitats (Mac et al., 1998). Today, approximately 1% of the forested landscape of the South is occupied by openings such as barrens, prairies, and glades. In most cases these areas are very small, and they are not integrated across the landscape (Mac et al., 1998; Stein et al., 2000) as they once were.

Among open habitat types, prairies seem to have suffered the greatest losses. Settlers saw these relatively flat, treeless, and fertile areas as productive and easy to clear. In Kentucky, less than 200 acres of an original 3 million acres of native prairie remain (Noss et al., 1995). In Texas, Louisiana, Florida, Mississippi, and Arkansas nearly 99% of acres originally in prairie types have been lost (Noss et al., 1995).

The majority of glades that survive today tend to occur in mountainous regions that were never converted to agriculture and they typically have very stony soil. There is no information on the total area in glades throughout the region, but estimates are that less than half of the original glade habitat in the region survives intact, and the majority of that which remains is ecologically compromised due to either the presence of exotic species or the lack of fire. In Tennessee, approximately half of all the area in cedar glades has been converted (Noss et al., 1995). Limestone glades throughout the region have been disturbed at higher rates (Noss et al., 1995), probably because they are more commonly located at lower elevations and in areas of gentler topography.

High-elevation grassy balds are mountaintop treeless areas. Although the mountains on which these open areas occur are not high enough to have an alpine plant community, various edaphic and historic circumstances have conspired to keep these areas treeless. Grassy balds tend to support herb-rich, communities that require frequent disturbance (Greller, 1988). Their ecological origin is still a matter of debate. About 50% of the area that was occupied by grassy balds in 1900 remains today (Mac et al., 1998).

Almost all of the wet, hardwood forests, such as those that occur in bottomlands and hammocks on the tropical Coastal Plain have declined to approximately 20% of their pre-settlement cover (Noss et al., 1995; Mac et al., 1998). A slightly larger percentage of the original floodplain forests have survived (Noss et al., 1995), but most of it was cleared at some time in the past and has returned to forested cover in the last century. In the last 25 years, accelerated efforts have been made to restore floodplain forest, especially in the Mississippi Valley.

The Southeast comprises only 16% of the land area of the lower 48 United States, but it contains 36% of all wetlands and 65% of forested wetlands. About 78% of all wetlands in the Southeast has been altered to some degree (Noss et al., 1995).

Unique or isolated wetlands have fared worse overall. Although the Southeastern United States has the highest diversity of carnivorous plants in the world, the habitat in which these plants occur has declined by approximately 97%. Reed wetlands, known as canebrakes, have been reduced by more than 98% (Mac et al., 1998). Mountain bogs, especially those in the Southern Appalachians and Blue Ridge, are home to a great variety of unique native plant species. Although approximately 10% of these bogs remain, few are in fully functioning ecological condition (Mac et al., 1998).

Pocosins, upland wetland that occur on the Coastal Plain, have been reduced to about 20% of their original area (Mac et al., 1998; Noss et al., 1995). Similarly, only about 10% of the original Atlantic white cedar forests, which require frequent, low-intensity fires and are typically only seasonally wet, are left (Noss et al., 1995).

In the early 1600s there were approximately 220 million acres of wetlands in the lower 48 states (Mitch and Gosselink, 1993). Nationwide, over half of wetland acres have been converted to other uses. The degree of wetland loss has been less on the Coastal Plains, thanks in part to restoration and conservation activities that began in the 20th century. Today, only 28 of Coastal Plain wetlands have been permanently converted (Noss et al., 1995), but a significantly higher proportion have been impacted by human management and exotic plant species

The degree of loss of wetlands varies widely among States within the South (Table 2), and is complicated by the large-scale alterations of wetlands and hydrology conducted by humans.

Table 2. Thousands of acres of timberland in southern states, by ownership class.

State	Percent loss	State	Percent loss
Alabama	50	Mississippi	59
Arkansas	72	North Carolina	49
Florida	46	Oklahoma	67
Georgia	23	South Carolina	27
Kentucky	81	Tennessee	59
Louisiana	46	Texas	52
		Virginia	42

Countless acres of wetland have been drained either for agriculture, pasture, or urbanization, and countless other acres were lost during stream channelization, diking, or deforestation (Mac et al., 1998; Mitch, Gosselink, 1993 and Noss et al., 1995). The rate of wetland conversion was greatest (Mitch and Gosselink, 1993) from the 1950s through the mid 1970s. Since the 1970s the States with the greatest rate of wetland loss nationwide are all in the South: Arkansas, Florida, Mississippi, North Carolina, and South Carolina (Mitch and Gosselink, 1993).

The condition of the native plant communities discussed above is reflective of the condition of the majority of native plant communities in the South. In fact, it is exceptionally rare to find “pristine” plant communities. Even the most remote places have been affected by invasive exotic plants, introduced disease organisms, changes in community structure and function stemming from altered fire and hydrological regimes, and even changes in the local seed and pollen dispersing animals.

Rare Plant Species in the Southern Region

Plant communities, whether rare or common, are comprised of species that share similar ecological needs and tolerances. The diversity of plant species in the South is rivaled in North America only by the California flora. This diversity is due in part to a broad array of species that are either highly localized in their distribution or are very sparsely distributed over large areas.

Two widely accepted determinants of rarity in plant species is protection under the Endangered Species Act of 1973 (ESA) or the determination of imperiled status by TNC (ABI, 2001).

Within the assessment area, approximately 115 plant species are listed as either threatened or endangered under the ESA (USDI, 2001). Of this number, 52 occur in Florida. Those species are clustered in the Appalachicola and Lake Wales Ridge areas. The Southern Appalachians contain the next greatest concentration of threatened and endangered plant species.

Figures 1 and 2 show the distribution of rare plant taxa in the South by equal-area hexagons and Counties, respectively. These maps were derived from data held by State Heritage Programs and represent the occurrences of vascular plant species with a TNC rarity rank of G1-G2. These are species considered to be critically imperiled (Stein et al., 2000) based on the number, size, and condition of populations known to exist. The distribution of rare taxa is used here as a proxy for the distribution of plant diversity. Low diversity plant communities such as agricultural lands or beaches rarely contain uncommon taxa, whereas there is a worldwide pattern of uncommon species being associated with highly diverse plant communities. The occurrence data represented in Figures 1 and 2 should not be interpreted as the distribution of plant species on a trajectory toward extinction. Most of the rare plants in the South (or the world for that matter) are species that are naturally rare (Rabinowitz, 1981). These data are, in all likelihood, incomplete in that private lands may be under-surveyed for rare plants and some states have generally better surveys than others. However, Figures 1 and 2 represent the best available data at this time and are more than adequate to elucidate the overall pattern of species diversity and rarity in the South.

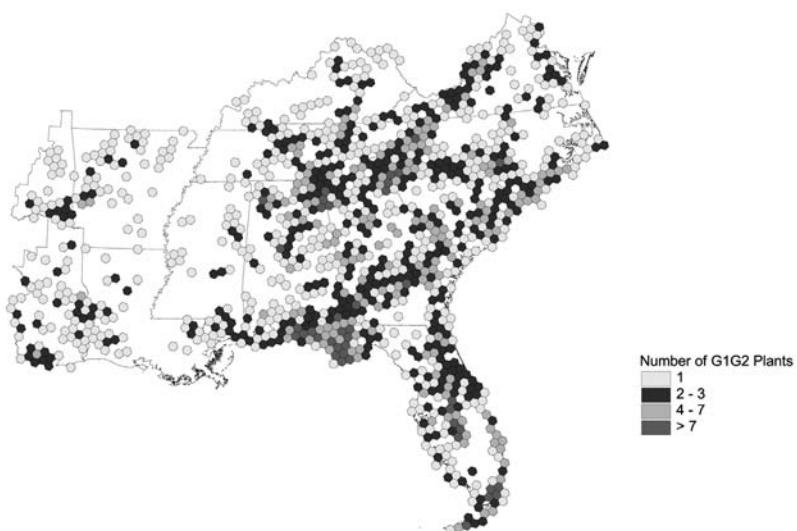


Figure 1. Distribution of imperiled vascular plant species in the South based on the number of occurrences in equal-area hexagons.

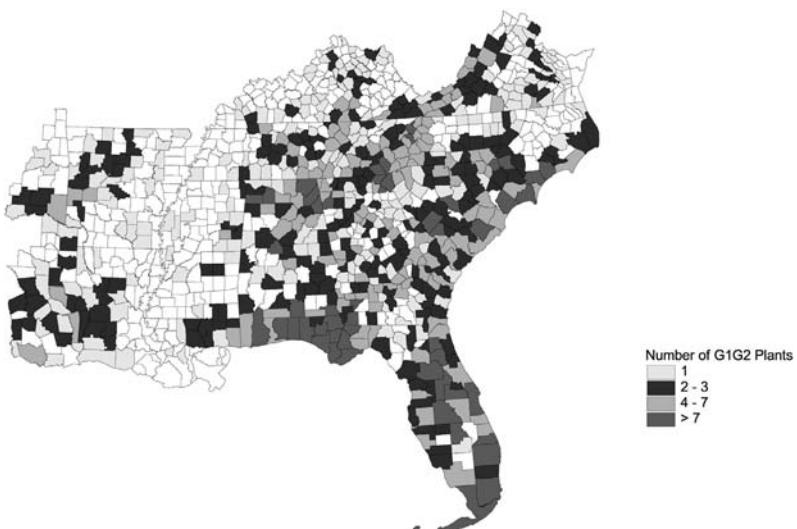


Figure 2. Distribution of imperiled vascular plant species in the South based on the number of occurrences in counties.

Figures 1 and 2 display three hotspots of plant diversity in the South; the southern Appalachian Mountains, the Apalachicola lowlands of the Florida Panhandle, and the Lake Wales Ridge region of central Florida. The Southern Appalachians are a refuge for a wide range of species in genera with generally more northerly affinities. Many of the

rare taxa in the Southern Appalachians are thought to be relicts from periods of glaciation in the distant past. The Lake Wales Ridge hotspot is a portion of Florida that was submerged during times of rising sea-levels, such as during the hypsithermal period from 8,700 to 5,000 BP. Many of the rare plants on Lake Wales ridge are thought to have been more widely distributed in the past. The Appalachicola lowlands plant diversity hotspot is more difficult to explain. Although the area has a striking diversity of habitats such as karst features, a variety of bogs, and wiregrass communities, these factors alone are unlikely to be the cause of the richest endemic flora in the South. Some scientists have suggested that some combination of habitat diversity, generally markedly low levels of soil nutrients, and a long history of frequent fires has made the area a challenge for most plant species and an opportunity for the evolution of specialized taxa.

Other areas with important levels of plant diversity in the South include the Coastal Plain, the Ozark-Ouachita Highlands, and the Cumberland Plateau.

Although most of the rare plant species in the South are species that are naturally rare, forest fragmentation and land conversion has significantly impacted the distribution and abundance of a large number of species. Other factors associated with human density such as over-harvesting and hydrologic alterations have diminished many species that were formerly common.

Many of the plant diversity hotspots represented in Figures 1 and 2 occurs primarily or largely on public land. This result highlights the importance of public land for the conservation of rare plants. Although not all public land management practices favor rare plants, in many places public land is the only place in which rare plant conservation is politically or economically possible.

KEYWORDS

- **Bartram**
- **Exotic species**
- **Glacial periods**
- **Southern Appalachians**

ACKNOWLEDGMENTS

Figures 1 and 2 were developed by the Association for Biodiversity Information (ABI) Natural heritage Central Database, Arlington, VA., USA (2001). Throughout the above discussion, work currently being accomplished by ABI has been attributed to its parent organization, “The Nature Conservancy”, for the sake of continuity. This chapter would not have been possible without the selfless work of thousands of professional and amateur botanists during the last 300 years. They have sweated and braved insects and briars to document, describe, and extol the flora of the Southeast. The chapter is dedicated to the thousands of young people that have yet to discover their love affair with botany.

Chapter 15

Predicting Pleistocene Climate from Vegetation in North America

C. Loehle

INTRODUCTION

Climates at the Last Glacial Maximum (LGM) have been inferred from fossil pollen assemblages, but these inferred climates are colder for eastern North America than those produced by climate simulations. It has been suggested that low CO₂ levels could account for this discrepancy. In this study biogeographic evidence is used to test the CO₂ effect model. The recolonization of glaciated zones in eastern North America following the last ice age produced distinct biogeographic patterns. It has been assumed that a wide zone south of the ice was tundra or boreal parkland (Boreal-Parkland Zone or BPZ), which would have been recolonized from southern refugia as the ice melted, but the patterns in this zone differ from those in the glaciated zone, which creates a major biogeographic anomaly. In the glacial zone, there are few endemics but in the BPZ there are many across multiple taxa. In the glacial zone, there are the expected gradients of genetic diversity with distance from the ice-free zone, but no evidence of this is found in the BPZ. Many races and related species exist in the BPZ which would have merged or hybridized if confined to the same refugia. Evidence for distinct southern refugia for most temperate species is lacking. Extinctions of temperate flora were rare. The interpretation of spruce as a boreal climate indicator may be mistaken over much of the region if the spruce was actually an extinct temperate species. All of these anomalies call into question the concept that climates in the zone south of the ice were extremely cold or that temperate species had to migrate far to the south. An alternate hypothesis is that low CO₂ levels gave an advantage to pine and spruce, which are the dominant trees in the BPZ, and to herbaceous species over trees, which also fits the observed pattern. Thus climate reconstruction from pollen data is probably biased and needs to incorporate CO₂ effects. Most temperate species could have survived across their current ranges at lower abundance by retreating to moist microsites. These would be microrefugia not easily detected by pollen records, especially if most species became rare. These results mean that climate reconstructions based on terrestrial plant indicators will not be valid for periods with markedly different CO₂ levels. The Pleistocene LGM period of 18,000 years ago has been widely interpreted as a time of bitter cold in eastern North America when tundra and boreal forest extended hundreds of miles south of the ice sheets and the temperate forest of the East retreated to the southern coastal plain, to Florida, and westward into Texas and Mexico (Davis, 1983, 1984; Davis and Shaw, 2001; Deevey, 1949; Delcourt and Delcourt, 1984, 1993; Jacobson et al., 1987; Maher et al., 1998; Maxwell and Davis, 1972; Over-peck et al., 1992; Prentice et al., 1991; Ritchie, 1987; Royall et al., 1991;

Schoonmaker and Foster, 1991; Tallis, 1991; Watts, 1970, 1971, 1973, 1979, 1980a, 1980b; Watts and Stuvier, 1980; Webb et al., 1988, 1993; Whitehead, 1973; Wilkins et al., 1991). This reconstruction, which may be called the standard model, is commonly presented in textbooks (e.g., Bradley, 1999; Delcourt and Delcourt, 1993; Pielou, 1991; Ritchie, 1987; Tallis, 1991). The standard model is based largely on pollen and macrofossil records. These pollen data have been interpreted qualitatively in some cases, and in other cases transfer functions or response surface models (e.g., Farrera et al., 1999; Nakagawa et al., 2002; Peyron et al., 1998; Tarasov et al., 1999; Webb et al., 1993, 1997) have been used to infer climate from pollen composition, with similar results. By any of these three methods, the inferred LGM climate is much colder than that simulated (Ganopolski et al., 1998; Huntley et al., 2003; Kageyama et al., 2001; Pinot et al., 1999; Webb et al., 1993, 1997). Determining the correct paleotemperature is important for calibrating general circulation models (e.g., Crowley, 2000; Farrera et al., 1999; Ganopolski et al., 1998; Pinot et al., 1999). In addition, the vegetation composition at this time is consistently described as having no analogs in modern periods. In this chapter it is argued that both of these anomalies arise from a combination of ambiguous pollen interpretation and the effects of low ambient CO₂ at the LGM which would have altered the relative dominance of different taxa in a manner that mimics colder and drier climates. Biogeographic and phylogenetic data are cited as independent tests of the CO₂ effect model.

The idea that glacial levels of CO₂ could affect vegetation structure and composition was proposed over a decade ago (e.g., Beerling, 1996; Beerling and Woodward, 1993; Bennet and Willis, 2000; Cole and Monger, 1994; Cowling, 1999; Cowling and Sykes, 1999; Farquhar, 1997; Harrison and Prentice, 2003; Jolly and Haxeltine, 1997; Polley et al., 1993, 1995; Street-Perrott et al., 1997). Nevertheless, LGM climate as estimated from pollen data generally still does not factor in CO₂ effects (e.g., Elenga et al., 2000; Nakagawa et al., 2002; Peyron et al., 1998; Tarasov et al., 1999) and discussions of simulated LGM climate versus vegetation (the anomaly problem) either do not mention the role of CO₂ (e.g., Ganopolski et al., 1998; Pinot et al., 1999) or only touch on this effect (e.g., Crowley, 2000; Huntley et al., 2003).

To date, the discussion has featured dueling models. Regression models based on pollen (which seem straightforward) are pitted against simulations of plant communities under low CO₂. It is tempting to be wary of the simulations because they cannot be directly verified. In what follows, several types of independent evidence are presented that imply that LGM climates inferred from pollen data are colder than the likely actual climates. These various types of evidence are shown to reconcile with effects of low ambient CO₂, thereby supporting the models of CO₂ effects. The results of the analysis also bear on questions of glacial refugia. The focus of this study is largely eastern North America for the sake of concreteness, but no-analog vegetation found elsewhere at the LGM can be explained by similar mechanisms.

GEOLOGIC ANOMALIES

By analogy with European mountains, it has been assumed (e.g., Deevey, 1949; Delcourt and Delcourt, 1984) that the southern Appalachians, particularly the Great

Smoky Mountains, should have been covered by permanent ice caps and should have generated glaciers. Mountain glaciers produce obvious signs like U-shaped valleys, striations, and moraines. None of these signs has ever been found in the southern Appalachians south of Pennsylvania. More recent treatments assume tundra but not glaciers on the peaks.

Tundra soils produce distinctive signs, due to processes such as solifluction, cryoturbation, ice wedges, and stone sorting. While such soil disturbances can coexist with trees (as in taiga) the lack of such signs rules out permafrost and therefore contradicts the existence of tundra. No such signs have been found more than a few miles south of the edge of the ice between Wisconsin and Pennsylvania. For example, Braun (1951) noted that at a site only 15 miles south of the ice margin in Ohio, evidence of congeliturbation, although present, did not indicate severe frost action, an observation also supported by Wolfe (1951). Black (1976) concluded that there is no evidence for permafrost in Illinois, Indiana, and Ohio south of the ice margin. A number of sources (Burns, 1958; Gooding, 1963; Grüger, 1972a, 1972b) found evidence for temperate forest vegetation right up to the ice margin throughout the period of the last glacial advance in Illinois, Indiana, and Ohio. Farther north and west, in Minnesota, Montana, and the Dakotas, the southern glacial margins were frozen to their beds and signs of permafrost are present (Clayton et al., 2001; Mickelson et al., 1983). Péwé (1983) compiled abundant evidence for permafrost from the driftless area of Wisconsin and from across Pennsylvania, but none from central Illinois east to the edge of Pennsylvania. Péwé (1983) shows some evidence for patterned ground on the high peaks of the Appalachians, but only as far south as the southern Virginia-West Virginia border. Periglacial forms south of this either are undated or are not reliably identified (Péwé, 1983). Denny (1951) documented frost action in Pennsylvania that diminishes with distance from the ice margin. There is thus evidence for a band of permafrost and periglacial climate in Pennsylvania and Wisconsin but none in between. This reconstruction makes sense because the area of western Pennsylvania is higher in elevation than are southern Ohio, Indiana, and Illinois. Thus, permafrost (and tundra) was not a universal feature of the ice front region in eastern North America as implied by most reconstructions. Wider periglacial zones did of course occur in Europe.

The lack of tundra puts limits on how cold this region could have been.

TAXONOMIC ANOMALIES

A key species in the reconstruction of LGM climates in eastern North America is spruce (*Picea*). The pollen of various species of spruce are very difficult to distinguish. Since all existing spruce species in eastern North America require a cold climate, they have generally been considered together (see, e.g., Davis and Shaw, 2001). In this context, pollen profiles dominated by spruce and herbaceous species have been classified as boreal parkland.

It is now apparent, however, that a major portion of the southern range of spruce may have been occupied by a now-extinct species (Jackson and Weng, 1999) which was common in the lower Mississippi valley and east at least to western Georgia. Given that fossil stumps of this species can be found growing with oaks and with

strictly southern species, this was likely a temperate spruce. This spruce pollen is thus not necessarily indicative of cold climates. This may have biased climate reconstructions at the LGM to a considerable degree.

A similar problem occurs with sedges (*Carex*). This taxonomic group has been interpreted as diagnostic of boreal or tundra habitats. However, sedges are also common in grasslands and forests and would have been favored under a low CO₂ atmosphere (Beerling, 1996; Beerling and Woodward, 1993; Jolly and Haxeltine, 1997; Robinson, 1994; Saxe et al., 1998), as will be discussed below.

In both cases, the species have been considered key indicators or diagnostic species of boreal forest but lumping species is known to produce misleading indicators of climate (Finkelstein et al., 2006). The fact that their pollen is found mixed with that of temperate species such as oaks across the “boreal parkland” (e.g., Webb et al., 1988, 1993, 1997), further calls into question their indicator status. When models are created for climate, the climatic tolerance of individual species is used to create a regression model that predicts climate, or a biome is defined from the dominant species, and that biome is used as an indicator of climate. Inclusion of ambiguous genera such as spruce in the model will create a cold bias. This analysis also helps explain part of the no-analog vegetation situation.

BIOGEOGRAPHIC ANOMALIES

The geographic distributions of plants can reveal a great deal about the history of a region. This is true of evidence for refugia, genetic gradients, distributions of endemics, and the existence of races and subspecies.

Glacial Refuge Anomalies

The traditional model assumes that temperate forest was forced to retreat to the far south (Gulf Coast region) by the cold (e.g., Bennett, 1985; Cain et al., 1998; Delcourt and Delcourt, 1984, 1993; Watts and Stuvier, 1980). Hewitt (2004) posits an Appalachian refuge, but without citing any data. When we look for fossil evidence for the migration of midwestern temperate forest species to the Gulf Coast and Florida (the refugia for deciduous forest in the traditional model), we encounter a striking lack of evidence. Lakes in the Gulf Coast area, Texas, and Florida either do not date back far enough (i.e., not as far as the LGM) or show a continuous presence of oak-pine associations (including such shrubs as wax myrtle at Sheelar Lake) similar to those present today (Bryant, 1977; Bryant and Holloway, 1985; Delcourt and Delcourt, 1993; Givens and Givens, 1987; Jackson and Givens, 1994; Watts, 1973; Watts and Stuvier, 1980). The most recent pollen maps do not show any distinct refugia for northern temperate species, most of which are either undetectable or diffusely rare across the region (Williams et al., 2004). It has been shown (Froyd, 2005; McLachlan and Clark, 2004) that rare species can be hard to detect in pollen samples. Thus, data do not support the existence of a Florida or Gulf Coast refuge. The area in question consists of the geographically unbroken piedmonts and coastal plains of Georgia, South Carolina, Florida, and Alabama. Except for fish, there are no real barriers keeping species from

moving from east to west across these areas, and they indeed today share a common flora. There is no mechanism to create distinct refuges, as there was in Europe.

A major phylogeographic study (Soltis et al., 2006) shows that there is a lack of genetic structure concordance across taxa. The genetic structure of a species defines how unique alleles vary over space. Concordance of pattern should occur if species shared a common refuge, as did occur in Europe where concordance is more evident.

While it is evident that some species did retreat to the Gulf Coast (Soltis et al., 2006), there is no evidence for a comprehensive southern refuge for the bulk of the species.

A related fact is that the high elevation rock outcrop flora of the southern Appalachians would be expected to be derived from boreal/alpine elements that would have moved in during the LGM. Instead, Wiser (1998) found very little overlap between the flora in the high elevation southern Appalachians and high elevation flora in the Presidential Range mountains of New Hampshire and the White Mountains, while the latter do show the expected affinities to alpine/tundra floras.

Genetic Gradients

After the last glacial advance 18,000 years ago, and the beginning of the Holocene, the North American ice cap began a rapid retreat (Pielou, 1991; Tallis, 1991). Plants colonizing the previously glaciated lands could potentially spread rapidly over a very large newly deglaciated area. Such rapid spread causes particular effects on population genetic structure. If a species was confined to a very small refuge with a small population, a genetic bottleneck combined with homogenization will produce a population with very little genetic differentiation (Hewitt, 2004). As this population spreads across the deglaciated zone, the genetic uniformity will be maintained for a long time. For trees, only a few hundred generations have passed since the LGM, too little for either random mutation or local adaptation to have had much effect. An example of this phenomenon is red pine (*Pinus resinosa*), which occurs largely in previously glaciated terrain and which has very low heterozygosity (Burns and Honkala, 1990). Almost no species can be found matching this pattern south of the ice sheets (Soltis et al., 2006), suggesting that species during the Pleistocene were not confined to small refugia.

If, on the other hand, a species invades the deglaciated zone from a wide front with a large initial population and high heterozygosity, the rapid migration will cause a gradual loss of rare alleles as the distance from the refugial population increases (Hewitt, 2004). This pattern is seen in lodgepole pine (*Pinus contorta*), which has high heterozygosity in the unglaciated refugia and progressive decrease in heterozygosity with distance north (Burns and Honkala, 1990). The same genetic structure is seen in black spruce (*Picea mariana*) in the formerly glaciated zone, in which northernmost populations have low mitochondrial DNA diversity (Gamache et al., 2003).

In contrast to these expectations for rapidly migrating populations, species in the unglaciated (but supposedly Pleistocene-era tundra or boreal zone) East show deep DNA divergence (which means long isolation) between separate sub-populations, weak or absent genetic gradients, and high genetic richness (Hewitt, 2004; Soltis et al., 2006). If these species had been excluded by cold from the area south of the Great

Lakes (southern Appalachians) and migrated back into this region, genetic diversity gradients should exist but they do not.

Distribution of Endemics

If tundra and boreal forest/parkland formed zones south of the ice margin in the East, then temperate vegetation would have had to retreat to the far south to refuges, as is commonly assumed. The data on endemism seriously constrain theories proposing large-scale migrations of vegetation at the LGM. Many endemics occur just south of the ice margin; almost none are found within the glaciated area (e.g., Baskin and Baskin, 1986, 1989; Estill and Cruzan, 2001; Qian, 1999; Schofield, 2004), and those that do generally are either hybrids or are recently derived. The only possible explanation for this observation is that the unglaciated East had climates that these endemic species could tolerate throughout the glacial period.

Local endemics are species or discrete races with a restricted geographic distribution (Brown and Gibson, 1983; Cain, 1944). Endemics may arise in a number of ways. In some cases, a species that is originally widespread may become progressively restricted to a unique habitat as climate changes, physiography changes, or other species out-compete it. It thereby becomes a relict species. For example, a swamp-dwelling species might progressively lose habitat during periods of uplift when swamps are drained.

A second type of origin of endemics involves evolution in a unique habitat, leading to the generation of a race, a subspecies, and ultimately a new species. For example, in the tropics the species on isolated volcanic peaks are often derived from tropical lowland elements and are unrelated to those on other mountains. Striking examples occur in cave populations of animals, with the cave population diverging drastically in morphology from the normal population. How many specialized species could have evolved since the LGM is unknown, but data on speciation rates from the fossil record (Levin and Wilson, 1976) suggest that very few species could have originated in such a short time. This is especially true for trees, which have only experienced a few hundred generations since the LGM. Telltale signs of recent origin for endemics are the proximity of the parent species, incomplete genetic isolation, and limited morphologic divergence. These signs of recent origin do not apply to most endemics in the unglaciated regions but do apply to those found in the recently glaciated regions.

Maps of the distribution of endemic species show that recently glaciated regions lack endemics (Estill and Cruzan, 2001; Qian, 1999; Schofield, 2004). The high frequency of endemics in central Alaska led Hultén (1937) to propose that a large portion of Beringia remained unglaciated and served as a refuge, a view since substantiated by geological studies. As Braun (1950, 1951) noted, the southern Appalachians, particularly the region of the Cumberland Plateau and the Blue Ridge escarpment, is a center of both species richness and endemism for the eastern United States. Both occur because this region has an ancient floristic history (since at least middle Tertiary time) as the only area of large mountains in the East. The mountains have a wide variety of soils, climates (often moist and equable), and topographies that both provide habitat for many species and encourage endemism by imposing barriers that isolate populations.

Many endemics can be found in this region across life forms including trees (*Abies fraseri*), ferns, shrubs, herbaceous plants, fish (Gilbert, 1987), salamanders, crayfish, and centipedes. The extremely limited distributions of many endemics in the southern Appalachians pose a real difficulty for any theory of local extinctions, distant refuges, and subsequent re-immigration, as does evidence concerning the time of their origin. Given the large number of endemics, such a major forced migration should have caused many extinctions, but very few Pleistocene extinctions of land plants in eastern North America have been documented (Potts and Behrensmeyer, 1992; Roy et al., 1996). As a group, these endemics exhibit many signs of ancient origin, including high morphologic divergence from related species, geographic isolation from congeners, complete reproductive isolation, and edaphic specialization. A number of endemic plants, for example, have no close relative in the East, and their nearest congener is in China or Europe (e.g., Qian, 1999). The very few endemics in glaciated regions do not exhibit these signs.

A telling example of the significance of endemics is a community centered on sinkhole ponds in the Blue Ridge Mountain region of Virginia (Church et al., 2003). In this small 1,350 ha area are found disjunct and endemic populations of 70 plants and animals. The adaptation of these species to this unique habitat would prohibit these species from migrating under a changing climate.

To conclude endemics in the unglaciated region have origins that predate the LGM and therefore that they survived the glacial climate in the locations where they are found, whereas endemics in glaciated regions have a recent origin.

Distribution of Races and Subspecies

The geographic distribution of races and subspecies provides invaluable information for interpreting past climatic changes. Even though most race and subspecies distinctions cannot be made using pollen data, such distinctions cannot be ignored. If multiple distinct races of a species can be identified, then (1) each must have had a separate glacial refuge to prevent introgression; or (2) they all evolved since the LGM; or (3) the bulk of the range of the species was not disturbed during the glacial period. If they had all been jammed together in a single refuge, they would have merged into a single species. Because, as noted, finding even a single glacial refuge in the far south has proven problematic, finding multiple refugia within which each of the distinct races of these many species could be preserved would seem doubly problematic. This is particularly true due to the geography of eastern North America. The Gulf Coastal Plain lacks significant east–west barriers, so most plant species today occur across most of the plain if they occur there at all. There is no mechanism by which different races or related species of plant could be geographically isolated during the Pleistocene in this region.

Fagus grandifolia has three distinct races, the gray, red, and white beeches (Bennett, 1985; Camp, 1950; Cooper and Mercer, 1977), which differ on multiple traits. These three races readily hybridize where their ranges overlap. Beech in formerly glaciated territory appear to originate from populations west of the Appalachian Mountains and just south of the ice sheet (McLachlan et al., 2005). Florida red maple

(*Acer rubrum*) is genetically distinct from other red maple populations (McLachlan et al., 2005). Genetic data for red maple further support refugia close to the ice (Soltis et al., 2006). Evidence is also accumulating for southern Appalachian refugia for various animal species (e.g., Austin et al., 2002, 2004; Brant and Ortí, 2003; Church et al., 2003). The genetics data for *Liriodendron* point to a number of distinct races. Virginia pine has two distinct races, as defined by genetics data (Parker et al., 1997); one to the northwest and one to the southeast of the main Appalachian axis. Although Parker et al. (1997) attribute the origin of the northwestern population to a postglacial migration from the southeast, based on its lower genetic diversity (often found in colonizing populations), the substantial number of unique alleles found in various northwestern populations suggest that it is not merely a genetic subset of the southeastern population. Rather, this gives a suggestion of prolonged isolation, possibly throughout the Pleistocene, in which case it had to remain to the northwest of the mountains during the LGM to remain separate from the southeastern race. Fraser fir, found in the southern Appalachians, will hybridize with balsam fir, but the only hybrids at present are near West Virginia. In all of these cases the distinct races would have required the existence of multiple refuges in order to persist.

We may similarly look at closely related species for evidence of LGM distributions. Sugar maple and Florida maple are not well separated morphologically. Where their ranges overlap, they hybridize readily. If sugar maple had migrated to the Gulf Coast, it would have occupied the same range as Florida maple, in which case the two species would have merged. A similar problem exists with Virginia pine. This pine is only kept from crossing with *Pinus clausa* (a Gulf Coast species) by their considerable geographic separation (Parker et al., 1997). If Virginia pine had moved down into Florida at the LGM, it would have merged with this closely related species. These examples are just a sample of the many cases where closely related species are today prevented from crossing by geographic separation but would have merged if crowded together in a single coastal plain refuge. The races of these several species do not coincide geographically, further multiplying the number of required refuges. Phylogenetic analysis of these species leads to the same conclusion (Soltis et al., 2006).

DISCUSSION

The anomalies documented here are not trivial. The biogeographic patterns expected for rapidly migrating species are observed in the glaciated zone, but rarely in the unglaciated zone. Boreal zone species exhibit low genetic variation or diversity gradients, but species in the unglaciated zone show high genetic richness, no gradients, and deep (ancient) DNA divergence between populations. Endemics are very rare in the glaciated zone, and endemics in this zone appear to be mostly recent (due to hybridization, for example). Sister species, subspecies, and races are common across the unglaciated East but not in the glaciated zone. To keep all of these groups separate during the Pleistocene so that they did not hybridize/homogenize out of existence would require multiple LGM refuges, but no such refuges have been found in the far South. If the pollen data say one thing and the biogeographic data flatly contradict this inference, what can we conclude?

There is a possible resolution to this paradox based on climate and CO₂ factors. First, glacial climates are not simply a general cooling of the climate. Glacial epoch climates were cooler, but more so in summer (Pielou, 1991). Thus, the existence of “boreal” species farther south than today does not mean that climates were bitterly cold. Second, a factor that also varied during the glacial period is CO₂, which was so low that it caused CO₂ starvation. CO₂ starvation does not affect all species equally.

At the LGM, about 18,000 years ago, CO₂ levels were very low, less than 200 ppm. This low level of CO₂ constitutes a severe deficiency for growth, and may have shifted competitive dominance between different plant types, as well as affecting overall vegetation biomass (Levis and Foley, 1999). Altered CO₂ levels could also affect pollen production, which would bias biome reconstruction. C4 grasses have a strong advantage over other plants under low CO₂ (Cole and Monger, 1994; Farquhar, 1997; Polley et al., 1993, 1995), though not under low temperatures. Robinson (1994) shows that the extent of stomatal regulation plants exhibit in response to CO₂ level varies by taxa in a manner that suggests that this trait is a modern adaptation. Ancient taxa such as conifers exhibit little stomatal responsiveness compared to angiosperms. The benefit of stomatal responsiveness trades off against greater water loss at low CO₂ levels resulting from stomates being open longer (Drake et al., 1997; Saxe et al., 1998). Where water is not limiting, a cool low CO₂ climate should favor C3 grasses, forbs, and angiosperm trees over conifers (Beerling, 1996; Beerling and Woodward, 1993; Jolly and Haxeltine, 1997; Robinson, 1994; Saxe et al., 1998). This suggests that mesic broadleaf trees should have remained competitive in mesic microsites during the Pleistocene. On most upland sites, however, a low CO₂ climate would favor xeric species such as conifers, which have consistently low stomatal conductance, and herbaceous species, including grasses and sedges. Oaks, as trees with intermediate responsiveness, would have persisted as well on upland sites. We in fact observe extensive open conifer forest replacement of broadleaf forest at the LGM in eastern North America, with a persistent oak component. Simulation studies also show that the reduced water use efficiency expected at the LGM would produce a xeric, open forest (with low leaf area index) south of the ice dominated by conifers (Cowling, 1999; Harrison and Prentice, 2003; Jolly and Haxeltine, 1997; Levis and Foley, 1999). This explains the “no-analog” open forest (or parkland) often remarked upon. This discussion shows that species that currently occur together would diverge in their responses under the LGM climate, and that climate predictions based on their occurrence would not be correct. Exact simulations of vegetation would need to be based on more detailed studies of the physiology of each species, few of which have been done.

Simulations of the effect of low CO₂ levels show that it could cause a major lowering of alpine treelines (Bennett and Willis, 2000; Cowling and Sykes, 1999; Street-Perrott et al., 1997). Jolly and Haxeltine (1997) simulated the montane ecotone for African mountains and showed that the entire LGM lowering of treeline at their study site is consistent with the effect of low CO₂ without assuming any drop in temperature. Lower treelines at the LGM have typically been taken as evidence for colder temperatures, but could really be the result of changes in CO₂ levels. They would thus not correspond to snowline depressions (e.g., Greene et al., 2002) or other indicators of temperature.

If lowered CO₂ affected vegetation in eastern North America, what changes would be expected? Grasses and other grassland species would be expected to expand their range into forest, perhaps creating parkland. Forest communities would shift to drought tolerant species such as pines and oaks. Significantly, mesic microsites such as coves, north slopes, and stream valleys would provide refuges for mesic species across their entire original range, especially in the highly dissected Appalachian region, because the effect of CO₂ would be similar to an overall drying. High precipitation, midelevation regions such as the Cumberland Plateau and the Southern Blue Ridge Escarpment would provide prominent refuges. In contrast, topography provides much less protection against extreme cold. An LGM African montane site (Jolly and Haxeltine, 1997) shows this pattern, with low levels of tropical montane forest pollen persisting in LGM profiles dominated by ericaceous shrubs. While drought-tolerant boreal elements such as white spruce (or the extinct temperate *Picea crutchfieldii*) could have moved south under a cooler, low CO₂ climate, most eastern boreal trees are not drought tolerant and would not have been able to move far south. We observe exactly this pattern of change in the pollen record, with most boreal trees being absent from the supposed boreal parkland south of the ice. Several authors have proposed that the southern Appalachians were a refuge (e.g., Braun, 1950, 1951; Church et al., 2003; Harvill, 1973; Hewitt, 2004). This view now has increased plausibility. This analysis provides independent biogeographic support for model-based projections of the effects of CO₂ on LGM vegetation. Other factors such as altered fire regimes and herbivores could also impact the details of reconstructions.

The above discussion has implications for the locations of refugia in Europe. The traditional reconstruction of LGM Western Europe has been a treeless steppe-tundra as far south as southern France (Stewart and Lister, 2001), with presumed refugia in the Mediterranean areas such as the Iberian and Italian peninsulas (Stewart and Lister, 2001). However, fossils of thermophilous trees and mammals have been dated to the LGM in Belgium, England, Hungary, Slovakia, and Germany, among other places (Stewart and Lister, 2001; Willis and van Andel, 2004). In addition, populations of numerous trees and animals such as Scots pine in Scotland (Stewart and Lister, 2001) and hedgehog in Germany (Willis and Whittaker, 2000) are either genetically distinct from Mediterranean populations (Scots pine) or have no southern relatives (hedgehog). For these species, it appears that they persisted throughout the LGM in refugia scattered across Europe in what is assumed to be a climate too cold for them to persist. Stewart and Lister (2001) interpret these refugia, most of those known being fossils from cave sites in steep valleys, as thermal refuges from the ice age cold. However, steep valleys are not particularly known for providing thermal refuges in present-day northern habitats. In fact, in hilly terrain, it is south and west facing midslopes and ridges that are warmest, not valleys, which receive less sun and are subject to cold air drainage. It is more likely that these valleys provided mesic refuges from the combined effects of a drier climate and the reduced water use efficiency caused by the CO₂ effect. These mesic plant communities would provide a home for the animals found there. Just such steep valley and stream-margin refuges are found today in dry regions across the world. It is noteworthy that tree species that went extinct in Europe were less drought tolerant than surviving species (Svenning, 2004), as predicted by the CO₂

effect model. To an even greater extent than in North America, the reduction of tree cover due to the CO₂ effect may have given a false impression of extreme dryness in Europe. It is not asserted that all aspects of the situation are analogous.

There are a number of implications of these results for the practice of climate reconstruction and for testing climate models against historical proxy data. Regression or response surface approaches (e.g., Farrera et al., 1999; Nakagawa et al., 2002; Peyron et al., 1998; Tarasov et al., 1999; Web et al., 1993, 1997) implicitly assume that species' climate responses are stable over time, but the CO₂ effect model suggests that this assumption is violated. The use of plant functional types to reconstruct either biomes (e.g., Elenga et al., 2000) or climate (e.g., Peyron et al., 1998; Tarasov et al., 1999) has the same issues as response surface models and also assumes that similar species will respond to a changing climate as a group. It was shown above that normally co-occurring species will show divergent responses to CO₂ change. Very few studies can be found that take explicit account of CO₂ effects when reconstructing vegetation and climate (e.g., Jolly and Haxeltine, 1997; Guiot et al., 2000; Levis et al., 1999). Regression and response surface models, including models of plant functional types, also are unable to account for other changes in the conditions at the LGM. For instance, seasonal distributions of temperature were not strictly analogous to latitudinal shifts. That is, the colder temperatures at the LGM did not produce seasonal shifts that correspond to a simple northward movement of place. This throws off calibration of models of plant response to climate. Annual variability may have been altered, fire regimes may have been different, and in North America the presence of large herbivores that later went extinct really should not be ignored. All of these factors could have shifted the community composition in ways that would give the impression of some climate effect.

Testing of general circulation models is likewise affected by the CO₂ response of vegetation. While there is some recognition that CO₂ could affect pollen interpretations, the range of weight given to this issue ranges from none (Nakagawa et al., 2002; Pinot et al., 1999) to a consideration of only the C3 versus C4 effect (Crowley, 2000; Farrera et al., 1999). Few climate model comparisons to paleotemperature proxies have fully factored in the CO₂ effects on leaf area, plant types, and water use efficiency in the temperature reconstruction. The treatment of vegetative cover and transpiration rate within climate simulation models is probably also deficient in this regard, though published descriptions of this model component are not usually adequate to determine exactly how vegetation is modeled.

In summary, the entire interpretation of LGM vegetation and climate in eastern North America may have been biased by the CO₂ effect. It was neither as cold nor as dry as has been assumed. Lower treelines were probably caused by the CO₂ effect. The "no-analog" conifer woodlands are the direct result of changes in water use efficiency between taxa and the lumping of spruce species and sedge species into generic-level categories. The presence of spruce in eastern North America was not an indicator of a boreal climate. Massive vegetation dislocations and migrations did not occur prior to ice melt, and the entire LGM unglaciated region acted as a refuge for species of the eastern deciduous forest. This explains the almost complete lack of tree species extinc-

tions in this region. It also suggests that the use of plant remains to predict climate for any period of the past in which CO₂ level was appreciably different from today may lead to incorrect conclusions unless the effect of CO₂ on relative growth rates is accounted for.

KEYWORDS

- *Fagus grandifolia*
- Last Glacial Maximum
- Local endemics
- Periglacial climate
- Permafrost

ACKNOWLEDGMENTS

This work was sponsored by the U.S. Department of Energy, Office of Energy Research, Office of Health and Environmental Research, Program for Ecosystem Research, under contract W-31-109-Eng-38 and by the National Council for Air and Stream Improvement, Inc. Some of the concepts about endemic distributions and difficulties with the pollen record have been expressed in correspondence and talks by H. Iltis dating back to the 1960s, including a talk in 1994 at the Ecological Society of America meeting in Knoxville, TN. Helpful reviews were provided by H. Birks, M. Davis, J. Hoffecker, H. Iltis, R. K. Peet, J. Ritchie, J. Tallis, and T. Webb III. H. Iltis provided many references. And edited by: C. Hatte.

Chapter 16

Past Vegetation Patterns of New Mexico's Rio Del Oso Valley

Richard D. Periman

INTRODUCTION

Humans have interacted with the landscape and ecosystem of New Mexico's Rio del Oso Valley for thousands of years. Throughout the Holocene, various cultures have dramatically affected and altered the Rio del Oso. An interdisciplinary research approach, incorporating geomorphology, paleobotany, archaeology, and history, provides a broad range of methodologies and data sets of past landscape dynamics. Integrating such data sets in three-dimensional Geographical Information Systems (GISs) models of past vegetation and landscape conditions may enable a view of anthropogenic ecosystem change. Analyses of past land use through landscape models, geoarchaeology, and other methods can provide a greater understanding for current and future ecosystem management. Landscape history and geoarchaeology may be applied in a forum that integrates concerns regarding the management of ecosystems. Since the end of the Pleistocene, humans have manipulated and domesticated global resources, transforming the world into a system of interconnecting and overlapping cultural landscapes (Caseldine and Hatton, 1993; Edwards, 1988; Norton, 1989; Simmons, 1988). Interactions between human cultures and environments have induced worldwide ecological change. Historic and current industrial activities produce measurable increases in carbon dioxide, greenhouse gases, and inorganic and organic nitrogen (Rossignol and Wandsnider, 1992; Schmidt, 1998; Vitousek et al., 1997; Willey and Sabloff, 1993). Culturally induced environmental changes are compounded by increased occurrences of flooding, prolonged drought, and possible famine (McIntosh et al., 2000). On a global scale, scientific analyses of environmental change suggest that the momentum of alterations in ecological processes continues to accelerate (Schmidt, 1998; Vitousek et al., 1997).

Hunter-gatherers initiated cumulative processes of broad-scale environmental change, influenced biodiversity, and maintained a heterogeneous landscape mosaic with the use of controlled burns over thousands of years, as evidenced by pollen and charcoal-particle analyses (Delcourt and Delcourt, 1996). In the Peruvian Andes, paleobotanical research shows widespread deforestation and burning to clear land for agriculture 4,000 years ago (Chepstow-Lusty et al., 1998). In the Appalachian Mountains of North America, the landscape was managed during the late Holocene with the cultivation and systematic care of chestnut oak (*Quercus prinus*) forests (Delcourt and Delcourt, 1997; Delcourt et al., 1998). Along the coastal areas of the Americas, indigenous peoples altered forests, grasslands, wetlands, and river valleys. The introduction

of intensive European agriculture and industry in existing Native American ecosystems superimposed further change and conflict onto landscape histories (Deneven, 1992:370; Mannion, 1991; Whitmore and Turner, 1990:416).

Human-induced changes in ecoprocesses, such as agriculture and controlled burning, affect plant-community composition, habitat area, and carbon, nutrient, water, and decomposition cycles. Changes in ecosystem processes, vegetation patterns, and geomorphology create variations in the configuration of landscape components (Hobbs, 1997). Fossil-pollen assemblages, microscopic charcoal, and faunal extinctions are indications of human-induced ecological change (Chambers, 1993a, 1993b; Walker and Singh, 1993). Knowledge of previous ecological conditions and capacities can be incorporated into environmental restoration and planning by land managers.

Cultural and paleoecological landscape approaches were integrated to reconstruct the environmental history of the Rio del Oso Valley, a tributary of the Rio Chama River, in north-central New Mexico, U.S.A. Human occupation has affected the conditions of the current Rio del Oso ecosystem. Cultural and ecological landscape interactions in the Rio del Oso have varied through time, from hunter-gatherers, pre-Columbian farmers, Hispanic ranchers, and commercial grazing pressures. Interdisciplinary methods identified cumulative anthropogenic effects, and provided the data sets used to reconstruct a long-term ecological history of the valley. Visual models of past landscape vegetation were created using GISs.

A set of definitions for ecosystems and anthropogenic landscapes that incorporate cultural and ecological environmental concepts were integrated in this Rio del Oso paleolandscape study. An ecosystem is “composed of physical-chemical-biological processes active within a space-time unit of any magnitude, i.e., the biotic community plus its abiotic environment” (Lindeman, 1942:400). An ecological landscape is a spatial matrix consisting of organisms, populations, and ecosystems (Allen and Hoekstra, 1992:56). Anthropogenic, cultural landscapes are formed within, and are integral to, ecological spatial matrices; this includes process as well as form, both changing through time. New anthropogenic landscapes are created within, and as part of, the ecosystem and spatial matrix.

The Rio del Oso Valley contains a rich archaeological record, resulting from over 7,000 years of human occupation. During the Archaic period (5500 B.C.–A.D. 600), hunter-gatherers in the valley dispersed seeds, selectively harvested plant and animal species, and used controlled burns to increase vegetative resource yields. Archaic archaeological data has been buried under 1–2 m of sediment, but significant arroyo erosion has enabled site discovery. Between the late Archaic and the Coalition period (A.D. 600–1200), there was less human occupation (Anscheutz, 1995), although archaeological evidence may simply remain buried, or unexposed by erosion.

Between 1200 and 1600 A.D., Puebloan groups built villages and agricultural features, causing significant landscape change in the Rio del Oso (Anscheutz, 1995). These activities placed selective pressures on local vegetation and animal populations. Pueblos cleared land for agriculture, constructed fields and rain conservation

structures, built carved terraces with cobble-mulch gardens to conserve water, and collected wood resources for fuel and construction. In the 1600s, Spanish settlers superimposed their own landscape patterns over those created by the Puebloans (Wozniak et al., 1992). They used existing pueblos, constructed roads, farms, and fortified settlements, and built systems of agricultural water ditches called *acéquias*. In the early 20th century, the U.S. Department of Agriculture, Forest Service, assumed administration of the area, adding another pattern of human use and environmental change.

Geology of the Rio Del Oso

The Rio del Oso is a perennial stream, fed by snow melt, rainfall, and springs. Its narrow watershed begins in the volcanic Jemez Mountains in north-central New Mexico, on the northern slope of Chicoma, the highest peak at an elevation of 3,524 masl. The headwaters of the Rio del Oso drainage originate in the Tschicoma Formation dacite. The northeastern margin of the Jemez consists of a series of extrusive flows and domes of late Tertiary volcanic rock (Dethier et al., 1988). The lower Rio del Oso Valley cuts into the indurated Ojo Caliente sandstone of the Tesuque Formation, late Tertiary Santa Fe Group. Exposed across the lower Oso Valley are north-south trending dikes of black Lobato Basalt (Dethier and Demsey, 1984). Pleistocene gravels occur at various elevations along the Rio Chama's valley margins, overlaying eroded Tertiary bedrock. This represents former channel and floodplain positions of the Rio Chama and its tributaries.

Gravel terraces indicate that the Rio Chama Valley has cut down 120 m during the past 600,000 years (Dethier et al., 1988; Dethier and Reneau, 1995). Pleistocene gravel deposits form a prominent terrace along the north slope, about 43 m above the Rio del Oso's low floodplain. The Rio del Oso flows 26 km northeast through the Jemez, dropping to 1743 masl, where it enters the Rio Chama River at the settlement of Chili. At this confluence, the watershed is less than 3 km across. The vegetation in Rio del Oso Valley consists of an Engelmann spruce (*Picea engelmannii*) cork bark fir forest at high elevations, Ponderosa pine (*Pinus ponderosa*), and Douglas fir (*Pseudotsuga menziesii*) at mid-elevations, and pinyon pine (*Pinus edulis*), juniper (*Juniperus monosperma*) woodlands, and juniper-shrub grasslands at lower elevations. Cottonwoods (*Populus fremontii*), aspen (*Populus tremuloides*), willow (*Salix exigua*), and other riparian vegetation are found in the stream bottoms (Figure 1). The surrounding mesas and terraces support a variety of mixed grasses, shrubs, and cacti. Although cryptogamic crusts provide some protection, plant cover is sparse, and underlying sandy sediments are exposed and vulnerable to erosion from wind and rain. Weather records from Los Alamos County show that annual rainfall ranges from 89 to 28 cm, from the high to low elevations of the eastern Jemez mountain slopes (Reneau et al., 1996). Along the lower, northern portion, the drainage is bordered by five mesas separated by deeply incised arroyos, which remain dry much of the year. Although heavy seasonal rains flood the arid tributaries, the water rapidly drains away, leaving little moisture behind. The southern edge of the valley is etched by erosion, deep arroyos, and colluvial slopes.

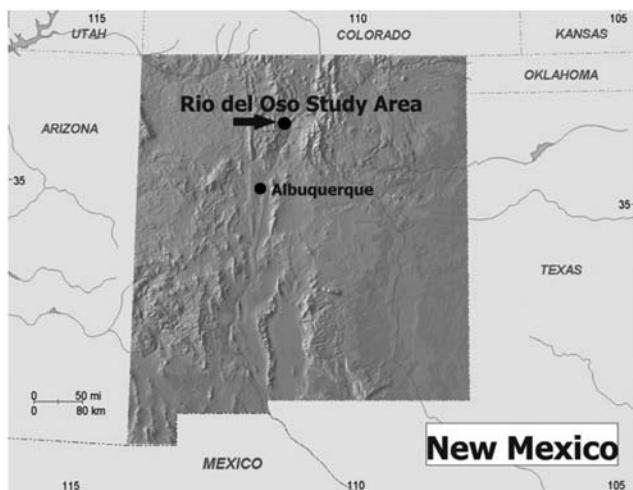


Figure 1. Study area in the lower portion of the Rio del Oso watershed, northern New Mexico.

The Rio del Oso floodplain has low terraces formed by deposition and scouring, currently 90–215 m wide in the lower 4 km of the stream valley, ending at the Rio Chama confluence. These terraces are approximately 1 m above the channel margin, and vegetated by grasses and scattered cottonwood, salt cedar (*Tamarix ramosissima*), Russian olive (*Elaeagnus angustifolia*), willow (*Salix exigua*), and elm (*Ulmus spp.*) trees. During intermittent periods of consistent flow in the Rio del Oso, water runs 55–120 cm deep, through two connecting channels incised in the floodplain deposits. Alternate channels occur seasonally, from snow meltwater in the Jemez Mountains during early spring and runoff from summer thunderstorms and monsoon rainstorms (Periman, 2001).

The sediments in the main channel of the Rio del Oso are composed of coarse quartz sand and gravel-sized clasts of basalt. The sand in this main channel is medium to quite coarse, pale brown in color, and contains basalt clasts; ranging from fine, to very coarse sand, and to granule-cobble-boulder gravels. The grains are subrounded, poorly sorted, and include carbonate-cemented sand aggregates about 1 mm in diameter, which form weak laminae and small crossbeds in floodplain deposits adjacent to the channel (Periman, 2001).

MATERIALS AND METHODS

The calibration of the fossil-pollen spectra from the Rio del Oso used pollen surface samples from modern vegetation communities and plant-density ratios in reconstructing models of past plant abundances. The relationship between vegetation and pollen assemblages are influenced by pollen production, dispersal patterns, deposition, preservation, and identification (Birks et al., 1988; Davis, 1963; Jackson, 1994). These processes are influenced by physical, chemical, biological, and behavioral factors (Sugita, 1994). Temporal and spatial variations affect comparison of corresponding

pollen and plant assemblages. Mathematical models of pollen-vegetation relationships were defined with the ratio between pollen and vegetational percentages, described as a taxon's R-value (Davis, 1963; Prentice and Parsons, 1983; Webb et al., 1981).

Analytical reconstruction of past environmental conditions in the Rio del Oso used analogues of current vegetation communities and ecosystems. Fossil-pollen assemblages represent the composition of vegetation that may have surrounded the site at the time of deposition (Jackson, 1994; Jackson and Kearsley, 1998). This pollen analysis used production and dispersion indices before the percentages derived from the pollen data to estimate past forest compositions. Variations in pollen assemblages corresponded with observable variations in vegetation composition. Comparable pollen-vegetation relationships in R-value calibration of pollen production and pollen-to-plant ratios estimated past and current plant abundances. Interpretation of pollen data provided indicators of the magnitude of vegetational population and change through time (Andersen, 1970; Birks and Gordon, 1985).

Preliminary fieldwork began in 1996, with the collection of sediment samples from a 5-m, exposed section of stratified alluvial deposits (Section A). The vertical arroyo wall was cleaned and faced using hand tools. Pollen, phytoliths, microscopic charcoal, and samples for radiocarbon dating were extracted from 24 strata defined by color and sediment composition. Prehistoric, archaeological hearth features were uncovered during preparation of the vertical section, and bulk sediment samples were radiocarbon dated. In 1999, another section (B), located 130 m east of Section A, was excavated with a backhoe. Here, an 8-m profile of alluvium was sampled for pollen, phytolith, microscopic charcoal, and datable material, and 17 strata were defined by color and sediment composition. Residues of pollen, phytoliths, and microscopic charcoal were concentrated from both Sections A and B (Periman, 2001). A third profile, Section C, located on the north side of the drainage, also was sampled. However, only paleobotanical residues from Sections A and B were used for landscape reconstructions due to their locations within contiguous remnant sediments.

Extended pollen and phytolith counts of 400 grains were obtained for the majority of the samples. Exceptions include phytolith samples 6, 7, 8, and 9 from Section A, which contained very few phytoliths, and pollen samples 16 and 17, near the base of Section B, did not yield sufficient pollen to reach a 400-grain count, and were excluded from further analysis. The pollen record was divided into zones based on cluster analysis. Microscopic charcoal particles were counted in size categories of 5–10 μm , 10–25 μm , 25–50 μm , 50–100 μm , and 100 μm . The number of particles and the geometric mean of the total area of the charcoal particles within each size-class were calculated to determine cross-sectional areas (Delcourt and Delcourt, 1996; Paterson et al., 1987).

In 2000, a stratified sampling strategy was used to identify two suitable surface sample plots within areas of four different vegetation types in order to calculate pollen and vegetation ratios. These areas included grassland, juniper grassland, juniper-pinyon grassland, and cottonwood-willow riparian habitats. After identifying specific areas, random Global Positioning Systems (GPS) coordinates were used within the range of the vegetation polygons to select the southeast corners of the sample plots.

Each sample plot measured 50 meters, delineated within a 10-m grid. Surface soil samples were taken at each 10 m intersection within the grid, for later analysis of pollen, phytoliths, and charcoal. The vegetation was identified, measured, and quantified using the point-centered quarter method (Cottam and Curtis, 1956:455). The mean distances determine mean area and density per unit area. All other absolute measures were computed from the density per unit area figure. Total basal area per acre was obtained by multiplying the number of trees-per-acre by the mean basal-area-per-tree. Absolute values for the number of trees and basal-area-per-unit area of any species were determined by multiplying the relative figures for density and dominance by total trees-per-hectare to determine density, and by total basal-area-per-acre to determine dominance (Cottam and Curtis, 1956:457). The measurements of tree species, shrubs, and ground cover at each quarter of each quadrat, and the percentage of exposed soil were quantified. After laboratory analysis, these data were used to calculate R-value calibrations for each of the vegetation types.

Three-dimensional visual reconstructions of past Rio del Oso landscapes were produced using the tree and shrub density figures derived from the surface pollen R-values. A digital elevation model (DEM) was used as the base map. The Santa Fe National Forest, USDA, provided digital vegetation data, land type, watershed coverages, DEMs, and made available all archaeological site records and field maps. With these data sources, GIS was used to develop map layers of vegetation, land forms, watershed, and archaeological sites. The simulations were created using a visual simulation system that can produce three-dimensional landscape perspective scenes as seen through a 35-mm camera with a 50-mm lens, providing definition of up to 14 basic tree and shrub forms, and control of the crown width, height, and trunk diameter.

Rio Del Oso Geomorphology

The late Holocene alluvium of the lower Rio del Oso consists of 4–8 m of fine sand with intercalated cumulic A-horizon paleosols that are preserved and exposed along a narrow, 5-m-high terrace on the north and south margins of the lower Oso Valley, Section A. Upstream, the valley is narrower and the terrace and late Holocene alluvium have been almost completely removed by erosion, except for an occasional terrace remnant. Radiocarbon dates show that the age of the alluvial deposits range from 4060 B.C. to A.D. 1768. Deposition in the valley slowed after A.D. 1768, the valley fill became incised, and the Rio del Oso channel and floodplain dominated the valley floor, leaving a 5–8-m-high terrace, a remnant of the Holocene fill. Prehistoric river sediments prior to A.D. 1768 have not been identified in any of the remnant alluvium. Slow sediment buildup resulted in the development of seven brown-to-dark grayish-brown cumulic A-horizon paleosols in Section A, separated by zones of fine sand. Some cumulic paleosols also developed in fine sand. Soil B-horizons, visible secondary carbonates, and chemical weathering are not associated with these cumulic soils. The upper boundaries of the paleosols do not form sharp contacts with overlying fluvial sand due to secondary bioturbation of the sediment. The stratigraphic sequence on the southern side of the Rio del Oso Valley appears to have been deposited continuously, without hiatus, in a slowly aggrading floodplain environment. Evidence of erosion in the 6,000+ year sequence, as seen throughout much of the Southwest, is not present.

The alluvium beneath the Holocene terrace on the northern side of lower Oso Valley differs considerably from the sequence on the southern side. In the lower half of Section C, a light yellowish-brown sand includes layers of clay-silt, exhibiting laminae and small crossbeds, with a single 25-cm cumulic A-horizon paleosols. Radiocarbon dates suggest correlation with palaeosol two on the southern side of the valley. The upper meter of the terrace alluvium is marked by numerous rodent burrows. The presence of laminae and crossbeds, poor sorting, and alternating sand and silt layers indicate higher stream velocity in Section C than was seen in sediments on the southern side of the valley. During the late Holocene, the Rio del Oso channel may have been closer to the northern perimeter of the valley than to the southern, similar to the present situation.

These cumulic soils develop during slow deposition of clastic particles by overbank alluvium or spreading flow. Bioturbation mixes the particles with decomposing plant material, and this process produces an organic-rich A-horizon. During periods of slower deposition, thickened, cumulic A-horizon soils form on floodplains and valley floors.

During the late Holocene, in the slowly aggrading environment of the lower Rio del Oso Valley, seven cumulic A-horizon paleosols, ranging from 5 to 62 cm thick, were formed. The paleosols from the top of the terrace down to the base of the exposure were numbered 1–7, and collectively, they make up 41% of the stratigraphic record. The paleosols developed in fine-to-medium quartz sand are dark grayish-brown (in contrast to the pale brown sand between paleosols), and are visually distinct in the outcrop exposures due to this dark color. The paleosols are moderately calcareous, with some carbonates occurring as filaments following small roots. However, the paleosols are no more calcareous than the sandy alluvium that separates them. None of the paleosols exhibit the secondary pedogenic B-horizon characteristics of weakly developed soils, such as blocky structural or chroma differences. The basal and upper boundaries of the paleosols have been obscured by the fill of rodent burrows and other bioturbation. The paleosols in Section B appear to match those of A, although more exposure of the terrace surface would be needed to substantiate this. The Rio del Oso alluvium has been well dated and documented by present standards; greater chrono-logic resolution could identify subtle differences in sedimentation rates for paleosols versus nonpaleosols zones.

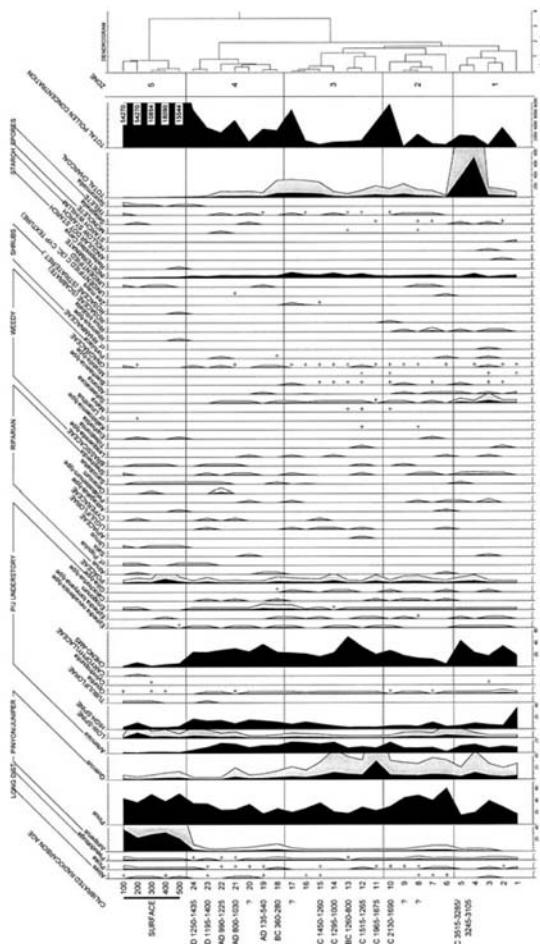
The net sedimentation rate for Section A is 7.65 cm per century, a low value compared to other alluvial records in the region. As a consequence of slow sedimentation, seven cumulic A-horizon soils formed on the valley floor, between 4060 B.C. and A.D. 1768. Two subtrends in the sedimentation occur in the terrace alluvium. The four ages taken from the upper meter of fill show a relatively rapid net sedimentation rate of 14.3 cm per century between A.D. 950 and 1350, twice that of the overall average for the Rio del Oso terrace. This was preceded by a period of much slower deposition that appears to correlate with a period of drought throughout the Southwest. The stratigraphic interval between 1.15 and 2.15 m has a net sedimentation rate of 4.88 cm per century from 1260 B.C. to A.D. 540.

In alluvial Section A, 20 samples were collected for radiocarbon dating, and 10 were collected from Section B. Section A contained three hearth features. The close-interval

samples from Section A were submitted for radiocarbon dating as bulk sediment and the solid organic matter (humus) was dated. The more precise method of accelerator mass spectrometry (AMS) was used for radiocarbon dates in Section B.

Paleobotanical Analyses

The stratified pollen data from Sections A and B were each divided into five pollen zones based on a sum of squares analysis, with four zones representing subsurface pollen deposition, and one zone that originated from five surface control samples. All of the subsurface zones differ substantially in pollen content from the surface samples. Radiocarbon dates placed the base of the subsurface record in Section B at 4060 B.C. and the top of the record at A.D. 1768. Section A dates range from 3515 B.C. at the base to A.D. 1350 at the top of the section. Pollen diagrams were structured to reflect samples of similar vegetation types (Figures 2 and 3).



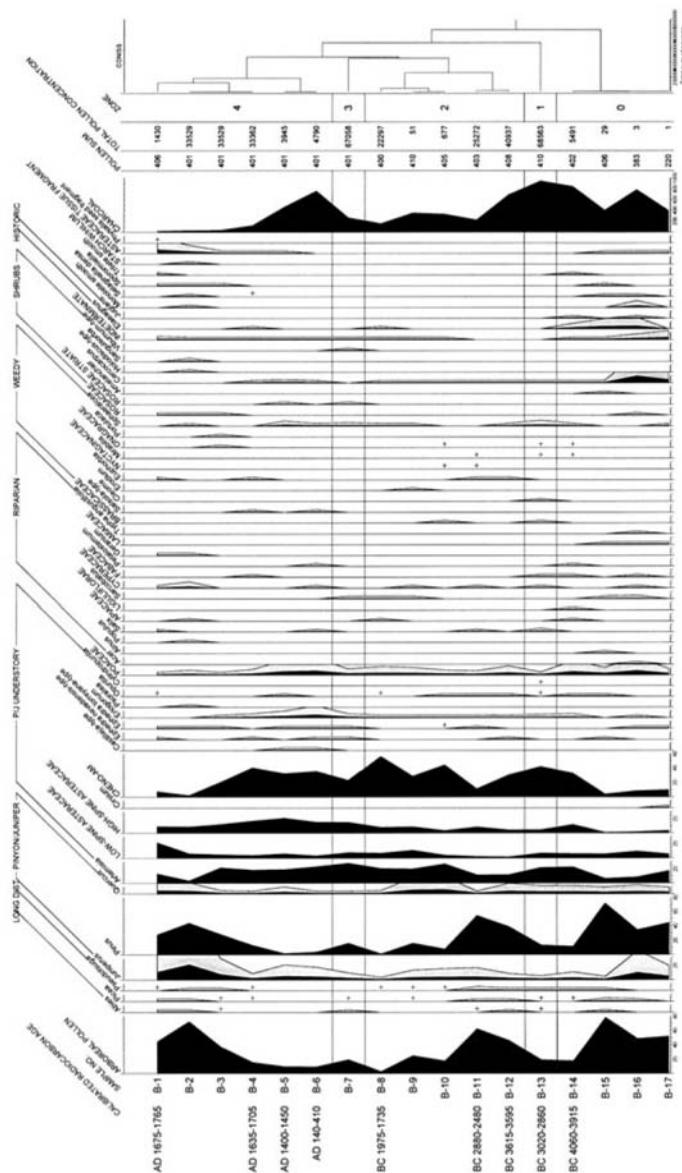


Figure 3. Section B pollen percent diagram, charcoal shown as particle counts.

For the surface samples, pollen concentrations in the samples were approximately 17,000–25,000 pollen grains per cc of sediment. The samples collected in the blue grama (*Bouteloua gracilis*) grassland with sparse oak (*Quercus gambelii*) included moderate quantities of juniper and pine pollen, and elevated quantities of oak, sagebrush (*Artemisia tridentata*), and grass pollen (Figure 4). Pollen concentrations are

considered high when they reach approximately 50,000 pollen grains per cc (Scott Cummings, 2001). A visual comparison shows that the pollen samples from blue grama grasslands are most similar to each other. Using the same quantitative method, samples from juniper with blue grama grass community display similarities to one another; and samples from juniper with pinyon community resemble each other. Samples from the juniper with blue grama grass vegetation type exhibited slightly elevated juniper and pine pollen frequencies compared to the blue grama grassland samples. High-spine Asteraceae, Chenopodium (*Chenopodium* spp.) and amaranth (*Amaranthus* spp.) pollen frequencies were elevated, while sagebrush and grass pollen frequencies were depressed in these areas.

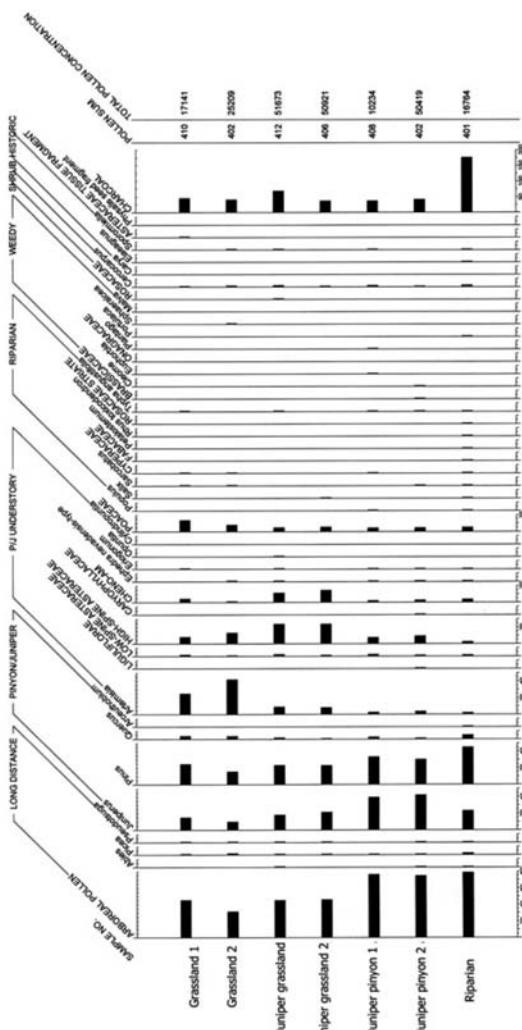


Figure 4. Pollen percent diagram, surface samples, charcoal shown as particle counts.

Rio Del Oso Paleo-Landscape

Archaeological evidence of the early Archaic period (ca. 5500 B.C.–A.D. 600) in the Rio del Oso Valley is scarce on the modern surface. Archaic-period material can be found in the arroyos and tributary drainages bordering the main river channel. Hearths, lithic scatters, and ground-stone artifacts characterize such sites. Exposed Archaic sites provide a chronicle of the geological forces responsible for their burial and the past configuration of the long-concealed landscape. The total area of Archaic sites is approximately 30 ha, distributed among 128 sites, and many of these contain components of Puebloan-period occupations, showing a continuity of place usage. Campsites and stone-tool manufacturing areas make up most of the Archaic period sites.

Palaeobotanical analyses of the Rio del Oso's past vegetation communities revealed that, although oak was more common than it is today, there appears to have been little differentiation between the species identified in current surface plots. This landscape was dominated by grassland, and juniper, ranging from approximately one tree-per-hectare to two trees-per-hectare, while the per-hectare density of pine in the lower drainage was lower. Juniper and pine pollen preserved in the fossil record likely represents an influx of particles from higher-elevation woodland. Oak pollen in the fossil record far exceeds current levels in the lower portions of the drainage, while no mature oak trees were present in the surface-vegetation sample plots. Although R-values were not calculated due to the lack of current oak samples, the consistent prevalence of oak in the fossil-pollen record enabled estimates of oak density in visual models of past Rio del Oso landscapes (Figure 5).

Vegetation change during the Archaic period occurred as fluctuating frequencies of specific taxa, and fires were common. The high occurrence of late Archaic microscopic charcoal (10–25 µm particles) suggests burning of grassland on a landscape scale. These smaller particles represent tiny airborne particles from landscape-level fires, while larger microscopic charcoal particles suggest localized fires (Clark and Royall, 1996; Delcourt and Delcourt, 1996).

During the Puebloan-period occupation of the valley (ca. A.D. 1300–1600), agriculture went from being a supplemental to a dominant food source. The Rio del Oso floodplain was at least 5–8 m higher than current levels, and the frequency of recorded Puebloan archaeological sites in the lower valley is the highest of any period (237 sites in approximately 196 ha). The fossil pollen and phytolith record shows high levels of disturbance-related vegetation, and sedimentation rates increased to 14+ cm per century. In Section A, this increase in floodplain aggradation represents a near-doubling of the sedimentation rate, from a mean rate of 7.65 cm per century. In the upper portion of Section B, from approximately A.D. 1400–1765, the sedimentation rate increased to 16.42 cm per century.

The influence of fire in the Rio del Oso landscape appears to have diminished between approximately A.D. 400 and 1600, counter to paleoenvironmental studies in which large-scale increases in microscopic charcoal suggest human presence and environmental manipulation (Caseldine and Hatton, 1993). Paleobotanical evidence in the Rio del Oso landscape indicates that it was dominated by grassland with scattered juniper rather than woodland before significant impact by horticulturalists. Pine and

juniper density were within the same range as during the Archaic period. The Puebloan landscape consisted of open grasslands, dominated visually by a central village, with fields dispersed across the floodplain, and supplementary agricultural features on Pleistocene terraces. The population of surrounding areas and centralized villages, as well as the creation of fields and frequently used trails, exposed soil to wind and water erosion and likely resulted in greater sedimentation on the valley floor.



Figure 5. Model of the Rio del Oso's Archaic period (1000 B.C.) vegetation patterns. This landscape was dominated by grassland with oak. Juniper ranged from approximately one tree per hectare to two trees per hectare, while the per-hectare density of pine in the lower drainage was lower. The sedimentation rate ranged from 5 cm per century to 8 cm per century.

The Spanish-period landscape (A.D. 1700–1846) differed greatly from its predecessors. Forty-six archaeological sites recorded from this period cover an area of 63 ha. Colonial documents from the 18th century reveal that by the 1730s, Spanish settlers lived in the upper portion of the Rio del Oso canyon (Wozniak et al., 1992). These communities imposed an Old World model of land use on the Rio del Oso landscape. Colonists depended on subsistence farming, and introduced ranching and livestock, including horses, cattle, sheep, and goats. The Spanish-period landscape consisted of rectangular fields with straight lines of construction, grazing herds of large, domestic ungulates, and encroaching juniper. This created a composite landscape built upon the ruins of Pueblo land-use systems.

In the lower Rio del Oso Valley, Spanish-period juniper density increased to 10+ trees per hectare, while disturbance vegetation, once used by Puebloan peoples, declined to the lowest levels observed during the past 6,000 years, and fire levels fell sharply (Figure 6). Alluvial aggradation decreased in the lower portion of the valley, and terminated after A.D. 1765. The valley floor had been stable, with steady sediment accumulation during the previous 7,000 years; however, Spanish-period erosion dramatically altered the Rio del Oso ecosystem. With the arrival of Spanish livestock in the 17th century, down-cutting of the floodplain likely began. What had been a rich,

meandering stream and riparian area became a hydrological system of braided channels, characterized by a network of constantly shifting, low-sinuosity courses. Such systems are characterized by a network of channels without clearly defined over-bank deposition. This is evident in the laminated, cross-laminated, and desiccation-cracked fine sands and silts from the lower strata of Sections B and C. The remnants of the former floodplain are now arid, supporting mostly juniper and cholla.



Figure 6. Model of Rio del Oso vegetation patterns at the end of the Puebloan period (A.D. 1500). Settlement and sedimentation increased to as high as 16 cm per century by the time the pueblos were abandoned and tree densities were at approximately four trees per hectare. Puebloan agricultural features are modeled as light-colored polygons in mid-ground. Massive erosion began with the introduction of Spanish livestock and village development in the valley. By 1750, juniper densities were already at more than 10 trees-per-hectare.

There is a lack of data documenting ecosystem change in the Rio del Oso landscape after the Mexican American War of 1846–1848. Historic paleobotanical records of Southwestern landscapes primarily consist of information from early expeditions of the U.S. Army. Gaps in this vegetation history may be filled by research of sediments from relict paleochannels, peat bogs, or other sources. Although sedimentation increased in the Rio del Oso Valley before Spanish settlement, this aggradation was followed by erosion and the expansion of the New Mexican cattle industry in the 1870s. Overgrazing caused massive decreases in vegetation cover, plant vigor, and the suppression of natural fires. The Rio del Oso landscape had already been exposed and eroded by the time of the 1890s Southwestern drought, and vulnerability to the forces of wind and water erosion only increased. Reduced upland and valley plant cover led to an estimated loss of at least 6 inches of vital topsoil (Wildeman and Brock, 2000:19–22). The floodplain, stripped of vegetation by grazing, became a labyrinth of arroyos when seasonal rains returned. Presently, the Rio del Oso landscape has juniper densities of 50 trees per hectare and pine densities of 15 trees per hectare (Figure 7). Many watersheds affected by this 1890s ecological crash have never recovered (Wildeman and Brock, 2000:19–22).



Figure 7. Today the Rio del Oso landscape exhibits juniper densities of 80 trees per hectare, while pine is now near 15 trees per hectare. Most of the alluvium accumulated before 1750 has been removed by broad-scale erosion to a depth 6–8 m below the A.D. 1500 surface, and Pleistocene gravels lay exposed in a braided stream channel.

CONCLUSION

This Rio del Oso landscape history incorporated interdisciplinary environmental sciences vital to ecological analyses, including paleoenvironmental data, geoarchaeology, paleoecology, and spatial modeling. Geomorphological and paleobotanical analyses of the sedimentary processes, vegetational composition, and structure of past landscapes, provided further insight into the role of humans in the creation of that structure. Various cultures inhabited the Rio del Oso, affecting landscape through the use of applied fire, plant manipulation, agriculture, transportation, and shelter construction. The Rio del Oso was imprinted by hunter-gatherers, Pueblo settlements, and forces of the world economic system with European colonization. Archaeological information was instrumental in providing direct spatial and temporal understanding of human influences on the evolution of the Rio del Oso landscape. The nature and extent of anthropogenic landscape change during the past 7,000 years in the Rio del Oso Valley may reveal how past human activities continue to influence ecosystem characteristics, and provides perspective regarding present and future environmental conditions.

KEYWORDS

- **Acéquias**
- **Anthropogenic landscapes**
- **Ecosystems**
- **Rio del Oso**
- **Taxon's R-value**

ACKNOWLEDGMENTS

I would like to thank Dr. Charles A.I. French, Dr. Gianna Ayala, and the geoarchaeology graduate students at the McBurney Laboratory of Geoarchaeology, Department of Archaeology, University of Cambridge, for inviting me to participate in this publication. I thank Dr. Stephen A. Hall for his analysis and expert advice concerning geomorphology of the Rio del Oso. Thanks to Dr. Linda Scott Cummings for her paleobotanical analysis. I thank Dr. Deborah Finch and the Ecology, Diversity, and Sustainability of Soil, Plant, and Human Resources of the Rio Grande Basin research program for assistance in funding this research, and Project Leader Dr. Joseph A. Tainter. My gratitude goes to Karin Periman for her generous editorial assistance.

Chapter 17

East Asian Monsoon and Paleoclimatic Data Analysis

J. Guiot, Hai Bin Wu, Wen Ying Jiang, and Yun Li Luo

INTRODUCTION

First we review several syntheses of paleodata (pollen, lake-levels) showing the climate variations in China and Mongolia from the last glacial maximum to present and in particular the precipitation increase at mid Holocene related to enhanced monsoon. All these results concur to a much enhanced monsoon on most of China during the first half of the Holocene. Second we present, in some details, a temporal study of a core (Lake Bayanchagan, Inner Mongolia) located in an arid region at the edge of the present East Asian Monsoon (EAM) influence and then sensitive to climatic change. This study involves pollen data together with other macro-remains and stable isotope curve to obtain a robust climate reconstruction. This study shows a long wet period between 11,000 and 5,000 years BP divided in two parts, a warmer one from 11,000 and 8,000 (marked by large evapotranspiration) and a cooler one more favorable to forest expansion. Third, we present a spatial study based on pollen data only and covering all China and Mongolia at 6,000 years BP, but using a mechanistic modeling approach, in an inverse mode. It has the advantage to take into account environmental context different from the present one (lower atmospheric CO₂, different seasonality). This study shows temperature generally cooler than present one in southern China, but a significant warming was found over Mongolia, and a slightly higher in northeast China. Precipitation was generally higher than today in southern, northeast China, and northern Mongolia, but lower or similar to today in northwest China and north China. Enhanced EAM was then found in the southern half of China and in northeast China.

The past 21,000 years are a very interesting time period as it contains two extreme states of the climate. The Last Glacial maximum (LGM, 21,000 years BP) is a cold and generally dry period driven by enlarged ice sheets and low atmospheric CO₂. The mid-Holocene period (6,000 years BP), generally warmer and wetter than the present one, is considered as orbital forced period with perihelion in northern summer/autumn and greater-than-present axial tilt (Berger, 1978) but free of major ice-sheet and relatively high CO₂ (taking as reference the pre-industrial present time). These two periods have been chosen as key time periods by the Palaeoclimate Modeling Intercomparaison Project, PMIP (Joussaume and Taylor, 1995). The mid-Holocene, with its high summer insolation, is a period of high land-sea contrast and consequently enhanced monsoon (Braconnot et al., 2002). It is of particular interest for climate modelers to test their simulations through palaeodata from the monsoonal regions.

The EAM is one of the most active components of the global climate system, influencing a large area of China and its surrounding countries. In China and surrounding countries, a megathermal period was reconstructed from 9,500 to 4,000 yrs ago (Shi et al., 1993). However, many recent studies have shown that Holocene climatic changes were asynchronous across China (An, 2000; An et al., 2006; He et al., 2004). The Holocene optimum was defined as EAM precipitation maximum, occurring ca. 11,000–9,000 years ago in northeastern China, 11,000–8,000 years ago in north-central and northern east-central China, 8,000–6,000 years ago in central China, and ca. 3,500 years ago in southern China (An, 2000). The reason for debate on Holocene climatic variations is that complexity of the EAM, and different responses of environmental proxies to climatic changes (Wang et al., 2003; Wei and Gasse, 1999). Therefore, more precisely dated palaeo-records and improved quantitative reconstruction are required to provide quantitative insights into the processes of climatic changes, and their links to the EAM.

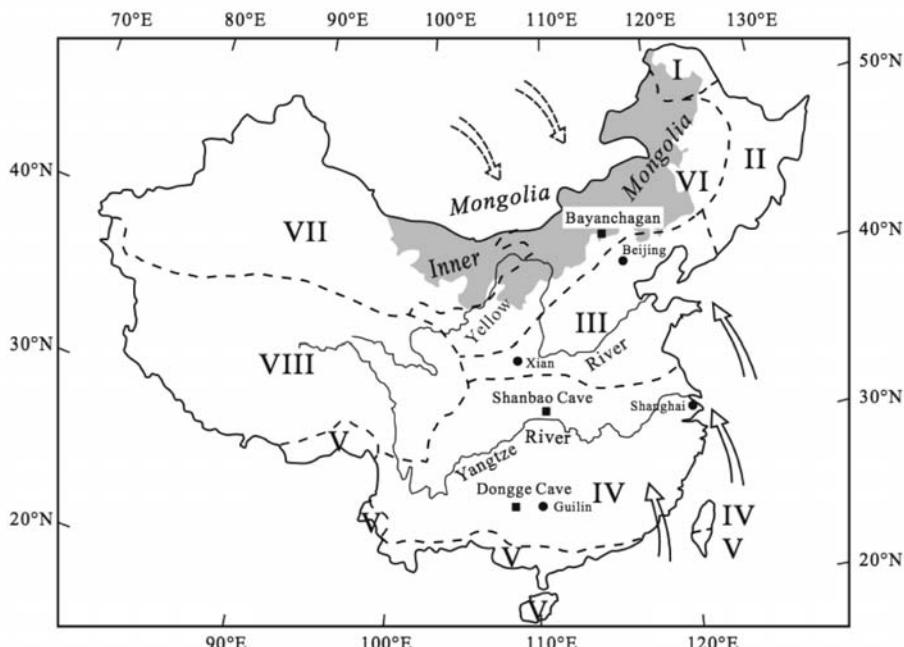


Figure 1. Location of studied site and modern vegetation zones in China (after Jiang et al., 2006). I, Cold-temperate conifer forest; II, Temperate mixed conifer-broadleaved forest; III, Warm-temperate broadleaved deciduous forest; IV, Subtropical evergreen broadleaved forest; V, Tropical rainforest and seasonal rainforest; VI, Steppe; VII, Desert; VIII, Tibet-Qinghai cold, and highland vegetation. The dashed and solid arrows indicate winter monsoon and the dominant direction of the summer monsoon precipitation belt, respectively.

The goal of this chapter is threefold. First we explore the temporal variability of a record located in a sensitive region at the northern edge of the EAM using a multiproxy

approach. Second we explore the spatial variability of the Chinese climate at 6 ka BP, when EAM is assumed to be the strongest. Third we illustrate a new methodology of climate reconstruction based on vegetation model inversion.

The temporal study is based on a core sampled in Lake Bayanchagan (Inner Mongolia) (Jiang et al., 2006) (Figure 1). This region is particularly sensitive to climate variations as it is located at the edge of the present EAM. Their results suggest that this region was dominated by steppe vegetation throughout the Holocene, except for the period 9,200 to 6,700 years BP, when forest patches were relatively common. This period can then be correlated to enhanced EAM. But these findings need to be confirmed by a multiproxy analysis. We will synthesize in the first part of this chapter an statistical approach based, in addition to pollen, on isotopic data and concentration of a green algae species (Jiang et al., 2008). This study will focus on the timing of this enhanced EAM period.

This approach based on detailed time series in a sensitive region will be completed by a spatial analysis based also on pollen data but done with the newest tools involving a process model able to relate vegetation and climatic variations (Luo et al., 2008). The strong feature of this approach is to be able to take into account the large differences existing between present and mid-Holocene conditions as (i) climate seasonality, possibly resulting in lack of modern analogues, or (ii) atmospheric CO₂ close to pre-industrial concentration but significantly lower than the present one. This spatial analysis will be first replaced in the context of previously published data syntheses at the sub-continental scale.

Data Syntheses

An interesting story has been depicted by Ren and Beug (2002) in the northern half of China (North of Yangtze River) for the whole Holocene. Forests generally expanded in the early Holocene times, reaching their maximum at 6 or 4 ka BP, with a maximum in Central China, and then regressed during the Late Holocene. An exception was found for Northeast China where the maximum development of forest occurred during the last 4,000 or 2,000 years. They concluded that, if the EAM enhancement seems to be responsible of the forest expansion at the beginning of the Holocene, disturbance by human activities may be responsible of the forest decline after 6 ka BP.

This picture was completed by the study of Yu et al. (1998) who analyzed the vegetation variations at the biome level for the whole China, but restricted at the 6 ka BP period. In eastern China at 6 ka BP, forest shifted northwards, with broadleaved evergreen forest extended about 300 km and temperate deciduous forest about 500–600 km beyond their present northern limit. In Northwestern China, the area of desert and steppe vegetation was reduced as compared to present. They concluded that these shifts were likely a response to enhanced Asian monsoon.

Lake levels data are less susceptible to be influenced by human disturbances. Yu et al. (2003) proposed a story of the lake levels since the Last Glacial Maximum (LGM, about 21 ka BP). This compilation showed LGM conditions much drier than today in eastern China but somewhat wetter in western China. These east-west differential patterns of climate conditions were completely different from the modern dry-wet

conditions with a north-south opposition. During the Holocene, at the mid-Holocene, both regions were wetter than present. Modern dry conditions returned after 5–4 ka BP depending on the region. Then if humans played a role in the forest decline in the Late Holocene, they simply accentuated a climatic trend. Atmospheric general circulation models (AGCM) coupled with land surface process model showed that the dry conditions in eastern China resulted from less summer precipitation due to the Pacific Subtropical High occupying eastern China and the decline in the summer monsoon.

More at north, in Mongolia, Tarasov et al. (1999) reconstructed, from pollen, warmer, and wetter at 6 ka BP conditions for the northern part of the country, in agreement with higher lake levels. In the central part of the country, warmer and drier conditions prevailed (inferred from pollen, no lake data being available). But these dry conditions are likely due to more evapotranspiration and not necessarily to less precipitation.

A Multi-Proxy Technique to Reconstruct Climate Time Variability in Inner Mongolia

The syntheses presented above are based either on pollen data or lake levels data. A multi-proxy approach is now presented to reinforce and precise these results (Jiang et al., 2008). It is based on a record taken from Lake Bayanchagan (115.21° E, 41.65° N, 1355 m a.s.l, Figure 1) in Inner Mongolia, which is today almost completely dry due to anthropogenic water use, with only small patches of shallow water maintained by summer rain. It is situated at the current northern edge of the summer monsoon. The mean annual temperature in the area is about 3°C , and total annual precipitation is 300–400 mm. About 70% of the precipitation occurs during the summer. The data used are pollen taxa counted for 90 pollen assemblages and 2,066 surface samples. The taxa are grouped into 17 plant functional types (e.g., boreal evergreen conifers, steppics, grass, temperate summergreen trees, etc.) to reduce the number of variables and also to consider together taxa which respond in the same way to climatic variations. These plant functional types (PFT) are used to reconstruct climate by the modern analogue method (PFT-MAT) proposed by Davis et al. (2003) and Jiang et al. (2006). The climatic variables considered are the temperature of the coldest month (MTCO), the temperature of the warmest month (MTWA), the annual precipitation (MAP), the ratio actual evapotranspiration over potential evapotranspiration (α). These variables are calculated by linear interpolation from meteorological stations (Jiang et al., 2006) and α is obtained from monthly temperature, precipitation, and sunshine variables using the Priestley-Taylor equation (Prentice et al., 1992).

To these proxies, are added total pollen concentrations, Pediastrum (a green algae that indicate shallow lake water) concentrations and $\delta^{18}\text{O}$ of authigenic carbonate, that is on the $<40\ \mu\text{m}$ fraction (Jiang et al., 2008). These three proxies show a similar general pattern during the Holocene (Figure 2a). Before 11,000 cal years BP, there is no Pediastrum in the lake. Pollen concentrations are lower than 2×10^5 grains/ml. All $\delta^{18}\text{O}$ values of authigenic carbonate are between -3 and -1‰ VPDB. Similar values are found after 5 ka BP and in between, there is high concentrations of pollen and Pediastrum and low $\delta^{18}\text{O}$ values. As Jiang et al. (2008) have shown that these three variables are controlled by balance between precipitation and evaporation, they can be synthetized a common signal, given here by their first principal component (Figure 2c).

Jiang et al. (2008) have used PFT-MAT constrained by the first principal component PC1 (Figure 2c) as an indicator of α , a variable directly related to the water stress. This constrained analysis has already been proposed with different proxies by Cheddadi et al. (1996); Guiot et al. (1993); Magny et al. (2001); Seret et al. (1992). For each fossil pollen spectrum, analogues were selected from the modern pollen spectra dataset subject to a broad consistency requirement according to α values. If we note the difference $\delta\alpha$ between α of the analogue and the modern α_o at the lake (56%), only the analogues i with a $\delta\alpha_i$ compatible with PC1 at time t , denoted C_t , were retained. This compatibility is defined as follows:

$$\begin{aligned} C_t &> 2 \text{ and } \delta\alpha_i > 10\% \\ C_t &< -2 \text{ and } \delta\alpha_i < -10\% \\ -2 \leq C_t \leq 2 &\text{ and } -10\% \leq \delta\alpha_i \leq 10\% \end{aligned} \quad (1)$$

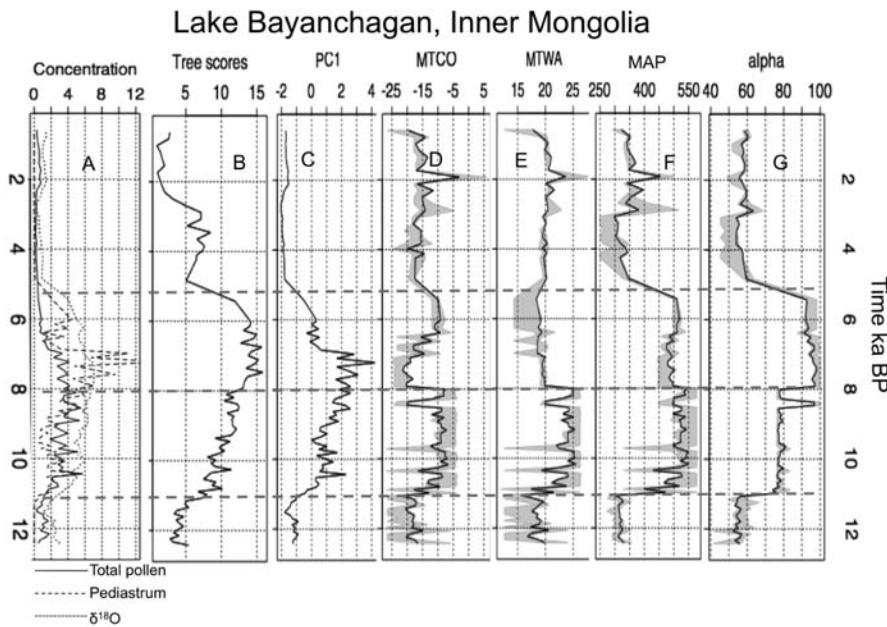


Figure 2. Comparison of a few proxies and the climatic reconstructions in Lake Bayanchagan (Inner Mongolia, China). (A) total pollen (10^5 grains ml^{-1}) and Pediastrum concentration (10^4 grains ml^{-1}), $\delta^{18}\text{O}$ of authigenic carbonate multiplied by -1 ; (B) tree scores, that is sum of the square root of the arboreal taxa percentages; (C) first principal component of the three proxies of (A); (D) mean temperature of the coldest month reconstruction; (E) mean temperature of the warmest month reconstruction; (F) total annual precipitation; (G) α , the ratio between actual, and equilibrium evapotranspiration. The climate reconstructions are represented with the uncertainties, given by the range of the analogues. After Jiang et al. (2008).

Figure 2 shows the results obtained for the Lake Bayanchagan core: the reconstructed climatic variables are compared with the constraint PC1 and the scores of the arboreal pollen taxa (Figure 2b). This enables one to question the direct relationship

often proposed between an increase in tree components of pollen assemblages and a warmer and wetter climate (Liu et al., 2002; Shi et al., 1993; Xiao et al., 2004). So, the highest tree scores of trees during the Holocene in Lake Bayanchagan occurred between 8,000 and 5,500 cal years BP (Figure 2b). However, the peak period of trees was not in phase with the warmest and wettest climate reconstructed between 11,000 and 8,000 cal years BP (Figure 2d–f), suggesting that a single climatic variable is not the triggering factor. In contrast, variations in tree components and α were consistent (Figures 2b and g). α is an integrated measure of annual amount of growth-limiting drought stress on plants related to both temperature and precipitation, and is one of primary factors influencing vegetation distributions (Prentice et al., 1992). The similarity in tree components and α variations inferred from our study indicates that it is also the main controlling factor for growth of trees over the Holocene in Inner Mongolia. α does not reach its maximum before 8 ka BP even if MAP is maximum because evaporation is too strong. The water stress is minimum only when temperature has decreased by a few degrees. The most favorable period for forest development is then between 8 ka and 5 ka BP.

The MAP record during the Holocene at Lake Bayanchagan is similar to $\delta^{18}\text{O}$ records of stalagmite calcite from Dongge Cave and Shanbao Cave (Figure 1) in EAM regions (Dykoski et al., 2005; Shao et al., 2006). Shifts in $\delta^{18}\text{O}$ values of the stalagmite from the cave largely reflect changes in $\delta^{18}\text{O}$ values of meteoric precipitation at the site, which in turn relates to changes in the amount of precipitation. The $\delta^{18}\text{O}$ results show that monsoon precipitation increased dramatically at the start of the Holocene (~11,500 cal years BP) and remained high for ~6,000 cal years BP (Dykoski et al., 2005). This timing is consistent with other paleoclimatic records in EAM regions (Zhou et al., 2004, 2005). Both the Lake Bayanchagan data and stalagmite $\delta^{18}\text{O}$ records from Dongge Cave and Shanbao Cave show the termination of monsoon precipitation maximum was abrupt between 6,000 and 4,400 cal years BP.

The second warm and humid period at Lake Bayanchagan centered at 6,000 cal years BP. This event was characterized by increased MTCO, decreased MTWA and high precipitation (Figures 2d–f). These results agree with a marked increase in winter temperatures across eastern China at 6,000 cal years BP estimated from pollen data (Yu et al., 1998) and simulated by climatic model (Yu et al., 2003). The short-term cold event between 8,500 and 8,300 cal years BP was characterized by decreases in both winter and summer temperature (Figures 2d–f). Even if such event has been recorded in several places and in particular in the GRIP and GISP2 records (Alley et al., 1997; Rohling and Palike, 2005), it cannot be considered as significant in our reconstruction, as several such peaks are reconstructed during the Holocene.

An Inverse Modeling Technique to Reconstruct Climate Spatial Variability in China

Multi-proxy approach is a way to produce robust paleoclimatic information but, as it is based on modern data using a statistical approach, it does not solve all the problems. The reconstruction methods are built upon the assumption that plant-climate interactions remain the same through time, and implicitly assume that these interactions are independent of changes in atmospheric CO₂. This assumption may lead to a

considerable bias, as polar ice core records show that the atmospheric CO₂ concentration has fluctuated significantly over the past (EPICA, 2004). At the same time, a number of physiological and palaeoecological studies (Cowling and Sykes, 1999; Farquhar, 1997; Jolly and Haxeltine, 1997) have shown that plant-climate interactions are sensitive to atmospheric CO₂ concentration. Therefore, the use of mechanistic vegetation models has been proposed to deal with these problems (Guiot et al., 2000). Wu et al. (2007) have improved the approach based on the BIOME4 model to provide better spatial and quantitative climate estimates from pollen records and correct for CO₂ bias to pollen-based climate reconstructions in Eurasia and Africa. The same method is quickly presented here for Eastern Asian data.

MATERIALS AND METHODS

The pollen data used have been compiled by the BIOME6000 project (Prentice and Jolly, 2000) for three key periods: 0 k, 6 k, and 21 ka BP to classify pollen assemblages into a set of vegetation types. For the study described here, a subset containing 601 sample sites for 0 ka BP and 116 sites for 6 ka BP from China and Mongolia were used (MCPD, 2000, 2001; Tarasov et al., 1998). The selection of the 6 ka BP samples is based following the BIOME6000 convention. Among them, 84 sites have a good age control, that is either with at least two dates encompassing 6 ka BP at less than 2,000 years distance.

BIOME4 is a physiological-process global vegetation model, with a photosynthesis scheme that simulates the response of plants to changed atmospheric CO₂ and by accounting for the effects of CO₂ on net assimilation, stomatal conductance, leaf area index and ecosystem water balance. It is driven by monthly temperature, precipitation, sunshine, by absolute minimum temperature, CO₂ concentration and soil texture. The principle of the model inversion is to estimate the input to BIOME4, the monthly climate, given that we know some information related to the output of the model, biome scores derived from pollen in our case (Prentice et al., 1996). This inversion, which uses a Monte-Carlo-Markov-Chain algorithm to explore possible combinations of climate parameters, allows an assessment of the probability of different anomalies, and therefore the investigation of different scenarios which may result in similar vegetation pattern. The procedure is described in Wu et al. (2007). As Guiot et al. (2000), they showed that several solutions were possible for the LGM climate in Western Europe where a mixture of steppes and tundra existed. As these biomes have no clear analogues today, reconstructions based on statistical methods will tend to choose the least poor match or fail to find a real match. With the inverse modeling, Wu et al. (2007) showed that a climate significantly warmer than inferred with modern CO₂ levels was the most probable. The overestimation of MTCO anomalies was about 10°C. Moreover uncertainties were also underestimated with the statistical methods.

Validation

We present here an analysis of Chinese mid-Holocene data (Luo et al., 2008). In a first step, the ability of this inversion scheme to reproduce the modern climate of China is evaluated, using the 601 modern spectra available. The statistical squared

correlations (R^2) between actual and reconstructed climate variables at the sample sites are presented in Figure 3. These R^2 are very large, generally above 0.67, except for MTWA which then does not appear to be a key factor to explain the modern vegetation distribution in China. The straight line between estimates and observations is expected to have an intercept of 0 and a slope of 1. The slope is slightly biased for MTWA, GDD and MAP. The intercepts are biased for MTCO, MTWA, and MAT, showing a tendency to overestimate the cold climates. There is also large error in estimating MAP and α in cold desert sites of the Tibet Plateau, where α below 60% are frequently estimated below 20%, that is values typical of warmer deserts.

The 6,000 Years BP Climate

For the 6 ka BP period, the atmospheric CO_2 concentration is set to 270 ppmv (EPICA, 2004). The results (MAP, MAT, α) are presented as maps of anomalies versus present climate (Figure 4). Large circles indicate significant differences from the modern values. The results show that, in most of the sites at 6 ka BP, the changes in precipitation and α were significantly different from modern values, while most of temperature changes are not. This is due to the larger uncertainty on the reconstructed temperature, which indicates a larger tolerance range of the vegetation to thermal variables while hydrological variables were more limiting factors. Annual temperature were generally lower than present one in southern China, but a significant warming was found over Mongolia, and a slight warming in northeast China.

Validation of the inverse modelling method on modern samples

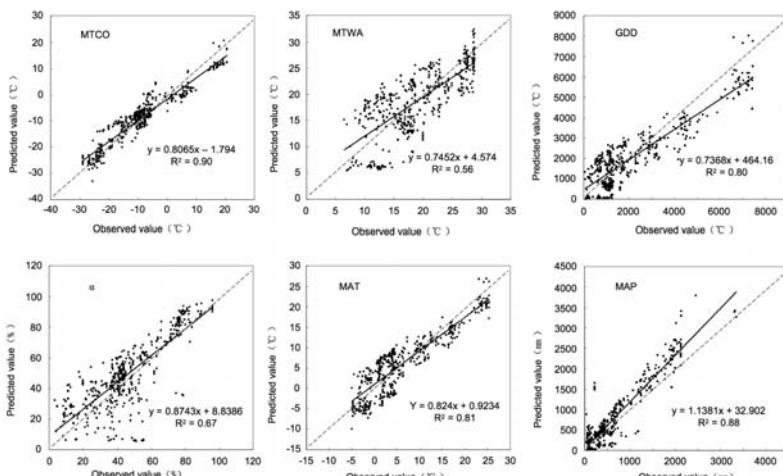
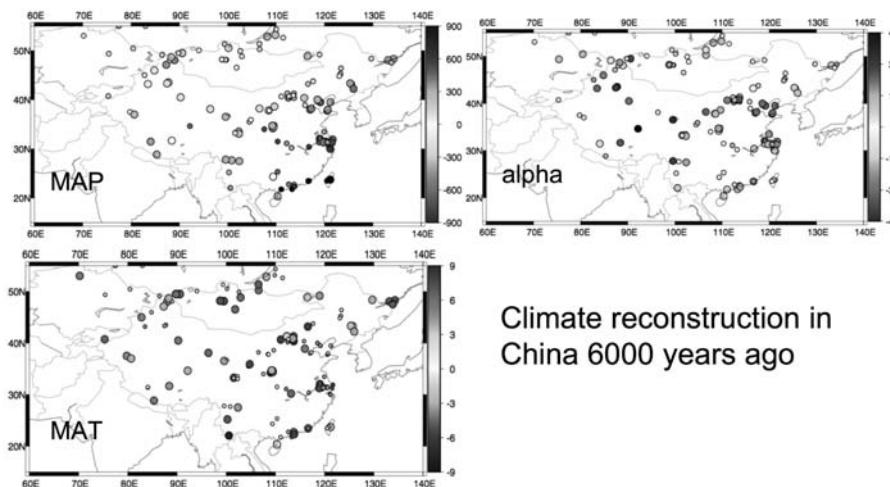


Figure 3. Validation of the inverse BIOME4 modeling on the 601 samples of the modern database of China and Eurasia. The six reconstructed variables are compared to the observed climate: MTCO (mean temperature of the coldest month reconstruction), present value = -20°C , (MTWA) mean temperature of the warmest month reconstruction, present value = 17°C , (GDD) growing degree days above 5°C , present value = 1500° days, (α), the ratio between actual and equilibrium evapotranspiration, present value = 30%, (MAT) mean annual temperature, present value = 3°C , (MAP) total annual precipitation, present value = 350 mm.



Climate reconstruction in
China 6000 years ago

Figure 4. Reconstruction of the climate in China 6,000 years ago using inverse modeling method: (α), the ratio between actual and equilibrium evapotranspiration in %, (MAT) mean annual temperature $^{\circ}\text{C}$, (MAP) total annual precipitation in mm. All the values are given in departures from present climate. Large circles indicate high significance levels (95%), small circles indicate no significance.

Hydrological variables have a much more coherent distribution. MAP was generally higher than today in Southern, Northeast China, and Northern Mongolia, but lower or similar to today in northwest China and north China. Alpha was considerably higher than today in North China, and slightly higher than present in northeast China. In contrast, drier conditions are shown in Northwest China and Mongolia.

Lake Bayanchagan is situated in a zone where most of the sites had a positive anomaly of MAP whereas a few ones had a negative one. This is broadly consistent with the reconstruction of Figure 2e where MAP was found 200 mm higher than at present. The anomaly of α for this zone is significantly positive, between +15 and 30% in agreement with Figure 2f where α was found 30% higher than at present. For these two variables, Lake Bayanchagan reconstruction provide values at the upper limit of the inverse modeling. The MAT appears also higher than at present, in good agreement with the reconstruction of Figures 2c–d. The reconstructions based on the inverse modeling are then approximately consistent with the Lake Bayanchagan, at least for the majority of surrounding sites, but the multiproxy statistical approach infers values at the wetter limit of the inverse modeling. When compared to Tarasov et al. (1999), Figure 4 shows also wetter and warmer conditions on northern Mongolia and warmer and drier conditions in the central part of the country.

DISCUSSION

We have explored the temporal variability of the Lake Bayanchagan record located in a sensitive region at the northern edge of the EAM. The use of a multiproxy approach coupled with robust statistics have enlightened the complexity of the climatic signal. A key problem in this respect is the timing of the monsoon enhancement. Monsoon

increase is translated in terms of increased precipitation. Then the period of maximum EAM occurred between 10.5 and 8 ka BP. A too rapid interpretation of the tree pollen curve should put this maximum between 8 and 5 ka BP. It is clear that precipitation was higher than at present time across the two periods. But, extension of forest depends as well of temperature than precipitation, and our quantitative evaluation of several proxies show a more complex behavior than Dongge and Shanbao cave series. This may also be due to spatial differences, the caves being located at much more lower latitudes than the lake (Fig. 1). This is confirmed by the post-5 ka BP decrease in the lake records where precipitation returns to the Late-Glacial level, while in the cave record, $\delta^{18}\text{O}$ remains at an intermediate level. This might be explained by a rapid northward advance of the northern limit of the summer monsoon at 11.5 ka BP (beyond 41° N) followed by a slow retreat, falling back south of Lake Bayanchagan by 5 ka, while the caves, being further south, remain under the monsoon influence. This illustrates well that Lake Bayanchagan, at the northern edge of the EAM zone, is a sensitive record of the monsoon signal.

A second implication concerns the physical mechanisms. EAM enhancement is related to summer radiation which is maximum at 9 ka BP and rapidly decreases to be at 6 ka BP on the same level than at 12 ka BP (An, 2000; Berger, 1978). When a large number of climate model simulations are compared (Braconnot et al., 2002), a robust feature is that the extension of the monsoon is related to the Eurasian continent warming. This might explain why the 8–5 ka BP period is characterized by a slight decrease of EAM accompanied by a decrease of temperature more marked in this northern lake than in lower latitudes. Maximum temperature of the warmest month falls by 5°C at 8 ka BP (but keep a level above the present one), which shows a mitigation of the impact on vegetation of the monsoon weakening by a sharp reduction of the evapotranspiration.

The analysis of the spatial variability of the Chinese climate at 6 ka BP—even if the 6 ka period is not the period of maximum monsoon enhancement—permits to replace the timing found for Lake Bayanchagan in a larger context. Figure 3 shows that some sites in the region of this lake have already a reduced precipitation, while α , which represents the water availability for vegetation, is still higher than at present. This is still a period favorable to maintain forest, even with a precipitation reduction. Annual temperature distribution shows higher values than at present time in northern China, but lower in southern and central China where monsoon had still a higher influence. This illustrates well the fact that northern China was more at that time under the influence of the Eurasian continent while the rest of China was under still the influence of the ocean through the Pacific Subtropical High.

A last point is the use of a new methodology of climate reconstruction based on vegetation model inversion. As already mentioned, this mechanistic model offers the possibility to escape from too constraining modern conditions as high atmospheric CO_2 concentrations or a climate seasonality different from modern one (in relation with isolation). The climatic maps obtained for 6 ka BP confirmed previous results based on modern analogues, likely because CO_2 concentration is sufficiently high. Wu et al. (2007) have shown that, for the Last Glacial Maximum conditions, biases are

introduced by the fact that CO₂ is sufficiently low to have limited vegetation productivity in a comparable amplitude than climate change.

CONCLUSION

Various syntheses have been done on Chinese paleodata using various methods. All converges to reconstruct intensification of EAM in China at 6 ka BP, especially on eastern part of China. Northern China cores indicate an intensified monsoon between 10 and 5 ka BP. After 8 ka BP, a cooler climate induced a less strong water stress, favoring the largest extend of the forest. This two-step division of mid-Holocene has been possible, thanks to a multi-proxy approach enabling more robust inference. Nevertheless, all approach involving modern analogues has its own limit when extrapolation is done on periods with characteristics very different from the present reference period. Then the use of mechanistic models in an inverse mode enables one to control the effect of external variables, such as atmospheric CO₂.

The results based on inverse modeling are coherent with the previous syntheses. They show that a pattern of higher precipitation is clear on eastern half of China. On Western part of China, the situation is less contrasted with higher precipitation on southwest and lower on northwest. The Eastern China situation is related to enhanced summer monsoon associated with the Pacific Subtropical High bringing warm and moist marine air from the West Pacific Ocean to Eastern China. The situation of South Western China can be related to the Indian summer monsoon bringing most marine air from the Indian Ocean to Southern Tibetan Plateau and Southern China lowlands. North Western regions are sheltered from these monsoon changes by the Tibetan Plateau and are dominated by the Westerlies and Asian winter monsoon. Increased land-sea contrast due to higher summer insolation at mid-Holocene will then influence more strongly Eastern China.

KEYWORDS

- Atmospheric general circulation models (AGCM)
- Last Glacial maximum (LGM)
- Multi-proxy approach
- Plant functional types (PFT)
- Steppe vegetation

ACKNOWLEDGMENTS

This research has been partly funded by a grant of the French Ministry of Research to two authors, the 5th PCRD EU project MOTIF (EVK2-CT-2002-00153), by the European Science Foundation, EUROCLIMATE/DECVEG and the French ANR PICC.

Edited by: Ryuji Tada.

Chapter 18

Evolutionary Divergence Times in the Annonaceae

Yvonne C.F. Su and Richard M.K. Saunders

INTRODUCTION

Phylogenetic analyzes of the Annonaceae consistently identify four clades: a basal clade consisting of *Anaxagorea*, and a small “ambavioid” clade that is sister to two main clades, the “long branch clade” (LBC) and “short branch clade” (SBC). Divergence times in the family have previously been estimated using non-parametric rate smoothing (NPRS) and penalized likelihood (PL). Here we use an uncorrelated log-normal (UCLD) relaxed molecular clock in BEAST to estimate diversification times of the main clades within the family with a focus on the Asian genus *Pseuduvaria* within the SBC. Two fossil calibration points are applied, including the first use of the recently discovered Annonaceae fossil *Futabanthus*. The taxonomy and morphology of *Pseuduvaria* have been well documented, although no previous dating or biogeographical studies have been undertaken. Ancestral areas at internal nodes within *Pseuduvaria* are determined using dispersal-vicariance analysis (DIVA) and weighted ancestral area analysis (WAAA).

The divergence times of the main clades within the Annonaceae were found to deviate slightly from previous estimates that used different calibration points and dating methods. In particular, our estimate for the SBC crown (55.226.9 Mya) is much younger than previous estimates ($62.553.1 \pm 3.6$ Mya and ca. 58.76 Mya). Early diversification of *Pseuduvaria* was estimated to have occurred 15–8 Mya, possibly associated with the “mid-Miocene climatic optimum.” *Pseuduvaria* is inferred to have originated in Sundaland in the late Miocene, ca. 8 Mya; subsequent migration events were predominantly eastwards towards New Guinea and Australia, although several migratory reversals are also postulated. Speciation of *Pseuduvaria* within New Guinea may have occurred after ca. 6.5 Mya, possibly coinciding with the formation of the Central Range orogeny from ca. 8 Mya.

Our divergence time estimates within the Annonaceae are likely to be more precise as we used a UCLD clock model and calibrated the phylogeny using new fossil evidence. *Pseuduvaria* is shown to have dispersed from Sundaland after the late Miocene. The present-day paleotropical distribution of *Pseuduvaria* may have been achieved by long-distance dispersal, and speciation events might be explained by global climatic oscillations, sea level fluctuations, and tectonic activity.

The Annonaceae are a large pantropical family of flowering plants, consisting of ca. 135 genera and ca. 2,500 species in predominantly tropical and subtropical lowland forests (APG, 2003). The phylogeny of the family has previously been reconstructed based on morphological (Doyle and Le Thomas, 1994, 1996, 1997) and molecular

data (Doyle et al., 2004; Mols et al., 2004; Richardson et al., 2004). Four main clades are consistently recognised in the molecular analyses: two of these clades (consisting of *Anaxagorea* and the small “ambaviod” clade) form a heterogeneous basal grade, basal to two large clades known as the “long branch clade” (LBC) and “short branch clade” (SBC) to reflect differing rates of nucleotide substitutions (Pirie et al., 2006; Richardson et al., 2004).

Evolutionary divergence times based on molecular data have been estimated for the Annonaceae as a whole (Doyle et al., 2004; Richardson et al., 2004) as well as for several individual genera and clades, including *Anaxagorea* (Scharaschkin and Doyle, 2005), “Andean-centred” genera in the SBC (Pirie et al., 2006), *Guatteria* (Erkens et al., 2007), and an African clade including *Isolona* and *Monodora* (Couvreur et al., 2008a). These studies were based on between two and five commonly used chloroplast regions (*matK*, *trnL*-F, *trnT*-L, *rbcL*, and *psbA-trnH* spacer). Comparatively few fossil calibration points were used in these studies due to the scarcity of unequivocal Annonaceae fossils. The most widely applied calibration point is the stem age of the Magnoliaceae, based on the fossil *Archaeanthus* from North America, which either provides a minimum age of 98 Mya in the early Cenomanian (Dilcher and Crane, 1984), or 100 Mya in the late Albian (Gröcke et al., 2006). Other possible calibration points include: 68 Mya for the split between the ambaviods and the combined LBC-SBC clade, derived using *Anonaspermum* seeds from Nigeria (Chesters, 1955); 112 Mya for the split between the Eupomatiaceae and Annonaceae, using the fossil *Endressinia* (Mohr and Bernardes-de-Oliveira, 2004); and 120 Mya for core Magnoliales, based on Aptian granular monosulcate pollen (Ward et al., 1989). In addition, secondary calibration points have been derived from earlier studies of divergence times, including 82 Mya (Wikström et al., 2001) and 90.93 Mya (unpublished estimate by M.D. Pirie) for the stem of the Annonaceae. Richardson et al. (Richardson et al., 2004) have accordingly estimated the Annonaceae stem age at 90.6 ± 1.3 Mya, based on three calibration points (Richardson et al., 2004). The age estimates for the family have enabled significant palaeobiogeographical insight, although the most important work has focused on dispersal patterns in Africa and South America (western Gondwana) since the genera studied were largely neotropical.

The program r8s (Sanderson, 2004) has been widely adopted in studies of divergence times in the Annonaceae (Erkens et al., 2007; Richardson et al., 2004; Scharaschkin and Doyle, 2005). This program incorporates both non-parametric rate smoothing (NPRS) methods and penalized likelihood (PL) methods. NPRS methods estimate ages through a smoothing criterion (Sanderson, 1997), whereas PL is a semiparametric approach (Sanderson, 2002) that combines parametric methods with the robustness of non-parametric methods. Other dating methods utilized in studies of the Annonaceae (Pirie et al., 2005) include PAML (Yang, 1997) and MULTIDIVTIME (Thorne and Kishino, 2002), based on Bayesian dating methods. BEAST (Bayesian Evolutionary Analysis Sampling Trees) (Drummond and Rambaut, 2008) is the only software that simultaneously co-estimates phylogeny, node ages, and substitution rates. BEAST has recently been used to estimate the origins of the East African lineages within the Annonaceae (Couvreur et al., 2008a).

The present study focuses on the palaeotropical genus *Pseuduvaria*, which is nested within the SBC. The genus consists of 56 species of trees and treelets (inclusive of *Craibella phuyensis*, which has recently been shown to be congeneric with *Pseuduvaria* (Su et al., 2008, *in press*), and three newly described species (Su et al., *in press*)). A comprehensive monograph of the genus has been completed (Su and Saunders, 2006) and the phylogeny of *Pseuduvaria* has been inferred using five chloroplast regions (Su et al., 2008). The genus was previously estimated to have evolved at least ca. 16 Mya ago based upon a published chronogram (Richardson et al., 2004), although this may not represent the true age of the genus since the study included only five *Pseuduvaria* species (inclusive of *C. phuyensis*).

The Malesian phytogeographical region is separated into two main subregions by Huxley's line and Wallace's line (Figure 1: (Hall, 2001)): western Malesia, which includes Peninsular Malaysia, Sumatra, Java, Borneo and Palawan; and eastern Malesia, which includes Sulawesi, the Lesser Sunda Islands, the Moluccas, and New Guinea. Borneo and New Guinea represent the two most important centres of plant species richness and endemism within Malesia (Wiffin, 2002). *Pseuduvaria* is widely distributed from IndoChina to New Guinea and NE Australia (Figure 1, inset), with the main centre of diversity in New Guinea (with 20 species) and a secondary centre in Peninsular Malaysia (with 16 species, including three recently described species) (Su et al., *in press*). In an earlier molecular phylogenetic study of *Pseuduvaria*, five clades were recognized, with the three basal clades occurring in the western Malesia (Su et al., 2008). A continuous land mass connecting IndoChina, Thailand, Peninsular Malaysia, Borneo, Sumatra, and Java (collectively known as Sundaland) was formed during the Early Mesozoic (Hall and Morley, 2004), separated from the AustralianNew Guinea plate by a major ocean barrier. The collision between the Asian and New Guinea land masses during the Cenozoic may have promoted the dispersal of *Pseuduvaria* species from Sundaland to New Guinea. The present paper aims to use molecular dating techniques to test whether this biogeographical hypothesis is supported by evidence of past geological events.

We use Bayesian molecular dating techniques in BEAST to estimate the divergence times of major Annonaceae *clades* (including the date of origin of *Pseuduvaria* as well as clades within *Pseuduvaria*). Five chloroplast DNA regions, *psbA-trnH* spacer, *trnL-F*, *matK*, *rbcL*, and *atpB-rbcL* spacer, were used in all analyses. The divergence times of major Annonaceae clades were estimated using a dataset ("matrix A") with a broad taxonomic sampling across the basal grade, LBC and SBC of the Annonaceae, together with representatives of related families in the Magnoliales. Two fossil calibration points were used to date the divergence times within the Annonaceae: the fossil *Archaeanthus* (Dilcher and Crane, 1984), providing a minimum age of 98 Mya for the stem of the Magnoliaceae (J.A. Doyle, pers. comm.); and the Late Cretaceous fossil *Futabanthus*, providing a minimum age of 89 Mya for the split between *Anaxagorea* and the combined ambavioid-LBC-SBC clade (Takahashi et al., 2008). The ages of clades within *Pseuduvaria* were subsequently estimated using a smaller dataset ("matrix B") that included all *Pseuduvaria* species available (54 species) and selected members of the SBC. The age estimates inferred from BEAST analyses using matrix A were used as prior information in the subsequent analyses using matrix B. An additional objective

was to identify the most likely biogeographical origin of *Pseuduvaria*, and to infer subsequent dispersal patterns using dispersal-vicariance analysis (DIVA) (Ronquist, 1996) and weighted ancestral area analysis (WAAA) (Hausdorf, 1998).

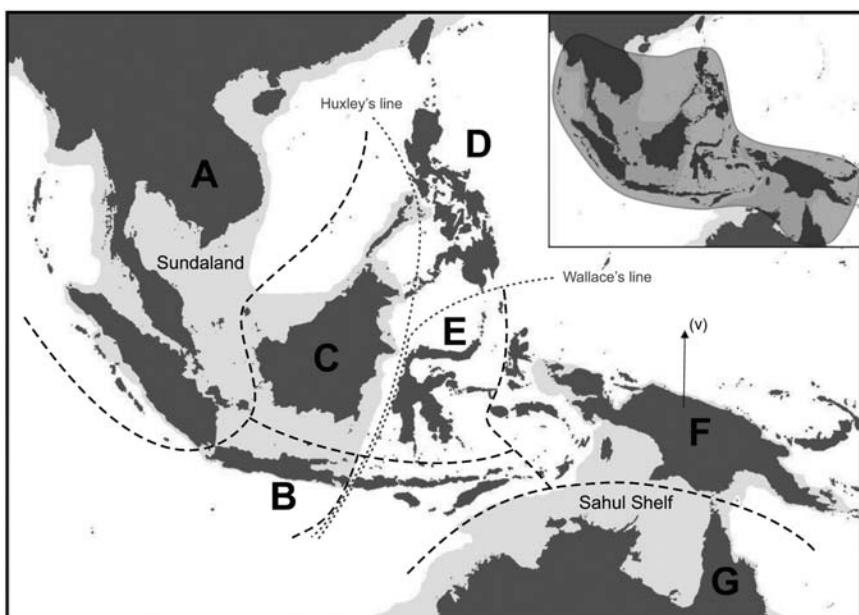


Figure 1. Main biogeographical areas used in analyses, and (inset) distribution of *Pseuduvaria* in SE Asia. Eight areas are outlined and labelled: A, China, IndoChina, Burma, Thailand, Peninsular Malaysia, and Sumatra; B, Java and Bali; C, Borneo and Palawan; D, Philippines (excluding Palawan); E, Sulawesi; F, New Guinea; and G, Australia. The shaded light v area represents shallow continental shelves, as modified from (Hall, 2001).

MATERIALS AND METHODS

Taxon Sampling and Sequence Data

We used two separate datasets: matrix A (for higher-level analyses) and matrix B (for species-level analyses within *Pseuduvaria*). Both sets of analyses were based on chloroplast DNA sequences from five regions (*psba-trnH* spacer, *trnL-F*, *matK*, *rbcL*, and *atpB-rbcL* spacer). The materials and methods for DNA extraction, PCR amplification and sequencing of *Pseuduvaria* are detailed in earlier publications (Su et al., 2008, in press). The sequences for the outgroups were downloaded from GenBank.

Matrix A comprised a total of 83 species. The ingroup included 22 of the total 56 *Pseuduvaria* species, representing all five main clades identified in our previous phylogenetic study (Su et al., 2008), and 56 species from 46 different Annonaceae genera (inclusive of the basal grade, LBC, and SBC). The outgroups included were *Coelocaryon preussii* (Myristicaceae), *Eupomatiabennettii* (Eupomatiaceae), *Liriodendron chinense* (Magnoliaceae), *Magnolia kobus* (Magnoliaceae), and *Persea americana* (Lauraceae). Matrix B consisted of 69 taxa, with 54 *Pseuduvaria* species. The selection of

outgroups was based on the phylogeny generated by BEAST using matrix A (Figure 2A). The SBC genera selected as outgroups were *Alphonsea*, *Haplostichanthus*, *Miliusa*, *Monocarpia*, *Neo-uvaria*, *Orophea*, *Polyalthia*, and *Sapranthus*, representing 15 species.

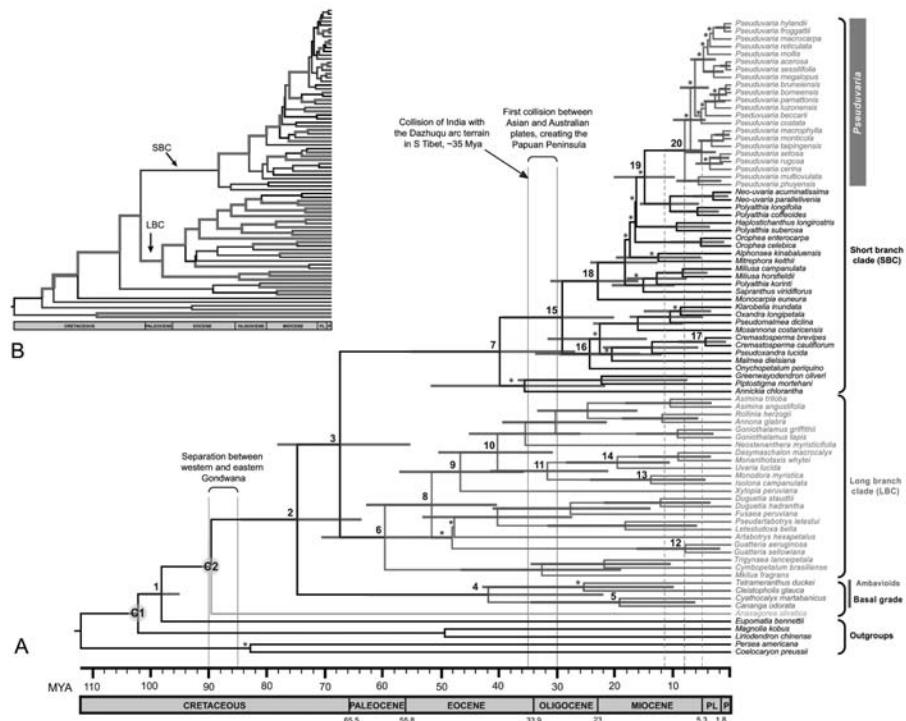


Figure 2. Chronograms of Annonaceae: maximum clade credibility trees from the BEAST analysis of matrix A. (A) Posterior estimates of divergence times were inferred using partitioned analyses based on five chloroplast regions, a UCLD model, and two fossils as minimum age constraints: an *Archaeanthus* fossil date of 98 Mya (calibration C1), and a *Futabanthus* fossil date of 89 Mya (calibration C2). Nodes are posterior mean ages (Mya), with node bars representing the 95% HPD intervals (see Table 2 for details). Bayesian PP < 0.95 are indicated by asterisks above branches. Geological time-scale abbreviations: PL, Pliocene; P, Pleistocene. The dashed line represents the estimated tMRCA of *Pseuduvaria* crown, with a 95% HPD indicated by the grey vertical lines. (B) Rate of evolution across all branches. Broad lines indicate lineages where the posterior rate is greater than the mean rate; narrow lines indicate lineages where the posterior rate is lower than the mean rate.

For matrices A and B, sequences were edited and assembled in SeqManPro using DNAStar Lasergene 8 (DNAStar), and aligned manually using BioEdit ver. 7.09 (Hall, 1999) and Se-Al ver. 2.0a11 (Rambaut, 1996). Ambiguously aligned positions were excluded from the analyzes. The combined five-region data sets in matrices A and B are composed of 4860 and 4361 characters, respectively (see Table 1 for data on individual gene regions).

Table 1. Length and best-fitting model for data partition analyses based on two different datasets (matrices A and B).

	Aligned length (bp)		Sequence evolutionary model	
	Matrix A	Matrix B	Matrix A	Matrix B
psbA-trnH spacer	572	441	GTR+G	GTR+G
trnL-F	1064	937	GTR+G	GTR+I
matK	823	811	GTR+G	GTR+G
rbcL	1344	1344	GTR+I+G	HKY+I+G
atpB-rbcL spacer	1057	828	GTR+G	GTR+I
Combined data	4860	4361	GTR+I+G	GTR+I+G

GTR, general time-reversible model; HKY, the Hasegawa, Kishino, and Yano model; I, proportion of invariant sites; G, gamma distribution.

Divergence Time Estimation

MrModelTest ver. 2.3 (Nylander, 2008) was used to determine the appropriate DNA substitution model and gamma rate heterogeneity using the Akaike Information Criterion (AIC). For both matrices, MrModelTest was performed for each gene region and combined regions (Table 1). For the combined five chloroplast DNA regions in matrices A and B, the GTR+I+G was determined as the best-fitting statistical model.

The tree topology, node ages and substitution rates were simultaneously estimated using a Bayesian MCMC (Markov chain Monte Carlo) approach as implemented in BEAST ver. 1.4.8 (Drummond and Rambaut, 2007, 2008). For matrix A analyses, exponential priors were selected for the two calibration points (see discussion below): the *Archaeanthus* calibration used 98 Mya (labelled C1 in Figure 2A) as zero offset, with an exponential mean of 1; and the *Futabanthus* calibration used 89 Mya (labeled C2 in Figure 2A) as zero offset, with an exponential mean of 1. A Yule speciation tree prior was furthermore specified; this prior has been recommended for species-level phylogenies, and assumes a constant rate of speciation per lineage (Drummond and Rambaut, 2008). An uncorrelated lognormal distributed relaxed clock (UCLD) model was employed, which allows evolutionary rates to vary along branches within lognormal distributions (Drummond et al., 2006). The five chloroplast regions were partitioned (see Table 1 for the different models of substitution), allowing the incorporation of different models of substitution for each gene independently; the xml input file based on matrix A was edited manually for gene region partition following the tutorial by T.L.P. Couvreur (Couvreur et al., 2008a; Couvreur, 2008).

Two independent MCMC runs were performed, each run of 50 million generations, with sampling every 5,000 generations. The two separate runs were then combined (following the removal of 10% burn-in) using LogCombiner ver. 1.4.8 (Drummond and Rambaut, 2007, 2008). Adequate sampling and convergence of the chain to stationary distribution were confirmed by inspection of MCMC samples using Tracer ver. 1.4 (Rambaut, 2007). The effective sample size (ESS) values of all parameters were greater than 200, which were considered a sufficient level of sampling. The

sampled posterior trees were summarized using TreeAnnotator ver. 1.4.8 (Drummond and Rambaut, 2007, 2008) to generate a maximum clade credibility tree (maximum posterior probabilities) and calculate the mean ages, 95% highest posterior density (HPD) intervals, posterior probabilities (PP) and substitution rates for each node. The BEAST topology (Figure 2A) was visualized with FigTree ver. 1.2 (Rambaut, 2008). The MCMC analysis based on matrix A was also performed without sequence data; the prior distribution can therefore be compared with the posterior distribution in order to examine the influence of the data and prior, showing that the results are not influenced by the chosen priors alone (Drummond et al., 2006).

The time to most recent common ancestor (tMRCA) of *Monocarpia* and *Pseuduvaria* inferred from the MCMC runs of matrix A were used as prior information in the subsequent BEAST analyses using matrix B in order to estimate the divergence times within *Pseuduvaria* lineages. A normal probability distribution was employed as priors for matrix B analyses, since the 95% HPD intervals from previous estimation (based on matrix A) can also be accommodated as uncertainty for this analysis. The normal probability distribution is thought to reflect uncertainty in secondary calibration points (Bergh and Linder, 2009; Couvreur et al., 2008a; Ho, 2007;). Normal distribution priors were therefore applied to calibration points at two nodes: the MRCA of *Monocarpia*, with a mean age of 23 Mya and a standard deviation of 3.78 Mya (labelled C3 in Figure 3); and the MRCA of *Pseuduvaria*, with a mean age of 8 Mya and a standard deviation of 1.64 Mya (labelled C4 in Figure 3). The xml input file based on matrix B was edited manually for partitioned chloroplast sequence data. A UCLD model with Yule speciation tree prior was employed. Two independent MCMC runs of 30 million generations were performed, with sampling every 3,000 generations. The two separate runs were then combined (following removal of 10% burn-in) using LogCombiner ver. 1.4.8 (Drummond and Rambaut, 2007, 2008). The runs were checked for convergence using Tracer ver. 1.4(Rambaut, 2007), and the ESS values of all parameters shown to be greater than 500, which was considered sufficient. The BEAST topology (Figure 3) was summarized and visualized as described above.

Fossil Calibrations

The most commonly used calibration point in previous studies of the Annonaceae has been the fossil *Archaeanthus* (Richardson et al., 2004; Pirie et al., 2006) which provides an age of 98 Mya for the stem of the Magnoliaceae. Other calibration points previously employed for the Annonaceae are 120 Mya for the node of the core Magnoliales using the pollen fossil *Lethomasites* (Scharaschkin and Doyle, 2005), 112 Mya for the split between the Eupomatiaceae and Annonaceae using the *Endressinia* fossil (Mohr and Bernardes-de-Oliveira, 2004), and 68 Mya for the common ancestor of the ambavioids and combined LBC-SBC clade using the Maastrichtian seed genus *Anonaspermum* from Nigeria and Sierra Leone (Chesters, 1955).

Lethomasites was not used in the present study since its placement within the Magnoliales is doubtful (Doyle, 2004). The inclusion of *Lethomasites* would also require broader taxonomic sampling of other families within the Magnoliales (including Degeneriaceae and Himantandraceae), and the large resultant dataset would require considerably longer computational time. The *Endressinia* fossil was not reliable

as its relationship with the Eupomatiaceae is doubtful (Couvreur et al., 2008a). The *Anonaspermum* seeds were similarly avoided as calibration points in the present study as the position of this fossil within the Annonaceae is uncertain (Couvreur et al., 2008b).

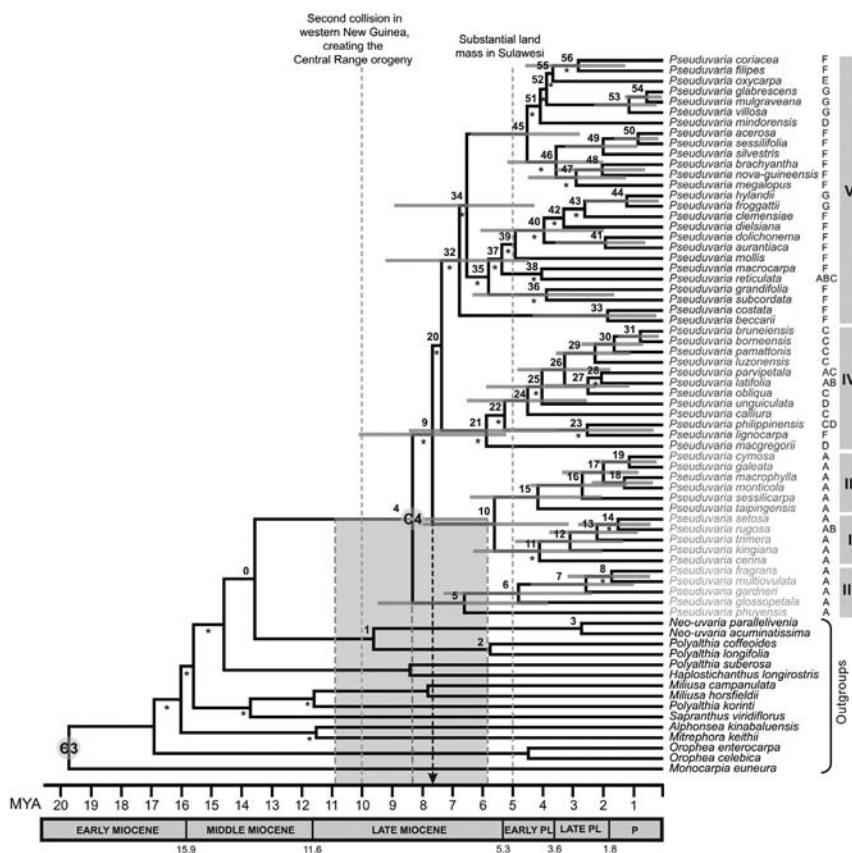


Figure 3. Chronogram of *Pseuduvaria*: maximum clade credibility tree from the BEAST analysis of matrix B. Posterior estimates of divergence times were inferred using partitioned analyses based on five combined chloroplast regions, a UCLD model, and two secondary calibration points (labelled C3 and C4). Nodes are posterior mean ages (Mya), with node bars representing the 95% HPD intervals (see Table 3 for details). Bayesian PP < 0.95 are indicated by asterisks below branches. Letters to the right of taxon names represent current geographical distributions (defined in Figure 1). Geological time scale abbreviations: PL, Pliocene; P, Pleistocene. The dashed line represents the estimated tMRCA of *Pseuduvaria* crown, with a 95% HPD indicated by the grey bar; black dashed line represents the earliest possible migration event for *Pseuduvaria* out of Sundaland.

The inclusion of a greater number of fossil calibration points can reduce bias and result in more accurate age estimates, assuming that the chosen points are mutually consistent (Hug and Roger, 2007). The fossil record of the Annonaceae is very limited, however, and only two fossils, *Archaeanthus* and *Futabanthus*, were regarded as sufficiently reliable. The phylogenetic position of *Archaeanthus* (Dilcher and Crane,

1984) appears to be reliable as it shares several characters with the Magnoliaceae, including distinctive stipules, elongated receptacle, and fruits with numerous follicles (Couvreur et al., 2008a; Doyle et al., 2004; Dilcher and Crane, 1984). The phylogenetic position of the recently discovered Late Cretaceous fossil flower, *Futabanthus asamigawaensis* (Takahashi et al., 2008), is similarly reliable as it shares many morphological similarities with the Annonaceae, including hermaphroditic, hypogynous flowers with six tepals in two whorls of three, an androecium consisting of numerous stamens with flattened connectives that extend above the thecae, and a gynoecium of numerous free carpels. Takahashi et al. [(2008): 914] postulated that *Futabanthus* has “a position within crown-group Annonaceae, perhaps as sister or near the base of all extant taxa except *Anaxagorea*.” Inner staminodes are absent in *Futabanthus* but are present in the basal genus *Anaxagorea*; the lack of inner staminodes is therefore possibly synapomorphic for the rest of the Annonaceae (plesiomorphic in *Anaxagorea*). We set *Futabanthus* as a minimum age constraint of 89 Mya for the split between *Anaxagorea* and the combined ambavioid-LBC-SBC clade (i.e., the crown age of the Annonaceae, labelled as C2 in Figure 2A).

Previous dating studies furthermore utilized secondary calibration points that were extracted from earlier estimates, viz.: the age of 82 Mya for the stem of the Annonaceae based on Wikström et al. (2001) using the NPRS method (Erkens et al., 2007; Richardson et al., 2004); and the age of 90.93 Mya for the stem of the Annonaceae based on unpublished estimations by M.D. Pirie. Secondary calibration points generated by these previous Annonaceae studies were avoided as a different molecular dating method and calibration points were employed in this study.

Historical Biogeography

Ancestral biogeographical areas occupied by *Pseuduvaria* were inferred using two methods: dispersal-vicariance analysis (DIVA) (Ronquist, 1996) and weighted ancestral area analysis (WAAA) (Hausdorf, 1998). Eight main biogeographical areas represented in the current distributions of *Pseuduvaria* species were used for the ancestral area reconstructions (Figure 1): (A) southern China, Indochina, Thailand, Burma, Peninsular Malaysia, and Sumatra; (B) Java and Bali; (C) Borneo and Palawan (identified as a single biogeographical unit following (Takhtajan, 1986)); (D) Philippines (excluding Palawan); (E) Sulawesi; (F) New Guinea and the Moluccas; (G) Australia; and also the outgroup distribution (H) India and Sri Lanka (not shown in Figure 1). Species distribution data were derived from our monograph of *Pseuduvaria* (Su and Saunders, 2006).

The DIVA was implemented using ver. 1.2 of the program (Ronquist, 1996), in which ancestral distributions are inferred by minimizing the number of dispersal and extinction events. It assumes vicariance is the default mode of speciation, thus it does not assign a cost for vicariance but counts steps for dispersal and extinction events. The tree topology resulting from the BEAST analyses based on matrix B was used as DIVA requires a fully resolved tree. The sister clade with four species (*Neo-uvaria acuminatissima*, *Neo-uvaria parallelivenia*, *Polyalthia coffeoides* and *Polyalthia longifolia*) was also included in the biogeographical analysis. The geographical distributions estimated for

internal nodes were optimized by constraining the maximum number of unit areas to two, hence restricting the number of unit areas in ancestral distributions.

The WAAA estimates ancestral areas using reversible parsimony, and weights areas in plesiomorphic branches more than apomorphic branches (Hausdorf, 1998). Probability indices (PI) were calculated for each area at each node by counting the number of weighted gain steps (GSW) and weighted loss steps (LSW) on the cladogram resulting from the BEAST analyses (also based on matrix B). The PI is the ratio of LGW and LSW; biogeographical areas with a value of less than 0.2 are not considered as part of the ancestral area.

DISCUSSION

Historical biogeography of Major Annonaceae Clades

The divergence times within the Annonaceae differ slightly from previous age estimates based on NPRS and PL methods as implemented in r8s (see Table 2). Our results are comparable, however, with previous BEAST analyses: the *Isolona-Monodora* clade is dated at 13.9 Mya (95% HPD: 24.34.4 Mya; Figure 2A: node 13) in this study, consistent with estimates of ca. 14 Mya in a previous study using BEAST (Couvreur et al., 2008a).

Previous studies have suggested that the Annonaceae are likely to have originated in western Gondwana (Richardson et al., 2004; Scharaschkin and Doyle, 2005). The break-up of the Gondwanan supercontinent into western Gondwana (Africa and South America) and eastern Gondwana (Australia, Antarctica, Madagascar, and India) began 180150 Mya (Hallam, 1994; Scotese, 1988), but it was not until 9085 Mya that the two became fully separated (Ali and Aitchison, 2008). Biotic interchange between South America and Africa had essentially ceased by the latter half of the Late Cretaceous (8065 Mya) (Hallam, 1994; Scotese, 1988), although it has been suggested that limited connection between the two continents was possible for sometime afterwards via the island chains of the Rio Grande Rise-Walvis Ridge and Sierra Leone Rises (Morley, 2003). The earliest divergence in the Annonaceae appears to have occurred between 98.0 Mya (95% HPD: 101.594.9 Mya; Figure 2A: node 1) and 89.5 Mya (Figure 2A: calibrated node C2 at a minimum age of 89 Mya for the oldest known crown group fossil, *Futabanthus*), probably after the separation of the two landmasses.

Anaxagorea is shown to be sister to all other members of the Annonaceae. The age of the split between the *Cananga**Cyathocalyx**Cleistopholis**Tetrameranthus* clade and the combined LBC-SBC clade is estimated at 74.7 Mya (95% HPD: 84.463.6 Mya, PP = 1.00; Figure 2A: node 2). In previous studies this node has been used as a minimum age constraint of 68 Mya. The split age estimated here may correspond to the origin of seeds of the *Anonaspermum* type, and our results furthermore indicate that the age of this node could be older than 68 Mya. The tMRCA of the *Cananga**Cyathocalyx**Cleistopholis**Tetrameranthus* clade is estimated at 41.8 Mya (95% HPD: 62.422.3 Mya, PP = 1.00; Figure 2A: node 4). The genera in this clade are currently distributed in geographically disparate regions: *Cananga* occurs in Southeast Asia and Australia (Heusden van, 1992); *Cyathocalyx* in Southeast Asia, extending east to Fiji

(Wang and Saunders, 2006a, 2006b); *Cleistopholis* in Africa (Heusden van, 1992); and *Tetrameranthus* in South America (Westra, 1985). The ancestors of this clade may have dispersed between Africa, Asia, and South America from 74.7 Mya (95% HPD: 84.463.6 Mya, PP = 1.00; Figure 2A: node 2) to 41.8 Mya (95% HPD: 62.422.3 Mya, PP = 1.00; Figure 2A: node 4). This disjunct distribution may have been the result of Eocene and early Oligocene cooling (Pennington and Dick, 2004; Richardson et al., 2004).

The LBC and SBC constitute over 90% of all Annonaceae genera. The age of the split between the LBC and SBC lineages is estimated at 67.3 Mya (95% HPD: 78.155.2 Mya, PP = 1.00; Figure 2A: node 3), and the mean tMRCA of the LBC at 59.6 Mya (95% HPD: 70.548.1 Mya, PP = 1.00; Figure 2A: node 6). The ancestors of the majority of LBC genera (*Guatteria-Asimina*) appear to have evolved ca. 51.5 Mya (95% HPD: 62.840.3 Mya, PP = 1.00; Figure 2A: node 8). These genera are mainly distributed in South America and Africa, although several occur in Asia, including *Artobotrys*, *Goniothalamus*, *Uvaria*, and *Xylopia*; this suggests that the LBC originated in South America and/or Africa-Madagascar [(Doyle et al., 2004); Figure 7]. The wide geographical distribution of boreotropical taxa could be associated with the combined effects of plate tectonics and global climatic changes. The warming period during the late Paleocene-early Eocene thermal maximum, which peaked around the early Eocene Climatic Optimum (5250 Mya) (Zachos et al., 2001), may have promoted the northward dispersal of tropical plants (Morley, 2003). This coincides with the suggested date for the clade consisting of all members (node 8) of the LBC except the basal *Mkilua-Trigynaea* lineage in the early Eocene. In a recent paleontological review, most modern subgroups of mammalian orders were shown to have appeared during the Paleocene-Eocene thermal maximum, suggesting that rapid environmental changes may have evolutionary significance (Gingerich, 2006). During this period (ca. 56 Mya: (Morley and Dick, 2003)), a Greenland land bridge may have connected North America and Eurasia, enabling plant dispersals between the two continents (Morley, 2003). Another possible dispersal route was across the Tethys seaway between Africa-Arabia and Asia, allowing tropical plant dispersal until the late Eocene-early Oligocene (Pennington and Dick, 2004).

The mean tMRCA of the SBC is estimated at 39.8 Mya (95% HPD: 55.226.9 Mya, PP = 1.00; Figure 2A: node 7). The ancestral biogeographical origin of the SBC cannot be postulated here without a more extensive sampling of genera, although the inferred areas could be Africa-Madagascar as shown in the *Annickia-Piptostigma-Greenwayodendron* clade ((Doyle et al., 2004); Figure 7). Most evolutionary diversifications within the SBC seemed to have initiated ca. 29.1 Mya (95% HPD: 39.220.2 Mya, PP = 1.00; Figure 2A: node 15). The rate of molecular evolution during the early divergence of the SBC (arrowed in Figure 2B, indicated by line) is about four times slower than that during the early divergence of the LBC (arrowed in Figure 2B, shown as thick line).

The SBC, as shown in Figure 2A, consists of small clades from Africa (a clade consisting *Annickia*, *Piptostigma*, and *Greenwayodendron*) and South to Central America (*Onychopetalum-Cremastosperma*), with all other members from Asia. Diversification

within the SBC ca. 29.1 Mya (Figure 2A: node 15) resulted in a South American clade consisting of *Cremastosperma*, *Klarobelia*, *Malmea*, *Mosannonia*, *Onychopetalum*, *Oxandra*, *Pseudomalmea*, and *Pseudoxandra*. The descendants of SBC genera could have dispersed from South America into Asia (since *Monocarpia euneura* occurs in Borneo (Mols and Keßler, 2000)) via Africa and the Indian Plate; this presumably occurred between 29.1 (95% HPD: 39.220.2 Mya, PP = 1.00; Figure 2A: node 15) and 23.0 Mya (95% HPD: 31.116.0 Mya, PP = 1.00; Figure 2A: node 18). The alternative dispersal scenario (from Africa into Asia via Europe) is less likely because most Asian Annonaceae occur in warm and moist tropical rainforests. It is therefore more probable that the SBC genera dispersed from Africa to Asia via India, and that the dispersal between Africa and India occurred before the collision between Asia and India. Many past biogeographical studies have postulated that the collision between Asia and India is occurred 50–55 Mya (Leech et al., 2005; Rowley, 1996; Zhu et al., 2004). More recent work (Aitchison et al., 2007), however, suggests that this Early Eocene event marks the collision of India's northern margin with a sub-equatorially located island arc (remnants of which are today preserved as the Dazhuqu arc terrain in southern Tibet), with the main continent-continent collision taking place ca. 34 Mya. However, the passage of the subcontinent may have taken the north-eastern corner of the block (Ali and Aitchison, 2005) close to Western South-East Asia, and thus faunal and floral exchanges between the two geographical regions may have been possible from the late Paleocene (ca. 57 Mya) onwards (Ali and Aitchison, 2008).

Phylogeny and Historical Biogeography of *Pseuduvaria*

The BEAST topology of *Pseuduvaria* (Figure 3) is similar to those resulting from previous ML and Bayesian analyses (see Figure 3 in (Su et al., 2008)). Although the phylogenetic positions of *Pseuduvaria* species in clades I, II, and III remain identical, there are nevertheless some minor topological discrepancies between the trees. In the present study, clade III is shown to be sister to all other representatives of *Pseuduvaria*, whereas in previous analyses (Su et al., 2008) this position was occupied by clade I. This is unlikely to have a significant impact on biogeographical interpretations, however, because the species in clades I, II, and III are all distributed in geographical regions that were formerly part of the Sunda landmass. Another difference between the present and previous analyses relates to the phylogenetic position of two New Guinea species, *P. becarrii* and *P. costata*. These species are shown to be sister to all other members of clade V in the present study, but were basal within clade IV in the previous analyses (Su et al., 2008); the position of these two species is not statistically supported, although we suggest that they are more likely to belong to clade V as this clade is predominantly composed of New Guinea species. Several species in clade V (*P. oxycarpa*, *P. filipes*, *P. coriacea*, *P. mindorensis*, *P. reticulata*, and *P. mollis*) are furthermore located in different positions in different analyses, but their positions lack statistical support. Again, this may have little effect on biogeographical interpretations as most species in the clade are from New Guinea (except *P. mindorensis* from the Philippines, *P. oxycarpa* from Sulawesi, and *P. reticulata*, which is widespread in Malesia).

Late Miocene Origin in Sundaland

Biogeographical reconstructions using DIVA and WAAA both suggest that *Pseuduvaria* originated in Sundaland (Figure 6: node 4). Initial BEAST analyses based on matrix A (with a broad taxonomic sampling) gave an estimated tMRCA for the *Pseuduvaria* lineage of 8.0 Mya (95% HPD: 11.55.0 Mya, PP = 1.00; Figure 2A: node 20); this node was subsequently selected as one of the two secondary molecular calibration points for *Pseuduvaria* in analyses based on matrix B (Figure 3: C4). These later analyses suggest a late Miocene origin, with a mean tMRCA of 8.3 Mya (95% HPD: 11.05.8 Mya, PP = 1.00; Figure 3: node 4). The former estimation from matrix A analyses, with a mean age of 8.0 Mya, will be used in all subsequent discussions presented here since this age was assigned as a calibration prior to the tMRCA of *Pseuduvaria* for matrix B analyses.

Early divergence of *Pseuduvaria* could have occurred between 14.9 Mya (95% HPD: 20.29.7 Mya, PP < 0.95; Figure 2A: node 19) and 8.0 Mya (95% HPD: 11.55.0 Mya; Figure 2A: node 20) during the Miocene, possibly associated with the “mid-Miocene climatic optimum”. This warming phase occurred ca. 1715 Mya (Zachos et al., 2001; Figure 2), and led to the extensive growth of megathermal vegetation throughout most regions in Sundaland (Morley, 2007). Subsequent global cooling during the middle Miocene climate transition (14.213.8 Mya) gradually resulted in the reduction of sea surface temperatures and the expansion of the Antarctic ice-sheets (Shevenell et al., 2004), which continued until ca. 6 Mya, in the early Pliocene (Zachos et al., 2001). These cold global temperatures were associated with a decline in CO₂ levels, affecting the productivity of terrestrial vegetation (Kürschner et al., 2008), and resulted in the contraction of moist megathermal vegetation to the tropics and the concomitant expansion of grasslands (Morley, 2007). This was followed by a warming phase that continued until ca. 3.2 Mya (Zachos et al., 2001). This climatic change was reflected by a steady increase in rainforest diversity in Southeast Asia until the mid Pliocene (Heane, 1991; Morley, 2007). Subsequent diversifications of *Pseuduvaria* gave rise to clade III at ca. 6.6 Mya (95% HPD: 9.53.8 Mya; Figure 3: node 5) and clades III collectively at ca. 5.6 Mya (95% HPD: 8.13.2 Mya; Figure 3: node 10); this might be linked to these global climatic oscillations, with evolutionary divergence promoted by the significantly warmer and wetter climate.

Diversification in Sundaland

Biogeographical reconstructions of the ancestral distributions of clades III using DIVA and WAAA suggest that all three clades originated in Sundaland (Figure 4: nodes 5, 11, 15). Clade III consists of five species, *P. phuyensis*, *P. glossopetala*, *P. gardneri*, *P. multiovulata*, and *P. fragrans*; the tMCRA is estimated at ca. 6.6 Mya (node 5 in Table 3 and Figure 3). Clades I and II originated at a similar time, with tMRCA estimates of 4.1 and 4.2 Mya, respectively (nodes 11 and 15 in Table 3 and Figure 3).

Species in clades I and II are mostly endemic to Peninsular Malaysia or Thailand, with the exception of four comparatively widespread species, viz.: *P. trimera* (southern China, Vietnam, Myanmar, and Thailand), *P. setosa* (Thailand and Peninsular Malaysia), *P. rugosa* (Myanmar, Thailand, Peninsular Malaysia, Sumatra, and Java), and *P. macrophylla* (Peninsular Malaysia and Sumatra) (Su et al., 2008). *Pseuduvaria*

rugosa is estimated to have diverged around 1.5 Mya (95% HPD: 2.80.4 Mya, PP<0.95; Table 3: node 14), at a similar time to *P. macrophylla* at ca. 1.3 Mya (95% HPD: 2.40.3 Mya, PP = 0.99; Table 3: node 18). In the middle and late Miocene, land connected Sumatra-Java to mainland Southeast Asia (Hall, 1998). Further reduction of sea levels during the Pleistocene (to a maximum depth of 120 m below the present level) exposed additional land area that connected Peninsular Malaysia, Sumatra, Java, and Borneo (Inger and Voris, 2001). Tropical forests also gradually expanded, as shown by the arrival of orangutans in Java (Bird et al., 2005). At that time, *P. macrophylla* and *P. rugosa* were likely to have been able to disperse unimpeded between Sumatra, Java, and Peninsular Malaysia; *P. trimera* would similarly have been able to advance northwards into southern China. Seed dispersal of *P. rugosa* is known to be facilitated by orangutans (Su and Saunders, 2006).

Dispersal Out of Sundaland

DIVA and WAAA indicated several possible ancestral distributions for clades IV and V collectively (Figure 6: node 20). According to DIVA, New Guinea, Borneo-New Guinea, or Philippines-New Guinea are potential ancestral areas, whereas WAAA highlighted New Guinea (PI = 1.13) and the Philippines (PI = 0.62). The age of node 9 is estimated at ca. 7.7 Mya, although this node is not statistically supported (PP<0.95; Table 3). This age estimate (indicated by a dotted black line in Figure 3) could represent the earliest migration event for *Pseuduvaria* out of Sundaland to the Philippines and/or New Guinea (Figure 4).

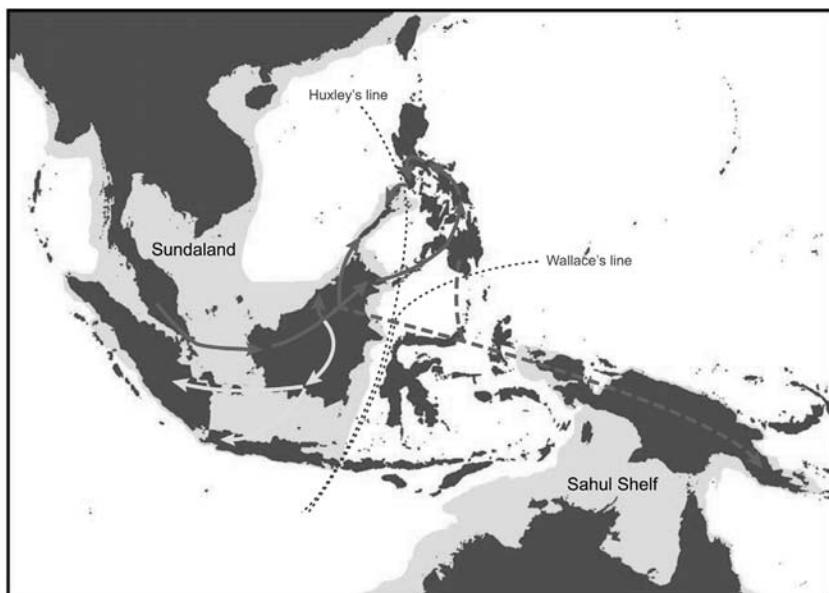


Figure 4. Mapping of ancestral distributions on *Pseuduvaria*. Letters (above lines) represent DIVA results, and letters (below lines) are WAAA results. For the latter results, letters in brackets represent areas with lower probability indices. Letters to the right of the species names represent current species distributions (as defined in Figure 1).

The tMRCA of clade IV is estimated at ca. 5.9 Mya (95% HPD: 8.53–5.5 Mya, PP < 0.95; Figure 3: node 21). WAAA indicated that the Philippines is the most likely ancestral area, with the highest PI value of 1.7 (Table 3 and Figure 6: node 21), although Borneo also returned a relatively high PI of 0.6. Similar results were obtained using DIVA, with the Philippines or Borneo–Philippines as the most likely ancestral areas (the third suggested scenario, with New Guinea–Philippines as the ancestral area, is improbable since the two areas are not contiguous). The timing of dispersal events from Sundaland to the Philippines may be associated with the docking of the Philippine Sea Plate with the Sunda Block, which was initiated in the late Miocene–Pliocene (Pubellier et al., 2003).

The WAAA and DIVA results are complex within clade IV, but appear to show probable dispersals from Borneo and/or the Philippines eastwards into New Guinea (dotted arrows in Figure 4) and westwards into continental Southeast Asia, Sumatra, and/or Java (arrows in Figure 4). The eastward migration into New Guinea may have occurred between node 22 (mean tMRCA: 5.3 Mya; without HPD, PP < 0.95; Table 3) and node 23 (mean tMRCA: 2.5 Mya; 95% HPD: 5.30–3.3 Mya, PP < 0.95; Table 3). These dispersals from Sundaland to New Guinea may have been achieved by three routes: island hopping after the collision of the Australian Plate with the Philippine Plate; stochastic dispersals across the Makassar Straits; or dispersal of montane taxa via island-hopping following the uplift of New Guinea and other islands (Morley, 2003). The dispersal of *Dacrycarpus* (Podocarpaceae) is a suggested example of the third route: *Dacrycarpus* dispersed to New Guinea from Australia in the mid-Miocene and then dispersed to Borneo ca. 3.5 Mya in the mid-Pliocene (Morley, 2003). *Pseuduvaria* species generally occur at low altitudes, suggesting that this third route is less likely to be important, although *P. lignocarpa* can occur at altitudes over 1,000 m (Su and Saunders, 2006).

The WAAA determined that the most probable ancestral areas at node 24 (Figure 6) are either Borneo (with the highest PI of 1.79; Table 3) or the Philippines (PI = 0.48; Table 3); similar results were obtained using DIVA. The ancestral distributions for node 25 (Figure 6) were the same as those for node 24, although node 25 is not statistically supported. A westward migration from Borneo–Philippines into mainland Southeast Asia, Sumatra, and/or Java may have occurred by nodes 26, 27 and/or 28 (arrows in Figure 4).

Colonization of New Guinea Via Long-Distance Dispersal

Clade V contains almost half of the genus, with most species currently distributed in New Guinea. The earliest suggested date for the colonization of New Guinea by *Pseuduvaria* species is 7.7 Mya (95% HPD: 10.15–2.2 Mya, PP < 0.95; Figure 6: node 9), although this clade lacks statistical support (Table 3). If colonization of New Guinea by node 9 is accepted, it is necessary to invoke a subsequent reverse westward dispersal for clades I and II (Figure 6: node 10). A more conservative scenario involves a later colonization of New Guinea after 7.4 Mya (Figure 6: node 20), although this node is also unsupported.

A recent plate tectonic modeling of New Guinea during the Cenozoic ((van Ufford and Cloos, 2005)) has shown major changes in the motion of the Pacific Plate beginning

around 43 Mya (Eocene), with the initiation of two subduction systems, generating the Inner and Outer Melanesian arcs. During the Oligocene (ca. 3530 Mya), the first significant collision event began and the northern part of the Australia plate was subducted underneath the Inner Melanesian arc, creating the “Peninsular orogeny” and generating the Papuan Peninsula (van Ufford and Cloos, 2005). As a result of a drop in sea level (ca. 90 m) during the Oligocene, central, and western New Guinea became largely sub-aerial.

In the late Miocene (ca. 10 Mya), a second collision event initiated in western New Guinea, creating the “Central Range orogeny” (van Ufford and Cloos, 2005). At that time, the Sunda and the Sahul Shelves may have reached closest proximity, although floristic migration may have been unlikely since New Guinea was still largely submerged, and deep ocean barriers existed with no land bridges (Hall, 2001; Rowe et al., 2008). The best opportunity for biota to “island hop” across Wallacea appears to have been during the last 5 million years (Hall, 2001), due to: (1) the formation of a substantial landmass in Sulawesi from ca. 5 Mya (Hall, 2001); (2) the connection between New Guinea and Sundaland via the Banda and Sunda arcs ((Hall, 1998), p. 122); (3) the increasing number of volcanic islands in East Indonesia (Barby and Pierce, 2007); (4) the exposure of the Sunda and Sahul Shelves caused by falling sea level (up to 120 m lower than present levels) during the Pleistocene glaciations (Inger and Voris, 2001); and (5) the accretion of microcontinental island arc fragments at the northern edge of the Australian craton portion of New Guinea (Barby and Pierce, 2007; Hall, 2001). Migration of *Pseuduvaria* from Sundaland to New Guinea may therefore have been facilitated by a series of stepping stones, possibly via Borneo (dotted arrow in Figure 5) or the Lesser Sunda Islands (dotted orange arrow in Figure 5). *Pseuduvaria* also

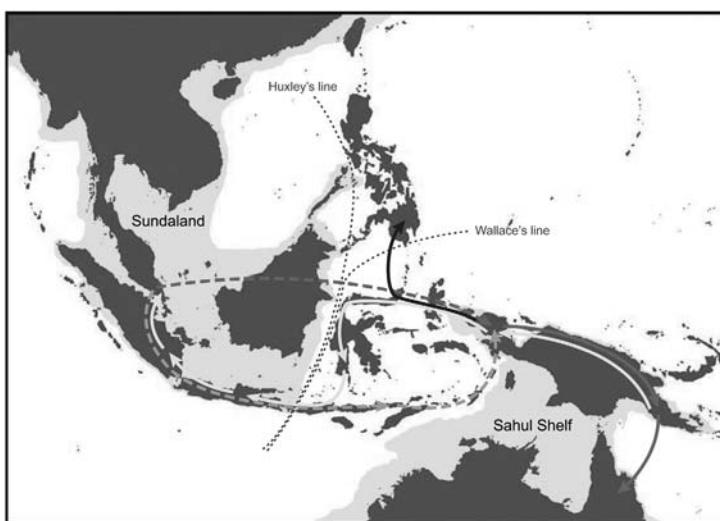


Figure 5. Mapping of ancestral distributions on *Pseuduvaria* (continued). Letters (above lines) represent DIVA results, and letters (below lines) are WAAA results. For the latter results, letters in brackets represent areas with lower probability indices. Letters to the right of the species names represent current species distributions (as defined in Figure 1).

occurs in the Aru islands (*P. aurantiaca*) and New Britain (*P. macrocarpa*). Similar biogeographical patterns are found in other plant groups such as the tribe Aglaiae (Meliaceae) which originated in western Malesia and subsequently dispersed eastwards to the Aru and Kai islands, New Guinea, New Britain, Australia, Fiji, and other Pacific islands (Muellner et al., 2008); migration between the Sunda and Sahul shelves was likely to have been achieved via long-distance dispersal by birds (Muellner et al., 2008).

Diversification Patterns in New Guinea and Australia

Pseuduvaria may have begun to diversify in New Guinea after node 34 from ca. 6.5 Mya (95% HPD: 9.04–3 Mya, PP < 0.95; Figure 3), resulting in two main lineages with tMRCA of ca. 5.8 Mya (Table 3: node 35) and ca. 4.5 Mya (Table 3: node 45). Coincident with this timeframe, there was a gradual transition between global cooling and warming phases: the cooling phase lasted until ca. 6 Mya, followed by a warming phase with moist climate until ca. 3.2 Mya (Zachos et al., 2001). A cooling period subsequently started again with a drier climate in the late Pliocene (Morley, 2002), followed by several cycles of glacial-interglacial episodes from ca. 1.8 Mya during the Pleistocene (Johnson, 2004; Poulin et al., 2002). These climatic oscillations, combined with the changes in sea level, may have promoted *Pseuduvaria* speciation in New Guinea.

Pseuduvaria exhibits a high level of endemism (95%) in New Guinea (Su and Saunders, 2006). This could be associated with the Central Range orogeny in New Guinea, which developed from ca. 8 Mya; this collisional orogenesis led to the formation of a ca. 1,300 km-long mountainous backbone from the Bird's Neck to the Papuan Peninsula (Ufford and Cloos, 2005) with some peaks over 5,000 m (Hill and Hall, 2003). Speciation within New Guinea may therefore have been promoted by mountain ranges over 4,000 m acting as physical barriers and influencing temperatures, and a continuously hot and wet climate with rainfalls of over 2,500 mm per annum (Beebe and Cooper, 2002). *Pseuduvaria* species in New Guinea diversified morphologically in response to these physical and climatic barriers, with evolutionary shifts from unisexual into bisexual flowers, elongation of flowering and fruiting peduncles (favoring fruit-bat frugivory), and the evolution of distinct pollination systems (Su et al., 2008; Su and Saunders, 2006). The New Guinea orogeny is analogous to the Miocene uplift of the Andean range, where different elevations (lowland or highland habitats) have impacted species diversification in the Neotropics (Young et al., 2002).

There is also evidence of dispersal of *Pseuduvaria* species from New Guinea to northern Australia (Figures 7 and 5), as shown in node 44 (*P. froggattii* and *P. hylandii*) and node 53 (*P. villosa*, *P. mulgraveana* and *P. glabrescens*). These clades are estimated to have evolved after 2.6 Mya (Table 3: node 43) and 3.9 Mya (Table 3: node 52), respectively. *Pseuduvaria* would have been able to disperse between New Guinea and Australia during the Pleistocene (arrows in Figure 5): glaciations caused a lowering of oceans to at least 120 m below the present level (Heaney, 1986; Morley and Flenley, 1987; Voris, 2000), and as a result New Guinea and northern Australia formed a continuous land mass (Rowe et al., 2008; Voris, 2000) with open forest vegetation

growing across the Torres Strait land bridge (between eastern New Guinea and northern Australia) between 17,000 and 8,000 years ago (Rawlings, 2004).

Although clade V is largely composed of New Guinea and Australian species, there are three significant exceptions. *Pseuduvaria reticulata* is widespread in Peninsular Malaysia, Sumatra, Java, Lesser Sunda islands, and Borneo, indicating an apparent westwards recolonization of Sundaland between 5.4 Mya (Table 3: node 37) and 4.1 Mya (Table 3: node 38). The most likely migration route from the Sahul Shelf to Sundaland (Java) may have been achieved via Sulawesi (Morley, 2002) (arrows in Figure 5); dispersal across the wide Makassar Straits seems unlikely since these straits have been shown to be an effective barrier between eastern and western Malesia (Van Tol and Gassmann, 2007).

A northward migration from New Guinea to the Philippines, possibly via Sulawesi (brown arrow in Figure 5), is shown by *P. mindorensis* after ca. 4.1 Mya (Table 3: node 51), although this node lacks statistical support. Another westwards reversal event is evident in *P. oxyacarpa*, which may have migrated from New Guinea to Sulawesi between 3.9 (Table 3: node 52) and 3.7 Mya (Table 3: node 55). A substantial landmass was present in the Sulawesi region between ca. 105 Mya (Hall, 2001), but it was not until the last 5 Ma that the three distinct fragments (peninsular NE, mainland NE, and W Sulawesi) became united (Marshall, 1983), facilitating the migration of plants and animals across Wallacea (Hall, 2001).

RESULTS

Phylogeny and Age Estimations

The BEAST analyses of matrix A resulted in a robust phylogeny of the Annonaceae (Figure 2A), which is largely consistent with the topologies of previous phylogenetic analyses (Richardson et al., 2004) and unpublished phylogenies using MrBayes (based on two and seven genes: L.W. Chatrou, pers. comm.). The mean ages of major nodes, with 95% HPD intervals and Bayesian PP values, are indicated in the chronogram (Figure 2A) and Table 2. The mean rate of evolution is 0.0006 substitutions per site per million years (95% HPD: 5.23E-4–7.10E-4). The “birth rate” (i.e., speciation rate) indicated by the Yule prior is 0.041 (95% HPD: 0.031–0.052). The coefficient of variation is 0.69 (95% HPD: 0.56–0.84), indicating that substitution rate heterogeneity across the tree and that a relaxed clock model is most appropriate (Drummond et al., 2007).

The BEAST chronogram (Figure 2A) is composed of four major clades: *Anaxagorea* (sister to all other members of the Annonaceae); a clade consisting of *Cananga*, *Cleistopholis*, *Cyathocalyx*, and *Tetrameranthus*; the long branch clade (LBC); and the short branch clade (SBC). Each of these clades received strong statistical support (Table 2: PP = 1.00). BEAST analyses based on matrix A gave an estimated tMRCA for *Pseuduvaria* of 8.0 Mya (95% HPD: 11.5–5.0 Mya, PP = 1.00; Figure 2A: node 20). The MCMC result of the posterior distribution for the tMRCA of *Pseuduvaria* (Table 2: node 20) is shifted ca. 44 Mya later, indicating that the sequence data strongly influences date estimation.

The BEAST analyses of matrix B generated a well-resolved phylogeny of *Pseuduvaria*. The mean ages, 95% HPD intervals and Bayesian PP values of all nodes

within *Pseuduvaria*, are indicated in the chronogram (Figure 3) and Table 3. The mean rate of evolution is 0.0005 substitutions per site per million years (95% HPD: 3.57E-4-7.024E-4). The birth rate indicated by the Yule prior is 0.188 (95% HPD: 0.122-0.270). The coefficient of variation is 0.543 (95% HPD: 0.365-0.723), indicating that

Table 2. Prior and posterior age distributions of major nodes of Annonaceae based on matrix A using BEAST analyses

Node	Posterior		Prior		Previous estimates	Bayesian PP
	Mean (Mya)	95% HPD (Mya)	Mean (Mya)	95% HPD (Mya)		
C 1: Magnoliaceae stem age	102.10	106.20-98.0	98.89	100.09-98.00	–	1.00
1: Annonaceae stem age	98.01	101.46-94.91	–	–	90.6 ± 1.3 by NPRS [7], 82 by PL [9]	1.00
C2: Annonaceae crown age	89.46	90.41-89.0	90.05	92.20-89.00	–	1.00
2: split between Cananga-Tetrameranthus and both LBCSBC clade	74.66	84.4-63.62	82.58	91.88-69.53	–	1.00
3: split between LBC and SBC	67.33	78.08-55.22	–	–	66.7-56.6 ± 2.3 by NPRS [7]	1.00
4: Cananga-Tetrameranthus	41.80	62.35-22.25	39.09	71.91-5.75	51.3-43.8 ± 5.1 by NPRS[7]	1.00
5: Cananga-Cyathocalyx	19.45	34.52-6.42	–	–	–	1.00
6: LBC crown node	59.59	70.5-48.07	61.95	85.26-38.61	60.2-51.1 ± 2.3 by NPRS [7]	1.00
7: SBC crown node	39.84	55.17-26.86	70.82	88.09-52.69	62.5-53.1 ± 3.6 by NPRS[7], 58.76 by PL [8]	1.00
8: Guatteria-Asimina	51.54	62.81-40.32	–	–	–	1.00
9: Xylopia-Asimina	46.62	57.1-35.68	–	–	–	0.97
10: Isolona-Asimina	40.23	50.41-30.71	–	–	–	1.00
11: Isolona-Dasymaschalon	31.65	41.45-21.18	–	–	–	1.00
12: Guatteria	7.94	16.3-1.91	–	–	I 1.4 ± 1.4 by PL[10]	1.00
13: Isolona-Monodora	13.85	24.32-4.43	–	–	Mean = ca. 14 by BEAST [11]	1.00
14: Uvaria-Dasymaschalon	19.61	28.42-10.59	–	–	14.8-12.4 ± 2.4 by NPRS[7]	1.00
15: Onychopetalum-Pseuduvaria	29.08	39.19-20.18	–	–	–	1.00
16: Onychopetalum-Karobelia	24.36	37.75-15.77	–	–	37.65 by NPRS[8], 24.76 by	1.00
17: Crematosperma	4.44	9.09-0.90	–	–	22.33 by NPRS, 7.17 by PL[8], 16.49 by Multidivtime [8]	1.00
18: Monocarpia-Pseuduvaria	22.97	31.14-16.01	–	–	–	1.00
19: Neo-uvaria-Pseuduvaria	14.9	20.2-9.7	–	–	–	< 9.5
20: Pseuduvaria	8.01	11.5-5.0	52.11	74.09-29.04	ca. 16 by NPRS[7]	1.00

Node numbers refer to Figure 2A. Mean ages and 95% HPD intervals of divergence times are in millions of years before the present. Age estimates were compared with earlier dating studies for the Annonaceae. Bayesian PP are also indicated. C I and C2 represent the calibration nodes (see texts for details).

Table 3. Posterior estimates of divergence time estimates (Mya) in *Pseuduvaria* based on matrix B using BEAST analyses, with results of ancestral area reconstructions using dispersal-vicariance analysis (DIVA) and weighted ancestral area analysis (WAAA).

	Age estimates	DIVA	WAAA								Bayesian PP	
			A	B	C	D	E	F	G	H		
0	13.6	18.49- 8.85	AH	0.40	0.08	0.40	0.17	0.03	0.43	0.06	0.33	1.00
1	9.64	14.10- 5.29	CH	—	—	1	—	—	—	—	1	1.00
2	5.76	9.37- 2.41	H	—	—	—	—	—	—	—	∞	1.00
3	2.72	5.34- 0.59	C	—	—	∞	—	—	—	—	—	1.00
4	8.34	10.95- 5.82	A	0.93	0.09	0.25	0.29	0.04	0.52	0.07	—	1.00
5	6.62	9.50- 3.83	A	∞	—	—	—	—	—	—	—	0.99
6	4.82	7.31- 2.36	A	∞	—	—	—	—	—	—	—	1.00
7	2.57	4.40- 0.97	A	∞	—	—	—	—	—	—	—	1.00
8	1.72	3.18- 0.44	A	∞	—	—	—	—	—	—	—	<0.95
9	7.68	10.13- 5.23	AC AD AF	0.55	0.11	0.30	0.38	0.05	0.68	0.08	—	<0.95
10	5.62	8.06- 3.15	A	∞	0.09	—	—	—	—	—	—	0.99
11	4.12	6.32- 2.04	A	∞	0.12	—	—	—	—	—	—	<0.95
12	3.10	4.93- 1.33	A	∞	0.18	—	—	—	—	—	—	0.99
13	2.20	3.79- 0.85	A	∞	0.33	—	—	—	—	—	—	0.96
14	1.50	2.83- 0.42	A	∞	1	—	—	—	—	—	—	<0.95
15	4.17	6.43- 2.05	A	∞	—	—	—	—	—	—	—	0.99
16	2.70	4.40- 1.22	A	∞	—	—	—	—	—	—	—	1.00
17	1.99	3.36- 0.82	A	∞	—	—	—	—	—	—	—	0.97
18	1.30	2.37- 0.34	A	∞	—	—	—	—	—	—	—	0.99
19	1.13	2.17- 0.22	A	∞	—	—	—	—	—	—	—	0.98
20	7.38	—	FCF DF	0.2	0.09	0.41	0.62	0.05	1.13	0.09	—	<0.95
21	5.88	8.46- 3.54	DCD DF	0.07	0.05	0.6	1.7	—	0.18	—	—	<0.95
22	5.28	—	CD CD CF	0.09	0.07	1.31	0.71	—	0.33	—	—	<0.95
23	2.52	5.34- 0.32	CF DF	—	—	1	1	—	1	—	—	<0.95
24	4.51	6.53- 2.54	C CD	0.12	0.09	1.79	0.48	—	—	—	—	1.00

Table 3. (Continued)

25	4.04	5.89- 2.25	D CD	0.18	0.12	0.58	1.6	—	—	—	—	<0.95
26	3.28	4.86- 1.76	CAD CD	0.33	0.18	1.6	0.33	—	—	—	—	1.00
27	2.50	3.94- 1.13	CAC	1	0.33	3	—	—	—	—	—	0.99
28	2.06	—	AAC BC	∞	1	1	—	—	—	—	—	<0.95
29	2.27	3.57- 1.09	CD	—	—	1	1	—	—	—	—	1.00
30	1.63	2.73- 0.67	C	—	—	∞	—	—	—	—	—	0.99
31	0.77	1.49- 0.15	C	—	—	∞	—	—	—	—	—	1.00
32	6.78	9.24- 4.47	F	0.09	0.09	0.09	0.12	0.07	2.84	0.11	—	<0.95
33	1.85	4.35- 0.23	F	—	—	—	—	—	∞	—	—	0.96
34	6.54	8.96- 4.29	F	0.12	0.12	0.12	0.18	0.09	1.88	0.15	—	<0.95
35	5.81	—	F	0.18	0.18	0.18	—	—	4.56	0.07	—	<0.95
36	3.88	6.34- 1.63	F	—	—	—	—	—	∞	—	—	<0.95
37	5.37	—	F	0.33	0.33	0.33	—	—	2.54	0.09	—	<0.95
38	4.05	—	F	1	1	1	—	—	1	—	—	<0.95
39	4.93	—	F	—	—	—	—	—	8.32	0.12	—	<0.95
40	3.97	6.08- 1.97	F	—	—	—	—	—	5.55	0.18	—	<0.95
41	1.92	3.59- 0.60	F	—	—	—	—	—	∞	—	—	1.00
42	3.31	—	F	—	—	—	—	—	3	0.33	—	<0.95
43	2.61	—	FG	—	—	—	—	—	3	0.33	—	<0.95
44	1.22	2.61- 0.15	G	—	—	—	—	—	—	∞	—	1.00
45	4.53	6.42- 2.78	F	—	—	—	0.33	0.12	1.16	0.18	—	1.00
46	3.57	5.18- 2.00	F	—	—	—	—	—	∞	—	—	<0.95
47	2.90	4.50- 1.24	F	—	—	—	—	—	∞	—	—	<0.95
48	2.03	3.57- 0.61	F	—	—	—	—	—	∞	—	—	0.98
49	2.00	3.26- 0.86	F	—	—	—	—	—	∞	—	—	1.00
50	0.83	1.64- 0.16	F	—	—	—	—	—	∞	—	—	1.00
51	4.09	—	DF	—	—	—	1.00	0.18	0.18	0.33	—	<0.95
52	3.88	—	FG	—	—	—	—	0.33	0.33	1	—	<0.95
53	1.13	2.31- 0.22	G	—	—	—	—	—	—	∞	—	1.00
54	0.55	1.27- 0.04	G	—	—	—	—	—	—	∞	—	0.99
55	3.67	—	EF	—	—	—	—	1	1	—	—	<0.95
56	2.82	4.59- 1.26	F	—	—	—	—	—	∞	—	—	0.99

Node numbers refer to Figure 3. Posterior mean ages and 95% HPD intervals of divergence times are in millions of years before the present. DIVA reconstructions are shown for maxareas = 2. WAAA reconstructions are shown with probability indices (PI). biogeographical areas are indicated by capital letters (as defined in Figure 1). Bayesian PP are also indicated.

substitution rate heterogeneity across the tree and that a relaxed clock model is appropriate (Drummond et al., 2007). The phylogeny of *Pseuduvaria* is segregated into five clades (Figure 3: clades I–V, numbering follows (Su et al., 2008)). The BEAST topology of *Pseuduvaria* is largely consistent with those resulting from earlier ML and Bayesian analyzes [(Su et al., 2008); Figure 3].

Historical Biogeography of *Pseuduvaria*

The BEAST tree generated using matrix B was used as input for DIVA and WAAA (results summarized in Table 3). The DIVA and WAAA results were largely congruent: the inferred ancestral distributions are shown on the tree (Figures 6 and 7). The DIVA reconstructions required at least 20 dispersal events to explain the present-day distribution when the maximum number of areas was restricted to two at each node.

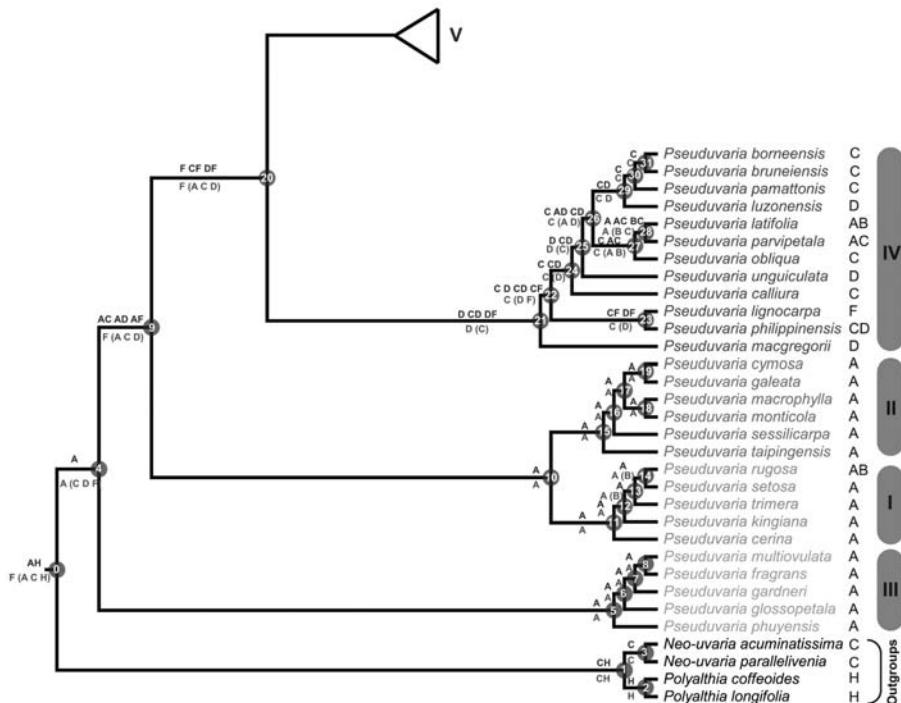


Figure 6. Proposed biogeographical scenarios for *Pseuduvaria* clade IV. Arrows represent possible migration routes (see text for details).

Both DIVA and WAAA indicate that the most likely ancestral area for *Pseuduvaria* was Sundaland (Figure 6: node 4). Diversification of *Pseuduvaria* within Sundaland gave rise to clades I, II, and III. The inferred ancestral area of the combined clade IV–V was ambiguous, however (Figure 6: node 20): DIVA indicated that the ancestral area was either New Guinea, Borneo/New Guinea, or Philippines-New Guinea; whereas WAAA suggested that it was either New Guinea (with the highest PI of 1.13; Table 3)

or the Philippines (PI = 0.62). It is noteworthy, however, that node 20 lacks statistical support.

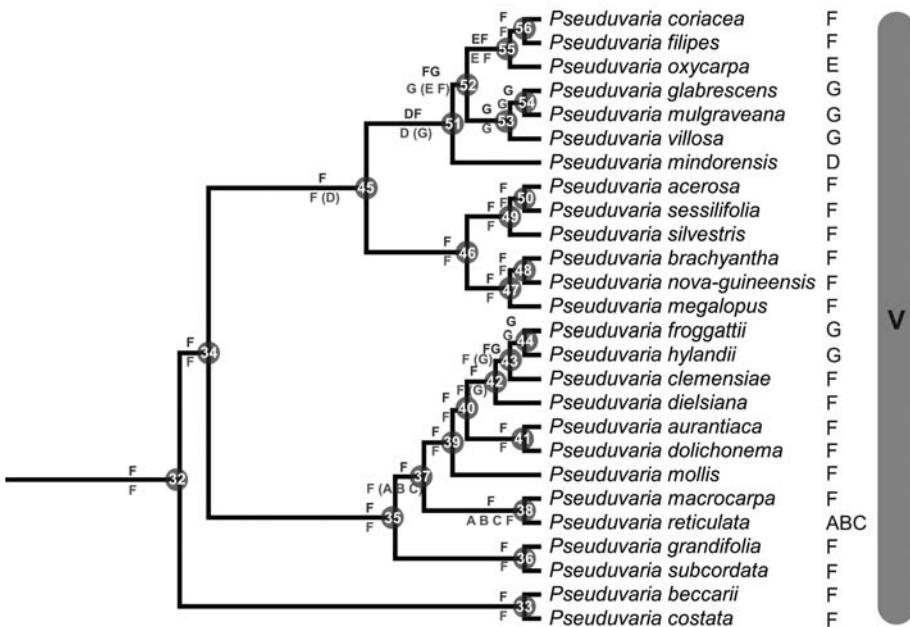


Figure 7. Proposed biogeographical scenarios for *Pseuduvaria* clade V. Arrows represent possible migration routes (see text for details).

The inferred ancestral area for the common ancestor of clade IV (Figure 6: node 21) is also ambiguous. DIVA indicated the inferred distribution was either the Philippines, BorneoPhilippines, or PhilippinesNew Guinea; whereas WAAA indicated that the Philippines was the most probable ancestral area, with the highest PI of 1.70 (Table 3). Within clade IV, a series of dispersal events could have occurred between Borneo and the Philippines, and also possibly between the Philippines and New Guinea. Furthermore, reverse westwards movements of *Pseuduvaria* from Borneo and/or the Philippines can also be inferred (Figure 6: node 28), as shown by the distribution of two species, *P. latifolia* and *P. parvipetala*, in Java and Sumatra.

The DIVA and WAAA results were congruent regarding the inferred ancestral distribution of clade V, both indicating that New Guinea was the most likely region (Figure 7: node 32). Subsequent speciation events occurred within New Guinea, which became the secondary centre of diversity in *Pseuduvaria*. Within clade V, dispersal events presumably occurred from New Guinea to Australia, although a westward migration also seems to have occurred from New Guinea to western Malesia, as evident in one species (*P. reticulata*).

CONCLUSION

The divergence times of the main clades within the Annonaceae (derived here using BEAST) deviate slightly from those previously determined using different calibration points and dating methods. In particular, our estimate for the SBC crown is 39.8 Mya (95% HPD: 55.226.9 Mya), which is much younger than previous estimates of $62.553.1 \pm 3.6$ Mya (Richardson et al., 2004) and 58.76 Mya (Pirie et al., 2006). The present study differs from previous research as it utilizes an uncorrelated lognormal relaxed clock model and is calibrated using the recently described fossil, *Futabanthus*; these factors are likely to enable more precise divergence time estimates within the Annonaceae.

Pseuduvaria is shown to have originated in Sundaland in the late Miocene, ca. 8 Mya (95% HPD: 11.55.0 Mya). Subsequent migration events were predominantly eastwards towards New Guinea and Australia, although several migratory reversals are also postulated. The suggested dispersal episodes are broadly consistent with the available geological data.

The migration of *Pseuduvaria* species into New Guinea from Sundaland and/or the Philippines is of particular biogeographical interest. The DIVA and WAAA reconstructions of the most probable ancestral areas at each node suggest that the earliest migration of *Pseuduvaria* species into New Guinea may have occurred as early as 7.7 Mya (95% HPD: 10.15.2 Mya), although this node is not statistically supported. If such a scenario is correct, it is necessary to invoke long-distance dispersal events since the geological data suggest that biotic migration between the two landmasses would have been unlikely: New Guinea was largely submerged during this period and the Sunda and Sahul landmasses were separated by deep ocean barriers without land bridges (Hall, 2001; Rowe et al., 2008). It was only within the past 5 Ma that significant opportunities may have existed for biotic migration into New Guinea from Sundaland (Hall, 2001), and significantly all the other postulated migrations between New Guinea and Sundaland occurred within this period: from Sundaland to New Guinea between nodes 22 and 23 in clade IV (i.e., after 5.3 Mya); from New Guinea to Sundaland between nodes 37 and 38 in clade V (i.e., after 5.4 Mya); and from New Guinea to Sulawesi between nodes 52 and 55 in clade V (i.e., after 3.9 Mya).

Our previous study of morphological character evolution within *Pseuduvaria* (Su et al., 2008) indicated that the ancestors of clade V were likely to have had fruits with large monocarps (individual fruit segments derived from the maturation of individual carpels within the flower); this inference holds true irrespective of whether the New Guinea species *P. beccarii* and *P. costata* are included in clade IV (as indicated in our previous study (Su et al., 2008)) or included in clade V (present analysis). There is a dearth of observational reports on frugivory and seed dispersal of *Pseuduvaria* species (Su and Saunders, 2006), although it appears that seeds of smaller-fruited species are bird-dispersed whereas those of larger-fruited species are dispersed by primates or fruit bats. It is probable that the long-distance dispersal events hypothesised to explain the early colonization of New Guinea by *Pseuduvaria* species may have been facilitated by fruit bats. Significantly, the majority of New Guinea species of *Pseuduvaria* possess elongate flowering peduncles and/or pedicels (Su et al., 2008): separation of

fruits from the foliage (thereby ensuring that the fruits are conspicuous) is a common feature of plants adapted for fruit-bat frugivory (Marshall, 1983).

KEYWORDS

- **Annonaceae fossils**
- **BEAST topology**
- **NPRS methods**
- **Weighted ancestral area analysis**

AUTHORS' CONTRIBUTIONS

YCFS collected and analysed the molecular data. RMKS conceived the study and obtained funding. Both authors wrote the chapter and have read and approved the final manuscript.

ACKNOWLEDGMENTS

This research was supported by a grant from the Hong Kong Research Grants Council (7358/03M). We thank Lars Chatrou for sending unpublished DNA sequences of two outgroups, Justin Bahl, Dhanesekaran Vijaykrishna, Alexei Drummond, Thomas Couvreur, and Gavin Smith for helpful advice and discussions on using BEAST, Jason Ali for discussions on the geological history of Southeast Asia, and Jim Doyle and two anonymous reviewers for their valuable comments on the manuscript.

Chapter 19

Early Cretaceous of Northern Japan

Toshihiro Yamada, Harufumi Nishida, Masayoshi Umebayashi,
Kazuhiko Uemura, and Masahiro Kato

INTRODUCTION

Molecular phylogenetic analyses have identified Trimeniaceae, a monotypic family distributed only in Oceania, as among the earliest diverging families of extant angiosperms. Therefore, the fossils of this family are helpful to understand the earliest flowering plants. Paleobotanical information is also important to track the historical and geographical pathways to endemism of the Trimeniaceae. However, fossils of the family were previously unknown from the Early Cretaceous, the time when the angiosperm radiated. In this study, we report a seed from the late Albian (ca. 100 million years ago) of Japan representing the oldest known occurrence of Trimeniaceae and discuss the character evolution and biogeography of this family.

A structurally preserved seed was collected from the early Late Albian Hikag-enosawa Formation of the Yezo Group, which was deposited in palaeolatitudes of 35–40°N. The seed has a multilayered stony exotesta with alveolate surface, parenchymatous mesotesta, and operculate inner integument, which are characteristic to extant trimeniaceous seeds. However, the seed differs from extant seeds, that is, in its well-developed endosperm and absence of antiraphal vascular bundle. Thus, the seed would be a new genus and species of Trimeniaceae.

The fossil seed indicates that seed coat characters were conserved for 100 million years or more in Trimeniaceae. It also suggests that the antiraphal vascular bundle and perisperm originated secondarily in Trimeniaceae as previously inferred from the phylogeny and character distribution in the extant basalmost angiosperms. The fossil seed provides the first evidence that Trimeniaceae was distributed in a midlatitude location of the Northern Hemisphere during the Early Cretaceous, when angiosperms radiated extensively, supporting a hypothesis that the extant austral distribution is relict.

Trimeniaceae is a small family consisting of only the genus *Trimenia*, with eight species known from Celebes to eastern Australia and the southwest Pacific (Kubitzki et al., 1993; Wanger and Lorence 1999). *Trimenia* consists of shrubs and lianas that have unisexual or bisexual flowers with numerous tepals and stamens and a single carpel (Endress and Sampson, 1983; Kubitzki et al., 1993; Money et al., 1950; Wanger and Lorence 1999). A solitary pendant ovule is enclosed in the ovary and develops into a seed with a stony seed coat contained in a berry (Endress and Sampson, 1983; Endress and Igersheim, 1997; Kubitzki et al., 1993; Prakash, 1998; Yamada et al., 2003). The family was placed in the Laurales based on floral structures, such as the single

carpel and ovule per flower (Endress and Sampson, 1983; Money et al., 1950), but molecular phylogenetic studies have identified it as one of the earliest diverging families of extant angiosperms, along with the Amborellaceae, Nymphaeales (Cabombaceae and Nymphaeaceae), Hydatellaceae, Austrobaileyaceae, and Illiciaceae (Jansen et al., 2007; Moore et al., 2007; Qiu et al., 1999; Saarela et al., 2007). Thus, clarifying the primitive character states of Trimeniaceae would be helpful in understanding the earliest flowers. In particular, information on key innovations of angiosperms, for example, fruit, bitegmic seeds, and endosperm, would provide clues to solving the origin of angiosperms, a central mystery of plant evolution (Friedman, 2006; Frohlich, 2006).

Great efforts have been made to find evidence of early angiosperms from the Early to mid-Cretaceous, when the angiosperms radiated extensively (Friis et al., 1999; Frumin and Friis, 1999; Friis et al., 2000, 2001, 2006; Sun et al., 1998; Sun et al., 2002). The early emergence of Nymphaeales (Friis et al., 2001; Friis et al., 2003) and Illiciaceae (Frumin and Friis, 1999) in the fossil record partly supports the phylogenetic framework inferred from molecular data. However, direct evidence that the other earliest diverging families emerged in the early phase of angiosperm radiation (Friis et al., 2006) has not been found, although paleobotanical records (Frumin and Friis, 1999; Friis et al., 2001), as well as phylogeny (Jansen et al., 2007; Moore et al., 2007; Qiu et al., 1999; Saarela et al., 2007), imply that these families should have diverged by the mid-Cretaceous (Schneider et al., 2004).

From the austral distribution of Trimeniaceae, and the similar distribution of Amborellaceae and Austrobaileyaceae (Kubitzki et al., 1993), neobotanists have inferred the Gondwana origin of these families (Bailey, 1949; Cronquist, 1998), but accumulated palynological data have indicated that a different explanation is needed for this distribution. These data indicate that angiosperms originated in low paleolatitudes (20°N to 20°S) no later than the Hauterivian, Early Cretaceous, about 132 million years ago (Crane and Lidgard, 1989; Hughes et al., 1989) and spread toward the poles over time (Axelrod, 1954; Crane and Lidgard, 1989; Doyle, 1984). Pollen records provide temporal and spatial distribution patterns of angiosperms, with the implication that the Trimeniaceae could have migrated to Laurasia, a continental mass in the Northern Hemisphere. However, pollen is difficult to assign to extant families due to the frequent convergent evolution of pollen morphology, such as the number and position of apertures (Friis et al., 2006; Sampson, 2007). Moreover, extant Trimenia pollens are eurypalynos with regard to the number of aperture even in a same species (Sampson, 2007), obscuring diagnostic pollen characters in familiar level. Thus, findings from other reproductive structures, such as flowers, fruits, and seeds, are needed to confirm the past existence of the Trimeniaceae in the Northern Hemisphere.

We report the oldest seed of Trimeniaceae from the Early Cretaceous Yezo Group in Hokkaido, northern Japan, and discuss character evolution and biogeography of the Trimeniaceae.

Geological Settings and Other Paleobotanical Data of Collection Site

A structurally preserved seed was found within a calcareous siltstone nodule collected at Pombetsu, Mikasa City, Hokkaido, Japan (Figure 1). The Hikagenosawa Formation

of the Yezo Group outcrops in the area and contains ammonoids indicating a date in the early Late Albian (Takashima et al., 2004; Narita et al., 2008). Palaeomagnetic studies show that these sediments were deposited in palaeolatitudes of 35–40°N and that the sedimentary basin was located on the eastern side of Laurasia (Takashima et al., 2004; Kodama et al., 2002). Because the formation mainly consists of offshore sediments, there are few palaeobotanical records within it (Takahashi et al., 1995; Saiki, 1997; Takahashi and Suzuki, 2003), but a subtropical climate with dry seasons seems to have prevailed in the area, as suggested by the occurrence of a cheirolepidiaceous conifer (Saiki, 1997) and a variety of ephedroid palynomorphs (Takahashi et al., 1995).

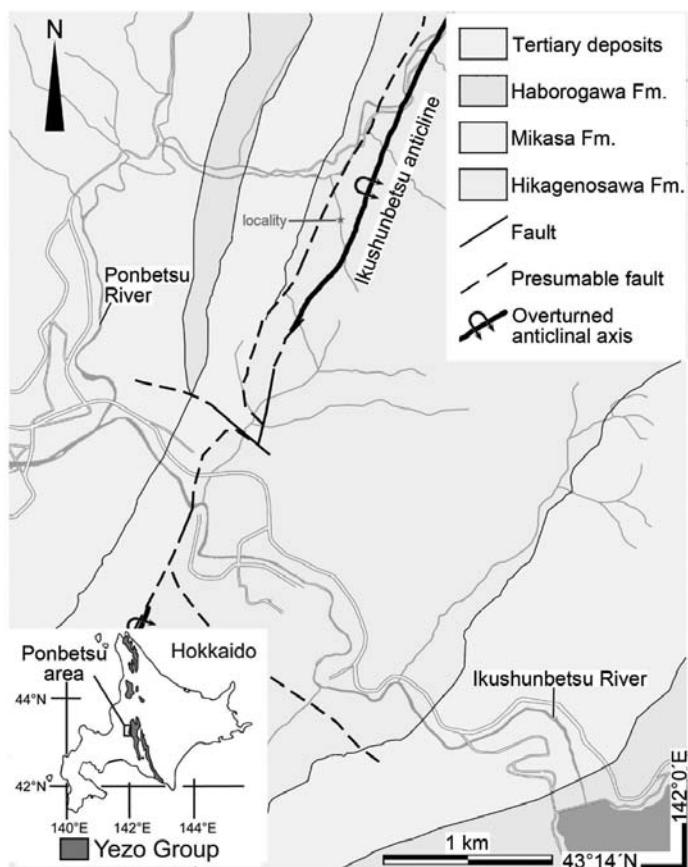


Figure 1. Geological map of Ponbetsu area and locality of the trimeniaceous seed. Ponbetsu area are boxed in large-scale map of Hokkaido on the left-bottom corner. Geological map is redrawn from Narita et al. 31.

MATERIALS AND METHODS

Materials (Yamada 001002) and methods for plastic sections and SEM micrograph of *Trimenia moorei* (extant) were described previously (Yamada et al., 2003). The fossil

seed was sectioned in planes tangential to the raphe, and the extant seed was sectioned in planes parallel to the raphe. A part of the fossil seed, 1.5 mm thick, was cut into 38 successive peel sections, that is, one section is ca. 40 µm thick. For reconstruction, the outline of the seed surface was traced, and surface views were compiled manually by overlying successive camera lucida drawings. The slides of serial sections and the nodule containing the seed are stored in the National Museum of Nature and Science, Tokyo, Japan, as NSM-PP-9176.

DISCUSSION

Relationships of the Fossil Seed

Although seed structures have not been examined in all *Trimenia* species, available information on four species (*T. moorei*, *T. neocalledonica*, *T. papuana*, and *T. weinmannifolia*) suggests that Trimeniaceae is uniform family in regard of ovule and seed structures (Endress and Sampson, 1983; Kubitzki et al., 1993; Yamada et al., 2003). The seed of extant *Trimenia* is characterized by the testa, consisting of a lignified multilayered exotesta with an alveolate surface and a nonlignified multilayered mesotesta (Figures 2b, d, i, j) (Endress and Sampson, 1983; Kubitzki et al., 1993; Yamada et al., 2003). Although many families possess seed coats with a hardened exotesta, none other than Trimeniaceae is known to possess a multilayered exotesta (Corner, 1976; Endress and Sampson, 1983; Kubitzki et al., 1993; Yamada et al., 2003.). The seeds of some taxa are similar in appearance to the Trimeniaceae seed in that the testa consists of outer sclerotic and inner non-sclerotic cell layers. These taxa include the genus *Nuphar* (Nymphaeaceae) (Collinson, 1980; Yamada et al., 2001), the families Buxaceae (Corner, 1976), Simmondsiaceae (Tobe et al., 1992), Melianthaceae (Doweld, 2001), Myrtaceae (Corner, 1976, Landrum and Sharp, 1989), and Theaceae (Corner, 1976), and the order Sapindales (Corner, 1976; Doweld, 1996). However, in these taxa, the stony outer layer is composed of a single-layered columnar exotesta and the outer part of the isodiametric mesostestal cells, whereas the exotesta of *Trimenia* is composed uniformly of sclerotic isodiametric cells, which are derivatives of the outer epidermis of the outer integument (Yamada et al., 2003). Thus, to our knowledge, no other seeds can be compared to the fossil seed.

The fossil seed is not distinguishable from that of extant *Trimenia* in terms of shape, size, anatropy, bitegmy, endo, and exostomic micropyle, crushed inner integument with an operculum, lignified multilayered exotesta composed of isodiametric cells, alveolate surface of the exotesta, or nonlignified multilayered mesotesta (Figure 2). These remarkable similarities unequivocally indicate the close affinity of the fossil seed to the Trimeniaceae. The fossil seed differs from *Trimenia* seed in the absence of antiraphal vascular bundle (Figure 2e) and in the size of membranous structures surrounding the nucellus (Figures 2c, g). Thus, a new genus of Trimeniaceae should be assigned to the fossil seed. The seed also differs from *Trimenia* seed in having a stalklike structure at the nucellar base (Figure 2a). This stalklike structure might represent a diagnostic feature of the seed, but the structure could be artificially formed by shrinkage.

Divergence Time of Trimeniaceae

Molecular phylogeny indicates that the Trimeniaceae diverged after the Nymphaeales and before the Illiciaceae (Jansen, 2007; Moore et al., 2007; Qiu et al., 1999; Saarela et al., 2007). Thus, diversification of the Trimeniaceae in the Early to earliest Late Cretaceous (from 125 to 90 million years ago) is indicated (Schneider et al., 2004) by the Late Barremian flower of the Nymphaeales from Portugal (Friis et al., 2001) and the Cenomanian to Turonian seeds of the Illiciaceae from Kazakhstan (Frumin and Friis, 1999). Although Trimeniaceae-like triporate or polyporate pollen was collected in the Albian to Cenomanian of Brazil (Dettmann and Jarzen, 1990; Herngreen, 1993), the Barremian of Portugal (Friis et al., 1999; Friis et al., 2000) and the Campanian to Maastrichtian (83–65 million years ago) of Australia and Antarctica (Dettmann and Jarzen, 1990; Dettmann, 1994), the possibility could not be ruled out that these are assigned to other families (Dettmann and Jarzen, 1990; Friis et al., 2006; Sampson, 2007; Walker and Doyle, 1994). *Longstrethia varidentata*, a foliar species reported from the Cenomanian of Nebraska, is similar to leaves of Trimeniaceae in presence of an intermarginal vein, but *L. varidentata* would represent a stem-group taxon of Trimeniaceae–Illiciaceae clade because some characters are also shared with Illiciaceae (Upchurch and Dilcher, 1915). Therefore, no unequivocal record of Trimeniaceae in the Cretaceous existed. The fossil seed provides the first unequivocal evidence of Trimeniaceae 100 million years ago that fills the gap between molecular data and paleobotanical records.

Character Evolution in Trimeniaceae

The fossil seed coat structures are strikingly similar to those of *Trimenia* (Figure 2). Although these characters were acquired 100 million or more years ago and are conserved in Trimeniaceae, the mode of nutrient storage of the fossil seed would have differed considerably from that of the extant *Trimenia*. The *Trimenia* embryo is enclosed by the endosperm, which is further surrounded by the perisperm, an additional storage tissue (Figures 2d, i) (Endress and Sampson, 1983; Prakash, 1998; Yamada et al., 2003). However, the fossil seed has two membranous structures inside the nucellus, and the area encircled by the inner structure corresponds in size to the dormant mature embryo of extant Trimeniaceae (Figures 2g, i). If this comparison is correct, the outer structure may be comparable to the endosperm membrane. In albuminous seeds without a perisperm, the epidermis of the nucellus is either completely disintegrated by the enlarged endosperm or vestigially retained as a crushed layer, whereas in perispermous seeds, it is retained as the epidermis of the perisperm (Corner, 1976; Prakash, 1998; Yamada et al., 2003). The fossil has a relatively well preserved nucellar epidermis, which may imply the presence of a perisperm. The putative endosperm in the fossil occupies a large area inside the nucellus, and thus the endosperm would have played a major storage role. Among the earliest families to diverge, only Nymphaeales (Yamada et al., 2001), Hydatellaceae (Rudall et al., 2007), and Trimeniaceae (Endress and Sampson, 1983; Prakash, 1998; Yamada et al., 2003) have a diploid maternal perisperm, which occupies a larger area than the diploid or polyploid fertilized endosperm. This character variation implies that the storage function was largely taken

over by the perisperm secondarily (Doyle and Endress, 2000; Floyd and Friedman, 2000) and that the fossil seed may have been on the way toward perispermity.

The fossil seed vasculature is different from that of extant *Trimenia*. In *Trimenia*, the vascular bundle supplied from the fruit wall enters the raphe of the seed and extends to the antiraphal side beyond the chalaza (Endress and Sampson, 1983). Lack of antiraphal vascular bundles in the fossil seed suggests that the antiraphal bundle is a derived character in the Trimeniaceae. This is consistent with the hypothesis, based on character distribution in extant basalmost angiosperms, that the antiraphal bundle is a derived character in the angiosperm seed (Prakash, 1998; Tobe et al., 2000).

So far, we know only the seed of this trimeniaceous fossil, but the areoles sculptured in the seed surface may imply that the fossil seed was contained in a berry because the pressure of berry endocarp cells forms areoles on seeds of extant *Trimenia* (Endress and Sampson, 1983; Endress and Igersheim, 1997; Kubitzki et al., 1993; Money et al., 1950; Prakash, 1998; Wanger and Lorence, 1999; Yamada, 2003).

Biogeography

Extant Trimeniaceae species are distributed in eastern Australia and an island chain stretching from Celebes to the Moluccas, New Guinea, New Caledonia, Fiji, Samoa, and the Marquesas (Endress and Sampson, 1983; Kubitzki et al., 1993; Wanger and Lorence, 1999). Along with the Trimeniaceae, other basalmost families (Amborellaceae, Hydatellaceae, Austrobaileyaceae) and many eumagnoliid families (e.g., Degeneriaceae, Monimiaceae, Winteraceae) now grow only in derivative fragments of Gondwanaland (Endress and Sampson, 1983; Hamann, 1998; Kubitzki et al., 1993). Contrary to their extant austral distribution, palynological records suggest that early angiosperms originated in the low latitudes (Crane and Lidgard, 1989; Hughes et al., 1991) and migrated both northward and southward (Crane and Lidgard, 1989; Doyle, 1984). Thus, the paleobotanical data indicate that the current austral distributions do not reflect the cradle of angiosperms but the area of conservation (Crane and Lidgard, 1989; Doyle, 1984). At the same time, the poleward migration model predicts the past occurrence of these basalmost angiosperm families in the Northern Hemisphere. Some pollen grains, tentatively assigned to the Amborellaceae (Hughes and McDougall, 1987) or Trimeniaceae (Friis et al., 1999, 2000, 2006) are reported from the Northern Hemisphere, but familial assignations of pollen are difficult (Friis et al., 2006; Sampson, 2007; Walker and Doyle, 1994). The fossil seed reveals that Trimeniaceae occurred in the eastern margin of Laurasia in the Northern Hemisphere during the Albian period and indicates a reduction to the relict area subsequent to the hypothesized bipolar migration.

RESULTS

The seed is ellipsoid, 3 mm long and 2.2 mm thick. It has an inner integument and an outer seed coat (testa) (Figures 2a, c, e). The micropyle is formed by both inner integument and testa (endostome and exostome, respectively) and adjacent to the hilum, which contains tracheids and sclerenchymatous fibers, that is, the seed is anatropous (Figures 2e, f). These tracheids and fibers are found only in sections through the raphal

side, thus the vascular bundle does not extend to the antiraphal side beyond the chalaza (Figure 2e). The inner integument is almost crushed, except in the endostomic region, where the inner integument is thickened to form the operculum (Figures 2c, f). The testa consists of an outer (exotesta) and inner (mesotesta) part. The exotesta comprises one to five layers of isodiametric cells, while the mesotesta is made up of four to five layers of longitudinally elongated cells. The exotesta is lignified to give mechanical strength to the seed coat (Figure 2c). In a nearly tangential section through the seed surface, there are polygonal openings in the exotesta, indicating the presence of polygonal areoles on the exotestal surface (asterisks in Figure 2h). This configuration was further examined by reconstructing a three-dimensional image of the seed surface by compiling serial peel sections, revealing that the exotestal surface is alveolate (Figure 2k). Inside the inner integument are three membranous structures. The outermost structure, connected to the chalaza, is the nucellar epidermis (Figure 2a and arrowheads in Figures 2c, g). The innermost structure is observed only in sections through the center of the seed and encircles an area 200 µm in length and 40 µm in width (arrowhead 3 in Figure 2g).

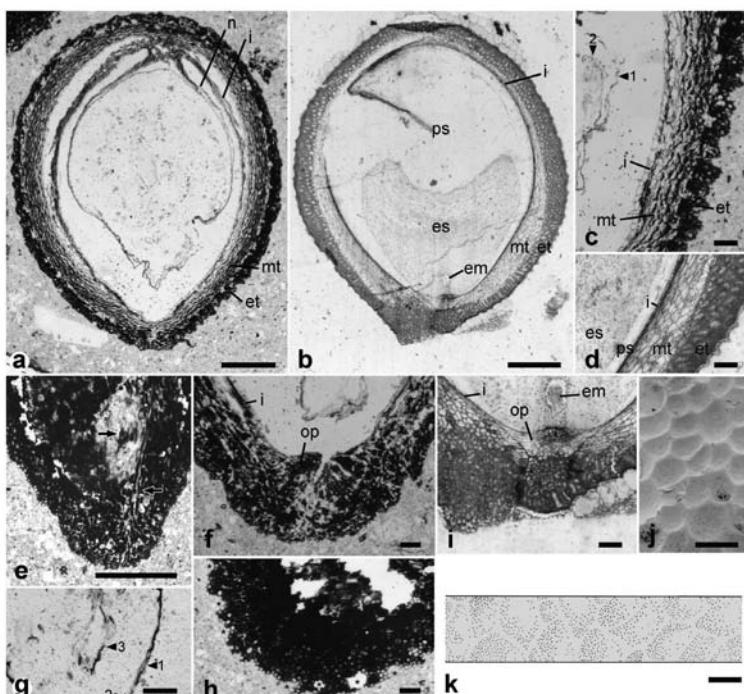


Figure 2. Fossil and extant seeds of Trimeniaceae share many morphological characters. Section of fossil (a, c, e–h) and seed of *Trimenia moorei* (b, d, i); surface of *T. moorei* (j) and fossil (k) seed. Close-up (c) of the micropylar part of (a); (g) is near apex of the nucellus. em, embryo; es, endosperm; et, exotesta; i, inner integument; mt, mesotesta; op, operculum; ps, perisperm. (c, g) arrowheads indicate the nucellar epidermis (1), endosperm membrane (2) and embryo (3). (e) Arrows show tracheids and fibers observed in the hilum. (h) Asterisks indicate areoles seen in nearly paradermal sections of the surface. Scale bars in a and b, 500 µm; e, 200 µm; others, 100 µm.

CONCLUSION

Fossil seed of Trimeniaceae is described from the Early Cretaceous (ca. 100 million years ago) Yezo Group in Hokkaido, northern Japan. The seed, which is the oldest yet found for the family, indicates; (1) Some seed coat structures of Trimeniaceae have been conserved for about 100 million years, including the multilayered stony exotesta with alveolate surface, parenchymatous mesotesta, and operculate inner integument. (2) The secondary origins of the perisperm and antiraphal vascular bundle. (3) Trimeniaceae was distributed in a midlatitude location of the Northern Hemisphere during the Early Cretaceous, when angiosperms radiated extensively, supporting that the extant austral distribution is relict.

KEYWORDS

- **Early Cretaceous**
- **Fossil seed**
- **Trimeniaceae**
- ***Trimenia moorei***

AUTHORS' CONTRIBUTIONS

Toshihiro Yamada performed the field survey, found the fossil and collected data on the fossil and extant seeds. Harufumi Nishida performed the field survey and provided facilities for making peel sections. Masayoshi Umebayashi drew the reconstruction of the fossil. Kazuhiko Uemura collected data on fossils and provided facilities for making peel sections. Masahiro Kato organized the study and collected data on extant seeds. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

We thank K. Kobayashi and N. Kobayashi for kindly providing us lodgings. The Sorachi Forest Management Office permitted us to conduct our field survey within its boundary. This research was supported by Grants-in-Aid for scientific research from the Japan Society for the Promotion of Science to Toshihiro Yamada and Masahiro Kato.

Permissions

Chapter 1: Paleobotany of Livingston Island was originally published as “Paleobotany of Livingston Island: The first report of a Cretaceous fossil flora from Hannah Point” by USGS-2007. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 2: Targeted Isolation Sequence Assembly and Characterization of White Spruce was originally published as “Targeted isolation, sequence assembly and characterization of two white spruce (*Picea glauca*) BAC clones for terpenoid synthase and cytochrome P450 genes involved in conifer defence reveal insights into a conifer genome” in *BioMed Central Ltd* 8:6, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 3: Expressed Sequence Tag Analysis in *Cycas* was originally published as “Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant” in *BioMed Central Ltd* 11:18, 2003. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 4: Complete Chloroplast Genome Sequence of a Tree Fern *Alsophila Spinulosa* was originally published as “Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes” in *BioMed Central Ltd* 6:11, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 5: Reverse Genetic Analyses of Gene Function in Fern Gametophytes was originally published as “A systemic gene silencing method suitable for high throughput, reverse genetic analyses of gene function in fern gametophytes” in *BioMed Central Ltd* 4:16, 2004. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 6: Distribution and Dynamics of Hayscented Fern was originally published as “Distribution and dynamics of hyscented fern following stand harvest” by USDA-2008. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 7: Glossopterid Seed Ferns from the Late Permian, Antarctica was originally published as “Ovule-bearing reproductive organs of the glossopterid seed ferns from the Late Permian of the Beardmore Glacier region, Antarctica” by USGS-2007. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 8: Unveiling Cryptic Species Diversity of Flowering Plants was originally published as “Unveiling cryptic species diversity of flowering plants: successful biological species identification of Asian *Mitella* using nuclear ribosomal DNA sequences” in *BioMed Central Ltd* 5:16, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 9: Pollen Development in *Annona Cherimola* Mill. (Annonaceae) was originally published as “Pollen development in *Annona cherimola* Mill. (Annonaceae). Implications for the evolution of aggregated pollen” in *BioMed Central Ltd* 10:29, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 10: DNA Barcoding in the Cycadales was originally published as “DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads” in *PLoS ONE* 11:7, 2007. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 11: Chloroplast Genomes in Ancestral Grasses was originally published as “Episodic Evolution and Adaptation of Chloroplast Genomes in Ancestral Grasses” in *PLoS ONE* 4:24, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 12: EST Analysis in *Ginkgo biloba* was originally published as “EST analysis in *Ginkgo biloba*: an assessment of conserved developmental regulators and gymnosperm specific genes” in

BioMed Central Ltd 10:15, 2005. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 13: Frequent Fires in Ancient Shrub Tundra was originally published as “Frequent Fires in Ancient Shrub Tundra: Implications of Paleorecords for Arctic Environmental Change” in *PLoS ONE* 3:5, 2008. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 14: History of Native Plant Communities in the South was originally published as “The History of Native Plant Communities in the South” by *USDA-2002*. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 15: Predicting Pleistocene Climate from Vegetation in North America was originally published as “Predicting Pleistocene Climate from Vegetation in North America” in *Clim* 2:12, 2007. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 16: Past Vegetation Patterns of New Mexico’s Rio Del Oso Valley was originally published as “Modeling Landscapes and Past Vegetation Patterns of New Mexico’s Rio Del Oso Valley” by *USDA-2005*. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 17: East Asian Monsoon and Paleoclimatic Data Analysis was originally published as “East Asian Monsoon and paleoclimatic data analysis: a vegetation point of view” in *Clim* 2:20, 2008. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 18: Evolutionary Divergence Times in the *Annonaceae* was originally published as “Evolutionary divergence times in the Annonaceae: evidence of a late Miocene origin of *Pseuduvaria* in Sundaland with subsequent diversification in New Guinea” in *BioMed Central Ltd* 7:2, 2009. Reprinted with permission under the Creative Commons Attribution License or equivalent.

Chapter 19: Early Cretaceous of Northern Japan was originally published as “Oldest record of Trimeniaceae from the Early Cretaceous of northern Japan” in *BioMed Central Ltd* 5:8, 2008. Reprinted with permission under the Creative Commons Attribution License or equivalent.

References

1

- Askin R.A. (1989). Endemism and heterochrony in the Late Cretaceous (Campanian) to Paleocene palynofloras of Seymour Island, Antarctica: Implications for origins, dispersal and palaeoclimates of southern floras, in origins and evolution of the Antarctic Biota, In J.A. Crame (Ed.). *Geol. Soc. Lond. Spec. Publ.* **47**:107–119.
- Askin R.A. and Spicer R.A. (1995). The Late Cretaceous and Cenozoic History of Vegetation and Climate at Northern and Southern High Latitudes: A comparison. *Effects of Past Global Change on Life, Panel on Effects of Past Global Change on Life*, In The Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources, National Research Council (Ed.), 156–173 pp., Washington, D.C., USA, National Academy Press.
- Birkenmajer K. and Zastawniak E. (1989). Late Cretaceous-Early Tertiary Floras of King George Island, West Antarctica: Their stratigraphic distribution and palaeoclimatic significance. *Origins and Evolution of the Antarctic Biota*, In J.A. Crame (Ed.). *Geol. Soc. Lond. Spec. Publ.* **47**:227–240.
- Cantrill D.J. and Poole I.P. (2002). Cretaceous patterns of floristic change in the Antarctic Peninsula. *Palaeobiogeography and Biodiversity Change: A Comparison of the Ordovician and Mesozoic-Cenozoic Radiations*. In A. Owens and J.A. Crame (Eds.). *Geol. Soc. Lond. Spec. Publ.* **194**:141–152.
- Cantrill D.J. and Poole I.P. (2005). Taxonomic turnover and abundance in Cretaceous to Tertiary wood floras of Antarctica: Implications for changes in forest ecology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **215**:205–219.
- Chapman J.L. and Smellie J.L. (1992). Cretaceous fossil wood and palynomorphs from Williams Point, Livingston Island, Antarctic Peninsula, *Rev. Palaeobot. Palynol.* **74**:163–192.
- Crame J.A. (1992). Late Cretaceous palaeoenvironments and biotas: An Antarctic perspective, *Ant. Sci.* **4**:371–382.
- Dettmann M.E. (1986). Significance of the Cretaceous-Tertiary spore genus Cyatheacidites in tracing the origin and migration of Lophosoria (*Filicopsida*), *Spec. Pap. Paleontol.* **35**:63–94.
- Dettmann M.E. (1989). Antarctica: Cretaceous cradle of austral temperate rainforests? *Origins and Evolution of the Antarctic Biota*. In J.A. Crame (Ed.). *Geol. Soc. Lond. Spec. Publ.* **47**:89–105.
- Dettmann M.E., Molna R.E., Douglas J.G., Burger D., Fielding C., Clifford H.T., Francis J.E., Jell P., Rich T., Wade W., Rich P.V., Pledge N., Kemp A., and Rozefelds A. (1992). Australian Cretaceous terrestrial faunas and floras: biostratigraphic and biogeographic implications. *Cret. Res.* **13**:207–262.
- Dettmann M.E. and Thomson M.R.A. (1987). Cretaceous palynomorphs from the James Ross Island area, Antarctica—a pilot study. *Bull. B.A.S.* **77**:13–59.
- Douglas J.G. and Williams G.E. (1982). Southern polar forests: The Early Cretaceous floras of Victoria and their palaeoclimatic significance. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **39**:171–185.
- Hathway B. (2000). Continental rift to back-arc basin: Jurassic–Cretaceous stratigraphical and structural evolution of the Larsen Basin, Antarctic Peninsula. *J. Geol. Soc. Lond.* **157**:417–432.
- Hayes P.A. (1999). Cretaceous angiosperm floras from Antarctica. Unpublished Ph.D. thesis, Leeds, UK, University of Leeds.
- Hobbs G.J. (1968). The geology of the South Shetland Islands. IV. The Geology of Livingston Island. *B.A.S. Sci. Rep.* **47**.
- Pallàs R., Soriano C., Zheng X., Sàbat F., and Casas J.M. (1999). Volcanic Stratigraphy of Hannah Point, Livingston Island, South Shetland Islands, Antarctica. *Acta Geologica Hispanica* **34**(4):323–328.
- Poole I.P. and Cantrill D.J. (2001). Fossil woods from Williams Point Beds, Livingston Island, Antarctica: a Late Cretaceous southern high latitude flora. *Palaeontology* **44**:1081–1112.

- Smellie J.L., Pallàs R., Sàbat F., and Xiangsheng Z. (1996). Age and correlation of volcanism in central Livingston Island, South Shetland Islands: K-Ar and geochemical constraints. *J.S. Am. Earth Sci.* **9**(3/4):265–272.
- Spicer R.A., Rees P.M., and Chapman J.L. (1993). Cretaceous phytogeography and climate signals. *Phil. Trans. Roy. Soc. Lond., Ser. B* **341**:277–286.
- Truswell E.M. (1990). Cretaceous and Tertiary vegetation of Antarctica: A palynological perspective. *Antarctic Palaeobiology: Its role in the reconstruction of Gondwana*. In T.N. Taylor and E.L. Taylor (Eds.), pp. 71–88, Berlin, Springer Verlag.
- Xiangshen Z., Sàbat F., and Smellie J.L. (1996). Mesozoic-Cenozoic Volcanism on Livingston Island, South Shetland Islands, Antarctica: Geochemical Evidences for Multiple Magma Generation Processes. *Korean J. Polar Res.* **7**(1/2):35–45.
- AGI. (2008). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796–815.
- Ahuja M.R. and Neale D.B. (2005). Evolution of genome size in conifers. *Silvae Genetica* **54**:126–137.
- Bedon F., Levasseur C., Grima-Pettenati J., Seguin A., and MacKay J. (2009). Sequence analysis and functional characterization of the promoter of the *Picea glauca* Cinnamyl Alcohol Dehydrogenase gene in transgenic white spruce plants. *Plant Cell Rep.* **28**:787–800.
- Bérubé Y., Zhuang J., Ralph S., Rungis D., Bohlmann J., and Ritland K. (2007). Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genetics & Genomics* **3**:251–259.
- Bohlmann J. (2008). Insect-induced terpenoid defenses in spruce. *Induced Plant Resistance to Herbivory*. In Springer Science (Ed.), pp. 173–187.
- Byun-McKay A., Godard K.A., Toudefallah M., Martin D.M., Alfaro R., King J., Bohlmann J., and Plant A.L. (2006). Wound-induced terpene synthase gene expression in Sitka spruce that exhibit resistance or susceptibility to attack by the white pine weevil. *Plant Physiol.* **140**:1009–1021.
- Chen N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics Chapter 4*(Unit 4):10.
- Cui L., Wall P.K., Leebens-Mack J.H., Lindsay B.G., Soltis D.E., Doyle J.J., Soltis P.S., Carlson J.E., Arumuganathan K., and Barakat A., et al. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**:738–749.
- De Bodt S., Maere S., and Peer Y. (2005). Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**:591–597.
- de la Bastide M. and McCombie W.R. (2007). Assembling genomic DNA sequences with PHRAP. *Curr. Protoc. Bioinformatics Chapter 11*(Unit 11):14.
- Eulgem T., Rushton P.J., Robatzek S., and Somssich I.E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**:199–206.
- Ewing B. and Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
- Fäldt J., Martin D., Miller B., Rawat S., and Bohlmann J. (2003). Traumatic resin defense in Norway spruce (*Picea abies*): Methyl jasmonate-induced terpene synthase gene expression, and cDNA cloning and functional characterization of (+)-3-carene synthase. *Plant Mol. Biol.* **51**:119–133.
- Friedmann M., Ralph S.G., Aeschliman D., Zhuang J., Ritland K., Ellis B.E., Bohlmann J., and Douglas C.J. (2007). Microarray gene expression profiling of developmental transitions in Sitka spruce (*Picea sitchensis*) apical shoots. *J. Exp. Bot.* **58**:593–614.
- Friesen N., Brandes A., and Heslop-Harrison J.S. (2001). Diversity, origin, and distribution of retrotransposons (gypsy and copia) in conifers. *Mol. Biol. Evol.* **18**:1176–1188.
- Giuliano G., Pichersky E., Malik V.S., Timko M.P., Scolnik P.A., and Cashmore A.R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci. USA* **85**:7089–7093.

- Godard K.A., Byun-McKay A., Levasseur C., Plant A., Séguin A., and Bohlmann (2007). Testing of a heterologous, wound- and insect-inducible promoter for functional genomics studies in conifer defense. *Plant Cell Rep.* **26**:2083–2090.
- Goldsborough A.P., Albrecht H., and Stratford R. (1993). Salicylic acid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved amongst stress-inducible genes. *Plant J.* **3**:563–571.
- Hamberger B. and Bohlmann J. (2006). Cytochrome P450 mono-oxygenases in conifer genomes: Discovery of members of the terpenoid oxygenase superfamily in spruce and pine. *Biochem. Soc. Trans.* **34**:1209–1214.
- Hoen D.R., Park K.C., Elrouby N., Yu Z., Mohnhabir N., Cowan R.K., and Bureau T.E. (2006). Transposon-mediated expansion and diversification of a family of ULP-like genes. *Mol. Biol. Evol.* **23**:1254–1268.
- Holliday J.A., Ralph S.G., White R., Bohlmann J., and Aitken S.N. (2008). Global monitoring of autumn gene expression within and among phenotypically divergent populations of Sitka spruce (*Picea sitchensis*). *New Phytol.* **178**:103–122.
- Holligan D., Zhang X., Jiang N., Pritham E.J., and Wessler S.R. (2006). The transposable element landscape of the model legume *Lotus japonicus*. *Genetics* **174**:2215–2228.
- Huang X. and Madan A. (1999). CAP3: A DNA sequence assembly program. *Genome Res.* **9**:868–877.
- Hudgins J.W., Ralph S.G., Franceschi V.R., and Bohlmann J. (2006). Ethylene in induced conifer defense: cDNA cloning, protein expression, and cellular and subcellular localization of 1-aminocyclopropane-1-carboxylate oxidase in resin duct and phenolic parenchyma cells. *Planta* **224**:865–877.
- IRGSP. (2005). The map-based sequence of the rice genome. *Nature* **436**:793–800.
- Isidore E., Scherrer B., Bellec A., Budin K., Faivre-Rampant P., Waugh R., Keller B., Caboche M., Feuillet C., and Chalhoub B. (2005). Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region. *Funct. & Integr. Geno.* **5**:97–103.
- Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., Casagrande A., Choisne N., Aubourg S., Vitulo N., Jubin C., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**:463–467.
- Jameson N., Georgelis N., Fouladbash E., Martens S., Hannah L.C., and Lal S. (2008). Helitron mediated amplification of cytochrome P450 monooxygenase gene in maize. *Plant Mol. Biol.* **67**:295–304.
- Juretic N., Hoen D.R., Huynh M.L., Harrison P.M., and Bureau T.E. (2005). The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res.* **15**:1292–1297.
- Jurka J., Kapitonov V.V., Pavlicek A., Klonowski P., Kohany O., and Walichiewicz J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**:462–467.
- Keeling C.I. and Bohlmann J. (2006). Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytol.* **170**:657–675.
- Keeling C.I., Weisshaar S., Lin R.P., and Bohlmann J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc. Natl. Acad. Sci. USA* **105**:1085–1090.
- Klotz K.L. and Lagrimini L.M. (1996). Phytohormone control of the tobacco anionic peroxidase promoter. *Plant Mol. Biol.* **31**:565–573.
- L'Homme Y., Seguin A., and Tremblay F.M. (2000). Different classes of retrotransposons in coniferous spruce species. *National Research Council Canada/Conseil national de recherches Canada* **43**:1084–1089.
- Lai Z., Vinod K., Zheng Z., Fan B., and Chen Z. (2008). Roles of *Arabidopsis* WRKY3 and WRKY4 transcription factors in plant responses to pathogens. *BMC Plant Biol.* **8**:68.
- Lescot M., Dehais P., Thijs G., Marchal K., Moreau Y., Peer Y., Rouze P., and Rombauts S. (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucl. Acids Res.* **30**:325–327.

- Lippert D., Chowrira S., Ralph S.G., Zhuang J., Aeschliman D., Ritland C., Ritland K., and Bohlmann J. (2007). Conifer defense against insects: Proteome analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils (*Pissodes strobi*). *Proteomics* **7**:248–270.
- Lippert D.N., Ralph S.G., Phillips M., White R., Smith D., Hardie D., Gershenson J., Ritland K., Borchers C.H., and Bohlmann J. (2009). Quantitative iTRAQ proteome and comparative transcriptome analysis of elicitor-induced Norway spruce (*Picea abies*) cells reveals elements of calcium signaling in the early conifer defense response. *Proteomics* **9**:350–367.
- Lippert D., Zhuang J., Ralph S., Ellis D.E., Gilbert M., Olafson R., Ritland K., Ellis B., Douglas C.J., and Bohlmann J. (2005). Proteome analysis of early somatic embryogenesis in *Picea glauca*. *Proteomics* **5**:461–473.
- Liu J.J. and Ekramoddoullah A.K. (2003). Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex. D. Don.). *Mol. Genet. Genomics* **270**:432–441.
- Liu J.J. and Ekramoddoullah A.K. (2004). Characterization, expression and evolution of two novel subfamilies of *Pinus monticola* cDNAs encoding pathogenesis-related (PR)-10 proteins. *Tree Physiol.* **24**:1377–1385.
- Liu J.J. and Ekramoddoullah A.K. (2009). Identification and characterization of the WRKY transcription factor family in *Pinus monticola*. *Genome* **52**:77–88.
- Martin D.M., Fäldt J., and Bohlmann J. (2004). Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiol.* **135**:1908–1927.
- Martin D.M., Gershenson J., and Bohlmann J. (2003). Induction of volatile terpene biosynthesis and diurnal emission by methyl jasmonate in foliage of Norway spruce. *Plant physiol.* **132**:1586–1599.
- Martin D., Tholl D., Gershenson J., and Bohlmann J. (2002). Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of Norway spruce stems. *Plant Physiol.* **129**:1003–1018.
- McKay S.A., Hunter W.L., Godard K.A., Wang S.X., Martin D.M., Bohlmann J., and Plant A.L. (2003). Insect attack and wounding induce traumatic resin duct development and gene expression of (-)-pinene synthase in Sitka spruce. *Plant Physiol.* **133**:368–378.
- Miller B., Madilao L.L., Ralph S., and Bohlmann J. (2005). Insect-induced conifer defense. White pine weevil and methyl jasmonate induce traumatic resinosis, de novo formed volatile emissions, and accumulation of terpenoid synthase and putative octadecanoid pathway transcripts in Sitka spruce. *Plant Physiol.* **137**:369–382.
- Morse A.M., Peterson D.G., Islam-Faridi M.N., Smith K.E., Magbanua Z., Garcia S.A., Kubisiak T.L., Amerson H.V., Carlson J.E., Nelson C.D., and Davis J.M. (2009). Evolution of genome size and complexity in Pinus. *PLoS ONE* **4**:e4332.
- Namroud M.C., Beaulieu J., Juge N., Laroche J., and Bousquet J. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**:3599–3613.
- Pavy N., Boyle B., Nelson C., Paule C., Giguere I., Caron S., Parsons L.S., Dallaire N., Bedon F., Berube H., et al. (2008). Identification of conserved core xylem gene sets: Conifer cDNA microarray development, transcript profiling and computational analyses. *New Phytol.* **180**:766–786.
- Pavy N., Paule C., Parsons L., Crow J.A., Morenczy M.J., Cooke J., Johnson J.E., Noumen E., Guillet-Claude C., Butterfield Y., et al. (2005). Generation, annotation, analysis and database integration of 16,500 white spruce EST clusters. *BMC Genomics* **6**:144.
- Pelgas B., Beauseigle S., Achere V., Jeandroz S., Bousquet J., and Isabel N. (2006). Comparative genome mapping among *Picea glauca*, *P. mariana* × *P. rubens* and *P. abies*, and correspondence with other Pinaceae. *Theor. Appl. Genet.* **113**:1371–1393.
- Phillips M.A. and Croteau R.B. (1999). Resin-based defenses in conifers. *Trends Plant Sci.* **4**:184–190.

- Phillips M.A., Walter M.H., Ralph S.G., Dabrowska P., Luck K., Uros E.M., Boland W., Strack D., Rodriguez-Concepcion M., Bohlmann J., and Gershenson J. (2007). Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (*Picea abies*). *Plant Mol. Biol.* **65**:243–257.
- Ralph S.G., Chun H.J., Kolosova N., Cooper D., Oddy C., Ritland C.E., Kirkpatrick R., Moore R., Barber S., Holt R.A., et al. (2008). A conifer genomics resource of 200,000 spruce (*Picea spp.*) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9**:484.
- Ralph S.G., Hudgins J.W., Jancsik S., Franceschi V.R., and Bohlmann J. (2007b). Aminocyclopropane carboxylic acid synthase is a regulated step in ethylene-dependent induced conifer defense. Full-length cDNA cloning of a multigene family, differential constitutive, and wound- and insect-induced expression, and cellular and subcellular localization in spruce and Douglas fir. *Plant Physiol.* **143**:410–424.
- Ralph S.G., Jancsik S., and Bohlmann J. (2007a). Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea spp.*). *Phytochemistry* **68**:1975–1991.
- Ralph S., Park J.Y., Bohlmann J., and Mansfield S.D. (2006). Dirigent proteins in conifer defense: Gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea spp.*). *Plant Mol. Biol.* **60**:21–40.
- Ralph S.G., Yueh H., Friedmann M., Aeschliman D., Zepnik J.A., Nelson C.C., Butterfield Y.S., Kirkpatrick R., Liu J., Jones S.J., et al. (2006). Conifer defence against insects: Microarray gene expression profiling of Sitka spruce (*Picea sitchensis*) induced by mechanical wounding or feeding by spruce budworms (*Choristoneura occidentalis*) or white pine weevils (*Pissodes strobi*) reveals large-scale changes of the host transcriptome. *Plant Cell. Environ.* **29**:1545–1570.
- Rizzon C., Ponger L., and Gaut B.S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput. Biol.* **2**:e115.
- Ro D.K., Arimura G., Lau S.Y., Piers E., and Bohlmann J. (2005). Loblolly pine abietadienol/abietadienal oxidase PtAO (CYP720B1) is a multifunctional, multisubstrate cytochrome P450 monooxygenase. *Proc. Natl. Acad. Sci. USA* **102**:8060–8065.
- Robatzek S. and Somssich I.E. (2001). A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. *Plant J.* **28**:123–133.
- Rouster J., Leah R., Mundy J., and Cameron-Mills V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. *Plant J.* **11**:513–523.
- Rungis D., Bérubé Y., Zhang J., Ralph S., Ritland C.E., Ellis B.E., Douglas C., Bohlmann J., and Ritland K. (2004). Robust simple sequence repeat markers for spruce (*Picea spp.*) from expressed sequence tags. *Theor. Appl. Genet.* **109**:1283–1294.
- Salamov A.A. and Solovyev V.V. (2000). *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**:516–522.
- Trapp S.C. and Croteau R.B. (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158**:811–832.
- Tuskan G.A., Difazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**:1596–1604.
- Zhang Y. and Wang L. (2005). The WRKY transcription factor superfamily: Its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* **5**:1.

3

Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.

Bogdanovic M. (1973). Chlorophyll formation in the dark. *Physiol. Plant* **29**:17–18.

- Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., and Benning C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* **17**:170–180.
- Brenner E.D., Martinez-Barboza N., Clark A.P., Liang Q.S., Stevenson D.W., and Coruzzi G.M. (2000). *Arabidopsis* mutants resistant to S(+)-beta-methyl-alpha, beta-diaminopropionic acid, a cycad-derived glutamate receptor agonist. *Plant Physiol.* **124**:1615–1624.
- Brenner E.D., Stevenson D.W., and Twigg R.W. (2003). Cycads: Evolutionary innovations and the role of plant-derived neurotoxins. *Trends Plant Sci.* **8**:446–452.
- Brownson D., Mabry T., and Leslie S. (2002). The cycad neurotoxic amino acid, beta-N-methylamino-L-alanine (BMAA), elevates intracellular calcium levels in dissociated rat brain cells. *J. Ethnopharmacol.* **82**:159–167.
- Chamberlain C. (1919). The Living Cycads. Chicago, University of Chicago Press.
- Chamovitz D.A. and Glickman M. (2002). The COP9 signalosome. *Curr. Biol.* **12**:R232.
- Chan R.L., Gago, G.M., Palena, C.M., and Gonzalez, D.H. (1998). Homeoboxes in plant development. *Biochim. Biophys. Acta.* **1442**:1–19.
- Charlton T.S., Marini A.M., Markey S.P., Nors托g K., and Duncan M.W. (1992). Quantification of the neurotoxin 2-amino-3-(methylamino)-propanoic acid (BMAA) in cycadalea. *Phytochemistry* **31**:3429–3432.
- Chiu J.C., Brenner E.D., DeSalle R., Nitabach M.N., Holmes T.C., and Coruzzi G.M. (2002). Phylogenetic and expression analysis of the glutamate-receptor-like gene family in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **19**:1066–1082.
- Chory J. and Wu D. (2001). Weaving the complex web of signal transduction. *Plant Physiol.* **125**:77–80.
- Crane P.R. (1985). Phylogenetic analysis of seed plants and the origin of angiosperms. *Annls. Missouri Bot. Gardens* **72**:716–793.
- Daly D.C., Cameron K.M., and Stevenson D.W. (2001). Plant systematics in the age of genomics. *Plant Physiol.* **127**:1328–1333.
- Duncan M.W., Kopin I.J., Crowley J.S., Jones S.M., and Markey S.P. (1989). Quantification of the putative neurotoxin 2-amino-3-(methylamino)-propanoic acid (BMAA) in Cycadales: Analysis of the seeds of some members of the family Cycadaceae. *J. Anal. Toxicol.* **13**:suppl A–G.
- Eshed Y., Baum S.F., and Bowman J.L. (1999). Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell* **99**:199–209.
- Eshed Y., Baum S.F., Perea J.V., and Bowman J.L. (2001). Establishment of polarity in lateral organs of plants. *Curr. Biol.* **11**:1251–1260.
- Ewing B. and Green P. (1998). Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome. Res.* **8**:186–194.
- Ewing B., Hillier L., Wendl M.C., and Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome. Res.* **8**:175–185.
- Foster A.S. and Gifford E.M. (1974). Comparative Morphology of Vascular Plants, 2nd Edition. San Francisco, WH Freeman.
- Gao Z. and Thomas B.A. (1989). A review of fossil cycad megasporophylls, with new evidence of *Crossozamia pomel* and its associated leaves from the lower Permian of Taiyuan, China. *Rev. Palaeobot. Palynol.* **60**:205–223.
- Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**:92–100.
- Hellmann H. and Estelle M. (2002). Plant development: Regulation by protein degradation. *Science* **297**:793–797.
- Index of full-length sequences [http://genomics.nybg.org/sequences/full_length].
- Khabazian I., Bains J.S., Williams D.E., Cheung J., Wilson J.M., Pasqualotto B.A., Pelech S.L., Andersen R.J., Wang Y.T., Liu L., et al. (2002). Isolation of various forms of sterol beta-D-glucoside from the seed of *Cycas circinalis*: Neurotoxicity and implications for ALS-parkinsonism dementia complex. *J. Neurochem.* **82**:516–528.
- Kirst M., Johnson A.F., Baucom C., Ulrich E., Hubbard K., Staggs R., Paule C., Retzel E., Whetten R., and Sederoff R. (2003). Apparent homology of expressed genes from wood-form-

- ing tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **100**:7383–7388.
- Kurland L.T. (1972). An appraisal of the neurotoxicity of cycad and the etiology of amyotrophic lateral sclerosis on Guam. *Fed. Proc.* **31**:1540–1543.
- Lacombe B., Becker D., Hedrich R., DeSalle R., Hollmann M., Kwak J.M., Schroeder J.I., Le Novere N., Nam H.G., Spalding E.P., et al. (2001). The identity of plant glutamate receptors. *Science* **292**:1486–1487.
- Lam H.M., Chiu J., Hsieh M.H., Meisel L., Oliveira I.C., Shin M., and Coruzzi G. (1998). Glutamate-receptor genes in plants. *Nature* **396**:125–126.
- Loconte H. and Stevenson D.W. (1990). Cladistics of the Spermatophyta. *Brittonia* **42**:197–211.
- Mamay S.H. (1969). Cycads: Fossil evidence of late paleozoic origin. *Science* **164**:295–296.
- Martienssen R. and McCombie W.R. (2001). The first plant genome. *Cell* **105**:571–574.
- Mayer K. and Mewes H.W. (2002). How can we deliver the large plant genomes? Strategies and perspectives. *Curr. Opin. Plant Biol.* **5**:173–177.
- Murray B. (1998). Nuclear DNA amounts in gymnosperms. *Ann. Bot.* **82** (Suppl A):3–15.
- Nixon K., Crepet W., Stevenson D.W., and Friis E. (1994). A reevaluation of seed plant phylogeny. *Annl. Missouri. Bot. Garden* **81**:484–583.
- Norstog K.J. and Nicholls T.J. (1997). The biology of the cycads. Ithaca, NY, Cornell University Press.
- Ohlrogge J. and Benning C. (2000). Unravelling plant metabolism by EST analysis. *Curr. Opin. Plant Biol.* **3**:224–228.
- Ohri D. and Khoshoo T. (1986). Genome size in gymnosperms. *Plant Syst. Evol.* **153**:119–132.
- Peer W., Silverthorne J., and Peters J.L. (1996). Developmental and light-regulated expression of individual members of the light-harvesting complex b gene family in *Pinus palustris*. *Plant Physiol.* **111**:627–634.
- Putterill J., Robson F., Lee K., Simon R., and Coupland G. (1995). The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**:847–857.
- Rudd S., Mewes H.W., and Mayer K.F. (2003). Sputnik: A database platform for comparative plant genomics. *Nucleic Acids Res.* **31**:128–132.
- Schwechheimer C. and Deng X.W. (2001). COP9 signalosome revisited: A novel mediator of protein degradation. *Trends Cell Biol.* **11**:420–426.
- Seawright A.A., Ng J.C., Oelrichs P.B., Sani Y., Nolan C.C., Lister A.T., Holton J., Ray D.E., and Osborne R. (1999). Recent toxicity studies in animals using chemicals derived from cycads. In biology and conservation of cycads—Proceedings of the Fourth International Conference on Cycad Biology 1996. Beijing: International Academic Publishers.
- Slater G.S.C. (2000). Algorithms for the Analysis of ESTs. PhD thesis. Cambridge, MA: University of Cambridge.
- Soltis D.E., Soltis P.S., and Zanis M.J. (2002). Phylogeny of seed plants based on evidence from eight genes. *Am. J. Bot.* **89**:1670–1681.
- Spencer P.S., Hunn P.B., Nugon J., Ludolph A.C., Ross S.M., Roy D.H., and Robertson R.C. (1987). Guam amyotrophic lateral sclerosis-Parkinsonism-dementia linked to a plant excitant neurotoxin. *Science* **237**:517–522.
- Sputnik *Cycas rumphii* [<http://mips.gsf.de/proj/sputnik/cycad>]
- Stevenson D.W. (1981). Observations on ptyxis, phenology, and trichomes in the Cycadales and their systematic implications. *Am. J. Bot.* **68**:1104–1114.
- Stevenson D.W. (1990). Morphology and systematics of the Cycadales. *Mem. NY Bot. Garden* **57**:8–55.
- Suarez-Lopez P., Wheatley K., Robson F., Onouchi H., Valverde F., and Coupland G. (2001). *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**:1116–1120.
- The *Arabidopsis* Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796–815.

- Treutlein J. and Wink M. (2002). Molecular phylogeny of cycads inferred from rbcL sequences. *Naturwissenschaften* **89**:221–225.
- VecScreen [http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html]
- Vega A. and Bell E.A. (1967). Alpha-amino-beta-methylaminopropionic acid, a new amino acid from seeds of *Cycas circinalis*. *Phytochemistry* **6**:759–762.
- Villanueva J.M., Broadhurst J., Hauser, B.A., Meister R.J., Schneitz K., and Gasser C.S. (1999). INNER NO OUTER regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes Dev.* **13**:3160–3169.
- Warrilow A.G. and Hawkesford M.J. (2000). Cysteine synthase (O-acetylserine (thiol) lyase) substrate specificities classify the mitochondrial isoform as a cyanoalanine synthase. *J. Exp. Bot.* **51**:985–993.
- Warrilow A.G. and Hawkesford M.J. (2002). Modulation of cyanoalanine synthase and O-acetylserine (thiol) lyases A and B activity by beta-substituted alanyl and anion inhibitors. *J. Exp. Bot.* **53**:439–445.
- Wei N. and Deng X.W. (1992). COP9: A new genetic locus involved in light-regulated development and gene expression in *Arabidopsis*. *Plant Cell* **4**:1507–1518.
- Whetten R., Sun Y.H., Zhang Y., and Sederoff R. (2001). Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Mol. Biol.* **47**:275–291.
- Whiting M.G. (1963). Toxicity of cycads. *Econ. Bot.* **17**:271–302.
- Yu J., Hu S., Wang J., Wong G.K., Li S., Liu B., Deng Y., Dai L., Zhou Y., Zhang X., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**:79–92.
- Cai Z., Guisinger M., Kim H.G., Ruck E., Blazier J.C., McMurtry V., Kuehl J.V., Boore J., and Jansen R.K. (2008). Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.*
- Cai Z., Penaflor C., Kuehl J.V., Leebens-Mack J., Carlson J.E., dePamphilis C.W., Boore J.L., and Jansen R.K. (2006). Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* **6**:77.
- Chumley T.W., Palmer J.D., Mower J.P., Fourcade H.M., Calie P.J., Boore J.L., and Jansen R.K. (2006). The complete chloroplast genome sequence of *Pelargonium × hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **23**:2175–2190.
- Conant D.S., Stein D.B., Valinski A.E.C., Sudarsanam P., and Ahearn M.E. (1994). Phylogenetic implications of chloroplast DNA variation in the Cyatheaceae. *I. Syst. Bot.* **19**:60–72.
- Conant G.C. and Wolfe K.H. (2008). GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24**:861–862.
- Ewing B. and Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
- Gautheret D. and Lambert A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J. Mol. Biol.* **313**:1003–1011.
- Gawel N. and Jarret R. (1991). A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Mol. Biol. Rep.* **9**:262–266.
- Haberle R.C., Fourcade H.M., Boore J.L., and Jansen R.K. (2008). Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **66**:350–361.
- Hasebe M. and Iwatsuki K. (1992). Gene localization on the chloroplast DNA of the maiden hair fern; *Adiantum capillus-veneris*. *J. Plant Res.* **105**:413–419.
- Janssen T., Bystríková N., Rakotondrainibe F., Coomes D., Labat J.N., and Schneider H. (2008). Neoendemism in Madagascan scaly tree ferns results from recent, coincident diversification bursts. *Evolution* **62**:1876–1889.
- Kuroda H., Suzuki H., Kusumegi T., Hirose T., Yukawa Y., and Sugiyama M. (2007). Translation of psbC mRNAs starts from the downstream GUG, not the upstream AUG, and requires the

- extended Shine-Dalgarno sequence in tobacco chloroplasts. *Plant Cell Physiol.* **48**:1374–1378.
- Kurtz S., Choudhuri J.V., Ohlebusch E., Schleiermacher C., Stoye J., and Giegerich R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **29**:4633–4642.
- Laslett D. and Canback B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**:11–16.
- Lung B., Zemann A., Madej M.J., Schuelke M., Techritz S., Ruf S., Bock R., and Huttenhofer A. (2006). Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res.* **34**:3842–3852.
- Ogihara Y., Terachi T., and Sasakuma T. (1988). Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci. USA* **85**:8573–8577.
- Ovcharenko I., Loots G.G., Hardison R.C., Miller W., and Stubbs L. (2004). zPicture: Dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res.* **14**:472–477.
- Palmer J.D. (1985). Comparative organization of chloroplast genomes. *Annu. Rev. Genet.* **19**:325–354.
- Palmer J.D., Nugent J.M., and Herbon L.A. (1987). Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. USA* **84**:769–773.
- Pfitzinger H., Weil J.H., Pillay D.T., and Guillemaut P. (1990). Codon recognition mechanisms in plant chloroplasts. *Plant Mol. Biol.* **14**:805–814.
- Pombert J.F., Lemieux C., and Turmel M. (2006). The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol.* **4**:3.
- Pombert J.F., Otis C., Lemieux C., and Turmel M. (2005). The chloroplast genome sequence of the green alga *Pseudodoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Mol. Biol. Evol.* **22**:1903–1918.
- Pryer K.M., Schneider H., Smith A.R., Cranfill R., Wolf P.G., Hunt J.S., and Sipes S.D. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**:618–622.
- Pryer K.M., Schneider H., Zimmer E.A., and Ann Banks J. (2002). Eciding among green plants for whole genome studies. *Trends Plant Sci.* **7**:550–554.
- Pryer K.M., Schuettpelz E., Wolf P.G., Schneider H., Smith A.R., and Cranfill R. (2004). Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *Am. J. Bot.* **91**:1582–1598.
- Raubeson L.A. and Jansen R.K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255**:1697–1699.
- Raubeson L.A. and Jansen R.K. (2005). Chloroplast genomes of plants. *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*. In R.J. Henry (Ed.). London: CABI Publishing, pp. 45–68.
- Raubeson L.A. and Stein D.B. (1995). Insights into fern evolution from mapping chloroplast genomes. *Am. Fern J.* **85**:193–204.
- Roper J.M., Kellon Hansen S., Wolf P.G., Karol K.G., Mandoli D.F., Everett K.D.E., Kuehl J., and Boore J.L. (2007). The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). *Am. Fern J.* **97**:95–106.
- Schneider H., Schuettpelz E., Pryer K.M., Cranfill R., Magallon S., and Lupia R. (2004). Ferns diversified in the shadow of angiosperms. *Nature* **428**:553–557.
- Shimada H. and Sugiura M. (1991). Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genomes. *Nucleic Acids Res.* **19**:983–995.
- Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome:

- its gene organization and expression. *EMBO J.* **5**:2043–2049.
- Smith A.R., Pryer K.M., Schuettpelz E., Korall P., Schneider H., and Wolf P.G. (2006). A classification for extant ferns. *Taxon* **55**:705–731.
- Stein D.B., Conant D.S., Ahearn M.E., Jordan E.T., Kirch S.A., Hasebe M., Iwatsuki K., Tan M.K., and Thomson J.A. (1992). Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. *Proc. Natl. Acad. Sci. USA* **89**:1856–1860.
- Sugiura M., Hirose T., and Sugita M. (1998). Evolution and mechanism of translation in chloroplasts. *Annu. Rev. Genet.* **32**:437–459.
- Sugiura C. and Sugita M. (2004). Plastid transformation reveals that moss tRNA^{Arg}-CCG is not essential for plastid function. *Plant J.* **40**:314–321.
- Taquist H., Cui Y., and Ardell D.H. (2007). TFAM 1.0: An online tRNA function classifier. *Nucleic Acids Res.* **35**:W350–353.
- Tatusova T.A. and Madden T.L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**:247–250.
- Wakasugi T., Nishikawa A., Yamada K., and Sugita M. (1998). Complete nucleotide sequence of the plastid genome from a fern, *Psilotum nudum*. *Endocyt Cell Res.* **13**(Suppl):147.
- Wakasugi T., Tsudzuki T., and Sugita M. (2001). The genomics of land plant chloroplasts: Gene content and alteration of genomic information by RNA editing. *Photosynth Res.* **70**:107–118.
- Wolf P.G. and Roper J.M. (2008). Structure and evolution of fern plastid genomes. *Biology and Evolution of Ferns and Lycophytes*. In T.A. Ranker and C.H. Haufler (Eds.). Cambridge, UK, Cambridge University Press, pp. 59–174.
- Wolf P.G., Rowe C.A., and Hasebe M. (2004). High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* **339**:89–97.
- Wolf P.G., Rowe C.A., Sinclair R.B., and Hasebe M. (2003). Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. *DNA Res.* **10**:59–65.
- Wyman S.K., Jansen R.K., and Boore J.L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**:3252–3255.
- ## 5
- Ashrafi K., Chang F.Y., Watts J.L., Fraser A.G., Kamath R.S., Ahringer J., and Ruvkun G. (2003). Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**:268–272.
- Banks J.A. (1997). The TRANSFORMER genes of the fern *Ceratopteris* simultaneously promote meristem and archegonia development and repress antheridia development in the developing gametophyte. *Genetics* **147**:1885–1897.
- Banks J.A. (1999). Gametophyte development in ferns. *Annu. Rev. Plant Phys.* **50**:163–186.
- Banks J.A., Hickok L., and Webb M.A. (1993). The Programming of Sexual Phenotype in the Homosporous Fern *ceratopteris-richardii*. *Int. J. Plant Sci.* **154**:522–534.
- Chatterjee A. and Roux S.J. (2000). *Ceratopteris richardii*: A productive model for revealing secrets of signaling and development. *J. Plant Growth Regul.* **19**:284–289.
- Chuang C.F. and Meyerowitz E.M. (2000). Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *P. Natl. Acad. Sci. USA* **97**:4985–4990.
- Cogoni C. and Macino G. (1997). Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **94**:10233–10238.
- Cogoni C. and Macino G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**:166–169.
- Crete P., Leuenberger S., Iglesias V.A., Suarez V., Schob H., Holtorf H., van Eeden S., and Meins F. (2001). Graft transmission of induced and spontaneous post-transcriptional silencing of chitinase genes. *Plant J.* **28**:493–501.

- Eberle J.R. and Banks J.A. (1996). Genetic interactions among sex-determining genes in the fern *Ceratopteris richardii*. *Genetics* **142**:973–985.
- Elder G.H. and Roberts A.G. (1995). Uroporphyrinogen decarboxylase. *J. Bioenerg. Biomembr.* **27**:207–214.
- Fraser A.G., Kamath R.S., Zipperlen P., Martinez-Campos M., Sohrmann M., and Ahringer J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**:325–330.
- Gonczy P., Echeverri C., Oegema K., Coulson A., Jones S.J., Copley R.R., Duperon J., Oegema J., Brehm M., Cassin E., Hannak E., Kirkham M., Pichler S., Flohrs K., Goessen A., Leidel S., Alleaume A.M., Martin C., Ozlu N., Bork P., and Hyman A.A. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**:331–336.
- Hannon G.J. (2002). RNA interference. *Nature* **418**:244–251.
- Hansson A., Kannangara C.G., von Wettstein D., and Hansson M. (1999). Molecular basis for semidominance of missense mutations in the XANTHA-H (42-kDa) subunit of magnesium chelatase. *Proc. Natl. Acad. Sci. USA* **96**:1744–1749.
- Hiei Y., Ohta S., Komari T., and Kumashiro T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**:271–282.
- Hu G.S., Yalpani N., Briggs S.P., and Johal G.S. (1998). A porphyrin pathway impairment is responsible for the phenotype of a dominant disease lesion mimic mutant of maize. *Plant Cell* **10**:1095–1105.
- Jefferson R.A., Kavanagh T.A., and Bevan M.W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *Embo. J.* **6**:3901–3907.
- Ketting R.F. and Plasterk R.H. (2000). A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* **404**:296–298.
- Klahre U., Crete P., Leuenberger S.A., Iglesias V.A., and Meins F. Jr. (2002). High molecular weight RNAs and small interfering RNAs induce systemic posttranscriptional gene silencing in plants. *Proc. Natl. Acad. Sci. USA* **99**:11981–11986.
- Klink V.P. and Wolniak S.M. (2000). The efficacy of RNAi in the study of the plant cytoskeleton. *J. Plant Growth Regul.* **19**:371–384.
- Klink V.P. and Wolniak S.M. (2001). Centrin is necessary for the formation of the motile apparatus in spermatids of *Marsilea*. *Mol. Biol. Cell* **12**:761–776.
- Maeda I., Kohara Y., Yamamoto M., and Sugimoto A. (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**:171–176.
- Misquitta L. and Paterson B.M. (1999). Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): A role for nautilus in embryonic somatic muscle formation. *Proc. Natl. Acad. Sci. USA* **96**:1451–1456.
- Mourrain P., Beclen C., Elmayan T., Feuerbach F., Godon C., Morel J.B., Jouette D., Lacombe A.M., Nikic S., Picault N., Remoue K., Sanial M., Vo T.A., and Vaucheret H. (2000). *Arabidopsis SGS2* and *SGS3* genes are required for post-transcriptional gene silencing and natural virus resistance. *Cell* **101**:533–542.
- Osteryoung K.W. and McAndrew R.S. (2001). The Plastid Division Machine. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**:315–333.
- Osteryoung K.W., Stokes K.D., Rutherford S.M., Percival A.L., and Lee W.Y. (1998). Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial *ftsZ*. *Plant Cell* **10**:1991–2004.
- Palauqui J.C. and Balzergue S. (1999). Activation of systemic acquired silencing by localised introduction of DNA. *Curr. Biol.* **9**:59–66.
- Pickford A.S., Catalano C., Cogoni C., and Macino G. (2002). Quelling in *Neurospora crassa*. *Adv. Genet.* **46**:277–303.
- Reski R. (2002). Rings and networks: the amazing complexity of *FtsZ* in chloroplasts. *Trends Plant Sci.* **7**:103–105.
- Stout S.C., Clark G.B., Archer-Evans S., and Roux S.J. (2003). Rapid and efficient suppression of gene expression in a single-cell model system, *Ceratopteris richardii*. *Plant Physiol.* **131**:1165–1168.

6

- Stoutjesdijk P.A., Singh S.P., Liu Q., Hurlstone C.J., Waterhouse P.A., and Green A.G. (2002). hpRNA-mediated targeting of the *Arabidopsis FAD2* gene gives highly efficient and stable silencing. *Plant Physiol.* **129**:1723–1731.
- Strain E., Hass B., and Banks J.A. (2001). Characterization of mutations that feminize gametophytes of the fern *Ceratopteris*. *Genetics* **159**:1271–1281.
- Strepp R., Scholz S., Kruse S., Speth V., and Reski R. (1998). Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein *FtsZ*, an ancestral tubulin. *Proc. Natl. Acad. Sci. USA* **95**:4368–4373.
- Suzuki M., Ario T., Hattori T., Nakamura K., and Asahi T. (1994). Isolation and characterization of two tightly linked catalase genes from castor bean that are differentially regulated. *Plant Mol. Biol.* **25**:507–516.
- Timmermans M.C., Maliga P., Vieira J., and Messing J. (1990). The pFF plasmids: Cassettes utilising CaMV sequences for expression of foreign genes in plants. *J. Biotechnol.* **14**:333–344.
- Tsai C.W. and Wolniak S.M. (2001). Cell cycle arrest allows centrin translation but not basal body formation during spermiogenesis in *Marssilea*. *J. Cell Sci.* **114**:4265–4272.
- Vaucheret H., Beclin C., and Fagard M. (2001). Post-transcriptional gene silencing in plants. *J. Cell Sci.* **114**:3083–3091.
- Voinnet O., Vain P., Angell S., and Baulcombe D.C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell* **95**:177–187.
- von Wettstein D., Gough S., and Kannangara C.G. (1995). Chlorophyll Biosynthesis. *Plant Cell* **7**:1039–1057.
- Wesley S.V., Helliwell C.A., Smith N.A., Wang M.B., Rouse D.T., Liu Q., Gooding P.S., Singh S.P., Abbott D., Stoutjesdijk P.A., Robinson S.P., Gleave A.P., Green A.G., and Waterhouse P.M. (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**:581–590.
- Allen D. and Bowersox T.W. (1989). Regeneration in oak stands following gypsy moth defoliations. In G. Rink and C.A. Budelsky (Eds.). *Proceedings, 7th Central Hardwood Forest Conference*, 1989 March 5–8; Carbondale, IL. Gen. Tech. Rep. NC-132. St. Paul, MN, U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, pp. 67–73.
- Engelman H.M. and Nyland R.D. (2006). Interference to hardwood regeneration in northeastern North America: Assessing and countering ferns in northern hardwood forests. *Nor. J. Appl. For.* **23**:166–175.
- George L.O. and Bazzaz F.A. (1999). The fern understory as an ecological filter: Emergence and establishment of canopy-tree seedlings. *Ecology* **80**:833–845.
- Groninger J.W. and McCormick L.H. (1991). Invasion of a partially cut oak stand by hay-scented fern. In L.H. McCormick. and K.W. Gottschalk (Eds.). *Proceedings, 8th Central Hardwood Forest Conference*, 1991 March 4–6; University Park, PA. Gen. Tech. Rep. NE-148. U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, pp. 585–586.
- Horsley S.B., McCormick L.H., and Groninger J.W. (1992). Effects of timing of oush application on survival of hardwood seedlings. *Nor. J. Appl. For.* **9**:22–27.
- Hughes J.F. and Fahey T.J. (1991). Colonization dynamics of herbs and shrubs in disturbed northern hardwood forest. *J. Ecol.* **79**:605–616.
- Johnson P.S., Jacobs R.D., Martin A.J., and Goebel E.D. (1989). Regenerating northern red oak: Three successful case histories. *Nor. J. Appl. For.* **6**:174–178.
- Marquis D.A. (1979). Shelterwood cutting in Allegheny hardwoods. *J. For.* **77**:140–144.
- Marquis D.A. (1994). *Quantitative Silviculture for Hardwood Forests of the Alleghenies*. Gen. Tech. Rep. NE-183. Radnor, PA, U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, pp. 129–246.
- McWilliams W.H., Stout S.L., Bowersox T.W., and McCormick L.H. (1995). Adequacy of advance treeseedling regeneration in Pennsylvania's forests. *Nor. J. Appl. For.* **12**:187–191.

Messier C.T., Honer T.W., and Kimmins J.P. (1989). Photosynthetic photon flux density, red: far-red ratio, and minimum requirement for survival of *Gaultheria shallon* in western red cedar-western hemlock stands in coastal British Columbia. *Can. J. For. Res.* **19**:1470–1477.

Steiner K.C. and Joyce B.J. (1999). Survival and growth of a *Quercus rubra* regeneration cohort during five years following masting. In J.W. Stringer and D.L. Loftis (Eds.). *Proceedings, 12th central hardwood forest conference*, 1999 February 28–March 2; Lexington, KY. Gen. Tech. Rep. SRS-24. Asheville, NC, U.S. Department of Agriculture, Forest Service, Southern Research Station, pp. 255–257.

7

Anderson J.M. and Anderson H.M. (1985). Palaeoflora of Southern Africa: *Prodromus of South African Megafloras Devonian to Lower Cretaceous*, A. A. Balkema, Rotterdam.

Barrett P.J. and Elliot D.J. (1973). Reconnaissance geologic map of the Buckley Island Quadrangle, Transantarctic Mountains, Antarctica, *Antarctic Geol. Map (U.S.G.S.)*, A-3.

Collinson J.W., Hammer W.R., Askin R.A., and Elliot D.H. (2006). Permian-Triassic boundary in the central Transantarctic Mountains, Antarctica. *Geol. Soc. Am. Bull.* **118**:747–763.

Cúneo N.R., Isbell J., Taylor E.L., and Taylor T.N. (1993). The *Glossopteris* flora from Antarctica: Taphonomy and paleoecology, 12 Congrès Intl. Géol. Carb.-Perm., Comptes Rendus, 2, In S. Archangelsky (Ed.), pp. 13–40.

Doyle J.A. (2006). Seed ferns and the origin of angiosperms. *J. Torrey Bot. Soc.* **133**:169–209.

Farabee M.J., Taylor E.L., and Taylor T.N. (1991). Late Permian palynomorphs from the Buckley Formation, central Transantarctic Mountains, Antarctica. *Rev. Palaeobot. Palynol.* **69**:353–368.

Galtier J.M. and Phillips T.L. (1999). The acetate peel technique, *Fossil Plants and Spores: Modern Techniques*, In T.P. Jones and N.P. Rowe (Eds.), London, UK, Geological Society, pp. 67–70.

Gould R. and Delevoryas T. (1977). The biology of *Glossopteris*: Evidence from petrified seed

bearing and pollen-bearing organs. *Alcheringa* **1**:387–399.

Hermsen E., Taylor, T.N., Taylor E.L., and Stevenson D.W. (2007). Cycads from the Triassic of Antarctica: Permineralized cycad leaves. *Int. J. Plant Sci.* **168**(7) 1099–111.

Hilton J. and Bateman R.M. (2006). Pteridosperms are the backbone of seed-plant phylogeny. *J. Torrey Bot. Soc.* **133**:119–168.

Holmes W.B.K. (1973). On some fructifications of the *Glossopteridales* from the Upper Permian of N.S.W. *Proc. Linn. Soc. New South Wales* **98**:131–141.

Klavins S.D., Taylor E.L., Krings M., and Taylor T.N. (2001). An unusual, structurally preserved ovule from the Permian of Antarctica. *Rev. Palaeobot. Palynol.* **115**:107–117.

Le Goc M.J. (1914). Observations on the centripetal and centrifugal xylem in the petioles of cycads. *Ann. Bot.* **28**:183–193.

McLoughlin S. (1990a). Late Permian glossopterid fructifications from the Bowen and Sydney Basins, eastern Australia. *Géobios* **23**:283–297.

McLoughlin S. (1990b). Some Permian glossopterid fructifications and leaves from the Bowen Basin, Queensland Australia. *Rev. Palaeobot. Palynol.* **62**:11–40.

McLoughlin S. (1993). Glossopterid megafossils in Permian Gondwanic non-marine biostratigraphy. *Gondwana Eight: Assembly, Evolution and Dispersal*, In R.H. Findlay, R. Unrug, M.R. Banks, and J.J. Veevers (Eds.), Rotterdam, Netherlands, A.A. Balkema, pp. 253–264.

McManus H.A., Taylor E.L., Taylor T.N., and Collinson J.W. (2002). A petrified *Glossopteris* flora from Collinson Ridge, central Transantarctic Mountains: Late Permian or Early Triassic? *Rev. Palaeobot. Palynol.* **120**:233–246.

Neish P.G., Drinnan A.N., and Cantrill D.J. (1993). Structure and ontogeny of Vertebraria from silicified Permian sediments in East Antarctica, *Rev. Palaeobot. Palynol.* **79**:221–244.

Nishida H., Pigg K.B., Kudo K., and Rigby J.F. (2007). New evidence of reproductive organs of *Glossopteris* based on permineralized fossils from Queensland, Australia. I. Ovulate organ

- Homevaleia gen. nov. *J. Plant Res.* **120**:539–549.
- Nixon K.C., Crepet W.L., Stevenson D., and Friis E.M. (1994). A reevaluation of seed plant phylogeny. *Ann. Mo. Gard.* **81**:484–533.
- Pant D.D. (1977). The plant of *Glossopteris*. *J. Ind. Bot. Soc.* **56**:1–23.
- Pant D.D. and Pant R. (1987). Some *Glossopteris* leaves from Indian Triassic beds. *Palaeontographica* **205B**:165–178.
- Pant D.D. and Singh R.S. (1974). On the stem and attachment of *Glossopteris* and *Gangamopteris* leaves. Part II—structural features. *Palaeontographica* **147B**:42–73.
- Pigg K.B. (1990). Anatomically preserved *Glossopteris* foliage from the central Transantarctic Mountains. *Rev. Palaeobot. Palynol.* **66**:105–127.
- Pigg K.B. and McLoughlin S.L. (1997). Anatomically preserved *Glossopteris* leaves from the Bowen and Sydney basins, Australia. *Rev. Palaeobot. Palynol.* **97**:339–359.
- Pigg K.B. and Nishida H. (2006). The significance of silicified plant remains to the understanding of *Glossopteris*-bearing plants: An historical review. *J. Torrey Bot. Soc.* **133**:46–61.
- Pigg K.B. and Taylor T.N. (1993). Anatomically preserved *Glossopteris* stems with attached leaves from the central Transantarctic Mountains. *Amer. J. Bot.* **80**:500–516.
- Retallack G.J. and Dilcher D.L. (1981). Arguments for a glossopterid ancestry of angiosperms. *Paleobiology* **7**:54–67.
- Rigby J.F. (1962). On a collection of plants of Permian age from Bara-bala, Queensland. *Proc. Linn. Soc. New South Wales* **87**:341–351.
- Rigby J.F. (1978). Permian glossopterid and other cycadopsis fructifications from Queensland. *Geol. Surv. Queensland Pub.* **367**:3–21.
- Schopf J.M. (1976). Morphologic interpretation of fertile structures in glossopterid gymnosperms. *Rev. Palaeobot. Palynol.* **21**:25–64.
- Shah B.A. (2000). Revision of age of Parsora Formation, Johilla-Son Valley, Rewa Basin, M.P. *Indian Minerals* **54**:223–232.
- Surange K.R. and Chandra S. (1973). Denkania indica gen. et sp nov. -A glossopteridean fructification from the Lower Gondwana of India. *Palaeobotanist* **20**:264–268.
- Surange K.R. and Chandra S. (1975). Morphology of the gymnospermous fructifications of the *Glossopteris* flora and their relationships. *Palaeo-ontographica* **149B**:153–180.
- Taylor E.L. and Taylor T.N. (1992). Reproductive biology of the Permian *Glossopteridales* and their suggested relationship to flowering plants. *Proc. Nat. Acad. Sci. USA* **89**:11495–11497.
- Taylor E.L., Taylor T.N., and Collinson J.W. (1989). Depositional setting and paleobotany of Permian and Triassic permineralized peatfrom the central Transantarctic Mountains, Antarctica. *Int. J. Coal Geol.* **12**:657–679.
- Thomas H.H. (1958). Lidgettonia, a new type of fertile *Glossopteris*. *Bull. Brit. Mus. (Nat. Hist.) Geol.* **3**:179–189.
- 8**
- Aguilar J.F., Rosselló J.A., and Feliner G.N. (1999). Molecular evidence for the compilospecies model of reticulate evolution in *Armeria* (*Plumbaginaceae*). *Syst. Biol.* **48**:735–754.
- Aguilar J.F., Rosselló J.A., and Feliner G.N. (1999). Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (*Plumbaginaceae*). *Mol. Ecol.* **8**:1341–1346.
- Alvarez I. and Wendel J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**:417–434.
- Bailey C.D., Carr T.G., Harris S.A., and Hughes C.E. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol. Phylogenet. Evol.* **29**:435–455.
- Bänfer G., Moog U., Fiala B., Mohamed M., Weising K., and Blattner F.R. (2006). A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance dispersals. *Mol. Ecol.* **15**:4409–4424.
- Barton N.H. (1979). Dynamics of hybrid zones. *Heredity* **43**:341–359.

- Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.K., and Das I. (2007). Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* **22**:148–155.
- Buckler-IV E.S., Ippolito A., and Holtsford T.P. (1997). The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics* **145**:821–832.
- Chase M.W., Cowan R.S., Hollingsworth P.M., Berg C., Madriñán S., Petersen G., Seberg O., Jørgensen T., Cameron K.M., Carine M., Pedersen N., Hedderson T.A.J., Conrad F., Salazar G.A., Richardson J.E., Hollingsworth M.L., Barraclough T.G., Kelly L., and Wilkinson M. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon* **56**:295–299.
- Chase M.W., Salamin N., Wilkinson M., Dunwell J.M., Kesanakurthi R.P., Haidar N., and Savolainen V. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Phil. Trans. R Soc. B* **360**:1889–1895.
- Coyne J.A. and Orr H.A. (2004). *Speciation*. Sunderland, Massachusetts, Sinauer Associates.
- Diamond J.M. (1992). Horrible plant species. *Nature* **360**:627–628.
- Fehrer J., Gemeinholzer B., Chrtk J., and Bräutigam S. (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (Hieracium, Cichorieae, Asteraceae). *Mol. Phylogenet. Evol.* **42**:347–361.
- Gerton J.L., DeRisi J., Shroff R., Lichten M., Brown P.O., and Petes T.D. (2000). Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97**:11383–11390.
- Grundt H.H., Kjølner S., Borgen L., Rieseberg L.H., and Brochmann C. (2006). High biological species diversity in the arctic flora. *Proc. Natl. Acad. Sci. USA* **103**:972–975.
- Huelsenbeck J.P. and Ronquist F. (2001). MR-BAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Kardolus J.P., van Eck H.J., and Berg R.G. (1998). The potential of AFLPs in biosystematics: A first application in *Solanum* taxonomy (Solanaceae). *Plant Syst. Evol.* **210**:87–103.
- Knowles L.L. and Carstens B.C. (2007). Delimiting species without monophyletic gene trees. *Syst. Biol.* **56**:887–895.
- Kress W.J. and Erickson D.L. (2007). A Two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* **6**:e508.
- Kress W.J., Wurdack K.J., Zimmer E.A., Weigt L.A., and Janzen D.H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**:8369–8374.
- Kron K.A., Gawen L.M., and Chase M.W. (1993). Evidence for introgression in azaleas (Rhododendron; Ericaceae): Chloroplast DNA and morphological variation in a hybrid swarm on Stone Mountain, Georgia. *Am. J. Bot.* **80**:1095–1099.
- Lahaye R., Bank M., Bogarin D., Warner J., Populin F., Gigot G., Maurin O., Duthoit S., Barraclough T.G., and Savolainen V. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA* **105**:2923–2928.
- Levin D.A. (1979). The nature of plant species. *Science* **204**:381–384.
- Lowry D.B., Modliszewski J.L., Wright K.M., Wu C.A., and Willis J.H. (2008). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Phil. Trans. R Soc. B* **363**:3009–3021.
- Mayr E. (1992). A local flora and the biological species concept. *Am. J. Bot.* **79**:222–238.
- Monaghan M.T., Balke M., Pons J., and Vogler A.P. (2006). Beyond barcodes: Complex DNA taxonomy of a South Pacific Island radiation. *Proc. R. Soc. B* **273**:887–893.
- Moyle L.C., Olson M.S., and Tiffin P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution* **58**:1195–1208.
- Nicolè F., Tellier F., Vivat A., and Till-Bottraud I. (2007). Conservation unit status inferred for plants by combining interspecific crosses and AFLP. *Conserv. Genet.* **8**:1273–1285.
- Nylander J.A.A. (2004). MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Okuyama Y., Fujii N., Wakabayashi M., Kawakita A., Ito M., Watanabe M., Murakami N., and Kato M. (2005). Nonuniform concerted evolution

- and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Mol. Biol. Evol.* **22**:285–296.
- Okuyama Y., Kato M., and Murakami N. (2004). Pollination by fungus gnats in four species of the genus *Mitella* (Saxifragaceae). *Bot. J. Linn. Soc.* **144**:449–460.
- Okuyama Y., Pellmyr O., and Kato M. (2008). Parallel floral adaptations to pollination by fungus gnats within the genus *Mitella* (Saxifragaceae). *Mol. Phylogenet. Evol.* **46**:560–575.
- Pellmyr O.P., Segraves K.A., Althoff D.M., Balcázar-Lara M., and Leebens-Mack J. (2007). The phylogeny of yuccas. *Mol. Phylogenet. Evol.* **43**:493–501.
- Petit R.J., Pineau E., Demesure B., Bacilieri R., Ducousoo A., and Kremer A. (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc. Natl. Acad. Sci. USA* **94**:9996–10001.
- Posada D. and Crandall K.A. (1998). MODELT-EST: Testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- R Development Core Team (2008). R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing.
- Rieseberg L.H. and Brouillet L. (1994). Are many plant species paraphyletic? *Taxon* **43**:21–32.
- Rieseberg L.H. and Soltis D.E. (1991). Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trend. Plant* **5**:65–84.
- Rieseberg L.H., Wood T.E., and Baack E.J. (2006). The nature of plant species. *Nature* **440**:524–527.
- Savolainen V. and Chase M.W. (2003). A decade of progress in plant molecular phylogenetics. *Trends Genet.* **19**:717–724.
- Shishido R., Sano Y., and Fukui K. (2000). Ribosomal DNAs: an exception to the conservation of gene order in rice genomes. *Mol. Gen. Genet.* **263**:586–591.
- Simmons M.P. and Ochoterena H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**:369–381.
- Soltis D.E., Soltis P.S., Pires J.C., Kovarik A., Tate J.A., and Mavrodiev E. (2004). Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biol. J. Linn. Soc.* **82**:485–501.
- Swofford D.L. (2002). PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts, Sinauer Associates.
- Syring J., Farrell K., Businsky R., Cronn R., and Liston A. (2007). Widespread genealogical non-monophyly in species of *Pinus* subgenus *Strobus*. *Syst. Biol.* **56**:163–181.
- Taberlet P., Coissac E., Pompanon F., Gielly L., Miquel C., Valentini A., Vermat T., Corthier G., Brochmann C., and Willerslev E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucl. Acids Res.* **35**:e14.
- The Angiosperm Phylogeny Group (1998). An ordinal classification for the families of flowering plants. *Ann. MO Bot. Gard.* **85**:531–553.
- The Angiosperm Phylogeny Group (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**:399–436.
- Vogler A.P. and Monaghan M.T. (2007). Recent advances in DNA taxonomy. *J. Zool. Syst. Evol. Res.* **45**:1–10.
- Wakabayashi M. (1973). A note on the genus *Mitella* of Japan. *Acta Phytotax Geobot.* **25**:136–153.
- Wakabayashi M. (1977). A note on *Mitella stylosa* and allied species (Saxifragaceae). *Acta Phytotax Geobot.* **28**:111–122.
- Wakabayashi M. (1992). Embryology of Japanese *Mitella* (Saxifragaceae) and its taxonomic significance. *Bot. Mag. Tokyo* **105**:485–502.
- Wakabayashi M. (2001). Saxifragaceae 13. *Mitella*. *Flora of Japan*, Vol. IIb. In K. Iwatsuki, D.E. Boufford, and H. Ohba (Eds.). Tokyo, Kodansha, pp. 70–75.
- Wendel J.F., Schnabel A., and Seelanan T. (1995). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **92**:280–284.

- Whittall J.B., Hellquist C.B., Schneider E.L., and Hedges S.A. (2004). Cryptic species in an endangered pondweed community (*Potamogeton*, *Potamogetonaceae*) revealed by AFLP markers. *Am. J. Bot.* **91**:2022–2029.
- Wu J.Z., Mizuno H., Hayashi-Tsugane M., Ito Y., Chiden Y., Fujisawa M., Katagiri S., Saji S., Yoshiaki S., Karasawa W., Yoshihara R., Hayashi A., Kobayashi H., Ito K., Hamada M., Okamoto M., Ikeda M., Ichikawa Y., Katayose Y., Yano M., Matsumoto T., and Sasaki T. (2003). Physical maps and recombination frequency of six rice chromosomes. *Plant J.* **36**:720–730.
- Yatabe Y., Kane N.C., Scotti-Saintagne C., and Rieseberg L.H. (2007). Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**:1883–1893.
- Zhang D. and Sang T. (1999). Physical mapping of ribosomal RNA genes in peonies (*Paeonia*, *Paeoniaceae*) by fluorescent in situ hybridization: implications for phylogeny and concerted evolution. *Am. J. Bot.* **86**:735–740.
- 9**
- APG II (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**:399–436.
- Baker H.G. and Baker I. (1979). Starch in angiosperm pollen grains and its evolutionary significance. *Am. J. Bot.* **66**:591–600.
- Barrett S. (2008). Major evolutionary transitions in flowering plant reproduction: an overview. *Int. J. Plant Sci.* **169**:1–5.
- Bedinger P. (1992). The remarkable biology of pollen. *Plant Cell* **4**:879–887.
- Blackmore S. and Crane P.R. (1988). The systematic implications of pollen and spore ontogeny. *Ontogeny and Systematics*. In C.J. Humphries (Ed.), London, UK, British Museum (Natural History), pp 83–115.
- Blackmore S., Wortley A.H., Skvarla J.J., and Rowley J.R. (2007). Pollen wall development in flowering plants. *New Phytologist* **174**:483–498.
- Brewbaker J.L. (1967). Distribution and phylogenetic significance of binucleate and trinucleate pollen grains in angiosperms. *Am. J. Bot.* **54**:1069–1083.
- Dahl A.O. and Rowley J.R. (1965). Pollen of *Degeneria vitiensis*. *J. Arnold Arbor., Harvard University* **46**:308–323.
- Davis C.C., Endress P.K., and Baum D.A. (2008). The evolution of floral gigantism. *Curr. Opin. Plant Biol.* **11**:49–57.
- Doyle J.A. and Le Thomas A. (1994). Cladistic analysis and pollen evolution in Annonaceae. *Acta Bot. Gallica* **141**:149–170.
- Doyle J.A. and Le Thomas A. (1997a). Phylogeny and geographic history of Annonaceae. *Geog. Phys. et Quat.* **51**:353–361.
- Doyle J.A. and Le Thomas A. (1997b). Significance of palynology for phylogeny of Annonaceae: experiments with removal of pollen characters. *Plant Systemat. Evol.* **206**:133–159.
- Eilat D. and Fischel R. (1991). Recurrent utilization of genetic elements in V regions of anti-nucleic acid antibodies from autoimmune mice. *J. Immunol.* **147**:361–368.
- Enns L.C., Kanaoka M.M., Torii K.U., Comai L., Okada K., and Cleland R.E. (2005). Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Mol. Biol.* **58**:333–349.
- Erdtman G. (1960). The acetolysis method. *Svensk Botanisk Tidskrift* **54**:561–564.
- Farr C.H. (1918). Cell division by furrowing in *Magnolia*. *Am. J. Bot.* **5**:379–395.
- Farr C.H. (1922). Quadripartition by furrowing in *Sisyrinchium*. *Bull. Torrey Bot. Club* **49**:51–61.
- Feder N. and O'Brien T.P. (1968). Plant microtechnique: some principles and new methods. *Am. J. Bot.* **55**:123–142.
- Franchi G.G., Bellani L., Nepi M., and Pacini E. (1996). Types of carbohydrate reserves in pollen: Localization, systematic distribution and ecophysiological significance. *Flora* **191**:143–159.
- Francis K.E., Lam S.Y., and Copenhagen G.P. (2006). Separation of *Arabidopsis* pollen tetrads is regulated by *QUARTET1*, a pectin methyl-esterase gene. *Plant Physiol.* **142**:1004–1013.

- Furness C.A., Rudall P.J., and Sampson F.B. (2002). Evolution of microsporogenesis in angiosperms. *Int. J. Plant Sci.* **163**:235–260.
- Gabarayeva N.I. (1995). Pollen wall and tafetum development in *Anaxagorea brevipes* (Annonaceae)—sporoderm substructure, cytoskeleton, sporopollenin precursor particles, and the endexine problem. *Rev. Palaeobot. Palynol.* **85**:123–152.
- Gottberger G. (1999). Pollination and evolution in Neotropical Annonaceae. *Plant Spec. Biol.* **14**:143–152.
- Grayum M.H. (1985). Evolutionary and ecological significance of starch storage in pollen of the Araceae. *Am. J. Bot.* **72**:1565–1577.
- Guzzo F., Baldan B., Bracco F., and Mariani P. (1994). Pollen development in *Liriodendron tulipifera*—some unusual features. *Can. J. Bot.* **72**:352–358.
- Harder L. and Johnson S. (2008). Function and evolution of aggregated pollen in angiosperms. *Int. Jour. Plant Sci.* **169**:59–78.
- Heslop-Harrison Y. (1977). Pollen-stigma interaction—pollen-tube penetration in *Crocus*. *Ann. Bot.* **41**:913–922.
- Heslop-Harrison Y. and Heslop-Harrison J. (1992). Germination of monocolpate angiosperm pollen: evolution of the actin cytoskeleton and wall during hydration, activation and tube emergence. *Ann. Bot.* **69**:385–394.
- Higuchi H. and Utsunomiya N. (1999). Floral differentiation and development in cherimoya (*Annona cherimola* Mill.) under warm (30/25 degrees C) and cool (20/15 degrees C) day/night temperatures. *J. Japan. Soc. Hort. Sci.* **68**:707–716.
- Hughes J. and McCully M.E. (1975). Use of an optical brightener in study of plant structure. *Stain Tech.* **50**:319–329.
- Knox J.P. (1997). The use of antibodies to study the architecture and developmental regulation of plant cell walls. *Int. Rev. Cytol.* **171**:79–120.
- Le Thomas A. (1980). Ultrastructure characters of the pollen grains of African Annonaceae and their significance for the phylogeny of primitive angiosperms (first part). *Pollen et Spores* **22**:267–342.
- Le Thomas A. (1981). Ultrastructure characters of the pollen grains of African Annonaceae and their significance for the phylogeny of primitive angiosperms (second part). *Pollen et Spores* **23**:5–36.
- Le Thomas A., Morawetz W., and Waha M. (1986). Pollen of paleo- and neotropical Annonaceae: Definition of the aperture by morphological and functional characters. *Pollen and Spores: Form and Function*. In S. Blackmore and I.K. Ferguson (Eds.). Academic Press, London, UK, pp:375–388.
- Lora J., Herrero M., and Hormaza J.I. (2009). The coexistence of bicellular and tricellular pollen in *Annona cherimola* Mill. (Annonaceae): Implications for pollen evolution. *Am. J. Bot.* **96**:802–808.
- Lora J., Oteyza M.A.P., Fuentetaja P., and Hormaza J.I. (2006). Low temperature storage and *in vitro* germination of cherimoya (*Annona cherimola* Mill.) pollen. *Sci. Hort.* **108**:91–94.
- Maheshwari P. (1950). An introduction to the embryology of angiosperms, New York, USA McGraw-Hill.
- McCormick S. (2004). Control of male gametophyte development. *Plant Cell* **16**:S142–S153.
- Mena C.G., Testillano P.S., Gonzalez-Melendi P., Gorab E., and Risueño M.C. (1994). Immunoelectron microscopy detection of RNA in plant nucleoli. *Exp Cell Res.* **212**:393–408.
- Nadot S., Furness C.A., Sannier J., Penet L., Triki-Teurtroy S., Albert B., and Ressayre A. (2008). Phylogenetic comparative analysis of microsporogenesis in angiosperms with a focus on monocots. *Am. J. Bot.* **95**:1426–1436.
- Neelam A. and Sexton R. (1995). Cellulase (endo beta-1,4 glucanase) and cell-wall breakdown during anther development in the sweet pea (*Lathyrus odoratus* L.): isolation and characterization of partial cDNA clones. *J. Plant Physiol.* **146**:622–628.
- Pacini E. and Franchi G.G. (1988). Amylogenesis and amyloylation during pollen grain development. *Sexual Reproduction in Higher Plants*. In M. Cresti, P. Gori, and E. Pacini (Eds.) Springer, Berlin, Heidelberg, New York, pp 181–186.

- Pacini E. and Franchi G.G. (1999). Pollen grain sporoderm and types of dispersal units. *Acta Soc. Bot. Pol.* **68**:299–305.
- Pacini E., Franchi G.G., and Hesse M. (1985). The tapetum—its form, function, and possible phylogeny in embryophyta. *Plant Syst. Evol.* **149**:155–185.
- Periasamy K. and Kandasamy M.K. (1981). Development of the anther of *Annona squamosa* L. *Ann. Bot.* **48**:885–893.
- Pirie M.D., Vargas M.P., Botermans M., Bakker F.T., and Chatrou L.W. (2007). Ancient paralogy in the cpDNA trnL-F region in Annonaceae: implications for plant molecular systematics. *Am. J. Bot.* **94**:1003–1016.
- Preuss D., Rhee S.Y., and Davis R.W. (1994). Tetrad analysis possible in *Arabidopsis* with mutation of the QUARTET (*QRT*) genes. *Science* **264**:1458–1460.
- Ratnayake R.M.C.S., Gunatilleke I.A.U.N., Wijesundara D.S.A., and Saunders R.M.K. (2006). Reproductive biology of two sympatric species of *Polyalthia* (Annonaceae) in Sri Lanka. I. Pollination by curculionid beetles. *Int. J. Plant Sci.* **167**:483–493.
- Rhee S.Y. and Somerville C.R. (1998). Tetrad pollen formation in quartet mutants of *Arabidopsis thaliana* is associated with persistence of pectic polysaccharides of the pollen mother cell wall. *Plant Journal* **15**:79–88.
- Rhee S.Y., Osborne E., Poindexter P.D., and Somerville C.R. (2003). Microspore separation in the quartet 3 mutants of *Arabidopsis* is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. *Plant Physiology* **133**:1170–1180.
- Rosell P., Herrero M., and Sauco V.G. (1999). Pollen germination of cherimoya (*Annona cherimola* Mill.). In vivo characterization and optimization of in vitro germination. *Sci. Hort.* **81**:251–265.
- Rosell P., Sauco V.G., and Herrero M. (2006). Pollen germination as affected by pollen age in cherimoya. *Sci. Hort.* **109**:97–100.
- Rowley J.R. (1964). Formation of the pore in pollen of *Poa annua*. *Pollen physiology and fertilization*. In H.F. Linskens (Ed.). North-Holland Publishing Company, Amsterdam, Netherlands, pp:59–69.
- Rudall P.J. and Bateman R.M. (2007). Developmental bases for key innovations in the seed-plant microgametophyte. *Trends Plant Sci.* **12**:317–326.
- Sabatini D.D., Bensch K., and Barnett R.J. (1963). Cytochemistry and electron microscopy—preservation of cellular ultrastructure and enzymatic activity by aldehyde fixation. *J. Cell Biol.* **17**:19–58.
- Sampson F.B. (1969). Cytokinesis in pollen mother cells of angiosperms, with emphasis on *Laurelia-novae-zelandiae* (Monimiaceae). *Cytologia* **34**:627–634.
- Satpute G.K., Long H., Seguí-Simarro J.M., Risueño M.C., and Testillano P.S. (2005). Cell architecture during gametophytic and embryogenic microspore development in *Brassica napus* L. *Acta Physiol. Plant.* **27**:665–674.
- Scott R.J., Spielman M., and Dickinson H.G. (2006). Stamen development: Primordium to pollen. *The Molecular Biology and Biotechnology of Flowering*. In B.R. Jordan (Ed.), Wallingford, UK, CAB International, pp 298–331.
- Solís M.T., Pintos B., Prado M.J., Bueno M.A., Raska I., Risueño M.C., and Testillano P.S. (2008). Early markers of in vitro microspore re-programming to embryogenesis in olive (*Olea europaea* L.). *Plant Sci.* **174**:597–605.
- Soltis D.E., Soltis P.S., Endress P.K., and Chase M.W. (2005). *Phylogeny and Evolution of Angiosperms*. Sunderland, Massachusetts, USA, Sinauer Associates Incorporated.
- Su Y.C. and Saunders R.M. (2003). Pollen structure, tetrad cohesion and pollen-connecting threads in *Pseuduvaria* (Annonaceae). *Bot. J. Linn. Soc.* **143**:69–78.
- Testillano P.S., Corredor E., Solís M.T., Chakrabarti N., Raska I., and Risueño M.C. (2007). Changes in nuclear architecture and DNA methylation pattern accompany the developmental programmed cell death of the tapetum. Proceedings of the 6th. Plant Genomics European Meeting. Tenerife, Spain, pp 9–10.
- Testillano P.S., Fadon B., and Risueño M.C. (1995). Ultrastructural localization of the polysaccharidic component during the sporoderm

- ontogeny of the pollen grain. *Rev. Paleobot. Palynol.* **85**:53–62.
- Tsou C.H. and David J. (2003). Comparative development of aseptate and septate anthers of Annonaceae. *Am. J. Bot.* **90**:832–848.
- Tsou C.H. and Fu Y.L. (2002). Tetrad pollen formation in *Annona* (Annonaceae): proxeine formation and binding mechanism. *Am. J. Bot.* **89**:734–747.
- Tsou C.H. and Fu Y.L. (2007). Octad pollen formation in *Cymbopetalum* (Annonaceae): the binding mechanism. *Plant Syst. Evol.* **263**:13–23.
- Walker J.W. (1971a). Unique type of angiosperm pollen from the family Annonaceae. *Science* **172**:565–567.
- Walker J.W. (1971b). Pollen morphology, phytogeography and phylogeny of the Annonaceae. *Contributions of the Gray Herbarium of Harvard University* **202**:1–132.
- Walker J.W. and Doyle J.A. (1975). Bases of angiosperm phylogeny—palynology. *Ann. Mo. Bot. Gard.* **62**:664–723.
- Williams J.H., Friedman W.E., and Arnold M.L. (1999). Developmental selection within the angiosperm style: Using gamete DNA to visualize interspecific pollen competition. *Proc. Natl. Acad. Sci. USA* **96**:9201–9206.
- Wu H. and Cheung A.Y. (2000). Programmed cell death in plant reproduction. *Plant Mol. Biol.* **44**:267–281.
- Xiang-Yuan X. and Demason D.A. (1984). Relationship between male and female gametophyte development in rye. *Am. J. Bot.* **71**:1067–1079.
- Zavada M. (1984). Pollen wall development of *Austrobaileya maculata*. *Bot. Gaz.* **145**:11–21.
- ## 10
- Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Blaxter M. (2004). The promise of a DNA taxonomy. *Phil. Trans. Royal Soc. B: Biol. Sci.* **359**:669–679.
- Chase M.W., Cowan R.S., Hollingsworth P.M., van den Berg C., Madriñan S., et al. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon* **56**:295–299.
- Chase M., Salamin N., Wilkinson M., Dunwell J., Kesanakurthi R., et al. (2005). Land plants and DNA barcodes: Short-term and long-term goals. *Phil. Trans. Royal Soc. B: Biol. Sci.* **360**:1889–1895.
- Chaw S.M., Walters T.W., Chang C.C., Hu S.H., and Chen S.H. (2005). A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. *Mol. Phylogenetic Evol.* **37**:214–234.
- Cho Y., Mower J.P., Qiu Y.L., and Palmer J.D. (2004). Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci.* **101**:17741–17746.
- CITES (May 2007). Convention on International Trade in Endangered Species of Wild Fauna and Flora.
- DasGupta B., Konwar K.M., Mandoiu I.I., and Shvartsman A.A. (2005a). DNA-BAR: Distiguisher selection for DNA barcoding. *Bioinformatics* **21**:3424–3426.
- DasGupta B., Konwar K., Mandoiu I.I., and Shvartsman A.A. (2005b). Highly scalable algorithms for robust string barcoding. *Inter. J. Bioinfo. Res. and Appl.* **1**:145–161.
- Donaldson J.S. (2003). *Cycads: Status Survey and Conservation Action Plan*. Gland, Switzerland, IUCN—the World Conservation Union.
- Doyle J.J. and Doyle J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**:11–15.
- Driscoll H.E. and Barrington D.S. (2007). Origin of Hawaiian *Polystichum* (Dryopteridaceae) in the context of a world phylogeny. *Am. J. Bot.* **94**:1413–1424.
- Ebach M.C. and Holdrege C. (2005). More Taxonomy, Not DNA Barcoding. *BioScience* **55**:823824.
- Gompert Z., Nice C.C., Fordyce J.A., Forister M.L., and Shapiro A.M. (2006). Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molec. Ecol.* **15**:1759–1768.

- Gonzalez D. and Vovides A.P. (2002). Low intra-lineage divergence in *Ceratozamia* (Zamiaceae) detected with nuclear ribosomal DNA ITS and chloroplast DNA *trnL-F* non-coding region. *Syst. Botany* **27**:654–661.
- Hebert P.D.N., Alina C., Shelley L.B. and Jeremy R.D. (2003). Biological identifications through DNA barcodes. *Proc. Royal Soc. B: Biol. Sci.* **270**:313–321.
- Hebert P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrapetes fugerator*. *Proc. Natl. Acad. Sci. USA* **101**:14812–14817.
- Hebert P.D.N., Stoeckle M.Y., Zemlak T.S., and Francis C.M. (2004). Identification of Birds through DNA Barcodes. *PLoS Biol.* **2**:e312.
- Hermsen E.J., Taylor T.N., Taylor E.L., and Stevenson D.W. (2006). Cataphylls of the Middle Triassic cycad *Antarcticycas schopfii* and new insights into cycad evolution. *Am. J. Botany* **93**:724–738.
- Hill K.D., Chase M.W., Stevenson D.W., Hills H.G., and Schutzman B. (2003). The families and genera of cycads: a molecular phylogenetic analysis of cycadophyta based on nuclear and plastid DNA sequences. *Inter. J. Plant Sci.* **164**:933–948.
- Hill K. and Stevenson D.W. (2004). *World List of Cycads*. Sydney, Royal Botanic Gardens.
- Jakob S.S. and Blattner F.R. (2006). A chloroplast genealogy of *Hordeum* (Poaceae): Long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol. Biol. Evol.* **23**:1602–1612.
- Jones D.L. (1993). *Cycads of the world*. In J. Young (Ed.). Smithsonian Institution Press.
- Kress W.J. and Erickson D.L. (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. *Pub. Lib. Sci. Biol.* **2**:e508.
- Kress W.J., Wurdack K.J., Zimmer E.A., Weigt L.A., and Janzen D.H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **102**:8369–8374.
- Lidholm J. and Gustafsson P. (1991). A three-step model for the rearrangement of the chloroplast *trnK-psbA* region of the gymnosperm *Pinus contorta*. *Nucl. Acids Res.* **19**:2881–2887.
- Lidholm J., Szmidt A., and Gustafsson P. (1991). Duplication of the *psbA* gene in the chloroplast genome of two *Pinus* species. *Molec. and Gen. Genet.* **226**:345–352.
- Little D.P. and Stevenson D.W. (2007). A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* **23**:1–21.
- Maggini F., Marrocco R., Gelati M.T., and de Dominicis R.I. (1998). Lengths and nucleotide sequences of the internal spacers of nuclear ribosomal DNA in gymnosperms and pteridophytes. *Plant Syst. and Evol.* **213**:199–205.
- Meier R., Shiyang K., Vaidya G., and Ng P.K.L. (2006). DNA barcoding and taxonomy in *Diptera*: A tail of high intraspecific variability and low identification success. *Syst. Biol.* **55**:715–728.
- Meyer C.P. and Paulay G. (2005). DNA Barcoding: Error rates based on comprehensive sampling. *Pub. Lib. Sci. Biol.* **3**:e422.
- Moritz C. and Cicero C. (2004). DNA Barcoding: Promise and pitfalls. *Pub. Lib. Sci. Bio.* **2**:279–354.
- Palmer J.D. and Herbon L.A. (1988). Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28**:87–97.
- Presting G.G. (2006). Identification of conserved regions in the plastid genome: Implications for DNA barcoding and biological function. *Can. J. Bot.* **84**:1434–1443.
- Ratnasingham S. and Hebert P.D.N. (2007). BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). Molecular Ecology Notes, **7**:355–364.
- Rubinoff D. (2006). Utility of Mitochondrial DNA Barcodes in Species Conservation. *Conser. Bio.* **20**:1026–1033.
- Stoeckle M. (2003). Taxonomy, DNA, and the Bar Code of Life. *BioScience* **53**:796–797.
- Thompson J.D., Higgins D.G., and Gibson T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment

- through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**:4673–4680.
- Timmis J.N., Ayliffe M.A., Huang C.Y., and Martin W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**:123–135.
- Ward R.D., Zemlak T.S., Innes B.H., Last P.R., and Hebert P.D.N. (2005). DNA barcoding Australia's fish species. *Philos. Trans.: Biol. Sci.* **360**:1847–1857.
- Will K., Mishler B., and Wheeler Q. (2005). The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Syst. Bio.* **54**:844–851.
- Will K.W. and Rubinoff D. (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* **20**:47–55.
- Whitelock L.M. (2002). *The Cycads*. Portland, OR, Timber Press, Inc.
- Wolf P.G., Rowe C.A., Sinclair R.B., and Hasebe M. (2003). Complete Nucleotide Sequence of the Chloroplast Genome from a Leptosporangiate Fern, *Adiantum capillus-veneris* L. *DNA Res.* **10**:59.
- ## 11
- Akaike H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*. In Petrov B.N. and Csaki F. (Eds.). Budapest, Akademiai Kiado, pp. 267–281.
- Aris-Brosou S. and Yang Z. (2003). Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Mol. Biol. Evol.* **20**:1947–1954.
- Bousquet J., Strauss S.H., Doerksen A.H., and Price R.A. (1992). Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. *Proc. Natl. Acad. Sci. USA* **89**:7844–7848.
- Bremer K. (2002). Gondwanan evolution of the grass alliance of families (Poales). *Evolution* **56**:1374–1387.
- Brown J.W., Rest J.S., Garcia-Moreno J., Sorenson M.D., and Mindell D.P. (2008). Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol.* **6**:6.
- Chaw, S.M., Chang C.C., Chen H.L., and Li W.H. (2004). Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* **58**:424–441.
- Drummond A.J., Ho S.Y.W., Phillips M.J., and Rambaut A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS. Biol.* **4**:e88.
- Erixon P. and Oxelman B. (2008). Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS. ONE* **3**:e1386.
- Gaut B.S., Muse S.V., Clark W.D., and Clegg M.T. (1992). Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *J. Mol. Evol.* **35**:292–303.
- Goldman N. and Yang Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- Guisinger M.M., Kuehl J.V., Boore J.L., and Jansen R.K. (2008). Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* **105**:18424–18429.
- Hasegawa M., Thorne J., and Kishino H (2003). Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* **78**:267–283.
- Herendeen P.S. and Crane P.R. (1995). The fossil history of the monocotyledons. *Monocotyledon: Systematics and Evolution*. In P.J. Rudall, P.J. Cribb, D.F. Cutler, and C.J. Humphries (Eds.). London, Royal Botanic Gardens, Kew, pp. 1–21.
- Huelsenbeck J.P., Larget B., and Swofford D. (2000). A compound Poisson process for relaxing the molecular clock. *Genetics* **154**:1879–1892.
- Jacobs B.F., Kingston J.D., and Jacobs L.L. (1999). The origin of grass-dominated ecosystems. *Ann. Mo. Bot. Gard.* **86**:590–643.
- Jansen R.K., Cai Z., Raubeson L.A., Daniell H., and dePamphilis C.W., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies

- genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* **104**:19369–19374.
- Kishino H., Thorne J.L., and Bruno W.J. (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* **18**:352–361.
- Kitazoe Y., Kishino H., Waddell P.J., Nakajima N., Okabayashi T., et al. (2007). Robust time estimation reconciles views of the antiquity of placental mammals. *PLoS. ONE* **2**:e384.
- Leebens-Mack J., Raubeson L.A., Cui L., Kuehl J.V., and Fourcade, M.H., et al. (2005). Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Mol. Biol. Evol.* **22**:1948–1963.
- Lepage T., Bryant D., Philippe H., and Lartillot N. (2007). A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**:2669–2680.
- Linder H.P. and Rudall P.J. (2005). Evolutionary history of Poales. *Annu. Rev. Ecol. Evol. Syst.* **36**:107–124.
- Martin W., Deusch O., Stawski N., Grunheit N., and Goremykin V. (2005). Chloroplast genome phylogenetics: Why we need independent approaches to plant molecular evolution. *TRENDS in Plant Sci.* **10**:203–209.
- Matsuoka Y., Yamazaki Y., Ogihara Y., and Tsunewaki K. (2002). Whole chloroplast genome comparison of rice, maize, and wheat: Implications for chloroplast gene diversification and phylogeny of cereals. *Mol. Biol. Evol.* **19**:2084–2091.
- Moore M.J., Bell C.D., Soltis P.S., and Soltis D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* **94**:19363–19368.
- Muse S.V. and Gaut B.S. (1994). A likelihood approach for comparing synonymous and non-synonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**:715–724.
- Muse S.V. and Gaut B.S. (1997). Interlocus comparisons of the nucleotide substitution process in the chloroplast genome. *Genetics* **146**:393–399.
- Nikaido M., Matsuno F., Hamilton H., Brownell R.L. Jr., Cao Y., et al. (2001). Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* **98**:7384–7389.
- Piperno D.R. and Sues H.D. (2005). Dinosaurs dined on grass. *Science* **310**:1126–1128.
- Prasad V., Stroemberg C.A.E., Alimohammadian H., and Sahni A. (2005). Dinosaur coprolites and the early evolution of grasses and grazers. *Science* **310**:1177–1180.
- Rannala B. and Yang Z. (2007). Inferring speciation times under an episodic molecular clock. *Syst. Biol.* **56**:453–466.
- Renner S.S., Grimm G.W., Schneeweiss G.M., Stuessy T.F., and Ricklefs R.E. (2008). Rooting and dating maples (*Acer*) with an uncorrelated-rates molecular clock: Implications for North American/Asian disjunctions. *Syst. Biol.* **57**:795–808.
- Sanderson M.J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**:1218–1232.
- Sanderson M.J., Thorne J.L., Wikstrom N., and Bremer K. (2004). Molecular evidence on plant divergence times. *Amer. J. Bot.* **91**:1656–1665.
- Smith S.A. and Donoghue M.J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**:86–89.
- Thorne J.L., Kishino H., and Painter I.S. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**:1647–1657.
- Vicentini A., Barber J.C., Aliscioni S.S., Giussani L.M., and Kellogg E.A. (2008). The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biol.* **14**:2963–2977.
- Wolfe K.H., Gouy M.Y., Yang W., Sharp P.M., and Li W-H (1989). Date of the monocot-dicot divergence estimated from chloroplast chloroplast DNA sequence data. *Proc. Natl. Acad. Sci. USA* **86**:6201–6205.
- Yang Z. (2006). *Computational Molecular Evolution*. Oxford, Oxford University Press.

- Yang Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**:1586–1591.
- Yang Z. and Bielawski B. (2000). Statistical methods for detecting molecular adaptation. *TREE* **15**:496–503.
- Yang Z. and Rannala B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**:212–226.
- Yang Z., Wong W.S.W., and Nielsen R. (2005). Bayes empirical Bayes inference of amino acids sites under positive selection. *Mol. Biol. Evol.* **22**:1107–1118.
- Yoder A.D. and Yang Z. (2004). Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* **13**:757–773.
- Zhang J., Nielsen R., and Yang Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**:1–8.
- Brenner E.D., Stevenson D.W., and Twigg R.W. (2003). Cycads: evolutionary innovations and the role of plant-derived neurotoxins. *Trends Plant Sci.* **8**:446–452.
- Brenner E.D., Stevenson D.W., McCombie R.W., Katari M.S., Rudd S.A., Mayer K.F., Palenchar P.M., Runko S.J., Twigg R.W., Dai G., Martienssen R.A., Benfey P.N., and Coruzzi, G.M. (2003). Expressed sequence tag analysis in *Cycas*, the most primitive living seed plant. *Genome Biol.* **4**:R78.
- Chamberlain C.J. (1935). *Gymnosperms. Structure and Evolution*. Chicago, University of Chicago Press.
- Chamovitz D.A. and Glickman M. (2002). The COP9 signalosome. *Curr. Biol.* **12**:R232.
- Chaw S.M., Parkinson C.L., Cheng Y., Vincent T.M., and Palmer J.D. (2000). Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl. Acad. Sci. USA* **97**:4086–4091.
- Chinn E. and Silverthorne J. (1993). Light-dependent chloroplast development and expression of a light-harvesting chlorophyll a/b-binding protein gene in the gymnosperm *Ginkgo biloba*. *Plant Physiol.* **103**:727–732.
- Chinn E., Silverthorne J., and Hohtola A. (1995). Light-regulated and organ-specific expression of types 1, 2, and 3 light-harvesting complex b mRNAs in *Ginkgo biloba*. *Plant Physiol.* **107**:593–602.
- Curtis-Prior P., Vere D., and Fray P. (1999). Therapeutic value of *Ginkgo biloba* in reducing symptoms of decline in mental function. *J. Pharm. Pharmacol.* **51**:535–541.
- Donoghue M.J. and Doyle J.A. (2000). Seed plant phylogeny: Demise of the anthophyte hypothesis? *Curr. Biol.* **10**:R106–9.
- Doyle J.A. and Donoghue M.J. (1986). Seed plant phylogeny and the origin of angiosperms: An experimental cladistic approach. *Botan. Rev.* **52**:321–431.
- Ewing B. and Green P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.

12

- Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Bierhorst D.W. (1971). *Morphology of Vascular Plants*. Macmillan.
- Bogdanovic M. (1973). Chlorophyll formation in the dark. *Physiol. Plant* **29**:17–18.
- Bowe L.M., Coat G., and dePamphilis C.W. (2000). Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA* **97**:4092–4097.
- Bowman J.L., Smyth D.R., and Meyerowitz E.M. (1991). Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**:1–20.
- Brenner E.D. and Stevenson D.W. (2005). Genomic approaches to understand the origins of seeds. *Landscape, Genomics and Transgenic Conifers*. In C.G. Williams (Ed.). Amsterdam, Kluwer-Springer.

- Ewing B., Hillier L., Wendl M.C., and Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
- Fagard M., Boutet S., Morel J.B., Bellini C., and Vaucheret H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. *Proc. Natl. Acad. Sci. USA* **97**:11650–11654.
- Foster A.S. and Gifford E.M. (1974). *Comparative Morphology of Vascular Plants*, 2nd ed. D. Kennedy and R.B. Park (Eds.). San Francisco, W. H. Freeman and Company.
- Friedman W.E. (1987). Morphogenesis and experimental aspects of growth and development of the male gametophyte of *Ginkgo biloba* in vitro. *Amer. J. Bot.* **74**:1816–1830.
- Gold P.E., Cahill L., and Wenk G.L. (2003). The lowdown on *Ginkgo biloba*. *Sci. Am.* **288**:86–91.
- Goodrich J., Puangsomlee P., Martin M., Long D., Meyerowitz E.M., and Coupland G. (1997). A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386**:44–51.
- Hasebe M. (1997). Molecular phylogeny of *Ginkgo biloba*: Close relationship between *Ginkgo biloba* and cycads. *Ginkgo biloba-a Global Treasure*. In T. Hori, R.W. Ridge, W. Tulecke, P. Del Tredici, J. Tremouillaux-Guiller, and H. Tobe (Eds.), Tokyo, Springer-Verlag, pp. 173–181.
- Hayama R. and Coupland G. (2000). The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. *Plant Physiol.* **135**:677–684.
- Hellmann H. and Estelle M. (2002). Plant development: regulation by protein degradation. *Science* **297**:793–797.
- Hori T., Ridge R.W., Tulecke W., Del Tredici P., Tremouillaux-Guiller J., and Tobe H. (1997). *Ginkgo biloba-A Global Treasure*. Tokyo, Springer.
- Huang X. and Madan A. (1999). CAP3: A DNA sequence assembly program. *Genome Res.* **9**:868–877.
- Ikeno S. and Hirase S. (1897). Spermatozooids in gymnosperms. *Ann. Bot.* **11**:344–345.
- Iseli C., Jongeneel C.V., and Bucher P. (1999). ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 138–148.
- Jager M., Hassinan A., Manuel M., Guyader H.L., Deutsch J. (2003). *MADS-box* genes in *Ginkgo biloba* and the evolution of the AGAMOUS Family. *Mol. Biol. Evol.* **20**:842–854.
- Kardailsky I., Shukla V.K., Ahn J.H., Dagenais N., Christensen S.K., Nguyen J.T., Chory J., Harrison M.J., and Weigel D. (1999). Activation tagging of the floral inducer FT. *Science* **286**:1962–1965.
- Katz A., Oliva M., Mosquna A., Hakim O., and Ohad N. (2004). FIE and CURLY LEAF polycomb proteins interact in the regulation of *HOME* box gene expression during sporophyte development. *Plant J.* **37**:707–719.
- Kirst M., Johnson A.F., Baucom C., Ulrich E., Hubbard K., Staggs R., Paule C., Retzel E., Whetten R., and Sederoff R. (2003). Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **100**:7383–7388.
- Kohler C. and Grossniklaus U. (2002). Epigenetic inheritance of expression states in plant development: The role of Polycomb group proteins. *Curr. Opin. Cell. Biol.* **14**:773–779.
- Lagercrantz U. and Axelsson T. (2000). Rapid evolution of the family of *CONSTANS LIKE* genes in plants. *Mol. Biol. Evol.* **17**:1499–1507.
- Lin W.C., Shuai B., and Springer P.S. (2003). The *Arabidopsis LATERAL ORGAN BOUNDARIES*-domain gene *ASYMMETRIC LEAVES2* functions in the repression of *KNOX* gene expression and in adaxial-abaxial patterning. *Plant Cell* **15**:2241–2252.
- Magollón S. and Sanderson M.J. (2006). Relationships among seed plants inferred from highly conserved genes: Sorting conflicting phylogenetic signals among ancient lineages. *Am. J. Bot.* **89**:1991–2006.
- Noh Y.S., Bizzell C.M., Noh B., Schomburg F.M., and Amasino R.M. (2004). EARLY FLOWERING 5 acts as a floral repressor in *Arabidopsis*. *Plant J.* **38**:664–672.

- Peer W., Silverthorne J., and Peters J.L. (1996). Developmental and light-regulated expression of individual members of the light-harvesting complex *b* gene family in *Pinus palustris*. *Plant Physiol.* **111**:627–634.
- Plant GDB. Plant Genomic Database [<http://www.plantgdb.org/>].
- Rothwell G.W. and Holt B. (1997). Fossils and phenology in the evolution of *Ginkgo biloba*. *Ginkgo-biloba-A Global Treasure*. In T. Hori, R.W. Ridge, W. Tulecke, P. Del Tredici, J. Trémouillaux-Guiller, and H. Tobe (Eds.). Tokyo, Springer-Verlag, pp. 223–230.
- Rudd S. (2003). Expressed sequence tags: Alternative or complement to whole genome sequences? *Trends Plant Sci.* **8**:321–329.
- Rudd S. (2005). openSputnika database to ESTablish comparative plant genomics using unsaturated sequence collections. *Nucleic Acids Res.* **33** (Database Issue):D622–627.
- Saijo Y., Sullivan J.A., Wang H., Yang J., Shen Y., Rubio V., Ma L., Hoecker U., and Deng X.W. (2003). The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. *Genes Dev.* **17**:2642–2647.
- Seo H.S., Yang J.Y., Ishikawa M., Bolle C., Ballessteros M.L., and Chua N.H. (2003). LAF1 ubiquitination by COP1 controls photomorphogenesis and is stimulated by SPA1. *Nature* **423**:995–999.
- Shuai B., Reynaga-Pena C.G., and Springer P.S. (2002). The lateral organ boundaries gene defines a novel, plant-specific gene family. *Plant Physiol.* **129**:747–761.
- Suzuki G., Yanagawa Y., Kwok S.F., Matsui M., and Deng X.W. (2002). *Arabidopsis* COP10 is a ubiquitin-conjugating enzyme variant that acts together with COP1 and the COP9 signalosome in repressing photomorphogenesis. *Genes Dev.* **16**:554–559.
- The Institute for Genomic Research (TIGR) [<http://www.tigr.org/>].
- Theissen G., Becker A., Di Rosa, A., Kanno A., Kim J.T., Munster T., Winter K.U., and Saedler H. (2000). A short history of *MADS-box* genes in plants. *Plant Mol. Biol.* **42**:115–149.
- Weigel D., Ahn J.H., Blazquez M.A., Borevitz J.O., Christensen S.K., Fankhauser C., Ferrandiz C., Kardailsky I., Malancharuvil E.J., Neff M.M., Nguyen J.T., Sato S., Wang Z.Y., Xia Y., Dixon R.A., Harrison M.J., Lamb C.J., Yanofsky M.F., and Chory J. (2000). Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**:1003–1013.
- Zhang P., Tan H.T., Pwee K.H., and Kumar P.P. (2004). Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *Plant J.* **37**:566–577.
- Zhou Z. and Zheng S. (2003). The missing link in *Ginkgo* evolution. *Nature* **423**:821–822.

13

- Abbott M.B., Finney B.P., Edwards M.E., and Kelts K.R. (2000). Lake-level reconstruction and paleohydrology of Birch Lake, Central Alaska, based on seismic reflection profiles and core transects. *Quat. Res.* **53**:154–166.
- ACIA (2004). *Impacts of a Warming Arctic: Arctic Climate Impact Assessment*. Cambridge, Cambridge University Press.
- Anderson P.M. and Brubaker L.B. (1994). Vegetation history of North Central Alaska: A mapped summary of Late-Quaternary pollen data. *Quat. Sci. Rev.* **13**:71–92.
- Anderson P.M., Edwards M.E., and Brubaker L.B. (2004). Results and paleoclimate implications of 35 years of paleoecological research in Alaska. *Dev. Quat. Sci.* **1**:427–440.
- Bond-Lamberty B., Peckham S.D., Ahl D.E., and Gower S.T. (2007). Fire as the dominant driver of Central Canadian boreal forest carbon balance. *Nature* **450**:89–92.
- Brubaker L.B., Garfinkel H.L., and Edwards M.E. (1983). A late Wisconsin and Holocene vegetation history from the central Brooks Range: Implications for Alaskan paleoecology. *Quat. Res.* **20**:194–214.
- Chapin F.S., McGuire A.D., Randerson J., Pielke R.S., Baldocchi D., et al. (2000). Arctic and boreal ecosystems of Western North America as components of the climate system. *Glob. Change Biol.* **6**:211–223.

- de Groot W.J. and Wein R.W. (1999). *Betula glandulosa* Michx. response to burning and postfire growth temperature and implications of climate change. *Int. J. Wildland Fire* **9**:51–64.
- Dugle J.R. (1966). A taxonomic study of Western Canadian species in genus *Betula*. *Canadian J. Bot.* **44**:929–1007.
- Earle C.J., Brubaker L.B., and Anderson P.M. (1996). Charcoal in northcentral Alaskan lake sediments: relationships to fire and late-Quaternary vegetation history. *Rev. Paleobot. Palyn.* **92**:83–95.
- Edwards M.E., Brubaker L.B., Lozhkin A.V., and Anderson P.M. (2005). Structurally novel biomes: a response to past warming in Beringia. *Ecology* **86**:1696–1703.
- Eugster W., Rouse W.R., Pielke R.A.S., McFadden J.P., Baldocchi D.D., et al. (2000). Land-atmosphere energy exchange in Arctic tundra and boreal forest: available data and feedbacks to climate. *Global Change Bio.* **6**:84–115.
- Gavin D.G., Hu F.S., Lertzman K., and Corbett P. (2006). Weak climatic control of stand-scale fire history during the late Holocene. *Ecology* **87**:1722–1732.
- Higuera P.E., Peters M.E., Brubaker L.B., and Gavin D.G. (2007). Understanding the origin and analysis of sediment-charcoal records with a simulation model. *Quat. Sci. Rev.* **26**:1790–1809.
- Johnson E.A. and Gutsell S.L. (1994). Fire frequency models, methods and interpretations. *Adv. Eco. Res.* **25**:239–287.
- Kasischke E.S., Williams D., and Barry D. (2002). Analysis of the patterns of large fires in the boreal forest region of Alaska. *Inter. J. Wildland Fire* **11**:131–144.
- Kaufman D.S., Ager, T.A., Anderson N.J., Anderson P.M., Andrews J.T., et al. (2004). Holocene thermal maximum in the Western Arctic (0–180 degrees W). *Quat. Sci. Rev.* **23**:529–560.
- Liljedahl A., Hinzman L., Busey R., and Yoshikawa K. (2007). Physical short-term changes after a tussock tundra fire, Seward Peninsula, Alaska. *J. Geophy. Res.-Earth Surf.* **112**. doi: 10.1029/2006JF000554.
- Lynch J.A., Clark J.S., and Stocks B.J. (2004). Charcoal production, dispersal and deposition from the Fort Providence experimental fire: Interpreting fire regimes from charcoal records in boreal forests. *Can. J. Forest Res.* **34**:1642–1656.
- Mack M.C., Schuur E.A.G., Bret-Harte M.S., Shaver and G.R., Chapin F.S. (2004). Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization. *Nature* **431**:440–443.
- McGuire A.D., Chapin F.S., Walsh J.E., and Wirth C. (2006). Integrated regional changes in arctic climate feedbacks: Implications for the global climate system. *Ann. Rev. Envir. Res.* **31**:61–91.
- Oechel W.C., Vourlitis G.L., Hastings S.J., Zulueta R.C., Hinzman L., et al. (2000). Acclimation of ecosystem CO₂ exchange in the Alaskan Arctic in response to decadal climate warming. *Nature* **406**:978–981.
- PALE members (1994). Research Protocols for PALE: Paleoclimates of Arctic Lakes and Estuaries. *PAGES workshop report* **1**:53.
- Post W.M., Emanuel W.R., Zinke P.J., and Stangenberger A.G. (1982). Soil carbon pools and world life zones. *Nature* **298**:156–159.
- Racine C., Allen J.A., and Dennis J.G. (2006). Long-term monitoring of vegetation change following tundra fires in Noatak National preserve, Alaska. *Arctic Network of Parks Inventory and Monitoring Program*. Alaska Region, National Park Service. NPS/AKRARCN/NRTR-2006/02, p. 37.
- Reimer P.J., Baillie M.G.L., Bard E., Bayliss A., Beck J.W., et al. (2004). IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* **46**:1029–1058.
- Rhodes A.N. (1998). A method for the preparation and quantification of microscopic charcoal from terrestrial and lacustrine sediment cores. *The Holocene* **8**:113–117.
- Shaver G.R., Giblin A.E., Nadelhoffer K.J., Thieler K.K., Downs M.R., et al. (2006). Carbon turnover in Alaskan tundra soils: effects of organic matter quality, temperature, moisture and fertilizer. *J. Eco.* **94**:740–753.

- Shiu W.K. and Bain L.J. (1982). Experiment size and power comparisons for two-sample Poisson tests. *J. Appl. Stat.* **31**:130–134.
- Sitch S., McGuire A.D., Kimball J., Gedney N., Gamon J., et al. (2007). Assessing the carbon balance of circumpolar Arctic tundra using remote sensing and process modeling. *Ecolo. Appl.* **17**:213–234.
- Stocks B.J., Mason J.A., Todd J.B., Bosch E.M., Wotton B.M., et al. (2002). Large forest fires in Canada, 1959–1997. *J. Geophys. Res.-Atmo.* **108**.
- Tape K., Sturm M., and Racine C. (2006). The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. *Global Change Bio.* **12**:686–702.
- Telford R.J., Heegaard E., and Birks H.J.B. (2004). The intercept is a poor estimate of a calibrated radiocarbon age. *The Holocene* **14**:296–298.
- Thoman D.R. and Bain L.J. (1969). Two sample tests in the Weibull distribution. *Technometrics* **11**:805–815.
- Tinner W., Hu F.S., Beer R., Kaltenrieder P., Schreuer B., et al. (2006). Postglacial vegetational and fire history: Pollen, plant macrofossil and charcoal records from two Alaskan lakes. *Veget. Hist. Archaeobot.* **15**:279–293.
- Walker D.A., Raynolds M.K., Daniels F.J.A., Einarsson E., Evlvebak A., et al. (2005). The circumpolar Arctic vegetation map. *J. Veget. Sci.* **16**:267–282.
- Walker M.D., Wahren C.H., Hollister R.D., Henry G.H.R., Ahlquist L.E., et al. (2006). Plant community responses to experimental warming across the tundra biome. *Proc. Natl. Acad. Sci. USA* **103**:1342–1346.
- Wright H.E., Mann D.H., and Glaser P.H. (1984). Piston corers for peat and lake sediments. *Ecol-ogy* **65**:657–659.
- Zimov S.A., Schuur E.A.G., and Chapin F.S. (2006). Permafrost and the Global Carbon Budget. *Science* **32**:1612–1613.
- Anderson M., Bourgeron P., Bryer M.T., Crawford R., Engleking L., Faber-Langendoen D., Gallyoun M., Goodin K., Grossman D.H., Landaal S., Metzler K., Patterson K.D., Pyne M., Reid M., Sneddon L., and Weakley A.S. (1998). International classification of ecological communities: Terrestrial vegetation of the United States. *The National Vegetation Classification System: List of Types*, Vol. II. Arlington, VA, The Nature Conservancy, P. 502.
- Association for biodiversity information (ABI). (2001). NatureServe. <http://www.natureserve.org/>.
- Bailey R.G. (1994). *Ecoregions of the United States*, revised edition. U.S. Department of Agriculture Forest Service, Washington, DC (1:7,500,000 map).
- Bailey R.G. (1998). *Ecoregion Map of North America*. Miscellaneous Publication 1548. USDA Forest Service, U. S. Department of Agriculture Washington, DC (1:15,000,000 map).
- Bartram W. (1791). Travels through North and South Carolina, Georgia, East and West Florida, the Cherokee country, the extensive territories of the Muscogulges or Creek Confederacy, and the country of the Choctaws: Containing an account of the soil and natural productions of those regions, together with observations on the manners of the Indians. Philadelphia, PA., James and Johnson, p. 824.
- Carrier L. (1923). *The Beginnings of Agriculture in America*. New York, McGraw Hill.
- Christensen, N.L. (1988). Vegetation of the South Eastern Coastal Plain. *North American Terrestrial Vegetation*. In M.G. Barbour and W.D. Billings (Eds.). New York, Cambridge University Press.
- Cowell C.M. (1998). Historical Change in Vegetation and Distribution on the Georgia Piedmont. *Am. Midland Natura.* **140**:78–89.
- Croker T.C. Jr. (1987). *Longleaf Pine: A History of Man and a Forest*. USDA Forest Service. Forestry Report R8-FR7, Southern Region, Atlanta.
- Delcourt H.R. (1979). Late Quaternary vegetation history of the Eastern Highland Rim and adjacent Cumberland Plateau of Tennessee. *Ecol. Monog.* **49**:255–280.

- Delcourt H.R. (1987). The Impact of prehistoric agriculture and land occupation on natural vegetation. *TREE* 2(2):39–44.
- Delcourt P.A. and Delcourt, H.R. (1993). Paleo-climates, Paleo-vegetation, and Paleo-floras during the Late Quaternary. *Flora of North America*, vol. 1. New York, Oxford University Press.
- Delcourt, H.R. and Delcourt, P.A. (1997). Pre-columbian Native American use of fire on Southern Appalachian landscapes. *Conser. Bio.* 11:1010–1014.
- Delcourt P.A. and Delcourt H.R. (1998). Paleo-ecological insights on conservation of biodiversity: A Focus on species, ecosystems, and landscapes. *Ecol. Appl.* 8:921–934.
- FLEPPC (2001). *Lists of Florida's Invasive Species*. Florida Exotic Pest Plant Council, <http://www.fleppc.org/>.
- Greller A.M. (1988). Deciduous forests. *North American Terrestrial Vegetation*. In M.G. Barbour and W.D. Billings (Eds.). New York, Cambridge University Press.
- Grossman D.H., Faber-Langendoen D., Weakley A.S., Anderson M., Bourgeron P., Crawford R., Goodin K., Landaal S., Metzler K., Patterson K.D., Pyne M., Reid M., and Sneddon. L. (1998). International classification of ecological Communities: Terrestrial vegetation of the United States. *The National Vegetation Classification System: Development, Status, and Applications*, Vol. I. Arlington, VA, The Nature Conservancy.
- Kuchler A.W. (1985). *Potential Natural Vegetation*, Revised. National Atlas of the United States, US Geological Survey, Reston VA (1:7,500,000 map).
- Mac M.J., Opler, P.A.; Haecker, C.E.P., and Doran, P.D. (Eds.) (1998). Status and Trends of the Nation's Biological Resources. Vols. 1 and 2. United States Department of the Interior, U.S. Reston, VA, Geological Survey.
- Martin P.S. and Klein R.G. (Eds.) (1984). *Quaternary Extinctions*. Tucson, AZ, University of Arizona Press.
- Milancovic M. (1941). Canon of Insolation and the Ice-Age Problem. Royal Serbian Academy Special Publication 132. cited in Delcourt and Delcourt 1993.
- Mitch W.J. and Gosselink J.G. (1993). *Wetlands*. New York, Van Nostrand Reinhold.
- Noss R.F., LaRoe E.T., and Scott J.M. (1995). *Endangered Ecosystems of the United States: A Preliminary Assessment of Loss and Degradation*. Biological Report 28, Washington, DC, USDI National Biological Service.
- Nuttall, T. (1821). *A Journal of Travels in the Arkansaw Territory during the year 1819*. Philadelphia, PA, T.H. Palmer.
- Palmer E.J. (1926) The Ligneous Flora of Hot springs National Park and Vicinity. *J. Arnold Arbor.* 7:104–135.
- Prasada A.M. and Iverson L.R. (1999). A Climate Change Atlas for 80 Forest Tree Species of the Eastern United States [database] <http://www.fs.fed.us/ne/delaware/atlas/index.html>, Northeast Research Station, Delaware, Ohio, USDA Forest Service.
- Rabinowitz D. (1981). Seven forms of rarity. *The Biological Aspects of Rare Plant Conservation*. In H. Syng (Ed.). New York, Wiley.
- Reveal J.L. and Pringle, J.S. (1993). Taxonomic botany and floristics. *Flora of North America*, vol. 1. New York, Oxford University Press.
- Ricketts T.H., Dinerstein E., Olson D.M., Louks C.J., Eichbaum W., DellaSala D., Kavanagh K., Hedao P., Hurley P.T., Carney K.M., Abell R., and Walters S. (1999). *Terrestrial Ecoregions of North America: A Conservation Assessment*. Washington, D.C., Island Press.
- Schlarbaum S.E., Hebard F., Spaine P.C., and Kamalay J.C. (1997). Three American Tragedies: Chestnut Blight, Butternut Canker, and Dutch Elm Disease, pp. 45–54, In K.O. Britton. *Exotic Pests of Eastern Forests*, Proceedings (April 1997). Tennessee Exotic Pest Plant Council, Nashville.
- Smith E.B. (1994). *Keys to the Flora of Arkansas*. Fayetteville, AR, University of Arkansas Press.
- Stein B.A., Kutner L.S., and Adams J.S. (Eds.) (2000). *Precious Heritage: The Status of Biodiversity in the United States*. New York, Oxford University Press.

- The Nature Conservancy (TNC). (1999). *Ecoregional Map of the United States*. Arlington, VA, TNC.
- University of Illinois. (1997). An Introduction to the Early Cultures of the American Bottom. <http://www.ncsa.uiuc.edu/Cyberia/RiverWeb/History/Cahokia/>.
- USDI, Fish and Wildlife Service. (2001). Species Information: Threatened and Endangered Animals and Plants. <http://endangered.fws.gov/wildlife.html#Species>.
- Walker L.C. and Oswald B.P. (1999). *The Southern Forest: Geography, Ecology, and Silviculture*. Boca Raton, FL, CRC Press.
- Watts W.A. (1971). Postglacial and interglacial vegetation history of Southern Georgia and Central Florida. *Ecology* **52**:676–690.
- Watts W.A. (1980). The Late Quaternary Vegetation History of the Southeastern United States. *Ann. Rev. Eco. and Syst.* **11**:387–409.
- Williams G.W. (2000). The USDA Forest Service—The First Century. USDA Forest Service Publication FS-650, Washington, DC, USDA Forest Service.
- Woods F.W. and Shanks R.E. (1959). Natural Replacement of Chestnut by Other Species in the Great Smoky Mountains National Park. *Ecology* **40**:349–361.
- Wright H.E. Jr. (1981). Vegetation east of the Rocky Mountains 18,000 years ago. *Quaternary Res.* **15**:113–125.
- Wunderlin R.P. and Hansen B.F. (2000). Flora of Florida. *Pteridophytes and Gymnosperms*, Vol. 1. Gainesville, University Press of Florida.
- 15**
- Austin J.D., Lougheed S.C., and Boag P.T. (2004). Discordant temporal and geographic patterns in maternal lineages of Eastern North American frogs, *Rana castanea* (Ranidae) and *Pseudacris crucifer* (Hylidae). *Molec. Phylog. Evo.* **32**:799–816.
- Austin J.D., Lougheed S.C., Neidrauer L., Chek A.A., and Boag P.T. (2002). Cryptic lineages in a small frog: The post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molec. Phylog. and Evo.* **25**:316–329.
- Baskin J.M. and Baskin C.C. (1986). Distribution and geographical/evolutionary relationships of cedar glade endemics in South Eastern United States. *ASB Bull.* **33**:138–154.
- Baskin J.M. and Baskin C.C. (1989). Cedar glade endemics in Tennessee, and a review of their autecology. *J. Tennessee Academy Sci.* **64**:63–74.
- Beerling D.J. (1996). Ecophysiological responses of woody plants to past CO₂ concentrations, *Tree Physiol.* **16**:389–396.
- Beerling D.J. and Woodward F.I. (1993). Ecophysiological responses of plants to global environmental change since the Last Glacial Maximum. *New Phytologist* **125**:641–648.
- Bennett K.D. (1985). The spread of *Fagus grandifolia* across eastern North America during the last 18,000 years. *J. Biogeogr.* **12**:147–164.
- Bennett K.D. and Willis K.J. (2000). Effect of global atmospheric carbon dioxide on glacial-interglacial vegetation change. *Global Ecol. Biogeogr.* **9**:355–361.
- Black R.F. (1976). Periglacial features indicative of permafrost: Ice and soil wedges. *Quat. Res.* **6**:3–26.
- Bradley R.S. (1999). *Paleoclimatology: Reconstructing climates of the Quaternary*, 2nd edition, Amsterdam, Elsevier Academic Press.
- Brant S.V. and Ortí G. (2003). Phylogeography of the northern short-tailed shrew, *Blarina brevicauda* (*Insectivora: Soricidae*): Past fragmentation and postglacial recolonization. *Molec. Ecol.* **12**:1435–1449.
- Braun E.L. (1950). *Deciduous Forests of Eastern North America*. London, Hafner Press.
- Braun E.L. (1951). Plant distribution in relation to the glacial boundary. *Ohio J. Sci.* **51**:139–146.
- Brown J.H. and Gibson A.C. (1983). *Biogeography*. St. Louis, MO, C. V. Mosby Co.
- Bryant V.M., Jr. (1977). A 16,000 year pollen record of vegetation change in central Texas, *Pollenology* **1**:143–156.
- Bryant V.M. and Holloway R.G. (1985). *A late Quaternary Paleoenvironmental Record of Texas: An Overview of the Pollen Evidence*. In V.M. Bryant and R.G. Holloway (Eds.). Dallas,

- American Association of Stratigraphic Palynologists Foundation, pp. 39–70.
- Burns G.W. (1958). Wisconsin age forests in western Ohio: II, Vegetation and burial conditions. *Ohio J. Sci.* **58**:220–230.
- Burns R.M. and Honkala B.H. (1990). *Silvics of North America*, Vol. 1, Conifers, USDA Handbook 654.
- Cain S.A. (1944). *Foundations of Plant Geography*. New York, Harper & Row.
- Cain M.L., Damman H., and Muir A. (1998). Seed dispersal and the Holocene migration of woodland herbs. *Ecol. Monogr.* **68**:325–347.
- Camp W. H. (1950). A biogeographic and para-genetic analysis of the American beech (*Fagus*). *Yearbook of the American Philosophical Society*, pp. 166–169.
- Church S.A., Kraus J.M., Mitchell J.C., Church D.R., and Taylor D.R. (2003). Evidence for multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum*. *Evolution* **57**:372–383.
- Clayton L., Attig J.W., and Mickelson D.M. (2001). Effects of late Pleistocene permafrost on the landscape of Wisconsin, USA. *Boreas* **30**:173–188.
- Cole D.R. and Monger H.C. (1994). Influence of atmospheric CO₂ on the decline of C4 plants during the last deglaciation. *Nature* **368**:533–536.
- Cooper A.W. and Mercer E.P. (1977). Morphological variation in *Fagus grandifolia* Ehrh. in North Carolina. *J. Elis. Mit. Sci. Soc.* **93**:136–149.
- Cowling S.A. (1999). Simulated effects of low atmospheric CO₂ on structure and composition of North American vegetation at the Last Glacial Maximum. *Global Ecol. Biogeogr.* **8**:81–93.
- Cowling S.A. and Sykes M.T. (1999). Physiological significance of low atmospheric CO₂ for plant-climate interactions. *Quat. Res.* **52**:237–242.
- Crowley T.J. (2000). CLIMAP SSTs re-visited. *Clim. Dyn.* **16**:241–255.
- Davis M.B. (1983). Quaternary history of deciduous forests of Eastern North America and Europe. *Ann. Miss. Bot. Gar.* **70**:550–563.
- Davis M.B. (1984). Holocene vegetational history of the Eastern United States, 166–181. *Late-Quaternary Environments of the United States*, Vol. 2. In H.E. Wright Jr., (Ed.). London, The Holocene, Long-man.
- Davis M.B. and Shaw R.G. (2001). Range shifts and adaptive responses to quaternary climate change. *Science*, **292**:673–679.
- Deevey Jr., E.S. (1949). Biogeography of the Pleistocene. *Geol. Soc. Am. Bull.* **60**:1315–1416.
- Delcourt H.R. and Delcourt P.A. (1984). Ice age haven for hardwoods. *Nat. Hist.* **84**:22–28.
- Delcourt P.A. and Delcourt H.R. (1993). Paleoclimates, paleovegetation, and paleofloras during the late Quaternary. *Flora of North America*. In Flora of North America Editorial Committee (Ed.). New York, Oxford University Press, pp. 71–94.
- Denny C.S. (1951). Pleistocene frost action near the border of the Wisconsin drift in Pennsylvania. *Ohio J. Sci.* **51**:116–125.
- Drake B.G., González-Meler M.A., and Long S.P. (1997). More efficient plants: A consequence of rising atmospheric CO₂? *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **48**:609–639.
- Elenga H., Peyron O., Bonnefille R., Jolly D., Cheddadi R., Guiot J., Andrieu V., Bottema S., Buchet G., de Beaulieu J.L., Hamilton A.C., Mailey J., Marchant R., Perez-Obiol R., Reille M., Rioullet G., Scott L., Straka H., Taylor D., Van Campo E., Vincens A., Laarif F., and Jonson H. (2000). Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp. *J. Biogeogr.* **27**:621–634.
- Estill J.C. and Cruzan M.B. (2001). Phytogeography of rare plant species endemic to the southeastern United States. *Castanea* **66**:3–23.
- Farquhar G.D. (1997). Carbon dioxide and vegetation. *Science* **278**:1411.
- Farrera I., Harrison S. P., Prentice I. C., Ramstein G., Guiot J., Bartlein P. J., Bonnefille R., Bush, M., Cramer W., von Grafenstein U., Holmgren K., Hooghiemstra H., Hope G., Jolly D., Lauritzen S.-E., Ono Y., Pinot S., Stute M., and Yu G. (1999). Tropical climates at the last glacial maximum: A new synthesis of terrestrial palaeoclimate data. I. Vegetation, lake-levels and geochemistry. *Clim. Dyn.* **15**:823–856.

- Finkelstein S. A., Gajewski K., and Viau A. E. (2006). Improved resolution of pollen taxonomy allows better biogeographical interpretation of post-glacial forest development: Analyses from the North American Pollen Database. *J. Ecol.* **94**:415–430.
- Froyd C.A. (2005). Fossil stomata reveal early pine presence in Scotland: Implications for post-glacial colonization analyses. *Ecology* **86**:579–586.
- Gamache I., Jaramillo-Correa J.P., Payette S., and Bousquet J. (2003). Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: Imprint of a founder effect associated with postglacial colonization. *Molec. Ecol.* **12**:891–901.
- Ganopolski A., Rahmstorf S., Petoukhov V., and Claussen M. (1998). Simulation of modern and glacial climates with a coupled global model of intermediate complexity. *Nature*, **391**:351–356.
- Gilbert C.R. (1987). Zoogeography of the freshwater fish fauna of southern Georgia and peninsular Florida. *Brimleyana* **13**:25–54.
- Givens C.R. and Givens F.M. (1987). Age and significance of fossil white spruce (*Picea glauca*), Tunica Hills, Louisiana-Mississippi. *Quat. Res.* **27**:283–296.
- Gooding A.M. (1963). Illinoian and Wisconsin glaciations in the White-water basin, southeastern Indiana, and adjacent areas. *J. Geol.* **71**:665–682.
- Greene A.M., Seager R. and Broecker W.S. (2002). Tropical snow-line depression at the Last Glacial Maximum: Comparison with proxy records using a single-cell tropical climate model. *J. Geophys. Res.* **107(D8)**, doi: 10.1029/2001JD000670.
- Grüger E. (1972a). Late Quaternary vegetational development in south-central Illinois. *Quaternary Res.* **2**:217–231.
- Grüger, E. (1972b). Pollen and seed studies of Wisconsinan vegetation in Illinois, USA. *Geol. Soc. Am. Bull.* **83**:2715–2734.
- Guiot J., Torre F., Jolly D., Peyron O., Boreux J.J., and Cheddadi R. (2000). Inverse vegetation modeling by Monte Carlo sampling to reconstruct palaeoclimates under changed precipitation seasonality and CO₂ conditions: application to glacial climate in Mediterranean region. *Ecol. Modell.* **127**:119–140.
- Harrison S.P. and Prentice C.I. (2003). Climate and CO₂ controls on global vegetation distribution at the last glacial maximum: Analysis based on palaeovegetation data, biome modeling and palaeoclimate simulations. *Glob. Chang. Biol.* **9**:983–1004.
- Harvill Jr., A.M. (1973). Phytogeography of the Virginias and the equilibrium concept of landscape. *Castanea* **38**:266–268.
- Hewitt G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B* **359**:183–195.
- Hultén E. (1937). *Outline of the History of Arctic and Boreal Biota During the Quaternary Period*. Stockholm, Bokforlags Aktiebolag Thule.
- Huntley B., Alfano M.J., Allen J.R.M., Pollard D., Tzedakis P.C., de Beaulieu J.-L., Grüger E., and Watts B. (2003). European vegetation during Marine Oxygen Isotope Stage-3. *Quat. Res.* **59**:195–212.
- Jackson S.T. and Givens C.R. (1994). Late Wisconsinan vegetation and environment of the Tunica Hills regions, Louisiana and Mississippi. *Quat. Res.* **41**:316–325.
- Jackson S.T. and Weng C. (1999). Late Quaternary extinction of a tree species in eastern North America. *PNAS* **96**:13847–13852.
- Jacobson Jr., G.L., Webb III T., and Grimm E.C. (1987). Patterns and rates of vegetation change during the deglaciation of eastern North America. *The geology of North America*, Vol. K-3 In W.F. Ruddiman and H.E. Wright Jr. (Eds.). North America and adjacent oceans during the last deglaciation, Geological Soc. Am., pp. 277–288.
- Jolly D. and Haxeltine A. (1997). Effect of low glacial atmosphere CO₂ on tropical African montane vegetation. *Science* **276**:786–788.
- Kageyama M., Peyron O., Pinot S., Tarasov P., Guiot J., Joussaume S., and Ramstein G. (2001). The Last Glacial Maximum climate over Europe and western Siberia: a PMIP comparison between models and data. *Clim. Dyn.* **17**:23–43.
- Levin D.A. and Wilson A.C. (1976). Rates of evolution in seed plants: Net increase in diversity of chromosome numbers and species

- numbers through time. *Proc. Natl. Acad. Sci.* **73**:2086–2090.
- Levis S. and Foley J.A. (1999). CO₂, climate, and vegetation feedbacks at the Last Glacial Maximum. *J. Geophys. Res.* **104**:31191–31198.
- Maher Jr., L.J., Miller N.G., Baker R.G., Curry B.B., and Mickelson D.M. (1998). Paleobiology of the sand beneath the Valders diamicton at Valders, Wisconsin. *Quat. Res.* **49**:208–221.
- Maxwell J.A. and Davis M.B. (1972). Pollen evidence of Pleistocene and Holocene vegetation on the Allegheny Plateau, Maryland. *Quat. Res.* **2**:506–530.
- McLachlan J.S. and Clark J.S. (2004). Reconstructing historical ranges with fossil data at continental scales. *Forest Ecol. Manage.* **197**:139–147.
- McLachlan J.S., Clark J.S., and Manos P.S. (2005). Molecular indicators of tree migration capacity under rapid climate change. *Ecology* **86**:2088–2098.
- Mickelson D.M., Clayton L., Fullerton D.S., and Borns Jr., H.W. (1983). The late Wisconsin glacial record of the Laurentide ice sheet in the United States. *Late-Quaternary Environments of the United States*, Vol. 1. In H.E. Wright Jr. (Ed.). The late Pleistocene, S.C. Porter (Ed.). Minneapolis, University of Minnesota Press, pp. 3–37.
- Nakagawa T., Tarasov P.E., Nishida K., Gotanda K., and Yasuda Y. (2002). Quantitative pollen-based climate reconstruction in central Japan: Application to surface and Late Quaternary spectra. *Quat. Sci. Rev.* **21**:2099–2113.
- Overpeck J.T., Webb R.S., and Webb III T. (1992). Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology* **20**:1071–1074.
- Parker K.C., Hamrick J.L., Parkers A.J., and Stacey E.A. (1997). Allozyme diversity in *Pinus virginiana* (Pinaceae): Intraspecific and interspecific comparisons. *Am. J. Botany* **84**:1372–1382.
- Péwé T.L. (1983). The periglacial environment in North America during Wisconsin time. *Late-Quaternary Environments of the United States*. In H.E. Wright Jr., Vol. 1, The late Pleistocene, S.C. Porter (Ed.). Minneapolis, University of Minnesota Press, 157–189.
- Peyron O., Guiot J., Cheddadi R., Tarasov P., Reille M., de Beaulieu Bottema S., and Andrieu V. (1998). Climatic reconstruction in Europe for 18,000 yr B.P. from pollen data. *Quat. Res.* **49**:183–196.
- Pielou E.C. (1991): *After the Ice Age*. Chicago, University of Chicago Press.
- Pinot S., Ramstein G., Harrison S.P., Prentice I.C., Guiot J., Stute M., and Joussaume S. (1999). Tropical paleoclimates at the Last Glacial Maximum: Comparison of Paleoclimate Modeling Intercomparison Project (PMIP) simulations and paleodata. *Clim. Dyn.* **15**:857–874.
- Polley H.W., Johnson H.B., Marino, B.D., and Mayeux, H.S. (1993). Increase in C3 plant water-use efficiency and biomass over glacial to present CO₂ concentrations. *Nature* **361**:61–64.
- Polley H.W., Johnson H.B., and Mayeux H.S. (1995). Nitrogen and water requirements of C3 plants grown at glacial to present carbon dioxide concentrations. *Functional Ecol.* **9**:86–96.
- Potts R. and Behrensmeyer A.K. (1992). Late Cenozoic terrestrial ecosystems. *Terrestrial Ecosystems Through Time*. In A.K. Behrensmeyer, J.D. Damuth, W.A. DiMichele, R. Potts, H.-D. Sues, and S.L. Wing (Eds.). Chicago, University of Chicago Press, pp. 419–436.
- Prentice I.C., Bartlein P.J., and Webb III T. (1991). Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology* **72**:2038–2056.
- Qian H. (1999). Spatial pattern of vascular plant diversity in North America north of Mexico and its floristic relationship with Eurasia. *Ann. Bot.* **83**:271–283.
- Ritchie J.C. (1987). *Postglacial Vegetation of Canada*. Cambridge, Cambridge University Press.
- Robinson J.M. (1994). Speculations on carbon dioxide starvation, late Tertiary evolution of stomatal regulation and floristic modernization. *Plant Cell Environ.* **17**:7345–7354.
- Roy K., Valentine J.W., Jablonski D., and Kidwell S.M. (1996). Scales of climatic variability and time averaging in Pleistocene biotas: Implications for ecology and evolution. *Trends Ecol. Evol.* **11**:458–463.

- Royall P.D., Delcourt P.A., and Delcourt H.R. (1991). Late Quaternary paleoecology and paleoenvironments of the central Mississippi alluvial valley. *Geol. Soc. Am. Bull.* **103**:157–170.
- Saxe H., Ellsworth D.S., and Heath, J. (1998). Tree and forest functioning in an enriched CO₂ atmosphere. *New Phytologist* **139**:395–436.
- Schofield W.B. (2004). Endemic genera of bryophytes of North America (North of Mexico). *Preslia Praha* **76**:255–277.
- Schoonmaker P.K. and Foster D.R. (1991). Some implications of paleoecology for contemporary ecology. *Botanical Rev.* **57**:204–245.
- Soltis D.E., Morris A.B., McLachlan J.S., Manos P.S., and Soltis P.S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molec. Ecol.* **15**:4261–4293.
- Stewart J.R. and Lister A.M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* **16**:608–613.
- Street-Perrott F.A., Huang Y., Perrott R.A., Eglington G., Barker P., Khelifa L.B., Harkness, D.D., and Olago D.O. (1997). Impact of lower atmospheric carbon dioxide on tropical mountain ecosystems. *Science* **278**:1422–1426.
- Svenning J.-C. (2003). Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora. *Ecol. Lett.* **6**:646–653.
- Tallis J.H. (1991). *Plant Community History*. New York, Chapman and Hall.
- Tarasov P.E., Peyron O., Guiot J., Brewer S., Volkova V.S., Bezusko L.G., Dorofeyuk N.I., Kvavadze E.V., Osipova I.M., and Panova N.K. (1999). Last Glacial Maximum climate of the former Soviet Union and Mongolia reconstructed from pollen and plant macrofossil data. *Clim. Dyn.* **15**:227–240.
- Watts W.A. (1970). The full-glacial vegetation of northwestern Georgia. *Ecology* **51**:17–33.
- Watts W.A. (1971). Postglacial and interglacial vegetation history of southern Georgia and central Florida. *Ecology* **52**:676–689.
- Watts W.A. (1973). The vegetation record of a mid-Wisconsin interstadial in North West Georgia. *Quaternary Res.* **3**:257–268.
- Watts W.A. (1979). Late Quaternary vegetation of central Appalachia and the New Jersey coastal plain. *Ecol. Monogr.* **49**:427–469.
- Watts W.A. (1980a). The late Quaternary vegetation history of the southeastern United States. *Ann. Rev. Ecol. Syst.* **11**:387–409.
- Watts W.A. (1980b). Late Quaternary vegetation history at White Pond on the inner coastal plain of South Carolina. *Quat. Res.* **13**:187–199.
- Watts W.A. and Stuiver M. (1980). Late Wisconsin climate of northern Florida and the origin of species-rich deciduous forest. *Science* **210**:325–327.
- Webb III T., Anderson K.H., Bartlein P.J., and Webb R.S. (1997). Late Quaternary climate change in eastern North America: A comparison of pollen-derived estimates with climate model results. *Quat. Sci. Rev.* **17**:587–606.
- Webb III T., Bartlein P.J., Harrison S.P., and Anderson K.H. (1993). Vegetation, lake levels, and climate in eastern North America for the past 18,000 years. *Global climates Since the Last Glacial Maximum*. In H.E. Wright Jr., J.E. Kutzbach, T. Webb III, W.F. Ruddiman, F.A. Street-Perrott, and P.J. Bartlein (Eds.). Minneapolis, University of Minnesota Press, pp. 415–467.
- Webb III T., Bartlein P.J., and Kutzbach J.E. (1988). Climatic change in eastern North America during the past 18,000 years: Comparisons of pollen data with model results. *The geology of North America*, Vol. K-3. In W.F. Ruddiman and H.E. Wright Jr. (Eds.). North American and adjacent oceans during the last deglaciation. Geological Society of America, pp. 447–462.
- Whitehead D.R. (1973). Late Wisconsin vegetational changes in unglaciated Eastern North America. *Quat. Res.* **3**:621–631.
- Wilkins G.R., Delcourt P.A., Delcourt H.R., Harrison F.W., and Turner M.R. (1991). Paleoecology of central Kentucky since the Last Glacial Maximum. *Quat. Res.* **36**:224–239.
- Williams J.W., Shuman B.N., Webb III T., Bartlein P.J., and Leduc P.L. (2004). Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecol. Monogr.* **74**:309–334.

Willis K.J. and van Andel T.H. (2004). Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quat. Sci. Rev.* **23**:2369–2387.

Willis, K.J. and Whittaker R.J. (2000). The refugial debate. *Science* **287**:1406–1407.

Wiser S.K. (1998). Comparison of southern Appalachian high-elevation outcrop plant communities with their northern Appalachian counterparts. *J. Biogeogr.* **25**:501–513.

Wolfe J.N. (1951). The possible role of microclimate. *Ohio J. Sci.* **51**:134–138.

16

Allen T.F.H. and Hoekstra T.W. (1992). *Toward a Unified Ecology*. New York, Columbia University Press.

Andersen S.T. (1970). The relative pollen productivity and pollen representation of North European trees, and correction factors for tree pollen spectra determined by surface pollen analyses from forests. *Danmarks Geologiske Undersøgelse, Raekke II* **96**:1–199.

Anschuetz K.F. (1995). *Preliminary Report for the 1993 Field Season*. University of Michigan, Rio del Oso Archaeological Survey, Española Ranger District, Santa Fe National Forest, Report on File. Santa Fe, NM, USDA Forest Service, Santa Fe National Forest Supervisor's Office.

Birks J.H.B. and Gordon A.D. (1985). *Numerical Methods in Quaternary Pollen Analysis*. London, Academic Press, pp. 202–204.

Birks H.J.B., Line J.M., and Persson T. (1988). Quantitative estimation of human impact on cultural landscape development. *The Cultural Landscape Past, Present and Future*. In H.H. Birks, H.J.B. Birks, P.E. Kaland, and D. Moe (Eds.). Cambridge, Cambridge University Press, pp. 229–240.

Caseldine C. and Hatton J. (1993). The development of high moorland on Dartmoor: Fire and the influence of Mesolithic activity on vegetation change. *Climate Change and Human Impact on the Landscape: Studies in Paleoecology and Environmental Archaeology*. In F.M. Chambers (Ed.). London, Chapman & Hall, pp. 119–131.

Chambers F.M. (1993a). *Climate Change and Human Impact on the Landscape: Studies in Paleoecology and Environmental Archaeology*. London, Chapman & Hall.

Chambers F.M. (1993b). Introduction, climate change and human impact: Relationships and interaction. *Climate Change and Human Impact on the Landscape: Studies in Paleoecology and Environmental Archaeology*. In F.M. Chambers (Ed.). London, Chapman & Hall, p. 3.

Chepstow-Lusty A.J., Bennett K.D., Fjeldsa J., Kendall A., Galiano W., and Herrera A.T. (1998). Tracing 4,000 years of environmental history in the Cuzco Area, Peru, from the pollen record. *Mt Res Dev* **18**:159–172.

Clark J.S. and Royall P.D. (1995). Particle-size evidence for source areas of charcoal accumulation in late Holocene sediments of eastern North American lakes. *Quat. Res.* **43**:80–89.

Clark J.S. and Royall P.D. (1996). Local and regional sediment charcoal evidence for fire regimes in pre-settlement North-Eastern North America. *J. Ecol.* **84**: 365–382.

Cottam G. and Curtis J.T. (1956). The use of distance measures in phytosociological sampling. *Ecology* **37**: 451–460.

Davis M.B. (1963). On the theory of pollen analysis. *American J. Sci.* **261**:897–912.

Delcourt H.R. and Delcourt P.A. (1996). Pre-settlement landscape heterogeneity: Evaluating grain of resolution using general land office survey data. *Lands. Ecol.* **11**:363–381.

Delcourt H.R. and Delcourt P.A. (1997). Pre-Columbian Native American use of fire on southern Appalachian landscapes. *Cons. Bio.* **11**:1010–1014.

Delcourt P.A., Delcourt H.R., Ison C.R., Sharp W.E., and Gremillion K.J. (1998). Prehistoric human use of fire, the Eastern Agricultural Complex, and Appalachian oak-chestnut forests: Paleoecology of Cliff Palace Pond, Kentucky. *Am. Antiquity* **63**:263–278.

Deneven W.M. (1992). The pristine myth: The landscape of the Americas in 1492. *Ann. Ass. Am. Geogr.* **82**:369–385.

- Dethier D.P. and Demsey K.A. (1984). *Erosional History and Soil Development on Quaternary Surfaces, Northwest Española Basin, New Mexico, Guidebook 35*. New Mexico Geological Society, Albuquerque, NM.
- Dethier D.P. and Reneau S.L. (1995). *Quaternary History of the Western Española Basin, New Mexico, Guidebook 46*. Albuquerque, NM, New Mexico Geological Society.
- Dethier D.P., Harrington C.D. and Aldrich M.J. (1988). Late Cenozoic rates of erosion in the western Española Basin, New Mexico: Evidence from geologic dating of erosion surfaces. *Bull. Geol. Soc. Am.* **100**:928–937.
- Edwards K.J. (1988). The hunter-gatherer, agriculture transition and the pollen record in the British Isles. *The Cultural Landscape—Past, Present and Future*. In H.H. Birks, H.J.B. Birks, P.E., Kaland, and D. Moe (Eds.). Cambridge, Cambridge University Press, pp. 255–266.
- Hobbs R. (1997). Future landscapes and the future of landscape ecology. *Land. Urb. Plan.* **37**:1–9.
- Jackson S.T. (1994). Pollen and spores in Quaternary lake sediments as sensors of vegetation composition: Theoretical models and empirical evidence. *Sediments of Organic Particles*. In A. Traverse (Ed.). Cambridge, Cambridge University Press, pp. 253–286.
- Jackson S.T. and Kearsley J.B. (1998). Quantitative representation of local forest composition in forest-floor pollen assemblages. *J. Ecol.* **86**:474–490.
- Lindeman R.L. (1942). The trophic-dynamic aspect of ecology. *Ecology* **23**:399–418.
- Mannion A.M. (1991). *Global Environmental Change: A Natural And Cultural Environmental History*. New York, Wiley.
- Mcintosh R.J., Tainter J.A., and Mcintosh S.K. (Eds.) (2000). *The Way the Wind Blows: Climate, History and Human Action*. New York, Columbia University Press.
- Norton W. (1989). *Explorations in the Understanding of Landscape: A Cultural Geography*. Westport, CT, Greenwood Press.
- Parsons R.W. and Prentice I.C. (1981). Statistical approaches to R-Values and the pollen-vegetation relationship. *Rev. Palaeobot. Palynol.* **32**:127–152.
- Pearsall D.M. (2000). *Paleoethnobotany: A Handbook of Procedures*. New York, Academic Press.
- Periman R.D. (2001). The changing cultural landscapes of New Mexico's Rio Del Oso valley. Doctoral dissertation, Albuquerque, University of New Mexico.
- Prentice I.C. and Parsons R.W. (1983). Maximum likelihood linear calibration of pollen spectra in terms of forest composition. *Biometrics* **39**:1051–1057.
- Reneau S.L., McDonald E.V., Gardner J.N., Kolbe T.R., Carney J.S., Watt P.M., and Longmire P.A. (1996). Erosion and deposition on the Pajarito Plateau, New Mexico, and implications for geomorphic responses to late Quaternary climatic changes, Guidebook 47. Albuquerque, NM, New Mexico Geological Society.
- Rossignol J. and Wandsnider L. (1992). *Space, Time, and Archaeological Landscapes*. New York, Plenum Press.
- Schmidt K. (1998). Coming to grips with the world's greenhouse gases. *Science* **281**:504–506.
- Scott Cummings L. (2001). *Palyenology, Phytolith, and Microscopic Charcoal Analyses of Rio Del Oso, New Mexico*. Unpublished report, Golden, CO, Paleo Research Institute.
- Simmons I.G. (1988). Vegetation change during the Mesolithic in the British Isles: Some amplifications. *Climate Change and Human Impact on the Landscape: Studies in Paleoenvironment and Environmental Archaeology*. In F.M. Chambers (Ed.). London, Chapman & Hall, pp. 109–118.
- Sugita S. (1994). Pollen representation of vegetation in Quaternary sediments; theory and method in patchy vegetation. *J. Ecol.* **82**:881–897.
- Vitousek P.M., Mooney H.A., Lubchenco J., and Melillo J. (1997). Human domination of the Earth's ecosystem. *Science* **277**:494–499.
- Walker D. and Singh G. (1993). Earliest palynological records of human impact on the world's vegetation. *Climate Change and Human Impact on the Landscape: Studies in Paleoenvironment and*

- Environmental Archaeology*. In F.M. Chambers (Ed.). London, Chapman & Hall, pp. 101–108.
- Webb T., Howe S.E., Bradshaw R.H.W., and Heide K.M. (1981). Estimating plant abundances from pollen percentages: The use of regression analysis. *Rev. Palaeobot. Palynol.* **34**:269–300.
- Whitmore T.M. and Turner B.L. (1990). Landscapes of cultivation in Mesoamerica on the eve of the conquest. *Ann. Ass. Am. Geogr.* **82**:402–442.
- Wildeman G. and Brock J.H. (2000). Grazing in the Southwest: History of land use and grazing since 1540. *Livestock Management in the American Southwest*. In R. Jemison and C. Raish (Eds.). Amsterdam, Elsevier, pp. 1–25.
- Willey G.R. and Sabloff J.A. (1993). *A History of American Archaeology*. New York, W.H. Free-man.
- Wozniak F.J., Kemrer M.F., and Carrillo C.M. (1992). History and ethnohistory along the rio chama. Albuquerque, NM, U.S. Army Corps of Engineers, Albuquerque, New Mexico District.
- Cowling S. and Sykes M. (1999). Physiological significance of low atmospheric CO₂ for plant-climate interactions. *Quat. Res.* **52**:237–242.
- Davis B.A.S., Brewer S., Stevenson A.C., Guiot J., et al. (2003). The temperature of Europe during the Holocene reconstructed from pollen data, *Quat. Sci. Rev.* **22**:1701–1716.
- Dykoski C.A., Edwards R.L., Cheng H., Yuan D.X., Cai Y.J., Zhang M.L., Lin Y.S., Qing J.M., An Z.S., and Revenaugh J. (2005). A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China, *Earth Planet. Sci. Lett.* **233**:71–86.
- EPICA (2004). Eight Glacial cycles from an Antarctic ice core. *Nature* **429**:623–628.
- Farquhar G D. (1997). Carbon dioxide and vegetation. *Science* **278**:1411–1411.
- Guiot J. Beaulieu de J.L., Cheddadi R., David F., Ponel P., and Reille M. (1993). The climate in Western Europe during the last Glacial/Interglacial cycle derived from pollen and insect re-mains. *Palaeogeogr. Palaeocl.* **103**:73–93.
- Guiot J., Torre F., Jolly D., Peyron O., Boreux J.J., and Cheddadi R. (2000). Inverse vegetation modeling by Monte Carlo sampling to reconstruct palaeoclimates under changed precipitation seasonality and CO₂ conditions: application to glacial climate in Mediterranean region. *Ecol. Model.* **127**:119–140.
- He Y., Theakstone W., Zhang Z., Zhang D., Yao T., Chen T., Shen Y., and Pang H. (2004). Asynchronous Holocene climatic change across China. *Quat. Res.* **61**:52–63.
- Jiang W., Guo Z., Sun X., Wu H., Chu G., Yuan B., Hatté C., and Guiot J. (2006). Reconstruction of climate and vegetation changes of Lake Bayanchagan (Inner Mongolia): Holocene variability of the East Asian monsoon. *Quat. Res.* **65**:411–420.
- Jolly D. and Haxeltine A. (1997). Effect of low glacial atmospheric CO₂ on tropical African montane vegetation. *Science* **276**:786–788.
- Joussaume S. and Taylor K. (1995). Status of the PMIP. *Proceedings of the First International AMIP Scientific Conference*, Geneva, WMO, pp 425–430.

17

- Alley R.B., Mayewski P.A., Sowers T., Stuiver M., Taylor K.C., and Clark P.U. (1997). Holocene climatic instability: A prominent, widespread event 8,200 years ago. *Geology* **25**:483–486.
- An C., Feng Z.D., and Barton L. (2006). Dry or humid? Mid-Holocene humidity changes in arid and semi-arid China. *Quat. Sci. Rev.* **25**:351–361.
- An Z.S. (2000). The history and variability of the East Asian paleomonsoon climate. *Quat. Sci. Rev.* **19**:171–187.
- Berger A. (1978). Long-term variations of caloric insolation resulting from the earth's orbital elements. *Quat. Res.* **9**:139–167.
- Braconnot P., Loutre M.F., Dong B., Joussaume S., and Valdes P. (2002). How the simulated change in monsoon at 6 ka BP is related to the simulation of the modern climate: results from the Paleoclimate Modeling Intercomparison Project. *Clim. Dynam.* **19**:107–121.
- Cheddadi R., Yu G., Guiot J., Harrison S.P., and Prentice I.C. (1996). The climate of Europe 6,000 years ago. *Clim. Dynam.* **13**:1–9.

- Liu H.Y., Xu L.H., and Cui H.T. (2002). Holocene history of desertification along the woodland-steppe border in northern China. *Quat. Res.* **57**:259–270.
- Magny M., Guiot J., and Schoellhammer P. (2001). Quantitative reconstruction of Younger Dryas to mid-Holocene paleoclimates at Le Locle, Swiss Jura, using pollen and lake-level data. *Quat. Res.* **56**:170–180.
- MCPD (Members of China Quaternary Pollen Database) (2000). Pollen-based Biome Reconstruction at Middle Holocene (6 ka BP) and Last Glacial Maximum (18 ka BP) in China. *J. Integr. Plant Biol.* **42**:1201–1209.
- MCPD (Members of China Quaternary Pollen Database) (2001). Simulation of China Biome Reconstruction Based on Pollen Data from Surface Sediment Samples. *J. Integr. Plant Biol.* **43**:201–209.
- Prentice I.C. and Jolly D. (2000). Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *J. Biogeogr.* **27**(3):507–519.
- Prentice I.C., Cramer W., Harrison S.P., Leemans R., Monserud R.A., and Solomon A.M. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* **19**:117–134.
- Prentice I.C., Guiot J., Huntley B., Jolly D., and Cheddadi R. (1996). Reconstructing biomes from palaeoecological data: A general method and its application to European pollen data at 0 and 6 ka. *Clim. Dynam.* **12**:185–194.
- Ren G.Y. and Beug H.J. (2002). Mapping Holocene pollen data and vegetation of China. *Quat. Sci. Rev.* **21**:1395–1422.
- Rohling E.J. and Palike H. (2005). Centennial-scale climate cooling with a sudden cold event around 8,200 years ago. *Nature* **434**:975–979.
- Seret G., Guiot J., Wansard G., Beaulieu J.L.d., and Reille M. (1992). Tentative Palaeoclimatic Reconstruction Linking Pollen and Sedimentology in La Grande Pile (Vosges, France). *Quat. Sci. Rev.* **11**:425–430.
- Shao X., Wang, Y., Cheng H., Kong X., Wu J., and Lawrence E. (2006). Long-term trend and abrupt events of the Holocene Asian monsoon inferred from a stalagmite $\delta^{18}\text{O}$ record from Shennongjia in Central China. *Chinese Sci. Bull.* **51**:221–228.
- Shi Y.F., Kong Z.Z., Wang S.M., Tang L.Y., Wang F.B., Yao T.D., Zhao X.T., Zhang P.Y., and Shi S.H. (1993). Mid-Holocene Climates and Environments in China. *Global Planet. Change* **7**:219–233.
- Tarasov P.E., Webb T.I., Andreev A.A., Afanas'eva N.B., Berezina N.A., Bezusko L.G., Blyakharchuk T.A., Bolikhovskaya N.S., Cheddadi R., Chernavskaya M.M., Cher-nova G.M., Dorofeyuk N.I., Dirksen V.G., Elina G.A., Filimonova L.V., Glebov F.Z., Guiot J., Gunova V.S., Harrison S.P., Jolly D., Khomutova V.I., Kavadze E.V., Osipova I.M., Panova N.K., Prentice I.C., Saarse L., Sevastyanov D.V., Volkova V.S., and Zernitskaya V.P. (1998). Present-day and mid-Holocene biomes reconstructed from pollen and plant macrofossil data from the former Soviet Union and Mongolia: Global change-BIOME 6000. *J. Biogeogr.* **25**:1029–1053.
- Tarasov P.E., Guiot J., Cheddadi R., Andreev A.A., Bezusko L.G., Blyakharchuk T.A., Dorofeyuk N.I., Filimonova L.V., Volkova V.S., and Zernitskaya V.P. (1999). Climate in Northern Eurasia 6,000 Years Ago Reconstructed From Pollen Data. *Earth Planet. Sci. Lett.* **171**:635–645.
- Wang B., Clemens S., and Liu P. (2003). Contrasting the Indian and East Asian monsoons: implications on geologic timescales. *Mar. Geol.* **201**:5–21.
- Wei K. and Gasse F. (1999). Oxygen Isotopes in Lacustrine Carbonates of West China Revisited: Implications for Post Glacial Changes in Summer Monsoon Circulation. *Quat. Sci. Rev.* **18**:1315–1334.
- Wu H., Guiot J., Brewer S., and Guo Z. (2007). Climatic changes in Eurasia and Africa at the Last Glacial Maximum and mid-Holocene: reconstruction from pollen data using inverse vegetation modelling. *Clim. Dynam.* **29**:211–229.
- Xiao J.L., Xu Q.H., Nakamura T., Yang X.L., Liang W.D., and Inouchi Y. (2004). Holocene vegetation variation in the Daihai Lake region of north-central China: a direct indication of the Asian monsoon climatic history. *Quat. Sci. Rev.* **23**:1669–1679.

- Yu G., Prentice I.C., Harrison S.P., and Sun X.J. (1998). Pollen-based biome reconstructions for China at 0 and 6,000 years. *J. Biogeogr.* **25**:1055–1069.
- Yu G., Xue B., Liu J., and Chen X. (2003). LGM Lake records from China and an analysis of climate dynamics using a modelling approach. *Global Planet. Change* **38**:223–256.
- Zhou W.J., Yu X.F., Jull A.J.T., Burr G., Xiao J.Y., Lu X.F., and Xian F. (2004). High-resolution evidence from southern China of an early Holocene optimum and a mid-Holocene dry event during the past 18,000 years. *Quat. Res.* **62**:39–48.
- Zhou W.J., Xie S.C., Meyers P.A., and Zheng Y.H. (2005). Reconstruction of late glacial and Holocene climate evolution in southern China from geopolips and pollen in the Dingnan peat sequence. *Org. Geochem.* **36**:1272–1284.
- ## 18
- Aitchison J.C., Ali J.R., and Davis A.M. (2007). When and where did India and Asia collide? *J Geophys Res.* **112B**:Art. B05423.
- Ali J.R. and Aitchison J.C. (2008). Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Sci. Rev.* **88**:145–166.
- Ali J.R. and Aitchison J.C. (2005). Greater India. *Earth-Sci. Rev.* **72**:169–188.
- Angiosperm Phylogeny Group (APG) (2003). An update of the Angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**:399–436.
- Barby M.F. and Pierce N.E. (2007). Systematics, biogeography and diversification of the Indo-Australian genus *Delias* Hübner (Lepidoptera: Pieridae): phylogenetic evidence supports an “out-of-Australia” origin. *Syst. Entomol.* **32**:2–25.
- Beebe N.W. and Cooper R.D. (2002). Distribution and evolution of the *Anopheles punctulatus* group (Diptera: Culicidae) in Australia and Papua New Guinea. *Int. J Parasitol.* **32**:563–574.
- Bergh N.G. and Linder H.P. (2009). Cape diversification and repeated out-of-Southern Africa dispersal in paper daises (Asteraceae-Gnaphalieae). *Mol. Phylogenet. Evol.* **51**:5–18.
- Bird M.I., Taylor D., and Hunt C. (2005). Paleoenvironments of insular Southeast Asia during the Last Glacial Period: a savanna corridor in Sundaland. *Quat. Sci. Rev.* **24**:2228–2242.
- Chesters K.I.M. (1955). Some plant remains from the Upper Cretaceous and Tertiary of West Africa. *Ann. Mag. Nat. Hist.* **12**:489–504.
- Dilcher D.L. and Crane P.R. (1984). Archaeanthus: an early angiosperm from the Cenomanian of the Western Interior of North America. *Ann. Missouri. Bot. Gard.* **71**:351–361.
- Couvreur T.L.P. (2008). [<http://tlpcouvreur.googlepages.com/beastpartitioning>] BEAST partitioning: a 7 step tutorial.
- Couvreur T.L.P., Chatrou L.W., Sosef M.S.M., and Richardson J.E. (2008a). Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* **6**:54.
- Couvreur T.L.P., Richardson J.E., Sosef M.S.M., Erkens R.H.J., and Chatrou L.W. (2008b). Evolution of syncarpy and other morphological characters in African Annonaceae: a posterior mapping approach. *Mol. Phylogenet. Evol.* **47**:302–318.
- Doyle J.A. and Le Thomas A. (1996). Phylogenetic analysis and character evolution in Annonaceae. *Bull. Mus. Hist. Nat., Paris, B., Adansonia* **18**:279–334.
- Doyle J.A. and Le Thomas A. (1997). Significance of palynology for phylogeny of Annonaceae: experiments with removal of pollen characters. *Pl. Syst. Evol.* **206**:133–159.
- Doyle J.A. and Le Thomas A. (1994). Cladistic analysis and pollen evolution in Annonaceae. *Acta. Bot. Gal.* **141**:149–170.
- Doyle J.A., Sauquet H., Scharaschkin T., and Le Thomas A. (2004). Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *Int. J. Pl. Sci.* **165**:S55–S67.
- Drummond A.J. and Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**:art. 214.
- Drummond A.J. and Rambaut A. (2008). [<http://beast.bio.ed.ac.uk/>] BEAST ver. 1.4.8.

- Drummond A.J., Ho S.Y.W., Phillips M.J., and Rambaut A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**:699–710.
- Drummond A.J., Ho S.Y.W., Rawlence N., and Rambaut A. (2007). *A Rough Guide to BEAST 1.4.Auckland*: University of Auckland.
- Erkens R.H.J., Chatrou L.W., Maas J.W., Niet T., and Savolainen V. (2007). A rapid diversification of rainforest trees (*Guatteria*; Annonaceae) following dispersal from Central to South America. *Mol. Phylogenet. Evol.* **44**:399–411.
- Gingerich P.D. (2006). Environment and evolution through the Paleocene-Eocene thermal maximum. *Trends. Ecol. Evol.* **21**:246–253.
- Gröcke D.R., Ludvigson G.A., Witzke B.L., Robinson S.A., Joeckel R.M., Ufnar D.F., and Ravn R.L. (2006). Recognizing the Albian-Cenomanian (OAE1d) sequence boundary using plant carbon isotopes: Dakota Formation, Western Interior Basin, USA. *Geology* **34**:193–196.
- Hall R. (1998). The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In *Biogeography and geological evolution of SE Asia*. Edited by: Hall R., Holloway J.D. Leiden: Backhuys Publishers99–131.
- Hall R. (2001). Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In *Faunal and Floral Migrations and Evolution in SE Asia-Asutralasia*. Edited by: Metcalfe I., Smith J.M.B., Morwood M., Davidsson I.D.. Lisse: Swets and Zeitlinger Publisher 35–56.
- Hall R. (2001). Extension during late Neogene collision in east Indonesia and New Guinea. *J. Virtual Explorer* **4**:1–13.
- Hall R. and Morley C.K. (2004). Sundaland Basins. In *Continent-Ocean Interactions within the East Asian Marginal Seas. Geophysical Monograph*. Vol **149**. Edited by: Clift P., Wang P., Kuhnt W., Hayes D.E. USA, Washington DC: American Geophysical Union, pp 55–85.
- Hall T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95–98.
- Hallam A. (1994). *An outline of Phanerozoic biogeography*. Oxford: Oxford University Press.
- Hausdorf B. (1998). Weighted ancestral area analysis and a solution of the redundant distribution problem. *Syst. Biol.* **47**:445–456.
- Heaney L.R. (1986). Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. *Biol. J. Linn. Soc.* **28**:127–165.
- Heaney L.R. (1991). A synopsis of climatic and vegetational change in Southeast Asia. *Clim. Change* **19**:53–61.
- Heusden van E.C.H. (1992). Flowers of Annonaceae: morphology, classification, and evolution. *Blumea (suppl 7)*:1–218.
- Hill K.C. and Hall R. (2003). Mesozoic-Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. In *Evolution and Dynamics of the Australian Plate*. Edited by: Hillis R.R., Müller R.D.. Sydney: Geological Society of Australia Special Publication 22 and Geological Society of America Special Paper **372**:259–283.
- Ho S.Y.W. (2007). Calibrating molecular estimates of substitution rates and divergence times in birds. *J Avian. Biol.* **38**:409–414.
- Hug L.A. and Roger A.J. (2007). The impact of fossils and taxon sampling on ancient molecular dating analyses. *Mol. Biol. Evol.* **24**:1889–1897.
- Inger R.F. and Voris H.K. (2001). The biogeographical relations of the frogs and snakes of Sundaland. *J. Biogeogr.* **28**:863–891.
- Johnson D.J. (2004). *The geology of Australia*. Cambridge, Cambridge University Press.
- Kürschner W.M., Kvaèek Z., and Dilche D.L. (2008). The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. *Proc. Natl. Acad. Sci. USA* **105**:449–453.
- Leech M.L., Singh S., Jain A.K., Klempener S.L., and Manickavasagam R.M. (2005). The onset of India-Asia continental collision: early, steep subduction required by the timing of UHP metamorphism in the western Himalaya. *Earth Planet. Sci. Letters* **234**:83–97.
- Marshall A.G. (1983). Bats, flowers and fruit: evolutionary relationships in the Old World. *Biol. J. Linn. Soc.* **20**:115–135.

- Mohr B.A.R. and Bernardes-de-Oliveira M.E.C. (2004). *Endressinia brasiliiana*, a magnolielean angiosperm from the Lower Cretaceous Crato Formation (Brazil). *Int. J. Plant Sci.* **165**:1121–1133.
- Mols J.B. and Keßler P.J.A. (2000). The genus *Monocarpia* (Annonaceae) in Borneo including a new species *Monocarpia borneensis*. *Bot. Jahrb. Syst.* **122**:233–240.
- Mols J.B., Gravendeel B., Chatrou L.W., Pirie M.D., Bygrave P.C., Chase M.W., and Keßler P.J.A. (2004). Identifying clades in Asian Annonaceae: monophyletic genera in the polyphyletic *Miliuseae*. *Amer. J. Bot.* **91**:590–600.
- Morley R.J. (2002). Tertiary vegetational history of Southeast Asia, with emphasis on the biogeographical relationships with Australia. In *Bridging Wallace's Line: The Environmental and Cultural History and Dynamics of the SE-Asian-Australian Region*. Edited by: Kershaw P., David B., Tapper N., Penny D., Brown J.. Reiskirchen: Catena Verlag49–60.
- Morley R.J. (2003). Interplate dispersal paths for megathermal angiosperms. *Persp. Pl. Ecol. Evol. Syst.* **6**:5–20.
- Morley R.J. (2007). Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In *Tropical Rainforest Responses to Climatic Changes*. Edited by: Flenley J.R., Bush M.B. Berlin: Springer1–31.
- Morley R.J. and Dick C.W. (2003). Missing fossils, molecular clocks, and the origin of the Melastomataceae. *Amer. J. Bot.* **90**:1638–1644.
- Morley R.J. and Flenley J.R. (1987). Late Cainozoic vegetational and environmental changes in the Malay Archipelago. In *Biogeographical Evolution of the Malay Archipelago*. Vol. 4. In T.C. Whitmore (Ed.). Oxford Monographs on Biogeography, UK, Oxford, Oxford University Press, pp 50–59.
- Muellner A.N., Pannell C.M., Coleman A., and Chase M.W. (2008). The origin and evolution of Indomalesian, Australasian and Pacific island biotas: insights from Aglaiaeae (Meliaceae, Sapindales). *J. Biogeogr.* **35**:1769–1789.
- Nylander JAA: [<http://www.abc.se/~nylander/>] MrModeltest ver. 2.3. 2008.
- Pennington R.T. and Dick C.W. (B2004). The role of immigrants in the assembly of the South American rainforest tree flora. *Phil. Trans. Roy. Soc. Lond.* **359**:1611–1622.
- Pirie M.D., Chatrou L.W., Erkens R.H.J., Maas J.W., Niet T., Mols J.B., and Richardson J.E. (2005). Phylogeny reconstruction and molecular dating in four Neotropical genera of Annonaceae: the effect of taxon sampling in age estimations. In Plant species-level systematics: new perspectives on pattern and process. *Regnum Vegetabile*. Vol. **143**. In F.T. Bakker, L.W. Chatrou, B. Gravendeel, P.B. Pelser (Eds.). Germany, Koenigstein: Koeltz, pp. 150–174.
- Pirie M.D., Chatrou L.W., Mols J.B., Erkens R.H.J., and Oosterhof J. (2006). “Andean-centred” genera in the short-branch clade of Annonaceae: Testing biogeographical hypotheses using phylogeny reconstruction and molecular dating. *J. Biogeogr.* **33**:31–46.
- Poulin E., Palma A.T., and Féral J.P. (2002). Evolutionary versus ecological success in Antarctic benthic invertebrates. *Trends. Ecol. Evol.* **17**:218–222.
- Pubellier M., Ali J., and Monnier C. (2003). Cenozoic Plate interaction of the Australia and Philippine Sea Plate: “hit-and-run” tectonics. *Tectonophysics* **363**:181–199.
- Rambaut A. (1996). Se-Al. [<http://tree.bio.ed.ac.uk/software/seal/>] Sequence Alignment Editor, ver. 2.0a11.
- Rambaut A. (2008). [<http://tree.bio.ed.ac.uk/software/figtree/>] FigTree ver. 1.2.
- Rambaut A. and Drummond A.J. (2007): [<http://tree.bio.ed.ac.uk/software/tracer/>] Tracer ver. 1.4.
- Rawlings L.H., Barker D., and Donnellan S.C. (2004). Phylogenetic relationships of the Australo-Papuan Liassis pythons (*Reptilia*: Macrostomata), based on mitochondrial DNA. *Austral J. Zool.* **52**:215–227.
- Richardson J.E., Chatrou L.W., Mols J.B., Erkens R.H.J., and Pirie M.D. (2004): Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Phil. Trans. Roy. Soc. Lond., B* **359**:1495–1508.

- Ronquist F. (1996). DIVA (DIspersal Vicariance Analysis) HOMEPAGE. [<http://www.ebc.uu.se/systzoo/research/diva/diva.html>] DIVA ver. 1.2.
- Rowe K.C., Reno M.L., Richmond D.M., Adkins R.M., and Steppan S.J. (2008). Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old endemic rodents (*Muroidea*: Murinae). *Mol. Phylogenet. Evol.* **47**:84–101.
- Rowley D.B. (1996). Age of initiation of collision between India and Asia: a review of stratigraphic data. *Earth Planet Sci. Lett.* **145**:1–13.
- Sanderson M.J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**:1218–1231.
- Sanderson M.J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**:101–109.
- Sanderson M.J. (2004). [<http://loco.biosci.arizona.edu/r8s/>] r8s version 1.71.
- Scharaschkin T. and Doyle J.A. (2005). Phylogeny and historical biogeography of *Anaxagorea* (Annonaceae) using morphology and non-coding chloroplast sequence data. *Syst. Bot.* **30**:712–735.
- Scotese C.R., Gahagan L.M., and Larson R.L. (1988). Plate tectonic reconstructions of the Cretaceous and Cenozoic ocean basins. *Tectonophysics* **155**:27–48.
- Shevenell A.E., Kennett J.P., and Lea D.W. (2004). Middle Miocene southern ocean cooling and Antarctic cryosphere expansion. *Science* **305**:1766–1770.
- Su Y.C.F., Chaowasku T., and Saunders R.M.K. An extended phylogeny of *Pseuduvaria* (Annonaceae) with descriptions of three new species and a reassessment of the generic status of Oreomitra. *Syst. Bot.*, in press.
- Su Y.C.F. and Saunders R.M.K. (2006). Monograph of *Pseuduvaria* (Annonaceae). *Syst. Bot. Monogr.* **79**:1–204.
- Su Y.C.F., Smith G.J.D., and Saunders R.M.K. (2008). Phylogeny of the basal angiosperm genus *Pseuduvaria* (Annonaceae) inferred from five chloroplast DNA regions, with interpretation of morphological character evolution. *Mol. Phylogenet. Evol.* **48**:188–206.
- Takahashi M., Friis E.M., Uesugi K., Suzuki Y., and Crane P.R. (2008). Floral evidence of Annonaceae from the Late Cretaceous of Japan. *Int. J. Plant. Sci.* **169**:908–917.
- Takhtajan A. (1986). *Floristic regions of the world*. California, University of California Press.
- Thorne J.L. and Kishino H. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* **51**:689–702.
- Van Tol J. and Gassmann D. (2007). Zoogeography of freshwater invertebrates of Southeast Asia, with special reference to Odonata. In *Bioogeography, Time and Place—Distributions, Barriers and Islands*. In Renema W. (Ed.). Dordrecht, Springer:45–91.
- van Ufford A.Q. and Cloos M. (2005). Cenozoic tectonics of New Guinea. *Amer. Assoc. Petrol. Geol. Bull.* **89**:119–140.
- Voris H.K. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *J. Biogeogr.* **27**:1153–1167.
- Wang R.J. and Saunders R.M.K. (2006a). The genus *Cyathocalyx* (Annonaceae) in the Philippines. *Syst. Bot.* **31**:285–297.
- Wang R.J. and Saunders R.M.K. (2006b). A synopsis of *Cyathocalyx* species (Annonaceae) in Peninsular Malaysia, Sumatra, and Borneo, with descriptions of two new species. *Bot. J. Linn. Soc.* **152**:513–532.
- Ward J.V., Doyle J.A., and Hotton C.L. (1989). Probable granular magnoliid angiosperm pollen from the Early Cretaceous. *Pollen Spores* **31**:101–120.
- Westra L.Y.T. (C 1985). Studies in Annonaceae. IV. A taxonomic revision of *Tetrameranthus* R. E. Fries. *Proc Kon Ned Akad Wetensch* **88**:449–482.
- Wiffin T. (2002). Plant Biogeography of the SE AsianAustralian Region. In Bridging Wallace's Line: the environmental and cultural history and dynamics of the SE AsianAustralian Region. *Advances in Geoeiology*. Volume 34. In P. Kershaw, B. David, N. Tapper, D. Penny, J. Brown

- (Eds.). Germany Reiskirchen, Catena Verlag, pp. 61–82.
- Wikström N., Savolainen V., and Chase M.W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proc Roy Soc Lond B* **268**:2211–2220.
- Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comp. Appl. Biosci.* **13**:555–556.
- Young K.R., Ulloa C.U., Luteyn J.L., and Knapp S. (2002). Plant evolution and endemism in Andean South America: an introduction. *Bot. Rev.* **68**:4–21.
- Zachos J., Pagani M., Sloan L., Thomas E., and Billups K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**:686–693.
- Zhu B.D., Kidd W.S.F., Rowley D.B., Currie B.S., and Shaffique N. (2004). Age of initiation of the India-Asia collision in the East-Central Himalaya. *J. Geol.* **113**:265–285.
- ## 19
- Axelrod D.I. (1954). Poleward migration of early angiosperm floras. *Science* **130**:203–207.
- Bailey I.W. (1949). Origin of angiosperms: Need for a broadened outlook. *J Arnold Arbor* **30**:64–70.
- Collinson M.E. (1980). Recent and Tertiary seeds of the Nymphaeaceae sensu lato with a revision of *Brasenia ovula* (Brong.) Reid et Chandler. *Ann. Bot.* **46**:603–632.
- Corner E.J.H. (1976). *The Seeds of Dicotyledons*, Vol. 1, 2. Cambridge, Cambridge University Press.
- Crane P.R. and Lidgard S. (1989). Angiosperm diversification and paleolatitudinal gradients in Cretaceous floristic diversity. *Science* **246**:675–678.
- Cronquist A. (1988). *The Evolution and Classification of Flowering Plants*, 2nd edition. New York, New York Botanical Garden.
- Dettmann M.E. (1994). Cretaceous vegetation: The microfossil record. *History of the Australian Vegetation: Cretaceous to Recent*. In R.S. Hill (Ed.). Cambridge, Cambridge University Press, pp. 143–170.
- Dettmann M.E. and Jarzen D.M. (1990). The Antarctic/Australian rift valley: Late Cretaceous cradle of northeastern Australian relicts? *Rev. Palaeobot. Palynol.* **65**:131–144.
- Doweld A.B. (1996). The systematic relevance of fruit and seed anatomy and morphology of *Akania* (Akaniaceae). *Bot. J. Linn. Soc.* **120**:379–389.
- Doweld A.B. (2001). The systematic relevance of fruit and seed structure in *Bersama* and *Melianthus* (Melianthaceae). *Plant Syst. Evol.* **227**:75–103.
- Doyle J.A. (1984). Evolutionary, geographic, and ecological aspects of the rise of angiosperms. *Proc. 27th Int. Geol. Congr., Paleontol.* **2**:23–33.
- Doyle J.A. and Endress P.K. (2000). Morphological phylogenetic analysis of basal angiosperms: Comparison and combination with molecular data. *Int. J. Plant Sci.* **161**(Suppl 6):S121–153.
- Endress P.K. and Iggersheim A. (1997). Gynoecium diversity and systematics of the Laurales. *Bot. J. Linn. Soc.* **125**:93–168.
- Endress P.K. and Sampson F.B. (1983). Floral structure and relationships of the Trimeniaceae (Laurales). *J. Arnold Arbor.* **64**:447–473.
- Floyd S.K. and Friedman W.F. (2000). Evolution of endosperm developmental patterns among basal flowering plants. *Int. J. Plant Sci.* **161**(Suppl 6):S57–81.
- Friedman W.E. (2006). Embryological evidence for developmental lability during early angiosperm evolution. *Nature* **441**:337–340.
- Friis E.M., Doyle J.A., Endress P.K., and Leng Q. (2003). *Archaeofructus*-angiosperm precursor or specialized early angiosperm? *Trends Plant Sci.* **8**:369–373.
- Friis E.M., Pedersen K.R., and Crane P.R. (1999). Early angiosperm diversification: The diversity of pollen associated with angiosperm reproductive structures in early Cretaceous floras from Portugal. *Ann. Mo. Bot. Gard.* **86**:259–296.
- Friis E.M., Pedersen K.R., and Crane P.R. (2000). Reproductive structure and organization of basal angiosperms from the Early Cretaceous (Barremian or Aptian) of Western Portugal. *Int J Plant Sci* **161**(Suppl 6):S169–182.

- Friis E.M., Pedersen K.R., and Crane P.R. (2001). Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **410**:357–360.
- Friis E.M., Pedersen K.R., and Crane P.R. (2006). Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**:251–293.
- Frohlich M.W. (2006). Recent developments regarding the evolutionary origin of flowers. *Adv. Bot. Res.* **44**:63–127.
- Frumin S. and Friis E.M. (1999). Magnoliid reproductive organs from the Cenomanian/Turonian of North Western Kazakhstan: Magnoliaceae and Illiciaceae. *Plant Syst. Evol.* **216**:265–288.
- Hamann U. (1998). Hydatellaceae. *The Families and Genera of Vascular Plants IV. Flowering Plants. Monocotyledons. Alismatanae and Commelinanae (except Gramineae)*. In K. Kubitzki (Ed.). Berlin, Springer, pp. 231–234.
- Herngreen G.F.W. (1973). Palynology of Albian-Cenomanian strata of Borehole 1-QS-1-MA, state of Maranhao, Brazil. *Pollen Spores* **15**:515–555.
- Hughes N.F. and McDougall A.B. (1987). Records of angiospermid pollen entry into the English early cretaceous succession. *Rev. Palaeobot. Palynol.* **50**:255–272.
- Hughes N.F., McDougall A.B., and Chapman J.L. (1991). Exceptional new records of Cretaceous Hauterivian angiospermid pollen from Southern England. *J. Micropalaeontol.* **10**:75–82.
- Jansen R.K., Cai Z., Raubeson L.A., Daniell H., dePamphilis C.W., Leebens-Mack J., Müller K.F., Guisinger-Bellian M., Haberle R.C., Hansen A.K., Chumley T.W., Lee S.B., Peery R., McNeal J.R., Kuehl J.V., and Boore J.L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* **104**:19369–19374.
- Kodama K., Maeda H., Shigeta Y., Kase T., and Takeuchi T. (2002). Integrated biostratigraphy and magnetostratigraphy of the Upper Cretaceous System along the River Naiba in Southern Sakhalin, Russia. *J. Geol. Soc. Japan.* **108**:366–384.
- Kubitzki K., Rohwer J.G., and Bittrich V. (1993). The Families and Genera of Vascular Plants. Flowering Plants. *Dicotyledons*, Vol. II. Berlin, Springer.
- Landrum L.R. and Sharp W.P. (1989). Seed coat characters of some American Myrtinae (Myrtaceae): *Psidium* and related genera. *Syst. Bot.* **14**:370–376.
- Money L.I., Bailey I.W., and Swamy B.G.L. (1950). The morphology and relationships of the Monimiaceae. *J. Arnold Arbor.* **31**:372–404.
- Moore M.J., Bell C.D., Soltis P.S., and Soltis D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci.* **104**:19363–19368.
- Narita A., Yamada T., and Matsumoto M. (2008). Platanoid leaves from Cenomanian to Turonian Mikasa Formation, Northern Japan and their mode of occurrence. *Paleontol Res.* **12**:81–88.
- Prakash N. (1998). An embryological study of *Piptocalyx moorei* (Trimeniaceae). *Plant Form and Function*. In B. Bhatia, A.K. Shukla, H.L. Sharma (Ed.). New Delhi, Angkor Publishers, pp. 207–216.
- Qiu Y.L., Lee J., Bernasconi-Quadroni F., Soltis D.E., Soltis P.S., Zanis M., Zimmer E.A., Chen Z.D., Savolainen V., and Chase M.W. (1999). The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**:404–407.
- Rudall P.J., Sokoloff D.D., Remizowa M.V., Conran J.G., Davis J.I., Macfarlane T.D., and Stevenson D.W. (2007). Morphology of Hydatellaceae, an anomalous aquatic family recently recognized as an early-divergent angiosperm lineage. *Amer. J. Bot.* **94**:1073–1092.
- Saarela J.M., Rai H.S., Doyle J.A., Endress P.K., Mathews S., Marchant A.D., Briggs B.G., and Graham S.W. (2007). Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* **446**:312–315.
- Saiki K. (1997). *Frenelopsis pombetsuensis*: A new cheirolepidiaceous conifer from the Lower Cretaceous (Albian) of Hokkaido, Japan. *Paleontol Res.* **1**:126–131.
- Sampson F.B. (2007). Variation and similarities in pollen features in some basal angiosperms,

- with some taxonomic implications. *Plant Syst. Evol.* **263**:59–75.
- Schneider H., Schuettpelz E., Pryer K.M., Cranfill R., Magallón S., and Lupia R. (2004). Ferns diversified in the shadow of angiosperms. *Nature* **428**:553–557.
- Sun G., Dilcher D.L., Zheng S., and Zhou Z. (1998). In search of the first flower: A Jurassic angiosperm, *Archaeofructus*, from northeast China. *Science* **282**:1692–1695.
- Sun G., Ji, Q., Dilcher D.L., Zheng S., Nixon K.C., and Wang X. (2002). Archaefractaceae, a new basal angiosperm family. *Science* **296**:899–904.
- Takahashi K. and Suzuki M. (2003). Dicotyledonous fossil wood flora and early evolution of wood characters in the Cretaceous of Hokkaido, Japan. *IWA J.* **24**:269–309.
- Takahashi M., Takai K., and Saiki K. (1995). Ephedroid fossil pollen from the Lower Cretaceous (Upper Albian) of Hokkaido, Japan. *J. Plant Res.* **108**:11–15.
- Takashima R., Kawabe F., Nishi H., Moriya K., Wani R., and Ando H. (2004). Geology and stratigraphy of forearc basin sediments in Hokkaido, Japan: Cretaceous environmental events on the north-west Pacific margin. *Cret. Res.* **25**:365–390.
- Tobe H., Jaffré T., and Raven P.H. (2000) Embryology of *Amborella* (Amborellaceae): descriptions and polarity of character states. *J. Plant Res.* **113**:271–260.
- Tobe H., Yasuda S., and Oginuma K. (1992). Seed coat anatomy, karyomorphology, and relationships of *Simmondsia* (Simmondsiaceae). *Bot. Mag. Tokyo* **105**:529–538.
- Upchurch G.R. and Dilcher D.L. (1990). Cenomanian angiosperm leaf megafossils, Dakota Formation, Rose Creek Locality, Jefferson County, southeastern Nebraska. *U. S. Geol. Surv. Bull.* **155**.
- Walker J.W. and Doyle J.A. (1994). Ultrastructure and relationships of mid-Cretaceous polyporate and triporate pollen from Northern Gondwana. In *Ultrastructure of Fossil Spores and Pollen*. In K. Kurmann and J.A. Doyle (Eds.). London, Royal Botanic Gardens Kew pp. 161–172.
- Wanger W.L. and Lorence D.H. (1999). A revision of *Trimenia Seem.* (Trimeniaceae) in the Marquesas Islands with description of a new species, *Trimenia nukuhivensis*. *Adansonia* **21**:225–230.
- Yamada T., Imaichi R., and Kato M. (2001). Developmental morphology of ovules and seeds of Nymphaeales. *Amer. J. Bot.* **88**:963–974.
- Yamada T., Imaichi R., Prakash N., and Kato M. (2003). Developmental morphology of ovules and seeds of Austrobaileyales. *Aust. J. Bot.* **51**:555–564.

Research Progress in Botany

Vascular Plants and Paleobotany

The chapters within this book provide the reader with a fascinating perspective on the evolutionary history of plants, which has bearing upon the evolution of life in general. Some plants have remained remarkably unchanged throughout Earth's history; early ferns, for example, had developed by the Mississippian Period and conifers by the Pennsylvanian. Some plants of prehistory are the same ones around today and are thus living fossils, such as *Ginkgo biloba* and *Sciadopitys verticillata*. Other plants have changed radically, or have gone extinct entirely. The research described within this volume is important to the reconstruction of ancient ecological systems and climate, and it provides a fundamental basis to the study of green plant development and evolution. What is more, it offers the reader a unique window on the history of early terrestrial life.

About the Editors

Dr. Philip Stewart has a PhD in horticulture with a focus on the genetics of flowering in strawberries. He has worked in association with Cornell University's Grapevine Breeding Program, the Department of Horticulture at the University of Arkansas in Fayetteville, and the Horticultural Sciences Program at the University of Florida in Gainesville. He has contributed to multiple publications, including the International Journal of Fruit Science, Horticultural Science, Plant Science, and BMC Plant Biology. He has served as a member on the U.S. Rosaceae Genetics and Breeding Executive Committee, the North American Strawberry Growers' Association, and the Small Fruit Crop Germplasm Committee. Dr. Stewart is one of the inventors of the patented strawberry plant named DrisStrawSeven, and he currently works with the NCRA, State Agricultural Experiment Station Directors.

Professor Sabine Globig received her BA in 1972 at the American University School of International Service and her MS in horticulture and plant physiology in 1988 at Rutgers University. Presently, she is Professor of Biology at Hazard Community and Technical College in the Appalachian Mountains of Eastern Kentucky, where she specializes in human anatomy and physiology and plant sciences. She has also worked as an Adjunct Instructor of Biology at Union County College in New Jersey and at Rutgers University, and as a certified high school biology teacher. While at Rutgers, she worked as a plant physiology researcher at their AgBiotech Center and held the same position for DNA Plant Technologies Corp. She has given presentations at XXII International Conference on Horticultural Science, UC Davis, CA, 1987; and 1997 ISHS International Symposium on Artificial Lighting in Horticulture, Noordwijkerhout, Netherlands. She has also been included in several Who's Who entries.

Related Titles of Interest

- *Photosynthesis: Genetic, Environmental and Evolutionary Aspects*
- *Phytopathology in Plants*
- *Plant Physiology*
- *Reproductive Physiology in Plants*



Apple Academic Press

www.appleacademicpress.com

ISBN 978-1-926692-98-2



9 781926 692982