

# Plant Genome

## Biodiversity and Evolution

Volume 1, Part E  
Phanerogams - Angiosperm

# **PLANT GENOME: BIODIVERSITY AND EVOLUTION**

Volume 1, Part E  
**Phanerogams - Angiosperm**

# **Plant Genome**

## Biodiversity and Evolution

*Series Editors*

**A.K. Sharma and Archana Sharma**

Department of Botany

University of Calcutta, Kolkata

India

*Advisory Board*

---

**J. Dolezel**, Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovska, Czech Republic

**K. Fukui**, Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan

**R.N. Jones**, Institute of Biological Sciences, Ceredigion, Scotland, UK

**G.S. Khush**, 416 Cabrillo Avenue, Davis, California 95616, USA

**Ingo Schubert**, Institut fur Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany

**Canio G. Vosa**, Scienze Botanische, Pisa, Italy

---

Volume 1, Part A : **Phanerogams**

1, Part B : **Phanerogams**

1, Part C : **Phanerogams (Angiosperm-Dicotyledons)**

1, Part D : **Phanerogams (Gymnosperm) and (Angiosperm-Monocotyledons)**

1, Part E : **Phanerogams - Angiosperm**

2, Part A : **Lower Groups**

2, Part B : **Lower Groups**

# **Plant Genome**

## Biodiversity and Evolution

Volume 1, Part E  
**Phanerogams - Angiosperm**

*Editors*

A.K. SHARMA and A. SHARMA



**Science Publishers**

Enfield (NH)

Jersey

Plymouth

*CIP data will be provided on request.*

Science Publishers

234 May Street

Post Office Box 699

Enfield, New Hampshire 03748

United States of America

**www.scipub.net**

General enquiries : *info@scipub.net*

Editorial enquiries : *editor@scipub.net*

Sales enquiries : *sales@scipub.net*

Published by Science Publishers, Enfield, NH, USA

An imprint of Edenbridge Ltd., British Channel Islands

Printed in India

© 2008 reserved

ISBN 978-1-57808-507-1

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the publisher, in writing. The exception to this is when a reasonable part of the text is quoted for purpose of book review, abstracting etc.

This book is sold subject to the condition that it shall not, by way of trade or otherwise be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

## Preface to the Series “Plant Genome”

---

The term *genome*, the basic gene complement of an individual, is almost synonymous with the chromosome complement of both nucleus and organelles. Refinements in cellular, genetic and molecular methods in recent years have opened up unexplored avenues in genome research. The modern tools of gene and genome analyses, coupled with analysis of finer segments of gene sequences in chromosomes utilizing molecular hybridization, are now applied on a wider scale in different groups of plants, ranging from algae to angiosperms. This synergistic approach has made the study of biodiversity highly fascinating, permitting a deep insight into the molecular basis of genetic diversity. Simultaneous to the enrichment of fundamentals in systematics and phylogeny, the plant system, because of its inherent flexibility, has permitted genetic engineering and horizontal transfer of genes with immense importance in agriculture, horticulture and medicine.

Despite the fact that the data on plant genomics with its impact on the assessment of biodiversity and evolution show a logarithmic increase, a comprehensive series on the aspect covering all groups of plant kingdom is sadly lacking. In view of this lacuna, the present series on Plant Genomics: Biodiversity and Evolution has been planned. It aims to cover, in successive volumes, *comprehensive reviews, concepts and discussions on the results of genome analysis and their impact on systematics, taxonomy, phylogeny and evolution of all plant groups*. We have not gone out of our way to seek original articles, but in course of reviews and discussions, research articles, if any, are welcome.

February, 2003

A.K. Sharma  
Archana Sharma  
Series Editors



## Preface to this Volume

---

This last volume of this series contains chapters of general genomics covering certain basic aspects which have an impact on all groups on the one hand, and, specific crop based genomics on the other.

The discussion on repeated sequences DNA not only covers its diverse structure, but also their location including introns and transposons, as well as functional implications if any. The quantitative modulation of different sequence families, has also been recorded as also the correlation of genomes with environmental parameters. The change of genome size and organization, through alteration of redundant sequences affecting development dynamics, has been shown to confer adaptive value in selection. The relationship between repetitive DNA and quantitative genetics of continuous characters has been discussed.

In the next chapter on phylogeny and evolution of rice genome, the repetitive elements and genome duplication have been proved to be the principal factors of genome evolution in rice. This comprehensive chapter deals with the characteristics of rice genome, nature of sequences, structural peculiarities of functional segments, transposable elements in evolution, and relation of different gene families and allies in the light of synteny. The affinity between different rice genomes and its allies—its understanding will be of much use to the breeders to tap the existing genetic pool for improvement. A model presentation of rice genome in the light of *Arabidopsis* genome is of special interest.

A very detailed tree genome study has been presented in two species of oak – *Q. petraea* and *Q. robur* both being compatible but maintaining species identity. The molecular organization of the oak genome, chromosome number along with the fluorochrome and “G” banding, as well as karyotype analysis, have been utilized to deduce speciation and phylogeny. Finally, a genetic map of *Q. robur* has been presented. This

study in woody plant would pave the way for researchers on arboreal species.

In the other tree genus, the phylogeny, evolution, and diversification of the chestnut (*Castanea*) has been analysed with the aid of chloroplast genome and monophyletic status of the genus is confirmed. After origin in E. Asia, it is evidenced to have undergone intercontinental divergence and later between European and North American continent in the Tertiary period. This genus is a very good example to show the use of organellar DNA in tracing ancestry through chloroplast sequences in plants.

Another chapter on arboreals is on *Azaleas* which covers the entire taxonomy and phylogeny of the genus *Rhododendron* *sensu lato* of Ericaceae and the status of this taxon as indicated through molecular, genetic and cytological data. Of the eight subgenera, discussions have been centered around *Azalea*, comprising SW *Pentanthera* and *Tsutsui* subgenera, and *Rhododendron* comprising *Rhododendron* and *Hymenanthes* subgenera. The interspecific hybridization, pollination biology and ovule culture have been covered as also intergeneric hybridization between *Rhododendron* and allies. Molecular studies for analyzing relationship are also evidenced in the study of phylogeny through MatK sequences and further characterization by RFLP and PCR methods. The use of AFLP in characterizing genetic resources, and forming gene pool obtained from different sources have been highlighted. The aid of functional markers for assessing diversity and mapping of qualitative traits, as well as a comparative evaluation of RFLP, STMS and EST and other markers along with their transferability in *Rhododendron* and *Azalea* have been discussed. The chapter not only provides with the present state of art on molecular studies in assessing phylogeny and affinity across species and generic barriers but simultaneously reiterates the need for search of stress responsive genes and development of new cultivars.

In the study on phylogeny, diversity and evolution of Australian genus *Eucalyptus*, fossil data has been utilized in conjunction with molecular studies. Molecular phylogeny vindicates its origin in Late Cretaceous being more dominant in Cenozoic with increased aridity. Molecular data indicate high level of diversity at the population level maintained through mixed mating. Allopatric hybridization and distant pollen dispersal had been significant features in evolution. The variation of functional genes correlated with ecological trait specific markers, have enabled an understanding of genetic variations and ecological

adaptations. Problems associated with sequence variation in genes in intragenic recombination in relation to Cinnamoyl CoA reductase has been highlighted. The “Eucalyptus Global Initiative” program has been proposed, as the results can be utilized to mitigate negative impact of environment on several species facing extinction.

In Hamamelidaceae too, the comparative importance of both chloroplast and nuclear DNA sequences in the molecular phylogenetics has been precisely brought out. The use of both coding and non coding region of chloroplast reveals Hamamelidaceae as a distinct clade including *Eubucklandia* and *Rhodoleia* which were often excluded earlier. The nuclear rDNA ITS phylogeny too indicates that the last two genera form a weakly supported clade in Altingiaceae. Within Hamamelidaceae, the combined data of chloroplast and nuclear sequences show a distinct phylogenetic tree with six clades. However, there are still problems regarding interclade relationships. Contradictory evidences have been obtained in relation to relationship of *Shaniodendron* with *Parrotia* or with *Distyliopsis* from nuclear and organelle data. Though the suggestion of chloroplast capture by *Shaniodendron* has been suggested, the solution may still await a full genome analysis.

Of the legumes, extensive cytogenetic and molecular data have been presented in *Arachis* (peanut) and its allies. Gene mapping by FISH, molecular markers and amphidiploid synthesis have been taken to indicate *A. duranensis* and *A. hypogea* are the probable genome donors of peanut genome and *A. monticola* is the wild tetraploid progenitor. The deep seated number  $n = 9$  is derived and polyphyletic status of the section “rhizomatous” has been suggested.

The other legume dealt in this volume is lentil (*Lens*), where SSR repeats have aided in the characterization of composite collections. Molecular hybridization involved genomic DNA and synthetic oligonucleotides. Dinucleotide repeats which are most abundant have been used for screening and their difference from other legumes was evident. Markers were designed for genetic mapping and SSRs were mapped in linkage map.

With regard to evolution of sex chromosome, *Silene* has been shown to provide with a model system. It enables an analysis of gynodioecy along with dioecism, sex chromosomal DNA sequences and the basis of the sex bias in plants. In origin of sex chromosomes, the strategy of recombination arrest and accumulation of sexually opposite genes in the

initial phase has been well studied in this genus. The gene order and divergence of homologous X and Y linked genes, with progress in evolution are indicated providing concrete evidence of diversification in evolution with sex chromosome formation.

There are two chapters on ornamentals namely *Phlox* and *Dahlia*. The article on *Phlox* of Polemoniaceae has brought out once more the importance of non-coding DNA sequences of chloroplasts in the study of phylogeny and its reconstruction. Such a study in this genus was needed in view of the fact that a comparison of restriction site data of chloroplast DNA and ITS spacer sequences of nuclear ribosomes revealed divergence evidently with its impact on status and phylogeny. The new sequence data of several non-coding chloroplast DNA have aided an understanding of taxonomy of this genus to a great extent. The results have confirmed the status of *Microsteris* as allied to *Phlox*, questioned the monophyletic state of several groups, and suggested that annual taxa of Central Texas form a well supported clade.

The final chapter in the volume is on *Dahlia* of Asteraceae, the most successful family of dicotyledons. Taxonomically, the genus is subdivided into several sections though molecular phylogeny based on ITS, ETS sequences as well as GISH indicating the presence of three distinct clades with 16, 17 and 18 chromosomes, 17 being ancestral and the other numbers originating through aneuploidy. Pattern of recombination has been deduced from pachytene analysis, synaptonemal complex and secondary association. Polyploidy has arisen independently in different groups. The GISH analysis reveals pairing within and between chromosomes derived from parental species, indicating palaeopolyploid nature of the genus, the base number, possibly being eight.

This volume with its coverage of herbaceous agricultural and arboreal species and the sequence analysis of both nuclear and organelle genome, along with other genetic criteria will be of immense value in the study of plant genome—its structure, diversification, phylogeny and speciation.

Arun Kumar Sharma

Archana Sharma

# **Contents**

---

Preface to the Series	v
Preface to this Volume	vii
List of Contributors	xiii
<b>1. Repetitive DNA and Plant Evolution</b>	<b>1</b>
<i>M. Ceccarelli, V. Sarri and P.G. Cionini</i>	
<b>2. Evolution and Phylogenetic Relationship of the Rice Genome</b>	<b>15</b>
<i>Akhilesh K. Tyagi, Jitendra P. Khurana, Paramjit Khurana, Shubha Vij, Mukesh Jain and V. Ravi</i>	
<b>3. Genome Organization and Evolution in Genus <i>Quercus</i> (Fagaceae): Special Attention to Two European White Oaks <i>Quercus petraea</i> (Matt.) Liebl. and <i>Q. robur</i> L.</b>	<b>43</b>
<i>Vlatka Zoldos Pecnik</i>	
<b>4. Biodiversity and Evolution of <i>Castanea</i></b>	<b>79</b>
<i>Fenny Dane and Ping Lang</i>	
<b>5. Cytogenetic and Molecular Evidences on the Evolutionary Relationships Among <i>Arachis</i> Species</b>	<b>101</b>
<i>Graciela I. Lavia, Aveliano Fernández and Jose Guillermo Seijo</i>	
<b>6. Biodiversity, Genetic Enhancement and Molecular Approaches in Lentil</b>	<b>135</b>
<i>Michael Baum, Aladdin Hamwieh, Bonnie Furman, Ashtusosh Sarker and William Erskine</i>	

7. Phylogenetic Relationships of the Genus <i>Phlox</i> (Polemoniaceae): Comparing and Combining Data Sets from the Chloroplast Genome	153
<i>Carolyn J. Ferguson, Shannon D. Fehlberg, Kristen A. Ford and Suzanne C. Strakosh</i>	
8. Genus <i>Silene</i> (Caryophyllaceae) – Evolutionary Diversification and Sex Chromosome Formation	173
<i>Jitka Zluvova, Bohuslav Janousek, Roman Hobza, Jaroslav Mracek, Alexander Widmer and Boris Vyskot</i>	
9. Molecular Phylogenetics of Hamamelidaceae: Evidence from DNA Sequences of Nuclear and Chloroplast Genomes	227
<i>Jianhua Li</i>	
10. The Potential of Molecular Analysis and Interspecific Hybridization for Azalea Phylogenetic Research	251
<i>Jan De Riek, Valentina Scariot, Tom Eeckhaut, Ellen De Keyser, Nobuo Kobayashi and Takashi Handa</i>	
11. Phylogeny, Diversity and Evolution of Eucalypts	303
<i>Margaret Byrne</i>	
12. Dahlia: Cytogenetics and Evolution	347
<i>Brian Grant Murray</i>	
Author Index	365
Detailed Contents	393

## **List of Contributors**

---

### **Akhilesh K. Tyagi**

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi – 110021, India. Tel: + 91-11-24113216, Fax: + 91-11-24119430, E-mail: akhilesh@genomeindia.org

### **Aladdin Hamwieh**

International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box, Aleppo, Syria

### **Alexander Widmer**

ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology, Zurich, Switzerland

### **Ashutosh Sarker**

International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box, Aleppo, Syria

### **Aveliano Fernández**

Instituto de Botánica del Nordeste, Corrientes, Argentina. Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

### **Bohuslav Janousek**

Laboratory of Plant Developmental Genetics, Institute of Biophysics, Brno, Czech Republic

### **Bonnie Furman**

International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box, Aleppo, Syria

**Boris Vyskot**

Laboratory of Plant Developmental Genetics, Institute of Biophysics,  
612 56 Brno, Czech Republic. Tel: +420 541 517 204,  
Fax: +420 541 240 500, E-mail: vyskot@ibp.cz

**Brian Grant Murray**

School of Biological Sciences, The University of Auckland, Private  
Bag 92019, Auckland 1001, New Zealand. Tel: +64 9 373 7599 x  
88334, Fax: +64 9 373 7416, E-mail: b.murray@auckland.ac.nz

**Carolyn J. Ferguson**

Herbarium and Division of Biology, 116 Ackert Hall, Kansas State  
University, Manhattan, Kansas 66506-4901, USA. Tel: 785-532-  
3166, Fax: 785-532-6653, E-mail: ferg@ksu.edu

**Ellen De Keyser**

Institute for Agricultural and Fisheries Research (ILVO), Plant Unit,  
Caritasstraat, Melle, Belgium

**Fenny Dane**

Department of Horticulture, Auburn University, AL 36849, USA

**Graciela I. Lavia**

Instituto de Botánica del Nordeste, Corrientes, Argentina. Facultad  
de Ciencias Agrarias, Universidad Nacional del Nordeste, Corrientes,  
Argentina

**J. Guillermo Seijo**

Instituto de Botánica del Nordeste, C.C. 209, (3400) Corrientes,  
Argentina. Tel: + 54-3783-422006, Fax: + 54-3783-427131, E-mail:  
lavia@agr.unne.edu.ar, seijo@agr.unne.edu.ar

Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad  
Nacional del Nordeste, Corrientes, Argentina

**Jan De Riek**

Institute for Agricultural and Fisheries Research (ILVO), Plant Unit,  
Caritasstraat 21, 9090 Melle, Belgium. Tel + 32 9 272 28 81, Fax +  
32 9 272 29 01, E-mail: jan.deriek@ilvo.vlaanderen.be

**Jaroslav Mracek**

Laboratory of Plant Developmental Genetics, Institute of Biophysics,  
Brno, Czech Republic

**Jianhua Li**

Arnold Arboretum of Harvard University, Cambridge, MA, USA.  
College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang  
310058, China.

Arnold Arboretum and Harvard University Herbaria, 22 Divinity  
Avenue, Cambridge, MA 02138, USA. Tele: +617-496-6429; Fax:  
+617-495-9484; E-mail: jli@oeb.harvard.edu

**Jitendra P. Khurana**

Interdisciplinary Centre for Plant Genomics and Department of Plant  
Molecular Biology, University of Delhi South Campus, New Delhi –  
110 021, India

**Jitka Zluvova**

Laboratory of Plant Developmental Genetics, Institute of Biophysics,  
Brno, Czech Republic

**Kristen A. Ford**

Herbarium and Division of Biology, Kansas State University,  
Manhattan, Kansas USA

**M. Ceccarelli**

Dipartimento di Biologia Cellulare e Ambientale, Sezione di Biologia  
Cellulare e Molecolare, Università di Perugia, Perugia, Italy

**Margaret Byrne**

Science Division, Department of Environment and Conservation,  
Locked Bag 104, Bentley Delivery Centre, WA 6983,  
Australia. Tel: + 618 9334 0503, Fax + 618 93340327,  
E-mail: margaret.byrne@dec.wa.gov.au

**Michael Baum**

International Center for Agricultural Research in the Dry Areas  
(ICARDA), P.O. Box 5466, Aleppo, Syria. Tel: +963-21-2213433,  
Fax: +963-21-2213490, E-mail:m.baum@cgiar.org

**Mukesh Jain**

Interdisciplinary Centre for Plant Genomics and Department of Plant  
Molecular Biology, University of Delhi South Campus, New Delhi –  
110 021, India

**Nobuo Kobayashi**

Laboratory of Plant Breeding, Faculty of Life and Environmental Science, Shimane University, Nishikawatsu, Matsue, Shimane 690-8504, Japan

**P.G. Cionini**

Dipartimento di Biologia Cellulare e Ambientale della Università, Sezione di Biologia Cellulare e Molecolare, Via Elce di Sotto, I-06123 Perugia, Italy. Tel: +39 075 585 5703, Fax: +39 075 585 5759, E-mail: cionini@unipg.it

**Paramjit Khurana**

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi – 110 021, India

**Ping Lang**

Population Medicine and Diagnostic Services, College of Veterinary Medicine, Ithaca, Cornell University, Ithaca, NY.  
E-mail: danefen@auburn.edu

**Roman Hobza**

Laboratory of Plant Developmental Genetics, Institute of Biophysics, Brno, Czech Republic. ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology, Zurich, Switzerland

**Shannon D. Fehlberg**

Herbarium and Division of Biology, Kansas State University, Manhattan, Kansas USA

**Shubha Vij**

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi – 110 021, India

**Suzanne C. Strakosh**

Herbarium and Division of Biology, Kansas State University, Manhattan, Kansas USA

**Takashi Handa**

Graduate School of Life and Environmental Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan

**Tom Eeckhaut**

Institute for Agricultural and Fisheries Research (ILVO), Plant Unit,  
Caritasstraat, Melle, Belgium

**V. Ravi**

Interdisciplinary Centre for Plant Genomics and Department of Plant  
Molecular Biology, University of Delhi South Campus, New Delhi –  
110 021, India

**V. Sarri**

Dipartimento di Biologia Cellulare e Ambientale, Sezione di Biologia  
Cellulare e Molecolare, Università di Perugia, Perugia, Italy

**Valentina Scariot**

Department of Agronomy, Forest and Land Management –  
University of Turin, via Leonardo da Vinci 44, 10095 Grugliasco  
(TO), Italy

**Vlatka Zoldos Pecnik**

Department for Molecular Biology, Faculty of Science, University of  
Zagreb, Horvatovac 102a, 10000 Zagreb, Croatia. Tel: +385 1  
4606266, Fax: +385 1 4606286, E-mail: vzoldos@zg.biol.pmf.hr

**William Erskine**

International Center for Agricultural Research in the Dry Areas  
(ICARDA), P.O. Box, Aleppo, Syria



## Repetitive DNA and Plant Evolution

MARILENA CECCARELLI, VANIA SARRI and  
PIER GIORGIO CIONINI

Dipartimento di Biologia Cellulare e Ambientale, Sezione di Biologia Cellulare e Molecolare, Università di Perugia, Perugia, Italy

### ABSTRACT

Variations in the size and organization of the nuclear genome due to quantitative modulation of DNA repeats belonging to different sequence families have been observed within several angiosperms. These genomic changes are correlated with variations in the environmental parameters and/or given phenotypic characteristics of the plants. Intraspecific alterations in the nuclear DNA may take place on an evolutionary timescale, as the result of selection processes after the occurrence of uncontrolled events in the genome. However, experimental results obtained in some species have shown that these alterations may also be due to direct responses of plant genomes to environmental stimuli, that can occur under plant-level control within a short developmental period of a single generation. The possibility of altering genome size and organization through variations in redundancy of repeated DNA sequences and consequently modifying developmental dynamics and the phenotype is an evolutionary trait that allows plant species to adapt to different environmental conditions as well as to the variability of conditions in a given environment. The link between repetitive DNA and quantitative genetic determination of continuous characters may suggest models of evolutionary change that extend beyond the conventional view of evolution by allelic substitution.

**Key Words:** Plant species, repetitive DNA, environmental adaptation, phenotype, evolution

*Address for correspondence:* Prof. Pier Giorgio Cionini, Dipartimento di Biologia Cellulare e Ambientale della Università, Sezione di Biologia Cellulare e Molecolare, Via Elce di Sotto, I-06123 Perugia, Italy. Tel: +39 075 585 5703, Fax: +39 075 585 5759, E-mail: cionini@unipg.it

## INTRODUCTION

Repetitive DNA constitutes the largest fraction of nuclear DNA in most eukaryotes and is particularly abundant in plant genomes. Repeated DNA sequences can be divided into two major groups according to their spatial organization in the nucleus. The first group is made up of families of repeating units which are tandemly arranged and form long arrays preferentially associated with constitutive heterochromatin around the centromeres and/or at chromosome ends; however, they may also be found at intercalary sites. Ribosomal cistrons are a well-characterized class of tandem repeats which are localized in the nucleolus-organizing regions of the chromosomes. Tandem repeated sequences include mini- and microsatellites, depending on the length of the repeating unit, which is about 100 bp in minisatellites and 2-6 bp in microsatellites. These tandemly arranged repeats are widely distributed in eukaryotic genomes and, as a rule, exhibit high polymorphism. The other group of repeated sequences is made up of DNA elements with a dispersed organization in the genome, and includes mobile sequences such as transposons and retrotransposons. The latter undergo replication by the use of a RNA intermediate and reverse transcriptase. Retrotransposons are classified as retrovirus-like retroelements, when they contain long terminal repeats (LTRs) and non-LTR retroelements. Non-LTR retroelements display a highly variable structure and can be subdivided according to their size into long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Transposable DNA elements are widely dispersed in the genomes, but may be preferentially localized at or absent from given chromosome regions [37, 56].

A trait of repetitive DNA domains is their fluidity [26], i.e., the ability of sequences to rapidly change their redundancy – and, when transposable, shift their position along the genome – thereby altering the amount and organization of nuclear DNA. The ways in which quantitative modulations of repetitive DNA can occur have been suggested. Strand slippage during DNA replication and unequal intrachromatid exchange are major causes of redundancy variation of tandemly repeated DNA elements (microsatellite, minisatellite, or satellite sequences). Extrachromosomal rolling circle replication, followed by reinsertion into the genome is another process which may be involved in the initial formation of satellite repeats and their evolution. Replicative transposition is a well-known process by which mobile DNA elements

can vary their redundancy and position in the genome [24]. All these mechanisms are responsible for the increase of DNA content in the nucleus. Unequal homologous recombination and especially illegitimate recombination [30] generate deletions that can decrease the sequence copy number and genome size [6]. These genomic changes are relatively easy to detect, and alterations of the overall amount of nuclear DNA indicate that quantitative modulations in repetitive DNA domains have occurred.

It is now evident that redundancy modulations in repetitive DNA domains and the resulting changes in genome size and organization may not be restricted to species divergence. Even if intraspecific plasticity of the nuclear genome is not yet entirely accepted [36], data in the literature show that changes in the redundancy of repeated DNA sequences may differentiate not only strictly related species, but also populations within a species and even individual plants within one and the same population. These genomic changes may obviously take place on an evolutionary timescale, as the result of selection processes after the occurrence of uncontrolled events in the genome. However, experimental results obtained in some species such as *Linum usitatissimum* [27, 29], *Helianthus annuus* [14, 18, 54] and *Festuca arundinacea* [22, 23] have shown that such changes may be also due to direct responses of plant genomes to environmental stimuli, that can occur within the short developmental period of a single generation.

While it is known that given repeated sequences, such as those coding for ribosomal RNA or located at the chromosome ends, carry out a specific activity, the functional role of most repetitive DNA is controversial. However, it has been observed in several angiosperm species that intraspecific changes in the size and organization of the nuclear genome due to quantitative modulations of repeated DNA sequences are significantly correlated to variations in the environmental parameters and/or given phenotypic characteristics of the plants. These correlations suggest that repetitive DNA produces quantitative genetic variation that plays a role in species evolution. Data that were obtained in studies on *Linum usitatissimum*, *Vicia faba*, *Helianthus annuus*, *Festuca arundinacea*, *Dasyperym villosum* and other plant species, are summarized and discussed hereunder.

### ***Linum usitatissimum***

The case with flax is now classic. When plants carrying a given phenotype are germinated and grown under different temperatures and/or in unbalanced mineral nutrients, altered phenotypes (genotrophs) are produced [29]. These phenotypes are heritable and the alterations in the offspring remain stable for many generations. The genome size of the genotrophs can vary 1.16-fold. Variations in the redundancy of repeated DNA sequences which span the whole range from highly repetitive to low copy number and comprise 5S ribosomal RNA genes and insertion-like sequences, has been shown to account for quantitative changes in the nuclear DNA [25, 26, 57]. Since the plastic variety of flax (Stormont Cirrus) is predominantly self-fertilizing and nearly all the seeds planted can contribute to the next generation, the observed genomic changes cannot be explained by any form of selection from a genotypically heterogeneous population of plants [27]. These changes are direct responses of the genome to environmental stimuli.

### ***Vicia faba***

Local populations of this species scattered throughout the Mediterranean Basin differ 1.35-fold in the nuclear DNA content of meristematic cells [21]. Molecular analyses have shown that redundancy modulations of repeated DNA sequences account for the observed differences in genome size, and are positively correlated with the copy number of sequences that make up a family of tandemly arranged repeats, as well as with the redundancy of interspersed sequences which are mobile elements or their remnants [35, 39]. Genome size is negatively correlated with cell proliferation in the meristems, and positively with cell enlargement in differentiated tissues. These two developmental factors and the interactions between them can affect the dynamics of plant development. Plants with large genomes grow faster during the early developmental stages, since the genome size is positively correlated with both the germination power of the seeds and the growth rate of the epicotyl. However, the final size of these plants is smaller than that of plants with smaller genomes; the amount of DNA is negatively correlated with the height of the main stem as well as with the fresh weight of plants at anthesis [46].

### ***Helianthus annuus***

Variations in the nuclear DNA content in *H. annuus* have been reported for some time [12, 45, 47, 49]. Genome size of cultivars and lines of this species may differ up to 1.58-fold. Significant variations in the amount of nuclear DNA have even been observed between seedlings obtained from plants belonging to selfed lines. These genomic changes occur during reproduction in spite of selfing and homozygosity. Embryos that develop in different portions of the flowering head have different genome sizes, decreasing from the periphery to the middle of the head. No differences occur in the genome size of the anthers and pistils or meiotic cells that develop in different portions of the head. Alterations in the amount of nuclear DNA appear in the embryos at the heart stage; the DNA content then remains unchanged with subsequent further embryo development and seed germination. Since reproduction events occur in the head in centripetal succession, genomic rearrangements are probably due to differences in the microenvironments within the different portions of the head where embryogenesis takes place [13, 14].

Variations in the redundancy of repeated sequences, ribosomal DNA included, accompany the changes in genome size. These variations are accompanied by changes in DNA methylation and are complex, involving different fractions of repetitive DNA, most of which are amplified in the DNA of embryos that develop at the periphery of the head. However, other fractions are amplified in the DNA of embryos developing in the middle of the head [18, 48].

This continuous genotypic variation within the species can affect developmental dynamics and plant phenotype. Indeed, cells proliferate more slowly in the meristems of plants with a larger genome. The mitotic cycle takes 1 hour 30 minutes longer in plants that have higher DNA content compared to plants with lower DNA content. Moreover, significantly positive correlations have been found between genome size and cell enlargement in differentiated tissues as well as the flowering time and interval [48].

### ***Festuca arundinacea***

Findings obtained in fescue provide another example of intraspecific changes of the genotype in which repetitive DNA is involved. The changes are related to environmental variation and can affect plant phenotype. Highly significant differences in genome size (up to 32.3%)

have been found between natural Italian populations of hexaploid *F. arundinacea*. The genome size is directly correlated with the mean annual temperature, as well as the mean temperature during the coldest month at the stations, and inversely correlated with the latitudes [19]. These genomic changes are due to a direct response of the nuclear DNA to temperature during seed germination. The DNA content increases significantly as the seed germination temperature increases. In contrast, the DNA content does not undergo quantitative changes when the temperature is altered during other developmental stages [22]. Redundancy variations in different families of DNA repeats are involved in the above-described response of the *F. arundinacea* genome. As observed in *H. annuus*, most of these sequences are amplified in the nuclear DNA of seedlings raised at a higher temperature (30°C), while other repeated sequences are more frequent in the nuclear DNA of seedlings raised at a lower temperature (10°C). These fluid domains are probably made up of mobile elements, as suggested by the length, redundancy level, and scattered distribution in the genome of the varying sequences [23]. As in the other species already considered, the genomic changes that occur in *F. arundinacea* in response to temperature can affect cell proliferation in the meristems. The mitotic cycle time is 3 hours shorter in populations with smaller genomes compared to populations with larger genomes. The genome size is negatively correlated with developmental events which take place during the winter and early spring, such as seed germination power and early root and leaf growth, and positively correlated with the height of the highest culm and other quantitative characteristics of plant organs that develop later in the warmer season, as well as with the flowering time [20].

### ***Dasypyrum villosum***

Substantial quantitative alterations in the nuclear DNA occur in this species between and within natural populations. When the most variant values are considered, there is a 17.6% difference between the mean genome size of the populations and a 66.2% difference between the genome size of individual plants within a given population. A highly significant correlation exists between the genome size and the altitude at which the plants grow, and the differences in DNA content between individual plants are greater in populations from mountain sites. As in *H. annuus*, the plasticity of the *D. villosum* genome is particularly apparent during reproduction. Small genomes are somewhat 'dominant' in

determining the DNA amounts in the progeny. A large range of genome sizes was found in the half-sib progeny of plants having relatively large genomes, while in the half-sib progeny of plants having small genomes, the genome size of the individual plants was less divergent and similar to that of the mother plant [8].

The involvement of repetitive DNA in these genomic changes has been proved by the redundancy modulations of subtelomeric repeats, which are particularly apparent during seed germination and are related to variations in nuclear DNA content [34]. The mean genome size of each population is negatively correlated with the mean length and width of the leaves. The genome size of individual plants is positively correlated with the weight of the seed from which each of them originated and with their flowering interval [8].

## OTHER PLANT SPECIES

In several other flowering species, intraspecific changes in genome size due to quantitative modulations of repeated DNA sequences are correlated with modifications of the cellular and organismal phenotype and/or environmental parameters.

In *Pisum sativum*, where unscheduled DNA synthesis (extrachromosomal DNA) [61, 62] has earlier been reported, the 1C DNA content varies between lines from 3.93 to 5.07 pg and specific chromatin fractions are involved in these variations [15]. The genome size of pea lines is negatively correlated with the ploidy level achieved by chromosome endoreduplication in cotyledon parenchyma cells, but positively correlated with the growth of root and stems in the first developmental stages after seed germination [16, 17]. In *Zea mays*, lines from the higher latitudes of North America have significantly less DNA and heterochromatin than those from lower latitudes; correlations between altitude and the amounts of both heterochromatin (C bands) and DNA have also been reported [51, 55]. The genome size of populations of wild barley (*Hordeum spontaneum*) from ecologically and geographically different sites differs significantly and is positively correlated with January temperatures at the stations [58]. BARE-1 retrotransposons, which constitute a major, dispersed, active component of *Hordeum* genomes and whose role has been proved in their evolution [63], vary in redundancy in the nuclear DNA of plants from stations having microclimates that differ markedly in solar irradiation and aridity

[38]. In *Silene latifolia*, variation of DNA content due to redundancy modulation of repeated DNA sequences has been observed. The DNA amount is correlated with changes in a phenotypic character, i.e., flower size, whose ecological significance has been well documented [41]. It has been confirmed that the negative correlation between genome size and the size of the flower is a correlated response to artificial selection for flower size [42]. This negative correlation was also observed for leaf size in closely related species of *Silene* [43]. The amounts of nuclear DNA have been found to vary up to 0.21-fold in seven populations of *Trifolium repens* from New Zealand. As in *F. arundinacea*, the genome size is correlated with temperature; plants having larger genomes grow well at low temperatures, while plants with smaller genomes grow well at high temperatures [9]. Two further examples of intraspecific variations of genome size and organization which are correlated with environmental parameters are *Anthirrinum majus*, in which temperature affects transposon mobility and hence their redundancy and location in the genome [10], and *Microseris douglasii*, in which changes in the amount of nuclear DNA between populations is correlated to rainfall [53].

## CONCLUSION

The data cited above show a connection between repetitive DNA and the genetic determination of continuous characters. Quantitative alterations in DNA domains made up of repeated DNA sequences allow plants to adapt to changes in the environment. These genomic changes clearly constitute one way of species evolution, even if, at the level of current knowledge, this cannot be considered to be a general way. There are several widely distributed plant species (e.g., *Allium cepa*) which show a surprisingly high degree of stability in genome size [5].

How do changes in repetitive DNA domains influence phenotype? The role of genomic changes of this kind may simply be to produce alterations in the nucleotype, i.e., in the DNA mass and volume [3]. The results of comparative studies carried out in many angiosperms suggest that genome size, per se, can influence several cellular characteristics and activities, such as nuclear volume [1], cell volume [11, 52], mitotic cycle time [31, 59, 60], and the duration of meiosis [2]. In turn, the cellular phenotype can influence developmental dynamics and the organismal phenotype [4]. This may also be true at the intraspecific level. According to the nucleotype concept, the nature of the DNA sequences that

undergo quantitative variation is of little consequence, since the phenotype is affected by changes in nuclear DNA, independent of the genome-encoded informational content [3]. This concept does not exclude a 'selfish' nature of the varying sequences nor the correlated view that any possible advantages upon the host genome due to their self-replication are indirect [28, 50].

However, the specificity of the responses to environmental stimuli observed in some of the above species suggests that these genomic changes are under plant-level control and are not independent of the genome as a whole. Two examples should be recalled. In *F. arundinacea*, extrasythesis of both guanine+cytosine- and adenine+thymine-enriched DNA sequences occurs when seeds are germinated at 30°C, while a different repeat is amplified when germinated at 10°C [22, 23]. In *H. annuus*, different families of repeated sequences are amplified in the DNA of seedlings depending on whether the seeds from which they have originated developed in the middle of the head or at its periphery [18]. Moreover, differential sequence redundancy in the DNA from the two groups of seedlings, as well as in the DNA from different flax genotrophs [27], involves ribosomal DNA [48]. These specific responses suggest that redundancy modulations of repeated DNA sequences may have a role in gene regulation, perhaps through effects on the genome involving epigenetic processes [66]. It has been hypothesized that accumulation of noncoding, repetitive DNA motifs (tandemly arranged sequences, in particular) can influence the pattern of interaction between DNA and proteins [66]. Another hypothesis is that repetitive DNA can affect metabolic rates, giving plants a chance to adapt to certain habitats [64]. Moreover, several authors have suggested that nuclear architecture and the level of chromatin organization are directly related to gene expression [37]. Other authors have demonstrated the role played by tandemly repeated sequences in forming particular DNA structures involved in the regulation of transcriptional events [32] and how they affect chromatin despiralization [33]. Likewise, transposable DNA elements, which move within the genome, can alter the chromosome structure; e.g., may cause inversions and/or translocations [65]. DNA portions, such as exons and promoters, can also move into already existing sequences, which creates new gene functions [37].

The possibility of altering genome size and organization through redundancy modulations of repeated DNA sequences is an evolutionary trait that allows certain plant species to adapt to different environmental

conditions and to the variability of conditions in a given environment. An example of the former is the quantitative genome plasticity in *F. arundinacea*, where there are optimal genotypes for seed germination and plant development at given temperatures. The ability to optimize the genotype by varying the proportion of repetitive DNA, along with the larger-scale calibration of the genome size in relation to the latitude, as seen in changes in the ploidy levels (tetraploid fescues are found in France and hexaploids prevail in Italy, Spain, and Portugal, while octoploids and decaploids are confined to northern Africa [7]), may help explain the widespread geographical distribution of this species (northern Africa, west-central Asia, most of Europe, Iceland, and the Faeroes and Spitsbergen Islands [40]).

Genome plasticity which continuously produces intrapopulation changes in the genome size and organization in *H. annuus* and *D. villosum* may have a somewhat different adaptive role. In both species, plants with differing amounts of nuclear DNA also differ in given aspects of their development and, remarkably, in the flowering time and interval [8, 48]. Therefore, intrapopulation variability of genome size and organization may be seen as a factor which allows plant populations to withstand the variability of conditions within a given environment. This view is supported by the fact that the differences in genome size between individual plants may be greater in *D. villosum* populations from mountain sites, where the environment is particularly limiting and/or variable.

In summary, the link between repetitive DNA and quantitative genetic determination of continuous characters indicates one way by which several angiosperm species adapt to changes in the environment and suggest models of evolutionary change that extend beyond the conventional view of evolution by allelic substitution [44].

## References

- [1] Baetke KP, Sparrow AM, Nauman CH, Schwemmer SS. The relationships of DNA content to nuclear and chromosome volumes and to radiosensitivity. Proc Natl Acad Sci USA 1967; 58: 533-540.
- [2] Bennett MD. The duration of meiosis. Proc R Soc London B 1971; 178: 259-275.
- [3] Bennett MD. Nuclear DNA content and minimum generation time in herbaceous plants. Proc R Soc London B 1972; 181: 109-135.
- [4] Bennett MD. Variation in genomic form in plants and its ecological implications. New Phytol 1987; 106: 177-200.

- [5] Bennett MD, Bhandol P, Leitch IJ. Nuclear DNA amounts in angiosperms and their modern use – 807 new estimates. Ann Bot 2000; 86: 859-909.
- [6] Bennetzen J, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. Ann Bot 2005; 95: 127-132.
- [7] Borrill M, Tyler B, Loyd-Jones M. Studies in *Festuca*. I. A chromosome atlas of *bovinæ* and *scariosæ*. Cytologia 1972; 36: 1-14.
- [8] Caceres ME, De Pace C, Scarascia-Mugnozza GT, Kotsonis P, Ceccarelli M, Cionini PG. Genome size variations within *Dasyperym villosum*: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. Theor Appl Genet 1998; 96: 559-567.
- [9] Campbell BD, Caradus JR, Hunt CL. Temperature responses and nuclear DNA amounts of seven white clover populations which differ in early spring growth rates. New Zeal J Agric Res 1999; 42: 9-17.
- [10] Carpenter R, Hudson A, Robins T, Almeida J, Martin C, Cohen E. Genetic and molecular analysis of transposable elements in *Anthirrinum majus*. In: Nelson O, ed. Plant transposable elements. New York: Plenum Press, 1988: 69-80.
- [11] Cavalier-Smith T. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the C-value paradox. J Cell Sci 1978; 34: 247-278.
- [12] Cavallini A, Cionini PG. Nuclear DNA content in differentiated tissues of sunflower (*Helianthus annuus* L.). Protoplasma 1986; 130: 91-97.
- [13] Cavallini A, Zolfino C, Cionini G, Cremonini R, Natali L, Sassoli O, Cionini PG. Nuclear DNA changes within *Helianthus annuus* L.: cytophotometric, karyological and biochemical analyses. Theor Appl Genet 1986; 73: 93-100.
- [14] Cavallini A, Zolfino C, Natali L, Cionini G, Cionini PG. Nuclear DNA changes within *Helianthus annuus* L.: origin and control mechanism. Theor Appl Genet 1989; 77: 12-16.
- [15] Cavallini A, Natali L. Nuclear DNA variability within *Pisum sativum* (Leguminosae): cytophotometric analyses. Plant Syst Evol 1990; 173: 179-185.
- [16] Cavallini A, Natali L, Cionini G, Gennai D. Nuclear DNA variability within *Pisum sativum* (Leguminosae): nucleotypic effects on plant growth. Heredity 1993; 70: 561-565.
- [17] Cavallini A, Natali L. Chromosome endoreduplication variability in cultivars of *Pisum sativum* L. and its relationship to basic genome size. Cytobios 1994; 79: 181-188.
- [18] Cavallini A, Natali L, Giordani T, Durante M, Cionini PG. Nuclear DNA changes within *Helianthus annuus* L.: variations in the amount and methylation of repetitive DNA within homozygous progenies. Theor Appl Genet 1996; 92: 285-291.
- [19] Ceccarelli M, Falistocco E, Cionini PG. Variation in genome size and organization within hexaploid *Festuca arundinacea*. Theor Appl Genet 1992; 83: 273-278.
- [20] Ceccarelli M, Minelli S, Falcinelli M, Cionini PG. Genome size and plant development in hexaploid *Festuca arundinacea*. Heredity 1993; 71: 555-560.
- [21] Ceccarelli M, Minelli S, Maggini F, Cionini PG. Genome size variation in *Vicia faba*. Heredity 1995; 74: 180-187.
- [22] Ceccarelli M, Giordani T, Natali L, Cavallini A, Cionini PG. Genome plasticity during seed germination in *Festuca arundinacea*. Theor Appl Genet 1997; 94: 309-315.
- [23] Ceccarelli M, Esposto MC, Roscini C, Sarri V, Frediani M, Gelati MT, Cavallini A, Giordani T, Pellegrino RM, Cionini PG. Genome plasticity in *Festuca arundinacea*:

- direct response to temperature changes by redundancy modulation of interspersed DNA repeats. *Theor Appl Genet* 2002; 104: 901-907.
- [24] Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994; 371: 215-220.
  - [25] Cullis CA, Cleary W. Rapidly varying DNA sequences in flax. *Can J Genet Cytol* 1986; 28: 252-259.
  - [26] Cullis CA. DNA rearrangements in response to environmental stress. *Adv Genet* 1990; 28: 73-97.
  - [27] Cullis CA. The environment as an active generator of adaptive genomic variation. In: Lerner HR, ed. *Plant responses to environmental stresses*. New York, Basel: Marcel Dekker Inc, 1999: 149-160.
  - [28] Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 1980; 284: 601-603.
  - [29] Durrant A. The environmental induction of heritable changes in *Linum*. *Heredity* 1962; 17: 27-61.
  - [30] Ehrlich SD. Illegitimate recombination in bacteria. In: Berg DE, Howe MM, eds. *Mobile DNA*. Washington DC: American Society of Microbiology, 1989: 799-832.
  - [31] Evans GM, Rees H. Mitotic cycles in dicotyledons and monocotyledons. *Nature* 1971; 23: 350-351.
  - [32] Fabregat I, Koch K, Aoki T. Functional pleiotropy of an intramolecular triplex-forming fragment from the 3'-utr of the rat Pigr gene. *Physiol Genom* 2001; 5: 53-65.
  - [33] Fransz P, Soppe W, Schubert I. Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. *Chrom Res* 2003; 11: 227-240.
  - [34] Frediani M, Colonna N, Cremonini R, De Pace C, Delre V, Caccia R, Cionini PG. Redundancy modulation of nuclear DNA sequences in *Dasyperym villosum*. *Theor Appl Genet* 1994; 88: 167-174.
  - [35] Frediani M, Gelati MT, Maggini F, Galasso I, Minelli S, Ceccarelli M, Cionini PG. A family of dispersed repeats in the genome of *Vicia faba*: structure, chromosomal organization, redundancy modulation, and evolution. *Chromosoma* 1999; 108: 317-324.
  - [36] Greilhuber J. Intraspecific variation in genome size: a critical reassessment. *Ann Bot* 1998; 82 (Supplement A): 27-35.
  - [37] Heslop-Harrison JS. Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell* 2000; 12: 617-635.
  - [38] Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci USA* 2000; 97: 6603-6607.
  - [39] Maggini F, D'Ovidio R, Gelati MT, Frediani M, Cremonini R, Ceccarelli M, Minelli S, Cionini PG. *FokI* DNA repeats in the genome of *Vicia faba*: species specificity, structure, redundancy modulation, and nuclear organization. *Genome* 1995; 38: 1255-1261.
  - [40] Markgraf-Dannenberg I, Festuca L. In: Tutin TG, et al., eds. *Flora Europaea Vol 5*. Cambridge, London, New York, New Rochelle, Melbourne, Sidney: Cambridge University Press, 1980: 125-133.
  - [41] Meagher TR, Costich DE. Sexual dimorphism in nuclear DNA content and floral morphology in populations of *Silene latifolia* (Caryophyllaceae). *Am J Bot* 1994; 81: 1198-1204.

- [42] Meagher TR, Costich DE. Nuclear DNA content and floral evolution in *Silene latifolia*. Proc R Soc Lond B 1996; 263: 1455-1460.
- [43] Meagher TR, Costich DE. 'Junk' DNA and long-term phenotypic evolution in *Silene* section *Elisanthe* (Caryophyllaceae). Proc R Soc Lond B (Suppl) 2004; 271: S493-S497.
- [44] Meagher TR, Vassiliadis C. Phenotypic impacts of repetitive DNA in flowering plants. New Phytol 2005; 168: 71-80.
- [45] Michaelson M, Price, HJ. Nuclear DNA content variation within *Helianthus annuus* L. Am J Bot 1985; 72: 877-882.
- [46] Minelli S, Moscariello P, Ceccarelli M, Cionini PG. Nucleotype and phenotype in *Vicia faba*. Heredity 1996; 76: 524-530.
- [47] Nagl W, Capesius I. Molecular and cytological characteristics of nuclear DNA and chromatin for angiosperms systematics: basic data for *Helianthus annuus* (Asteraceae). Plant Syst Evol 1976; 126: 221-237.
- [48] Natali L, Cavallini A, Cionini G, Sassoli O, Cionini PG, Durante M. Nuclear DNA changes in *Helianthus annuus* L.: changes within single progenies and their relationships to plant development. Theor Appl Genet 1993; 85: 506-512.
- [49] Olsewska MJ, Osiecka R. The relationship between 2C DNA content, life cycle type, systematic position, and the dynamics of DNA endoreduplication in parenchyma nuclei during growth and differentiation of roots in some dicotyledonous herbaceous species. Biochem Physiol Pflanz 1983; 178: 581-599.
- [50] Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. Nature 1980; 284: 604-607.
- [51] Porter HC, Rayburn AL. B-chromosome and C-band heterochromatin variations in Arizona maize populations adapted to different altitudes. Genome 1990; 33: 659-662.
- [52] Price H, Sparrow A, Naumann A. Correlation between nuclear volume, cell volume and DNA content in meristematic cells of herbaceous angiosperms. Experientia 1973; 29: 1028-1029.
- [53] Price HJ, Chambers KL, Bachmann K. Geographic and ecological distribution of genomic DNA content variation in *Microseris douglasii* (Asteraceae). Bot Gaz 1981; 142: 415-428.
- [54] Price HJ, Johnston JS. Influence of light on DNA content of *Helianthus annuus* Linnaeus. Proc Natl Acad Sci USA 1996; 93: 11264-11267.
- [55] Rayburn AL, Auger JA. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. Theor Appl Genet 1990; 79: 470-474.
- [56] Santini S, Cavallini A, Natali L, Minelli S, Maggini F, Cionini PG. Ty1/copia and Ty3/gypsy-like DNA sequences in *Helianthus* species. Chromosoma 2002; 111: 192-200.
- [57] Schneeberger R, Cullis CA. Specific DNA alterations associated with the environmental induction of heritable changes in flax. Genetics 1991; 128: 619-630.
- [58] Turpeinen T, Kulmala J, Nevo E. Genome size variation in *Hordeum spontaneum* populations. Genome 1999; 42: 1094-1099.
- [59] Van't Hof J, Sparrow A. A relationship between DNA content, nuclear volume, and minimum mitotic cycle time. Proc Natl Acad Sci USA 1963; 49: 897-902.
- [60] Van't Hof J. Relationships between mitotic cycle duration, S period duration and the average rate of DNA synthesis in the root meristem of several plants. Exp Cell Res 1965; 39: 48-58.
- [61] Van't Hof J, Bjerknes CA. Cells of pea (*Pisum sativum*) that differentiate from G2 phase have extrachromosomal DNA. Mol Cell Biol 1982; 2: 339-345.

- [62] Van't Hof J, Bjerknies CA. Excision and replication of extrachromosomal DNA of pea (*Pisum sativum*). *Mol Cell Biol* 1983; 3: 172-181.
- [63] Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 1999; 11: 1769-1784.
- [64] Vinogradov AE. Buffering: a possible passive-homeostasis role for redundant DNA. *J Theor Biol* 1998; 193: 197-199.
- [65] Wessler SR, Bureau TE, White SE. LTR-retrotransposons and MITEs – important players in the evolution of plant genomes. *Curr Opin Genet Dev* 1995; 5: 814-821.
- [66] Zuckerkandl E. Junk DNA and sectorial gene repression. *Gene* 1997; 205: 323-343.

## Evolution and Phylogenetic Relationship of the Rice Genome

**AKHILESH K. TYAGI, JITENDRA P. KHURANA, PARAMJIT KHURANA, SHUBHA VIJ, MUKESH JAIN and V. RAVI**

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi – 110 021, India

### ABSTRACT

Cereal crops such as rice, wheat, maize and barley comprise the primary source of global nutrition. Rice has emerged as the model for other cereal genomes chiefly due to its small size. Comparison of the sequence information of rice with the available information from other plant genomes has shown that polyploidization and repetitive elements have been the two main factors influencing plant genome evolution. Comparative studies also point to the phylogenetic relationship of flowering plants, which is reflected in the form of synteny in closely related species. This relationship has already proved a valuable resource for genetic markers and identification of agriculturally important genes from other plant genomes as well. Genome information will not only assist in identification and isolation of orthologs, it will also serve as a rich source for evolutionary studies. Moreover, relationship between genomes of rice will help breeders tap the rich gene pool and existing genetic variability within the wild varieties of rice.

**Key Words:** Evolution, gene duplication, phylogeny, rice, synteny

**Abbreviations:** LSE= lineage-specific expansion, LTR= long-terminal repeat, MULE= Mutator-like element, mya= million years ago, TE= transposable element.

*Address for Correspondence:* Prof. Akhilesh K. Tyagi, Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi – 110021, India. Tel: + 91-11-24113216, Fax: + 91-11-24119430, E-mail: akhilesh@genomeindia.org

## INTRODUCTION

Rice is the staple food for the largest number of people on earth. Consumed by nearly half the world's population and covering approximately 9% of the total arable land, it represents the most important economic activity on Earth [57]. Rice belongs to the genus *Oryza* (family Poaceae, formerly Gramineae) comprising a total of 23 species grouped into four complexes (Table 1). Of these, two species (*O. sativa* and *O. glaberrima*) were domesticated and together provide more than one-fifth of the calories consumed by humans. The former species, grown worldwide, is commonly referred to as Asian rice while the latter, grown in parts of West Africa, is known as African rice. Asian rice (*O. sativa*) is believed to have originated from common wild rice *O. rufipogon* in the foothills of the Himalayas with the *indica* and *japonica* rice originating on the Indian and Chinese side, respectively [53]. Owing to the green revolution, production of rice has increased tremendously during the last four decades. An estimated increase of 130% in rice production during the period from 1966-2000 coincided with a 90% increase in population of the densely-populated low income countries [44]. However, rice production still needs a dramatic increase and by the year 2030, 40% increase in production is required to meet the needs of an ever-increasing population and the growing demand. To face this mammoth task, sound knowledge of the rice genome and a combination of breeding strategies as well as molecular biology are required [98, 105]. A high degree of variation is seen in the grass genomes with estimates for rice being 430 Mb in size while wheat has an estimated genome size of 16,000 Mb [3]. Possessing a genome 40 times smaller than wheat, rice was the obvious choice for genome sequencing. Four independent versions of the rice genome were produced by private and public funded groups, highlighting the importance of this crop [4, 32, 36, 122, 123] and hoping that the availability of the genome sequence might enable scientists to develop better rice varieties both in terms of yield and tolerance to various stresses. Moreover, the genome sequence would help in extrapolation to other cereals with much larger genomes owing to conservation in gene order – a phenomenon known as synteny [28, 64]. The 23 species of rice comprising of 10 genome types represent an enormous gene pool for the genetic improvement of rice cultivars. To use this wild rice germplasm, knowledge of the phylogenetic relationships between these genomes is essential. Phylogenetic analyses using nuclear

**Table 1.** Complexes, taxa, genomic characters and distribution of the genus *Oryza* (modified from Rice Almanac 3<sup>rd</sup> ed.)

Complex/Taxon	Genome group	Estimated size (Mb)*	Distribution
<b><i>O. sativa</i> complex</b>			
<i>O. glaberrima</i> (cultigen)	AA	809	West Africa
<i>O. barthii</i>	AA		Africa
<i>O. longistaminata</i>	AA		Africa
<i>O. sativa</i> (cultigen)	AA	389 <sup>\$</sup> ( <i>japonica</i> ), 466 <sup>#</sup> ( <i>indica</i> )	Worldwide
<i>O. nivara</i>	AA	760	Tropical Asia
<i>O. rufipogon</i>	AA	760	Tropical Asia
<i>O. meridionalis</i>	AA		Tropical Australia
<i>O. glumaepatula</i>	AA		South America
<b><i>O. officinalis</i> complex</b>			
<i>O. officinalis</i>	CC	1,201	Tropical Asia to Papua New Guinea
<i>O. eichingeri</i>	CC		East and West Africa
<i>O. rhizomatis</i>	CC		Sri Lanka
<i>O. minuta</i>	BBCC	1,691	Philippines, Papua New Guinea
<i>O. punctata</i>	BB, BBCC	539	Africa
<i>O. latifolia</i>	CCDD		Central and South America
<i>O. alta</i>	CCDD	1,000	Central and South America
<i>O. grandiglumis</i>	CCDD		South America
<i>O. australiensis</i>	EE	1,054	Australia
<b><i>O. meyeriana</i> complex</b>			
<i>O. granulata</i>	GG	907	South and Southeast Asia
<i>O. meyeriana</i>	GG		Southeast Asia
<b><i>O. ridleyi</i> complex</b>			
<i>O. longiglumis</i>	HHJJ		Papua New Guinea
<i>O. ridleyi</i>	HHJJ	1,568	Southeast Asia
<b>Unknown/Miscellaneous</b>			
<i>O. schlechteri</i>	HHKK	1,568	Papua New Guinea
<i>O. brachyantha</i>	FF	343	Africa

\*Oryza Map Alignment Project (OMAP; [114])

<sup>\$</sup>[36]

<sup>#</sup>[123]

*Adh1*, *Adh2* and chloroplastic *matK* genes revealed that the AA genome lineage (containing cultivated rice) is a recently diverged, rapidly radiating and a well-adapted lineage [30]. The EE genome species is closest to the DD genome progenitor which gave rise to the CCDD genome. The relatedness between the genome types is important in terms of ‘crossability’. The wild species have been categorized as primary, secondary and tertiary gene pools according to the ‘crossability’ with *Oryza sativa* [43]. Species of the AA genome are regarded as the primary gene pool and can be easily crossed with *O. sativa*. The secondary gene pool comprises of the genomes BB to EE, and the remaining (FF to KK) comprise the tertiary gene pool. The tertiary gene pool contributes less than one-third of the species diversity, but nearly half of the genomic diversity. Although these genomes are distantly related to the cultivated varieties, they have tremendous potential to provide novel genes.

The availability of the finished sequence of rice, along with a large amount of sequence information available from other plant species, offers tremendous potential for evolutionary analysis. Such studies show that several rearrangements in genomes have taken place, after the cereals diverged from their common ancestor, which have given a specific identity to each cereal genome. However, in each case, microcolinearity would have to be examined in order to exploit its use in the identification of orthologs [25]. Despite the signatures of colinearity even in distantly related plant species such as rice and *Arabidopsis*, the frequency of rearrangements is very high and its extent depends on the evolutionary relatedness of the plant species. In spite of these differences, map-based cloning has greatly benefited from the conservation of gene order in cereal genomes. In fact, sequence comparisons have also revealed the major role played by polyploidy and repeat elements in shaping the cereal genomes. The impact of this knowledge towards a better understanding of rice genome evolution, including creation and divergence of gene families, duplication, events related to individual gene amplification and elimination events, along with understanding the relationship of rice with other plant genomes are discussed here.

## **FEATURES OF THE RICE GENOME**

The map-based rice genome sequence reveals a genome size of 389 Mb comprising of an estimated 37,544 non-transposable element related genes [36]. Through manual curation of annotations involving evidence

of transcription for gene identification, the number of genes in rice is now estimated to be 32,000 [93]. Before the entire rice genome was completed, finished sequences for three chromosomes 1, 4 and 10 had been published [23, 80, 94]. International Rice Genome Sequencing Project (IRGSP) published their effort in the August 2005 issue of *Nature*. This finished map-based sequence will serve as a standard, not only for rice, but also for interpretation of genomes of other related species. More than 3,000 BAC/PAC clones were sequenced to approximately 10X coverage to achieve an error rate of less than one in 10,000 bp. The finished sequence still has 62 physical gaps remaining, including 9 centromeric and 17 telomeric gaps, adding up to around 18.1 Mb. The total genome size was obtained by adding the non-overlapping sequence with the estimated size of gaps resulting in a size of 388.8 Mb. The individual chromosomes varied in size ranging from 43.3 Mb (chromosome 1; [80]) to 22.4 Mb (chromosome 10; [94]). The genome sequence information represents approximately 95.3% of the genome and 98.9% of the euchromatin, and has been further extended in the form of Build 4.0 release (based on the data-freeze on January 25, 2005; <http://www.rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html>) representing 382 Mb finished sequence. Almost 35% of the genome comprises of repeat elements. Chromosomes 8 and 12 have the highest (38 and 38.3%, respectively), while chromosomes 1, 2 and 3 have the least (31, 29.8 and 29%, respectively) transposon content [36].

## Centromeres and Telomeres

Eukaryotic centromeres typically comprise of long stretches of repeat elements and are thus very difficult to sequence [17]. However, the finished rice genome sequence includes three sequenced centromeres (chromosomes 4, 5 and 8). The rice centromeres typically contain 155–165 bp CentO repetitive satellite DNA along with several retrotransposons, out of which Ty3/gypsy-like retrotransposons form the largest family [36,117,124]. About 200 ORFs (open reading frames) were predicted for chromosome 8 centromere, though centromeres are generally considered to be transcriptionally silent zones. Although only ~20% of these showed similarity to known proteins or full-length cDNAs, transcriptional activity could be experimentally confirmed at least for some [65]. Recently, sequencing of chromosome 3 centromere has also revealed extensive transcriptional activity, suggesting that centromeres could have evolved from gene containing regions [120].

Telomeres that are integral to chromosomal maintenance are also comprised of large stretches of repetitive DNA. Even though the finished level sequence is available for rice, not all telomeric and subtelomeric regions are present in the complete sequence. Only chromosome 2S telomeric sequence was included in the IRGSP pseudomolecule [36]. Towards this end, seven subtelomeric rice clones were sequenced and analyzed. The sequence had 25-100 copies of the telomere array sequence (CCCTAAA)<sub>n</sub> at one side and unique sequences on the other end. In addition, two chromosome-specific telomere-associated tandem repeats (TATR) could be identified on chromosome 7 (TATR7) as well as chromosome 10 (TATR10) [121]. In a recent study, fiber-fluorescent *in situ* hybridization and terminal restriction-fragment assay revealed that the average length of rice telomeres ranged from 5.1 to 10.8 kb. Further, telomeric as well as subtelomeric regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S were sequenced and characterized [63] in continuation of the IRGSP effort and revealed the existence of conserved telomere repeats. From a 500 kb region adjacent to the telomere repeats in each of the seven chromosomal ends, a total of 598 genes could be predicted out of which 199 were hypothetical, being merely computer predictions without any evidence of transcription or protein level homology. More than 50% of the predicted genes could be related to various biological processes such as metabolism, cell communication, regulation of gene expression and pathogenesis [63].

## **Organelle Genome Insertions**

Search for organelle genome insertions in the rice genome revealed 421 chloroplast (nuptDNA) and 1,191 mitochondrial (numtDNA) insertions amounting to 0.24 and 0.19% of the genome, respectively. The maximum amount of mitochondrial DNA insertions (~1%) was in chromosome 12, while chromosome 7 had the least (~0.03%). Chromosome 10 had the maximum amount of chloroplast DNA insertions (~0.8%) and chromosome 11 had the least (0.007%) [36].

## **Gene Density**

Depending on the number of genes predicted in different studies, the average gene density could vary from one gene per 9.9 to 11.5 kb [36,93]. Chromosome 3 had the highest gene density of one gene per 8.7 kb, while chromosome 12 had the lowest gene density of one gene per 11.6 kb. The average GC content was 43.6% with 54.2% GC in case of exons and

38.3% in case of introns [36]. At least 61% genes showed a significant match with either a rice EST (Expressed sequence Tag) or a full-length cDNA. Almost 90% of the *Arabidopsis* proteins had a homolog in rice, while only 71% of the rice proteins had an *Arabidopsis* homolog. When compared with other model organisms like *Drosophila*, human, *C. elegans*, yeast, *Synechocystis* and *E. coli*, the percentage of rice proteins with homologs was 38.1, 40.8, 36.5, 30.2, 17.6 and 10.2 %, respectively.

## Duplications

Analysis of duplications in the completed rice genome revealed three main classes of duplications – segmental, tandem and background. Almost 60% of the genome is duplicated [36] and duplications are spread over all rice chromosomes. However, chromosomes 11 and 12 share the largest and most recent duplication block. Chromosomal duplication has been analyzed at various stages of draft sequence [33, 71, 123] and genetic and physical mapping studies had earlier revealed the presence of duplicated regions at the distal ends of the short arms of chromosome 11 and 12 [116]. The latest study, however, was at the level of gene models estimated from the finished rice genome sequence. Comparison of chromosome 11 and 12 gene models revealed that the first 3 Mb region of both the chromosomes shared the maximum proportion of duplicated genes (~67 and 71% for chromosome 11 and 12, respectively). This block also seems to be recent in origin and the duplication event would have probably occurred about 7.7 mya [95], in contrast to the previous estimates of 25 and 21 mya [32, 123]. The differences in the estimates can be attributed to the sequence quality of finished sequence as compared to the whole-genome shotgun sequence [36]. The rest of the duplications have also been attributed to a whole genome duplication, which may have occurred 55-70 mya [71, 123].

## Comparison of Rice Genomes

Rice has the added advantage that genome sequence information is available for two of its subspecies, *japonica* and *indica*. *Indica* is the major subspecies of rice grown and consumed in the Asian subcontinent with an estimated genome size of 466.3 Mb and a total of 49,088 genes. Using EST data the number of genes comes to 40,216 [123]. The *japonica* genome is estimated to be 389 Mb with ~32,000 genes as against the previous estimate of 37,544 predicted genes [36, 93]. A comparison of *indica* and *japonica* genomes has shown that the major difference lies in

the intergenic regions [34, 88]. A comparison of the two genomes showed that only 72% of the sequences could be aligned which could probably explain the difference in the size of the two genomes [123]. The availability of the two genomes has also facilitated evolutionary studies. In fact, although they diverged only ~0.44 mya, significant differences are evident even between them [56].

A large-scale project entitled the 'Oryza Map Alignment Project' has been undertaken with the objective of making BAC/STC based physical maps of one cultivated and 11 wild rice species. These would be then aligned to the finished rice genome sequence. The 11 wild rice species represent six diploid genomes (AA, BB, CC, EE, FF and GG) and four tetraploid genomes (BBCC, CCDD, HHKK and HHJJ) and thus represent a broad variation in gene pool [1,114]. In order to gain insight into the extent of insertions and deletions in *O. sativa* vis-à-vis *O. niwara*, 1 Mb finished rice chromosome 3 was used to identify *O. niwara* BAC end pairs that mapped to sequence greater than 200 kb in length. The comparison revealed that either an insertion of ~200 kb had occurred in *japonica* or a deletion of the same size had taken place in the *O. niwara* genome. Similarly, orthologous region corresponding to the heading date locus, *Hd1*, were compared between these two genomes. The analysis revealed that the *japonica* *Hd1* region was about 80% larger than the corresponding *O. niwara* region [114]. In fact, detailed comparison reveals that the *O. sativa* ssp. *japonica* rice chromosome 3 is ~20% larger than *O. niwara* chromosome 3, thus supporting the concept of rapid genome expansion in cultivated rice [1]. Relative genome expansion was also studied by analyzing the *Adh1* orthologous region across different rice genomes. In this case also, the *O. sativa* genome was found to be larger by 28.0, 14.8, 11.3 and 9.4% relative to *O. punctata*, *O. rufipogon*, *O. glaberrima* and *O. niwara*, respectively [1]. Such comparisons across different rice genomes would not only aid evolutionary studies, but would also help in identifying useful alleles for crop improvement.

## THE IMPACT OF DUPLICATION ON EVOLUTION OF THE RICE GENOME

Polyplody is considered a key factor in the evolution of plants and their genomes [111]. It has been estimated that 50-80% of angiosperms are polyploids, including crop plants such as alfalfa (*Medicago sativa*), potato (*Solanum tuberosum*), wheat (*Triticum aestivum*), oat (*Avena sativa*), cotton

(*Gossypium hirsutum*), and coffee (*Coffea arabica*) [69, 111]. Some of the genetically diploid species are in fact ancient polyploids, for example, maize [29, 91], soybean [84], cabbage [48], and *Arabidopsis* [2].

The analyses of complete genome sequences of the *Arabidopsis thaliana* (dicot) and *Oryza sativa* (monocot), and of tree species *Populus trichocarpa*, revealed that genome duplication events are more common than previously anticipated [2, 36, 97]. The systematic analysis of the *Arabidopsis* genome sequence identified a large number of duplicated regions and showed that more than 80% of the *Arabidopsis* genes occur in genomic segments that have been duplicated at one time or another [10, 85, 106]. The genome has been proposed to be reshaped by three large-scale gene, and probably even entire genome, duplication events [85]. The analysis of draft genome sequence of *Populus* also suggested three genome duplication events: first at the base of angiosperms, second about 100-120 mya, and most recently, 60-65 mya [90, 97]. The evidence for polyploid origin of rice genome was provided more than 70 years ago [50] based on cytological observations. Later, rice was proposed as an allotetraploid that originated by hybridization between two species [66]. However, the recent identification of large-scale duplication events in the rice genome raised the question of rice polyploidy, because several different assumptions have been made regarding scale and timing of these duplications [70, 71, 86, 101, 109]. Based on the dating of duplicated segments identified from BAC assembly covering >70% of rice genome containing 15% of total rice genes, their distribution over different rice chromosomes, and comparison with the duplication history of *Arabidopsis*, it was concluded that rice is not an ancient polyploid, but an ancient aneuploid that has experienced the duplication ~70 mya [101]. This hypothesis was asserted based on a more completely assembled rice genome sequence, although more (20%) rice genes were predicted on duplicated regions [86]. However, based on largely the same *japonica* genome sequence, more extensive duplications (~62% of total rice genes) were found in the rice genome suggesting an ancient polyploidy rather than aneuploidy [70, 71]. Recently, the analysis of nearly complete rice genome sequence strongly supported the polyploid origin of the rice genome as a result of genome duplication that occurred ~70 mya [109]. However, a 3 Mb segment duplication of rice chromosomes 11 and 12 has been shown to have occurred 7.7 mya providing evidence for segmental duplications as well [95].

## ROLE OF TRANSPOSABLE ELEMENTS IN THE PROCESS OF EVOLUTION

Transposable elements (TEs) or transposons are the mobile genetic elements that constitute a significant proportion of plant genomes. Based on their different modes of mobility, TEs are classified into class I and class II elements [15]. The class I elements or retroelements move via RNA intermediate that is reverse transcribed prior to integration into the genome. These elements include Ty1-copia and Ty3-gypsy groups of LTR retrotransposons (retrotransposons flanked by long-terminal repeats), and LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) representing non-LTR transposons. The class II elements are the DNA transposons characterized by terminal inverted-repeats, including Ac-like, CACTA-like, and Mutator-like elements (MULE) superfamilies. The analysis of rice genome sequence using hidden Markov models revealed that the transposon content of *japonica* genome is as high as 35% [36]. The representatives from all the known transposon superfamilies are populated in the rice genome, including those that cannot be distinguished into either class I or II [36,96].

Genome sizes vary remarkably in plant species and are correlated to the amount of repetitive sequences [5, 26, 99, 107]. It has been suggested that rapid amplification of LTR-retrotransposons plays a major role in genome expansion [61, 78, 79, 112]. The plants with large genomes (e.g., maize) have numerous nested retrotransposons, which are less frequent in grasses with small genomes, e.g. rice [7, 21]. The class II elements outnumber the class I TEs more than two-fold in rice [36]. However, in maize, class I elements outnumber class II elements [60]. This differential amplification of LTR elements in maize and rice can explain the differential genomic expansion found between syntenic regions of maize and rice [12, 49, 88]. The removal of LTR-retrotransposons owing to unequal homologous recombination and illegitimate recombination has been suggested to be responsible for genome size reduction in rice [55]. More than 190 Mb of the LTR-retrotransposon sequences have been removed from rice genome in the last 8 million years [55]. The analysis of >1 Mb of orthologous regions showed that the genome sizes of both *indica* and *japonica* have expanded substantially, >2% and >6%, respectively, since their divergence from a common ancestor, mainly because of the amplification of LTR-retrotransposons; however,

subsequent deletions, insertions, and tandem duplications of all classes of DNA sequence have attenuated the expansion of their genomes [56]. These results indicate that LTR-retrotransposons have played a central role in evolution of the rice genome by generation and removal of genetic variations.

Retroposition of reverse transcribed spliced mRNAs gives rise to retrogenes. Such retrogenes are generally non-functional because of lack of regulatory sequences [11]. However, some of these acquire a new function by recruitment of certain regulatory sequences. Recently, the presence of 1,235 primary retrogenes has been reported in rice, out of which only 337 (27%) are pseudogenes and vast majority of others are functional [108]. In addition, 42% of these retrogenes were predicted to have chimeric protein coding sequence. Such chimerical gene structures are considered to be an important component of protein diversity [54, 72]. The duplication of gene fragments by MULEs in rice has also been documented and is considered as a mechanism for the creation of novel protein coding genes [41, 96]. Conversely, in a recent report, it has been demonstrated that MULE-mediated host genome duplications in rice result in the formation of pseudogenes that lack protein coding function. However, the possibility of regulation of host gene expression by transcribed pseudogenes has not been ruled out [42].

## **EVOLUTION OF GENE FAMILIES**

As discussed earlier, plant genomes have evolved essentially through segmental duplications and polyploidization, which gives rise to gene duplications. This genetic redundancy originating from polyploidy is erased subsequently by massive loss of duplicated genes [12, 49]. However, some of these duplicated genes are amplified, which give rise to paralogs and gene families [118]. The extent of paralog retention in plants is so important that generation of protein diversity in this kingdom has been hypothesized primarily through gene duplication rather than by alternative splicing, as has been proposed for vertebrates [35, 122]. The mechanism that leads to paralog retention is functional divergence of duplicated genes, which involves neofunctionalization (acquisition of a new function by one of the duplicated genes) or by subfunctionalization (the duplicated genes retain different subsets of the functionality of the ancestral gene) during evolution [27]. It has long been believed that gene duplication is the primary source for the evolution of genes with new

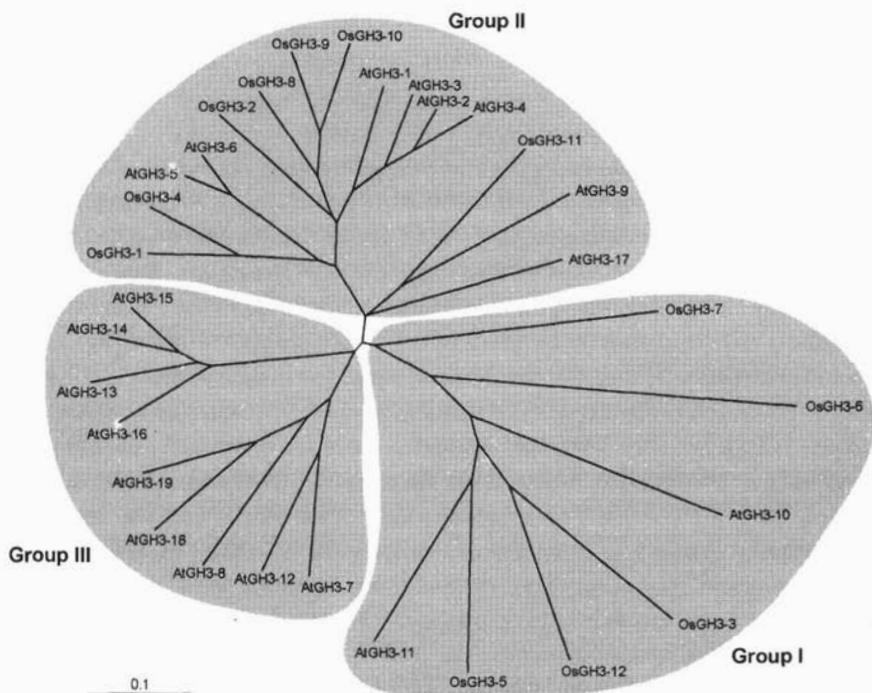
functions [89, 92]. Functionally redundant genes can also be retained when the dosage effect presents a selective advantage [38, 68, 103, 104]. For example, the gene products that participate together in a macromolecular complex or pathway require a dosage balance for proper function [14, 103, 104]. The preferential retention of duplicated genes encoding long complex proteins and their unexpectedly slow divergence (perhaps because of homogenization) suggest that a primary advantage of retaining duplicated genes may be the buffering of crucial functions [14]. Regardless of the mechanism of retention, the degree of gene family expansion in plants is substantially higher than in other eukaryotes. Recently, it has been demonstrated that transcription factor families have much higher expansion rates in plants than in animals [83]. The analysis of duplicated gene pairs in *Arabidopsis* revealed that the duplicated genes, which have been retained, are not evenly distributed among various functional categories [9], which indicates non-random gene loss events. The genes encoding proteins involved in signal transduction and transcription were found to be preferentially retained, whereas housekeeping genes have been lost preferentially. More than 50% of the duplicated pairs exhibit significantly different expression patterns [9], which suggest that functional diversification of the surviving duplicated genes is a major feature of the long-term evolution of polyploids.

The *Arabidopsis* and rice genome sequences revealed that a majority of plant genes belong to large gene families. Both tandem gene duplication and chromosomal segmental duplication followed by dispersal and diversification have been ascribed for expansion and evolution of gene families [38, 40, 51, 87]. The mechanisms, including transposition, homologous recombination, unequal crossing over, and reintegration of pre-mRNA/mRNA can also account for the expansion of gene families [21, 38, 40, 51, 62, 67, 118]. Recently, evidence was provided for ongoing individual gene duplications in rice, which provide ever-increasing raw material for studying gene genesis and their functions [123]. The size of some gene families varies among different organisms, which may be because of their lineage-specific expansion. It has been thus suggested that more subtle changes in gene content among species can be generated by lineage-specific expansions (LSEs) and contractions of a gene family [52, 82]. The computational analysis of LSE in five eukaryotic species, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana* revealed that a significant proportion (20% in yeast to 80% in *Arabidopsis*) of the proteins

encoded in these genomes belong to LSEs [52]. It has been suggested that genes encoding proteins with predicted organism-specific functions, such as pathogen and stress response, transcription regulation, protein degradation, protein modification, and signal transduction, are often subjected to LSE followed by fine-tuning through sequence diversification [47, 52]. Although the LSE of some of the transcription factor families has been shown, parallel expansion of transcription factor families was more pronounced in *Arabidopsis* and rice [83]. While LSE may be responsible for lineage-specific adaptation, the parallel expansion suggests common selection pressure contributing to the retention of certain genes in both lineages. Recently, the LSE of early auxin-responsive GH3 gene family has been reported in *Arabidopsis* vis-à-vis rice [39]. Although group I and II GH3 proteins exhibited parallel expansion in rice and *Arabidopsis*, group III proteins were completely absent in rice and other monocots (Fig. 1; [39]). These results suggested that either the group III GH3 genes evolved only in dicots, or they were lost in monocots after divergence of monocots and dicots, and may perform dicot-specific functions.

## **COMPARISON OF RICE WITH CEREAL GENOMES**

The cereals are known to have diverged from their common ancestor about 50-70 mya [70]. One of the first reports on cereal genome conservation showed that a common set of DNA probes could be used for diverse cereal species such as rice, wheat, rye, foxtail millet, sorghum and sugarcane. This finding, when reported for the first time, came as a surprise considering the fact that the cereals diverged from their common ancestor ~60 mya and the genome size can differ up to 40-folds [64]. Syntenic relationship between cereal genomes such as rice, wheat, maize, barley, rye and sorghum was also confirmed [28]. The large difference in sizes amongst cereal genomes can be attributed to the variation in the repetitive element content [31]. For instance, a small genome such as rice has ~35% repeat elements [36], but large genomes such as wheat and barley have more than 75% repetitive DNA [6]. This finding raised immense possibilities of the use of a smaller cereal genome as a reference, which would aid in the identification of target genes in other crop genomes. These preliminary studies reporting a strong syntenic relationship amongst cereal crops had, however, been done at a low resolution. For instance, 98% of wheat, barley and maize proteins could



**Fig. 1.** Phylogenetic relationship among rice (Os) and *Arabidopsis* (At) GH3 proteins. The unrooted tree was generated using ClustalX program by neighbor-joining method and displayed by Treeview. All the rice and *Arabidopsis* GH3 proteins clustered distinctly into three groups (I, II and III). Group III includes only *Arabidopsis* GH3 proteins. (Redrawn from [39]).

be identified in the draft sequence of the rice genome [32]. In fact, the initial data obtained at the molecular level, such as *sh2/a1* and *Adh1* maize loci and the corresponding region in rice and sorghum genomes [16] and also wheat and barley genomes corresponding to receptor-like kinases [24], suggested that colinearity is preserved even at the micro-level. However, recent studies carried out at a high resolution have shown several breaks in the colinearity, as seen at the sequence level. One of the first such reports focused on *rpg1* in barley and its ortholog in the rice genome [45]. In another study, a syntenic barley and rice BAC clone were sequenced to study the extent of conservation. At least four conserved regions containing four predicted genes could be identified in the analyzed region. Even though rearrangements were identified, the extent of conservation did prove useful for gene identification as well as

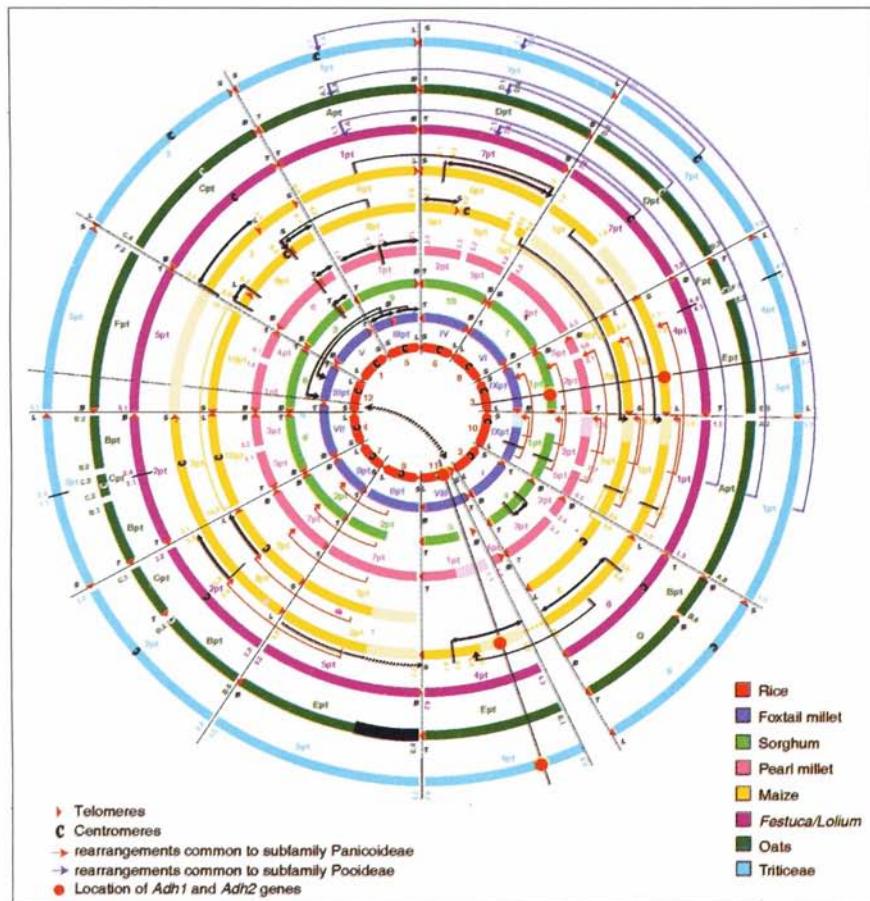
prediction of gene structure in barley [22]. The scale of analysis was further extended by considering four regions of maize, rice and sorghum, two of wheat and one of barley [8]. Gene rearrangements were observed in each case; in fact even in closely related species such as barley-wheat or maize-sorghum, at least 20% alterations could be observed. In a study undertaken to assess the extent of synteny between rice and maize, 2,629, maize molecular markers were used to identify orthologous regions in the rice genome. A general preservation of synteny was observed, however, several breaks such as insertions, deletions and duplications were observed at the micro-level [77]. In another comparison of maize and rice, two duplicated regions in the maize genome were compared with a single such orthologous region in rice. Syntenic segments could be identified on alignment of the three regions. However, differential contractions in the genic/intergenic regions, as well as expansion due to retrotransposable elements, could be identified in maize. Also, at least 9% of the predicted genes in maize were completely missing from the rice genome and almost 20% had moved to other locations [12]. A 212 kb contig of barley corresponding to the orthologous position of *Rph7* leaf rust resistance gene of rice was sequenced and analyzed for the extent of colinearity [13]. The five HGA family members were conserved at the orthologous positions, but the biggest difference was in the size of intergenic regions with that in barley being almost 40-folds compared to rice. Similarly, in a comparison of sorghum chromosome 3 and rice chromosome 1, conservation of gene order to a large extent with minor rearrangements, as well as one major chromosomal rearrangement, were observed [46]. In another study comparing rice and wheat at the level of gene models, the majority of rice gene models from chromosome 11 mapped to group 4 chromosomes of wheat indicating a common origin. However, many of these gene models which mapped to the distal end of wheat chromosome 4A (long arm) also mapped to the short arms of chromosomes 4B and 4D indicating significant rearrangements. A similar situation was seen in the case of chromosome 12 where most of the gene models mapped to wheat group 5 chromosomes while the rest were uniformly distributed on the remaining homeologous groups [95]. A comparative distribution of rice chromosome 11-12 gene homologs to the wheat homoeologous groups indicates different origins for the two chromosomes. This contradicts the earlier observation which supports the evolution of chromosomes 11 and 12 via polyploidization [71]. Comparison of the draft rice genome with 610 maize markers revealed a

complex pattern even at a low stringency [32]. In fact, extensive rearrangements have also been observed at the intraspecific level, as revealed by comparison of two maize lines around the *Bronze* locus as well as two wheat lines in a comparison of the low molecular weight glutelin locus. The studies also point to the fact that colinearity can be lost sometimes even between genomes representing the same species [18]. The preservation of colinearity amongst the grass genomes has definitely made a significant impact on the study of cereals using rice as a reference genome.

The rice genome, along with the sequence information available for other plant genomes as well as comparative genetic maps, would definitely serve as an integrated platform to understand cereal genome evolution, and also aid in the identification of useful orthologs. In fact, a large number of databases provide individual and comparative information of cereal crops including TIGR (<http://www.tigr.org>), RGP (<http://rgp.dna.affrc.go.jp>), IRFGC (<http://www.iris.irri.org/IRFGC>) and Gramene (<http://www.gramene.org>). The Gramene database is a valuable resource for cereal genomics. It is basically a rice-centric database but also provides comparative maps as well as related information on maize, sorghum, barley, wheat and oat [110]. Other than these, there is the NCBI Unigene database containing 44,367 entries representing almost the total number of genes known in rice. This information would further serve as a value addition to the sequence information since it would help in refining the intron-exon boundaries as well as define gene coding regions. This information would of course translate into better data for crop genome comparison. Interestingly, divergent regions have been observed even at the subspecies level amongst different cereals. Thus, comparisons even at the subspecies level would be valuable in identifying regions of evolutionary stability as well as divergence. At least macrocolinearity will help in broadly identifying the syntenic region in the crop of interest (Fig. 2).

## **COMPARISON OF MODEL PLANT GENOMES - RICE VERSUS ARABIDOPSIS**

Monocots and dicots diverged ~200 mya and are considered to be distantly related species [115]. Comparison of the model monocot and dicot genomes, i.e., rice and *Arabidopsis*, showed that 90% of *Arabidopsis* genes have a homolog in the rice genome; however, only 71% of the rice



**Fig. 2.** The updated Crop Circle depicting the current relationships between genomes of eight species from [20]. Arrows indicate rearrangements relative to rice. In some cases, arrows have been supplemented or replaced by numbered genome fragments. Reproduced with permission.

genes had a homolog in the *Arabidopsis* genome. The *Arabidopsis* genome has an overall GC content of 34.7% while rice genome has 43.6%. This higher GC content in rice was especially seen in the coding regions. A negative gradient has been observed in the majority of rice genes, typically being up to 25% richer in GC content at the 5' end in comparison to the 3' end [122]. Such a gradient has, however, not been seen in case of *Arabidopsis* genes.

Although rice and *Arabidopsis* are considered to be evolutionarily divergent, some claims of syntenic regions have been made between the

two genomes, albeit at a very low resolution. An initial comparison of *Arabidopsis* and rice showed that colinearity between these two plant genomes had been considerably eroded and confined to limited regions [19]. Thereafter, however, several studies have shown that colinearity between *Arabidopsis* and rice does exist [32, 58, 76, 100]. In fact, in a large-scale analysis of the complete *Arabidopsis* and the available rice genome sequence, a total of 60 syntenic regions could be identified between the two genomes. The extent of conservation was low with just ~17% genes conserved in homologous segments [76]. After availability of the draft rice genome, the number of colinear blocks identified between the two genomes rose to 137 [32]. The syntenic region encompassed 119 *Arabidopsis* proteins, showing an identity of at least 70% over a minimum stretch of 30 amino acids, indicating that synteny was maintained but to a lesser extent. With such weak signatures of synteny between the two genomes, it is difficult to be conclusive about its existence. However, a good conservation of primary gene structure has been observed between the two genomes [93]. In several cases, information from rice has proved useful in identification of *Arabidopsis* orthologs. For instance, rice EST sequence information for the gibberellin acid insensitive (GA1) gene from rice was used for the identification of its ortholog in the *Arabidopsis* genome [73].

An area of general interest relates to genes present either in rice or in *Arabidopsis*, since this would aid in better understanding of the monocot/dicot divergence. In a comparison of rice and *Arabidopsis* gene repertoire, 2,859 genes were identified as unique to rice. The detailed study of these monocot specific genes could help in understanding the monocot-dicot difference [36]. There are 5,663 rice and 3,402 *Arabidopsis* specific proteins [93]. Some of the proteins specific to rice include cereal specific proteins such as prolamins, hormone response proteins, defense proteins, chitinase precursor, seed allergen and starch branching enzyme. A comparison of the rice and *Arabidopsis* gene content has also revealed that some interesting genes are missing in one or the other. For instance, Toll-interleukin 1 receptor resistance (TIR) encoding genes are not present in the rice genome but are prevalent in the *Arabidopsis* genome [32].

The number of tandem gene families is also comparable between the rice (14%) and *Arabidopsis* (17%) genomes [36]. In a study involving comparison of Tandemly Arrayed Genes (TAGs), a striking similarity was seen between rice and *Arabidopsis* in terms of the proportion of genes that

are tandemly arrayed, number of genes within an array, number of arrays, and the scarcity of TAGs in the centromeric regions. Further, TAGs in both the genomes are enriched for membrane protein-encoding and stress-related genes, and under represented for transcription and DNA/RNA-binding-related genes [75].

## CONCLUSION

To date, complete sequence information is available for genomes of rice, *Arabidopsis* and recently *Populus* [2, 36, 97, 105] and to a large extent is also available for *Medicago*, lotus, tomato and maize [74]. It is generally accepted that considerable synteny exists between various plant genomes [81]. The grasses are represented by over 10,000 species, of which the more well-known members include rice, wheat, maize, rye and sorghum. Rice has an estimated genome size of ~400 Mb, which is smallest in comparison to other economically important cereals such as wheat (16,000 Mb), barley (5,000 Mb), maize (2,500 Mb), rye (7,600 Mb) and sorghum (735 Mb). This was the main reason for selection of rice as a representative cereal genome for sequencing. Due to the much larger size of the other cereal genomes, it is highly improbable that complete sequencing of these genomes will be achieved in the near future. Thus, in this context, the rice genome will gain utility in deciphering the evolutionary history of plant genomes, especially crop plants. Initial studies using cross-hybridizing genetic markers provided strong evidence for conservation and indicated the sharing of large syntenic blocks in rice, maize, sorghum, wheat, rye, sugarcane and other cereal genomes [28,113]. In fact, several important traits such as seed shattering, plant height and disease resistance could be mapped to syntenic regions amongst various grass genomes [102, 119].

Sequence analysis of rice and other plant genomes clearly points to the existence of an ancestral plant genome, due to which syntenic signatures are clearly evident even between the evolutionary divergent rice and *Arabidopsis* genomes. Studying the rice genome evolution in relation to other plants has not only aided in understanding the complex pattern of plant genome evolution, but has also been useful for map-based cloning. The rice genome would indeed prove valuable for identification of novel components controlling important traits such as abiotic stress tolerance, faster germination rate, or improving nutritional quality of syntenic crops. The cereals have been a particularly interesting subject

matter for evolutionary studies due to their relatively less time of divergence (~60 mya). Also, a large number of resources are available for cereals which greatly aid the understanding of their origin. Comparative studies gain significance in the case of cereal genomes, which have a large amount of repetitive DNA, and hence would greatly benefit both scientifically as well as economically from the information available in rice. Analyses to trace the roots of cereal genome evolution have shown that the cereals have evolved independently for at least two-thirds of the time after duplication. The differences in colinearity can be explained by the gene loss events occurring during this period [71]. In fact, alignment and analysis of all the available sequence information for all plant species would help in understanding the mechanism of plant evolution. Rice would definitely play a central role in the scenario. Methods aimed at genome-wide expression analysis would also aid in the understanding of evolution by revealing the pattern of gene expression across the plant kingdom. Identification of candidate genes may become an arduous task depending upon the level of synteny in the analyzed region, but the use of synteny as a valuable source of markers is undisputable [20]. To overcome the limitations posed by the large genome size in the majority of cereals, focus is being shifted to getting the transcriptome information which is reflected by the large number of cereal ESTs available in the NCBI database. Such information falls short in terms of not being able to provide information regarding the regulatory elements as well as low expressing genes. Some alternate methods which are being used to obtain sequence libraries rich in gene coding regions include methylation filtration and high C<sub>o</sub>t sequencing [59]. It is presumed that the genome sequence of 23 plant species is either partially available or will be completely available in the near future. However, these would represent only 13 out of 606 plant families. In future, carefully choosing representative family members would help in a much more balanced phylogenetic comparison of rice with the other members of the plant kingdom [37].

As part of the Oryza Map Alignment Project, physical maps are being made available for one cultivated and 11 wild rice species from diverse geographical distribution and ecological adaptation [1, 114]. The wild species would of course represent a rich gene pool whose potential has to be tapped in order to identify useful alleles for crop improvement.

## Acknowledgments

The authors are thankful to Prof. J. Messing, Rutgers University, USA for critical reading of the manuscript and valuable comments. The Department of Biotechnology, Government of India, New Delhi financially supported the research on rice genomics. Shubha Vij, Mukesh Jain and V. Ravi are grateful to the Council of Scientific and Industrial Research, New Delhi, for the award of research fellowship.

## References

- [1] Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, et al. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res* 2006; 16: 140-147.
- [2] Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; 408: 796-815.
- [3] Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 1991; 9: 208-218.
- [4] Barry GF. The use of the Monsanto draft rice genome sequence in research. *Plant Physiol* 2001; 125: 1164-1165.
- [5] Bennett MD, Leitch IJ. Nuclear DNA amounts in angiosperms. *Ann Bot* 1995; 76: 113-176.
- [6] Bennetzen JL, Freeling M. The unified grass genome: synergy in synteny. *Genome Res* 1997; 7: 301-306.
- [7] Bennetzen JL, SanMiguel P, Chen M, Tikhonov A, Francki M, Avramova Z. Grass genomes. *Proc Natl Acad Sci USA* 1998; 95: 1975-1978.
- [8] Bennetzen JL, Ramakrishna W. Exceptional haplotype variation in maize. *Proc Natl Acad Sci USA* 2002; 99: 9093-9095.
- [9] Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 2004; 16: 1679-1691.
- [10] Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 2003; 422: 433-438.
- [11] Brosius J. Retroposons-seeds of evolution. *Science* 1991; 251:753.
- [12] Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, et al. Uneven chromosome contraction and expansion in the maize genome. *Genome Res* 2006; 16: 1241-1251.
- [13] Brunner S, Keller B, Feuillet C. A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. *Genetics* 2003; 164: 673-683.
- [14] Chapman BA, Bowers JE, Feltus FA, Paterson AH. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication. *Proc Natl Acad Sci USA* 2006; 103: 2730-2735.

- [15] Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994; 371: 215-220.
- [16] Chen M, SanMiguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, et al. Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. *Proc Natl Acad Sci USA* 1997; 94: 3431-3435.
- [17] Cooke HJ. Silence of the centromeres-not. *Trends Biotechnol* 2004; 22: 319-321.
- [18] Delseny M. Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement. *Curr Opin Plant Biol* 2004; 7: 126-131.
- [19] Devos KM, Beales J, Nagamura Y, Sasaki T. *Arabidopsis-rice*: will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res* 1999; 9: 825-829.
- [20] Devos KM. Updating the 'crop circle'. *Curr Opin Plant Biol* 2005; 8: 155-162.
- [21] Du C, Swigoňová Z, Messing J. Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol Biol* 2006; 6: 62.
- [22] Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, et al. Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol* 2001; 125: 1342-1353.
- [23] Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, et al. Sequence and analysis of rice chromosome 4. *Nature* 2002; 420: 316-320.
- [24] Feuillet C, Keller B. High gene density is conserved at syntenic loci of small and large grass genomes. *Proc Natl Acad Sci USA* 1999; 96: 8265-8270.
- [25] Feuillet C, Keller B. Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Ann Bot (Lond)* 2002; 89: 3-10.
- [26] Flavell AJ, Pearce SR, Kumar A. Plant transposable elements and the genome. *Curr Opin Genet Dev* 1994; 4: 838-844.
- [27] Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999; 151: 1531-1545.
- [28] Gale MD, Devos KM. Plant comparative genetics after 10 years. *Science* 1998; 282: 656-659.
- [29] Gaut BS, Doebley JF. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci USA* 1997; 94: 6809-6814.
- [30] Ge S, Sang T, Lu BR, Hong DY. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA* 1999; 96: 14400-14405.
- [31] Goff SA. Rice as a model for cereal genomics. *Curr Opin Plant Biol* 1999; 2: 86-89.
- [32] Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 2002; 296: 92-100.
- [33] Guyot R, Keller B. Ancestral genome duplication in rice. *Genome* 2004; 47: 610-614.
- [34] Han B, Xue Y. Genome-wide intraspecific DNA-sequence variations in rice. *Curr Opin Plant Biol* 2003; 6: 134-138.
- [35] Hodgkin J. A view of mount *Drosophila*. *Nature* 2000; 404: 442-443.
- [36] International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* 2005; 436: 793-800.
- [37] Jackson S, Rounsley S, Purugganan M. Comparative sequencing of plant genomes: choices to make. *Plant Cell* 2006; 18: 1100-1104.

- [38] Jain M, Kaur N, Garg R, Thakur JK, Tyagi AK, Khurana JP. Structure and expression analysis of early auxin-responsive Aux/IAA gene family in rice (*Oryza sativa*). *Funct Integr Genomics* 2006a; 6: 47-59.
- [39] Jain M, Kaur N, Tyagi AK, Khurana JP. The auxin-responsive GH3 gene family in rice (*Oryza sativa*). *Funct Integr Genomics* 2006b; 6: 36-46.
- [40] Jain M, Tyagi AK, Khurana JP. Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*). *Genomics* 2006c; 88: 360-371.
- [41] Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 2004; 431: 569-573.
- [42] Juretic N, Hoen DR, Huynh ML, Harrison PM, Bureau TE. The evolutionary fate of MULE-mediated duplications of host gene fragments in rice. *Genome Res* 2005; 15: 1292-1297.
- [43] Khush GS. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 1997; 35: 25-34.
- [44] Khush GS. What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol Biol* 2005; 59: 1-6.
- [45] Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. *Plant Mol Biol* 1997; 35: 187-195.
- [46] Klein PE, Klein RR, Vrebalov J, Mullet JE. Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. *Plant J* 2003; 34: 605-621.
- [47] Kondrashov FA, Rogozin IB, Wolf YA, Koonin E. Selection in the evolution of gene duplications. *Genome Biol* 2002; 3: RESEARCH0008.0001-0008.0009.
- [48] Lagercrantz U, Lydiate DJ. Comparative genome mapping in *Brassica*. *Genetics* 1996; 144: 1903-1910.
- [49] Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, et al. Gene loss and movement in the maize genome. *Genome Res* 2004; 14: 1924-1931.
- [50] Lawrence WJC. The secondary association of chromosomes. *Cytologia* 1931; 2: 352-384.
- [51] Lecharny A, Boudet N, Gy I, Aubourg S, Kreis M. Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure. *J Struct Funct Genomics* 2003; 3: 111-116.
- [52] Lespinet O, Wolf YI, Koonin EV, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res* 2002; 12: 1048-1059.
- [53] Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci USA* 2006; 103: 9578-9583.
- [54] Long M, Betran E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 2003; 4: 865-875.
- [55] Ma J, Devos KM, Bennetzen JL. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 2004; 14: 860-869.
- [56] Ma J, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 2004; 101: 12404-12410.
- [57] Maclean JL, Dawe DC, Hardy B, Hettel GP, eds. *Rice almanac*. Los Baños, Philippines: International Rice Research Institute, 2002; 253 pp.

- [58] Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, et al. Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res* 2001; 11: 1167-1174.
- [59] Mayer K, Mewes HW. How can we deliver the large plant genomes? Strategies and perspectives. *Curr Opin Plant Biol* 2002; 5: 173-177.
- [60] Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, et al. Sequence composition and genome organization of maize. *Proc Natl Acad Sci USA* 2004; 101: 14349-14354.
- [61] Meyers BC, Tingey SV, Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 2001; 11: 1660-1676.
- [62] Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* 1998; 8: 1113-1130.
- [63] Mizuno H, Wu J, Kanamori H, Fujisawa M, Namiki N, Saji S, et al. Sequencing and characterization of telomere and subtelomere regions on rice chromosomes 1S, 2S, 2L, 6L, 7S, 7L and 8S. *Plant J* 2006; 46: 206-217.
- [64] Moore G, Devos KM, Wang Z, Gale MD. Cereal genome evolution. Grasses, line up and form a circle. *Curr Biol* 1995; 5: 737-739.
- [65] Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, et al. Sequencing of a rice centromere uncovers active genes. *Nat Genet* 2004; 36: 138-145.
- [66] Nayar NM. Origin and cytogenetics of rice. *Adv Genet* 1973; 17: 153-292.
- [67] Noel L, Moores TL, van Der Biezen EA, Parniske M, Daniels MJ, Parker JE, et al. Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of *Arabidopsis*. *Plant Cell* 1999; 11: 2099-2112.
- [68] Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, et al. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet* 2003; 19: 141-147.
- [69] Otto SP, Whitton J. Polyploid incidence and evolution. *Annu Rev Genet* 2000; 34: 401-437.
- [70] Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA. Structure and evolution of cereal genomes. *Curr Opin Genet Dev* 2003; 13: 644-650.
- [71] Paterson AH, Bowers JE, Chapman BA. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 2004; 101: 9903-9908.
- [72] Patthy L. Genome evolution and the evolution of exon-shuffling-a review. *Gene* 1999; 238: 103-114.
- [73] Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, et al. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 1999; 400: 256-261.
- [74] Rensink WA, Buell CR. Microarray expression profiling resources for plant genomics. *Trends Plant Sci* 2005; 10: 603-609.
- [75] Rizzon C, Ponger L, Gaut BS. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput Biol* 2006; 2: e115.
- [76] Salse J, Piegu B, Cooke R, Delseny M. Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res* 2002; 30: 2316-2328.
- [77] Salse J, Piegu B, Cooke R, Delseny M. New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. *Plant J* 2004; 38: 396-409.

- [78] SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996; 274: 765-768.
- [79] SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL. The paleontology of intergene retrotransposons of maize. *Nat Genet* 1998; 20: 43-45.
- [80] Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, et al. The genome sequence and structure of rice chromosome 1. *Nature* 2002; 420: 312-316.
- [81] Schmidt R. Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Mol Biol* 2002; 48: 21-37.
- [82] Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 2004; 16: 1220-1234.
- [83] Shiu SH, Shih MC, Li WH. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol* 2005; 139: 18-26.
- [84] Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, et al. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 1996; 144: 329-338.
- [85] Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 2002; 99: 13627-13632.
- [86] Simillion C, Vandepoele K, Saeys Y, Van de Peer Y. Building genomic profiles for uncovering segmental homology in the twilight zone. *Genome Res* 2004; 14: 1095-1106.
- [87] Song R, Llaca V, Linton E, Messing J. Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family. *Genome Res* 2001; 11: 1817-1825.
- [88] Song R, Llaca V, Messing J. Mosaic organization of orthologous sequences in grass genomes. *Genome Res* 2002; 12: 1549-1555.
- [89] Stephens SG. Possible significances of duplication in evolution. *Adv Genet* 1951; 4: 247-265.
- [90] Stokstad E. Poplar tree sequence yields genome double take. *Science* 2006; 313: 1556.
- [91] Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J. Close split of sorghum and maize genome progenitors. *Genome Res* 2004; 14: 1916-1923.
- [92] Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet* 2004; 38: 615-643.
- [93] The Rice Annotation Project. Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res* 2007; 17: 175-183.
- [94] The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 2003; 300: 1566-1569.
- [95] The Rice Chromosomes 11 and 12 Sequencing Consortia. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol* 2005; 3: 20.
- [96] Turcotte K, Srinivasan S, Bureau T. Survey of transposable elements from rice genomic sequences. *Plant J* 2001; 25: 169-179.
- [97] Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Tort. & Gray). *Science* 2006; 313: 1596-1604.

- [98] Tyagi AK, Khurana JP, Khurana P, Raghuvanshi S, Gaur A, Kapur A, et al. Structural and functional analysis of rice genome. *J Genet* 2004; 83: 79-99.
- [99] Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui K. Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol Biol* 1997; 35: 791-799.
- [100] van Dovedeerd AM, Hall CR, Bent EG, Johnson SJ, Bevan MW, Bancroft I. Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*. *Genome* 1999; 42: 887-892.
- [101] Vandepoele K, Simillion C, Van de Peer Y. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* 2003; 15: 2192-2202.
- [102] Varshney RK, Hoisington DA, Tyagi AK. Advances in cereal genomics and applications in crop breeding. *Trends Biotechnol* 2006; 24: 490-499.
- [103] Veitia RA. Gene dosage balance in cellular pathways: implications for dominance and gene duplicability. *Genetics* 2004; 168: 569-574.
- [104] Veitia RA. Paralogs in polyploids: one for all and all for one? *Plant Cell* 2005; 17: 4-11.
- [105] Vij S, Gupta V, Kumar D, Vydiathan R, Raghuvanshi S, Khurana P, et al. Decoding the rice genome. *Bioessays* 2006; 28: 421-432.
- [106] Vision TJ, Brown DG, Tanksley SD. The origins of genomic duplications in *Arabidopsis*. *Science* 2000; 290: 2114-2117.
- [107] Wang S, Liu N, Peng K, Zhang Q. The distribution and copy number of *copia*-like retrotransposons in rice (*Oryza sativa* L.) and their implications in the organization and evolution of the rice genome. *Proc Natl Acad Sci USA* 1999; 96: 6824-6828.
- [108] Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, et al. High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 2006; 18: 1791-1802.
- [109] Wang X, Shi X, Hao B, Ge S, Luo J. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol* 2005; 165: 937-946.
- [110] Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, et al. Gramene: a resource for comparative grass genomics. *Nucleic Acids Res* 2002; 30: 103-105.
- [111] Wendel JF. Genome evolution in polyploids. *Plant Mol Biol* 2000; 42: 225-249.
- [112] Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J* 2001; 26: 307-316.
- [113] Wilson WA, Harrington SE, Woodman WL, Lee M, Sorrells ME, McCouch SR. Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. *Genetics* 1999; 153: 453-473.
- [114] Wing RA, Ammiraju JS, Luo M, Kim H, Yu Y, Kudrna D, et al. The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol* 2005; 59: 53-62.
- [115] Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc Natl Acad Sci USA* 1989; 86: 6201-6205.
- [116] Wu J, Kurata N, Tanoue H, Shimokawa T, Umehara Y, Yano M, et al. Physical mapping of duplicated genomic regions of two chromosome ends in rice. *Genetics* 1998; 150: 1595-1603.

- [117] Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, et al. Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell* 2004; 16: 967-976.
- [118] Xu J-H, Messing J. Maize haplotype with a *helitron*-amplified cytidine deaminase gene copy. *BMC Genet* 2006; 7: 52.
- [119] Xu Y, McCouch SR, Zhang Q. How can we use genomics to improve cereals with rice as a reference genome? *Plant Mol Biol* 2005; 59: 7-26.
- [120] Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S, et al. Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell* 2006; 18: 2123-2133.
- [121] Yang TJ, Yu Y, Chang SB, de Jong H, Oh CS, Ahn SN, et al. Toward closing rice telomere gaps: mapping and sequence characterization of rice subtelomere regions. *Theor Appl Genet* 2005; 111: 467-478.
- [122] Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002; 296: 79-92.
- [123] Yu J, Wang J, Lin W, Li S, Li H, Zhou J, et al. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* 2005; 3: e38.
- [124] Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, et al. Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res* 2004; 32: 2023-2030.



# Genome Organization and Evolution in Genus *Quercus* (Fagaceae): Special Attention to Two European White Oaks *Quercus petraea* (Matt.) Liebl. and *Q. robur* L.

VLATKA ZOLDOS PECNIK

Department of Molecular Biology, Faculty of Science, University of Zagreb, Croatia

## ABSTRACT

*Quercus* is a fascinating tree genus. Even on the face value, oaks are impressive because of their immense size and very long life span. Oak trees are of enormous economical, anthropological and ecological significance. With around 600 species, this is the most species-rich genus in the family Fagaceae, which in whole comprises around 1,000 species. Oaks are widespread from the temperate zone of the Northern Hemisphere to Mediterranean and subtropical zones. The oaks probably represent the most extensively studied genus by different specialists, such as population geneticists, tree breeders, sylviculturists, cytogeneticists, molecular biologists and plant physiologists. The tremendous morphological diversity of oaks varies from deciduous to evergreen trees and shrubs. There are common cases where pairs of oaks show clear morphological and ecological differences but are poorly differentiated on genetic level, as for example the most popular and predominant European white oaks, *Q. petraea* (Matt.) Liebl. and *Q. robur* L. They can hybridize in nature, as many pairs of oak species of different taxonomical sections, and still remain distinct morphologically. Therefore, the concept of biological species has

been questionable as far as oaks are concerned, and a new concept of multispecies or ecological species has been proposed. This review has been restricted to *Q. petraea* and *Q. robur* with respect to interspecific hybridization and gene flow for oak genome structure. However, these two oaks, and other oak species in general, are still insufficiently explored in terms of molecular structure of their genomes. The present data will give a valuable image of oak genome structure, which will serve as guidelines for further research. Cytogenetic data are also scarce, but progress has been made recently in terms of chromosome number and genome size determination, karyotyping (after Giemsa C-banding and fluorochrome banding) and mapping of ribosomal DNA, retrotransposon-like sequences and telomeres on chromosomes of some oak species. Chromosomal and DNA markers are used across the genus *Quercus* to infer about karyotype/genome evolution and speciation. Also, the first comprehensive genetic map is given for *Q. robur*. In spite of the large volumes of literature published on oaks, this woody group of plants still requires further exploration.

**Key Words:** Evolution, gene flow, genome organization, genetic differentiation, interspecific hybridization, karyotypes, molecular markers, systematics, phylogeny, *Quercus*

**Abbreviations:** AFLP= Amplified Fragment Length Polymorphism, CMA= Chromomycin A3, cpDNA= chloroplast DNA, DAPI= 4',6-DiAmidino-2-PhenylIndole, FISH= Fluorescence In Situ Hybridisation, IGS= Intergenic Spacer, ITS= Internal Transcribed Spacer, NOR= Nucleolar Organizer Region, mtDNA= mitochondrial DNA, PCR= Polymerase Chain Reaction, QTL= Quantitative Trait Locus, RAPD= Random Amplified Polymorphic DNA, RDA= Representational Difference Analysis, rDNA= ribosomal DNA, SCAR= Sequence Characterized Amplified Regions, SNP= Single Nucleotide Polymorphism, SSR= Simple Sequence Repeat.

## INTRODUCTION

Genus *Quercus* comprises around 600 species, which forms a remarkable proportion of deciduous forests of the Northern Hemisphere and are important evergreen elements of the Mediterranean woodlands and subtropical vegetation. Oaks are of anthropological importance because of usage of timber as building material and acorns as animal and human food. A significant anthropological meaning of oak is also confirmed by the fact that oak is a part of almost all mythologies, from Celtic and Greek to Christian. Oaks also have an essential role in the ecosystems since many plant and animal species, such as fungi and different insects, depend on them for ecological niches and they probably represent the most extensively studied tree species in the world. For over six decades, researchers have focused on *Quercus* species for studies of interrelated

questions about the roles of hybridization and gene flow in the functioning of species integrities, population genetic structure and plant evolution. Oaks form an ideal case study of ecological genetic principles and are considered to be the cardinal genus in understanding the role of glacial refugia and dynamics of continental recolonization [83].

In this chapter, genome and chromosomal organization and evolution in oaks in general are reviewed including the meaning of interspecific hybridization and gene flow for oak genome structure, as related to the example of two most interesting and predominant European white oaks, *Q. petraea* and *Q. robur*. Other aspects of biology and genetics of oaks are extensively reviewed by other authors: evolutionary significance of interspecific hybridization and the role of hybridization as a mechanism of dispersal [44, 107], phylogeny, biogeography and systematics [86, 87], genetic diversity [77], and ecological genetics (i.e., phylogeography with implications for postglacial colonization and dispersal dynamics of European oaks) and population genetic structure [83].

## **SYSTEMATICS, PHYLOGENY AND HISTORY OF THE GENUS QUERCUS**

The family Fagaceae currently includes nine genera according to Manos, Doyle and Nixon [86]: *Fagus* L., *Castanea* L., *Castanopsis* Spach., *Chrysolepis* Hjelmquist, *Colombobalanus* (Lozano, Hdz-C. & Henao) Nixon & Crepet, *Formanodendron* (Camus) Nixon & Crepet, *Lithocarpus* Bl., *Quercus* L., and *Trigonobalanus* Forman. Fagaceae make a forest plant cover of temperate, seasonally dry regions of the Northern Hemisphere, and represent economically very important group of species. Although the family Fagaceae generally consists of low levels of species diversity [85], the genera *Quercus* and *Lithocarpus* are the two most species-rich genera. The whole family comprises approximately 1,000 species, of which about half are *Quercus* [143]. On the generic level, the centre of diversity is found in tropical Southeast Asia, but at the species level it appears to be distributed evenly between the seasonal subtropical and evergreen tropical forests of Central America (for genus *Quercus*) and southern continental Asia and the Malayan Archipelago (for subfamily Castaneoideae) [87].

The latest data reports between 531 [97] and 600 [143] *Quercus* species in comparison to some 200-300 species that appeared in the first list, established by de Candolle [40], and to only 12 species, described by Linnaeus in early 1753. Many more species are found in America compared to Europe and Asia, and apart from "good species" numerous

hybrids are described [42, 65, 67, 68, 115, 132]. The genus functions taxonomically according to Linnean or classical species concept, which relies on criteria of morphological distinctness. However, perplexing patterns of intraspecific morphological variation, partially related to interspecific hybridization [6, 7, 34, 59, 65, 108, 112, 113, 115, 121, 125, 127, 129, 130], result in a large difference in the number of described species and make taxonomy and systematics of *Quercus* rather complicated and uncertain. Oaks are a typical example of a taxon challenging the biological species concept of Mayr [89] which assumes that if a group of individuals has an independent phyletic identity they have little or no chance of exchanging genetic information with other groups of individuals. Accordingly, binomials in *Quercus* would mean populations that are easily recognized as different in morphological terms, but are still not genetically isolated. Darwin had attempted to delineate *Quercus* species, which resulted in his own rejection of the idea that the notion of species is fixity at all times [39].

The latest phylogenetic reconstruction on the basis of DNA sequence data showed that *Quercus* is closely related to the castaneoid genera and possibly derived from among them [87].

Manos, Doyle and Nixon [86] have given the most reliable phylogenetic reconstruction of the genus *Quercus* and suggested the following clades, resolved on the basis of combined chloroplast DNA (cpDNA) and ITS/5.8S (internal transcribed spacers) rDNA (ribosomal DNA) data, which mostly support morphological cladistic analysis of Nixon [97]:

- (1) Subgenus *Cyclobalanopsis* (cycle cup oaks) which embraces subtropical to tropical Southeast Asian species.
- (2) Subgenus *Quercus* (white oaks) comprising four sections: sections *Quercus* s.s. (*sensu* Manos, Doyle and Nixon [86]), *Protobalanus* (golden cup oaks, intermediate oaks) and *Lobatae* (red oaks) (*sensu* Nixon [97]), which include mostly New World species, and section *Cerris* which includes mostly Old World species.

Molecular phylogeny provided novel data and supported previous hypotheses:

- (1) In intra-generic relationships (*Cerris*- [*Lobatae*- {*Protobalanus* + *Quercus* *sensu stricto*}]) a novelty is the recognition of the strictly Eurasian monophyletic section *Cerris* (*Cerris* and *Ilex* groups *sensu* Nixon [97]) and its basal placement within subgenus *Quercus*. In previous taxonomical treatments species forming this group were

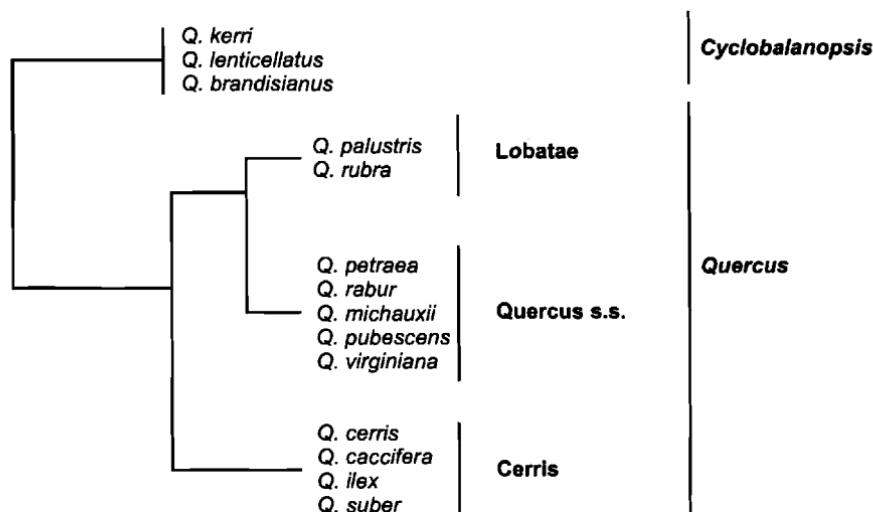
differently classified (recognized either at the sectional or subgeneric level [27, 122].

- (2) All the major oak lineages were supposed to have evolved locally at middle latitudes in the Tertiary period and now occupy areas within the distributional range of their fossil ancestors [4, 86].
- (3) New World origin for section *Quercus* s.s., which includes white oaks of the Northern Hemisphere, is suggested.

*Quercus* originated somewhere in Oligocene, as found in fossil deposits of North America [35, 36] so that the minimum age of several major oak groups recognized today would be 40 million years. Nevertheless, some doubtful fossils from Baltic amber of Northern Europe indicate even earlier existence in an uncertain Early Tertiary age [97]. Chloroplast DNA, together with fossil pollen records, and mitochondrial DNA (mtDNA) to a lesser extent, appeared extremely useful in tracing the history of oak species [24, 45-47, 52-55, 66, 69, 105, 106]. Oaks were confined to three main primary Pleistocene refugia in southern Europe – Iberian and Italian Peninsulas and the Balkans [66] – and to an additional but significant secondary refugia in Alps and Carpathian mountains (during the Younger Dryas period) [105, 106]. Oaks initialized colonization from these refugia somewhere in the late glacial maximum, around 13-11,000 BP. Analyses of cpDNA revealed genetic differentiation of the three refugia and outlined postglacial colonization patterns of oaks across the Europe [45, 55]. Phylogeography of oaks has been extensively reviewed by Lowe, Harris and Ashton [83].

## ECOLOGY OF OAKS

The worldwide distribution of *Quercus* species can be attributed to their phenotypic plasticity [133, 134] and genetic diversity [77, 117]. Oaks are monoecious, obligatory outcrossing and anemophilous trees. Floral characters associated with wind pollination are considered to be derived from characters associated with insect pollination and entomophilous ancestry is proposed for the family [87]. Oaks predominantly propagate by means of acorn dispersal over long distances by a large number of different animals. Jays are the principal long-distance dispersers [44] and the co-evolution of European jays and oaks has been shown [22, 107]. Recent molecular ecological studies point to extremely high levels of long-distance pollination and a limited level of acorn dispersal in oaks [43]. The dispersal abilities of oaks via pollen and seeds have been recently reviewed [107].



**Fig. 1.** Phylogenetic positions of the *Quercus* species studied cytogenetically by Zoldos and co-workers [148-150] and Chokchaichamnankit and co-workers [32]. The phylogenetic tree is a modified version of the phylogenetic tree constructed by Manos and co-workers [86].

However, the gravity of acorns and their short distance dispersal from mother trees, as well as longevity of pedunculate and sessile oaks, which is estimated to be over 1,000 years, are the main factors that one should bear in mind while considering ecological conditions in which the two species grow. Regarding a great distribution area of oaks (western, central and southeast Europe, western Asia and northern Africa), different ecological niches they inhabit and differences between the same niches in temporarily variable climatic, edaphic and hydrological conditions, it appears logical that certain communities of pedunculate and sessile oaks have been adapting genetically throughout hundreds of thousands of years and transmitting these adaptations on progeny. Any transportation of oaks from autochthonous to another ecological niche by man was stressful for the trees, which were consequently forced to grow in a non-native environment. Longevity of oaks means long-term adaptations to new ecological conditions and this can last for generations, i.e., several hundreds or even thousands of years. For this reason, the ecology of pedunculate and sessile as well as other species of oaks, has mostly been anthropologically influenced for the last 500 years, from the time of felling of oak trees because of timber demands, through industrial revolution, to the modern era when uncontrolled dispersal of seeds

throughout Europe started with the objective of improving the quality of oak forests, yet ignoring the origin of seeds and the conditions from which these seeds were taken out and brought into. Therefore, man plays the most influential role in the ecology of oaks. Moreover, this issue has to be taken into account especially when molecular studies are conducted on oaks, particularly the ones that consider intraspecific genetic structure.

## **INTERSPECIFIC HYBRIDIZATION IN OAKS: AN EXAMPLE OF EUROPEAN WHITE OAKS *Q. PETRAEA* AND *Q. ROBUR***

Interspecific hybridization has played an important role in plant evolution in general, most importantly by opening up the possibility for gene exchanges. Interspecific hybridization and gene flow in oaks is a matter of great interest and controversial opinions within oak specialists. Historical [52, 55, 106] and contemporary [6, 59, 65, 67, 68, 108, 115, 125, 127, 129-131] hybridization have been reported for many pairs of oaks, nevertheless, species remain distinguishable in morphology, physiology and ecology. The possibility of inter-crossing between oaks has raised the question of biological species concept [89] and led some authors to introduce the concept of ecological species, multispecies or species complexes [138]. Species complex represents a group of species linked by evolutionary adjusted gene flow as an evolving force operating within this complex, however eventually not related to the speciation process [103].

Sessile oak, *Q. petraea* (Matt.) Liebl. and pedunculate oak, *Q. robur* L. give an example of the two oaks that intercross freely and show a range of morphological and genetic overlaps, which led some authors to consider them as two ecotypes within one species of *Q. robur* [75, 114]. Their taxonomical entities were indeed questioned for decades. The *petraea-robur* taxonomy issue, fascinated even Darwin [39] and his contemporaries. The species status has been doubtful in the case of *Q. petraea* and *Q. robur*, as several leaf and fruit morphological characters are traditionally used for identification of the two morphologically different types [79]. Descriptive names – sessile and pedunculate – were given to these two oaks because they differ in length of the peduncle that bears the cup (short vs. long). Although these two predominant European oaks are largely sympatric [107] and coexist at mixed stands [74, 79], they occupy different ecological niches and show some habitat preferences.

A further peculiarity of these two species is their extremely low genetic differentiation in contrast to clear ecological preferences and

morphological differences for leaf and fruit structures observed throughout sympatric populations across Europe. A palette of genetic markers were used in numerous surveys for genetic differences such as rDNA [94, 104, 151], RAPD (random amplified polymorphic DNA) markers [18, 91], AFLP (amplified fragment length polymorphism) markers [31, 33, 73, 88], isozyme markers [8, 146], RDA (representational difference analysis) markers [150] and microsatellites [88, 93]. Genomes of the two species contain certain more or less important differences in allelic frequencies of nuclear markers [18, 33, 93, 117, 146], but until today no species-specific locus has been found for species identification.

Since *Q. petraea* and *Q. robur* show a high degree of allele sharing and no species-specific sequences, interpretation is that the two species are confined to ongoing hybridization including a certain level of gene flow and introgression [107]. Both controlled crosses and mating system analyses in natural conditions have shown that due to extensive pollen flow, interspecific crossings occur in nature [6, 7, 108, 111, 127, 129, 130], however unidirectionally-ovules of *Q. robur* being preferentially fertilized by *Q. petraea* pollen [1, 6, 127]. Differences in flowering time between the two oaks cause the asymmetric gene flow, which probably constitutes the prezygotic isolation of the two species [1]. Although there is no systematic study of the amount of hybrids in the natural stands, several studies have shown that hybridization rates and the number of morphological intermediates are actually low, i.e., include less than 5% of trees [2, 5, 33, 48]. In addition, even if the viable hybrids between *Q. petraea* and *Q. robur* are formed, their survival is questionable when confined to natural conditions under competition between seedlings of the pure species. Also, fertility of morphological intermediates is significantly lower than pure species [112, 113]. All these observations suggest a postzygotic breeding barrier between the two sympatric oaks.

Additionally, several molecular studies have demonstrated that contemporary interspecific gene flow is not so important between *Q. petraea* and *Q. robur*, otherwise the two interfertile closely related species with their overlapping distributions would have completely lost their separate entities by now. A clear differentiation of the gene pool in the two species has been achieved using microsatellite and AFLP markers [33, 93]. Both multilocus markers succeeded in clustering populations of the same species together, with the strong bootstrap support (value of 100%), as well as to assign the individual trees to the “right” species [33,

93]. A somewhat weaker differentiation was revealed using some isozymes and RAPDs [58, 146].

A debate still continues as to whether the lack of nuclear differentiation between the sessile and pedunculate oaks represents a recurrent gene flow or a shared ancestral polymorphism [82, 95, 96]. One group considers that the majority of the genome is shared between the two species due to introgressive gene flow, and that species integrity is secured by selection operating on a subset of loci [82, 124]. This hypothesis is based on a model where genes represent units of species differentiation, and genomes of two interfertile species are mosaics of impermeable (accumulate differences in response to selection) and permeable (share the same alleles due to introgression) regions to gene flow [144]. Therefore, “particular combinations of adaptive genes would remain associated, either due to fertility consequences or ecological selection pressures” [78, 83]. On the other hand, another set of researchers considers the evolutionary importance of the gene flow in these two oaks to be minor [95, 96]. They have used microsatellites, markers of choice in studies of hybridization and introgression, to study the amount and importance of the gene flow between *Q. petraea* and *Q. robur* [95]. They analyzed populations of both species in mixed stands from six geographical regions, covering their whole European distributional range. If the hybridization occurred frequently between the two species, gene flow would have reduced differentiation between the two oaks in close proximity. On the contrary, differentiation was higher for distantly located pairs of species. Yet, authors reported a similar level of interspecific differentiation in both cases, irrespective of the geographical origin of the two populations analyzed. They suggest that the reason why genomes of the two white European oaks differ only in allelic frequencies and not in species-specific alleles lies in the recent speciation event, i.e., such a situation would represent shared ancestral polymorphisms.

Analyses of rDNA in genomes of *Q. robur* and *Q. petraea* revealed no differences in spacer regions (ITS1-ITS2 and IGS). Spacers normally represent the rapidly evolving part of a rDNA cluster in contrast to conserved 18S, 5.8S and 26S rRNA genes coding for corresponding rRNA molecules. rDNA spacers are usually used for resolving specific phylogenetic questions across different phylogenetic spectra, i.e., ITS is applied for generic and subgeneric levels and IGS (intergenic spacer region) is used for suprageneric level [3]. Indeed, ITS spacers have been

proved as excellent molecular markers to trace evolutionary relationships between closely related species [30], and their utility in molecular systematics of oaks has been demonstrated [13, 118]. However, *Q. robur* and *Q. petraea* share the same ITS sequences, most probably because the split between them was too recent for ITS sequences to be differentiated [94], unlike some other white oaks where rDNA sequences were differentiated between species despite ongoing hybridization [12, 13, 142]. Sequencing of 5.8S and ITS2 regions revealed no species-specificity—three divergent rDNA families appeared in both genomes, presumably due to some ancient hybridization events that predated speciation and brought diverged rDNA families together [94]. Another rDNA spacer, IGS, has functional importance apart from the utility in phylogenetic studies, i.e., structural elements of IGS have a role in achieving the appropriate transcription level of rRNA genes. IGS can vary in length between closely related species, as well as among and within individuals of the same species and the length variants are due to variation in number of internal subrepeats present in IGS. The structure of IGS can be species-specific [136] or can be very similar for closely related species [19]. PCR (polymerase chain reaction) amplification of IGS region in genomes of *Q. robur* and *Q. petraea* and Southern blott revealed that the same spacer length variants were shared by genomes of both species. Moreover, sequence and structural organization of the predominant IGS spacer length variant were identical in both species [151].

Differences in allele frequency of some polymorphic markers can clearly distinguish these two species, favouring their treatment as “good species”. The nuclear differentiation between *Q. robur* and *Q. petraea* appears to be extremely insignificant, which is smaller than often found between closely related species [84]. The overall nucleotide divergence is 0.5% in overall genome, but it may augment to 3% in the most differentiated chromosome regions [18]. Poor genetic differentiation between the two species could be observed by a screening process of 2,800 RAPD fragments, only 2% of which exhibited significant allelic frequency difference [18]. Further exploration is needed in order to fully understand how these two interfertile closely related oaks maintain their distinct species status. Orr [101], in his review has pointed out that genetics of species differences can be “simple” or “complex” in several different ways, which suggested that these do not follow any regularity in different organisms. He has given a comprehensive view of species

differences that involve one or more (up to 19) genes. The most interesting data are that of the two species of monkeyflowers, *Mimulus lewisii* and *Mimulus cardinalis*. More than 80% of the phenotypic differences among F2 hybrids between these two species involve only a single QTL that affects carotenoid concentration.

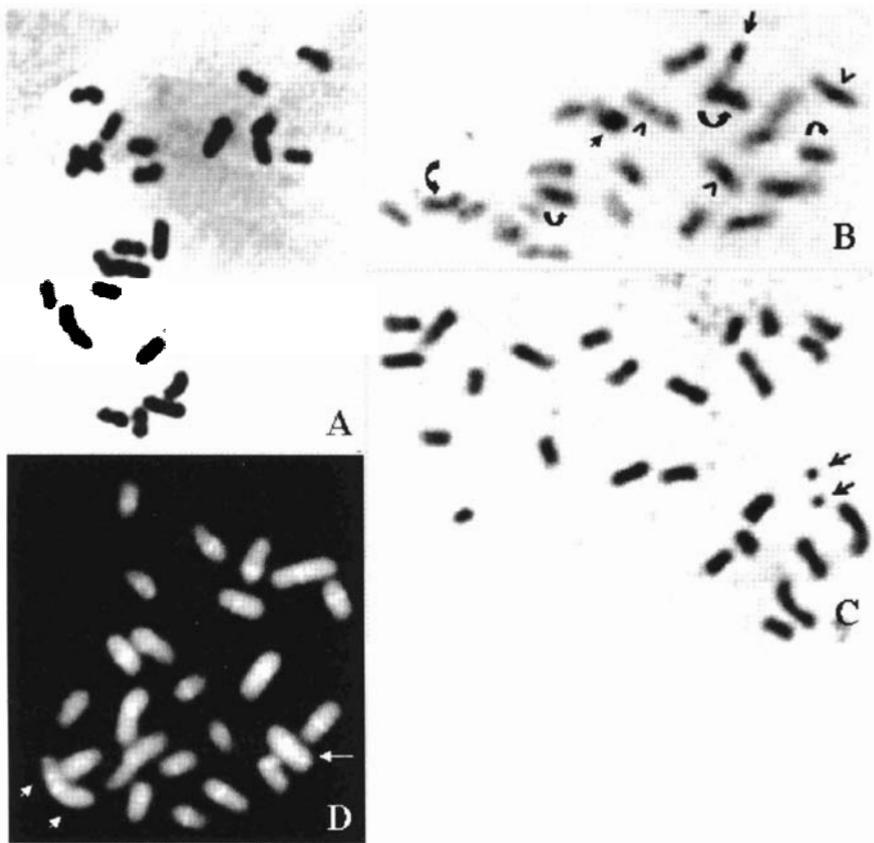
The related questions include: Are genomic regions involved in species differentiation rare (correspond to single copy genes, i.e., coding part of the genome) or frequent (correspond to repetitive sequences, i.e., noncoding part of the genome)? Are discriminative genomic regions distributed all over the genome or confined to a few hot spots? Analyses of RAPD fragments [16], SCAR (sequence characterized amplified regions) fragments [17, 18] and microsatellites [95, 96] revealed that informative genomic regions belong to highly polymorphic parts of the *petraea* and *robur* genomes. Genome-wide distribution of such interspecific discriminative regions is demonstrated when different experimental strategies are applied. Joint analyses of all 20 microsatellite loci, as well as analysis of individual microsatellite loci, have shown important differentiation between the two genomes and locate informative regions over six out of 12 haploid chromosomes [95]. Five quantitative trait loci (QTLs) that control species-discriminant leaf morphological traits were distributed over all linkage groups [116]. Finely, systematic genome scan based on 389 different markers (isozymes, AFLPs, SCARs, SNPs and microsatellites), for which allelic frequencies were estimated in pairs of populations of sessile and pedunculate oaks across their European distributional range, confirmed that genomic regions involved in interspecific differentiation are distributed all over the genome [124].

It is proposed that discriminative sequences revealed so far correspond to single copy genes involved in adaptation to the two different ecological niches of the two *Quercus* species [18]. The presumption that genomic regions involved in species differentiation are preferentially located in functional regions of the genome is further supported by Scotti-Saintagne [124] who used EcoRI/MseI and PstI/MseI enzyme combinations for generating AFLP markers. Namely, AFLP fragments digested by PstI, which cuts preferentially within the coding regions that are expressed, exhibited larger species differentiation than fragments digested by EcoRI, which cuts rather randomly in the genome. Ongoing surveys for candidate genes responsible for species differentiation are actually based on gene expression studies.

## KARYOTYPE EVOLUTION IN FAMILY FAGACEAE (AND IN GENUS QUERCUS IN PARTICULAR): CHROMOSOME NUMBER, KARYOTYPIC FEATURES AND GENOME SIZE

In general, groups of woody plants, especially tropical ones, have a tendency to be uniform in chromosome number in subgeneric, generic or even familial level. Relatively new but scarce data is available for genus *Quercus* [38, 98, 149], *Castanopsis*, *Lithocarpus* [32] and *Castanea* (Y.K. Lim, Personal communication). Diploid number of 24 chromosomes seems to be constant across the family Fagaceae, however, no extensive cytogenetic data are available due to technical difficulties in obtaining good chromosome preparations of woody plants. This chromosome number is considered to have a paleo-polyploid origin. Ancient polyploidy among woody plants, which has resulted in today's high basic chromosome number, i.e.,  $x = 12, 13$  or  $14$ , supposedly occurred in the Cretaceous and early part of the Tertiary period. Paleobotanical and molecular phylogenetic evidences suggest that all major oak lineages have evolved locally at middle latitudes in the Tertiary period [4]. Although the species of subgenus *Quercus* could be easily distinguished morphologically from the species of strictly Southeast Asian subgenus *Cyclobalanopsis*, and even if there were two initial centres of diversification for subgenus *Quercus* (America and Asia), the chromosome number is the same for all of the species analyzed so far, belonging to both subgenera, and/or American and strictly Euroasian sections. Therefore, extensive polyploidy must have happened very early in the evolution of the genus *Quercus*. This could be the case of other genera of the family Fagaceae, too, as supposed for woody angiosperms in general [126].

The tree of *Q. lenticellatus*, a species from the northern Thailand belonging to the subgenus *Cyclobalanopsis*, has  $2n = 2x = 14$ , which could serve as a rare contemporary evidence for the hypothetical original basic chromosome number of 7 in the genus *Quercus* [32]. Groups of oak species with many plesiomorphic characters (relicts of ancestral types) are abundant in Eurasia. In particular, in Southeast Asia and Indo-China one can find the largest number of the most primitive forms of *Castanopsis*, *Lithocarpus* and *Quercus* (subgenus *Cyclobalanopsis*) species compared to other regions of the world, even if characters of ancestral types (e.g., hermaphroditic flowers) are also described in American [133, 135] as well



**Fig. 2.** Orcein staining (A), Giemsa C-banding (B, C) and DAPI banding (D) of chromosomes of *Q. petraea* (A, B, C) and *Q. robur* (D). Note uniform morphology, similar size of the oak chromosomes, and hardly distinguishable centromere positions after orcein staining (A). In B, arrowheads indicate centromeric constitutive heterochromatin of some chromosomes, and arrows show telomeric constitutive heterochromatin corresponding to GC-rich 18S rDNA locus on the only subtelocentric chromosome pair of the chromosome complement. Curved arrows indicate several chromosomes that contain potentially more heterochromatin than just at centromeric position. Note unequal condensation of chromosomes. Arrows in C mark detached satellites from the NOR-bearing homologous chromosomes. Arrows in D indicate on intercalary DAPI bands, i.e., constitutive heterochromatin which is AT rich. Magnification  $\times 1300$  (A, B and C from [149]).

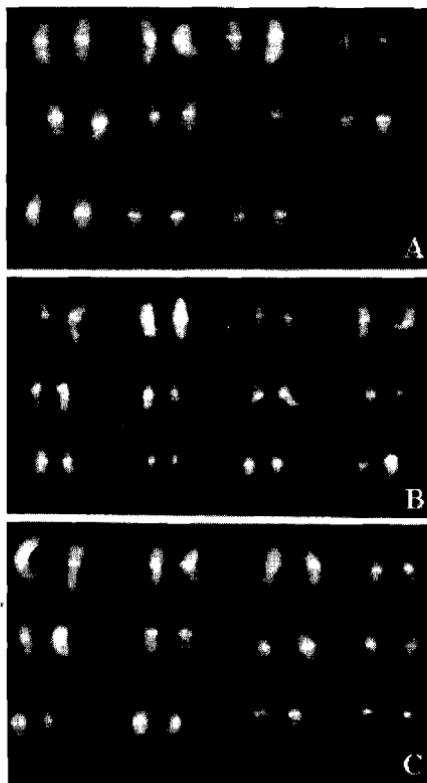
as European oaks [20, 21, 120]. Stebbins [126] had suggested that the "present rarity of species having low chromosome number could be due to the differential extinction of primitive diploids".

In contrast to ancient polyploidy, supposed to be involved in the creation of the high basic chromosome number in woody angiosperms, modern polyploidy is very rare, which is explained by ecological and historical factors of woody plants [126]. Throughout the Tertiary period woody plants did not have to cope with drastic environmental changes, and only when the climatic and edaphic conditions have become similar to those in their previous habitats, they have colonized new regions. Therefore, occasional modern polyploid individuals had no adaptive superiority over diploid ones. Rare cases of modern polyploidy are reported in Fagaceae; in *Castanopsis* and *Lithocarpus* species of natural populations in Thailand [32], in *Trigonobalanus* species [64], *Fagus* and *Quercus* species [25, 70]. Polyploid oak trees are described in the Voronezh region in south central Russia, in the oak stand that provides optimal edaphic conditions for these trees. These triploids ( $2n = 3x = 36$ ) of pedunculate oak were more than hundred years old and were of gigantic height, weak fertility and exhibited unusual morphological and anatomic features [25]. The origin of such triploidy in *Quercus* is explained by the mechanism of  $2n$  pollen formation due to parallel spindles in meiosis. Two *Castanopsis* mixoploid trees were also described, exhibiting diploid and hyperaneuploid, in addition to triploid, chromosome number in the same individual. These trees produced no flowers or acorns [32]. Polyploidy recorded in *Trigonobalanus verticillata* ( $2x = 44$ ) can be considered as an antique character together with other relict features that sign *Trigonobalanus*, the genus consisting of the most antique and relictual species of all in the family Fagaceae [87]. However, all these cases are exceptional from the general rule of a low frequency of the modern polyploidy for angiosperm woody plants.

Supernumerary chromosomes are found in *Q. petraea* [148]. Such supernumerary or B-chromosomes are found in numerous plant species. Bedi [11] reported even 3.9% out of 862 woody species of known chromosome number to have B-chromosomes. The number varies between populations within species, among individuals, and even between tissues and cells. One 2 or 3 supernumeraries have been recorded in root-tip cells of sessile oak. The number varied between individuals of the population analyzed, and within the same root-tip [148]. Usually, the size of B-chromosomes distinguishes them from A-chromosomes of the normal complement, however, in oak supernumerary chromosomes were of similar size to A-chromosomes. Ohri and Ahuja [99] reported one B-chromosome, much smaller than the smallest A-

chromosome, in karyotypes of *Q. petraea*, *Q. robur* and *Q. rubra*. The current author suggests that the B-chromosome reported by these authors actually represents satellites detached from the NOR-bearing (nucleolus organizer region) chromosomes. Secondary constriction in oak subtelocentric NOR-bearing chromosome is fragile and cytogenetic manipulation often results in detachment of the satellite (Fig. 2C). The observations made by the author of this chapter of 1, 2 or 3 supernumeraries in the French population of *Q. petraea* is the first report of this type and it is worth to explore further the ecological meaning of the presence of extra chromosomes in this population. Widespread occurrence of supernumerary chromosomes in plants and animals argue against the common opinion that they simply represent nuclear "parasites", and favour the opinion that they confer some fitness upon individuals and populations under certain environmental conditions [110].

In certain groups of organisms, the rate of chromosome evolution is correlated with the rate of speciation. According to White [141] karyotypic change occurred in over 90% of the speciation events and the initiation of species divergence is most frequently caused by structural chromosomal rearrangement. In fact, many closely related species can be distinguished by obvious karyotypic differences, and simple chromosome features such as number, size and shape are useful in karyotype analyses of many plant groups. In oaks, however, they reveal little information since the chromosomes are of similar morphology and small size – they range from 2.49  $\mu\text{m}$  for the smallest chromosome to 6.26  $\mu\text{m}$  for the largest chromosome of the complement [99] – hence they are difficult to analyze (Figs. 2A, B, C). Obvious chromosome change has played a minor role in karyotype evolution and speciation in the genus *Quercus*. Classical cytogenetic techniques, such as Giemsa C-banding, revealed very similar, if not identical, karyotypes for the two closely related European white oaks *Q. robur* and *Q. petraea* (subgenus *Quercus*, section *Quercus* s.s., sensu Manos, Doyle and Nixon [86]), Euroasian oak *Q. cerris* (subgenus *Quercus*, section *Cerris*, sensu Manos, Doyle and Nixon [86]) and American red oak *Q. rubra* (subgenus *Quercus*, section *Lobatae*, sensu Manos, Doyle and Nixon [86]). Their karyotypes consist of submetacentric (6) to metacentric (5) chromosomes, and only one subtelocentric chromosome [99, 149] (Fig. 3). According to Stebbins' classification [126] the karyotypes belong to 2b category, which indicates a moderate degree of asymmetry. Very similar karyotypes were also reported for nine *Quercus* species from China, characterized again with mostly metacentric and submetacentric (0, 2 or 4, depending on species)



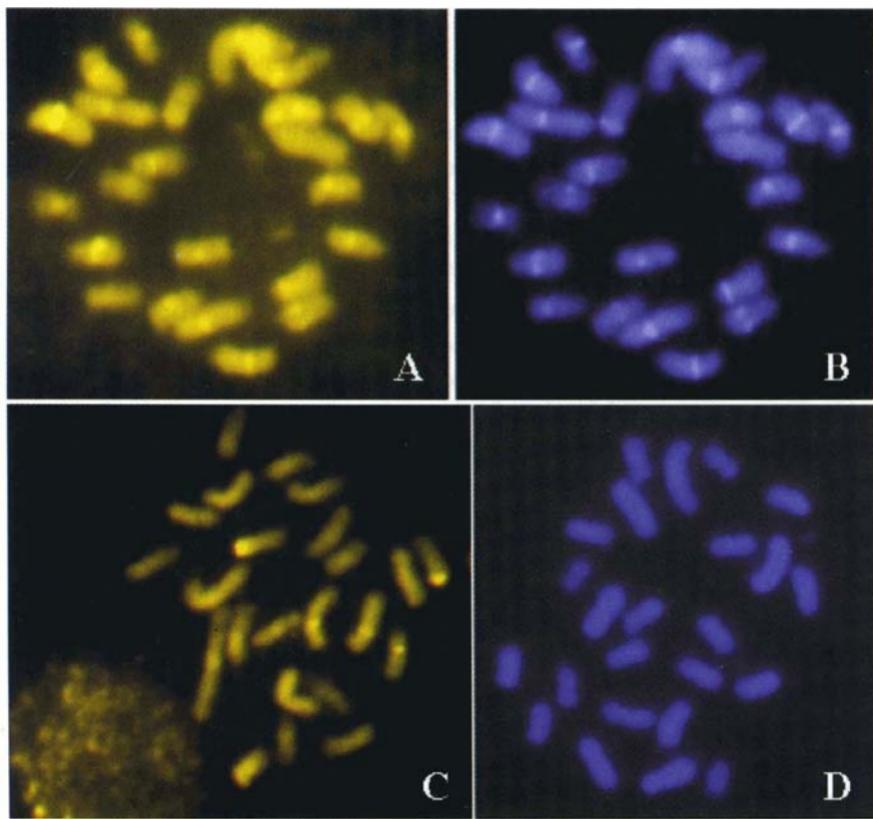
**Fig. 3.** Karyograms of *Q. robur* (A), *Q. cerris* (B) and *Q. pubescens* (C) obtained after chromosome pairing on the basis of the relative chromosome length and arm ratio, where DAPI bands were used as position of centromeres. Magnification x2300 (A, B, C from [149]).

chromosomes [28]. Only karyotypic differences involve the number/position of the SCs and absence of one subtelocentric chromosome in karyotypes of oaks from China. Almost constant karyotypes within the genus *Quercus* fit in the general view of chromosome evolution and speciation in woody genera [126]. Karyotype similarity within *Quercus* genus is by no means among the reasons for successful chromosome pairing in meiosis and proper segregation into functional gametes through interspecific hybridization of many pairs of oaks, as mentioned earlier.

Although karyotypes were established for several oak species [99, 149], corresponding chromosome pairs can not be distinguished between them only on the basis of Giemsa staining, due to homogeneous chromosome morphology, similar size and unequal chromosome

condensation. Even more advanced molecular-cytogenetic techniques such as fluorochrome banding, e.g., 4,6-diamidino-2-phenylindole (DAPI) Chromomycin 3 (CMA3), and fluorescence *in situ* hybridization (FISH), revealed only centromere and NOR positions, and provided just three markers, corresponding to rDNA loci, for unambiguous chromosome identification [149] (Figs. 4, 5E). Surprisingly, 5S and 18S-5.8S-26S (18S) rDNA loci were present in the same number and position in karyotypes of 11 European and American oaks [149]. These 11 oaks embraced three sections of subgenus *Quercus* (missing the fourth, i.e., *Protobalanus*, a sister group of *Quercus* s.s.). Eurasian plesiomorphic section *Cerris* *sensu* Camus (and according to Manos, Doyle and Nixon [86]) were represented by deciduous and evergreen species *Q. cerris* L., *Q. coccifera* L., *Q. ilex* L. and *Q. suber* L., the section *Quercus* s.s. (white oaks) were represented with *Q. michauxii* Nutt., *Q. petraea* (Matt.) Liebl., *Q. pubescens* Willd., *Q. robur* L. and *Q. virginiana* Ten. – the species with a majority of derived characters – and the section *Lobatae* (red oaks), characterized with the most morphological plesiomorphies, were represented by *Q. rubra* L. and *Q. palustris* Muench. The position of these 11 species, surveyed cytogenetically [149], on modified version of the phylogenetic tree constructed by Manos, Doyle and Nixon [86], is presented in Fig. 1. Interestingly, even if ITS+5.8S rDNA sequences have clearly resolved the clade corresponding to Eurasian section *Cerris* and the clade formed by species of the sections *Lobatae*, *Protobalanus* and *Quercus* s.s. on the phylogenetic tree [86], no differences were found on chromosomal rDNA maps of these two groups of species [149]. In all the 11 oaks analyzed there were always two 18S rDNA loci – the major locus positioned on the short arm of subtelocentric chromosome, and minor paracentromeric locus on one medium-sized metacentric chromosome – and only one 5S rDNA locus positioned at the paracentromeric region of second largest submetacentric chromosome (Figs. 5A, B). The identical rDNA maps of the oak species from three different sections is an unusual feature if one recalls the fact that the number/position of rDNA loci is usually species-specific and as such, these loci serve as markers in exploring karyo-evolutionary trends and/or phylogenetic relationships in groups of species.

More recently, rRNA genes were mapped on three more oak species from Southeast Asia, and different maps were obtained on comparisons with maps of 11 European and American oaks [32]. *Q. kerri* Craib and *Q. lenticellatus* Barnett possessed two 18S and one 5S rDNA loci, like all



**Fig. 4.** Chromosomes of *Q. pubescens* (A, B) and *Q. coccifera* (C, D) after fluorochrome banding with CMA (A, C) and DAPI (B, D). Note the difference in banding patterns between the two species. CMA and DAPI bands are co-localized at centromeric regions of all chromosomes in karyotype of *Q. pubescens* (A, B). In karyotype of *Q. coccifera* (C, D) there is no specific DAPI band, while CMA band are restricted to rDNA loci. Magnification  $\times 2300$  (A, B from [149]).

oaks studied so far, however, rDNA loci were positioned differently in these two species. In *Q. kerri* two 18S rDNA loci were located terminally on the long arm of one submetacentric and on one metacentric chromosomes, whilst in *Q. lenticellatus* one locus was positioned terminally on the metacentric chromosome and another locus was positioned intercalary (but quite close to centromere) on another metacentric chromosome. The only 5S rDNA locus was positioned on the metacentric chromosome bearing 18S rDNA locus in both species, however, in *Q. lenticellatus* it was juxtaposed to intercalary 18 rDNA locus.

*Q. brandisianus* Kurz displayed one 5S and one 18S rDNA locus positioned on different chromosomes, although the exact position is not reported [32]. The occurrence of only one 18S rDNA locus in this species is doubtful since no figure is shown. Minor 18S rDNA locus was sometimes difficult to map (to chromosomes of other oak species). Nevertheless, these new data have revealed some differences in karyotypes of oaks from subgenus *Quercus* studied by the current author and co-workers [149] and oaks from subgenus *Cyclobalanopsis* studied by others [32]; the oaks of these two subgenera are phylogenetically separated in two different sister clades [86] (Fig. 1). The species from these two subgenera are also distinguished by several flower characters which represent the primary source of evolutionary trends [97]. Oak species from subgenus *Cyclobalanopsis* display many plesiomorphic characters and this is the difference between them and the oak species from subgenus *Quercus*. Thus, it seems that rDNA loci are useful cytogenetic markers in oaks on subgeneric level (but not on sectional or species level).

It is worth to notice that karyotypes of three *Castanea* species (*C. mollisima*, *C. sativa*, and *C. crenata*) are very similar to those of *Quercus* species – they consist only of metacentric and submetacentric chromosomes continuously decreasing in size. In addition, rDNA maps are identical to those of oak species belonging to subgenus *Quercus* (Y. K. Lim, Personal communication). The closest relationship of *Quercus* to castaneoid genera within the family Fagaceae has been revealed by DNA sequence data [87].

Genome size is a parameter useful in various aspects of cell and molecular biology, ecology, systematics and evolution of species. Estimates of genome size in family Fagaceae are restricted only to several species of the genus *Quercus*, to *Fagus sylvatica* which contain 1.11 pg/2C [57] and *Castanea sativa* which contain 1.62 pg/2C (S Šiljak-Yakovlev and S Brown, Personal communication). As can be presumed from the constant chromosome number and very similar karyotypes in oaks, genome size is constant between the *Quercus* species studied so far. According to the authors, measurements, using flow cytometry, the values of DNA content in seven oaks were as follows: 1.88 pg/2C for *Q. petraea* and *Q. robur*, 1.91 pg/2C for *Q. cerris*, *Q. pubescens* and *Q. suber*, and 2.00 pg/2C for *Q. ilex* and *Q. coccifera*. Variation of DNA content was recorded in *Q. petraea* (1.88 to 1.91 pg/2C) and this was due to supernumerary chromosomes found in one population [148]. Somewhat lower values of DNA content of the same and different oak species were obtained using Feulgen microdensitometry [14, 62, 99, 100], and the differences could be

attributed to the different techniques used. Higher DNA content for two evergreen oaks, *Q. ilex* and *Q. coccifera*, up to 6% comparing to *Q. cerris*, *Q. petraea*, *Q. pubescens*, *Q. robur* and *Q. suber* is revealed by the author's estimates. Previous morphological classifications grouped these two species separately from other species forming *Cerris* group (i.e., *Q. cerris* and *Q. suber*) [27, 97, 123]. The phylogenetic tree based on cpDNA restriction maps separated the *Cerris* group and *Ilex* group [86], and restriction site analyses of rDNA region distinguished *Q. ilex* and *Q. coccifera* from other four Mediterranean oaks (*Q. cerris*, *Q. macrolepis*, *Q. trojana* and *Q. suber* [12]). Also, the difference in heterochromatin content is presumed after cytogenetic analyses [149] (Fig. 4), and positive correlation between genome size and heterochromatin content is well established. Genome size can be affected by life cycle and ecological factors, i.e., longer cell cycle and larger genome size are positively correlated, and are related to adaptations on stress in environment such as aridity, low temperatures or high altitudes [71, 109]. In our panel of species, there are species growing in strictly continental (*Q. petraea* and *Q. robur*), sub-Mediterranean (*Q. cerris* and *Q. pubescens*) and Mediterranean climate (*Q. ilex* and *Q. coccifera*). The trend of augmentation of DNA content can be noticed, i.e., increasing from continental to Mediterranean climate. Nevertheless, genome size, like karyotype analyses, had no important informative value for deriving at taxonomic relations in *Quercus* in contrast to the same parameters of many annual and perennial plants which can be used in resolving systematic problems [29, 137, 145].

## **MOLECULAR ORGANIZATION OF THE OAK GENOMES**

In general, plant genomes consist of a high proportion of repetitive DNA sequences that include tandemly, mostly highly repeated DNAs (satellites, microsatellites and minisatellites), and dispersed, moderately repeated DNAs that are mobile (mostly retrotransposons). Usually, larger genomes have a higher percentage of repetitive DNAs than smaller ones. Some plant genomes are characterized by a high amount of satellite DNA such as *Arabidopsis*, *Anemone*, *Triticum* and other Gramineae [15, 76, 90], and some contain a substantial amount of retroelements such as maize, barley or lily [60, 119, 140]. In evolutionary terms, repetitive noncoding sequences represent a fast evolving part of eukaryotic genomes, and as such they are often used in studies of genome evolution and systematics, especially satellite DNA.

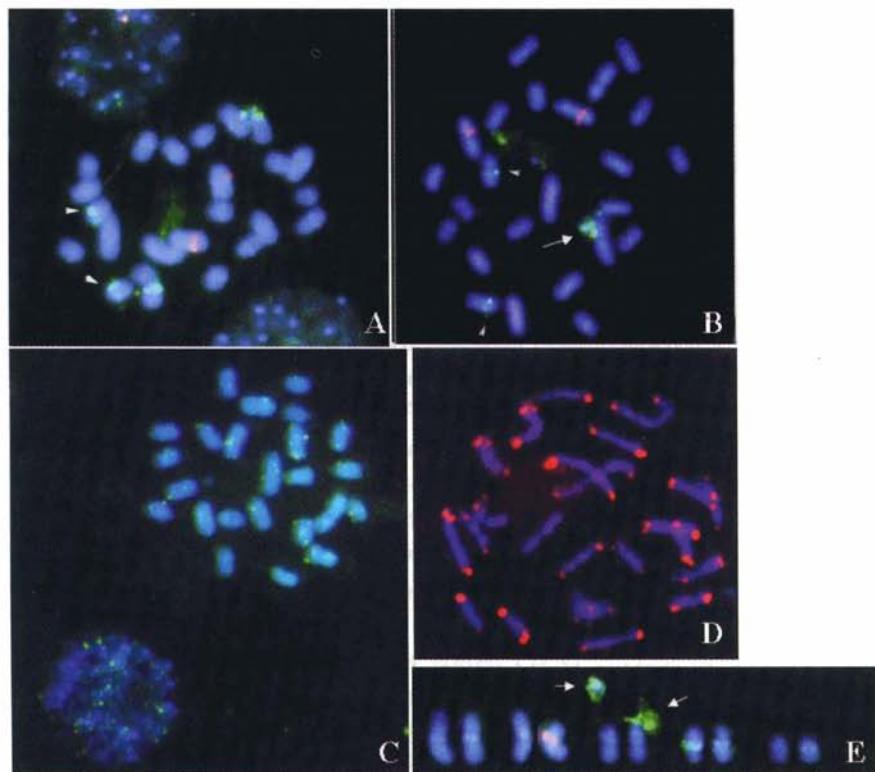


Fig. 5. Simultaneous FISH with 18S (green signals, digoxigenin-anti-digoxigenin, coupled with FITC) and 5S rDNA (red signals, fluoro-red) probes to chromosomes of *Q. petraea* (A, E) and *Q. michauxii* (B). Green signals represent two 18S rDNA loci; the major locus is located either terminally on the only subtelocentric chromosome (A, E) or on detached satellites (B, E, arrows), the minor paracentromeric locus is indicated with arrowheads in A and B. The part of the major locus is sometimes dispersed and can be seen as a hazy and/or punctuate hybridization signal (A, B, E). FISH using RDA fragment as a probe (labelled with digoxigenin-anti-digoxigenin, coupled with FITC). Green signals correspond to dispersed distribution of retrotransposon-like sequence on almost all metaphase chromosomes of *Q. robur* (compare with interphase nucleus) (C). *Arabidopsis*-type of telomeres (TTTAGGG) hybridized on chromosomes of *Q. petraea* using FISH. The telomeric probe was labelled with Cy3 (D). Five marker elements in oak karyotypes: the biggest and the smallest chromosomes, and chromosomes bearing 18S and 5S rDNA loci (E). In all figures chromosomes are counterstained with DAPI. Magnification  $\times 1600$  in A, B, C, D and  $\times 2300$  in E (A, B, E from [149], C from [150]).

Satellite DNA resides in constitutive heterochromatin that can be found at two basic chromosome areas – near centromeres and telomeres, and at satellites of NOR-bearing chromosomes. In karyotypes containing a large quantity of heterochromatin, it can also be found at intercalary chromosome regions [126]. The cytogenetic technique of choice for revealing the amount and distribution of constitutive heterochromatin is Giemsa C-banding, and C-banding patterns are also useful in karyotyping. This technique produced C-banded karyotypes of poor resolution in four oak species (*Q. cerris*, *Q. petraea*, *Q. pubescens* and *Q. robur*). The constitutive heterochromatin is mostly distributed near centromeres of all chromosomes and associated with 5S and 18S rDNA sites (Fig. 3). The same pattern was obtained after CMA banding, which stains GC-rich heterochromatin, and DAPI banding, which stains AT-rich heterochromatin. The CMA/DAPI banding pattern was the same in respect to chromosomal position of the bands in 11 species studied (referred to earlier in this chapter), however, a reduced number of bands appeared on the chromosomes of *Q. coccifera*, *Q. ilex*, *Q. palustris*, *Q. suber* and *Q. virginiana*, i.e., only CMA bands corresponding to rDNA loci were present, and no specific DAPI bands were observed at the centromeric regions [149] (Fig. 4C, D). These could represent either quantitative differences in the heterochromatin between different oaks belonging to different sections, or technical difficulties in chromosome banding, i.e., no differences were detected in the GC content between 11 species after assessment by flow cytometry [148].

In general, it can be stated that genomes of the oak species studied are characterized by a low amount of constitutive heterochromatin. The largest fraction was associated with 18S rDNA sites (i.e., NORs) and was exclusively GC rich, while the remaining centromeric heterochromatin consisted of both GC and AT intermingled repetitive sequences [149] (Fig. 4). Ohri and Ahuja [99] also reported heterochromatin at centromeric positions in all chromosomes of *Q. petraea*, *Q. robur* and *Q. rubra*, nevertheless they described intercalary and telomeric C-bands in some chromosomes of the complements. Rare and very thin intercalary C- and DAPI bands were seen by our proper efforts but the bands were of very poor resolution (Figs. 2B, D). Several entire heterochromatic chromosome arms were recorded in *Q. petraea*, *Q. robur* and *Q. rubra* [99]. Some chromosomes could have more heterochromatin than only at the centromeric position (Fig. 2B, curved arrows), albeit an unsatisfactory resolution of C-banding pattern was obtained by their and our proper

efforts to be sure in this statement. Support to the opinion that oak genomes contain a low amount of constitutive heterochromatin comes from the results of isolation of satellite DNA from *Q. petraea* and *Q. robur* genomes. We did not succeed in obtaining a ladder pattern characteristic for satellite DNA with any of the several endonucleases used for genomic DNA restriction. These results could indicate that a very low amount of satellite DNA is present in the oak genomes and that a large quantity of genomic DNA would be needed for its isolation (Zoldos, unpublished data).

Unfortunately, oaks are still poorly explored in terms of molecular structure of their genomes. It is possible that oak genomes contain much more dispersed sequences than tandemly repeated ones. In general, retrotransposons are ubiquitous sequences of the plant genomes, but the number of different families and copy number within a family of the mobile elements differ greatly from species to species. When genomes of two oaks of different systematic positions, *Q. robur* and *Q. suber*, were compared by subtractive cloning using the RDA technique, about 60% of obtained clones were *Q. robur* specific and were characterized as retrotransposon sequences after homology search in GenBank, and according to their molecular and chromosomal organization [150]. Even if this was only a very small representation of the genome, a random selection of clones, all belonging to transposable elements, could indicate that this type of sequences prevails in regions that diversify oak genomes. FISH using RDA sequences as probes revealed their dispersed distribution throughout the entire length of almost all chromosomes of pedunculate oak (Fig. 5C).

Oak chromosomes are also poorly analyzed by means of molecular-cytogenetic techniques as compared to chromosomes of economically important crops such as rye or wheat [37, 139]. Only rDNA, RDA (retrotransposon-like) and telomeric sequences are mapped on chromosomes of some oak species with the aid of FISH. Here we report for the first time FISH mapping of telomeric sequences TTTAGGG on chromosomes of *Q. petraea* and *Q. robur*. All 24 chromosomes were capped with Arabidopsis-type of telomeres, and the strength of the hybridization signals revealed that a different copy number of telomeric sequences was present on the ends of different chromosomes of the complement (Fig. 5D).

Genetic maps can also add to understanding of molecular organization and structure of plant genomes. The first genetic map in

genus *Quercus* and family Fagaceae as a whole is available for *Q. robur*, and it was constructed by the use of RAPD, SCAR, SSR (simple sequence repeat) isozyme and 5S rDNA markers [9]. Genetic maps are available so far only for a few forest and fruit trees, e.g., *Eucalyptus* [61], *Populus* [23], *Citrus* [26], *Malus* [63] and *Prunus* [41]. The genetic map of pedunculate oak represents one of the most extensive maps – it covers 85% of the genome – compared to other available maps of woody angiosperm species, which all cover less than 80% of the genomes. Reported male and female maps are of similar sizes (922 cM and 893 cM) and represent 12 linkage groups corresponding to 12 haploid oak chromosomes. The physical distance per unit of genetic distance in pedunculate oak amounts to 0.77x10<sup>6</sup> bp/cM. Enrichment of the genetic map of *Q. robur* has recently been done by positioning the 13 coding genes on 10 different chromosomes [108]. These are the first mapped coding genes on linkage maps in the family Fagaceae, and could serve as orthologous markers in comparative mapping studies. Fifteen microsatellite markers from this genetic map were already used in comparative mapping between *Quercus robur* and *Castanea sativa* (Mill.) and orthologous loci were identified, which will be useful as anchor loci in comparative mapping studies within the whole family Fagaceae [10]. The genetic linkage map of *Q. robur* is also important for exploring the molecular organization of genomic regions involved in interspecific differentiation between *Q. robur* and *Q. petraea*, and white oaks in general, as well as for identification of QTLs for adaptive traits. It is already established that four isozyme loci are distributed within the same linkage group (linkage group 2 on the map of *Q. robur*) in these two species [9, 92, 147].

Microsatellites or SSRs are short motifs of 1-5 pb in length that repeats in a genome 5 to 25 times at a certain locus. Larger clusters occur sometimes in the genomes so that they can be identified by FISH, such as in *Hordeum vulgare* [80, 102] or *Vicia faba* [56]. In spite of their great importance as markers in genetic mapping and fingerprinting, in studies of genetic structure of populations and in forensics, little is known about general distribution of microsatellites within plant genomes. About 50 microsatellite loci are discovered in some oak species, mostly in *Q. petraea* and *Q. robur*. About 18 microsatellite loci are reported by Steinkellner et al. [128], 32 loci by Kampfer and et al. [72] and 20 loci by Muir and Schlotterer [95] in *Q. robur*. These microsatellites are shared by genome of *Q. petraea*. Moreover, some of them are present in genomes of the oaks

from the same (*Quercus* s.s.) and other sections [49, 51] and microsatellite primers are usable across the oak species [128] and even across related genera within the family Fagaceae [10]. Thus, microsatellites are regarded as consensus markers across different oak species in the genus *Quercus*. This scarce data add to the observation of high genetic and genomic similarity of oaks of a given botanical section. Due to their genetic characteristics (polymorphic and codominant loci) microsatellites represent markers of choice in studies of genetic diversity within oak populations [88, 129], parentage analysis and reconstruction of the mating system and/or gene flow [43, 95, 96, 130], for verifying/falsifying maternal half-sib relationships from subsets of just a few seeds [81] and for verifying the species status [95].

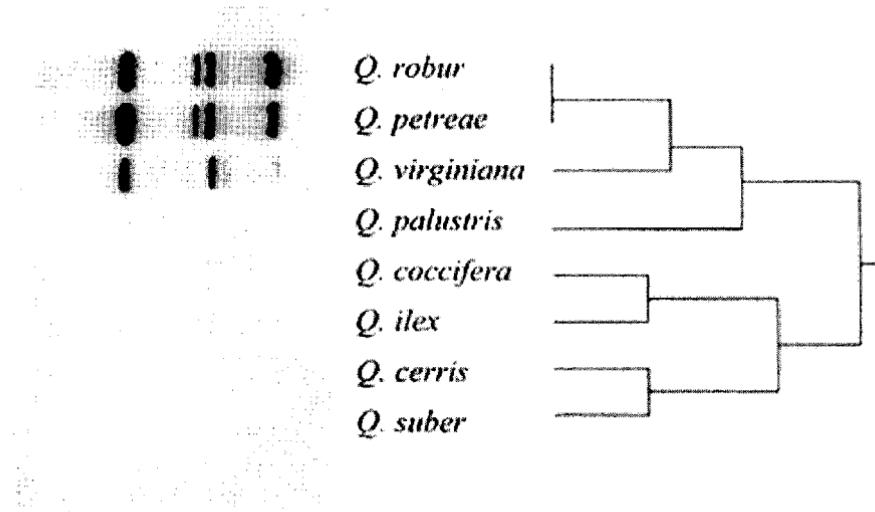
## **GENOME/KARYOTYPE EVOLUTION AND SPECIATION IN QUERCUS**

Ohri and Ahuja [1] have applied Jones' hypothesis [110] on oaks, which states that the presence of one subtelocentric and six metacentric chromosomes in their karyotypes indicates a high degree of structural rearrangements during karyotypic evolution. Taking into account this presumption and bearing in mind the striking similarity of modern karyotypes of 11 oak species (belonging to three different sections of subgenus *Quercus*) it can be inferred that karyotypic evolution is much older than the speciation events, at least on a subgeneric level. It is known that the amount/distribution of constitutive heterochromatin (satellite DNA) as well as number/position of rDNA loci are subjected to most rapid changes within genomes, however, these sequences have had or no little informative value in studies of karyo-evolutionary trends in oaks.

Local evolution of the several major oak lineages at middle latitudes in the Tertiary is well-established [4, 86]. Two initial centres of diversification are proposed for subgenus *Quercus*: three monophyletic sections *Lobatae*, *Protobalanus* and *Quercus* s.s. have evolved in America and diversified into about 300 species (with secondary radiation in Mexico, eastern North America, and Eurasia), and the single monophyletic section *Cerris* has evolved in Asia and diversified in only about 70 species (with secondary radiation in eastern Europe and South-western China) [97]. These two major groups of species within subgenus *Quercus* show similar patterns of morphological evolution that is speculated to represent either morphological convergences or shared

plesiomorphies [86] (and referred herein). Our and others' [99, 148, 149] reports on strikingly similar karyotypes of species from the subgenus *Quercus*, and the hypothesis that karyotypic and genome evolution had occurred before diversification of this group support the scenario of shared plesiomorphies.

Seven RDA sequences subtracted from the genome of *Q. robur* offered some information on the genome evolution in oaks [150]. All seven retrotransposon-like sequences were present in *Q. petraea* genome, too, but not in genomes of *Q. cerris*, *Q. coccifera*, *Q. ilex*, *Q. palustris* and *Q. suber*. Three out of these seven RDA sequences were, however, present in the genome of *Q. virginiana*. According to the molecular phylogeny of Manos, Doyle and Nixon [86], *Q. virginiana* is most closely related to *Q. petraea* and *Q. robur* of all 11 species analyzed in our work: three species belong to the section *Quercus* s.s. while the other five species belong to the sections *Cerris* and *Lobatae* (Fig. 1). These results show that retrotransposon-like sequences are more rapidly evolving fraction in oak genomes than satellite sequences, i.e., constitutive heterochromatin (referred to earlier in this chapter). Seven RDA sequences have probably appeared in *Q. petraea/Q. robur* genome rather recently in the course of



**Fig. 6.** Southern hybridization of genomic DNA (digested with Csp6I) of *Q. robur*, *Q. petraea*, *Q. virginiana*, *Q. palustris*, *Q. coccifera*, *Q. ilex*, *Q. cerris* and *Q. suber* using RDA sequence as a probe. The phylogenetic tree on the right was extracted from the consensus tree constructed by Manos and co-workers [86].

the genome evolution. The three RDA sequences also found in the genome of *Q. virginiana* were probably present in the common ancestor of the three species, and have evolved at a much lower rate than the remaining four RDA sequences present uniquely in *Q. petraea/Q. robur* genome, since the divergence of the two species and *Q. virginiana*. Alternatively, they could have appeared in the genome of the common ancestor later than the remaining four sequences, which are only present in *Q. petraea/Q. robur* genome. Southern hybridization revealed complex molecular organization of RDA sequences (several members of each family) and identical hybridization patterns of all seven sequences to genomic DNA of *Q. petraea* and *Q. robur* [150] (Fig. 6). Complex organization of RDA sequences suggests their very recent diversification within *Q. petraea/Q. robur* genome.

## CONCLUSION

Revealing the molecular events, which form the base for differentiation of the genomes, is pivotal for the understanding of speciation, evolution and adaptation of organisms. Chromosome evolution consists of chromosome changes on different scales (i.e., gross chromosome changes in the number and structure, and fine cryptic chromosome rearrangements) which are produced by continuous qualitative and quantitative DNA sequence changes. However, because of technical limitations, studies of karyotype evolution provide no insight into the underlying molecular mechanisms that caused the observed changes. On the other hand, sequence analyses of only representations of the genomes, i.e., analyses of patches of DNA sequences, are also insufficient to gain the entire picture of genetic architecture of species' dissimilarities. We are still far from understanding the genome organization and evolution of species in the fascinating tree genus *Quercus*. It has been shown in this chapter that gross karyotype changes have probably predated speciation events in the genus, and that genomes show high genetic similarity on a sectional and even subgeneric level. However, molecular-cytogenetic studies have to be carried out on many more species from subgenus *Cyclobalanopsis* since species belonging to this strictly Southeast Asian subgenus differ karyologically from the species from subgenus *Quercus*. Molecular structure and organization of oak genomes are still poorly described, and require considerable effort to be brought to light. However, present data provide hazy but a valuable image that could serve as a guideline for

further research. Also, the first genetic map in Fagaceae, given for *Q. robur*, covers 85% of its genome and includes coding genes besides noncoding genetic markers [9]. This map is one of the most comprehensive maps in woody angiosperm species in general and will be useful for exploring the molecular organization of genomic regions involved in interspecific differentiation between closely related white oaks *Q. robur* and *Q. petraea*. Linked coding genes could also serve as orthologous markers in comparative mapping studies in the family Fagaceae. Given the facts that oaks possess small genomes and that a large number of oak species is of great importance in industries as well as in ecology, we hope that complete genomes of some oak species will be sequenced soon. Comparative analyses of the entire genomes can provide a comprehensive view of large-scale changes in synteny and gene order, as well as an opportunity to investigate whole genomic changes [50].

## Acknowledgment

The author conveys his special thanks go to Ž. Borzan for critical reading of the manuscript, useful suggestions and fruitful discussions, as well as for providing some important literature, which certainly have added value to this chapter. The author is also grateful to S Šiljak-Yakovlev, N Bauer and D. Papeš for critical reading and advice. Sincere thanks to I. Biruš and J. Pužina for technical aid, and L. Zoldoš for language editing.

## References

- [1] Aas G. Kreuzungversuche mit Stiel- und Traubeneichen (*Quercus robur* L. und *Q. petraea* (Matt.) Liebl.). Allgemeine Forst und Jagdzeitung 1991; 162: 141-145.
- [2] Aas G. Taxonomical impact of morphological variation in *Quercus robur* and *Quercus petraea*: a contribution to the hybrid controversy. Ann Sci For 1993; 50: 107-113.
- [3] Alvarez I, Wendel JF. Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogen Evol 2003; 29: 417-434.
- [4] Axelrod DI. Biogeography of oaks in the Arco-Tertiary province. Ann Missouri Bot Gard 1983; 70: 629-657.
- [5] Bacilieri R, Ducoussو A, Kremer A. Genetic, morphological, ecological and phenological differentiation between *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. in a mixed stand of Northwest of France. Sylvae Genet 1995; 44: 1-10.
- [6] Bacilieri R, Ducoussо A, Kremer A. Comparison of morphological characters and molecular markers for the analysis of hybridization in sessile and pedunculate oak. Annales des Sciences Forestieres 1996; 53: 79-91.
- [7] Bacilieri R, Ducoussо A, Petit RJ, Kremer A. Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 1996; 50: 900-908.

- [8] Barreneche T, Bahrman N, Kremer A. Two dimensional gel electrophoresis confirms the low level of genetic differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *For Genet* 1996; 3: 89-92.
- [9] Barreneche T, Bodenes C, Lexer C, et al. A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theor Appl Genet* 1998; 97: 1090-1103.
- [10] Barreneche T, Casasoli M, Russell K, et al. Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* 2004; 108: 558-566.
- [11] Bedi YS, Gill BS, Bir SS. Cytological evolution in the woody taxa of Gamopetalae and Monochlamydeae. *J Cytol and Genet* 1985; 20: 162-203.
- [12] Bellarosa R, Delre V, Schirone B, Maggini F. Ribosomal RNA genes in *Quercus* spp. (Fagaceae). *Pl Syst Evol* 1990; 172: 127-139.
- [13] Bellarosa R, Simeone MC, Papini A, Schirone B. Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. *Mol Phylogen Evol* 2005; 34: 355-370.
- [14] Bennett MD, Smith JB. Nuclear DNA amount in angiosperms. *Philos Trans R Soc Lond B Biol Sci* 1991; 334: 309-345.
- [15] Bennetzen JL. The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology* 1996; 4: 347-353.
- [16] Bodenes C, Laigret F, Kremer A. Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor Appl Genet* 1996; 93: 348-354.
- [17] Bodenes C, Labbe T, Pradere S, Kremer A. General vs. local differentiation between two closely related white oak species. *Mol Ecol* 1997; 6: 713-724.
- [18] Bodenes C, Joandet S, Laigret F, Kremer A. Detection of genomic region differentiating two closely related oak species *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Heredity* 1997; 78: 433-444.
- [19] Borisjuk NV, Davidjuk YM, Kostishin SS, et al. Structural analysis of rDNA in the genus *Nicotiana*. *Plant Mol Biol* 1997; 35: 655-660.
- [20] Borzan Ž. Hermaphroditic, unseasonal flowering in the "green oak", growing in northern Dalmatia, Croatia. *Glas Sum Pokuse* 2000; 37: 425-439.
- [21] Borzan Ž, Stabentheiner E. Biological and taxonomical investigations of some oak species. *Acta Bot Croat* 2002; 61: 135-144.
- [22] Bossema I. Jays and oaks: an eco-ethological study of a symbiosis. *Behaviour* 1979; 70: 1-117.
- [23] Bradshaw HDH, Jr., Villar M, Watson BD, Otto KG, Stewart S, Stettler RF. Molecular genetics of growth and development in populus. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theor Appl Genet* 1994; 89: 167-179.
- [24] Brewer S, Cheddadi R, Beaulieu JL, Reille M. The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management* 2002; 156: 27-48.
- [25] Butorina AK. Cytogenetic study of diploid and spontaneous triploid oaks, *Quercus robur* L. *Ann Sci For* 1993; 50: 144s-150s.

- [26] Cai Q, Guy CL, Moore GA. Extension of the linkage map in *Citrus* using random amplified polymorphoid DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theor Appl Genet* 1994; 89: 606-614.
- [27] Camus A. *Les Chenes: Monographie du Genre Quercus. (et Lithocarpus).* Encyclopedie economique de sylviculture. Vol 6-8. Academie des Sciences, Paris. 1936-1954.
- [28] Cao M, Zhou Z-K. A karyotype analysis of 9 species of the *Quercus* from China. *Guizhou 2000*; 20: 341-346.
- [29] Cerbah M, Coulaud J, Siljak-Yakovlev S. rDNA organisation and evolutionary relationships in the genus *Hypochaeris* (Asteraceae). *J Hered* 1998; 89: 312-318.
- [30] Cerbah M, Souza-Chies T, Jubier MF, Lejeune B, Siljak-Yakovlev S. Molecular phylogeny of the genus *Hypochaeris* using internal transcribed spacers of nuclear rDNA: inference for chromosomal evolution. *Mol Biol Evol* 1998; 15: 345-354.
- [31] Cervera MT, Remington D, Frigerio JM, et al. Improved AFLP analysis of tree species. *Can J For Res* 2000; 30: 1608-1616.
- [32] Chokchaichamnankit P, Anamthawat-Jónsson K, Chulalaksananukul W. Differentiation of Fagaceae species based on chromosomal mapping of 18S-25S and 5S ribosomal genes. *Silvae Genetica* 2007; in press.
- [33] Coart E, Lamote V, De Loose M, Van Bockstaele E, Lootens P, Roldan-Ruiz I. AFLP markers demonstrate local genetic differentiation between two indigenous oak species (*Quercus robur* L. and *Quercus petraea* (Matt.) Liebl.) in Flemish populations. *Theor Appl Genet* 2002; 105: 431-439.
- [34] Craft KJ, Ashley MV, Koenig WD. Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *Amer J Bot* 2002; 89: 1792-1798.
- [35] Crepet WL, Nixon KC. Earliest megafossil evidence of Fagaceae: Phylogenetic and biogeographic implications. *Amer J Bot* 1989; 76: 842-855.
- [36] Crepet WL, Nixon KC. Extinct transitional Fagaceae from the Oligocene Catahoula Formation and their revolutionary significance. *Amer J Bot* 1989; 76: 1493-1505.
- [37] Cuadrado A, Schwarzacher T. The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma* 1999; 107: 587-594.
- [38] D'Emerico S, Bianco P, Medagli P, Schirone B. Karyotype analysis in *Quercus* spp. (Fagaceae). *Silvae Genetica* 1995; 44: 66-70.
- [39] Darwin C. *The origin of species by means of natural selection*, 6th ed. London, UK: John Murray, 1872.
- [40] de Candolle A. Etude sur l'espèce à l'occasion d'une révision de la famille des Cupulifères. *Bibliothèque Universelle et Revue Suisse (Archives des Sciences Physiques et Naturelles)* 1862; 15: 326-365.
- [41] Dirlewanger E, Proner V, Parvery C, Rothn C, Guye A, Moner R. Genetic linkage map of peach (*Prunus persica* (L.) Batsch) using morphological, RFLP, isoenzyme, RAPD and AFLP markers. *Theor Appl Genet* 1998; 97: 888-895.
- [42] Dodd RS, Kashani N. Molecular differentiation and diversity among the Californian red oaks (Fagaceae; *Quercus* section Lobatae). *Theoretical and Applied Genet* 2003; 107: 884-892.
- [43] Dow BD, Ashley MV. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol Ecol* 1996; 5: 615-627.
- [44] Ducousoo A, Michaud H, Lumaret R. Mating system and gene flow in oak species. *Annales des Sciences Forestières* 1993; 50: 91s-106s.

- [45] Dumolin-Lapegue S, Demesure B, Fineschi S, Corre L, Petit RJ. Phylogeographic Structure of White Oaks throughout the European Continent. *Genetics* 1997; 146: 1475-1487.
- [46] Dumolin-Lapegue S, Kremer A, Petit RJ. Are chloroplast and mitochondrial variation species independent in oaks? *Evolution* 1999; 53: 1406-1414.
- [47] Dumolin-Lapegue S, Pemonge M-H, Gielly L, Taberlet P, Petit RJ. Amplification of oak DNA from ancient and modern wood. *Mol Ecol* 1999; 8: 1-8.
- [48] Dupouey JL, Badeau V. Morphological variability of oaks (*Quercus robur* L. *Quercus petraea* (Matt.) Liebl. *Quercus pubescens* (Willd.) in northeastern France: preliminary results. *Ann Sci For* 1993; 50: 35s-40s.
- [49] Dzialuk A, Chybicki I, Burczyk J. PCR Multiplexing of Nuclear Microsatellite Loci in *Quercus* Species. *Plant Mol Biol Reporter* 2005; 23: 121-128.
- [50] Eichler EE, Sankoff D. Structural Dynamics of Eukaryotic Chromosome Evolution. *Science* 2003; 301: 793-797.
- [51] Fernandez JF, Sork VL, Gallego G, Lopez J, Bohorques A, Tohme J. Cross-Amplification of Microsatellite Loci in a Neotropical *Quercus* species and Standardization of DNA Extraction from Mature Leaves Dried in Silica Gel. *Plant Mol Biol Rep* 2000; 18: 397a-397e.
- [54] Ferris C. Ancient history of the common oak. *Tree News* 1996: 12-13.
- [55] Ferris C, King RA, Vainola R, Hewitt GM. Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity* 1998; 80: 584-593.
- [52] Ferris C, Oliver RP, Davy AJ, Hewitt GM. Native oak chloroplasts reveal an ancient divide across Europe. *Mol Ecol* 1993; 2: 337-344.
- [53] Ferris C, Oliver RP, Davy AJ, Hewitt GM. Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Mol Ecol* 1995; 4: 731-738.
- [56] Fuchs J, Strehl S, Brandes A, Schweizer D, Schubert I. Molecular-cytogenetic characterization of the *Vicia faba* genome-heterochromatin differentiation, replication patterns and sequence localization. *Chromosome Res* 1998; 6: 219-230.
- [57] Gallois A, Burrus M, Brown S. Evaluation of the nuclear DNA content and GC percent in four varieties of *Fagus sylvatica* L. *Ann For Sci* 1999; 56: 615-618.
- [58] Gomory D, Yakovlev I, Zhelev P, Jedinakova J, Paule L. Genetic differentiation of oak populations within the *Quercus robur*/*Quercus petraea* complex in Central and Eastern Europe. *Heredity* 2001; 86: 557-563.
- [59] Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama K. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Quercus laurina* (Fagaceae), two Mexican red oaks. *Amer J of Bot* 2004; 91: 401-409.
- [60] Grandbastien MA. Retroelements in higher plants. *Trends Genet* 1992; 8: 103-108.
- [61] Grattapaglia D, Sederoff R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 1994; 137: 1121-1137.
- [62] Greilhuber J. Critical reassessment of DNA content variation in plants. In: Brandham PE, ed. *Kew Chromosome Conference III: Proceedings of the 3rd Chromosome Conference*, held at the Royal Botanic Gardens, Kew, London, UK, 1988: 39-50.
- [63] Hemmat M, Weeden NF, Manganaris AG, Lawson DM. Molecular marker linkage map for apple. *The Journal of Heredity* 1994; 85: 4-11.

- [64] Hou D. Chromosome numbers of *Trigonobalanus verticillata* Forman (Fagaceae). *Acta Bot Neerl* 1971; 20: 543-549.
- [65] Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ. How discrete are oak species? Insight from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* 1997; 51: 747-755.
- [66] Huntley B, Birks HJB. An Atlas of Past and Present Pollen Maps of Europe, 0-13,000 Years Ago. Cambridge: Cambridge University Press, 1983.
- [67] Ishida TA, Hattori AK, Sato H, Kimura MT. Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): insights from morphological traits, amplified fragment length polymorphism markers and leaf minor composition. *Ameri J Bot* 2003; 90: 769-776.
- [68] Jensen J, Hokanson SC, Isebrands JG, Hancock JF. Morphometric variation in oaks of the Apostle Islands in Wisconsin: evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae) *Ameri J Bot* 1993; 80: 1358-1366.
- [69] Johnk N, Siegismund HR. Population stucture and post-glacial migration routes of *Quercus robur* and *Quercus petraea* in Denmark, based on chloroplast DNA analysis. *Scandinavian J For Sci* 1997; 12: 130-137.
- [70] Johnson H. Chromosome numbers of twin plants of *Quercus robur* and *Fagus silvatica*. *Hereditas* 1946; 32: 469-472.
- [71] Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *PNAS* 2000; 97: 6603-6607.
- [72] Kampfer S, Lexer C, Gloessl J, Steinkellner H. Characterization of (GA)<sub>n</sub> microsatellite loci from *Quercus robur*. *Hereditas* 1998; 129: 183-186.
- [73] Kelleher CT, Hodkinson TR, Douglas GC, Kelly DL. Species Distinction in Irish Population of *Quercus petraea* and *Q. robur*; Morphological versus Molecular Analyses. *Annals of Botany* 2005; 96: 1237-1246.
- [74] Kleinschmit J, Kleinschmit JGR. *Quercus robur* - *Quercus petraea*: a critical review of the species concept. *Glasnik za šumske pokuse* 2000; 37: 441-452.
- [75] Kleinschmit JRG, Bacilieri R, Kremer A, Kleinschmit J. Comparison of morphological and genetic traits of pedunculate oak (*Q. robur* L.) and sessile oak (*Q. petraea* (Matt.) Liebl.). *Silvae Genetica* 1995; 44: 256-269.
- [76] Konieczny A, Voytas DF, Cummings MP, Ausubel FM. A superfamily of *Arabidopsis thaliana* retrotransposons. *Genetics* 1991; 127: 801-809.
- [77] Kremer A, Petit RJ. Gene diversity in natural populations of oak species. *Ann Sci For* 1993; 50: 186s-202s.
- [78] Kremer A, Dupouey JL, Deans JD, et al. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Ann Sci For* 2002; 59: 777-787.
- [79] Kremer A, Dupouey JL, Deans JD, et al. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annales des Sciences Forestieres* 2002; 59:777-787.
- [80] Kubalkova M, Cihalinkova J, Dolezal J. Localisation of GAA microsatellites on barley chromosomes by direct cyclin-prins (C-prins). In: Maluszynska J, ed. *Plant Cytogenetics*. Katowice: Wydawnictwo Uniwersytetu Slaskiego, 1998: 173-177.
- [81] Lexer C, Heinze B, Steinkellner H, Kampfer S, Ziegenhagen B, Gloessl J. Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: detection of

- seed contaminations and inference of the seed parents from the offspring. *Theor Appl Genet* 1999; 99: 185-191.
- [82] Lexer C, Kremer A, Petit RJ. Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Mol Ecol* 2006; 15: 2007-2012.
- [83] Lowe A, Harris S, Ashton P, eds. Case studies in ecological genetics: Lycaenid butterflies, ragworts, bears, and oaks, first ed. Malden, USA: Blackwell Publishing, 2004.
- [84] Lowe AJ, Abbott RJ. Routes of origin of two recently evolved hybrid taxa: *Senecio vulgaris* var. *bibernius* and York radiate groundsel (Asteraceae). *Amer J Bot* 2000; 87.
- [85] Manos PS, Steele KP. Phylogenetic analyses of "higher" Hamamelididae based on plastid sequence data. *Amer J Bot* 1997; 84: 1407-1419.
- [86] Manos PS, Doyle JJ, Nixon KC. Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae). *Mol Phylogen Evol* 1999; 12: 333-349.
- [87] Manos PS, Zhou Z-K, Cannon CH. Systematics of Fagaceae: phylogenetics tests of reproductive trait evolution. *Int J Plant Sci* 2001; 162: 1361-1379.
- [88] Mariette S, Cottrell J, Csaikl UM, et al. Comparison of levels of genetic diversity detected with AFLP and microsatellite markers within and among mixed *Q. petraea* (Matt.) Liebl. and *Q. robur* L. stands. *Silvae Genetica* 2002; 51: 72-79.
- [89] Mayer E, ed., Systematics and origin of species. New York: Columbia University Press, 1942.
- [90] Mlinarec J, Papeš D, Besendorfer V. Ribosomal, telomeric and heterochromatin sequences localization in the karyotype of *Anemone hortensis* L. *Bot J Linn Soc* 2006; 150: 177-186.
- [91] Moreau F, Kleinschmit J, Kremer A. Molecular differentiation between *Q. petraea* and *Q. robur* assessed by random amplified DNA fragments. *Forest Genet* 1994; 1: 51-64.
- [92] Mueller-Starck G, Zanetto A, Kremer A, Herzog S. Inheritance of isoenzymes in sessile oak (*Quercus petraea* (Matt.) Liebl.) and offspring from interspecific crosses. *Forest Genet* 1996; 3: 1-12.
- [93] Muir G, Fleming CC, Schlotterer C. Species status of hybridizing oaks. *Nature* 2000; 405: 1016.
- [94] Muir G, Fleming CC, Schlotterer C. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol Biol Evol* 2001; 18: 112-119.
- [95] Muir G, Schlotterer C. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Mol Ecol* 2005; 14: 549-561.
- [96] Muir G, Schlotterer C. Moving beyond single-locus studies to characterize hybridization between oaks (*Quercus* spp.). *Mol Ecol* 2006; 15: 2301-2304.
- [97] Nixon KC. Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Ann Sci For* 1993; 50: 25s-34s.
- [98] Ohri D, Ahuja MR, Giemsa C-Banded Karyotype in *Quercus* L. (Oak). *Silvae Genetica* 1990; 39: 216-219.
- [99] Ohri D, Ahuja MR, Giemsa C-Banded Karyotype in *Quercus* L. (Oak). *Silvae Genetica* 1990; 39: 216-219.

- [100] Olszewska MJ, Osiecka R. The relationship between 2C DNA content, systematic position and the level of nuclear DNA endoreplication during differentiation of root parenchyma in some dicotyledonous shrubs and trees. Comparison with herbaceous species. *Biochem Physiol Pflanz* 1984; 719: 641-657.
- [101] Orr AH. The genetics of species differences. *Trends in Ecology & Evolution* 2001; 16: 343-350.
- [102] Pedersen C, Linde-Larsen I. Chromosomal locations of four minor rDNA loci and a marker microsatellite sequence in barley. *Chromosome Res* 1994; 2: 65-71.
- [103] Pernes J. Gestion des ressources Génétiques. Tome 2: Manuel. Paris, France: Agence de Cooperation Culturelle et Technique, 1984.
- [104] Petit RJ, Kremer A. Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Ann Sci For* 1993; 50: 41s-47s.
- [105] Petit RJ, Brewer S, Bordacs S, et al. Identification of postglacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 2002; 156: 49-74.
- [106] Petit RJ, Csaikl U, Bordacs S, et al. Chloroplast DNA variation in European white oaks: phytogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* 2002; 156: 5-26.
- [107] Petit RJ, Bodenes C, Ducousoo A, Roussel G, Kremer A. Hybridization as a mechanism of invasion in oaks. *New Phytologist* 2003; 161: 151-164.
- [108] Porth I, Scotti-Saintagne C, Barreneche T, Kremer A, Burg K. Linkage mapping of osmotic stress induced genes of oak. *Tree Genetics & Genomes* 2005; 1: 31-40.
- [109] Rayburn AL, Auger JA. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theor Appl Genet* 1990; 79: 470-474.
- [110] Rees H, Jones RN. *Chromosome Genetics*. London, UK: Edward Arnold, 1977.
- [111] Rushton BS. Artificial hybridization between *Q. robur* and *Q. petraea*. *Watsonia* 1977; 11.
- [112] Rushton BS. *Quercus robur* and *Q. petraea*: a multivariate approach to hybrid problems. I. Data cquisition, analysis and interpretation. *Watsonia* 1978; 12: 81-101.
- [113] Rushton BS. *Quercus robur* and *Q. petraea*: a multivariate approach to hybrid problems. II. The geographical distribution of population types. *Watsonia* 1979; 12: 209-224.
- [114] Rushton BS. The analysis of variation of leaf characters in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. Population samples from Northern Ireland. *Irish Forestry* 1983; 40: 52-77.
- [115] Rushton BS. Natural hybridization within the genus *Quercus* L. *Ann Sci for Suppl* 1993; 50: 73-90.
- [116] Saintagne C, Bodenes C, Barreneche T, Pot D, Plomion C, Kremer A. Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* 2004; 92: 20-30.
- [117] Samuel R, Pinsker W, Ehrendorfer F. Electrophoretic analysis of genetic variation within and between populations of *Quercus cerris*, *Q. pubescens*, *Q. petraea* and *Q. robur* (Fagaceae) from eastern Austria. *Bot Acta* 1995; 108: 290-299.
- [118] Samuel R, Bachmair A, Jobst J, Ehrendorfer F. ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and Northern American taxa of *Quercus* (Fagaceae). *Plant Syst Evol* 1998; 211.

- [119] SanMiguel P, Tikhonov A, Jin YK, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996; 274: 765-768.
- [120] Schirone B, Spada F. Anomalies in reproductive phenology and vegetation history: the case of SE Italy. *Coll Phytosoc* 1995; 14: 847-857.
- [121] Schnitzler J-P, Steinbrecher R, Zimmer I, Steigner D, Fladung M. Hybridization of European oaks (*Quercus ilex* x *Q. robur*) results in a mixed isoprenoid emitter type. *Plant, Cell and Environment* 2004; 27: 585-593.
- [122] Schwarz O. Entwurf einem natürlichen System der Culpuliferen und der Gattug *Quercus* L. *Notizbl Bot Gart Berlin* 1936; 13: 1-22.
- [123] Schwarz O. *Quercus*. London, UK: Cambridge University Press, 1964.
- [124] Scotti-Saintagne C, Mariette S, Porth I, et al. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* 2004; 168: 1615-1626.
- [125] Spellenberg R. On the hybrid nature of *Quercus Basaseachicensis* (Fagaceae, sect. *Quercus*). *Sida* 1995; 16: 427-437.
- [126] Stebbins GL. Chromosomal evolution in higher plants. London, UK: Edward Arnold, 1971.
- [127] Steinhoff S. Results of species hybridization with *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Annales des Sciences Forestieres*, Supplement 1993; 50: 137s-143s.
- [128] Steinkellner H, Fluch S, Turetschek E, et al. Identification and characterization of (GA/CT)<sub>n</sub>-microsatellite loci from *Quercus petraea*. *Plant Molec Biol* 1997; 33: 1093-1096.
- [129] Streiff R, Labbe T, Bacilieri R, Steinkellner H, Gloessl J, Kremer A. Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Mol Ecol* 1998; 7: 317-328.
- [130] Streiff R, Ducousoo A, Lexer C, Steinkellner H, Gloessl J, Kremer A. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Mol Ecol* 1999; 8: 831-841.
- [131] Tovar-Sanchez E, Oyama K. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. *Amer J Bot* 2004; 91: 1352-1363.
- [132] Tucker JM. Studies in *Quercus undulata* complex I. A preliminary statement. *Amer J of Bot* 1961; 48: 202-208.
- [133] Tucker JM. Hermaphroditic flowers in Californian oaks. *Madrono* 1972; 21: 482-486.
- [134] Tucker JM. Patterns of parallel evolution of leaf form in new world oaks. *Taxon* 1974; 23: 129-154.
- [135] Tucker JM, Neilson RP, Wullstein LH. Hermaphroditic flowering in Gambel oak. *Amer J Bot* 1980; 62: 1265-1268.
- [136] Unfried K, Schiebel K, Hemleben V. Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not in *Vigna angularis*. *Gene* 1991; 99: 63-68.
- [137] Valles J, Garnatje T. *Artemisia and Its Allies: Genome Organization and Evolution and their Biosystematic, Taxonomic, and Phylogenetic Implications in the Artemisiinae and Related Subtribes (Asteraceae, Anthemideae)*, first ed. Enfield, NH, USA: Science Publishers, I 2005.
- [138] VanValen L. Ecological species, multispecies and oaks. *Taxon* 1976; 25: 233-239.

- [139] Vershinin AV, Schwarzacher T, Heslop-Harrison JS. The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. *Plant Cell* 1995; 7: 1823-1833.
- [140] Vicient CM, Suoniemi A, Anamthawat-Jonsson K, et al. Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* 1999; 11: 1769-1784.
- [141] White MJD. Animal cytology and evolution. Cambridge, UK: Cambridge University Press, 1973.
- [142] Whittemore AT, Schaal BA. Interspecific gene flow in sympatric oaks. *Proc Natl Acad Sci USA* 1991; 88: 2540-2544.
- [143] Wielgorskaya T, ed. Dictionary of generic names of seed plants, 1st ed. New York: Columbia University Press, 1995.
- [144] Wu CI. The genic view of the process of speciation. *J Evol Biol* 2001; 14: 851-865.
- [145] Yokoka K, Roberts AV, Mottley J, Lewis R, Brandham PE. Nuclear DNA amounts in *Rosa*. *Ann Bot* 2000; 85: 557-561.
- [146] Zanetto A, Roussel G, Kremer A. Geographic variation of inter-specific differentiation between *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *For Genet* 1994; 99: 111-123.
- [147] Zanetto A, Kremer A, Mueller-Starck G, Hattemer HH. Inheritance of isozymes in Pedunculate oak (*Quercus robur* L.). *J Hered* 1996; 87: 364-370.
- [148] Zoldos V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev S. Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. *Genome* 1998; 41: 162-168.
- [149] Zoldos V, Papes D, Cerbah M, Panaud O, Besendorfer V, Siljak-Yakovlev S. Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. *Theor Appl Genet* 1999; 99: 969-977.
- [150] Zoldos V, Siljak-Yakovlev S, Papes D, Sarr A, Panaud O. Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. *Mol Genet Genomics* 2001; 265.
- [151] Bauer N, Horvat T, Birus I, Zoldos V. Ribosomal DNA intergenic spacer of sessile, *Quercus petraea* (Matt.) Lieb., and pedunculate, *Q. robur* L., oak. *Plant Mol Biol* in press.

## Biodiversity and Evolution of *Castanea*

FENNY DANE and PING LANG

Department of Horticulture, Auburn University, AL 36849 and  
Department of Population Medicine and Diagnostic Sciences, Cornell  
University, NY

### ABSTRACT

Evolutionary and phylogenetic studies of species in the genus *Castanea* using chloroplast DNA sequence data indicated that the genus is monophyletic with *C. crenata*, the Japanese chestnut, as basal. A unique westward expansion of extant *Castanea* species is hypothesized with *Castanea* originating in eastern Asia, followed by intercontinental dispersion and divergence between the Chinese (*C. mollissima*, *C. sequinii* and *C. henryi*) and European/North American species during the middle Eocene and a split between the European (*C. sativa*) and North American (*C. dentata* and *C. pumila*) species in the early Oligocene. Differentiation in North America and China might have occurred in the Miocene. On the American continent, the Ozark chinkapin (*C. pumila* var. *ozarkensis*), is basal and sister to the group comprising *C. pumila* var. *pumila*, the Allegheny chinkapin and *C. dentata*, the American chestnut. Colonization patterns from the Ozarks in eastern and southern direction and mixing of divergent cpDNA lineages in Appalachian mountain range reflect the impacts of the Tertiary, while the extant populations of the American chestnut are the result of post-glacial expansion from southern refugia.

**Key Words:** Biogeography, *Castanea*, chloroplast DNA, chestnut, chinkapin

The remarkable floristic similarities between temperate eastern Asia and eastern North America, known as the Arcto-Tertiary disjunction (or the Tertiary relict disjunction), have been widely recognized for over a century [3, 17]. The floristic similarity encompasses most of the world's

Northern Hemisphere temperate forest regions, including eastern and western Asia, Europe, and eastern and western North America. Of particular interest is the Asian-American floristic affinity. Approximately 65 genera of woody plants are distributed disjunctly in eastern Asia and eastern North America [70], and eastern Asia has approximately twice as many species, on average, as eastern North America [22, 56, 57, 58]. It is generally accepted that the disjunct pattern represents a relict distribution of taxa with a wider geographic range due to the maximum development of the temperate forests in the mid-Tertiary, around 32 million years before present (mybp), especially in the Miocene (25.5 mybp) [18]. Migration between Eurasia and North America were facilitated by the presence of the Bering Land Bridge (BLB) that connected eastern Asia with western North America throughout the mid-Tertiary and by a North Atlantic Land Bridge (NALB) during the late Eocene (45-38 mybp) [23, 43, 65, 66]. The deterioration of the temperate flora in western North America and in Europe during the Pliocene (5-1.8 mybp) and the Quaternary (1.8 mybp-present, culminating in repeated Pleistocene glaciations) caused the present distribution pattern [71, 72, 73]. The disjunct pattern between eastern Asia and eastern North America is the product of vicariance, dispersal, extinction, and speciation [70].

The genus *Castanea* Miller exhibits an Arcto-Tertiary disjunction. *Castanea* species are found in southern Europe and western Asia Minor, eastern China, Korea and Japan, and the eastern United States (Fig. 1). Closely related genera in the same biological family of Fagaceae, such as oak (*Quercus*) and beech (*Fagus*), are phylogenetically and biogeographically well-studied genera with an eastern Asia and eastern North America disjunct distribution pattern. According to Manos et al. [44], *Quercus* appears to have achieved a widespread distribution during the Upper Eocene (53-45 mybp) to Lower Oligocene (30-25 mybp) suggesting more recent connections between Asia and North America, perhaps involving the BLB. Fossil data generally support a North American origin during the Oligocene (38-25 mybp) with larger appearance in China in the Upper Miocene (25-20 mybp). In this case, a phylogenetic pattern of Eurasian white oaks nested within North American lineages would be expected, given the probable direction of migration [45, 46]. For *Fagus*, an Asian origin, is supported with bidirectional migration to Europe and North America, consistent with a paraphyletic assemblage of Asian species and intercontinental exchange

via the BLB [45]. Studies on differentiation of *Fagus* in western Eurasia suggested that *Fagus* originated in the northern Pacific Basin based on Eocene fossils, found in Kamchatka [10, 11]. From the northern Pacific Basin *Fagus* spread towards the east to western North America, and along the northern Paratethyan coast to Central Asia and Europe. Relationships within the subgenus *Fagus* could not be clearly resolved, possibly due to the low rate of diversification in *Fagus* during the early phase of range expansion of the genus in the Oligocene period, as indicated by the uniformity of leaf and cupule/nut fossils [10]. The genus *Juglans* belongs to the Juglandaceae, a sister family of Fagaceae. The fossil record suggests that the butternuts evolved by the early Oligocene (38-34 mybp) in North America. The presence of butternuts in Eurasia could be the result of migration from North America to Eurasia during the warming trend of the mid Oligocene (34-30 mybp) [62]. *Castanea* is a more recently diverged lineage of the Fagaceae than *Fagus* and *Quercus*, and it might share a similar biogeography history with *Fagus* [45].

*Castanea*, the genus of the chestnuts and chinkapins, is widely distributed in the temperate forests of the Northern Hemisphere. *Castanea* species are trees and shrubs with simple, ovate or lanceolate leaves with sharply-pointed, widely-spaced teeth, and with rounded sinuses between the teeth. The fruit is a paired nut enclosed within a spiny cupule. The flowers are catkins with two types: a male staminate type which tends to flower earlier, and a mixed type which has female flowers below a staminate catkin. All *Castanea* species are diploid ( $2n = 2x = 24$ ) and hybridize freely, although some interspecific  $F_1$ s usually suffer from low seed germination and male sterility [32]. Pollination is described as anemophilous, entomophilous, or a combination of both [32]. The limited seed dispersal by weight, which is typical of large and heavy seeds, is partially balanced by animals and humans who also contribute to their diffusion [13]. Morphological variation within *Castanea* is low, and their classification is based primarily on leaf morphology, and variation in fruit and cupule numbers within an inflorescence.

The genus comprises three sections and seven species [33]. Section *Eucastanon* comprises five species: *C. mollissima* BL., *C. seguinii* Dode. from China, *C. crenata* Sieb. & Zucc. from Japan, *C. dentata* (Marsh.) Brokh from North America, and *C. sativa* Mill. from Europe, all with the characteristics of three nuts per cupule. Chinese chestnut (*C. mollissima*), Japanese chestnut (*C. crenata*) and European chestnut (*C. sativa*) are

economically important as nut producers, and have been cultivated for centuries. The Chinese chestnut is the most commercially important species and currently produces one third to half of the world's chestnuts. *C. mollissima* is indigenous to 22 Chinese provinces and more than 300 cultivars have been recognized. Japanese chestnut (*C. crenata*) is found on the Japanese islands and Korean peninsula [32, 63]. European chestnut (*C. sativa*) is an ecologically and economically important tree species with a long cultural tradition in many areas of Europe. Selection and grafting of superior trees began at least 3,000 years ago. There are hundreds of varieties that span the ages of civilization [32]. Chestnut cultivation offers the advantage of both nut and timber production. In general, nuts of the European and Japanese species have fewer soluble sugars than the Chinese chestnut. *C. seguinii*, Seguin chestnut, is commercially less important because of its small nut size and occurs sympatrically with Chinese chestnut over central and southwest China. The American chestnut (*C. dentata*) is native to eastern North America, from southern Maine south to Georgia, Alabama and Mississippi, and west to southern Ontario and Michigan [59]. It was valued primarily as a timber and tannin source and only secondarily as a nut tree. The chestnut tree was one of the best for timber.

Section *Balanocastanon*, has a widely disputed taxonomy, with taxa exclusively found in the southeastern United States. It contains just one species, *C. pumila* Mill., with two varieties: var. *pumila* and var. *ozarkensis*, characterized by one nut per cupule. *C. pumila* var. *pumila*, the Allegheny chinkapin, has a wide distribution from southern New Jersey and Pennsylvania, west to Indiana and Missouri, and south to Florida and Texas. *C. pumila* var. *ozarkensis*, the Ozark chinkapin, has limited and fragmented distribution in the Ozark Mountains of eastern Oklahoma, southwest Missouri, and north-central Arkansas [33]. Currently, no chinkapins are cultivated in commercial orchards, however, dwarfing genes may be important for the introduction of smaller trees into chestnut orchards [51].

Section *Hypocastanon* consists of just one species, *C. henryi* (Skan) Rehder & Wilson, which is found in a restricted area in southeast China and is sometimes called the Chinese chinkapin since it is characterized by a single nut per cupule [36]. This species is also commercially less important because of its small nut size.

The geographic distribution of *Castanea* in China and the United States has a similar pattern, i.e., the geographically widespread species (*C.*

*mollissima* in China and *C. dentata* in the US) overlap the restricted species (*C. segoii* and *C. henryi* in China and *C. pumila* in the US; Fig. 1). Although floristic similarity exists between the intercontinental disjunct species, significant differences in blight resistance were found between eastern Asia and eastern North American species. The American species are blight susceptible, while the Asian species are blight resistant. Chestnut blight, caused by the fungus *Cryphonectria parasitica*, was introduced into North America on imported Asian *Castanea* species and was first observed in New York in 1904. Recent population genetic analyses provide evidence that the North American strains of the fungus originated in Japan [47]. The blight fungus is a wound parasite that occludes the phloem of susceptible trees, forming a girdling canker on the main trunk. Within the genus, *C. dentata* is the species that is the most susceptible to *Cryphonectria parasitica* and its destructive effects. However, infection by the fungus does not affect American chestnut root systems, only the aboveground stems. It is only by producing sprouts from the root collar that *C. dentata* has escaped extirpation from the Appalachian forest ecosystem. Since the introduction of the chestnut blight in America, the American chestnut has been reduced from an important timber and nut producing tree to an understory shrub [20, 28].

It is reasonable to assume that *Castanea* trees were more widespread in Tertiary times than they are today. Macrofossils of *Castanea* have been reported in the Rocky Mountain region of North America from the Paleocene to the early Eocene (<http://www.nhm.uio.no/palmus/galleri/montre/english/a31696.htm>, [19, 64] and throughout the Tertiary in European floras (International Organization of Palaeobotany, 1997; <http://www.paleodb.org>) [30]. In Asia, fossils have been found in Japan (Oligocene), North Korea (Miocene), northeastern, eastern and southwestern China (late Eocene, Miocene, and Pliocene) and western Asia (Oligocene) [74]. In eastern North America, Eocene leaves and cupules with strong affinities to modern species suggest long-term presence [5, 19, 45]. Although many fossil records are available, attempts to reconstruct phylogeographic history based on fossil distribution patterns may be misleading, because distinguishing characters such as locule number and style number have not been found, and pollen and leaf characters are too general to be of diagnostic value [19, 45]. Thus, the origin of the genus *Castanea* needs to be considered carefully and a clear phylogenetic framework is needed to elucidate complex biogeographic patterns.

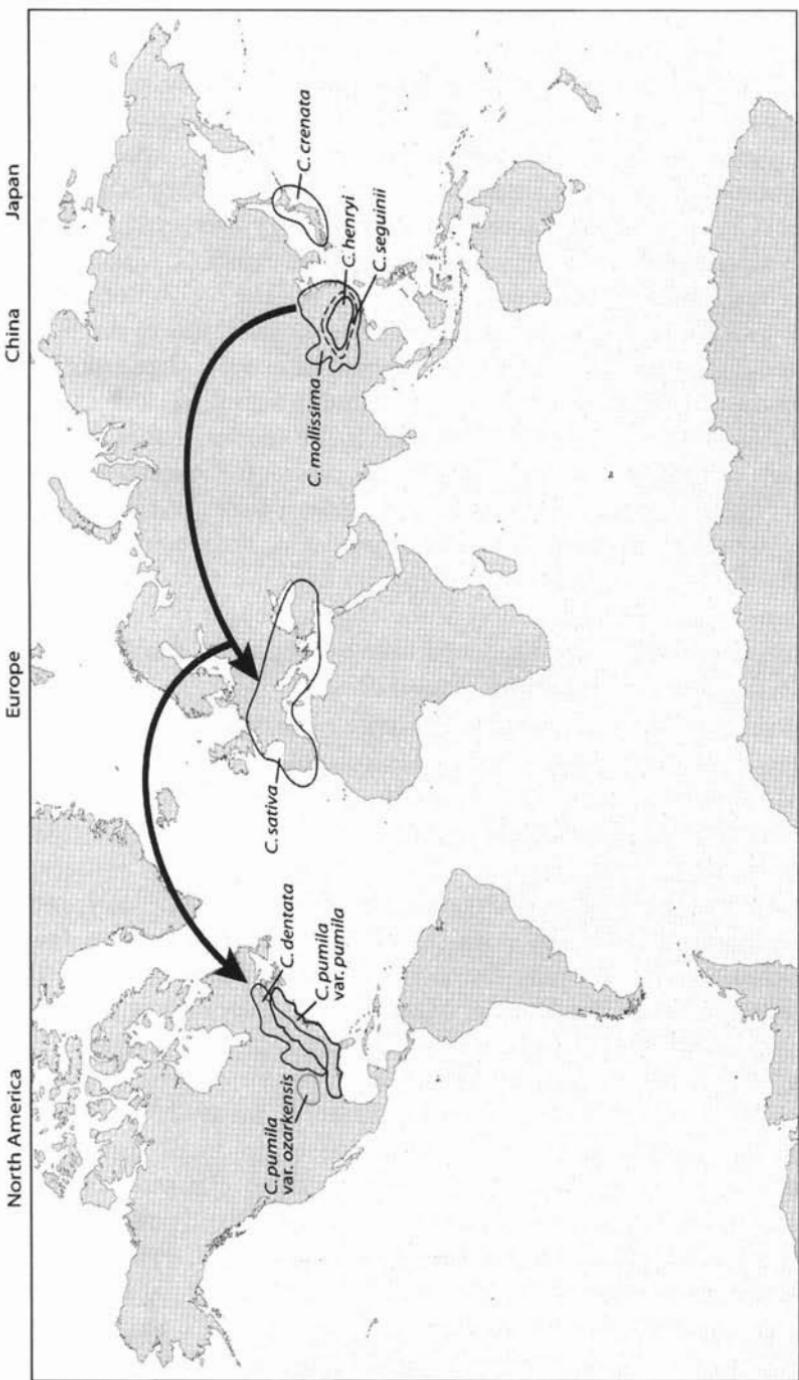


Fig. 1. Distribution of the genus *Castanea* and a hypothesized westward migration route indicated by arrows based on cpDNA sequence information from extant species.

Allozyme investigations conducted on European chestnut (*C. sativa*) provided strong evidence that Eastern Turkey is the center of origin of the European chestnut. A clear decline in total allozyme variation from East Turkey to West Turkey and thence into Italy was demonstrated. The evolution of European chestnut appears to be a complex mixture of long-range gene flow, natural and artificial selection, and local effects of isolation by distance [67, 68]. Allozyme and random amplified polymorphic DNA (RAPD) studies on American chestnut indicate that the genetic diversity of the surviving American chestnut populations is narrow as compared to other *Castanea* species [27, 29], while genetic variation among populations of the Ozark and Allegheny chinkapins is similar to that observed for other species in the genus [6]. Genetic variability found in *C. mollissima* is the highest among the examined *Castanea* species [7, 36]. The center of origin of the Chinese chestnut might be located in the Changjiang river area, particularly in Shennongjia, where several so-called living fossil species have been found. Warm climates during glaciation periods would have provided an ideal refugium for many ancient endemic plants and animals in that area [36]. Cultivated and wild Japanese chestnuts were identified using RAPD and restriction fragment length polymorphism (RFLP) markers and very low polymorphisms were detected [48]. However, in recent genetic diversity studies of wild chestnut populations distributed in northern Japan using simple sequence repeat (SSR) markers, extensive diversity in both wild and cultivated chestnuts was observed [63].

Historical factors, such as the last glacial period and the subsequent recolonization of species into lands newly released by ice, have in all probability played a major role in determining the present patterns of genetic variation in natural populations of the temperate zone [71, 73]. Ice ages occur at regular intervals of 100,000 years with warm interglacial periods lasting 15,000 to 20,000 years [2]. Knowledge of Pleistocene history, especially in eastern North America, is limited due to the relatively low abundance of fossil data from the late Tertiary and early to middle Pleistocene in this region [19]. At its maximum, the most recent Wisconsin ice sheet extended as far south as 40° N latitude in eastern North America. Many populations that had once inhabited areas north of the glacial boundary were restricted to smaller ranges south of the ice or to refugia. Interglacial pollen records of *Castanea* in fossil lakes in Florida date back to 35,000 to 30,000 B.P. [69]. Analyses of preserved pollen samples suggest that *Castanea* migrated very slowly northward and

that while oak-chestnut forests dominated central Appalachian more than 5,000 years ago, chestnuts arrived in Connecticut only 2,000 years ago [8]. The most recent review of eastern tree distributions at the last glacial maximum concludes that the occurrence of most temperate hardwoods is difficult to document using the fossil record, except for Lower Mississippi Valley sites [31].

Recolonization of Europe by forest tree species after the last glaciation is well documented in the fossil pollen record generating a patchy genetic structure through founder effects. Several studies investigated vegetation changes in relation to glacial cycles [12, 35] and detected patterns of fragmentation and dispersal resulting from climate changes during the Pleistocene epoch. Long term isolation in glacial refugia resulted in a strong genetic drift reflected in extant populations located close to glacial refugia [25]. The expansion from refugia since about 12,000 years was predicted to result in a loss of variation due to successive founder events. However, many species displayed high diversity in recolonized forests in Europe [54]. Rangewide population distribution of oaks (*Q. rubra*) in eastern North America was very low, indicative of different evolutionary histories of the oak species in Europe versus North America. Palynological evidence indicates that during the last glacial maximum *Q. rubra* has one major distribution range with populations located relatively far to the north, resulting in only modest movement northward when the climate improved [41]. Complementary paleobotanical and genetic data indicated that the European beech survived the last glacial period in multiple refuge areas in Europe and that the modern genetic diversity was shaped over multiple glacial-interglacial cycles [42]. *Castanea sativa* similarly survived the cold periods of the last glaciation in refugia's in mountainous regions of southern Europe together with other thermophilous European tree taxa [34].

In order to investigate the correspondence of genetic relationships and geographical distribution in plant species, variation at noncoding regions of the chloroplast genome have been used in plants and mitochondrial DNA (mtDNA) variation in animal species [1, 16]. Chloroplast DNA (cpDNA) is ideal for studies of historical patterns of gene flow, in particular migration and colonization routes of woody species; since it is non-recombining and maternally inherited thus haplotypes remain mostly unchanged when passed to the next generation [35, 50, 53, 55]. Colonization patterns which derive from seed dispersal are not blurred by pollen flow. The maternal inheritance implies that

effective population size is smaller for cpDNA than for nuclear DNA, and levels of differentiation are expected to be higher for cpDNA than for nuclear markers. Since recombination is absent, cpDNA responds as a single gene and the rate of evolution is relatively low. Using cpDNA variability of two sympatric oak species in Western France, Petit et al. [52] detected patches of oaks which are fixed for a single haplotype for both oak species, suggesting local interspecific sharing of the maternal genome at the time of colonization some 10,000 years ago. Chloroplast DNA markers similarly provide insights into the genetic dynamics underlying morphologically indistinguishable taxa such as *Castanea* species [38, 39]. DNA sequence data from different regions of the chloroplast genome (*tmT-L-F*, *rpl16*, *ndhF*, *ycf6-psbM*, *ycf9-tmGM*) were used for phylogeographic studies in the genus *Castanea*. Divergence among and within *Castanea* species is generally very low, consistent with results of the *matK* [45, 61] and *rbcL* region [14]. Possible explanations might be that the genus evolved slowly, or divergence in *Castanea* occurred relatively recently such that neutral markers would not have had sufficient time to become fixed in different species [39]. Higher levels of cpDNA sequence divergence were found in Asian species, as evidenced by an accumulation of many unique substitutions and indels in *C. crenata*. Haplotype divergence in Asia is consistent with species richness and points to the center of origin for *Castanea* [39]. Contemporary patterns of diversity suggest that the effects of climate change in the late Tertiary were less severe in eastern Asia and promoted diversification but were more severe in North America and may have caused widespread extinctions [58, 70, 22], at least in *C. dentata*. Sequence data analysis indicated that the genus is monophyletic, with *C. crenata* as basal and sister to the remaining species. The Chinese species are supported as a single monophyletic clade, sister to the North American (NA) and European clade. Sequence data reveal fragmentation of species into two main lineages, with the accumulation of unique substitutions in one main lineage shared by Chinese, European and North American species (Fig. 1). The Chinese species fragmented early and accumulated unique substitutions, while the European and NA species, which share several unique substitutions and insertions are more closely related to each other than to the Asian species [39]. This clearly indicates the direction of dispersal and migration in extant *Castanea* from Asia to Europe and onto North America via the BLB. According to Bayesian analysis of combined cpDNA data, the split of eastern Asia is

estimated to have occurred in the early Eocene (54.43 mybp); the divergence of European/NA species and Chinese species can be estimated to have occurred in the middle Eocene (42.55 mybp). The split in *Castanea* between European and NA species is estimated to have occurred at 39.14 mybp during the late Eocene. The differentiation in North America and Asia occurred in the early and late Miocene, respectively. Due to difficulties faced with the integration of the fossil record with studies of modern species, the analysis of dispersal-vicariance analysis (DIVA) was based solely on phylogenies of extant species [39].

## **PHYLOGEOGRAPHY OF AMERICAN *CASTANEA* SPECIES**

Unglaciated eastern North America is a large geologically and topologically complex area characterized by several terrestrial discontinuities such as the Appalachian mountain range and Apalachicola, Tombigbee and Mississippi river basins [60]. Delcourt [9] posited stable glacial refugia for mesic temperate species on isolated bluffs associated with alluvial valleys along the Gulf Coast. In order to reconstruct the biogeographic history of *C. dentata* and *C. pumila* on the American continent, cpDNA sequence data from many different populations were analyzed. Sequences from five variable cpDNA regions (>3,000 bp) of American chestnut trees indicate very little diversity. *C. dentata* displays a total of only three different haplotypes across the Appalachian mountain range with most populations fixed for the common haplotype (Fig. 2). The *C. dentata* haplotypes, which can be distinguished from other *Castanea* species by a large deletion at *tml*-F [38], show only minor nucleotide differences, either a transversion and/or a chloroplast simple sequence repeat (mononucleotide repeats in poly A/T stretches (cpSSR; Table 1), implicating relatively recent mutational changes (Fig. 2).

Chloroplast DNA variation in *C. pumila* populations (65 trees) is much higher. Several indels ranging in size from 2 bp to 72 bp, transversions, and cpSSRs accumulated, resulting in more than 25 different haplotypes (Fig. 3; Table 1), which indicates that the American chinkapin persisted during glacial periods possibly at low density populations in several regions.

Ancestral or ancient haplotypes at cpDNA regions can be identified using comparative sequence analysis. At the *rpl16*, known for its deep

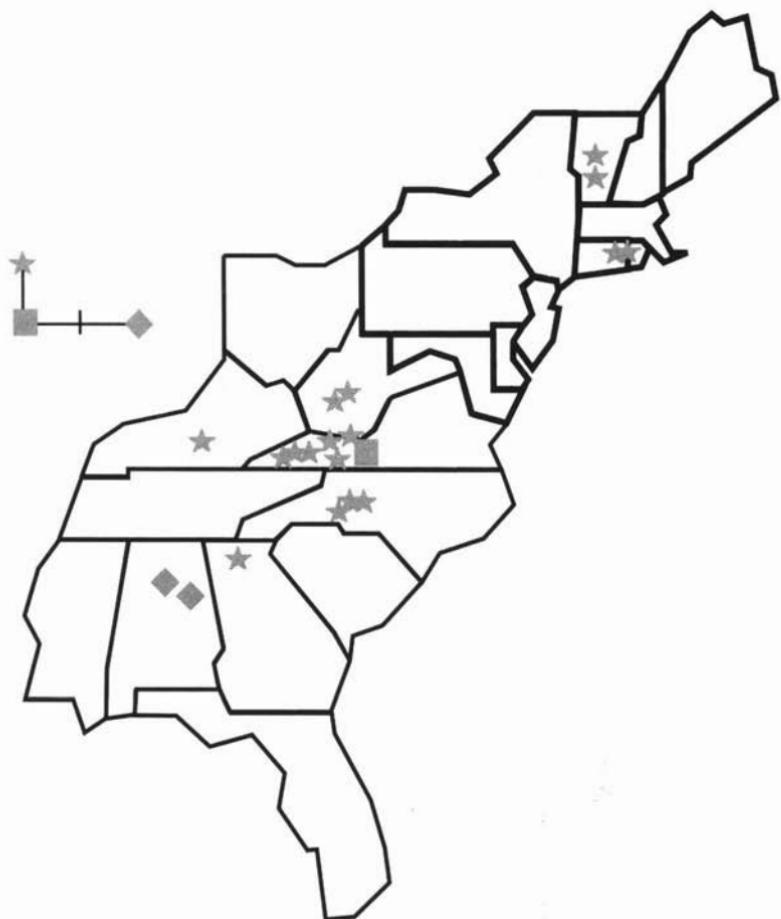


Fig. 2. Geographic distribution and frequency of chloroplast haplotypes of *Castanea dentata* on the North American continent. The haplotype network represents the minimum number of mutational events between the different haplotypes.

evolutionary splits in the genus [37], *tmL* and *ycf6-psbM* regions, the ancestral haplotypes are shared between *C. dentata* and the main Florida *C. pumila* var. *pumila* population. This large population is located west of the Apalachicola river basin (Fig. 3). Several indels and microsatellites at *ndhF* and *ycf9-tmGM* differentiate the Florida chinkapins from the American chestnut. The Florida population is quite widespread and appears persistent locally, thus maintaining ancient genetic structure. In a recent study of white oaks in California, haplotype richness was higher than that observed in European oaks [21] indicating that environmental

**Table 1.** Nucleotide substitutions and number of repeats (poly A/T) for five cpDNA regions and number of haplotypes of American *Castanea* species

Species	<i>C. pumila</i> var. <i>pumila</i>						<i>C. ozarkensis</i>			<i>C. dentata</i>			
	cpDNA regions	Tv	ID	Repeats	Haplotypes	Tv	ID	Repeats	Haplotypes	Tv	ID	Repeats	Haplotypes
ycf6	1	-	11, 12	3	-	11	1	-	-	-	-	12	1
rpl16	1	-	7, 8, 9	4	1	-	9	1	-	-	-	8	1
trnL	1	1	4	1	1	-	-	2	1	-	-	-	2
ndhF	2	4	8	2	-	-	-	2	1	1	-	-	2
ycf9	5	8, 10, 11, 12	9	2 (Ts)	2	9	2	-	-	10, 11	2	-	3
Total					20			5					

Tv = transversion, Ts = transition, ID = insertion/deletion; Sample locations and GenBank accession numbers are available from the senior author.

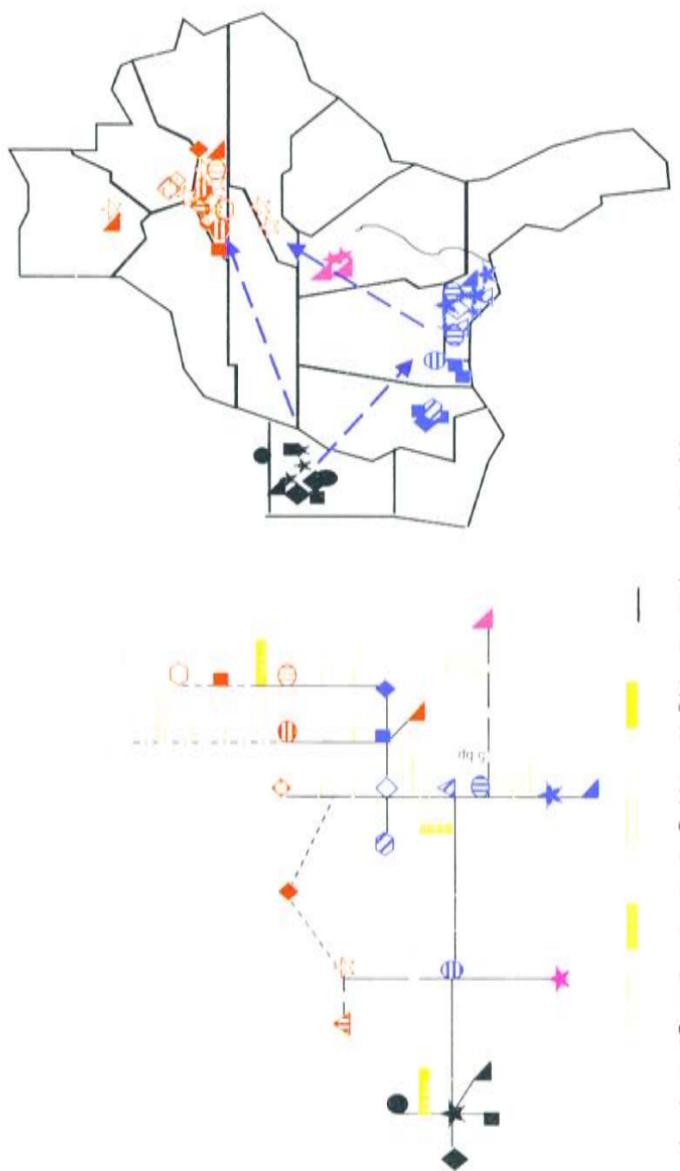


Fig. 3. Geographic distribution and frequency of chloroplast haplotypes of *Castanea pumila* on the North American continent. The haplotype network represents the minimum number of mutational events between the different haplotypes, indicating unique insertions and deletions (with length in bp), transversions, transitions and cSNPs. Branch lengths correspond to the number of mutational steps separating the haplotypes; the placement of nodes along the branches is arbitrary. The *C. pumila* var. *ozarkensis* haplotypes are shown in black, southern *C. pumila* var. *pumila* haplotypes in blue, Appalachian var. *pumila* haplotypes in red.

The geographic distribution and frequency of chloroplast haplotypes of *Castanea pumila* on the North American continent. The haplotype network represents the minimum number of mutational events between the different haplotypes, indicating unique insertions and deletions (with length in bp), transversions, transitions and cSNPs. Branch lengths correspond to the number of mutational steps separating the haplotypes; the placement of nodes along the branches is arbitrary. The *C. pumila* var. *ozarkensis* haplotypes are shown in black, southern *C. pumila* var. *pumila* haplotypes in blue, Appalachian var. *pumila* haplotypes in red.

changes during the Pleistocene and earlier epochs did not seriously affect California oak populations.

Haplotype sharing at cpDNA regions between the species similarly occurred at the *trnL* and *ndhF* regions. At the *trnL* region *C. dentata* and *C. pumila* share two haplotypes, while a unique 18 bp deletion characterizes additional haplotypes in *C. pumila* var. *ozarkensis* populations and one Virginia Allegheny chinkapin population. A large number of mutational changes separate the different *Castanea* haplotypes at the *ndhF* region, where only two of the eight *C. pumila* haplotypes are shared with *C. dentata*. Similarly, at the *ycf9* region two of the nine *C. pumila* haplotypes are shared with *C. dentata* (Table 1). Most of the chinkapin populations are fixed for particular haplotypes and many haplotypes are not shared among the different populations (Fig. 3). Sharing of haplotypes occurred between North Carolina and Virginia, Virginia and Ohio, and between Mississippi and Alabama (Fig. 3). The haplotypes encountered in the populations are mostly separated by a few mutational changes, with the exception of mountain populations in Virginia, where some haplotypes are separated by more than 12 mutational events.

Unique indels at *ndhF* and *ycf6* can be used to deduce the direction of colonization from the Ozark Mountains in a northeastern direction to the Appalachian mountain range, or via southern Mississippi, Alabama and northern Florida in north eastern direction to the Appalachian mountain range (Fig. 3). The haplotype distribution map shows strong geographic structuring of haplotypes with distinct regional groups.

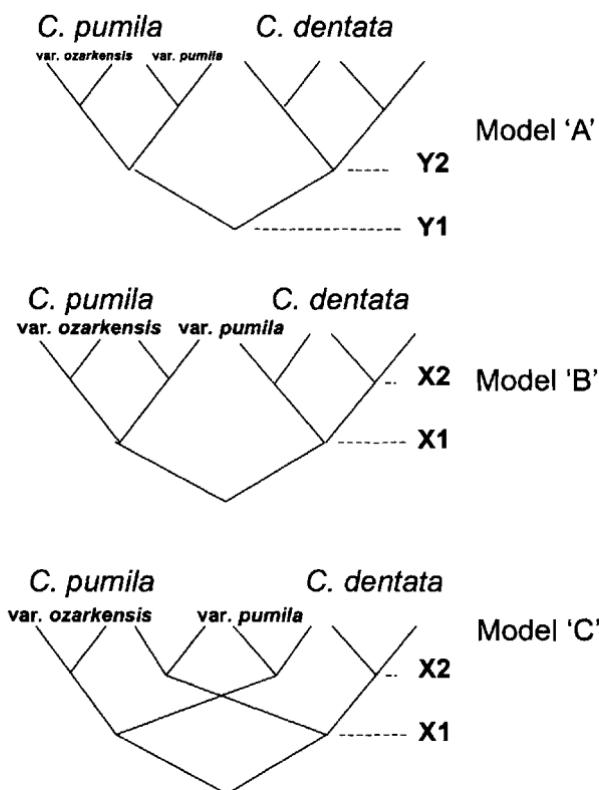
Low polymorphism and high fixation levels in woody species have been explained by low chloroplast mutation rates, low population size during glacial time and predominantly short-range dispersal [24]. One Virginia mountain chinkapin population holds haplotypes which differ by 1-12 mutational events from the common haplotype (Fig. 3: red star) of the Virginia mountain populations. These possibly have ancient origins mirroring larger historical population sizes. Sharing of chloroplast genomes has been reported in several plant species across species boundaries [24, 50, 53] generally as a result of introgression. This has also influenced the American *Castanea* species complex. Haplotype sharing can have occurred through hybridization in glacial refugia and/or during postglacial recolonization. This could explain the low number of mutational events and sharing of a unique insert at the *ndhF* region between one chinkapin haplotype from Virginia (red striped circle) and

another chinkapin haplotype from Georgia (purple triangle) with the *C. dentata* haplotypes, as illustrated in Fig. 3 using dotted lines.

High level of cpDNA diversity within *C. pumila* suggests that the species was little effected by glacial cycles, while fragmentation and range expansion perhaps by long-distance dispersal, have been important factors in the history of the American chestnut. Chestnut was once a leading species in eastern North America, often approaching 50% of the total stand basal area in upland forests from North Carolina to Connecticut [49]. It was one of the last deciduous tree species to arrive within its native range in the Holocene from a glacial refuge probably located on the southern coastal plane along rivers draining western Florida and Georgia. As the Quaternary is 2.5 Myr old, most DNA sequences will diverge little over the ice ages and few new mutations will distinguish postglacial haplotypes. Differences in such markers between postglacial haplotypes are likely due to sorting of mutations that have mostly originated in the more distant past [53]. Since there are more than 12 mutational differences between some haplotypes in Virginia, divergences in the order of many hundred thousand years are indicated for that region. The mixing of the two divergent cpDNA lineages reflects patterns which have established over several ice ages.

Phylogenetic analyses from combined data set indicated that on the North American continent the Ozark chinkapin, *C. pumila* var. *ozarkensis* is basal, sister to the group comprising *C. pumila* var. *pumila*, the Allegheny chinkapin and *C. dentata*, the American chestnut [39] with colonization patterns from the Ozarks to eastern states, or via southern states northward into the Appalachian mountain range (Fig. 3). The genetic structure observed in the *Castanea* species on the American continent reflects the impacts of the Tertiary characterized by most climatic and continental changes. The patchy distributions of the populations might reflect range contractions, expansions and migrations during successive glacial and interglacial periods as has been proposed for California oaks [21]. In North America, the only other oak species studied across its entire range is the northern red oak [41]. The species shows a lower genetic structure than the European oaks and is characterized by one major haplotype geographically widespread like *C. dentata*.

Several models can be used to explain the relationships within the American *Castanea* species complex (Fig. 4). Model 'A' of reciprocal



**Fig. 4.** Models explaining the relationship between the American *Castanea* species complex. Model 'A' of reciprocal monophyly indicates divergence between *C. dentata* and *C. pumila* at time point Y1. In models 'B' and 'C', *C. dentata* haplotypes are derived from within a clade of *C. pumila* between time points X1 and X2.

monophyly indicates divergence between *C. dentata* and *C. pumila* at time point Y1. This pattern would assume an equal number of mutational changes within each species. In model 'B' *C. dentata* haplotypes are derived from within a clade of *C. pumila* between time points X1 and X2. This model is consistent with the low number of *C. dentata* haplotypes (the daughter species) as compared to *C. pumila* (the mother species), and different number of mutational changes between the *C. pumila* lineages and *C. dentata*. In model 'C' two clades of haplotypes are shared between the species and the species are derived from a single large population that contained different haplotypes. Since the number of mutational changes and haplotypes within *C. dentata* is quite low in

comparison to other *Castanea* species, the most likely scenario might be explained by model 'C'. This model explains haplotype sharing and different number of mutational events between lineages.

The southern Appalachian region appears to be a refugial area for many species. Allozyme studies of widespread plant species in glaciated and non-glaciated regions of eastern North America report reduced genetic variation in glaciated regions compared to regions that remained ice-free during the Wisconsin glaciation [4, 40]. Allozyme studies of *Castanea* populations [7, 15] indicated high levels of genetic diversity in most chinkapin populations, while low levels were detected in *C. dentata* [29]. Little genetic exchange and low incidence of gene flow between the populations [15] was indicated. The western chinkapin populations have experienced considerable drift and accumulated unique mutations with little evidence of gene flow. Ancient haplotypes are evident in once more widespread southern populations suggesting a north-eastern ward expansion from those regions. Contemporary populations of American chestnut species might be the result of post-glacial expansion and thus show reduction in genetic diversity relative to now extinct refugial populations. If all present-day populations of *C. dentata* are indeed derived, both long-distance and stepping-stone models of migration predict a loss of genetic diversity in the recently founded populations [26]. During the last glacial maximum (LGM) *Pinus* dominated vegetation occurred extensively to 34°N and probably as far south as 30°N, with temperate hardwoods growing locally near the Lower Mississippi Valley. The Florida peninsula was occupied by open vegetation with warm-temperate species of *Pinus*. The paleovegetational data indicated highly extensive cooling in eastern North America at the LGM [31]. Sequence analysis of several nuclear regions (using primers based on EST sequence information; Dane, unpublished) indicated that *C. pumila* can be distinguished from *C. dentata* (and *C. sativa*) by unique substitutions, but clear discernable intraspecific phylogenetic patterns cannot be deduced yet. More information from the geological and climatic history of the earth and single-copy DNA sequences (SNPs) is needed in combination with the cpDNA data to provide a comprehensive view of evolutionary relationships and biogeography of this economically and ecologically important species.

## References

- [1] Avise JC. Molecular markers, natural history and evolution. Sunderland, MA: Sinauer Associates, 2004.
- [2] Bennett KD. Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology* 1990; 16: 11-21.
- [3] Boufford DE, Spongberg SA. Eastern Asian-eastern North American phytogeographical relationships. A history from the time of Linnaeus to the twentieth century. *Annals of the Missouri Botanical Garden* 1983; 70: 423-439.
- [4] Broyles SB. Postglacial migration and the loss of allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae). *Amer J of Bot* 1998; 85: 1091-1097.
- [5] Crepet WL, Daigleian CP. Castaneoid inflorescences from the Middle Eocene of Tennessee and the diagnostic value of pollen (at the subfamily level) in the Fagaceae. *Amer J Bot* 1980; 67: 739-757.
- [6] Dane F, Hawkins LK, Huang H. Genetic variation and population structure of *Castanea pumila* var. *ozarkensis*. *J Amer Soc for Hort Sci* 1999; 124: 666-670.
- [7] Dane F, Lang P, Huang H, Fu Y. Intercontinental genetic divergence of *Castanea* species in eastern Asia and eastern North America. *Heredity* 2003; 91: 314-321.
- [8] Davis MB. Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden* 1983; 70: 550-563.
- [9] Delcourt HR. Forests in peril. Tracking deciduous trees from ice age refuges into the greenhouse world. The McDonald and Woodward Publishing Company, Blacksburg, VA 2002.
- [10] Denk T, Grimm G, Stögerer K, Langer M, Hemleben V. The evolutionary history of *Fagus* in western Eurasia: Evidence from genes, morphology and the fossil record. *Plant Syst and Evol* 2002; 232: 213-236.
- [11] Denk T, Grimm GW, Hemleben V. Patterns of molecular and morphological differentiation in *Fagus* (Fagaceae): phylogenetic implications. *Amer J Bot* 2005; 92: 1006-1016.
- [12] Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ. Phylogeographic structure of white oaks throughout the European continent. *Genetics* 1999; 46: 1475-1487.
- [13] Fineschi S, Tauchini D, Villani F, Vendramin GG. Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* mill. (Fagaceae) throughout southern European countries. *Mol Ecol* 2000; 9: 1495-1503.
- [14] Frascaria N, Maggia L, Michaud M, Bousquet J. The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome* 1993; 36: 668-671.
- [15] Fu Y, Dane F. Allozyme variation in endangered Allegheny chinkapin, *Castanea pumila* var. *pumila*. *Ann Bot* 2003; 92: 223-230.
- [16] Gielly L, Taberlet P. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Mol Biol Evol* 1994; 11: 769-777.
- [17] Graham A. Outline of the origin and historical recognition of floristic affinities between Asia and eastern North America. In Graham A, ed. *Floristics and paleofloristic of Asia and eastern North America*. Amsterdam: Elsevier, 1972; 1-16.

- [18] Graham A. History of the vegetation: Cretaceous (Maastrichtian) Tertiary. In: F. O. N. A. E. Committee, ed. Flora of North America north of Mexico. New York: Oxford University Press, 1993; 57-70.
- [19] Graham A. Late Cretaceous and Cenozoic history of North American Vegetation. NY: Oxford University Press, 1999.
- [20] Graves AH. Relative blight resistance in species and hybrids of *Castanea*. *Phytopathology* 1950; 49: 1125-1131.
- [21] Grivet D, DeGuilloux M-F, Remy RJ, Sork VL. Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Mol Ecol* 15: 2006; 4085-4093.
- [22] Guo Q, Ricklefs R. Species richness in plant genera disjunct between temperate eastern Asia and North America. *Bot J Linn Soc* 2000; 134: 401-423.
- [23] Hamilton W. Cretaceous and Cenozoic history of the northern continents. *Ann Mo Bot Gard* 1983; 70: 440-458.
- [24] Heuertz MS, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L, Vendramin GG. Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (Oleaceae): roles of hybridization and life history traits. *Mol Ecol* 2006; 15: 2131-2140.
- [25] Hewitt GM. Some genetic consequences of ice ages and their role in divergence and speciation. *Biol J Linn Soc* 1996; 58: 247-276.
- [26] Hewitt GM. The genetic legacy of the European biota. *Biol J Linn Soc* 2000; 68: 87-112.
- [27] Huang H, Dane F, Norton JD. Allozyme diversity in Chinese, Seguin and American chestnut (*Castanea* spp.). *Theor Appl Genet* 1994; 88: 981-985.
- [28] Huang H, Carey WA, Dane F, Norton JD. 1996. Evaluation of Chinese chestnut cultivars for resistance to *Cryphonectria parasitica*. *Plant Disease* 1996; 80: 45-47.
- [29] Huang H, Dane F, Kubisiak TL. Allozyme and RAPD analysis of the genetic diversity and geographic variation in wild populations of the American chestnut (Fagaceae). *Amer J Bot* 1998; 85: 1013-1021.
- [30] International Organization of Palaeobotany. The plant fossil record (PFR v 2.2) database. <http://www.ibs.ucl.ac.uk/palaeo-bin/pfemap.pl>. 1997.
- [31] Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb III T, Williams JW, Hansen BCS. Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews* 2000; 19: 489-508.
- [32] Jaynes R. Chestnut. In: Moore J. ed. *Advances in Fruit Breeding*. West Lafayette, USA: Purdue University Press, 1975; 490-503.
- [33] Johnson GP. Revision of *Castanea* sect. *Balanocastanon* (Fagaceae). *Arnold Arboretum* 1988; 69: 25-49.
- [34] Krebs P, Conedera M, Pradella M, Torriani D, Felber M, Tinner W. Quaternary refugia of the sweet chestnut (*Castanea sativa* Mill.): an extended palynological approach. *Veget Hist Archaeolbotany* 2005; 13: 145-160.
- [35] King, RA, Ferris C. Chloroplast DNA phylogeography of *Alnus glutinosa*. *Mol Ecol* 1998; 7: 1151-1161.
- [36] Lang P, Huang H. Genetic variation and population structure of three endemic *Castanea* species in China. *Acta Horticulturae*. [Leuven, Belgium: International Society for Horticultural Science] 1999; 494: 269-276.

- [37] Lang P. Phylogeny and biogeography of *Castanea* (Fagaceae) based on chloroplast sequence data. PhD dissertation, Auburn University, AL, USA, 2004.
- [38] Lang P, Dane F, Kubisiak TL. Phylogeny of *Castanea* (Fagaceae) based on chloroplast trnT-L-F sequence data. *Tree Genet Genom* 2006; 2: 132-139.
- [39] Lang P, Dane F, Kubisiak TL, Huang H. Molecular evidence for an Asian origin and a unique westward migration of species in the genus *Castanea* via Europe to North America. *Mol Phylogen and Evol* 2007; 43: 49-59.
- [40] Lewis PO, Crawford DJ. Pleistocene refugium endemics exhibit greater allozyme diversity than widespread congeners in the genus *Polygonella* (Polynaceae). *Amer J Bot* 1995; 82: 141-149.
- [41] Magni CR, Ducoussou A, Caron H, Petit RJ, Kremer A. Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Mol Ecol* 2005; 14: 513-524.
- [42] Magri D, Vendramin GG, Comps B, Dupanloup I, Guberek T, Gomory D, Latalowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Remy RJ, De Beaulieu J-L. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist* 2006; 171: 199-221.
- [43] Manchester SR. Biogeographical relationships of North American Tertiary floras. *Annals of the Missouri Botanical Garden* 1999; 86: 472-522.
- [44] Manos PS, Doyle JJ, Nixon KC. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Mol Phylogen Evol* 1999; 12: 333-349.
- [45] Manos PS, Stanford AM. The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the Northern Hemisphere. *Int J Plant Sci* 2001; 162: S77-S93.
- [46] Manos PS., Zhou ZK, Cannon CH. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *Int J Plant Sci* 2001b; 162: 1361-1379.
- [47] Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S. Intercontinental population structure of the chestnut blight fungus *Cryphonectria parasitica*. *Mycologia* 1996; 88: 179-190.
- [48] Morimoto Y, Shimada T, Ozaki T, Nomura K, Araki H, Yoshida M. Identification and relationship of Japanese chestnut varieties by RAPD and southern blot analysis. *Sci. Rept. Fac. Agr. Kobe Univ.* 1997; 22: 63-69 (in Japanese).
- [49] Paillet FL. Chestnut: history and ecology of a transformed species. *Journal of Biogeography* 2002; 29: 1517-1530.
- [50] Palme AE, Vendramin GG. Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. *Mol Ecol* 2002; 11: 1769-1779.
- [51] Payne JA, Miller G, Johnson GP, Senter SD. *Castanea pumila* (L.) Mill.: An underused native nut tree. *HortScience* 1994; 239: 62, 130-131.
- [52] Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducoussou A, Kremer A. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc Nat Acad Sci USA* 1997; 94: 9996-10001.
- [53] Petit RJ, Csaikl U, Bordacs S, Burg K, Coart E, Cottrell J, Van Dam B, Deans JD, Glaz I, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Goicoechea PG, Jensen JS, Konig A, Lowe AJ, Madsen SF, Maryas G, Munro RC, Olalde M, Pemonge M-H,

- Popescu F, Slade D, Tabbener H, Tauchini D, Ziegenhagen B, Kremer A. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2,600 populations. *Forest Ecology and Management* 2002; 156: 5-26.
- [54] Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 2003; 300: 1563-1565
- [55] Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol Ecol* 2005; 14: 689-701.
- [56] Qian H, Ricklefs RE. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature (London)* 2000; 407: 180-186.
- [57] Qian H. A comparison of generic endemism of vascular plants between East Asia and North America. *Int J Plant Sci* 2001; 162: 191-199.
- [58] Qian H. A comparison of the taxonomic richness of temperate plants in East Asia and North America. *Amer J Bot* 2002; 89: 1818-1825.
- [59] Saucier JR. American Chestnut - an American wood. 1975.
- [60] Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. Comparative phylogeography of unglaciated eastern North America. *Mol Ecol* 2006; 15: 4261-4293.
- [61] Stanford A. The Biogeography and phylogeny of *Castanea*, *Fagus*, and *Juglans* based on matK and ITS Sequence Data. PhD dissertation, UNC, Chapel Hill, 1998.
- [62] Stanford, AM, Harden R, Parks CR. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *Amer J Bot* 2000; 87: 872-882.
- [63] Tanaka T, Yamamoto T, Suzuki M. Genetic diversity of *Castanea crenata* in northern Japan assessed by SSR markers. *Breeding Science* 2005; 55: 271-277.
- [64] Tidwell WD. Common fossil plants of western North America. Washington, Smithsonian Institution Press: 1998.
- [65] Tiffney, BH. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the northern hemisphere. *Journal of the Arnold Arboretum* 1985a; 66: 243-273.
- [66] Tiffney, BH. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum* 1985b; 66: 73-94.
- [67] Villani F, Pigliucci M, Benedettelli S, Cherubini M. Genetic differentiation among Turkish chestnut (*Castanea sativa* Mill.) populations. *Heredity* 1991; 66: 131-136.
- [68] Villani F, Pigliucci M, Cherubini M. Evolution of *Castanea sativa* Mill. in Turkey and Europe. *Genetical Research* 1994; 63: 109-116.
- [69] Watts WA, Hansen BCS. Pre-Holocene and Holocene pollen records of vegetation history from the Florida peninsula and their climatic implications. *Paleogeogr., Paleoclimatol., Paleoecol.* 1994; 109:163-176.
- [70] Wen J. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* 1999; 30: 421-455.
- [71] Wolfe JA. A paleobotanical interpretation of Tertiary climate in the Northern Hemisphere. *American Scientist* 1978; 66: 694-703.
- [72] Wolfe JA. Vicariance biogeography of angiosperms in relation to paleobotanical data. In: Rosen DE, ed. *Vicariance biogeography*. New York, NY: Columbia University Press 1981; 413-427.

- [73] Wolfe JA. Distribution of major vegetational types during the Tertiary. In: Broecker WS. ed. The Carbon Cycle and Atmospheric CO<sub>2</sub>: Natural Variation Archean to Present. Washington American Geophysical Union:1985; 357-375.
- [74] Zhou, Z. Fossils of the Fagaceae and their implications in systematics and biogeography. *Acta Phytotaxonomica Sinica* 1999; 37: 369-385.

# Cytogenetic and Molecular Evidences on the Evolutionary Relationships Among *Arachis* Species

GRACIELA I. LAVIA<sup>1,2</sup>, AVELIANO FERNÁNDEZ<sup>1,3</sup> and JOSE GUILLERMO SEIJO<sup>1,3</sup>

<sup>1</sup>Instituto de Botánica del Nordeste, Corrientes, Argentina

<sup>2</sup>Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste, Corrientes, Argentina

<sup>3</sup>Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

## ABSTRACT

The new cytogenetic and molecular data relating to the peanut, its secondary/tertiary gene pool, and the *Arachis* species that are being used as forage are integrated with the large existent biosystematic knowledge to combine information on the genetic origin of the cultigen, species relationships within *Arachis* section, the evolutionary relatedness of  $x=9$ , and the origin of polyploids of Rhizomatosae and Caulorrhizae sections. The allopolyploid nature of peanut was confirmed as having two complements with heterochromatic bands and two without them in all botanical varieties. Gene mapping by FISH, most molecular markers and amphidiploid resynthesis support *A. duranensis* and *A. ipaensis* as the most probable genome donors of *A. hypogaea*. Moreover, *A. monticola* is revealed as the wild tetraploid from which the peanut derived upon domestication rather than a wild derivative. Although the close relatedness of some species has been established, a clear phylogeny of section

Arachis has still to be delineated. Basic chromosome number  $x=9$  may have arisen from  $x=10$  independently in different sections. Rhizomatous species appears as a polyphyletic group in which the diploid *A. burkartii* seems to have a genome constitution differing from that of the tetraploids. Meiotic behavior of *A. pintoi* suggests that spontaneous polyploidization in *Arachis* might have occurred through unreduced gametes.

**Key Words:** Peanut, origin of polyploids, genome relationships, basic chromosome numbers

## INTRODUCTION

The genus *Arachis* (Leguminosae) is native to South America and has 80 formally recognized species. These entities were arranged into nine sections based on morphology, geographic distribution and cross compatibility [48, 107].

The species are distributed from South of Amazonas region towards De La Plata river at 34°S, and from the Atlantic Coast to the foothills of the Andes mountains, below 800 m, except *A. duranensis*, *A. batizocoi* and *A. monticola* that have been collected up to a height of 1,100-1,500 m. They grow spontaneously in a wide range of environments ranging from pure white or red sand to stony or clayey soils, and from xerophytic (Chaco and Cerrado) to very rainy climates (Pantanal).

It has been suggested that the genus originated in the Amambay mountains, in the boundary between Brazil and Paraguay. In this region grows *A. guaranitica* (Section Trierectoides) that presents several characters considered as ancestrals in the genus – trifoliated leaves, tuberiform hypocotyls, erect stems and welded stipules that form a long basal tube that covers up to two internodes of the stems [48]. The center of morphological and chromosomal variation is localized east of the Pantanal, on the Paraguay river [13, 48, 57], where species of five sections – Trierectoides, Erectoides, Rhizomatosa, Procumbentes and *Arachis* [48], and ten out of the 11 SAT chromosome types [13, 57] were found. Moreover, the prediction of species richness made by computer modeling, using 2,175 georeferenced populations and 36 climate variables, detected three high spots of wild *Arachis* diversity. All of them are included in Brazil, around Cuiaba and Campo Grande in Mato Grosso state, and in Serra Geral of Goias state [35].

Three species of *Arachis* are being cultivated by modern agriculturists. All of them belong to different sections, i.e., *A. hypogaea* (peanut) from section *Arachis*, *A. glabrata* from section *Rhizomatosa*, and *A. pintoi*

from section Caulorrhizae. Peanut is one of the most important grain legumes consumed directly or industrialized by humans, and it has been cultivated for over 5,000 years. Even though it originated in South America, it is adapted to a wide ecological range of tropical and subtropical regions and under diverse agricultural production systems in Asia, Africa, and the Americas [30]. The other two species are used as forage in the tropics and subtropics mainly of the Americas [14].

Significant progress has been made over the last 60 years in the biosystematics of cultivated and wild species of *Arachis*. Particular attention has been given to the evaluation of peanut germplasm and a number of lines with desirable sources of resistance to both biotic and abiotic hazards, and other valuable agronomic characters such as yield, fatty acid composition and flavors have been identified [1, 34, 67]. A considerable number of studies have been also undertaken to collect and characterize *Arachis* species with a high potential as forage, such as *A. glabrata* and *A. pintoi* [9].

Peanut genetic resources are available from landraces of *A. hypogaea* in various centers of diversity, breeding lines or material developed in different peanut producing areas. However, it has been generally recognized that the genetic variability of *A. hypogaea* is very limited [25, 39, 74]. Wild species are of particular concern for peanut improvement because they present a wide range of genetic diversity, which includes many desirable traits [94]. Several efforts have been directed at understanding the evolutionary relationships among wild species, and between them and *A. hypogaea*. Nevertheless, the use of the germplasm in genetic improvement of *A. hypogaea* is still scarce, restricting progress in broadening the genetic base of the cultigen.

This chapter reviews old and new problems associated with *Arachis*, such as species relationships within *Arachis* section, the genetic origin of peanut, the phylogenetic relationships of  $x=9$  species together with the mechanism that may have been involved in the reduction of chromosome number, and the origin of polyploids of *Rhizomatosae* and *Caulorrhizae* sections in the light of new cytogenetic and molecular genetic data. The discussion presented acquired relevance since it takes in account the monographic treatment of the genus [48] and the recently described new species [107].

## SPECIES RELATIONSHIPS WITHIN ARACHIS SECTION

Arachis section is distinguished from other sections by having axonomorphic roots, no rhizomes, vertical pegs and flowers without red veins in the back of the standard [48]. It is the largest section genetically diverse, with the widest geographical distribution, and several derived characters, including the species *A. hypogaea*.

The section is currently composed of 32 species, including *A. chiquitana* that still have to be formally incorporated in to this group [58]. It has the same distribution of the genus, except in the east, since species only grow up to the Tocantins, Paranaiba and Parana rivers. Some populations of *A. stenosperma* that grow near the Atlantic coast are beyond this area, but it is currently believed that they were transported from its original location in the Pantanal by ancient humans [48].

Different laboratories around the world have been working on several topics mainly pertaining to: (1) genome composition of the species included in section Arachis, (2) the relationship between species of different genomes, and finally (3) the relationship among species harboring the same genome?

These topics have been studied based mainly on morphology, cytological evidence, crossing experiments and chromosome pairing analyses in interspecific hybrids. These analyses have revealed that most of the species are diploids (27) or tetraploids (2) based on  $x=10$ , and that three are diploids based on  $x=9$ . They also provided evidence for the arrangement of the entities included in section Arachis as belonging to one of three different genome types.

The A genome species are characterized by symmetric karyotypes and the presence of a distinctly small chromosome pair with a differential condensation pattern and a large heterochromatic band in the centromere, namely, A chromosomes [13, 32, 33, 88]. This genome is present in 16 species, most of them perennials, except *A. duranensis*, *A. schinini* and *A. trinitensis* [13, 55, 59, 76]. The entities belonging to the non A genome (or usually referred as B genome) also have symmetric karyotypes but lack the small A chromosomes. This group includes 13 species, which in spite of being all annuals, formed a more heterogeneous group than those with the A genome [13, 48, 55, 56, 59, 90, 95, 98, 101]. Finally, the D genome, which has solely been proposed for *A. glandulifera*, is characterized by asymmetric karyotype and absence of A chromosomes.

**Table 1.** Genome composition of section Arachis species and their life cycle [13, 48, 57, 76, 101, 107]

	A genome perennials	A genome annuals	Non A genome annuals	D genome annuals
x=10	<i>A. cardenasii</i>	<i>A. duranensis</i>	<i>A. batizocoi</i>	<i>A. glandulifera</i>
	<i>A. chiquitana</i>	<i>A. schinini</i>	<i>A. benensis</i>	
	<i>A. correntina</i>	<i>A. trinitensis</i>	<i>A. cruziana</i>	
	<i>A. diogoi</i>		<i>A. gregoryi</i>	
	<i>A. helodes</i>		<i>A. hoehnei</i>	
	<i>A. herzogii</i>		<i>A. ipaensis</i>	
	<i>A. kempff-mercadoi</i>		<i>A. krapovickasii</i>	
	<i>A. kuhlmannii</i>		<i>A. magna</i>	
	<i>A. linearifolia</i>		<i>A. valida</i>	
	<i>A. microsperma</i>		<i>A. williamsii</i>	
x=9	<i>A. simpsonii</i>		<i>A. decora</i>	
	<i>A. stenosperma</i>		<i>A. palustris</i>	
	<i>A. villosa</i>		<i>A. praecox</i>	

[101]. The tetraploids are annual allopolyploids having the AABB genome formulae [33, 98]. Table 1 summarizes the species genome composition and their life cycle. Crosses among the entities of different genomes are usually unsuccessful, or when hybrid seeds are obtained, their seedlings are non viable, F1 plants never produce flowers, but even if they produce, pollen viability is very low [20].

The arrangement of species in three different genomes was supported by several studies using heterochromatic banding patterns, ribosomal DNA (rDNA) loci mapping by fluorescent in situ hybridization – FISH [88, Seijo & Robledo unpubl.] and molecular markers [40, 69, 70, 106]. These studies usually revealed a more homogeneous group within A genome species compared to that of the non A genome species, except in an amplified fragment length polymorphism (AFLP) analysis, in which the authors found that non A genome species have less interspecific variation than A genome species [69]. The latter result is unexpected since when morphological, cytological and cross compatibility data were put together or considered together, A genome species appeared as a

more compact group. Moreover, it has been even suggested that different, still not well characterized, genome groups may be included within the non A group of species [88], although further research is required to fully support this hypothesis. The so far monospecific D genome appears to have less genetic distance with the non A genome than with the A genome species according to the AFLP and the plastid *trnT-F* region data [106].

There is still a great deficit of data regarding the species relationship within each genome group. Although some groups of species have been detected as more closely related within each genome type, the fact that in many cases the interpopulation cross compatibility is similar to that observed in interspecific crosses [20, 48], that most species have similar karyotypes [13, 56, 57, 60] and, that the intraspecific variability of some morphological characters overlaps the limit of species [48, 107], hampered the establishment of clear evolutionary trends among species within a particular genome.

With the objective of finding species specific markers and to understand the phylogenetic pattern underlying in section *Arachis*, several cytogenetic and molecular analyses have been performed. Differential staining of chromosome with fluorochromes has shown that the entities harboring the A chromosome pair also have heterochromatic bands in all (or almost all) chromosomes of their complements, resulting in a homogeneous group also at the sub-chromosome structure [88]. However, the species within the non A genome showed more diversification because some of them have centromeric bands in all or most of their chromosomes (examples are *A. batizocoi* and *A. hoehnei*), while in other species the chromosomes are deprived of large blocks of heterochromatin (*A. ipaensis*, *A. magna* and *A. williamsii*) [88, Seijo unpubl.]. Therefore, the heterochromatin patterns evidenced that the non A genome group may be actually composed of species that differ in their genome constitution. *Arachis glandulifera* has a very particular heterochromatic pattern since, besides having large centromeric bands, it presents a series of interstitial bands, which are absent in all the so far analyzed species of section *Arachis* [87].

Physical mapping of rDNA by FISH evidenced that the 5S loci were always located at a paracentromeric position of only one chromosome pair, irrespective of whether the species harbored the A, B or D genome. Contrarily, the 18S–25S sites are diverse in size, number, and location [82, 88]. Within the A genome species, the loci varied from 4 to 6, but

they always mapped in a paracentromeric position. However, in the B and D genomes the sites ranged from 2 to 8 and they mapped in paracentric, interstitial or terminal chromosome portions. The variability in the 18S–25S arrays, together with the heterochromatic patterns, allowed the discrimination of most of the analyzed species within each of the three genomes [88, Robledo and Seijo unpubl.].

Molecular markers carried out within the section were based on total seed protein [3, 93], isozyme variation [22, 64, 103] and DNA marker analysis. Among the latter, most studies, including restriction fragment length polymorphism (RFLP) [25, 39, 40, 74], random amplified polymorphic DNA (RAPDs) [29, 25, 53], single sequence repeats (SSR) [84] and AFLP [16] analyzed species of many sections, and in spite of providing valuable data to support the infraspecific arrangement of the genus into nine sections, the number of species of section Arachis considered (usually less than 10) was insufficient to produce a conclusive robust phylogeny. Since most investigations were mainly focused on the search for the wild diploids that have participated in the origin of the allotetraploids of the section, wild species relatedness was inadequately discussed.

Recent studies on a larger number of diploid species were particularly focused on the species relationship within section Arachis. The most extensive report analysed SSRs in 27 wild species [71], in which the entities proposed as means group of species is believed to have the same genome having the same genome composition clustered together, as expected. Nevertheless, accessions of *A. duranensis* were grouped with the tetraploids, apart from the A genome diploids [70]. This clear delimitation of *A. duranensis* was also observed by Milla et al. [69] using AFLP. Large genomic differences between *A. duranensis* and the other A genome species would be unexpected given that this species crosses readily with other entities of the group [48, 102]. Also, it has a series of substitutions and indels in the plastid *tRNA*-F region, which are characteristic of the A genome species [106]. Other species associations within this genome are not clear since the clusters of species varied greatly depending on the DNA marker used.

Within the non A genome, SSR markers showed very limited genetic distance between *A. ipaensis*, *A. magna*, *A. williamsii* and *A. gregoryi* [70], in accordance with their chromosomal similarity [88] and cross compatibility [Simpson, pers. com.]. This group of species seems to be genetically different from that represented by *A. batizocoi*. In a recent

report of crossing assays, the genomes of *A. ipaensis* and *A. williamsii* have been proposed to be intermediate between the A and non A genome when *A. duranensis* and *A. batizocoi* were used as tester species [106]. Moreover, different reports using molecular markers [39, 69, 105] or molecular cytogenetics [88] suggested that *A. ipaensis* and *A. williamsii* are genetically different from the non A genome represented by *A. batizocoi*. In contrast, the analysis of the plastid *trnT-F* region showed that *A. ipaensis* nested within the non A genome lineage since it lacks the characteristic indels of the A genome species, however, it was unable to resolve the position of *A. williamsii* [106].

*Arachis batizocoi* and *A. cruziana* seem to be very closely related as revealed by AFLP and the plastid *trnT-F* markers [106] and by the fact that both taxa have heterochromatic bands in most of the chromosomes of their complements [82, 88]. One of the great controversies is centered on the genome identity of *A. hoehnei*. This species was traditionally included among the non A genome group, since it lacks the small A chromosome [13]. Moreover, its karyotype is very similar to that of *A. batizocoi*, because it has heterochromatic bands in the centromere of most chromosomes of the complement [Seijo unpubl.] and both species are cross compatible [106]. However, it possesses the characteristic indels pattern of the A genome species in the plastid *trnT-F* region.

## **GENETIC ORIGIN OF *A. HYPOGAEA***

*Arachis hypogaea* is a cultigen that shows a great morphological and physiological variability and is mainly adapted to the tropics and subtropics, although its cultivation extends into the warm areas of the temperate regions [26]. It was taxonomically divided in two subspecies, *hypogaea* and *fastigiata* Waldron based on cross compatibility and morphological characters, such as the overall growth habit, whether or not flowers occur on the main stem, and the interspersion pattern of reproductive and vegetative lateral branches. Subspecies *hypogaea* was further divided into *hypogaea* and *hirsuta* Köhler varieties, while ssp. *fastigiata* into *fastigiata*, *vulgaris* Harz, *aequatoriana* Krapov. et W.C. Gregory, and *peruviana* Krapov. et W.C. Gregory varieties [48]. This infraspecific classification was established mainly on indumentum of the leaflets, reticulum of the fruits and number of seeds per fruit. Also, a large number of landraces have been described, based mainly on the morphological characters of the fruits and seeds.

Several questions regarding the origin and domestication of peanut remain unresolved, such as (1) did the subspecies or varieties of the cultigen originate by a single or multiple events of hybridization/polyploidization? (2) which diploid species participated in the origin of *A. hypogaea* and (3) is peanut the result of an extant domesticated wild tetraploid.

### **Single vs. Multiple Origin**

According to the species distribution and morphological variability of the landraces, south of Bolivia and north of Argentina was traditionally proposed as the primary candidate location for the domestication of peanut [48], despite the fact that the oldest archeological fruits of *A. hypogaea*, dated around 5,000 years BP, were found near the Peruvian coast of the Huarmey valley (Fig. 1, [4]). This proposition is strongly supported since no wild *Arachis* species is presently found, there nor are they presumed to have occurred there spontaneously in earlier times.

However, pod samples found in the Casma valley (Pacific coast of Perú) dating between 3,500 and 3,800 years BP (Fig. 1), which have a very strong resemblance to three wild species presently found in southern Bolivia (*A. ipaensis*), western Brazil and east Bolivia (*A. magna*) and north of Argentina (*A. monticola*), have led to the proposal that ancient people of northwest Peru may have used wild *Arachis* species and probably cultivated them. The orchards of those people may have served as another possible site for the origin of *A. hypogaea* [89].

A multiple origin for different subspecies was early advanced based on morphology [48] and cross compatibility experiments [90, 91] since both taxa are not fully fertile. Krapovickas and Gregory [48] proposed an initial center of origin for the ssp. *hypogaea* in SW Bolivia and NW Argentina and a later differentiation of the ssp. *fastigiata* northernwards. This hypothesis also considers that introgression with sympatric wild species may have occurred, as the cultigen was dispersed along South America by human beings, leading to the origin of different subspecies and varieties [47]. However, a single origin of *A. hypogaea* varieties is mainly based in the low genetic variability detected in the amphidiploid [25, 40, 74].

Considerable biochemical and molecular data have been analyzed in the search of evidences for the events and mechanisms that led to the origin and diversification of the subspecies and varieties of *A. hypogaea*.

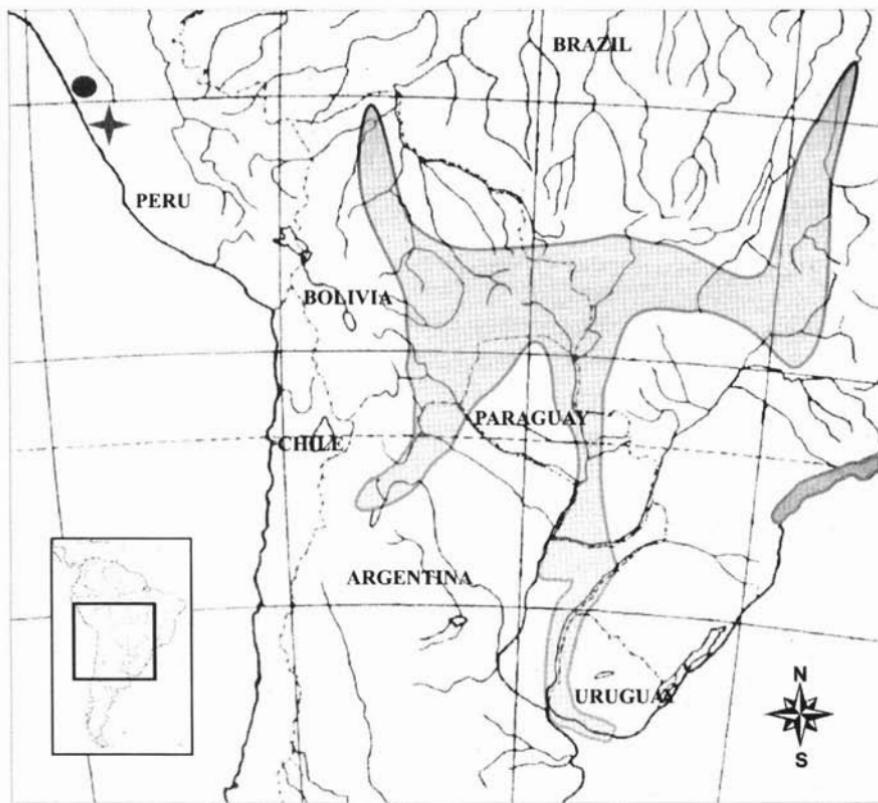


Fig. 1. Geographical distribution of the *Arachis* section and archeological *Arachis*. Gray shaded indicates the area of *Arachis* section. ● = Casma valley, ✪ = Huarmey valley.

Among them, protein based markers have produced results that would support both, the single and multiple origin of *A. hypogaea*. Lanham et al. [54] and Grosso et al. [24] found that the two subspecies may be distinguished using seed proteins, while Lu and Pickersgill [64] observed consistent dissimilarities in just two of the 13 putative isozyme loci analyzed and, Lacks and Stalker [52] were unable to detect any association between isozyme variation and subspecies or varieties. Therefore, available data is not conclusive.

Early reports using RFLP [25, 40], RAPD [25] and AFLP [28] have shown very limited polymorphism at the DNA level among the various botanical varieties of the two subspecies, and suggested that *A. hypogaea* may have originated as the result of a single polyploidization event. Further explanations of this low genetic variability have been attributed to (1) a

recent polyploidization event including one or few individuals of each diploid parental species associated with self-pollination [25], (2) chromosome doubling has built strong reproductive isolation, preventing gene flow between the tetraploid and related diploid species [109], (3) use of few accessions with low representation of the world germplasm collection, most of them available after several cycles of stock regeneration, resulting in a narrow genetic base [96], or 4) inadequacy of the procedures used to reveal the real variability [96].

Recent analyses on RAPD, AFLP and SSR have revealed some DNA polymorphism in *A. hypogaea* [12, 27, 31, 70, 71, 84]. The results obtained allowed the assignation of each peanut accession to specific botanical varieties with few exceptions. Discriminant function analysis of SSR markers highlights the relative integrity of the varieties, suggesting little introgression among them, and reveals that taxa delimitation on the basis of morphological characters is supported, although their hierarchical organization deserves revision [12, 70]. Based on these results, the early hypothesis of multiple origins has regained new support.

Classical chromosome analysis showed very similar karyotypes in all the varieties and it was unable to discriminate varieties or subspecies [61]. Chromosome mapping of heterochromatin and 45S and 5S rDNAs loci by FISH demonstrated a highly conserved karyotype in all the varieties, which is composed by half of the chromosomes with heterochromatic bands and the remaining without them. These complements also have exactly the same number, position and size of the rDNA loci. These results evidenced a great similarity of the infraspecific taxa of peanut and demonstrated that all the varieties may have originated from a hybridization of a diploid species bearing heterochromatic bands and another without them [88]. Since all the A genome species so far analysed have heterochromatic bands, the hypothesis which was based on SSR, that landraces of var. *hypogaea* could be derived from a straight genome duplication of a diploid species with A genome [70, 74] may have little support. Moreover, genome in situ hybridization (GISH) has also given evidence in favor of the hypothesis that considers all the varieties have been composed by two A and two non A chromosome complements [Seijo et al. unpubl.]. Since all the botanical varieties of the cultigen showed a similar pattern of genomic hybridization with the diploids assayed, it was suggested that the wild species which have participated in their origin would be the same. These results may be taken to interpret that there had been a unique

allotetraploid plant from which all the varieties presently known had originated, or alternatively, that there had been different allotetraploid plants, but all of them originated from the same two diploid species [88].

Genomic hybridization by GISH also showed the absence of alien chromosomes or chromosome segments in the complements of *A. hypogaea*. This result may be interpreted as additional evidence that further hybridizations with other diploid species did not occur after the establishment of the first amphidiploid plants [Seijo et al. unpubl.]. However, since it has been demonstrated that chromosome recombination, and not chromosome substitution, was the mechanism of introgression of diploid species characters into *A. hypogaea* genome [15, 106], limited introgression from another diploid species during domestication and dispersion of the ancestral tetraploid could not be fully ruled out. Comparison of the subspecies and varieties by means of high density linkage maps based on molecular markers, such as those being constructed based on RFLP [5] and microsatellites [71], would be required to test this hypothesis.

## **Wild Progenitors**

Peanut is an allotetraploid or segmental allotetraploid with  $2n = 4x$  ( $2A + 2B$ ) = 40 [13, 33, 98, 104] and considering wild species related to the cultigen, it is agreed that peanut progenitors should come from within section Arachis. However, the recognition of the wild progenitors of peanut has been a subject of great debate involving more than 10 diploid species belonging to either the A or non A genomes.

Since the early 1950s, when the first hybrid between *A. hypogaea* and the diploid *A. correntina* was obtained [43, 51], several other diploid species such as *A. villosa*, *A. cardenasii*, *A. helodes*, and *A. duranensis*, have produced hybrids with *A. hypogaea* and, thus, were proposed as putative progenitors of the tetraploids [44, 46, 85, 90, 97, 99]. According to morphological similarities and interspecific cross compatibilities, Gregory and Gregory [19] were the first to suggest *A. duranensis* and *A. cardenasii* as possible progenitors of the cultigen, but both are currently considered A genome species. On the other hand, Smartt et al. [98] advanced the possibility that *A. cardenasii* and *A. batizocoi*, with the proposed A and non A genome respectively, have many of the characters expected for the genome donors of *A. hypogaea*. At that time, *A. batizocoi* was the only species known with non A genome and was cross compatible

with *A. hypogaea*. Also, it produces hybrids with many A genome species, which varied in the degree of sterility, but all of them achieved fertility restoration after chromosome doubling [95].

In the first half of the 1990s, intensive efforts were made using biochemical and molecular data, with the objective of shedding light on which diploid species participated in the origin of peanut. RFLP data suggested that *A. batizocoi* would not be closely related to the cultigen, and that *A. duranensis* and *A. ipaensis* were the closest relatives of *A. hypogaea* [39], although the genome constitution of the latter diploid species was unknown. Further studies [40, 74] using molecular marker techniques supported the close relationship of *A. duranensis* and the more distant affinity of *A. batizocoi* with the cultigen. For this reason, it is generally argued that the latter wild species cannot be the donor of the non A genome to *A. hypogaea*.

Even though *A. batizocoi* could be easily eliminated as a progenitor of *A. hypogaea*, it was not clear which species without A chromosome pair was to be considered as parent of the tetraploid. This was finally resolved when it was determined that besides *A. batizocoi*, other diploid species of section Arachis had karyotypes without the A chromosomes [13], including *Arachis ipaensis*, which has been reported as having close affinity with the cultigen by RFLP [39].

Newer reports on molecular markers considered *A. villosa* (A genome) and *A. ipaensis* (non A genome) as the best candidates based on RAPDs and ISSR (inter-simple sequence repeat) analyses [84], while AFLP analysis demonstrated that at least three diploid species belonging to the A genome (*A. duranensis*, *A. simpsonii* and *A. helodes*) and two of the non A genome (*A. ipaensis* and *A. williamsii*) present small genetic distance with the cultigen [69]. Microsatellite analysis has shown that, although *A. duranensis* and *A. ipaensis* are closely related to the cultigen, a group of other species belonging to the A or non A genomes can be also referred as possible genome donors [70]. Moreover, classical chromosome studies have brought *A. williamsii* and *A. trinitensis* as new genome donor candidates [55, 49].

By determining the number of rDNA loci in diploid and tetraploid species and by GISH experiments, it has been proposed that *A. ipaensis* and *A. villosa* would be the most probable genome donors of *A. monticola* [82, 83]. However, as those authors found twice the number of 45S rDNA sites in *A. hypogaea* than in *A. monticola*, they contemplated the

hypothesis of amplification of minor gene clusters in the cultigen to maintain the same diploids as the parents of peanut. Physical mapping of these genes and the analysis of heterochromatin patterns in all the botanical varieties revealed exactly the same number and position of the loci in the chromosomes without heterochromatic bands of *A. hypogaea* (non A complement), compared to those observed in *A. ipaensis* [88]. *Arachis batizocoi* was discarded as genome donor of the non A complement of the tetraploids because it has a different mapping pattern of rDNA loci and heterochromatic bands in the centromeres of almost all chromosomes of the complement. On the other hand, the loci in the banded chromosomes of the cultigen (A complement) were the same as those observed in *A. duranensis*, *A. villosa* and *A. correntina*. From these three A genome species, *A. duranensis* was considered as the most probable genome donor, because it is annual and grows in the area which has been proposed as the most probable center of origin of the tetraploids. Moreover, the other two species grow in a region 1,000 km further from the center of origin of the cultigen, separated by the large xerophytic Chaco region which is deprived of *Arachis* species in most of its total range area. This distance is very significant for *Arachis* species since it was established that the mean rate of dispersion for an individual plant is around one or two meters per year due to geocarpy. Occasional dispersion by fluvialtile hydrochory is unexpected since they are distributed downstream at the center of origin in De La Plata river basin [88].

Double GISH also support the candidacy of *A. duranensis* and *A. ipaensis* as the genome donors of *A. hypogaea* since their DNA showed the most intense and uniform pattern of hybridization onto the chromosomes of the cultigen [Seijo et al. unpubl.]. Moreover, when *A. ipaensis* and *A. duranensis* are considered as the genome donors of *A. hypogaea*, a postpolyplodization rDNA loci duplication or a gene cluster amplification is not necessary because the number of sites found in the cultigen is just the sum of those observed in the diploids.

Phylogenetic analysis of the *tmtF* region of the plastid genomes in hybrids and their parental species demonstrated that these genomes in *Arachis* are maternally inherited. According to this analysis, the maternal donor of the cultigen has to be from one of the A genome species, since *A. hypogaea* nested means grouped with the A genome species analyzed, being *A. duranensis* among them [106]. However, the interspecific hybrids between these two species were predominantly successful when *A. duranensis* was used as male parent [20, 102]. In

addition, resynthesis of the allopolyploid by artificial hybridization and further chromosome doubling was successfully achieved when *A. duranensis* was used as the male donor, and *A. ipaensis* as the female donor [11]. The production of F1 hybrid between *A. duranensis* and *A. ipaensis*, together with the fact that the artificial amphidiploid obtained by chromosome doubling of the resultant F1 hybridizes with individuals of all the varieties of the cultigen and produce fertile offspring [11], additionally support the hypothesis that these diploids have been involved in the origin of the cultigen.

### **Is *A. monticola* the Wild Tetraploid Ancestor of *A. hypogaea*?**

*Arachis monticola* was described by Krapovickas and Rigoni [44] and was considered either as a wild tetraploid derivative of hybridization between *A. hypogaea* and a wild diploid species, or as the direct wild progenitor of the cultigen [48].

*A. monticola* as a derivative can be argued on the hypothesis that considers can be argued (means rejected) that this tetraploid actually grows is geographically isolated from all other diploid species, except some populations of *A. duranensis*. Therefore, if *A. monticola* is a wild derivative of *A. hypogaea* due to hybridization with a diploid species, the best candidate is *A. duranensis*. From a karyotype point of view and considering that *A. hypogaea* has half the number of chromosomes with heterochromatic bands and that *A. duranensis* has centromeric bands in all chromosomes, a hybrid between both species should have more than the half the number of chromosomes with heterochromatin. However, this is not the case, since the number of chromosomes with heterochromatic bands in *A. monticola* is exactly the same as that observed in the cultigen. Moreover, the 5S and 45S rDNA loci mapped identically in both species [88] and the DNA of *A. duranensis* hybridized only with 20 chromosomes of *A. monticola* in GISH experiments [Seijo et al. unpubl.]. Therefore, the only possibility that remains to be tested in order to completely reject the hypothesis of *A. monticola* as being a derivative of the cultigen, is that which considers introgression from *A. duranensis* into the genome of *A. hypogaea* involving just few genes (including those that conferred wild characteristics).

Considering molecular markers, *A. monticola* presents identical RFLP [39], AFLP [16, 69] and SSR [70] patterns than *A. hypogaea*. This is not completely unexpected since the overall morphology of the two taxa is

similar, except the isthmus that separates the seeds in *A. monticola*. Moreover, both species showed a high degree of similarity of protein eletrophoretic patterns [38], same karyotypes [13], and are known to be interfertile with no apparent sterility of the F1 hybrids [37]. Some authors, based on chromosome pairing behavior of *A. hypogaea* and *A. monticola* hybrids, have even suggested that separating these two taxa into individual species may not be warranted [37, 100].

Resynthesis of the allotetraploid from the hybrid *A. ipaensis* x *A. duranensis* and their morphological similarity with *A. monticola* [11] support hypothesis that *A. monticola* is the immediate wild ancestor of *A. hypogaea*.

## **ORIGIN OF X=9 BASIC CHROMOSOME NUMBER**

Chromosome numbers are currently known for 76 species out of the 80 formally recognized for *Arachis* (Table 2). Most of them are diploids based on  $x=10$  (67), a few (4) are diploids based on  $x=9$ , and some (5) are tetraploids with  $x=10$  as base number. All diploids with  $x=18$  belong to section *Arachis* (*A. decora*, *A. palustris* and *A. praecox*), except *A. porphyrocalyx* which is included in section *Erectoides*.

All the species with  $x=9$  belonging to section *Arachis* are very similar [108] and have a common karyotype formulae ( $16m+2sm$ ) with the same SAT chromosome type 3 [57]. They share significant morphological similarity with A genome species [10], although they lack the small A chromosome pair [55, 56, 75, 76].

Molecular data revealed that these species clustered together, but the relationship with other species of the section is still not well resolved. AFLP [69, 106] and microsatellites [70] showed that these taxa are genetically closer to some non A species, although RFLP analysis has shown that they have more genetic similarity with the A genome species [10]. In a phylogram constructed on the basis of plastid *trnT-F* region, *A. palustris* clustered closely to non A genome species, while *A. praecox* remained among species with non A or D genome. Crosses between  $x=9$  and  $x=10$  entities have generally produced aborted seeds [106] or F1 with very low pollen stainability [Simpson pers. com.], except one hybrid between *A. palustris* x *A. duranensis* in which 68% pollen stainability was reported [102].

From the morphological, cytogenetical and molecular evidences it is difficult to address the genome identity of  $x=9$  species. However, the fact

**Table 2.** Chromosome numbers in species of genus *Arachis* and the respective references for first counts. n/d = no data.

Species	2n	Reference	Ref. Sl. No.
<i>A. archeri</i> Krapov. et W.C. Gregory	20	Mendes (1947)	[68]
<i>A. appressipila</i> Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]
<i>A. batizocoi</i> Krapov. et W.C. Gregory	20	Krapovickas et al. (1974)	[47]
<i>A. benensis</i> Krapov., W.C. Gregory et C.E. Simpson	20	Fernández and Krapovickas (1994)	[13]
<i>A. benthamii</i> Handro	20	Mendes (1947)	[68]
<i>A. brevipetiolata</i> Krapov. et W.C. Gregory	n/d		
<i>A. burchellii</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. burkartii</i> Handro	20	Gregory et al. (1973)	[18]
<i>A. cardenasi</i> Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]
<i>A. chiquitana</i> Krapov., W.C. Gregory et C.E. Simpson	20	Lavia (2000)	[59]
<i>A. correntina</i> (Burkart) Krapov. et W.C. Gregory	20	Krapovickas and Rigoni (1949)	[41]
<i>A. cruziana</i> Krapov., W.C. Gregory et C.E. Simpson	20	Lavia (2000)	[59]
<i>A. cryptopotamica</i> Krapov. et W.C. Gregory	20	Lavia (2000)	[59]
<i>A. dardani</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. decora</i> Krapov., W.C. Gregory et Valls	18	Peñaloza et al. (1996)	[75]
<i>A. diogoi</i> Hoehne	20	Smartt and Gregory (1967)	[97]
<i>A. douradiana</i> Krapov. et W.C. Gregory	20	Lavia (2000)	[59]
<i>A. duranensis</i> Krapov. et W.C. Gregory	20	Krapovickas and Rigoni (1957)	[44]
<i>A. giacometti</i> Krapov., W.C. Gregory, Valls et C.E. Simpson	20	Lavia (1996)	[55]

Table 2 contd.

Table 2 contd.

<i>A. glabrata</i> Benth. var. <i>hagenbeckii</i> (Harms ex Kuntze) F.J. Herm.	40	Fernández and Krapovickas (1994)	[13]
<i>A. glabrata</i> Benth. var. <i>glabrata</i>	40	Gregory et al. (1973)	[18]
<i>A. glandulifera</i> Stalker	20	Stalker (1991a)	[101]
<i>A. gracilis</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. gregoryi</i> C.E. Simpson, Krapov. et Valls	20	Peñaloza and Valls (2005)	[76]
<i>A. guaranitica</i> Chodat et Hassl.	20	Fernández and Krapovickas (1994)	[13]
<i>A. hassleri</i> Krapov., Valls et C.E. Simpson	20	Peñaloza and Valls (2005)	[76]
<i>A. hatschbachii</i> Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]
<i>A. helodes</i> Martius ex Krapov. et Rigoni	20	Smartt and Gregory (1967)	[97]
<i>A. hermannii</i> Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]
<i>A. herzogii</i> Krapov., W.C. Gregory et C.E. Simpson	20	Lavia (2000)	[59]
<i>A. hoehnei</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. hypogaea</i> L.	40	Ghimpu (1930), Kawakami (1930)	[17, 36]
<i>A. interrupta</i> Valls et C.E. Simpson	20	Peñaloza and Valls (2005)	[76]
<i>A. ipaensis</i> Krapov. et W.C. Gregory	20	Gregory and Gregory (1979)	[20]
<i>A. kempff-mercadoi</i> Krapov., W.C. Gregory et C.E. Simpson	20	Singh and al. (1990)	[94]
<i>A. krapovickasii</i> C.E. Simpson, D.E. Williams, Valls et I.G. Vargas	20	Peñaloza and Valls (2005)	[76]
<i>A. kretschmeri</i> Krapov. et W.C. Gregory	20	Kretschmer and Wilson (1988)	[50]
<i>A. kuhlmanni</i> Krapov. et W.C. Gregory	20	Singh et al. (1990)	[92]
<i>A. lignosa</i> (Chodat et Hassl.) Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]

Table 2 contd.

Table 2 contd.

<i>A. linearifolia</i> Valls, Krapov. et C.E. Simpson	20	Peñaloza and Valls (2005)	[76]
<i>A. lutescens</i> Krapov. et Rigoni	20	Conagin (1963)	[7]
<i>A. macedoi</i> Krapov. et W.C. Gregory	20	Conagin (1964)	[8]
<i>A. magna</i> Krapov., W.C. Gregory et C.E. Simpson	20	Singh et al. (1990)	[92]
<i>A. major</i> Krapov. et W.C. Gregory	20	Mendes (1947)	[68]
<i>A. marginata</i> Gardner	n/d		
<i>A. martii</i> Handro	n/d		
<i>A. matiensis</i> Krapov., W.C. Gregory et C.E. Simpson	20	Fernández and Krapovickas (1994)	[13]
<i>A. microsperma</i> Krapov., W.C. Gregory et Valls	20	Lavia (1996)	[55]
<i>A. monticola</i> Krapov. et Rigoni	40	Krapovickas and Rigoni (1957)	[44]
<i>A. nitida</i> Valls, Krapov. et C.E. Simpson	40	Peñaloza and Valls (2005)	[76]
<i>A. oteroii</i> Krapov. et W.C. Gregory	20	Mendes (1947)	[68]
<i>A. palustris</i> Krapov., W.C. Gregory et Valls	18	Lavia (1996)	[55]
<i>A. paraguariensis</i> Chodat et Hassl. ssp. <i>paraguariensis</i>	20	Smartt and Gregory (1967)	[97]
<i>A. paraguariensis</i> ssp. <i>capibarensis</i> Krapov. et W.C. Gregory	20	Singh et al. (1990)	[92]
<i>A. pfluegeae</i> C.E. Simpson, Krapov. et Valls	20	Peñaloza and Valls (2005)	[76]
<i>A. pietrarelli</i> Krapov. et W.C. Gregory	20	Smartt and Gregory (1967)	[97]
<i>A. pintoi</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. porphyrocalix</i> Valls et C.E. Simpson	18	Peñaloza and Valls (2005)	[76]
<i>A. praecox</i> Krapov., W.C. Gregory et Valls	18	Lavia (1998)	[56]
<i>A. prostrata</i> Benth.	20	Mendes (1947)	[68]
<i>A. pseudovillosa</i> (Chodat et Hassl.) Krapov. et W.C. Gregory	40	Fernández and Krapovickas (1994)	[13]

Table 2 contd.

Table 2 contd.

<i>A. pusilla</i> Benth.	20	Fernández and Krapovickas (1994)	[13]
<i>A. repens</i> Handro	20	Conagin (1962)	[6]
<i>A. retusa</i> Krapov., W.C. Gregory et Valls	20	Lavia (1996)	[55]
<i>A. rigonii</i> Krapov. et W.C. Gregory	20	Krapovickas and Gregory (1960)	[45]
<i>A. schininii</i> Krapov., Valls et C.E. Simpson	20	Peñaloza and Valls (2005)	[76]
<i>A. seridoensis</i> Valls, C.E. Simpson, Krapov. et R. Veiga	20	Peñaloza and Valls (2005)	[76]
<i>A. setinervosa</i> Krapov. & W.C. Gregory	n/d		
<i>A. simpsonii</i> Krapov. et W.C. Gregory	20	Lavia (2000)	[59]
<i>A. stenophylla</i> Krapov. et W.C. Gregory	20	Singh et al. (1990)	[92]
<i>A. stenosperma</i> Krapov. et W.C. Gregory	20	Gregory and Gregory (1979)	[20]
<i>A. subcoriacea</i> Krapov. et W.C. Gregory	20	Lavia (2000)	[59]
<i>A. submarginata</i> Valls, Krapov. et C.E. Simpson	20	Peñaloza and Valls (2005)	[76]
<i>A. sylvestris</i> (A.Chev.) A. Chev.	20	Fernández and Krapovickas (1994)	[13]
<i>A. trinitensis</i> Krapov. et W.C. Gregory	20	Lavia (1996)	[55]
<i>A. triseminata</i> Krapov. et W.C. Gregory	20	Gregory et al. (1973)	[18]
<i>A. tuberosa</i> Bong. ex Benth.	20	Lavia 2001	[60]
<i>A. valida</i> Krapov. et W.C. Gregory	20	Fernández and Krapovickas (1994)	[13]
<i>A. vallsii</i> Krapov. et W.C. Gregory	20	Lavia (1996)	[55]
<i>A. villosa</i> Benth.	20	Krapovickas and Rigoni (1951)	[42]
<i>A. villosulicarpa</i> Hoehne	20	Mendes (1947)	[68]
<i>A. williamsii</i> Krapov. et W.C. Gregory	20	Lavia (1996)	[55]

that the three species with  $x=9$  from the section Arachis constitute a compact group, suggested that they may be phylogenetically related and a single origin for these species has been advanced. This proposal suggested a unique event of chromosome number reduction, after which, dispersion and speciation gave rise to the presently known species [10, 57].

*Arachis porphyrocalyx* is morphologically distinct from the other  $x=9$  species, it has a more asymmetric karyotype with a SAT chromosome type 8, and it is genetically isolated from species of section Arachis. Considering all these data, the hypothesis of an independent origin of  $x=9$  in section Erectoides has been proposed [76, 107].

The question as to how has the  $x=9$  derived from the  $x=10$  basic chromosome number, still remains unanswered. The mechanisms responsible for the diminution of chromosome numbers can be aneuploidy or dispoloidy, besides Robertsonian centromere fusion [21]. Aneuploidy implies diminution of the number of chromosomes with loss of DNA content; while dispoloidy reduces it by means of chromosomal rearrangement without causing considerable variations in the DNA amount.

It has been suggested that the  $2n=2x=18$  species included in Arachis section would have originated from a diploid species with  $x=10$  ( $2n=2x=20$ ), belonging either to the A [57] or to the non A genome group [106]. If aneuploidy has been the mechanism involved, it is expected that tolerance to chromosome loss would be directly related to the amount and quality of information borne by that chromosome. From a chromosome morphology approach, the best candidate to be lost is the A chromosome, since it has around half the size of the other chromosomes of the complement [13] and a large block of heterochromatin, which in some cases is equivalent to 40% of the total length [88]. Therefore, it is expected that the genetic information borne by this pair of chromosomes is much lower than any other is chromosome of the complement. According to this rationale, a plausible hypothesis would consider that the  $x=9$  species were derived from an A genome ancestor which has lost the A chromosome pair by aneuploidy [55, 56]. This hypothesis is supported by the fact that all  $x=9$  species of section Arachis lack the A chromosome.

On the other hand, if dispoloidy has been the mechanism of chromosome reduction without loss of coding information, a large number of chromosome rearrangements would be needed. This

assumption is based on the evidence that the karyotypes of these species are very symmetric and, therefore, the chromatin of the chromosomes to be lost has to be translocated in small pieces to all or almost all chromosomes. However, this sort of karyotype modification has not been reported in any *Arachis* species or their hybrids yet.

Studies on DNA amount showed that the average DNA content (2C) of the diploid species with  $x=10$  belonging to the *Arachis* section is 5.13 pg, whereas the average of those with  $x=9$  (*A. palustris*, *A. praecox* and *A. decora*) is 3.67 pg [Lavia and Fernández unpubl.]. This difference of average DNA content (1.46 pg) is greater than that expected for the loss of a unique chromosome pair. Also, the mean chromosome length is smaller in  $x=9$  species than in those based in  $x=10$ . These data suggest that if  $x=9$  species included in *Arachis* section was derived by chromosome number reduction, either by dispoloidy or aneuploidy, from  $x=10$  species included in the same section, the event would have been also accompanied by an extensive DNA content reduction in the whole complement.

Conclusive evidences are not available about the origin of the  $x=9$  species, in both section *Arachis* and section *Erectoides*. Researches on FISH and GISH to determine the genome composition of these species and the relationships with the other *Arachis* species are expected to clarify this question.

## **ORIGIN OF POLYPLOIDS IN RHIZOMATOSAE SECTION**

Section Rhizomatosae is currently defined on morphological basis, mainly because all the entities have rhizomes, and can be asexually propagated. Among this group of taxa, *A. burkartii* is the only diploid ( $2n=2x=20$ ), while the remaining, *A. glabrata*, *A. pseudovillosa* and *A. nitida* are tetraploids ( $2n=4x=40$ ) [13, 18, 76].

*Arachis glabrata* is the most widely distributed *Arachis* species. It grows spontaneously from northeast Argentina to the Tocantins State in Brazil, and from San Pablo and Mina Gerais States in Brazil up to the Paraguay river. Some accessions of this species have a great value as forage, and of particular interest is the tolerance to soils with low fertility and a wide range of pH levels [14]. Several cultivars, like “Florigraze” and “Arbrook” in the USA [80, 81], and “Prine” in Australia [78, 79], were developed and commercialized. All of them are included among the list

of legumes to be selected for the establishment of artificial pastures in the subtropics.

The other two tetraploid species, *A. pseudovillosa* and *A. nitida*, have a more restricted area and are distributed in northeast Paraguay (Amambay and Concepción Departments), and in the South of Mato Grosso do Sul, in Brazil [48, 107], although their areas do not seem to be coincident.

*Arachis burkartii* grows in Rio Grande do Sul (Brazil), north of Uruguay and northeast of Argentina, and it is associated with the Uruguay river basin [48]. The area overlaps, to some extent, that of *A. glabrata* in the northeast of Argentina. Although populations of both species have been found relatively close, they were never found in sympatry or even in parapatry. This pattern of distribution is probably related with soil preferences, since *A. burkartii* grows in rocky locations, while *A. glabrata* thrives in deeper red soils [Seijo, pers. obs.].

From a phylogenetic point of view, the species of section Rhizomatosae constitute an interesting biological model, because although there is only one diploid species, three tetraploids are recognized. If section Rhizomatosae were monophyletic, *A. burkartii* or some other unknown or extinct diploid would perhaps to be the progenitors of the polyploids, them being either auto or allotetraploids. An alternative rationale is that section Rhizomatosae is not monophyletic, therefore, a multiple origin of rhizomes in *Arachis* has to be contemplated. In the latter case, the genome donors of the rhizomatous tetraploids have to be searched in other sections.

In order to test these hypotheses several investigations have been carried out, including cross compatibility experiments, molecular analysis and cytogenetic studies. Crosses between species of this section have revealed that none of the tetraploids have produced F1 hybrids with the diploid *A. burkartii*, and that all the tetraploids have none or very low cross compatibility. Despite the many attempts to produce hybrids between these polyploid taxa, only two hybrid plants between *A. pseudovillosa* and *A. glabrata* were reported, although no data on pollen viability was provided [48]. However, tetraploids have produced F1 hybrids with tetraploid and diploid species of other sections like Erectoides (four species), Procumbentes (two species) and Arachis (five species). Many of these hybrids were unable to produce flowers, some F1 produced normal flowers with different pollen viability such as *A.*

*hermanii* x *A. pseudovillosa* (46.10%), *A. major* x *A. nitida* (29.60%), *A. batizocoi* x *A. nitida* (0.03%), *A. appressipila* x *A. glabrata* (22.20%), *A. rigonii* x *A. glabrata* (14.90%) [48], *A. hypogaea* x *A. glabrata* (26.00%), *A. duranensis* x *A. glabrata* (31.00%) and *A. diogoi* x *A. glabrata* (30.00%) [65, 66].

Such crossabilities demonstrate a high reproductive isolation among the tetraploids of section Rhizomatosae, and between them and the only diploid so far known within it, *A. burkartii*. This fact may be interpreted as the existence of great genomic differentiation among tetraploids and suggest that they have a multiple origin. On the other hand, the candidacy of *A. burkartii* as the diploid genome donor of the tetraploids may be not supported. Moreover, the intersectional hybrids support the hypothesis that the genome donors of the rhizomatous polyploids have to be from other sections, probably Arachis or Erectoides.

The phylogenetic relationships within section Rhizomatosae and of these species in the context of the genus have been investigated by different molecular markers. AFLP analysis in species of seven sections revealed a close association of *A. glabrata* with *A. major* and *A. paraguariensis* (both of section Erectoides), while *A. burkartii* associated with *A. repens* and *A. pintoi* of section Caulorrhizae [16]. Moreover, based on microsatellites and RAPDs, it has been suggested that *A. burkartii* may be not the ancestor of *A. glabrata*, *A. pseudovillosa* and *A. nitida* since rhizomatous species with different ploidy levels were located in different clusters [2, 72]. These dendograms suggest that section Rhizomatosae may be not monophyletic and that the rhizomes may have arisen more than once during the evolution of the genus.

From a cytogenetic point of view, all the plants of the three polyploid species analyzed invariably showed  $2n=4x=40$  chromosomes [13, 55, 62, 73, 76]. Meiotic analyses in different materials of *A. glabrata* have revealed chromosome association which ranged from 20 II to 4II + 8IV [55, 73, 94]. These findings are in accordance with the meiotic behavior expected in an autotetraploid species, although populations may have different diploidization stages.

Localization of rDNA genes by FISH showed that *A. glabrata* has twice the number of 45S and 5S rDNA loci than *A. burkartii*, however these clusters mapped differently in each species [63]. Therefore, with the chromosome markers analyzed, the candidacy of *A. burkartii* as a genome donor of *A. glabrata* may be arguable.

The genomic composition of the rhizomatous species is currently being investigated by GISH using DNA of different diploid species of sections, Erectoides and Arachis as probes into the chromosomes of the rhizomatous tetraploids in the search for species with high genome affinities, which may be considered as their putative parents.

## **ORIGIN OF THE TRIPLOID CYTOTYPE OF *A. PINTOI* (SECTION CAULORRHIZAE)**

Section Caulorrhizae includes species adapted to clonal propagation by means of stolons, in addition to have good seed production. The two species included in this section, *A. pintoi* and *A. repens* are diploids with  $2n=2x=20$  [6, 13].

*Arachis pintoi* produces high quality forage [23], and several cultivars have been released in Colombia and Brazil. Among them, cv. "Maní Forrajero Perenne" [86] is the most widely cultivated and is referred to as the "tropical lucerne". Cultivars of this species have high adaptability to acid and low fertility soils, high nutritive value and persistence in association with grasses of genus *Brachiaria* [77].

A triploid cytotype of *A. pintoi* has arisen spontaneously among a diploid cultivar implanted in Costa Rica [75]. Considering that it was isolated from a diploid population, in the absence of any other species of *Arachis*, it has been largely speculated that the triploid would have arisen by autopolyploidy. This event provides a particular opportunity to study how polyploidy can arise in the genus and the subsequent genomic processes that occurred after a natural polyploidization event.

Cytological analysis has demonstrated that the diploid cytotype has a karyotype formula composed of  $9m+1sm$  with a SAT chromosome type 2 [13], while the triploid has 30 chromosomes arranged in  $27m+3sm$  with three SAT chromosomes of the same type [Lavia et al. unpubl.]. Physical mapping of rDNA genes using FISH have revealed an active cluster of 45S rDNA exclusively in each secondary constriction of the SAT chromosomes and four 5S in metacentric chromosomes of median size. The number, position and size of the rDNA clusters in each haploid complement of the triploid were identical to those observed in the diploid [Lavia et Seijo 2006]. Therefore, the bulk of evidence demonstrated that the  $3x$  cytotype originated by autopolyploidy.

The question is how can an impaired polyploidy arise in a supposedly diploid population. Two pathways are possible (1) hybridization between

2x and 4x individuals in a mixed population, or (2) sexual polyploidization by means of unreduced gametes. Preliminary analysis has shown that all the individuals of this species are diploids; therefore, the existence of mixed populations can be discarded. On the other hand, pollen analysis has demonstrated that unreduced grains (2x and 4x) are produced spontaneously in diploid plants [63], which may support the origin of the 3x cytotype by one way sexual autopolyploidization.

## CONCLUSION

This chapter provides an overview of the cytogenetic and molecular markers information related to peanut, its secondary/tertiary gene pool, and the species of the genus that are being used as forage. *A. hypogaea* is an allotetraploid, having two complements with heterochromatic bands (A genomes) and other two without those bands (non A genomes). The six botanical varieties have the same karyotype constitution and identical patterns of rDNA loci mapped by FISH. These results were interpreted as due to a common origin, however, introgression by gene recombination can not be discarded as a mechanism involved in the differentiation of the varieties. DNA polymorphism recently detected among peanut varieties has been used as an argument to support their polyphyletic origin. Therefore, despite many efforts, the simple versus multiple origins of subspecies and botanical varieties are still under intense debate.

After the proposal of *A. hypogaea* as being an allotetraploid was accepted, reports from different laboratories generally proposed various diploid species as probable genome donors of the cultigen. However, on the basis of new cytogenetic and molecular evidences and on the resynthesis of the allopolyploid, a general consensus considers *A. ipaensis* and *A. duranensis* as the diploid species that have participated in the origin of the cultigen. Moreover, *A. monticola* is being accepted as the wild tetraploid from which *A. hypogaea* may have been derived upon domestication.

Outstanding conclusions from section *Arachis* include the consistency of results obtained from the loci mapped by FISH and the DNA markers, which confirmed the three genomes previously proposed based on cross compatibility assays. However, they further revealed the existence of other genomic groups, still not precisely defined, within the non A species. The bulk of data indicate that almost half of the species of the section belong to the A genome group and 13 taxa are included

in the non A genome group. These findings have enlarged the secondary/tertiary gene pool of peanut, either for the A or non A complements of the cultigen, since all or almost all the entities are potential candidates for gene transfer to the cultigen directly or by the construction of complex diploid hybrids.

Even though much progress has been made in the species characterization of section *Arachis*, there are still many cases in which the data is puzzling, and a clear phylogeny is still awaited. The situation becomes worse if it is taken in account that almost every expedition to collect wild *Arachis* species results in the finding of new taxa. Therefore, intensification of field collection of wild species, additional data on cross compatibility, geographical distribution as well as climatic and edaphic adaptations, may be needed to bring light on these relationships.

The existence of a new basic chromosome number in the genus was unexpected, but it was so far reported for three species of section *Arachis*, and in one of section *Erectoides*. This fact may indicates that the reduction of chromosome number might has occurred more than once in the evolutionary history of the *Arachis* species. Particular attention has to be directed to the entities harboring the  $x=9$ , to their genomic constitution, and their phylogenetic relationships.

Especial interest has been directed towards the biosystematics of the sections, which includes the forage species, since they are the ones with the highest frequency of polyploids. Among the *Rhizomatosae* species, molecular and cytogenetic approaches, demonstrated that the unique diploid species of the section, *A. burkartii*, may have different genome constitution from those harbored by the polyploids. Also, the tetraploids may have arisen from different diploid species in multiple events of polyploidization and the rhizomes may be a convergent character in *Arachis*. The microsporogenesis analysis of *A. pintoi* (section *Caulorrhizae*) has demonstrated a mechanism that can lead to the origin of individuals with additional chromosome complements in the genus, which is sexual polyploidization by means of unreduced gametes.

Progress in the understanding *Arachis* biosystematics was so far done by means of the convergence of information coming from areas as diverse as archeology, morphology, classical genetics, cytogenetics, biogeography, and molecular genetics. The construction of dense gene maps and comparative genomics are potentially powerful tools for solving outstanding problems in crop domestication and phylogenetics. The

information coming from this area, integrated with the existent background knowledge of the genus, will aid in the construction of an integrated genetic system for *Arachis*, and should facilitate the identification and utilization of valuable genes from wild species.

## Acknowledgment

The authors are grateful to Ing. Agr. Antonio Krapovickas for his valuable comments.

## References

- [1] Amin PW, Singh KN, Dwivedi SL. Thrips (*Frankliniella schultzei* Trybom) and termite (*Odontotermes* sp.) in groundnut (*A. hypogaea*). Peanut Science 1985; 12: 58-60.
- [2] Angelici C, Hoshino A, Bravo J, Gimenes M, Lopes CR. Genetic diversity in species of Rhizomatosa section, genus *Arachis*, using microsatellite molecular markers. Proc Plant & Animal Genomes 12nd Conference, San Diego, USA, 2004.
- [3] Bianchi-Hall CM, Keys RD, Stalker HT, Murphy JP. Diversity of seed storage protein patterns in wild peanut (*Arachis*, Fabaceae) species. Plant Syst Evol 1993; 186: 1-15.
- [4] Bonavia D. Precerámico peruano, Los Gavilanes, oasis en la historia del hombre. Ed Corporación Financiera de Desarrollo S.A. Lima, Perú: COFIDE e Instituto Arqueológico Alemán. 1982.
- [5] Burow MD, Simpson CE, Starr JL, Paterson AH. Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): Broadening the gene pool of monophyletic polyploid species. Genetics 2001; 159: 823-837.
- [6] Conagin CHTM. Espécies selvagens do gênero *Arachis*. Observações sobre os exemplares da coleção da seção de Citologia. Bragantia 1962; 21: 341-374.
- [7] Conagin CHTM. Número de cromosomos das espécies selvagens de *Arachis*. Bragantia 1963; 22: 125-129.
- [8] Conagin CHTM. Número de cromosomos em *Arachis* selvagem. Bragantia 1964; 23: 25-27.
- [9] Cook BG, Crosthwaite IC. Utilization of *Arachis* species as forage. In: Smartt J, ed. The Groundnut Crop: A scientific basis for improvement. London, UK: Chapman & Hall, London, 1994: 624-663.
- [10] Creste S, Tsai SM, Valls JFM, Gimenes MA, Romero Lopes C. Genetic characterization of Brazilian annual *Arachis* species from sections *Arachis* and *Heteranthae* using RAPD markers. Genet Resour Crop Evol 2005; 52: 1079-1086.
- [11] Fávero AP, Simpson CE, Valls JFM, Vello NA. Study of the evolution of cultivated peanut through crossability. Studies among *Arachis ipaensis*, *A. duranensis*, and *A. hypogaea*. Crop Sci 2006; 46: 1546-1552.
- [12] Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor Appl Genet 2004; 108: 1064-1070.

- [13] Fernández A, Krapovickas A. Cromosomas y evolución en *Arachis* (Leguminosae). Bonplandia 1994; 8: 187-220.
- [14] French EC, Prine GM, Ocumpaugh WR, Rice RW. Regional Experience with Forage *Arachis* in the United States. In: Kerridge PC, Hardy B, eds. Biology and Agronomy of Forage Arachis. Cali, Colombia: Centro Internacional de Agricultura Tropical, 1994: 169-186.
- [15] García GM, Stalker HT, Kochert G. Introgression analysis of an interspecific hybrid population in peanuts (*Arachis hypogaea* L.) using RFLP and RAPD markers. Genome 1995; 38: 166-176.
- [16] Gimenes MA, Lopez CR, Valls JFM. Genetic relationships among *Arachis* species based on AFLP. Genetic and Molecular Biology 2002; 25: 349-353.
- [17] Ghimpur V. Recherches cytologiques sur les génés: *Hordeum*, *Acacia*, *Medicago*, *Vitis* et *Quercus*. Arch d'Anat Microsc 1930; 26: 136-234.
- [18] Gregory WC, Gregory MP, Krapovickas A, Smith BW, Yarbrough JA. Structures and genetic resources of Peanuts. In: Wilson CT, ed. Peanuts – Culture and Uses. Stillwater, USA: Am Peanut Res Educ Soc, 1973: 47-134.
- [19] Gregory WC, Gregory MP. Groundnut. *Arachis hypogaea* (Leguminosae-Papilionatae). In: Simmonds NW, ed. Evolution of crop plants. London, UK: Longman Group, 1976: 151-154.
- [20] Gregory WC, Gregory MP. Exotic germ plasm of *Arachis* L. interspecific hybrids. J Heredity 1979; 70: 185-193.
- [21] Greilhuber J, Ehrendorfer F. Karyological approaches to plant taxonomy. ISI Atlas of Science: Plants and Animals 1988; 1: 289-297.
- [22] Greishammer U, Wynne JC. Isozyme variability in mature seeds of United States peanut cultivars. Peanut Science 1990; 18: 72-75.
- [23] Grof B. Forage attributes of the perennial groundnut *Arachis pintoi* in a tropical savanna environment in Colombia. Proc 15th Int Gras Congr, Kyoto, Japan, 1985: 168-170.
- [24] Gross NR, Krapovickas A, Pietrarelli JR, Guzmán CA. Las proteínas seminales del maní (*Arachis hypogaea*, Leguminosae) y su relación con las categorías infraespecíficas. Bonplandia 1994; 8: 221-233.
- [25] Halward TM, Stalker HT, Larue EA, Kochert G. Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related species. Genome 1991; 34: 1013-1020.
- [26] Hammons RO. The origin and history of the groundnut. In: Smartt J, ed. The groundnut Crop: A scientific basis for improvement. London, UK: Chapman & Hall, London, United Kingdom, 1994: faltan págs.
- [27] He G, Meng R, Newman M, Gao G, Pittman RN, Prakash CS. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biology 2003, 3:3. (<http://www.biomedcentral.com/1471-2229/3/3>).
- [28] Herselman L. Genetic variation among southern African cultivated peanut (*A. hypogaea* L.) genotypes as revealed by AFLP analysis. Euphytica 2003; 133: 319-327.
- [29] Hilu KW, Stalker HT. Genetic relationships between peanut and wild species of *Arachis* sect. *Arachis* (Fabaceae): Evidence from RAPDs. Plant Syst Evol 1995; 198: 167-178.
- [30] Holbrook CC, Isleib TG. Geographical distribution and genetic diversity in *Arachis hypogaea*. Peanut Science 2001; 28: 80-84.

- [31] Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean R, Kochert GD, Kresovich S. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. *Crop Sci* 1999; 39: 1243–1247.
- [32] Husted L. Cytological studies on the peanut, *Arachis*. I. Chromosome number and morphology. *Cytologia* 1933; 5: 109-117.
- [33] Husted L. Cytological studies on the peanut, *Arachis*. II. Chromosome number, morphology and behaviour, and their application to the problem of the origin of the cultivated forms. *Cytologia* 1936; 7: 396-423.
- [34] Isleib TG, Wynne JC, Nigam SN. Groundnut breeding. In: Smartt J, ed. *The Groundnut Crop: A scientific basis for improvement*. London, UK: Chapman & Hall, London, 1994: 552-623.
- [35] Jarvis A, Ferguson ME, Williams DE, Guarino L, Jones PG, Stalker HT, Valls JFM, Pittman RN, Simpson CE, Bramel P. Biogeography of wild *Arachis*: assessing conservation status and setting future priorities. *Crop Sci* 2003; 43: 1100–1108.
- [36] Kawakami J. Chromosome numbers in Leguminosae. *Bot Mag (Tokyo)* 1930; 44: 319-328.
- [37] Kirti PB, Murty UR, Bharathi M, Rao NGP. Chromosome pairing in F1 hybrid *Arachis hypogaea* L. x *A. monticola* Krap. et Rig. *Theor Appl Genet* 1982; 62: 139-144.
- [38] Klosová E, Turkova V, Smartt J, Pitterová K, Svachulová J. Immunochemical characterization of seed proteins of some species of the genus *Arachis* L. *Biol Plant* 1983; 25: 201–208.
- [39] Kochert G, Halward T, Branch WD, Simpson CE. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theor Appl Genet* 1991; 81: 565-570.
- [40] Kochert G, Stalker HT, Gimenes M, Galgaro L, Romero Lopes C, Moore K. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Amer J Bot* 1996; 83: 1282-1291.
- [41] Krapovickas A, Rigoni VA. Cromosomas de una especie silvestre de *Arachis*. IDIA (Informativo de Investigaciones Agrícolas) 1949; 2: 23-24.
- [42] Krapovickas A, Rigoni VA. Estudios citológicos en el género *Arachis*. *Revista de Investigaciones Agrícolas* 1951; 5: 289-293.
- [43] Krapovickas A, Rigoni VA. Cruzamientos interespecíficos en *Arachis* y consideraciones sobre el origen del maní cultivado. Proc 28th Congreso Panamericano de Agronomía, Piracicaba, Sao Paulo, Brazil, 1954: 266-267.
- [44] Krapovickas A, Rigoni VA. Nuevas especies de *Arachis* vinculadas al origen del maní. *Darwiniana* 1957; 11: 431-455.
- [45] Krapovickas A, Gregory WC. "Arachis Rigonii". Nueva especie silvestre de maní. *Revista de Investigaciones Agrícolas* 1960; 14: 157-160.
- [46] Krapovickas A. Evolution of the genus *Arachis*. In: Moav R, ed. *Agricultural genetics. Selected topics*. Jerusalem, Israel: National Council for Research and Development, Jerusalem, Israel, 1973: 135–151.
- [47] Krapovickas A, Fernández A, Seeligman P. Recuperación de la fertilidad en un híbrido interespecífico estéril de *Arachis* (Leguminosae). *Bonplandia* 1974; 11: 129-142.
- [48] Krapovickas A, Gregory WC. Taxonomía del género *Arachis* (Leguminosae). *Bonplandia* 1994; 8: 1-186.
- [49] Krapovickas A, Lavia GI. Advances in taxonomy in *Arachis*. Proc 32nd Am Peanut Res Educ Soc, Alabama, USA, 2000: 46.

- [50] Kretschmer AE, Jr, Wilson TC. A new seed-producing *Arachis* sp. with potential as forage in Florida. Proc 47th Soil and Crop Sci Soc Fla, 1988; 229-233.
- [51] Kumar LSS, D'Cruz R, Oke JG. A synthetic allohexaploid in *Arachis*. Curr Sci 1957; 26: 121-122.
- [52] Lacks G, Stalker HT. Isozyme analyses of *Arachis* species and interspecific hybrids. Peanut Science 1993; 20: 76-81.
- [53] Lanham PG, Fennell S, Moss JP, Powell W. Detection of polymorphic loci in *Arachis* germplasm using random amplified polymorphic DNA. Genome 1992; 35: 885-889.
- [54] Lanham PG, Forster BP, McNicol P, Moss JP, Powell W. Seed storage protein variation in *Arachis* species. Genome 1994; 37: 487-496.
- [55] Lavia GI. Estudios cromosómicos en *Arachis* (Leguminosae). Bonplandia 1996; 9: 111-120.
- [56] Lavia GI. Karyotypes of *Arachis palustris* and *A. praecox* (Section Arachis), two species with basic chromosome number  $x=9$ . Cytología 1998; 63: 177-181.
- [57] Lavia GI. Caracterización cromosómica del germoplasma de maní. PhD thesis, Facultad de Ciencias Exactas y Químicas (Universidad Nacional de Córdoba, Argentina). 1999a.
- [58] Lavia GI. Evidencias cromosómicas para la reubicación taxonómica de *Arachis chiquitana* y *A. vallii*. Genetics and Molecular Biology 1999b; 22 (Suppl.): 118.
- [59] Lavia GI. Chromosomes studies of wild *Arachis* (Leguminosae). Caryologia 2000; 53: 177-181.
- [60] Lavia GI. Chromosomal characterization of germplasm of wild species of *Arachis* L. belonging to sections Trierectoides, Erectoides and Procumbentes. Caryologia 2001; 54: 115-119.
- [61] Lavia GI, Fernández A. Karyotypic studies in *Arachis hypogaea* L. varieties. Caryologia 2004; 57: 353-359.
- [62] Lavia GI, Ortiz AM, Cabaña Fader A, Seijo JG. Estudios citogenéticos en *Arachis glabrata* Benth. Proc 5th Simposio de Recursos Genéticos para América Latina y el Caribe, Montevideo, R. O. del Uruguay, 2005: 64.
- [63] Lavia GI, Seijo JG. Avances en la caracterización citogenética de las secciones Caulorrhizae y Rhizomatoseae. Proc 5th Encuentro Internacional de Especialistas en *Arachis*, Río Cuarto, Córdoba, Argentina, 2006: 77-79.
- [64] Lu J, Pickersgill B. Isozyme variation and species relationship in peanut and its wild relatives (*Arachis* L. -Leguminosae). Theor Appl Genet 1993; 85: 550-560.
- [65] Mallikarjuna N. Gene introgression from *Arachis glabrata* into *A. hypogaea*, *A. duranensis* and *A. diogoi*. Euphytica 2002a; 124: 99-105.
- [66] Mallikarjuna N, Sastri DC. Morphological, cytological and disease resistance studies of the intersectional hybrid between *Arachis hypogaea* L. and *A. glabrata* Benth. Euphytica 2002b; 126: 161-167.
- [67] Mehan VK. Screening groundnut for resistance to seed invasion by *Aspergillus flavus* and aflatoxin production. Proc of the International Workshop, ICRISAT, Patancheru, India, 1987: 323-345.
- [68] Mendes AJT. Estudos citológicos no gênero *Arachis*. Bragantia 1947; 7: 257-268.
- [69] Milla SR, Isleib TG, Stalker HT. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. Genome 2005; 48: 1-11.

- [70] Moretzshon MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology* 2004; 4: 11. (<http://www.biomedcentral.com/1471-2229/4/11>).
- [71] Moretzshon MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimenes MA, Martins WS, Valls JFM, Grattapaglia D, Bertioli DJ. A microsatellite-based, gene rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor Appl Genet* 2005; 111: 1060-1071.
- [72] Nóbile PM, Gimenes MA, Valls JFM, Lopes CR. Genetic variation within and among species of genus *Arachis*, section Rhizomatosa. *Genet Resour Crop Evol* 2004; 51: 299-307.
- [73] Ortiz AM. Caracterización citogenética del germoplasma de *Arachis glabrata* y *A. burkartii* de la sección Rhizomatosa. Tesina de Grado. Facultad de Ciencias Exactas y Naturales y Agrimensura (Universidad Nacional del Nordeste). 2006.
- [74] Paik-Ro OG, Smith RL, Knauf DA. Restriction fragment length polymorphism evaluation of six peanut species within *Arachis* section. *Theor Appl Genet* 1992; 84: 201-208.
- [75] Peñaloza APS, Pozzobon MT, Valls JFM. Cytogenetic findings in wild species of *Arachis* (Leguminosae). Proc 42nd Congresso Nacional de Genética, Caxambu, Brasil, 1996: 42.
- [76] Peñaloza APS, Valls JFM. Chromosome number and satellite chromosome morphology of eleven species of *Arachis* (Leguminosae). *Bonplandia* 2005; 15: 65-72.
- [77] Pizarro EA, Rincón A. Regional experience with forage *Arachis* in South America. In: Kerridge PC, Hardy B, eds. *Biology and Agronomy of Forage Arachis*. Cali, Colombia, Centro Internacional de Agricultura Tropical, 1994: 144-157.
- [78] Prine GM. Forage possibilities in the genus *Arachis*. Proc Soil Crop Sci Soc Fla 1964; 24: 187-196.
- [79] Prine GM. Perennial peanuts for forage. Proc Soil Crop Sci Soc Fla 1972; 32: 33-35.
- [80] Prine GM, Dunavin LS, Moore JE, Roush RD. "Florigraze" rhizoma peanut: a perennial forage legume. Agricultural Experiment Station, Circular S-275, University of Florida, Gainsville. 1981.
- [81] Prine GM, Dunavin LS, Gennon RJ, Roush RD. "Arbrook" rhizoma peanut: a perennial forage legume. Agricultural Experiment Station, Circular S-332, University of Florida, Gainsville. 1986.
- [82] Raina SN, Mukai Y. Detection of a variable number of 18S-5.8S-26S and 5S ribosomal DNA loci by fluorescent in situ hybridization in diploid and tetraploid *Arachis* species. *Genome* 1999a; 42: 52-59.
- [83] Raina SN, Mukai Y. Genomic in situ hybridization in *Arachis* (Fabaceae) identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. *Plant Syst Evol* 1999b; 214: 251-262.
- [84] Raina SN, Rani V, Kojima T, Ogihara Y, Singh KP, Devarumath RM. RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification and phylogenetics relationships in peanut (*Arachis hypogaea*) cultivars and wild species. *Genome* 2001; 34: 1013-1020.
- [85] Raman VS. Studies in the genus *Arachis*. IX. A fertile synthetic tetraploid groundnut from the interspecific backcross *A. hypogaea* *A. villosa*. *Indian Oilseeds Journal* 1960; 4: 90-92.

- [86] Rincón CA, Cuesta PAM, Pérez RB, Lascano CE, Ferguson J. Maní Forrajero Perenne (*Arachis Pintoi* Krapovickas y Gregory): Una alternativa para ganaderos y agricultores. Boletín Técnico Instituto Colombiano Agropecuario. 1992; 219: 23.
- [87] Robledo G, Seijo JG. Mapeo físico de los genes ribosomales y regiones heterocromáticas en *Arachis glandulifera* (Leguminosae). Proc 5th Encuentro Internacional de Especialistas en *Arachis*, Río Cuarto, Córdoba, Argentina, 2006: 99.
- [88] Seijo JG, Lavia GI, Fernández A, Krapovickas A, Ducasse D, Moscone EA. Physical mapping of 5S and 18S-25S rRNA genes evidences that *Arachis duranensis* and *A. ipaensis* are the wild diploid species involved in the origin of *A. hypogaea* (Leguminosae). Amer J Bot 2004; 91: 2293-2303.
- [89] Simpson CE, Faries MJ. Advances in the characterization of diversity in section Arachis: archeological evidence, crossing results and their relationship in understanding the origin of *Arachis hypogaea* L. Proc 3rd Simposio de recursos genéticos para a America Latina e Caribe, 2001: 103-104.
- [90] Singh AK, Moss JP. Utilization of wild relatives in genetic improvement of *Arachis hypogaea* L. 2. Chromosome complements of species in the section Arachis. Theor Appl Genet 1984; 61: 305-314.
- [91] Singh AK. Putative genome donors of *Arachis hypogaea* (Fabaceae), evidence from crosses with synthetic amphidiploids. Plant Syst Evol 1988; 160: 143-151.
- [92] Singh AK, Venkateshwar A, Moss JP, Rau TPS. Chromosome number and karyomorphology of some new accessions of genus *Arachis*. IAN 1990; 8: 11-14.
- [93] Singh AK, Sivaramakrishnan S, Mengesha MH, Ramaiah CD. Phylogenetic relations in the section Arachis based on seed protein profile. Theor Appl Genet 1991; 82: 593-597.
- [94] Singh AK, Simpson CE. Biosystematics and genetic resources. In: Smartt J, ed. The Groundnut Crop: A scientific basis for improvement. London, UK: Chapman & Hall, 1994: 96-137.
- [95] Singh AK, Smartt JP. The genome donors of the groundnut/peanut (*Arachis hypogaea* L.) revisited. Genet Resour Crop Evol 1998a; 45: 113-118.
- [96] Singh AK, Smartt J, Simpson CE, Raina SN. Genetic variation vis-à-vis molecular polymorphism in groundnut, *Arachis hypogaea* L. Genet Resour Crop Evol 1998b; 45: 119-126.
- [97] Smartt J, Gregory WC. Interspecific cross-compatibility between the cultivated peanut *Arachis hypogaea* L. and other species of the genus *Arachis*. Oleagineux 1967; 22: 455-459.
- [98] Smartt J, Gregory WC, Gregory MP. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. Euphytica 1978; 27: 665-675.
- [99] Stalker HT, Wynne JC. Cytology of interspecific hybrids in section Arachis of peanuts. Peanut Science 1979; 6: 110-114.
- [100] Stalker HT, Moss JP. Speciation, cytogenetics, and utilization of *Arachis* species. Adv Agronomy 1987; 41: 1-40.
- [101] Stalker HT. A new species in section Arachis of peanuts with a D genome. Amer J Bot 1991a; 78: 630-637.
- [102] Stalker HT, Dhesi JS, Parry DC, Hahn JH. Cytological and interfertility relationships of *Arachis* section Arachis. Amer J Bot 1991b; 78: 238-246.
- [103] Stalker HT, Phillips TD, Murphy JP, Jones TM. Variation of isozyme patterns among *Arachis* species. Theor Appl Genet 1994; 87: 746-755.

- [104] Stebbins CD. Genetics, evolution, and plant breeding. Indian Journal of Genetics and Plant Breeding 1957; 17: 129-141.
- [105] Tallury SP, Milla SR, Copeland SC, Stalker HT. Genome donors of *Arachis hypogaea* L. Proc 33rd Am Peanut Res Educ Soc 2001: 60.
- [106] Tallury SP, Hilu KW, Milla SR, Friend SA, Alsaghir M, Stalker HT, Guandt D. Genomic affinities in *Arachis* section *Arachis* (Fabaceae): molecular and cytogenetic evidence. Theor Appl Genet 2005; 111: 1229-1237.
- [107] Valls JFM, Simpson CE. New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 2005; 14: 35-64.
- [108] Veiga RFA, Queiroz-Voltan RB, Valls JFM, Fávero AP, Barbosa W. Caracterização morfológica de acessos de germoplasma de quatro espécies brasileiras de amendoim silvestre. Bragantia 2001; 60: 167-176.
- [109] Young ND, Weeden NF, Kochert G. Genome mapping in legumes (Family Fabaceae). In: Paterson AH, ed. Genome Mapping in Plants. Austin, Texas, 1996: 212-227.

## Biodiversity, Genetic Enhancement and Molecular Approaches in Lentil

MICHAEL BAUM, ALADDIN HAMWIEH, BONNIE FURMAN,  
ASHTUSOSH SARKER and WILLIAM ERSKINE

International Center for Agricultural Research in the Dry Areas  
(ICARDA), P.O. Box, Aleppo, Syria

### ABSTRACT

Lentil is a nutritious food legume and is cultivated for its seed which has relatively higher contents of protein content. Currently, lentil is grown in over 48 countries worldwide on 4 million hectares and produced 4 million MT in 2005 growing season. ICARDA has created a composite collection of 1,000 lines to explore the genetic diversity of the global germplasm collections held by ICARDA. The methodology for establishing the composite collection combined classical hierarchical cluster analyses using agronomic traits and two-step cluster analyses using agro-climatological data linked to the geographical coordinates of the accessions' collection sites. The composite collection has been characterized using simple sequence repeat (SSR) markers. The SSR markers were developed and extracted from a genomic library of the line ILL5588 that was not enriched for SSRs. The genomic DNA was cloned and hybridized with synthetic oligonucleotides. The CA/GT di-nucleotide repeat was the most abundant among the six repeats used for screening. This repeat is different from the most abundant repeat motif in other legumes such as chickpea or pea. Primers were designed in flanking regions of the SSR repeats to develop markers that could be used for genetic mapping. SSRs were mapped into the genetic linkage map of lentil.

**Key Words:** Genetic diversity, composite collection, breeding, molecular mapping, microsatellite markers

## INTRODUCTION

Lentil is one of the oldest domesticated grain legumes in the world. It was first used for human nutrition in the Fertile Crescent of the Near East and West Asia before it spread to other regions of the world [12, 34, 81]. Lentil was an important crop in ancient times; its importance, however, decreased during the Roman Empire era in the East, when more cropland was devoted to crops such as grapes and wheat [74]. Lentil was disseminated throughout the Mediterranean region, Asia, and Europe during the Bronze Age, and was introduced to the United States in 1916 [80].

## BIODIVERSITY AND EVOLUTION IN LENTIL

Lentil (*Lens culinaris* Medicus subsp. *culinaris*) is a nutritious food legume and is cultivated for its seed which has relatively higher content of protein. Additionally, the plants residues can be fed to livestock. The traditional areas for lentil cultivation are the Middle East, Ethiopia, Indian subcontinent and Northern Africa. Currently, lentil, grown in over 48 countries worldwide on 4 million hectares, produced 4 million MT in 2005 growing season [27]. The major lentil producing regions are Asia (58% of the area) and the West Asia-North Africa region (37% of the acreage of developing countries).

All members of *Lens* are self-pollinating diploids ( $2n = 2x = 14$ ; [65]). The haploid genome size of the cultivated genome is 4,063 Mbp [2]. The genus *Lens* comprises seven taxa within four species including the cultivated type, *Lens culinaris* spp. *culinaris* [28]). Cultivated lentil includes two varietal types: small-seeded microsperma and large-seeded macrosperma [5]. *L. culinaris* subsp. *culinaris* is grouped with its three putative wild progenitors, *L. culinaris* Medikus subsp. *orientalis* (Boiss.) Ponert, *L. culinaris* Medikus subsp. *odemensis* (Ladizinsky) Ferguson, Maxted, van Slageren and Robertson, and *L. culinaris* subsp. *tomentosus* (Ladizinsky) Ferguson, Maxted, van Slageren and Robertson. The remaining wild species include *L. nigricans* (M. Bieb.) Godron, *L. ervoides* (Brign.) Grande, and *L. lamottei* Czeft.

*L. culinaris* subsp. *orientalis* is fully cross compatible with cultivated lentil [45, 51]. *L. nigricans* can also cross with the cultigen, yet seed set

in hybrids is considerably lower than a *L. culinaria* subsp. *culinaria* x *L. culinaria* subsp. *orientalis* cross [45]. *L. culinaria* subsp. *orientalis* accessions have been found to contain resistance to drought [32], cold [31, 50], wilt [7], and Ascochyta blight [6, 50].

The International Center for Agricultural Research in the Dry Areas (ICARDA) has a global mandate for research on lentil improvement. As such, ICARDA houses the world collection of *Lens*, which includes 10,800 accessions. This includes 8,860 accessions of cultivated lentil from 70 different countries, 1,373 ICARDA breeding lines, and 581 accessions of 6 wild *Lens* taxa representing 24 countries. From this collection, a composite germplasm set of about 1,000 accessions was identified (Fig. 1) as an activity under the Generation Challenge Program.

ICARDA was responsible for creating a 'reference' core collection for lentil as part of a large-scale CGIAR Generation Challenge Program that aims to explore the genetic diversity of the global germplasm collections held by the CGIAR research centers (<http://www.generationcp.org>). A global composite collection of 1,000 lentil accessions was established representing the overall genetic diversity and the agro-climatological range of lentil. Included accessions were compiled from landraces, wild relatives, and elite germplasm and cultivars representing the overall ICARDA collection in both distribution and type.

The methodology for establishing the composite collection combined classical hierarchical cluster analyses using agronomic traits and two-step cluster analyses using agro-climatological data linked to the geographical coordinates of the accessions' collection sites [30]. The hierarchical cluster analysis ensured that the level of variation found in the larger collection was maintained in the composite collection. In addition, 64 accessions of landraces, released cultivars, and breeding materials for their resistances to a number of stresses affecting lentil production were included in the composite collection.

The composite collection has been characterized with 24 SSR markers, and diversity parameters such as polymorphic information contents, allelic diversity, etc., have been determined. This could enable extracting core and referencing collection for use by breeders and geneticists.

## VARIABILITY IN KEY TRAITS

Work with world lentil germplasm that are conserved at ICARDA, reveals that substantial variability exists among landraces and wild species

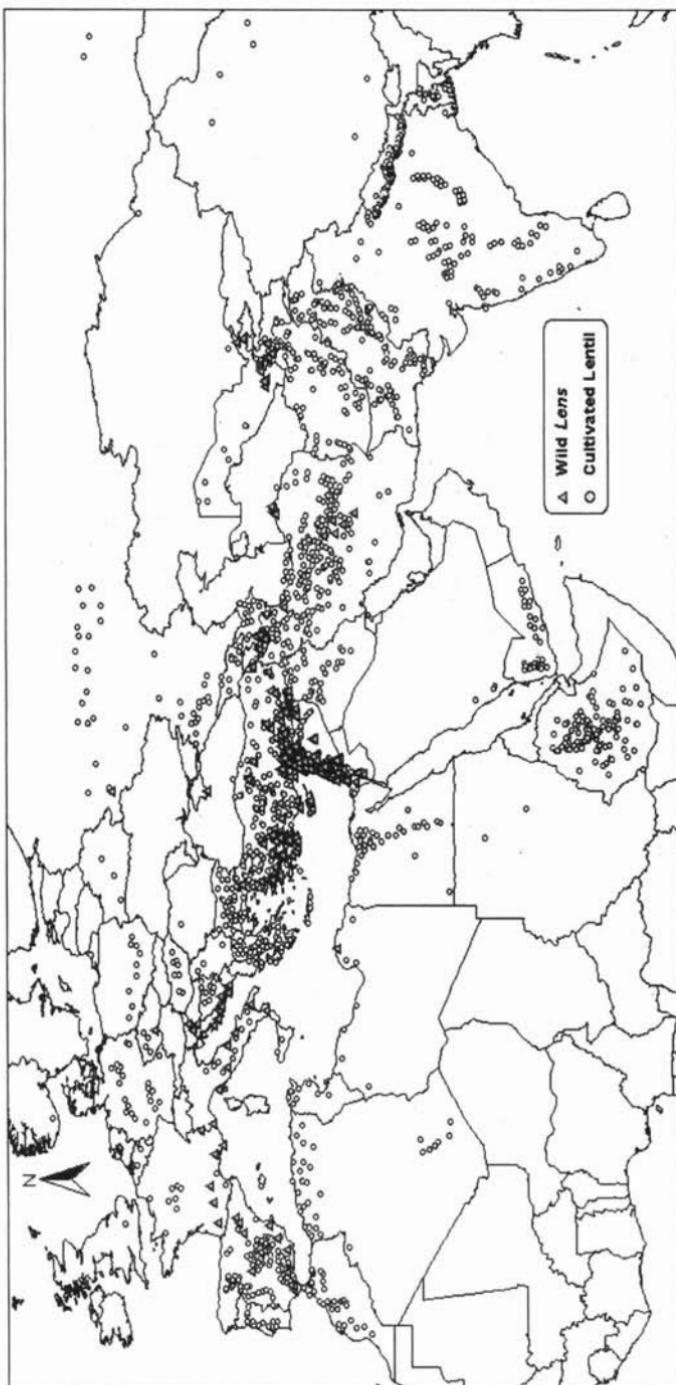


Fig. 1. Distribution of ICARDA lentil collection.

for economically important traits. Studies on genetic variability have been conducted since early 1980s among germplasm of diverse origin. Morphological traits were critically examined in variable climatic conditions for use in breeding and selection programs [3, 15, 16, 18, 19, 39, 49, 53, 62, 64, 67], responses in flowering to temperature and photoperiod [20, 22] and abiotic stresses [14, 20, 32, 43, 61, 62, 68, 79]. Genetic variability and sources of resistance to fungal diseases [1, 6, 7, 8, 37, 43, 47] and viruses [42] were reported.

Desirable variability in many of these traits exists within the crop gene pool allowing genetic enhancement through plant breeding. On the other hand, several other important traits, such as nitrogen fixation, resistance to pea leaf weevil (*Sitona* sp.) and aphids, and the broomrape (*Orobanche* sp.) are not currently addressable by breeding because of lack of sufficient genetic variation [63].

Evaluation of wild species for resistance to key stresses included winter hardiness, drought tolerance, and resistance to lentil vascular wilt and *Ascochyta* blight. Sufficient variability was noticed for these traits [6, 32, 47]. Additionally, variation in agronomic characters was also explored. For seed and straw yields, there was no striking variation within the wild species for transfer to the cultivated lentil. Clearly, direct selection of wild lentil germplasm for biomass yield under dry conditions is of little value.

## **MAJOR ACCOMPLISHMENTS IN GENETIC ENHANCEMENT**

### **a. Resistance to Biotic Factors**

Biotic stresses are the major production impediments of lentil globally. Among such stresses, diseases caused by various fungal pathogens are important and commendable progress has been made on genetic enhancement for disease resistance. Resistance to *Fusarium* wilt caused by *Fusarium oxysporum* fp. *lentis* has been identified among landraces and resistance sources have been utilized to develop resistant cultivars [61]. Lentil rust caused by *Uromyces fabae* is also a devastating disease, limiting crop productivity to a significant level globally. Using parents of Argentine and West Asian origins, rust resistant genotypes have been developed and released as commercial cultivars in Bangladesh, Ethiopia, Pakistan and Morocco [4, 9, 56, 69]. Genes for *Stemphylium* blight resistance present in ILL 4605 and ILL 6002 are transferred to landraces

from Bangladesh and India. Additionally, Ascochyta blight resistance genes from ILL 5588 became the key source of all resistant cultivars globally, including Australia and Canada. Several breeding lines developed at ICARDA possessing Botrytis Grey Mold (BGM) resistance have been reported in Australia [11]. Resistance sources to multiple viruses are being used in the breeding program at ICARDA [42].

### **b. Improving Tolerance to Abiotic Stresses**

Among abiotic stresses, drought and winter hardiness are widely known to reduce lentil yield. Genetic enhancement for drought tolerance is a key researchable area in ICARDA's crop improvement strategy [58]. Traits like early seedling establishment, early growth vigor and ground coverage, high biomass, early flowering and maturity, were taken into consideration while selecting drought tolerant genotypes. Selection in drought-prone sites is the key to success in identification of drought tolerant genotypes. Seedling shoot and root traits, such as taproot length and lateral root number, are important for drought tolerance and enormous variability has been observed in these characters [62]. Parents with these traits were used in hybridization and the resulting drought tolerant lines are under test with NARS [43]. Accessions which originate from USA, Turkey and Iran possess genes for winter-hardiness and have been used to develop winter-hardy lines that thrive well under -20°C in cold-prone areas of Turkey and Iran [61]. Although resistance sources to soil salinity, iron deficiency, boron toxicity deficiency, and waterlogging have been identified, these are yet to be utilized in lentil improvement.

### **c. Construction of Plant Type Suitable for Machine Harvest**

Traditional cultivars from West Asia and North Africa are prone to weak stems, and are unsuitable for mechanical harvest. ICARDA has developed and promoted regional national programs on lentil production package, which includes harvest mechanization and the use of improved cultivars with erect plant type having good standing ability. Such cultivars, which include *Idlib-2*, *Idlib-3* and *Idlib-4* in Syria, *Hala* and *Rachayya* in Lebanon, *IPA-98* in Iraq, *Saliana* and *Kef* in Tunisia, and *Firat-87* and *Sayran-96* in Turkey, have been developed through aggregation of genes conferring erect growth from multiple parents. On an average, mechanical harvesting combined with improved cultivars

having good standing ability reduces cost of cultivation by USD 100 per hectare [56].

#### **d. Genetic Base of Lentil in South Asia has been Widened**

Lentil germplasm from South Asia are *pilosa* type; short statured with narrow genetic base characteristic of morphological, phenological, agronomic, and biotic and abiotic stress resistance traits. Narrow variability in desirable traits has restricted the breeders from achieving any breakthrough in yield improvement for long. However, the South Asian national programs used exotic germplasm supplied by ICARDA in their breeding programs. Early, bold-seeded, disease resistant accessions from Latin America and West Asia were the prime sources of useful variability, which were efficiently used in widening the genetic base of lentils in the region [22]. This led to the development of improved cultivars and a manifold increase in yield in Bangladesh (*Barimasur-2* and *Barimasur-4*) [56, 59], Nepal (*Shekher* and *Sital*) and Pakistan (*Manshera-98*, *Shiraz-96*, *Masoor-93* and *Masoor-2002*). Bold-seeded and early maturing lines are now available in India and are being used in various programs. Early and extra-early materials are very important to fit in various cropping system niches, especially in rice fallow systems.

#### **e. Varietal Development**

Desirable genes for yield contributing traits have been assembled through crossbreeding using agronomically desirable parents. Various stress resistance genes also contributed to stabilize yield of new cultivars. For example, varieties with combined resistance to rust and *Stemphylium* blight diseases have been released in Bangladesh, rust and wilt resistant cultivars are now available with Ethiopian farmers, Morocco released rust resistant varieties and Syria, Iraq, Lebanon and Turkey released fusarium wilt resistant varieties. Drought tolerant and winter-hardy lentil varieties were released in Syria and Turkey respectively. In this endeavor, more than 100 lentil varieties have been released by 29 national programs, which emanated from ICARDA-supplied genetic materials.

#### **f. Development of Iron and Zinc-rich Lentil Varieties**

To address micronutrient malnutrition, most particularly in Southeast Asia and Africa, >1,000 lentil accessions were analyzed to study

variability for iron and zinc contents. Preliminary studies show that enormous variability exists for iron (41-109 mg/kg) and zinc (22-78 mg/kg) contents. Higher contents of these nutrients are mostly present in the breeding lines, where 3-4 parents were involved in their parentage [36].

High iron content varieties Idlib-2 (ILL 5883) (73 mg/kg) and Idlib-3 (ILL 6994) (72 mg/kg) were released in Syria, and Beleza (ILL 7711) (74 mg/kg) was released in Portugal.

High zinc content variety Beleza (56 mg/kg) was released in Portugal Ada'a (52 mg/kg) and high iron (82 mg/kg) and Zn (66 mg/ka) content variety Alemaya in Ethiopia and Meyveci-2001 (53 mg/kg) in Turkey. High micronutrient content Nepalese varieties are: Sisir (Fe-98 mg/kg; Zn-64 mg/kg); Khajurah-2 (Fe-94 mg/kg; Zn-54 mg/kg) and Barimasur-5 (Fe-86 mg/kg; Zn- 59 mg/kg) and Barimasur-6 (Fe-86 mg/kg; Zn-63 mg/kg) [35].

## **GENOME EVOLUTION IN LENTILS**

The evolution of the legume genomes can ideally be studied with the help of DNA markers [10, 82]. SSR markers are ideal for genotype identification [41, 48], however, they can also be used to study genome evolution [10, 71]. These microsatellites are composed of tandemly repeated, 2-5 nucleotide long, repeat units, which are reiterated from a few up to several thousand times at a particular locus. Microsatellites are among the fastest evolving DNA sequences and are abundant in all eukaryotic genomes [76]. SSRs frequently change their length by means of addition or deletion of repeat units; this is most probably due to the high propensity for slippage that repeat arrays demonstrate during DNA replication [40, 71]. The conserved sequences flanking these repeats can be used to design primers for use in amplifying the intervening microsatellite in individuals of the same species, and in some cases, across different genera of the same family.

In lentil, microsatellite markers using a genomic library have been developed [34]. Genomic DNA (ILL 5588) has been digested with *Sau*3AI to generate small fragments of genomic DNA for library construction. Since the objective was to isolate microsatellites that are abundant and well distributed in the genome, no enrichment for microsatellites was followed. The genomic library containing approximately 200,000 clones, was screened by two rounds of

hybridization using oligonucleotides complementary to microsatellite sequences  $(GT)_{10}$ ,  $(GA)_{10}$ ,  $(GC)_{10}$ ,  $(TA)_{10}$ ,  $(TAA)_5$  and  $(GAA)_8$ . This screen identified 243 clones (65.5%) that were used for sequencing. After sequencing, 173 (71.2%) sequences contained microsatellites. After examining the sequences of microsatellite-containing clones, the observed lentil microsatellites were classified as perfect, imperfect, compound or compound/imperfect, according to Weber [75], with slight modifications. Most of the repeats in lentil were of the simple/perfect type, in agreement with many studies of plant species [38, 78]. Among the simple/perfect type of microsatellites, dinucleotide repeats were more abundant in lentil than other types. Even though the lentil library was screened with six different probes, the  $(GT)_{10}$ ,  $(GA)_{10}$ ,  $(GC)_{10}$ ,  $(GAA)_8$ ,  $(TA)_{10}$ , and  $(TAA)_5$ , the CA/GT repeats were predominant, indicating that CA/GT repeats might be a major type of microsatellite in the lentil genome, but this repeat is comparatively less frequent in many other plants [44]. Tri-nucleotide repeats were recovered in only 11 (4.7%) of the microsatellite-containing clones, including 10 simple perfect types (GAA and TAA) and one simple imperfect type (TAA). In total, the simple/perfect repeats were predominant (56.8%), followed by compound/perfect (16.1%). The compound/imperfect (12.7%) occurred least often (Table 1). Among the perfect repeats,  $(CA/GT)_n$  motifs were the most abundant, comprising 24.2% of the isolated clones, followed by  $(AT/TA)_n$  repeats (8.9%). A similar work in chickpea [38, 77, 78] identified the tri-nucleotide repeat TAA as the most abundant microsatellite. Predominance of different SSR motifs indicates how fast this class of repetitive sequence changes in evolutionary times.

Of the 173 sequences of our lentil genomic library containing microsatellites, 31 were not useful for primer design, either because of AT-rich regions at the flanking sites or because the microsatellites were too close to the cloning sites. Therefore, only 142 clones could be used for primer design. A total of 156 primer pairs were designed. All the primer combinations were used to amplify the genomic DNA of different lentil accessions, including cultivated lentil ILL 5588. A total of 124 primer combinations yielded PCR (polymerase chain reaction) products and thus were considered functional. A comparison between the observed and expected molecular weight of PCR products revealed that 102 primer combinations amplified products having expected sizes, whereas the remaining primer combinations amplified products having unexpected sizes. In summary, from the construction and screening of the

**Table 1.** Frequency of various types of microsatellite motifs observed in the lentil genomic library

	Type	Microsatellite motif	Number	% Occurrence
Simple	Perfect	CA/GT	57	24.2
		CG/GC	2	0.8
		CT/GA	7	3.0
		CTT/GAA	3	1.3
		AT/TA	21	8.9
		ATT/TAA	7	3.0
		Others	37	15.7
	Total		134	56.9
	Imperfect	CA/GT	21	8.9
		CG/GC	0	0.0
		CT/GA	1	0.4
		CTT/GAA	0	0.0
		AT/TA	3	1.3
		ATT/TAA	1	0.4
		Others	8	3.4
	Total		34	14.4
Compound	Perfect		38	16.1
	Imperfect		30	12.7
	Total		68	28.8
	Total		236	

genomic library to the development of a working set of primers that could amplify microsatellites in lentil, only 124 (51%) of the original 243 sequenced clones yielded functional primer sets.

## MOLECULAR APPROACHES USED FOR LENTIL

Various lentil maps have been developed that differ with respect to the nature of crosses used (interspecific or intraspecific), molecular marker types and their coverage of the genome. Early genetic maps consisted of a small number of markers, mainly morphological and isozyme markers covering a small portion of the genome [46, 72] and often wide crosses were used for genetic mapping. Thereafter, many types of DNA-based markers were developed and used for genome mapping in lentil. RFLP

(restriction fragment length polymorphism) markers were the first DNA markers used to construct a genetic linkage map [36]. Subsequently, PCR-based markers, such as RAPDs (random amplified polymorphic DNA), were used to study diversity, phylogeny and taxonomy of *Lens* [28, 29, 65] and to develop linkage maps [25]. AFLP (amplified fragment length polymorphism) markers, after their discovery in 1995 [74], were used in lentil linkage mapping [13, 24] and to study genetic diversity [66].

At ICARDA an  $F_2$  population of a wide cross, L92-013 (ILL5588  $\times$  L692-16-1(s)), was used to develop RILs by advancing individual  $F_2$  plants [24] to  $F_6$  using SSD [24]. Using a LOD score of 4 and maximum distance of 25 cM, a total of 177 markers (89 RAPD, 79 AFLP, 6 RFLP, and 3 morphological markers) were mapped in seven major linkage groups covering 1,073 cM, with an average distance of 6 cM between adjacent markers. The morphological markers, pod indehiscence (*Pi*), Seed coat pattern (*Scp*), and flower color (*W*), were also mapped. Of the total linked loci, 8.4% deviated from the expected segregation, and more than 25% of the distorted loci clustered in one linkage group [24].

Another genetic linkage map of the lentil genome was constructed with 114 molecular markers, 100 RAPD, 11 inter simple sequence repeat (ISSR), and 3 resistance gene analogues (RGA), using an  $F_2$  population developed from a cross between lentil cultivars ILL5588 and ILL7537 that differ in their resistance to Ascochyta blight. The map consisted of 9 linkage groups comprising between 6 and 18 markers, and spanning 784.1 cM [52].

Another *Lens* sp. map contains 62 RAPDs, 29 ISSRs, 65 AFLPs, 4 morphological, and 1 microsatellite marker developed in 113  $F_2$  plants obtained from a single hybrid of *Lens culinaris* ssp. *culinaris* and *L. culinaris* ssp. *orientalis*. The map was grouped into ten linkage groups covering 2,172.4 cM with an average inter-marker distance of 15.87 cM [13].

The most recent map of lentil contained 283 markers distributed over 14 linkage groups (Table 2), including 39 microsatellite markers and 5 morphological markers of seed color pattern (*Scp*), flower color (*W*), pod indehiscence (*Pi*), Fusarium wilt resistance (*Fw*) and radiation frost tolerance locus (*Rf*) (Fig. 2). The map was constructed using 86 recombinant inbred lines derived from the cross ILL 5588  $\times$  L692-16-1(s), which had been previously used for the lentil linkage map [24, 25, 26].

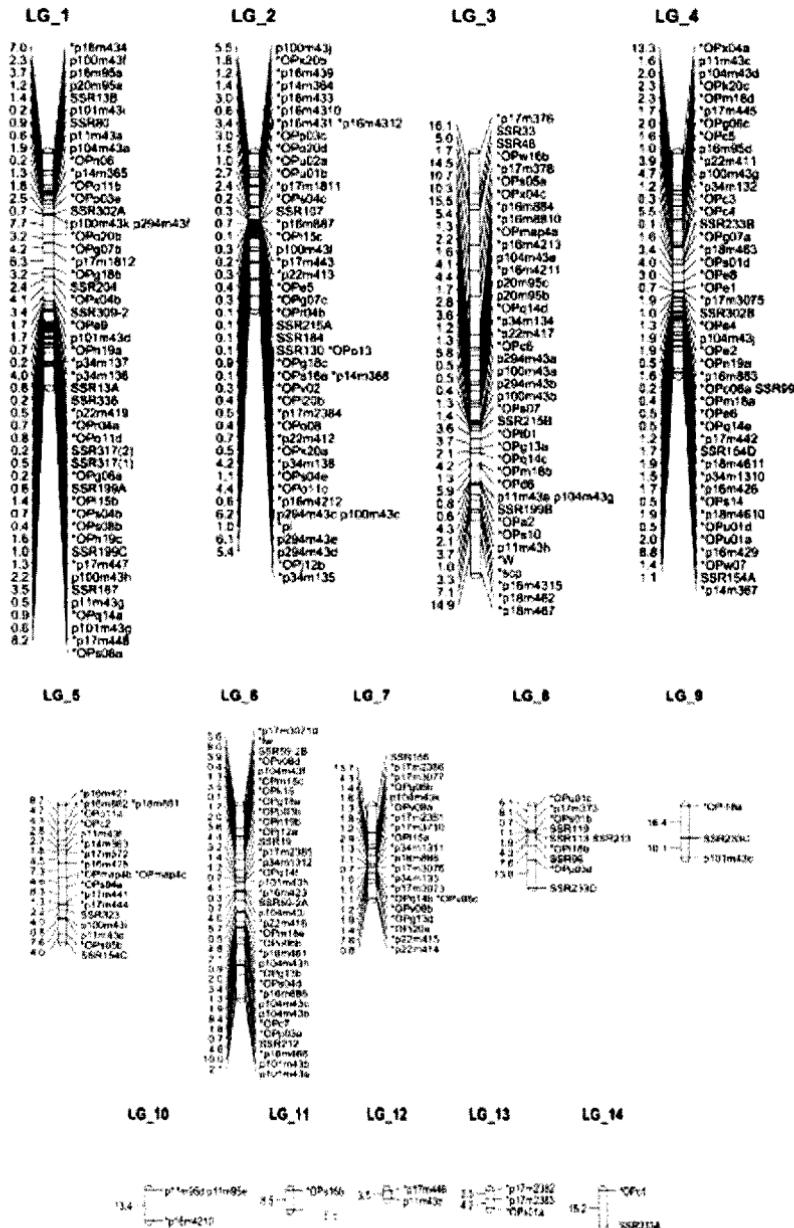
**Table 2.** Summary of the marker distribution on different linkage groups of the *Lens* sp. genomes

Linkage group	Length of the linkage group (cM)	Total number of markers	LOD score
1	93.3	50	4
2	63.5	47	5
3	171.9	41	5
4	90.1	45	4
5	69.6	19	4
6	98.3	35	4
7	47.4	21	4
8	41.6	10	5
9	26.5	3	4
10	13.4	3	4
11	8.5	2	4
12	3.5	2	3
13	7.7	3	3
14	15.2	2	3
<b>Total</b>	<b>750.5</b>	<b>283</b>	

The new map spanned 750.5 cM (Kosambi function) with an average distance between two markers of 2.7 cM. The Fusarium vascular wilt resistance was localized on LG 6, and this resistance gene was flanked by microsatellite marker SSR59-2B and AFLP marker p17m30710 by a distance of 8.0 cM, and 3.5 cM, respectively [33]. Further analysis for the association between these markers and *Fw* was confirmed. However, only SSR59-2B was closely linked with *Fw* at the estimated linkage distance of 19.7 cM.

## Acknowledgment

The authors gratefully acknowledge the technical support extended by, the genetic resource unit and the biotechnology staff towards the lentil breeding program. The authors' research was supported by grants to ICARDA from the German Federal Ministry of Economic Cooperation and Development (BMZ, Bonn, Germany).



**Fig. 2.** A genetic linkage map of *Lens* sp. based on microsatellite, AFLP, RAPD and morphological markers [33]. The marker names beginning with an asterisk are those markers mapped by [24]. The values on the left side of the individual linkage groups represent distance in centiMorgans calculated using the Kosambi mapping function [33].

## References

- [1] Agarwal SC, Singh K, Lal SS. Plant protection of lentil in India. In: Erskine W, Saxena MC, eds. Proceedings of the Seminar on Lentil in South Asia, 11-15 March, 1991, New Delhi, India. ICARDA, Aleppo, Syria, 1993: 147-167.
- [2] Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. *Plant Mol Biol* 1991; 9: 208-218.
- [3] Baidya BN, Eunus AN, Sen S. Estimations of variability and correlation in yield and yield contributing characters in lentil (*Lens culinaris*). *Environment and Ecology* 1988; 6(3): 694-697.
- [4] Bakr MA. Plant protection of lentil in Bangladesh. In: Erskine W, Saxena MC, eds. Proceedings of the Seminar on Lentil in South Asia, 11-15 March, 1991, New Delhi, India. ICARDA, Aleppo, Syria, 1993: 177-186.
- [5] Barulina HI. Lentil of the USSR and other countries. *Bull Appl. Bot. Gent. Plant Breeding (Lenin)* Suppl. 1930; 40:1-319.
- [6] Bayaa B, Erskine W, Hamdi A. Response of wild lentil to *Ascochyta fabae* f. sp. *lentis* from Syria. *Genetic Resources and Crop Evolution* 1994; 41: 61-65.
- [7] Bayaa B, Erskine W, Hamdi A. Evaluation of a wild lentil collection for resistance to vascular wilt. *Genetic Resources and Crop Evolution* 1995; 42: 231-235.
- [8] Bayaa B, Erskine W. Diseases of lentil. The pathology of food and pasture legumes (DJ, Allen Lenne JM, eds. should CAB read as CAB). CAB International 1998: 423-472.
- [9] Bejiga G, Tadesse N, Erskine W. We fixed rust, now wilt. *ICARDA Caravan*, 1998; 9: 12.
- [10] Choumane W, Winter P, Baum M, Kahl G. Conservation of microsatellite flanking sequences in different taxa of Leguminosae. *Euphytica* 2004; 138: 239-245.
- [11] CIPAL. Coordinated Improvement Program for Australian Lentils DPI, Horsham, Victoria, Australia, 2006.
- [12] Cubero JI. Origin, taxonomy and domestication 1981:15-38. In: Webb C., Hawtin G, eds. *Lentils*. CAB, London, UK.
- [13] Durán Y, Fratini R, García P, Pérez de la Vega M. An intersubspecific genetic map of *Lens*. *Theor Appl Genet* 2004; 108: 1265-1237.
- [14] Erskine W, Myveci K and Izgin N. Screening a world lentil collection for cold tolerance. *LENS Newsletter* 1981; 8: 5-8.
- [15] Erskine W, Witcombe, JR. Lentil Germplasm Catalog. International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria, 1984.
- [16] Erskine W, Williams PC, Nakhoul H. Genetic and environmental variation in the seed size, protein, yield and cooking quality of lentils. *Field Crops Res* 1985; 12: 153-161.
- [17] Erskine W, Choudhary MA. Variation between and within lentil landraces from Yemen Arab Republic. *Euphytica* 1986; 35: 695-700.
- [18] Erskine W, Adham Y, Holly L. Geographic distribution of variation in quantitative characters in a world lentil collection. *Euphytica* 1989; 43: 97-103.
- [19] Erskine W, Ellis RH, Summerfield RJ, et al. Characterization of responses to temperature and photoperiod for time to flowering in a world lentil collection. *Theor Appl Genet* 1990; 80: 193-199.

- [20] Erskine W, Hussain A, Tahir M, et al. Field evaluation of a model of photothermal flowering responses in a world lentil collection. *Theor Appl Genet* 1994; 88: 423-428.
- [21] Erskine W, Saxena NP, Saxena MC. Iron deficiency in lentil: Yield loss and geographic distribution in a germplasm collection. *Plant and Soil* 1993; 151: 249-254.
- [22] Erskine W, Chandra S, Chaudhury M, et al. A bottleneck in Lentil: Widening the Genetic Base in South Asia. *Euphytica* 1998; 101: 207-211.
- [23] Eujayl I, Baum M, Ershne W, et al. The use of RAPD markers for lentil genetic mapping and the evaluation of distorted F<sub>2</sub> segregation. *Euphytica* 1997; 96: 405-412.
- [24] Eujayl I, Baum M, Powell W, et al. A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. *Theor Appl Genet* 1998a; 97: 83-89.
- [25] Eujayl I, Erskine W, Bayaa B, et al. Fusarium vascular wilt in lentil: Inheritance and identification of DNA markers for resistance. *Plant Breeding* 1998b; 117: 497-499.
- [26] Eujayl I, Erskine W, Baum M, et al. Inheritance and linkage analysis of frost injury in a lentil population of recombinant inbred lines. *Crop Sci* 1999; 39: 639-642.
- [27] FAOSTAT Database 2005.
- [28] Ferguson ME, Maxted N, Van Slageren M, Robertson LD. A re-assessment of the taxonomy of *Lens* Mill. (*Leguminosae, Papilionoideae, Vicieae*). Bot J Linn Soc 2000; 133: 41-59. (En).
- [29] Ford R, Pang ECK, Taylor PWJ. Diversity analysis and species identification in *Lens* using PCR generated markers. *Euphytica* 1997; 96: 247-255.
- [30] Furman BJ. Methodology to establish a composite collection: Case study in lentil. *Plant Genetic Resources: Conservation and Utilization* 2006; 4(1): 2-12.
- [31] Hamdi A, Küsmenoglu I, Erskine W. Sources of winter hardiness in wild lentil. *Genetic Resources and Crop Evolution* 1996; 43: 63-67.
- [32] Hamdi A, Erskine, W. Reaction of wild species of the genus *Lens* to drought. *Euphytica* 1996; 91: 173-179.
- [33] Hamwieh A, Udupa SM, Choumane W, et al. A genetic linkage map of lentil based on microsatellite and AFLP markers and localization of Fusarium vascular wilt resistance. *Theor Appl Genet* 2005; 110: 669-677.
- [34] Hansen J, Renfrew JM. Palaeolithic-neolithic remains at Franchthi cave, Greece. *Nature* 1978; 71: 349-352.
- [35] Harvest Plus. HarvestPlus Challenge Program Annual report. ICARDA, Aleppo, Syria, 2006: 19.
- [36] Havey MH, Muehlbauer FJ. Linkages between restriction fragment length, isozyme and morphological markers in lentil. *Theor Appl Genet* 1989; 77: 395-401.
- [37] Khare MN, Agrawal SC, Jain AC. Diseases of lentil and their control. Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur, Madhya Pradesh India, 1991: 29.
- [38] Hüttel B, Winter P, Weising K, et al. Sequence-tagged microsatellite-site markers for chickpea (*Cicer arietinum* L.). *Genome* 1999; 42: 210-217.
- [39] Lakhani JP, Holker S, Misra R. Genetics of seedling vigor and hard seed in lentil. *LENS Newsletter* 1986; 13(2): 10-12.
- [40] Levinson G, Gutman GA. Slipped-strand misrepairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 1987; 4: 203-221.
- [41] Litt M, Luty JA. A hyper variable microsatellite revealed by *in-vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 1989; 44: 397-401.

- [42] Makkouk K, Kumari S, Sarker A, et al. Registration of six lentil germplasm lines with combined resistance to viruses. *Crop Sci* 2001; 41: 931-932.
- [43] Malhotra RS, Sarker A, Saxena MC. Drought Tolerance in Chickpea and Lentil - Present Status and Future Strategies. In: Srinivas C Rao, Ryan J, eds. Challenges and Strategies for Dryland Agriculture. J CSSA Special Publication 32, Madison, Wisconsin, USA, 2004; 257-274.
- [44] Morgante M, Olivieri AM. PCR-amplified microsatellites as markers in plant genetics. *Plant J* 1993; 3: 175-182.
- [45] Muehlbauer FJ, Slinkard AE. Genetics and Breeding Methodology. In: Webb C, Hawtin G, eds. *Lentils*. Slough, UK: Commonwealth Agricultural Bureaux, 1981; pp69-90.
- [46] Muehlbauer FJ, Weeden NF, Hoffman DL. Inheritance and linkage relationships of morphological and isozyme loci in lentil (*Lens Miller*). *J. Heredity* 1989; 80: 298-303.
- [47] Nasir M, Bretag T. Screening lentil for Ascochyta blight resistance. CLIMA Annual Report 1998; 14: 2-6.
- [48] Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1996; 1: 215 –222.
- [49] Ramgiry SR, Paliwal KK, Tomar SK. Variability and correlations of grain yield and other qualitative characters in lentil. *LENS Newsletter* 1989; 16 (1): 19-21.
- [50] Robertson LD, Singh KB, Erskine W, Abd El Moneim AM. Useful genetic diversity in germplasm collections of food and forage legumes from West Asia and North Africa. *Genetic Resources and Crop Evolution* 1996; 43: 447-460.
- [51] Robertson LD, Erskine W. Lentil. In: Fuccillo D, Sears L, Stapleton P eds. *Biodiversity in Trust*. Cambridge: Cambridge University Press, 1997: 128-138.
- [52] Rubeena Ford R, Taylor PWJ. Construction of an intraspecific linkage map of lentil (*Lens culinaris* ssp. *culinaris*). *Theor Appl Genet* 2003; 107: 910-916.
- [53] Sahai JP, Singh J, Agarwal I et al. Studies on variability for seed size, permeability of seed coat to water and germination in lentil. *LENS Newsletter* 1986; 13(2): 14-15.
- [54] Sangam L, Dwivedi SL, Blair MW, et al. Using Genomics to Exploit Grain Legume Biodiversity in Crop Improvement. In: Janick J, ed. *Plant Breeding Reviews Volume 26*. UK: John Wiley, 2006.
- [55] Sarker A, Erskine W. Harvesting Profits in Lentil Fields. ICARDA Caravan 1998; 9: 13-14.
- [56] Sarker A, Erskine W, Hassan MS, et al. Registration of "Barimasur-2" Lentil. *Crop Sci* 1999a; 39: 875.
- [57] Sarker A, Erskine W, Hassan MS, et al. Registration of "Barimasur-4" Lentil. *Crop Sci* 1999b; 39: 876.
- [58] Sarker A, Malhotra RS, Erskine W. Drought resistance in chickpea and lentil in Mediterranean environments. In Proc. LEGUMED Symposium, Rabat, Morocco, AEP-France, 2001; 119-126.
- [59] Sarker A, Kusmenoglu I, Aydogan A, et al. Winter hardiness in lentil and factors determining success of a lentil crop. Abs. Proc. 5<sup>th</sup> European Conf. On Grain Legumes, Dijon, France, 2004: 45.
- [60] Sarker A, Bayaa B, El-Hassan H, et al. New sources of resistance for fusarium wilt in lentil (*Lens culinaris* Medikus subsp. *culinaris*). *Lentil Research* 2004a; 1: 30-33.

- [61] Sarker A, Erskine W, Saxena MC. Global perspective on lentil improvement. In: Masood Ali, Singh BB, Shiv Kumar, Dhar V, eds. Pulses in New Perspective. Indian Society of Pulses Research and Development, IIPR, Kanpur, India, 2004b; 543p.
- [62] Sarker A, Erskine W, Singh M. Variation in root and shoot traits and their relationship in drought tolerance in lentil. *Genetic Resources and Crop Evolution* 2005; 52: 87-95.
- [63] Sarker A, Erskine W. Recent Progress in the ancient lentil. *Journal of Agric. Sciences, Cambridge* 2006; 144: 1-11.
- [64] Sarwar DM, Kaul AK, Quader M. Correlation studies in lentils. *LENS Newsletter* 1982; 9: 22-23.
- [65] Sharma SK, Dawson IK, Waugh R. Relationships among cultivated and wild lentils revealed by RAPD analysis. *Theor Appl Genet* 1995; 91: 647-654.
- [66] Sharma SK, Knox MR, Ellis THN. AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theor Appl Genet* 1996; 93: 751-758.
- [67] Sindhu JS, Mishra HO. Genetic variability in Indian *microsperma* type lentil. *LENS Newsletter* 1982; 9: 10-11.
- [68] Srivastava SP, Bhandari TMS, Yadav CR, et al. Boron deficiency in lentil: Yield loss and geographic distribution in a germplasm collection. *Plant and Soil* 2000; 219: 147-151.
- [69] Tufail M, Ahmed M, Ali A. Masoor-93: An ideal combination of characters for Punjab province, Pakistan *LENS Newsletter* 1995; 22 (1/2): 50-52.
- [70] Udupa SM, Robertson LD, Weigand F, et al. Allelic variation at (TAA)<sub>n</sub> microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. *Mol Gen Genet* 1999; 261: 354-363.
- [71] Udupa SM, Baum M. High mutation rate and mutational bias at microsatellite loci in chickpea (*Cicer arietinum* L) *MGG* 2001; 265: 1097-1103.
- [72] Vaillancourt RE, Slinkard AE. Inheritance of new genetic markers in lentil (*Lens Miller*). *Euphytica* 1992; 64: 227-236.
- [73] Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 1995; 23: 4407-4414.
- [74] Webb C, Hawtin G, eds. Lentils. CommonWealth Bureau AB, Page Bros (Norwich). UK: 1981.
- [75] Weber JL. Informativeness of human (dC-dA)<sub>n</sub> (dG-dT)<sub>n</sub> polymorphisms. *Genomics* 1990; 7: 524-530.
- [76] Weber J, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet* 1993; 2: 1123-1128.
- [77] Winter P, Pfaff T, Udupa SM, et al. Characterization and mapping of sequence tagged microsatellite sites in the chickpea (*Cicer arietinum* L) genome. *Mol Gen Genet* 1999; 262: 90-101.
- [78] Winter P, Beko-Iseppon AM, Huttel B, et al. A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* *C. reticulatum* cross: localization of resistance genes for *Fusarium* wilt races 4 and 5. *Theor Appl Genet* 2000; 101: 1155-1163.
- [79] Yau SK, Erskine W. Diversity of boron-toxicity tolerance in lentil growth and yield. *Genet Resources and Crop Evolution* 2000; 47: 55-61.
- [80] Youngman VE. Lentils - a pulse of the Palouse. *Economic Botany* 1967; 22: 135-139.

- [81] Zohary D, Hopf M. Domestication of plants in the World. Oxford, UK: Clarendon Press, 1988.
- [82] Zurawski G, Clegg MT. *rbcL* sequence data and phylogenetic reconstruction in seed plants: foreword. Ann Mo Bot Gard 1993; 80: 523–525.

# Phylogenetic Relationships of the Genus *Phlox* (Polemoniaceae): Comparing and Combining Data Sets from the Chloroplast Genome

CAROLYN J. FERGUSON<sup>1</sup>, SHANNON D. FEHLBERG<sup>2</sup>,  
KRISTEN A. FORD<sup>3</sup> and SUZANNE C. STRAKOSH<sup>4</sup>

<sup>1, 2, 3, 4</sup>Herbarium and Division of Biology, Kansas State University,  
Manhattan, Kansas USA

## ABSTRACT

The genus *Phlox* (Polemoniaceae) is a large, predominantly North American plant genus that presents interesting questions in terms of evolutionary diversification. The author's ongoing research focuses on the phylogeny and evolutionary biology of this group. This study reports new sequence data for several noncoding chloroplast DNA regions: *trnD-trnT*, *trnL-trnF* and *trnS-trnG*. These data are compared to previous data for the same samples from a restriction site study of the *Phlox* chloroplast genome. Although the restriction site data provide more phylogenetic information, the phylogeny based on data from all three sequence regions is substantially resolved and is in agreement with the phylogeny based on restriction site data. The combination of all sequence and restriction site data yields the strongest chloroplast phylogeny for *Phlox* to date, and provides further support for the taxonomic findings of previous studies.

**Key Words:** *Phlox*, Polemoniaceae, phylogeny, chloroplast DNA

*Address for correspondence:* Carolyn J. Ferguson, Herbarium and Division of Biology, 116 Ackert Hall, Kansas State University, Manhattan, Kansas 66506-4901, USA. Tel: 785-532-3166, Fax: 785-532-6653, E-mail: ferg@ksu.edu

## INTRODUCTION

*Phlox* L. is the largest genus of the mostly New World plant family Polemoniaceae. The group comprises approximately 65 species of subshrubs and perennial and annual herbs, and ranges across the major part of North America (with one species in Asia; [29]). Like many members of the family, *Phlox* has received considerable attention from systematists and evolutionary biologists [10, 11, 15, 16, 17, 18, 19]. It exhibits intriguing variation in terms of morphology, ecological affinities and ploidy levels; and is known for hybridization ([8, 9] and references therein). In addition, *Phlox* includes several species of horticultural significance, ranging from the tall *P. paniculata* (garden phlox) to the creeping *P. subulata* (moss phlox) to the showy annual *P. drummondii* (annual phlox; Fig. 1). Phylogenetic relationships of *Phlox* are of great interest because of the diversity exhibited by member species: inference of the patterns of diversification in the group improves the understanding of the evolutionary processes underlying those patterns.

Data from the chloroplast genome (chloroplast DNA, cpDNA) are widely used for phylogenetic inference of plant groups at many taxonomic levels [2, 14, 25]. At lower taxonomic levels (i.e., among closely related species), detection of adequate levels of variation for phylogeny reconstruction can be an issue. The mapping and scoring of cpDNA restriction sites is an effective approach to cpDNA phylogeny development at low taxonomic levels [13, 21]. However, this approach has been employed less widely in recent years, partly because it is very labor-intensive. Sequencing of cpDNA regions is fast and efficient, and a significant advantage of sequence data is that they can be readily employed in further research (whereas, the addition of new data to a restriction site data set must be accomplished with detailed comparison to the original, raw data, and ideally by the original researcher or set of researchers). Researchers over the years have developed and presented conserved primer pairs for amplification and sequencing of specific cpDNA regions [3, 12, 28] (refer to Shaw et al. [26] for additional references), and a multitude of systematists have employed these primer sets in phylogenetic studies of a diversity of plant groups. Shaw et al. [26] present a valuable study exploring the utility of various noncoding cpDNA regions for phylogeny reconstruction at lower levels. When employment of a single sequence region yields less phylogenetic information than is desired, sequencing of an additional region or regions



**Fig. 1.** The showy flowers of *Phlox* have made it a favorite ornamental for cultivation. Two species are shown here from natural populations: left, closeup of *P. pilosa*; right, *P. drummondii* subsp. *mcallisteri*. The *Phlox* corolla consists of five petals fused into a narrow tube below with flared lobes above; corolla color ranges from white to bright pink or lavender (rarely red or yellow); and flowers are generally sweetly scented.

can be carried out. When data are available, the combination of restriction site and sequence data can be expected to yield a strengthened phylogenetic hypothesis.

The present study builds on the *Phlox* cpDNA restriction site data of Ferguson and Jansen [9] by adding sequence data from three noncoding chloroplast regions (*tmD-tmT*, *tmL-tmF* and *tmS-tmG*) for a common set of samples. The four separate data sets (with equivalent sampling) and their resultant phylogenies are compared, as are the cpDNA restriction site versus combined sequence-based data sets and phylogenies. Finally, all of these data are combined for the strongest chloroplast-based phylogenetic hypothesis for *Phlox*. This contributes to the understanding of diversification of *Phlox*, and it also informs ongoing work, which predominantly uses sequence data to estimate relationships of the entire genus. More generally, this study provides a comparison of cpDNA restriction site data and sequence data in phylogeny reconstruction for closely related taxa.

## MATERIALS AND METHODS

### Taxon Sampling, DNA Extraction

*Phlox* samples included were a representative subset of those from a previous phylogenetic study employing restriction site data [9]. Furthermore, a sample of the putative sister genus, *Microsteris*, and a sample of *Leptosiphon* was included to serve as an outgroup (samples also in the 2002 study; [9]). Appropriate outgroup selection for phylogenetic investigations in the Polemoniaceae is facilitated by extensive phylogenetic data for the family (see [22], and references therein; [23]). Sample data (including taxonomy and voucher information) are presented in Table 1. Eastern North American *Phlox* taxa (including the most widely cultivated species) were best represented in the taxon set, although several western taxa were also included (Table 1). In most cases, the same DNA extracts used in the previous study were employed (these were extracted following a large scale CTAB extraction procedure followed by purification on cesium chloride/ethidium bromide gradients; [9]). In some cases, additional DNA was extracted from the original leaf material (archived at -80°C) following a small scale CTAB extraction procedure [20] (modified from [4]) without further purification.

### Restriction Site Data

Restriction site data for the samples in this study (Table 1) were obtained through a previous, larger study employing 18 restriction endonucleases, Southern blotting and filter hybridization with tobacco cpDNA probes, followed by autoradiography and inference of restriction sites [9]. While the samples were a subset of the original set, the data matrix included all characters from the original data set (as a consequence, some characters in the matrix were invariant). All included samples exhibited a single chloroplast type based on the restriction site data (i.e., there were no polymorphisms in the data set for any taxa [9]).

### Sequence Data

Following preliminary study of several noncoding cpDNA sequence regions for their utility in phylogeny reconstruction for *Phlox*, three regions were selected for use: the *trnD*<sup>GUC</sup>-*trnT*<sup>GGU</sup> intergenic spacer region, hereafter *trnD-trnT* (including the *trnY* and *trnE* genes; amplified using primers “*trnD* [tRNA-Asp (GUC)]” and “*trnT* [tRNA-Thr

Table 1. Phlox samples included in the phylogenetic studies

Taxonomy <sup>1</sup>	Sample <sup>3</sup>	Locality	Voucher <sup>4</sup>	Letter Symbol <sup>5</sup>	GenBank accession numbers <sup>6</sup>	
				tmD-tmT	tmL-tmF	tmS-tmG
<i>Phlox</i>						
Sect. <i>Annuae</i> A. Gray						
Subsect. <i>Speciosae</i>						
A. Gray						
<i>P. oklahomensis</i> Wherry	<i>P. oklahomensis</i>	Woods Co., OK	CF 269	AA	EF433203	EF433274
<i>P. nivalis</i> Sweet	<i>P. nivalis</i> subsp. <i>nivalis</i>	Liberty Co., FL	CF 279	X	EF433201	EF433272
Subsect. <i>Nanae</i> Wherry						
<i>P. nana</i> Nutt.	<i>P. nana</i> CH1	Chihuahua, MX	CF 245	FF	EF433204	EF433240
	<i>P. nana</i> TX1	Brewster Co., TX	CF 332	CC	EF433205	EF433241
	<i>P. noemeriana</i>	Burnet Co., TX	CF 87	S	EF433192	EF433263
Subsect. <i>Divaricatae</i>						
(Peter) Wherry						
<i>P. amoena</i> Sims	<i>P. amoena</i> GA2	Carlton Co., GA	CF 286	E	EF433211	EF433282
<i>P. divaricata</i> L.	<i>P. divaricata</i> NY	Genesee Co., NY	D. Goldman 912	D	EF433210	EF433281
	<i>P. divaricata</i> TN1	Maury Co., TN	G. Mayfield, sn, 10 Apr 95	A	EF433176	EF433248
<i>P. pattersoni</i> Prather	<i>P. pattersoni</i>	Nuevo Leon, MX	CF 256	Q	EF433182	EF433254
<i>P. pilosa</i> L.	<i>P. pilosa</i> subsp. <i>oxaricana</i> AR1	Crawford Co., AR	CF 153	U	EF433185	EF433257

Table 1 contd.

Table 1 contd.

<i>P. pilosa</i> subsp. <i>riparia</i>	Real Co., TX	CF 125	N	EF433177	EF433249	EF433213
<i>P. pilosa</i> subsp. <i>laisepala</i> TX1	Kerr Co., TX	CF 131	P	EF433178	EF433250	EF433214
<i>P. pilosa</i> subsp. <i>pilosa</i> TX1	Liberty Co., TX	CF 119	V	EF433179	EF433251	EF433215
<i>P. pilosa</i> subsp. <i>pilosa</i> TX2	Montgomery Co., TX	CF 121	O	EF433180	EF433252	EF433216
<i>P. pilosa</i> subsp. <i>pilosa</i> AL	Sumter Co., AL	CF 292	I	EF433209	EF433280	EF433245
<i>P. pilosa</i> subsp. <i>fulgida</i> MN	Waseca Co., MN	CF 204	C	EF433187	EF433258	EF433223
<i>P. pilosa</i> subsp. <i>deamii</i>	Christian Co., KY	CF 193	B	EF433181	EF433253	EF433217
Subsect. <i>Drummondianae</i>						
(Peter) Wherry			T	EF433191	EF433262	EF433227
<i>P. cuspidata</i> Scheele	Refugio Co., TX	CF 75	R	EF433190	EF433261	EF433226
<i>P. drummondii</i> Hook.	Lampasas Co., TX	CF 86				
Sect. <i>Phlox</i>						
Subsect. <i>Subulatae</i>						
(Peter) Wherry	Rutherford Co., TN	G. Mayfield, sn. 9 Apr 95	Z	EF433202	EF433273	EF433238
<i>P. bifida</i> Beck	<i>P. bifida</i> subsp. <i>stellaria</i>					

Table 1 contd.

Table 1 contd.

<i>P. subulata</i> L.	<i>P. subulata</i> subsp. <i>australis</i>	Grainger Co., TN	G. Wofford 96-4 CF 199	Y	EF433200	EF433271	EF433236
	<i>P. subulata</i> subsp. <i>subulata</i>	Lenawee Co., MI		W	EF433199	EF433270	EF433235
Subsect. <i>Stoloniferae</i> Wherry							
<i>P. stolonifera</i> Sims	<i>P. stolonifera</i>	Yancey Co., NC	G. Mayfield, sn, 26 Apr 95	GG	EF433198	EF433269	EF433234
Subsect. <i>Longifoliae</i> (Rydberg) Wherry							
<i>P. longifolia</i> Nutt. <i>calva</i>		Sublette Co., WY	CF 215	DD	EF433206	EF433277	EF433242
Subsect. <i>Phlox</i>							
<i>P. ovata</i> L. <sup>2</sup>	<i>P. ovata</i>	Grainger Co., TN	CF 173	HH	EF433183	EF433255	EF433219
	<i>P. pulchra</i>	Bibb Co., AL	CF 169	J	EF433184	EF433256	EF433220
<i>P. carolina</i> L. <i>angusta</i>	<i>P. carolina</i> subsp. <i>angusta</i>	Drew Co., AR	CF 159	K	EF433193	EF433264	EF433229
	<i>P. carolina</i> subsp. <i>turritella</i>	Escambia Co., FL	CF 331	H	EF433194	EF433265	EF433230
<i>P. glomerata</i> L.	<i>P. glomerata</i> subsp. <i>triflora</i> NC	Avery Co., NC	R. Mayfield, sn, 18 Jul 95	F	EF433195	EF433266	EF433231
	<i>P. glomerata</i> subsp. <i>interior</i>	Vanderburgh Co., IN	CF 194	L	EF433196	EF433267	EF433232
<i>P. maculata</i> L.	<i>P. maculata</i> subsp. <i>maculata</i>	Worth Co., IA	CF 206	G	EF433186		EF433222

Table 1 contd.

Table 1 contd.

Subsect. <i>Paniculatae</i> (Peter) Wherry	<i>P. amplifolia</i>	Knox Co., TN	E. Schilling, sn	M	EF433197	EF433268	EF433233
Sect. <i>Occidentales</i>							
A. Gray							
Subsect. <i>Albonarginatae</i> (Rydb erg) Wherry							
<i>P. abyssifolia</i> Greene	<i>P. abyssifolia</i> subsp. <i>abyssifolia</i>	Natrona Co., WY	D. Ferguson 1	EE	EF433207	EF433278	EF433243
Subsect. <i>Canescentes</i> (Rydb erg) Wherry							
<i>P. hoodii</i> Richardson	<i>P. hoodii</i> subsp. <i>hoodii</i>	Natrona Co., WY	D. Ferguson 2	BB	EF433208	EF433279	EF433244
<i>Microsteris gracilis</i> (Hook.) Greene	<i>M. gracilis</i>	San Mateo Co., CA	A. David 274	II	EF433188	EF433259	EF433224
<i>Leptosiphon nuttallii</i> (A. Gray) J.M. Porter & L.A. Johnson	<i>L. nuttallii</i>	Mono Co., CA	A. David 295	JJ	EF433189	EF433260	EF433225

<sup>1</sup>The general *Phlox* taxonomy presented follows that of Wherry ([29], the most recent comprehensive taxonomic treatment of the genus), with slight changes in taxon recognition as discussed in Ferguson and Jansen [9].

<sup>2</sup>*Phlox ovata* is considered the correct name for the taxon that has sometimes been considered *P. latifolia* [1, 30].

<sup>3</sup>For comparative purposes, sample names match those of Ferguson and Jansen [9], with the exception of *P. ovata* (refer to footnote 2).

<sup>4</sup>All voucher specimens are housed in the herbaria KSC and TEX; CF denotes C. Ferguson.

<sup>5</sup>Letter symbols correspond to those coded in Fig. 3.

<sup>6</sup>GenBank accession numbers are provided for all but a single sequence employed in this study. The tmL-tmF sequence of *P. maculata* included a short “string” of uncertainties (Ns); the sequence used is available from the first author, and an updated sequence will be made available via GenBank in the future.

(GGU)]" of Demesure et al. [3]); *tmL<sup>UAA</sup>-tmL<sup>UAA</sup>-tmF<sup>GAA</sup>*, hereafter *tmL-tmF* (including the *tmL* intron, the *tmL* 3' exon and the *tmL-tmF* intergenic spacer; amplified using primers "c" and "f" of Taberlet et al. [28]); and the *tmS<sup>GCU</sup>-tmG<sup>UUC</sup>* intergenic spacer, hereafter *tmS-tmG* (a single intergenic spacer; amplified using primers "trn S (GCU)" and "trn G (UCC)" of Hamilton [12]). Each 50- $\mu$ L Polymerase Chain Reaction (PCR) reaction contained 1X *taq* polymerase buffer (Promega, Madison, Wisconsin, USA), 2.5 mM MgCl<sub>2</sub>, 0.2 mM each dNTP (Promega), 0.2  $\mu$ M each primer, 1.25 units *taq* polymerase (Promega), and ~50 ng template DNA. All PCRs were conducted on MJ Research PTC-200 thermal cyclers. The PCR profiles generally consisted of 25-30 cycles of denaturing (1 minute at 95°C), annealing (1 minute at 56.5°C, 55°C and 52°C, respectively, for each region) and extension (1 minute at 72°C) steps, with a final five-minute extension step at 72°C. In many cases, the general PCR program was preceded by an initial five minute denaturing step, after which time the *taq* polymerase was added (a "hot start"), as this greatly increased product yield.

PCR products were purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, California, USA) and quantified by eye by comparison with a Low DNA Mass Ladder (Invitrogen Life Technologies, Carlsbad, California, USA) on an agarose gel. The same primers used for PCR were used for sequencing; in most cases for *tmL-tmF*, an additional, internal primer was also used—"d" of Taberlet et al. [28]. Approximately 60-70 ng of purified PCR product were used in a cycle sequencing reaction using an ABI Prism BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, California, USA; following the manufacturer's instructions, except that quarter reactions were used). Sequencing products were purified on Sephadex (G-50, molecular grade; Sigma, St. Louis, Missouri, USA) columns, dried by vacuum centrifuge and sent to the DNA Sequencing and Synthesis Facility at Iowa State University for gel run on an ABI automated sequencer. Both forward and reverse sequences were obtained for each region. Sequence data were edited using Sequencher (GeneCodes Corp., Ann Arbor, Michigan, USA) and aligned manually using Se-Al [24].

For each sequence region, a complete, aligned matrix was prepared for phylogenetic analysis with the gaps scored as missing data (refer to *Phylogenetic Analyses*, below). An additional data matrix was developed for each sequence region to allow scoring of insertion-deletion (indel) characters as described below:

- (1) for simple indel events greater than 1bp in length, all but one of the affected characters was excluded (i.e., removed from the matrix) so that the indel event would be scored only once when gaps were coded as a “fifth base”;
- (2) for rare, more complex indel events in which there were, additionally, phylogenetically informative characters within an indel region, an uncertainty (N) was scored in place of the gap (-) for the character in question for the appropriate taxa and the character was not excluded;
- (3) characters causing length variation in areas of repetitive DNA (poly-A/T regions), which may be expected to vary within taxa (even, potentially, within populations) were excluded; and
- (4) occasional areas involving gaps for which homology was difficult to assess were also excluded.

It is possible that some characters scored in the sequence data sets could correspond to particular restriction site characters also scored in this study (the restriction site characters reflect differences at the nucleotide sequence level). However, since the restriction site characters were drawn from across the chloroplast genome, any such characters (which would be effectively double scored in the combined restriction site and sequence data set) would be very few.

## **Phylogenetic Analyses**

Parsimony analyses were conducted for all data using PAUP\* 4.0b10 [27]. Basic heuristic searches were conducted with simple addition, TBR branch swapping, and the MULTREES option on. MAXTREES was set to 20,000. For all data sets including sequence data (refer below), analyses were conducted using the complete matrices (refer to *Sequence Data*, above) with gaps scored as missing data, and using the additional matrices with gaps scored as a “fifth base” (with characters excluded, as appropriate; refer above). Support for branches was evaluated by bootstrapping [6] using 1,000 “fast” bootstrap replicates in PAUP\*. In addition, a final parsimony analysis was conducted using the combined data matrix of all data examined (refer below): this analysis was conducted with 100 random additions, TBR branch swapping and the MULTREES option on, followed by full heuristic bootstrapping (with 1,000 pseudoreplicates, each with 10 random additions and TBR branch swapping).

Analyses were conducted for each data set separately (the restriction site data set and the six sequence data matrices [two per DNA region]). As the chloroplast genome is nonrecombining, it is appropriate to combine cpDNA data sets for a larger data set (still a single data source). However, incongruence length difference (ILD) tests [5] were conducted as implemented in PAUP\* (the partition homogeneity test) to test for homogeneity among the four separate data sets (using a data matrix with gaps in sequence data scored as a "fifth base"; refer above); 100 random partitions were analyzed using heuristic searches with 10 random additions, TBR branch swapping, and the MULTREES option on. Data from the three sequence regions were then combined and analyzed for a sequence-based phylogeny. Finally, all data—the restriction site data and all of the sequence data—were combined and analyzed for a *Phlox* cpDNA phylogeny based on all available data.

## RESULTS

The restriction site data set included 611 characters, 514 of which were variable and 169 of which were phylogenetically informative (Table 2), in this subset of the larger restriction site data set [9]. The sequenced regions varied in length from 1,038-1,093 bp for *trnD-trnT*, 889-1,009 bp for *trnL-trnF*, and 677-711 bp for *trnS-trnG*, and have been deposited in GenBank (accession numbers provided in Table 1). In terms of phylogenetic information, the *trnS-trnG* data set had the largest number of phylogenetically informative characters (34 when gaps were coded), followed by *trnL-trnF* (22) and *trnD-trnT* (21; Table 2). Scoring of gaps as characters (refer to Materials and Methods, Sequence Data, above) increased the numbers of characters in each sequence data set (Table 2), with gaps representing 19, 18 and 38% of the phylogenetically informative characters in the *trnD-trnT*, *trnL-trnF* and *trnS-trnG* data sets, respectively. The combined sequence data matrix yielded 310 variable and 77 phylogenetically informative characters (with gaps coded; compared with 514 and 169, respectively, in the restriction site matrix; Table 2). Complete information for each of the data matrices is presented in Table 2 for comparative purposes, including length of the complete (restriction site) or aligned (sequence) matrix, number of variable characters and number of phylogenetically informative characters.

Tree statistics for the phylogenetic analyses are presented in Table 2, including number of most parsimonious trees, tree length, CI excluding uninformative characters, and number of nodes (based on a strict

**Table 2.** Characteristics of data matrices and maximum parsimony tree statistics, by data set

Data set	Matrix length	# var. char. <sup>1</sup>	# PI char. <sup>2</sup>	# of trees <sup>3</sup>	Tree length	CI excl. uninfr. char. <sup>4</sup>	# nodes BS ≥ 80% <sup>5</sup>
Restriction site	611	514	169	70	659	0.54	12
<i>trnD-trnT</i> , gaps missing	1,167	87	17	3	99	0.80	1
<i>trnD-trnT</i> , gaps coded	1,028	104	21	3	119	0.78	2
<i>trnL-trnF</i> , gaps missing	1,053	79	18	20,000	84	0.86	3
<i>trnL-trnF</i> , gaps coded	805	126	22	8	222	0.82	4
<i>trnS-trnG</i> , gaps missing	828	89	21	20,000	97	0.78	3
<i>trnS-trnG</i> , gaps coded	650	119	34	74	143	0.72	7
Sequences combined, gaps missing	3,048	254	56	20,000	285	0.75	8
Sequences combined, gaps coded	2,483	310	77	2	362	0.72	8
All data combined, gaps missing	3,659	768	225	800	948	0.58	16
All data combined, gaps coded	3,094	824	246	240	1,028	0.58	15

<sup>1</sup>Number of variable characters.<sup>2</sup>Number of phylogenetically informative characters.<sup>3</sup>Number of most parsimonious trees.<sup>4</sup>Consistency Index, CI, excluding uninformative characters.<sup>5</sup>Number of nodes in strict consensus tree with bootstrap values greater than or equal to 80.

consensus tree) with bootstrap values 80 or higher. Figure 2 presents the strict consensus tree resulting from analysis of restriction site data alone. This phylogeny is a good representation of the restriction site-based phylogeny from the original, larger data set [9], and is a generally well-resolved and well-supported tree (Fig. 2). Strict consensus trees resulting

from analyses of each separate sequence region are presented in Fig. 3, with differences resulting from differential scoring of gaps (as missing data versus coded) indicated. Each of these phylogenies is lacking in resolution (with 13, 11 and 11 nodes resolved for *trnD-trnT*, *trnL-trnF* and *trnS-trnG*, respectively; Fig. 3). Interestingly, the *trnS-trnG* tree lacks substantial resolution in interior areas, despite the fact that this data set contained the greatest number of informative characters (Table 2). As expected, the ILD tests indicated no significant conflict in phylogenetic signal between or among all data sets ( $P=0.25$ ; refer to Figs. 2 and 3). There were very minor differences resulting from the differential scoring of gaps (Table 2; Fig. 3): in general, the scoring of gaps as a "fifth base" (with the appropriate characters excluded; refer to *Materials and Methods*, above) yielded phylogenies with slightly better resolution and higher bootstrap numbers (numbers of informative characters increased, and the values for CI [excluding uninformative characters] decreased), and the scoring of the gap information was favored.

The tree resulting from analysis of all sequence data (Fig. 4) reveals that these sequence data, as a whole, enable a fairly strong phylogenetic hypothesis for *Phlox*. However, they provide less phylogenetic information than the cpDNA restriction site data set employed (e.g., 20 nodes are resolved in the sequenced-based phylogeny versus 27 in the restriction site-based phylogeny; Figs. 2 and 4). Figure 5 presents the strict consensus tree based on the complete combined data; it is generally a well-resolved and well-supported phylogeny. There are minor differences between the trees developed from restriction site data, combined sequence data, and complete combined data (Figs. 2, 4 and 5), but these differences all occur in areas where support for relationships is weak. Overall, the phylogeny based on the complete combined data set (Fig. 5) provides excellent phylogenetic information for *Phlox*.

## CONCLUSION

The new and combined cpDNA data presented in this study enable an improved chloroplast-based phylogeny for the genus *Phlox*. Several phylogenetic findings with taxonomic implications have been noted in previous studies of the genus [7, 9] and are further supported here, as detailed below:

- (1) *Microsteris* is supported as sister to *Phlox*;
- (2) many of the major taxonomic groupings of Wherry [29] (Table 1) do not represent monophyletic groups, suggesting the genus is in

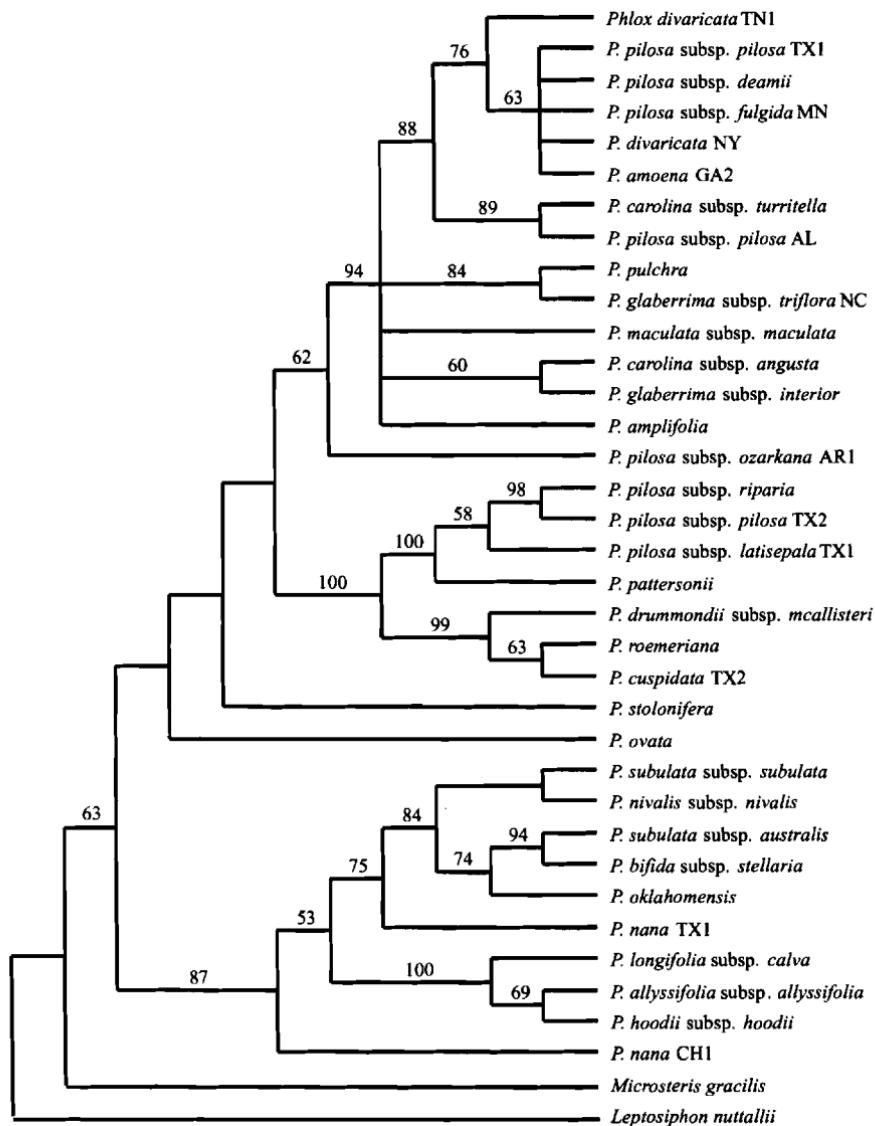
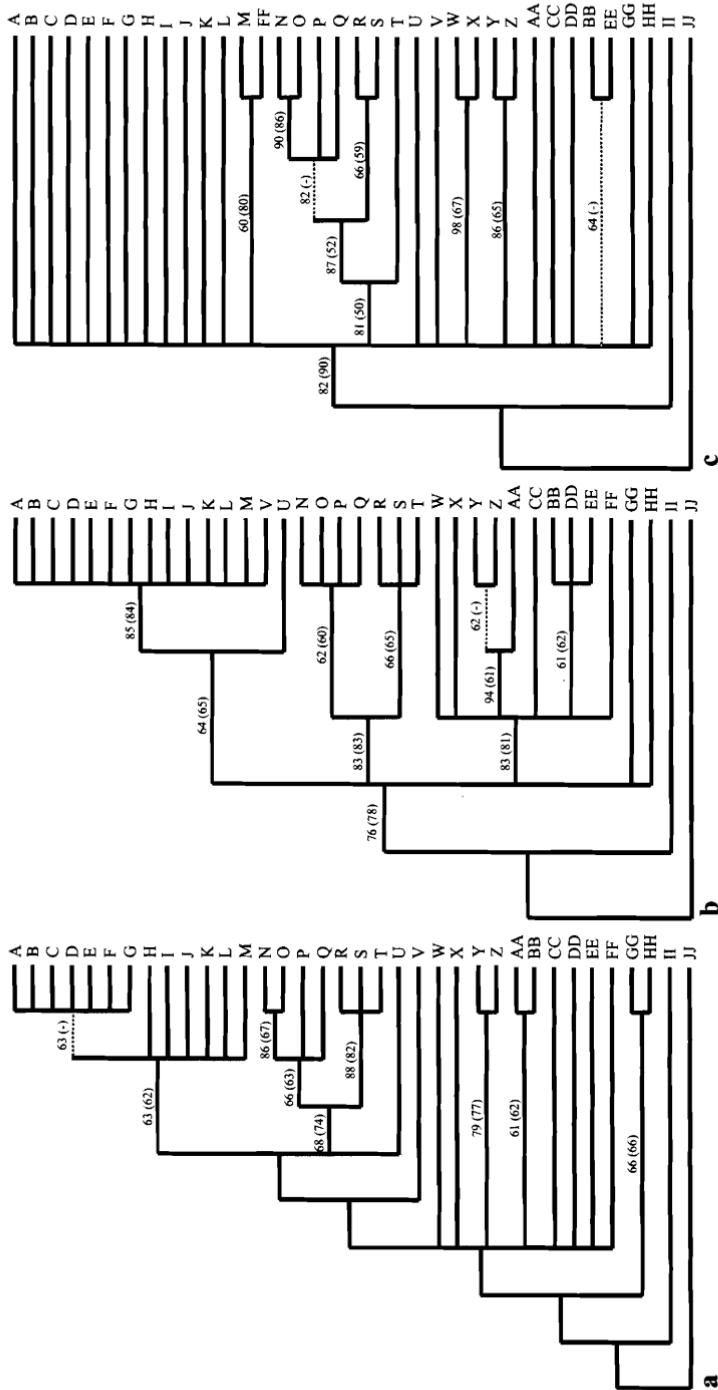


Fig. 2. Strict consensus tree resulting from analysis of cpDNA restriction site data (data from Ferguson and Jansen [9]). Numbers above branches are bootstrap values.

need of further taxonomic study and revision (which is in progress);

- (3) the low, "mat-forming" eastern taxa (*P. bifida*, *P. nivalis*, *P. oklahomensis* and *P. subulata*) resolve within a lineage of western



**Fig. 3.** Comparison of strict consensus trees from analyses of the three separate sequence data sets: (a) tmrD-trnT, (b) tmrL-trnF and (c) tmrS-trnG. Numbers above branches are bootstrap values from the analysis of the matrix with gaps coded (first number) and from that of the matrix with gaps scored as missing data (second number, in parentheses). Dotted lines denote branches that are not present in the strict consensus tree when gaps were scored as missing data.

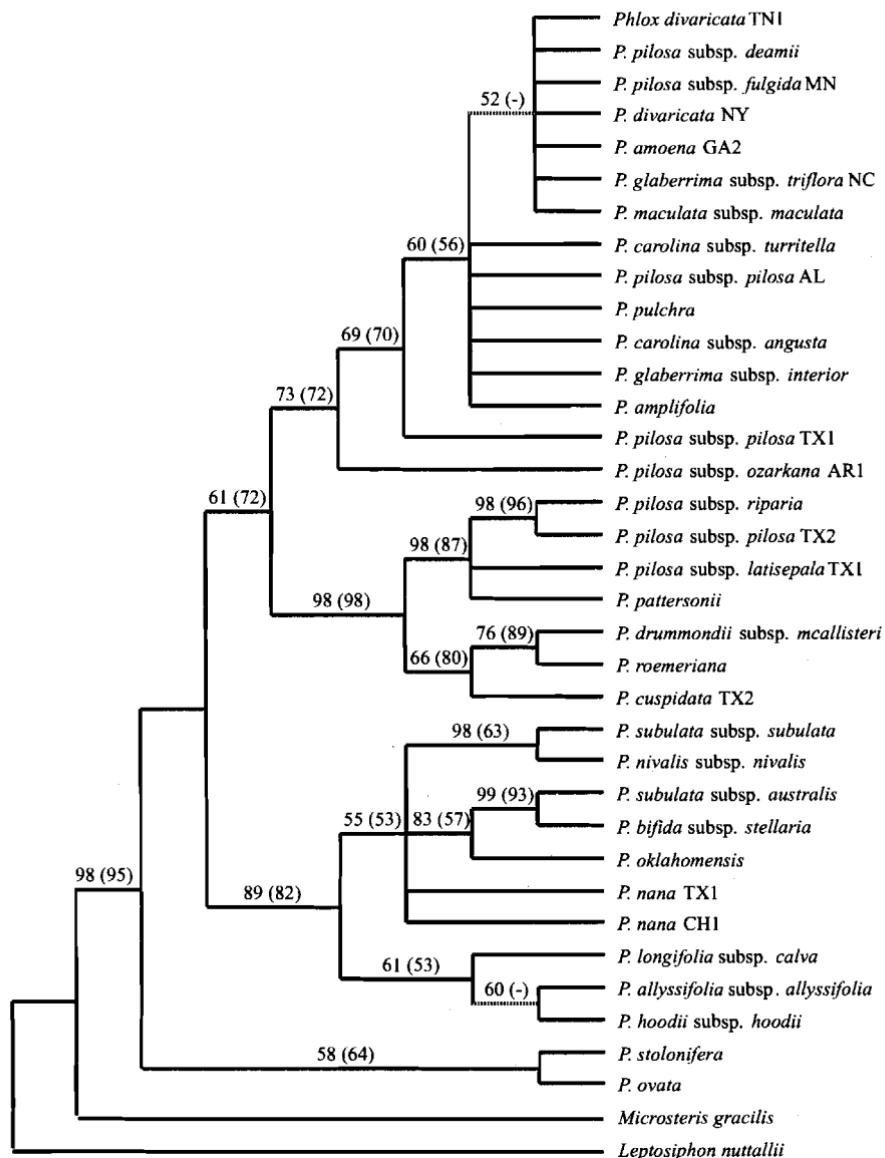
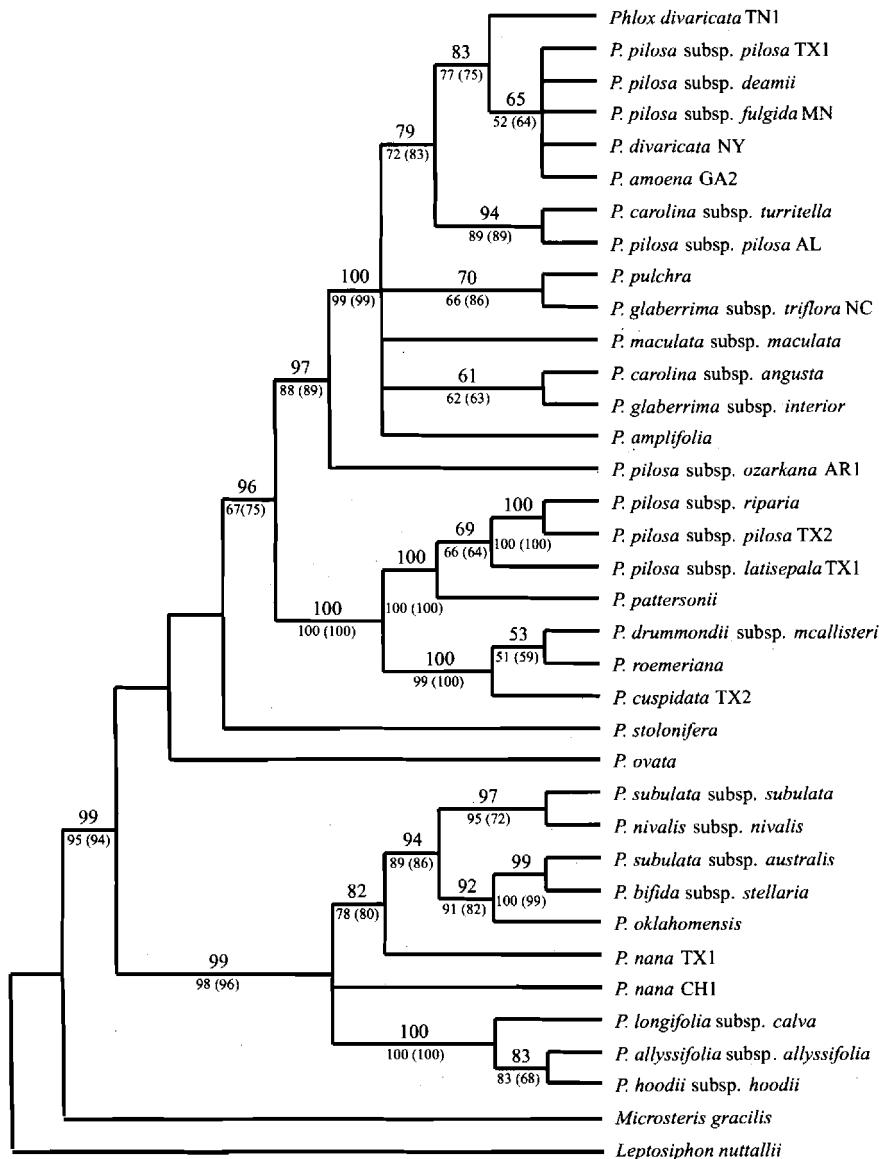


Fig. 4. Strict consensus tree resulting from analysis of combined sequence data (trnD-trnT, trnL-trnF and trnS-trnG). Numbers above branches are bootstrap values (as described in Fig. 3), and the branch shown with a dotted line is not present when gaps were scored as missing data.



**Fig. 5.** Phylogeny based on all available cpDNA data. The tree is a strict consensus tree resulting from analysis of restriction site data, and sequence data from trnD-trnT, trnL-trnF and trnS-trnG for the same set of samples. Numbers along branches are bootstrap values: the number above each branch is from the final, full heuristic analysis; the numbers below each branch are from the regular analyses and are ordered as described in Fig. 3.

North American taxa (including *P. alyssifolia*, *P. hoodii*, *P. longifolia* and *P. nana*, based on samples included in the present study), and are relatively distantly related to the remainder of the eastern species (this study and unpubl. data); and

- (4) the central Texas annual taxa (*P. cuspidata*, *P. drummondii* and *P. roemeriana*) form a well-supported clade.

In agreement with previous studies, the cpDNA data demonstrate the disparate phylogenetic placement of multiple samples of the *P. pilosa* complex in particular. Previous comparison of a cpDNA phylogeny based on restriction site data [9] with a nuclear-based phylogeny (based on sequence data from the nuclear ribosomal internal transcribed spacer, or ITS, region [7]) found significant incongruence between the two data sets from the different genomes with respect to placement of members of *P. pilosa* (as well as additional samples). This evidence is consistent with the hypothesis that past and possibly ongoing hybridization has affected relationships of *Phlox* [9]. The additional cpDNA data gathered as part of the present study aid in better understanding the very intriguing patterns of diversification in the genus *Phlox*, and further comparison of the cpDNA phylogeny to nuclear-based phylogenies will be valuable.

The cpDNA sequence data add to ongoing research exploring relationships (based on both the chloroplast and nuclear genomes) of the entire genus *Phlox*. Since sequence data are relatively easy to generate and because of the advantage they provide in ready combinability with new data, ongoing phylogenetic work on *Phlox* is sequence-based at present. Determination of the phylogenetic utility for *Phlox* of various cpDNA sequence regions relative to that of the original restriction site data is highly informative. In the present study the restriction site data [9] provided more phylogenetic information than the combined *tmD-tmT*, *tmL-tmF* and *tmS-tmG* sequence data. This suggests that addition of data from another noncoding cpDNA region or regions (and there are many candidate regions for which developed primers are already available; [26]) would be necessary to obtain equivalent phylogenetic resolution and support in a sequence-based tree relative to that based on the restriction site data. Jansen et al. [13], in comparing the utility of cpDNA restriction site and sequence data, found that generally multiple cpDNA regions must be sequenced and combined to yield a similar amount of phylogenetic information as is typically obtained from a restriction site study. The findings of the present study add further support to the value of cpDNA restriction site data in studies of

relationships at low taxonomic levels. Nonetheless, the *Phlox* phylogeny based on *tmD-tmT*, *tmL-tmF* and *tmS-tmG* is substantially resolved (e.g., it enables inference of all of the findings with taxonomic implications, discussed above), and it is in excellent agreement with the restriction site phylogeny. This clearly demonstrates the utility of these noncoding cpDNA sequence regions for phylogeny reconstruction in *Phlox*.

## Acknowledgment

Members of the Ferguson lab have provided technical support in the DNA sequencing work reported here, and the authors are particularly grateful to Audrey Atencio, Matt Fields, Danielle Froelich and Joe Lyon for their valuable assistance. Financial support for this research was provided by NSF (DEB-0089656), the KSU Center for the Understanding of Origins, and the Kansas Agricultural Experiment Station (Contribution No. 07-170-B). The restriction site study providing the basis for the comparative work reported here was funded by NSF (DEB-9623386, with R. Jansen).

## References

- [1] Brummitt RK. Report of the committee for spermatophyta: 38. *Taxon* 1993; 42: 687-697.
- [2] Chase MW, Soltis DE, Olmstead RG, et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Missouri Bot Gard* 1993; 80: 528-580.
- [3] Demesure B, Sodzi N, Petit, RJ. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molec Ecol* 1995; 4: 129-131.
- [4] Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 1987; 19: 11-15.
- [5] Farris JS, Källersjö M, Kluge AG, Bult C. Testing significance of incongruence. *Cladistics* 1995; 10: 315-319.
- [6] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; 39: 783-791.
- [7] Ferguson CJ, Krämer F, Jansen RK. Relationships of eastern North American *Phlox* (Polemoniaceae) based on ITS sequence data. *Syst Bot* 1999a; 24: 616-631.
- [8] Ferguson CJ, Levin DA, Jansen RK. Natural hybridization between an outcrossing and a selfing *Phlox* (Polemoniaceae): the maternal species of *F*<sub>1</sub> hybrids. *Plant Syst Evol* 1999b; 218: 153-158.
- [9] Ferguson CJ, Jansen RK. A chloroplast DNA phylogeny of eastern *Phlox* (Polemoniaceae): implications of congruence and incongruence with the ITS phylogeny. *Amer J Bot* 2002; 89: 1324-1335.

- [10] Grant V. Natural history of the phlox family: systematic botany. The Hague, Netherlands: Martinus Nijhoff, 1959.
- [11] Grant V, Grant KA. Flower pollination in the phlox family. New York: Columbia University Press, 1965.
- [12] Hamilton MB. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol Ecol 1999; 8: 513-525.
- [13] Jansen RK, Wee JL, Millie D. 1998. Comparative utility of chloroplast DNA restriction site and DNA sequence data for phylogenetic studies in plants. Soltis DE, Soltis PS, Doyle JJ, eds. Molecular systematics of plants II: DNA sequencing. Boston: Kluwer Academic, 1998: 87-100.
- [14] Kim K-J, Jansen, RK. *ndhF* sequence evolution and the major clades in the sunflower family. Proc Natl Acad Sci USA 1995; 92: 10379-10383.
- [15] Lendvai G, Levin DA. Rapid response to artificial selection on flower size in *Phlox*. Heredity 2003; 90: 336-342.
- [16] Levin DA. Hybridization between annual species of *Phlox*: population structure. Amer J Bot 1967; 54: 1122-1130.
- [17] Levin DA, Schaal BA. Reticulate evolution in *Phlox* as seen through protein electrophoresis. Amer J Bot 1970; 57: 977-987.
- [18] Levin DA. Consequences of long-term artificial selection, inbreeding, and isolation in *Phlox*: I. The evolution of cross-incompatibility. Evolution 1976; 30: 335-344.
- [19] Levin DA. Inbreeding depression in partially self-fertilizing *Phlox*. Evolution 1989; 43: 1417-1423.
- [20] Loockerman D, Jansen RK. 1996. The use of herbarium material for molecular systematic studies. In: Stuessy TF, Sohmer S, eds. Sampling the Green World. New York: Columbia University Press, 1996: 205-220.
- [21] Olmstead RG, Palmer JD. Chloroplast DNA systematics: a review of methods and data analysis. Amer J Bot 1994; 81: 1205-1224.
- [22] Porter JM, Johnson LA. A phylogenetic classification of Polemoniaceae. Aliso 2000; 19: 55-91.
- [23] Prather LA, Ferguson CJ, Jansen RK. Polemoniaceae phylogeny and classification: implications of sequence data from the chloroplast gene *ndhF*. Amer J Bot 2000; 87: 1300-1308.
- [24] Rambaut A. Se-Al: Sequence Alignment Editor. 1996. Available at <http://evolve.zoo.ox.ac.uk/>
- [25] Sang T, Crawford D, Stuessy TF. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Amer J Bot 1997; 84: 1120-1136.
- [26] Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Amer J Bot 2005; 92: 142-166.
- [27] Swofford DL. PAUP\*. Phylogenetic analysis using parsimony (\* and other methods), version 4.0b10. Sunderland, Massachusetts: Sinauer Associates, 2002.
- [28] Taberlet P, Gielly L, Pautou G, Bouvet J. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 1991; 17: 1105-1109.
- [29] Wherry ET. The genus *Phlox*. Morris Arb Monogr 1955; 3.
- [30] Wilbur RL. Lectotypification of *Phlox ovata* L. (Polemoniaceae). Taxon 1987; 36: 130-132.

# Genus *Silene* (Caryophyllaceae) – Evolutionary Diversification and Sex Chromosome Formation

JITKA ZLUVOVA<sup>1</sup>, BOHUSLAV JANOUSEK<sup>1</sup>, ROMAN HOBZA<sup>1,2</sup>, JAROSLAV MRACEK<sup>1</sup>, ALEXANDER WIDMER<sup>2</sup> and BORIS VYSKOT<sup>1</sup>

<sup>1</sup>Laboratory of Plant Developmental Genetics, Institute of Biophysics, Brno, Czech Republic

<sup>2</sup>ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology, Zurich, Switzerland

## ABSTRACT

The genus *Silene* has become an emerging model system for plant ecology, evolution and development. This chapter summarizes namely the progress of the main topics studied in this genus – sex determination in gynodioecious and dioecious plants, sex chromosome evolution on DNA sequence level, and mechanisms controlling sex bias in populations. The most studied dioecious species is *S. latifolia*, characterized by sex chromosomes of relatively recent origin (10–20 million years). The order of the genes on the *S. latifolia* X-chromosome and divergence between the homologous X- and Y-linked copies of these genes support the "evolutionary strata" model, with at least three consecutive expansions of the nonrecombining region on the Y-chromosome in this plant species. This finding increases the interest in this species, since *S. latifolia* provides a unique model to study processes taking part in the early sex chromosome evolution, e.g., recombination arrest and accumulation of the

sexually antagonistic genes. Another important species from this genus is represented by gynodioecious *S. vulgaris*. The study of the CMS in this species indicates that this model could be of the same importance among the gynodioecious plants as is *S. latifolia* among dioecious plants.

**Key Words:** *Silene*, gynodioecy, dioecy, sex ratio bias, sex chromosomes

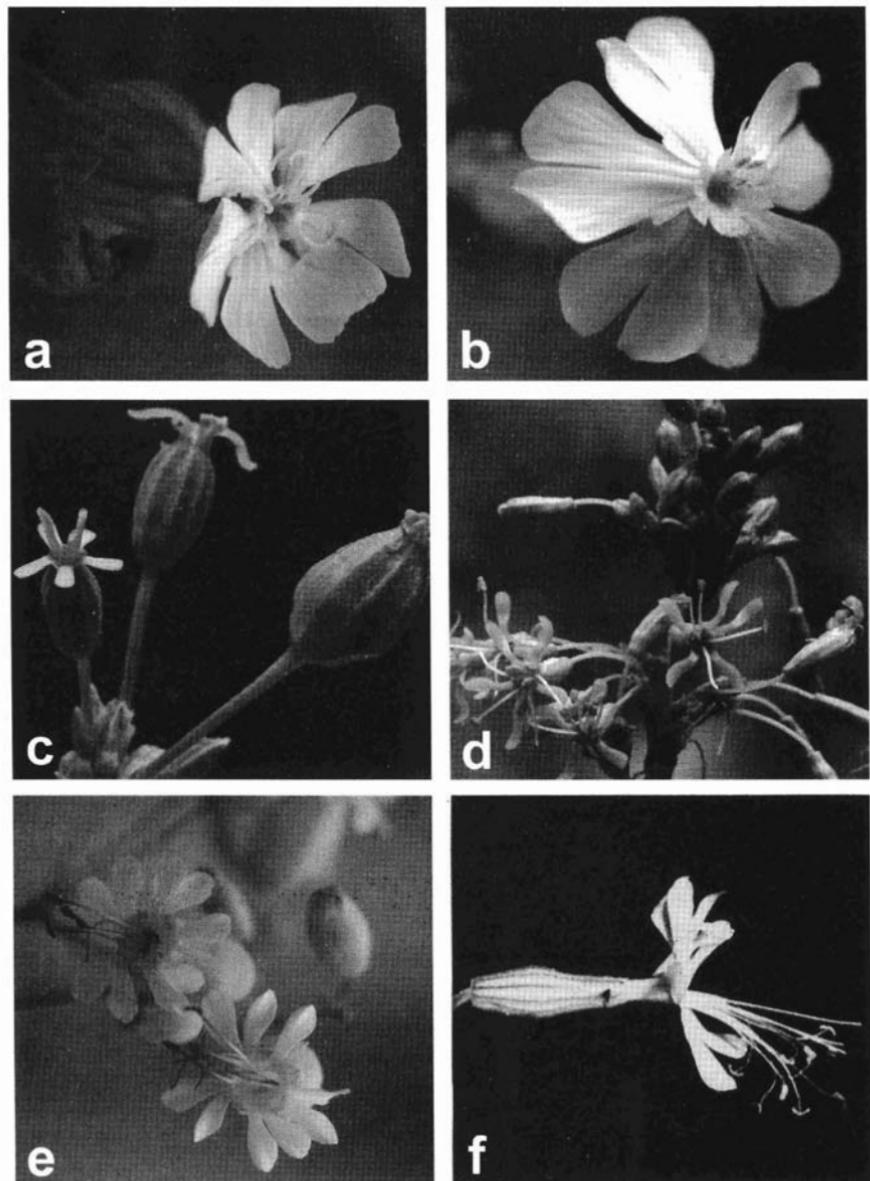
**Abbreviations:** BAC= bacterial artificial chromosome, CMS= cytoplasmic male sterility, DOP-PCR= degenerate oligonucleotide-primed PCR, FISH= fluorescence *in situ* hybridization

## INTRODUCTION

The plant genus *Silene* belongs to the family Caryophyllaceae in the order Caryophyllales. The family is mainly Holarctic with 86 genera and about 2,200 species that are classified into three subfamilies, i.e., Paronychioideae, Alsinoideae, and Caryophylloideae [22]. Complex and possibly homoplasious morphological characters have made the classification difficult. A recent molecular phylogeny of the family [76] found no strong support for monophyly of the three subfamilies, and Alsinoideae were found to be paraphyletic. The Paronychioideae were found to form a basal grade. If these results find support from other studies, major changes will be required in the family Caryophyllaceae in order to establish a phylogenetic subfamilial classification [76].

The genus *Silene* belongs to tribe Sileneae, regarded as being monophyletic by Fior et al. [76], but having a long history of taxonomic changes at the genus level [47, 149, 168]. Oxelman et al. [170] presented a revised generic classification of the tribe that is consistent with monophyletic, diagnosable groups identified in prior molecular phylogenetic analyses [60, 168, 169]. A total of eight genera were recognized. Of those, genus *Silene* is the largest. Morphological heterogeneity may indicate that the genus is not monophyletic [22, 170]. Phylogenetic patterns within the genus are still largely unclear due to frequent character reversals and parallelism [85] and a distinct phylogeny of the genus *Silene* is still lacking.

*Silene* species are annual, biennial or perennial herbs, rarely small shrubs. Flowers are monoecious, gynodioecious and rarely dioecious (Fig. 1). The approximately 700 species are subdivided into more than 20 sections and grow in Eurasia (ca. 600 spp.), Africa (ca. 40 spp.), and North America (ca. 50 spp.). The greatest diversity is found in Mediterranean areas [85] with two centres of distribution: (1) the Pamir and Hindukush region and adjacent areas, and (2) Transcaucasia, Azerbaijan and Kurdistan [22].



**Fig. 1.** Flowers of some *Silene* species. *S. latifolia* belongs to the most common dioecious species. It forms different individuals possessing either pistillate (female, a) or staminate flowers (male, b). Another *Silene* section with dioecious species is *Orites*, here represented by female (c) and male (d) inflorescences of *S. otites*. *S. vulgaris* (e) is a gynodioecious species forming hermaphrodite or male-sterile individuals. *S. viscosa* (f) is a fully hermaphrodite species.

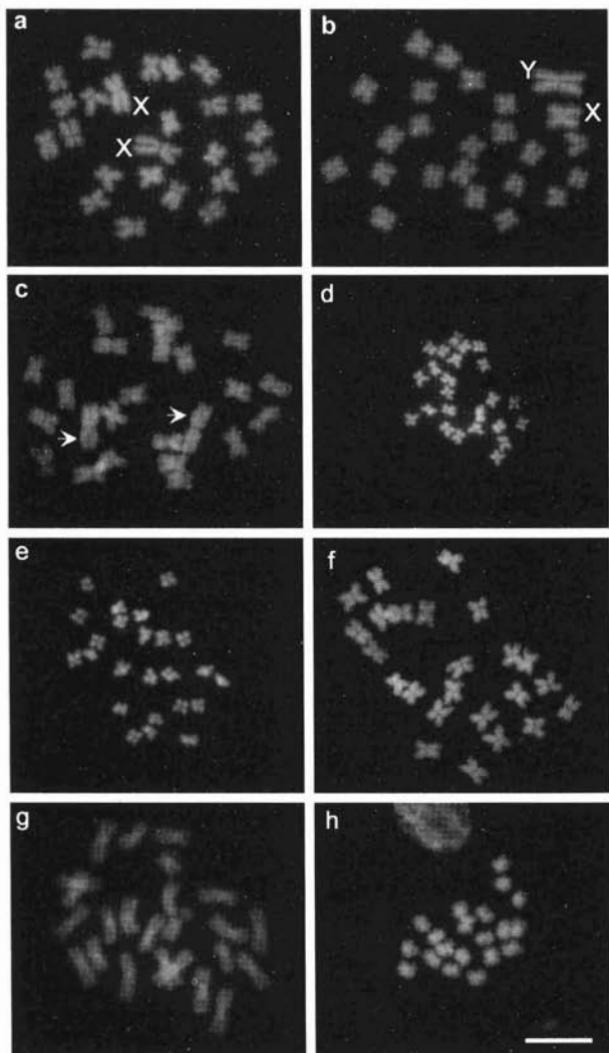
The basic chromosome number for tribe Sileneae is  $x=12$  ([22, 54], Figs. 2a-g), but members of *Silene* section *Conomorpha* have  $x=10$  ([53, 110, 111], Fig. 2h). *Silene* species vary greatly in genome size and structure. The nuclear genomes of *Silene* species are highly diversified as a result of numerous DNA amplifications and translocations [200]. The incidence of polyploidy is relatively low in *Silene*, but several arctic and subarctic taxa, a few Mediterranean taxa, and the majority of North American taxa are polyploids [115, 168, 174, 176].

## **SILENE – AN EMERGING MODEL SYSTEM FOR PLANT ECOLOGY, EVOLUTION AND DEVELOPMENT**

As a consequence of the great diversity of species, mating systems, ploidy levels, and ecological differences, the genus *Silene* is emerging as a model system for plant ecology, evolution and development.

Several classic experimental taxonomic studies on *Silene* [115, 177] have not only provided new insights into taxonomic relationships, but also provide valuable insights into the evolution of reproductive isolation among species [155], and thus in speciation. While reproductive isolation evolves gradually in homoploid divergence, polyploidy provides an instant reproductive isolation mechanism. Several cases of autopolyploidy have been reported, such as in *S. ciliata*, a Mediterranean species with discontinuous geographical distribution and populations ranging from diploids to 16-ploids [24], and *S. vulgaris* subspecies *macrocarpa*, which is tetraploid [13]. Although gene flow may occur between taxa that differ in ploidy level, autopolyploids often form morphologically and ecologically distinct taxa. Allopolyploidy, in which genomes from two distinct species are combined after hybridization in a newly formed polyploid, has also been reported [174, 175, 176] and may lead to diversification in some *Silene* lineages.

Given the wide range of habitats colonized by *Silene* species, ecological adaptation and divergence may play an important role in *Silene* diversification and speciation. Some species of *Silene*, such as *S. vulgaris* or *S. dioica*, are well-known for their tolerance to a range of heavy metals or to naturally heavy-metal rich serpentine soil [186, 187, 188, 228]. Adaptation to highly selective habitats, such as heavy metal rich soils, may represent a first step in ecological speciation as shown by the existence of serpentine endemics in *Silene* and many other plant groups [25]. Ecological adaptation and divergence are often quantitative in



**Fig. 2.** Mitotic metaphase chromosomes of several *Silene* species. The dioecious *S. latifolia* possesses large heteromorphic sex chromosomes, the two X-chromosomes in homogametic females (a) and one X-and one Y-chromosome (b) in males. The dioecious *S. dioecia* has a pair of large homomorphic sex chromosomes (indicated by arrows, c). A majority of *Silene* species harbours the same number of chromosomes,  $2n=24$ . They also include the gynodioecious *S. vulgaris* (d), and hermaphrodite *S. pendula* (e), *S. viscosa* (f), and *S. chalcedonica* (g). The different size of chromosomes (e.g., between *S. pendula* and *S. latifolia*) reflects large differences in the nuclear amount of DNA. *S. conica* (h) belongs to very few *Silene* species possessing only 20 chromosomes in mitotic metaphase. Slides were stained with DAPI, the bar represents 10  $\mu\text{m}$ .

nature and controlled by multiple genes. Recently, a quantitative trait locus (QTL) approach was used to study the number and genomic distribution of loci controlling traits that are presumably involved in serpentine adaptation [26, 27]. The results indicate that a few genes of major effect are responsible for serpentine adaptation in *S. vulgaris*, which provides the foundation for rapid evolution of ecotypes. Identification of the genes involved in habitat adaptation will be of great interest in the future and candidate gene approaches are a promising approach to identify these genes [7].

*Silene* species also attract a diverse group of pollinators. Of primary importance for pollinator attraction are floral form, colour and scent, and these traits are intensively being studied in European *Silene* species [63, 64, 65, 101, 102, 103, 104, 105, 156].

Pollinators are not only of importance to plants because they are essential pollen vectors for outcrossing species, but also because they may transmit pathogens while visiting flowers. An intensively studied example of such a vector-transmitted disease is the anther smut fungus *Microbotryum violaceum*. This fungus is an obligate parasite on a wide range of Caryophyllaceae and has been recorded on 92 species in Europe and on 21 species in North America [216]. Interactions between the fungus, its hosts, and the insect vector provide a fascinating arena for evolutionary processes [20, 21, 49].

## **REPRODUCTIVE SYSTEMS WITHIN THE GENUS *SILENE***

### **Overview of Reproductive Systems in Plants**

In plants, in order to prevent harmful effects of inbreeding depression, the species can be subdivided into two basic groups: either self-pollinating (autogamic) to eliminate harmful mutations from the population, or preventing autogamy and ensuring cross-pollination (allogamy) by different mechanisms, to keep harmful mutations in heterozygous state [15]. In sexually monomorphic (cosexual) species, the mechanisms involve dichogamy (differences in the timing of pollen dispersal from anthers and stigma receptivity), or different types of autoincompatibility [16]. Another mechanism is to produce unisexual flowers, either on the same (monoecy) or on different plants. The plant populations can either remain sexually monomorphic because all plants in the population are cosexual with unisexual flowers, or become sexually polymorphic because

plants of different sexual types are formed. The sexually polymorphic populations are either dioecious (exclusively males or females), gynodioecious (females and hermaphrodites), androdioecious (males and hermaphrodites), or trioecious or subdioecious (males, females and hermaphrodites).

### **Gynodioecy**

Gynodioecy, in 7.5% of European angiosperm species, is the second most frequent reproductive system after hermaphroditism present in 72% of species [182]. As females in a gynodioecious population lose 50% of their fitness, mechanisms keeping them in the population [46] include obligate outcrossing of females with elimination of inbreeding depression, and compensatory resource reallocation following the loss of the male sexual function (reviewed in [43]).

Theoretically, a nuclear recessive mutation causing male sterility can be maintained in a population if the females display at least two-fold seed set than the hermaphrodites to compensate the transmission of the dominant male-fertile allele to the next generation by the hermaphroditic plants [35]. However, the most frequent cause of the gynodioecy in natural populations is cytoplasmic male sterility (CMS [31]) genes which are located in the mitochondria. The sterility is usually maternally inherited [178] and can be maintained in the population without any assumption of female advantage in the seed set. Moreover, when females carrying the CMS gene have even a slight advantage over the hermaphrodites, the CMS can invade the population [127].

However, when a dominant nuclear allele restoring the original male fertility (a restorer of fertility; Rf) appears in the population, it will be positively selected, leading to the rapid spread of the Rf alleles and subsequently to its fixation in the population, and thus restitution of the hermaphroditic nature of the population. Evidence of the fixed Rf alleles was described in the crop plants, where CMS was recovered in the crosses between different lines or species [190]. Also, the mitochondrial genome of the hermaphroditic plant *Arabidopsis thaliana* contains several potential CMS genes [135]. As gynodioecy based on the combination of the cytoplasmic male sterility and restorers of fertility is relatively frequent [40], the studied populations either did not reach equilibrium and/or there exist other selective forces involved in the maintenance of the females in the population [145]. These forces include higher reproductive

fitness of females because of reallocation of resources from pollen to seeds [215] and/or decreased fitness of plants bearing restorer of fertility when a corresponding sterile cytoplasm is lacking [40]. Both the higher seed set in females, when compared to hermaphrodites [215], and the cost of restoration [10, 86] were demonstrated in several gynodioecious species.

At the histological level, the first anomalies in plants carrying sterile cytoplasm appear in the tapetal cells in a majority of cases, subsequently followed by abortion of microspores [31]. Concerning the genetic point of view, very often more different CMS genes coexist with several different Rf alleles [134] and the interactions between the different CMS and Rf genes seem to be complex. The original theoretical description assuming one dominant restorer of fertility acting on one CMS gene [40] seems to be thus insufficient and several new models (involving more parameters, including, for example, gene interactions or polygenic nature of the restorers [11, 42, 67, 185]) are proposed.

Genes responsible for the CMS have been found in several species [190]. Their common feature is that they code for fusion open reading frames (ORFs), as at least a part of the CMS-associated gene is related to known mitochondrial genes [31]. Recently, Rf genes from several species have been cloned, and most of them code for pentatricopeptide repeat proteins [18, 30, 113, 223]. However, the exact mechanism(s) of CMS is still unclear.

In addition to the CMS/Rf system, the gynodioecy also evolved several times independently by disruption of gametophytically controlled self-incompatibility by polyploidization. Due to strong inbreeding depression in the resulting self-compatible species, selection for outbreeding appeared, and male-sterile mutants invaded the population [151].

### ***Androdioecy***

Theoretical models indicate that the androdioecy evolution is more likely from dioecy than from hermaphroditism [234]. To date, most of the well-described examples of androdioecy really evolved from the dioecy [114, 171, 227, 236].

The rarity of the androdioecy as a sex determination system in plants indicates that the conditions for the invasion of female-sterile male plants in otherwise hermaphroditic population are very difficult. This was also

supported by mathematical models, which imply that the androdioecy may be maintained only if males effectively disperse more than twice as much pollen as hermaphrodites [35, 129]. The models also predict that the most favourable conditions for the invasion of males into otherwise hermaphroditic population occurs when this population is fully outcrossing (i.e., males are always the minority phenotype [41]).

### ***Trioeey or Subdioecy***

Trioeey is considered as an extreme form of gynodioecy at a late stage in the evolutionary transition to dioecy [16]. The trioeious populations are often defined as populations where at least some of the pollen-producing individuals produce seeds (inconstant males), but the females exhibit constant sex expression because they never produce pollen. In many trioeious species that evolved from gynodioecy, males express plasticity in the fruit production depending on the availability of resources (gender plasticity [59]). The gender plasticity could stabilize trioeey and hinder the evolution to dioecy, provided that seed production by inconstant males contributes to their overall fitness [15, 56, 59].

To determine which sex is heterogametic in several trioeious species, the phenomenon of gender plasticity was used [214, 231]. Self-fertilization of the inconstant males led in different species very often to the 3:1 or 2:1 ratio of males to females, which can be explained by the presence of one allele of a dominant suppresser of femaleness in the inconstant males, and the ratio depends on whether homozygous constitution of the suppresser of femaleness is viable or not.

### ***Dioecy***

Dioecy is present in 6% of the flowering plant species spread over 43% of families [179]. Its taxonomic distribution reveals that dioecy arose multiple times independently [43].

Genetic model of the evolution of dioecy from cosexuality suggests that genetic changes in at least two genetically linked genes are necessary to establish dioecy [35], one causing loss of the male function (and thus producing females) and the other causing loss of the female function (and thus creating males). The existence of inconstant males (but not females) in many trioeious species [59] led to the conclusion that a major recessive mutation leading to the formation of a gynodioecious population appears in the first step. In the arising gynodioecious

population, the functional gender of the hermaphroditic individuals is influenced by the frequency of females [57], and thus with the increasing frequency of females, they become functionally more male. As a result, there is a selective pressure on the hermaphrodites to invest more resources in the pollen production than to the ovules, thereby evolving a greater male bias [35]. The evolving modifier genes, which make the hermaphrodites more male, are also expected to reduce female fertility, unless they are sex-linked in their expression, and thus partial female-sterility factors are most likely to spread in a gynodioecious population if they are linked to the male-sterility gene [35, 161]. This step is followed by a selection for a tight linkage between the male-sterility locus and modifier loci because the recombinants harbouring the male-sterility allele (causing loss of the male functions) and modifier alleles (causing decrease of the female functions) would be fully male- and at least partially female-sterile. As many cases of dioecy have evolved from gynodioecy, at least the first step of the model seems to be relatively widespread.

However, also other ways to reach dioecy were also described. Several genera evolved dioecy using an avenue from self-incompatibility through polyploidization followed by gynodioecy and dioecy [151]. Another possible route is from monoecy [231]. If there exists an ability of a species to suppress male and female function in different flowers, the floral sex ratios between different plants can be subsequently adjusted to form male and female individuals [128, 143]. As a strong association of dioecy with monoecy was already found [179], this model also seems to be relevant. Moreover, in the genus *Acer*, the dioecy evolved from androdioecy [82]; it is the only known documented case of dioecy, which evolved via androdioecy.

Similarly to trioeccious species, most of the dioecious species have heterozygous male sex [222]. In some species, sex is determined by a simple mendelian genetic system based on the segregation of one locus, as in squirting cucumber (*Ecballium elaterium* [79]), or a few loci, as in annual mercury (*Mercurialis annua* [66]), whereas in other species, such as sorrel (*Rumex acetosa* [172]) or white campion (*Silene latifolia* [23, 232]), sex chromosomes evolved. Although the genetic basis of sex determination system in many species is known, available data on the molecular mechanisms of sex determination are rather limited.

## Reproductive Systems in the Genus *Silene*

The reproductive systems in the genus *Silene* range from hermaphroditism through many possible intermediate stages (gynomonoecy, gynodioecy, andromonoecy and trioecy) to the dioecy [61, 102].

### Gynodioecy

The most studied gynodioecious *Silene* species is *S. vulgaris* (bladder campion). It is a short-lived perennial plant native to Europe, which was introduced into northeastern and central North America after Europeans colonized the New World [67, 146], but probably before 1785 when its presence was for the first time described in this area [12, 52].

In the gynodioecy research in *S. vulgaris* Charlesworth and Laporte [42] analyzed natural populations in England based on comparisons of expected sex ratios in different inheritance models with observed ratios for various crosses and reciprocal crosses. Inheritance models expected strictly maternal inheritance of CMS genes and biparental transmission of nuclear male fertility restorers to the progeny. The comparative analysis of reciprocal crosses has shown presence of at least two, and more likely three, different CMS genes. The analysis of inheritance models for dominant or recessive nuclear fertility restorers had a less dependable outcome, but the results indicated that at least one recessive and several dominant nuclear loci are responsible for male fertility restoration. Similar results were obtained from North American populations of *S. vulgaris* [212].

Taylor et al. [212] studied for the first time the correlation between the among-population variance of mitochondrial DNA (mtDNA) haplotypes tested by molecular markers [164] in *S. vulgaris* with the gender variance. The conclusion that the mtDNA haplotypes differ in their likelihood of CMS restorability, and thus there should exist an association between the mtDNA haplotype and gender, resulted in the search for this association [146, 165, 204]. Olson and McCauley [165] assessed the population structure of mtDNA haplotypes in 250 individuals from 18 North American populations containing 13 mtDNA haplotypes, and found a strong statistical association between haplotypes and gender for several mtDNA haplotypes. When compared with North American populations, the overall genetic diversity and within-population diversity of mtDNA in Central European populations were much higher (20 mtDNA haplotypes found among 86 individuals from 5

populations). Most of the haplotypes were private, i.e., found only in one population [204]. Moreover, the association between mtDNA haplotypes and gender was not as clear as in the North American populations. These phenomena are explained by much larger population sizes in European populations when compared to the North American ones and/or by lower levels of disturbance in the European localities. However, it seems that the overall haplotype numbers have not been strongly decreased by the colonization process because the regional level mtDNA diversity of Central European populations was comparable to that found in western Virginia. A similar level of cpDNA (chloroplast DNA) restriction polymorphisms found in Virginian [147] and Central European populations [204] also supports these results.

*S. vulgaris* populations are characterized by a broad variation in sex ratios in different natural populations, with the proportion of females varying from 0 to 78% [42, 164, 204]. This sex ratio variation was found not only in ancestral European populations [42, 204], but also in large amounts in populations descending from plants introduced in North America [164]. This variation is most probably caused by different fitness of female and hermaphroditic individuals in various populations, including inbreeding depression in hermaphroditic individuals [12, 81], sex ratio and fitness within populations [145], or fine scale spatial structuring of local populations [167]. Higher inbreeding depression in hermaphrodites, when compared to females, is one of the hypotheses explaining the maintenance of females in gynodioecious populations, because females, contrary to hermaphrodites, are obligatory outcrossing. It is expected that the less inbred (and thus more fit) female plants produce more seeds or higher quality seeds. In *S. vulgaris*, inbreeding depression following self-fertilization was observed both in North American [12] and European populations [81]. Variability of sex ratio within population theoretically forms a negative feedback loop together with pollen limitation [166]. The expected higher production of seeds or higher quality seeds in females, when compared to hermaphrodites [215], should lead to the spread of CMS factors. When sex ratio becomes sufficiently female biased, pollen limitation is expected to inhibit the seed production of females and thus to slow the invasion of CMS factors [166]. In experimental populations of *S. vulgaris*, McCauley and Brock [144] studied the effect of sex ratio in population on fruit set and seeds per fruit by manipulation of sex ratio. It was established that with the increasing frequency of hermaphrodites in population, fruits to flower

ratios and production of seeds per female also increase. Similar results were obtained in natural populations where the relative fitness of females (measured as relative seed production and seed germination) decreased with the rarity of hermaphrodites in the population [145]. The fitness decrease was explained by the fact that females suffer from pollen limitation. This assumption is based on the work of Taylor et al. [210] where single females were placed at varying distances from the source population, which included hermaphrodites. Local variation in the sex ratio most likely influences the relative fitness of females and hermaphrodites by the reduction of females when they are clustered. This clustering can be generated in natural populations by the fact that seeds generally fall close to their mothers [126] and that *S. vulgaris* females generally produce more female progeny [212]. This idea was subsequently confirmed by finding that fine scale spatial structuring of both sex and mtDNA haplotypes significantly decreases female fruit production relative to hermaphrodites [167].

Very interesting results were obtained during the search for association between mtDNA and cpDNA haplotypes. This association was expected because of maternal co-inheritance of mtDNA and cpDNA, which has been described in many plant species [152, 178]. Although extremely strong association between mtDNA and cpDNA haplotypes was found in one *S. vulgaris* North American population [146], this association could not be so distinctly defined in Central European populations [204]. Houlston and Olson [96] described clear incongruence between mtDNA and cpDNA haplotypes, strongly suggesting that cpDNA and mtDNA in *S. vulgaris* are not always co-inherited. The most plausible explanation is based on the phenomenon of heteroplasmy, i.e., coexistence of both maternal and paternal mtDNA and/or cpDNA within one individual. Indirect evidence for occasional heteroplasmy of CMS factors in *S. vulgaris* was already described, because different offspring ratios produced by female and hermaphrodite flowers on the same individual were found [5]. McCauley et al. [148] published a complex analysis of mtDNA inheritance in *S. vulgaris* from 23 families where they have identified 6 haplotypes based on the presence or absence of *AhuI* and *MspI* restriction sites within *atpA* gene. Only one family (which contained five haplotypes, therefore, it could not have the effect of a haplotype) had strictly maternal inheritance of mtDNA. In four families, at least one individual was identified in each family which did not resemble the mother's haplotype and they also found several families

which resembled paternal haplotypes rather than maternal. It proved non-maternal inheritance of mitochondrial genome and also occurrence of heteroplasmy in *S. vulgaris*. A study performed on *S. vulgaris* natural populations [226] confirmed these results using a highly sensitive “knock back” method combined with quantitative PCR (polymerase chain reaction) to reveal cases of “cryptic heteroplasmy” where the level of the putative secondary haplotype was too low. According to the presented results, heteroplasmy in *S. vulgaris* seems to be much more widespread than expected previously, with the mitochondrial variation within individuals reaching in one population 26% of the total. It seems that the heteroplasmy in *S. vulgaris* may also influence *de novo* formation of CMS factors, as evidences for interlocus and intralocus recombination in mitochondrial genomes were described [96, 148]. It is, therefore, possible that sex determination in *S. vulgaris* is less complicated than expected by Charlesworth and Laporte [42] and Taylor et al. [212] as the interpretations were in both cases based on expectation of strictly maternal transmission of CMS factors. The work of Houlston and Olson [96] revealed yet another interesting point. Sequencing of two mtDNA genes in several *S. vulgaris* populations and their subsequent phylogenetic analysis showed very long internal branches and short external branches, which is explained by the maintenance of divergent haplotypes over a long period of time due to balancing selection. Hitchhiking effects of CMS factor(s) present in the mitochondrial genome in all probability preserved the allelic variation of mitochondrial haplotypes.

### **Trioecy**

Trioecious species are within the genus *Silene* relatively rare, involving *S. acaulis* [112], *S. dichotoma* [50] and *S. otites* [50]. The breeding system in *S. acaulis* is highly variable; different populations were described as dioecious [61], trioecious [89, 142, 173], and gynodioecious [57, 154, 196]. In the gynodioecious populations, plant gender seems to be under a cytonuclear control [108]. Precise information on the variability of reproductive systems in *S. dichotoma* and *S. otites* is missing.

### **Dioecy**

Dioecy evolved in the genus *Silene* at least three times independently, as can be inferred from its phylogenetic distribution [61]. Dioecious species fall into three distantly related sections *Oties* (*S. otites* and *S. pseudotites*),

*Nanosilene* (*S. acaulis* ssp. *bryoides*), and *Elisanthe* (*S. latifolia*, *S. dioica*, *S. diclinis*, *S. marizii* and *S. heuffelii*). All the well-studied dioecious *Silene* species belong to the section *Elisanthe*. Gender in these species is controlled by the XX/XY sex determination system, where the Y-chromosome carries at least two independent functions: one suppresses gynoecium development [231] whereas the other promotes anther formation [70].

## ***Silene latifolia* as a Model Plant**

### **Sex Bias in *S. latifolia***

The study of the sex ratio bias is connected with the genetic studies in *Silene* [50]. The quantity of pollen grains used for pollination was shown to influence the sex ratio proving the male heterogamety in *S. latifolia* and *S. dioica*. Further experiments indicated possible differences in the rate of the pollen tube growth between X- and Y-carrying gametes [50]. Carroll and Mulcahy [34], however, showed that these results cannot be generalized and that these results were probably influenced by the use of interspecific hybrids *S. latifolia* x *S. dioica*.

The important progress in the study of the genetic determination of the sex ratio in *S. latifolia* is connected with the work of Taylor [206] who surveyed maternal families from natural population as source plants genetically variable for sex ratio controlling genes. Crosses among plants from the most-female biased and the most-male biased populations showed that the sex ratio polymorphism was inherited through or expressed in the male parent. The inheritance of sex ratio investigated using a reciprocal crossing design showed a significant difference in the reciprocal crosses. The sex ratios produced by males generally resembled those of their male parents, indicating that the sex ratio modifier was Y-linked. The maternal parent also significantly influenced sex ratio, but only through an interaction with the genotype of the paternal parent. Sex ratio, therefore, is apparently controlled by several loci; the most probable model involves an interaction between loci that cause the female bias and a Y-linked locus that enhances the proportion of males in the progeny, i.e., that some Y-linked alleles enhance the production of males in female-biased populations, i.e., they are restorers. This model was further supported by the work done on the interspecific hybrids between *S. latifolia* and *S. dioica*. Sex ratio was paternally inherited (i.e., Y-linked),

but was also strongly influenced by the maternal parent through an interaction with the Y-chromosome. These results corroborated previous work on the inheritance of sex ratio within *S. latifolia*, and suggest that the sex ratio in *S. latifolia* and *S. dioica* has a similar genetic basis [207]. Examination of the maternal by Y-chromosome interaction revealed that the Y-chromosome of each species produced a more severe female bias in crosses with females of the opposite species. This was consistent with the hypothesis that the alleles expressed in the maternal parent cause the female bias, while the Y-linked alleles tend to restore sex ratio towards equality.

Later work, relying on the PCR genotyping of the fertilized ovules (three days after pollination) supported the view that the sex bias is realized during the pollen tube growth or early seed development [208]. This indicates that there could be a different viability of the X- and Y-carrying pollen grains, a competition between X- and Y-carrying gametes [50] or a different rate of survival of the male and female embryos in the very early phase of development.

In the extensive population study done by Taylor [209], two males and two females (differing in the sex ratio in their progenies) were used as tester plants to test 142 plants coming from eight natural American populations for the presence of the sex ratio modifiers. The variation among maternal parents explained little or no sex ratio variation among natural populations, but it was possible to explain about 80% variation in natural populations by variation among the paternal parents. The males showed relatively often “sex ratio bias phenotype” (0-20% of males in the progeny). By itself it does not mean that variations among maternal parents or maternal x paternal interactions do not exist but it means that the among-population variability in these alleles is relatively small. There was no discontinuous variation in the sex ratio among maternal parents as might be expected if the sex ratio distortion were under maternal control. There was, however, a discontinuous sex ratio variation among paternal parents. This could remind other systems involving sex chromosome meiotic drive as meiotic drive associated with the wild-type X-chromosomes of *Drosophila melanogaster* [51] and Y-associated meiotic drive in mosquitoes [92], which, however, leads to the excess of males in the progeny.

In the search for the putative mechanism of this segregation distortion in *S. latifolia* [213], the authors tested a size distribution of

pollen grains in the normal males and in the “sex ratio” males obtained from previous experiments. They found that “sex ratio” males show, in contrast to normal males, bimodal distribution of the pollen grain size. The smaller group of the pollen grains was later shown to be sterile. Thus, the sex ratio distortion in *S. latifolia* was probably accomplished by the prevalent abortion of the Y-linked gametes. This interpretation needs, however, yet more experimental support to be accepted. It was not yet shown that the aborted pollen grains are indeed pollen grains carrying the Y-chromosome, and some males also show decreased pollen fertility but no sex bias. The fact that no relationship exists between the size of pollen grains and sex ratio in the work of Carroll and Mulcahy [33] can be explained so that it is not the Y-chromosome itself which causes the small pollen grain size and the tendency to abortion, but it is the Y-linked “sex ratio” allele. That it is not just a Y-linked lethal allele, which causes sex ratio bias in *S. latifolia*, can be concluded from the competition experiments of Taylor et al. [211]. If the pollen of the sex biased male was used in an unlimited amount, it produced a similar amount of seeds as wild-type pollen (but with excess of females). In the competition with the wild-type pollen, the “sex ratio” pollen showed fewer offspring of both sexes. It is, therefore, possible to speculate that there is some X-linked allele in these males, which is expressed during the anther development that negatively influences performance of both X and Y bearing pollen grains, but its influence on Y-carrying gametes is more intense. Simultaneously, the observed effect of the female parents [206, 207], especially in the interspecific crosses [207], should be kept in mind.

Apart from the above discussed mechanisms, the sex ratio in natural populations can also be influenced by other factors, such as different male and female seed germination [133], and an increased plant density which seems to increase female bias due to different survivorship of males and females [131].

### ***Sex Organ Differentiation in Silene latifolia***

The flowers of the dioecious *Silene* species are potentially hermaphroditic in the early stages of flower development because they contain rudiments of organs of the opposite sex. It is, thus, possible to expect that the dioecious species evolved from the hermaphroditic species or from species bearing CMS [61]. As *S. latifolia*, is the most widely studied *Silene* species, it forms the subject matter of this chapter.

In the early stages of flower formation, the meristems of male and female flowers are identical [122]. In the female flowers, the stamens develop normally to the transition from the bilobal to the tetralobal stage, when sporogenous cells are formed. At this point, the anther development is stopped before the formation of parietal layers and later on the anther rudiments dedifferentiate [69, 70]. The anther development arrest can be thus explained as a loss-of-function mutation in some gene necessary for the progression from the relatively undifferentiated stage to the formation of the primary parietal layer. Contrary to the stamen rudiments, the gynoecium rudiment present in male flowers is a filamentous structure without any apparent anatomical relationship to the gynoecium [69, 70]. It is thus difficult to foretell which developmental pathway participates in the formation of the filament in the place of the gynoecium. The situation is complicated by the fact that the gynoecium formation in male plants is suppressed actively [122]. The gynoecium suppressor can be thus a “natural” suppressor of the gynoecium size, which in the male flowers is expressed in the place of gynoecium. An alternative (and also plausible) explanation is that the gynoecium suppressor represses some other developmental pathway(s) in female plants, but its copy present on the Y-chromosome is expressed in the fourth whorl and thus causes repression of the gynoecium formation.

Knowledge of the basic mechanism of sex determination is one of the biggest advantages of *S. latifolia* compared to other dioecious model species. Westergaard [229] in experiments with deletion mutants (which spontaneously appeared in his polyploid experimental populations) revealed two main functional regions of the Y-chromosome: a female organ suppressing region, and a region important for male fertility. He was also able to predict the existence and approximate position of the stamen-promoting region. The Y-deletion mutants with the early arrest of stamen development (asexuals) were subsequently independently found by Donnison et al. [62] and Farbos et al. [70]. Increased number of Y-chromosome specific molecular markers [8, 55, 62, 74, 94, 123, 140, 153, 157, 159, 162, 237] together with a broad collection of Y-chromosome deletion mutants [62] enabled localization of the functional regions on the Y-chromosome using the deletion map. Deletion mapping revealed that the Y-chromosome p-arm carries the female organ suppressing region and the stamen-promoting region, whereas the stamen promoting region is present on the q-arm [123, 238].

Although the sex determination system in dioecious *Silene* species possessing heteromorphic sex chromosomes is stable (based on the active role of the Y-chromosome), the sex expression can be in experimental populations or even in the natural populations modulated by genes localized outside the Y-chromosome and by external factors.

### ***Suppression of the Gynoecium Development in Males***

Shull [196] described in his populations of *S. dioica* (without any special treatment) two kinds of hermaphrodites: genetic hermaphrodites (transmitting the hermaphroditism to the next generation), and somatic hermaphrodites (non-transmitting this trait). It appears that genetic hermaphrodites were caused either by an epimutation in the gynoecium suppressor (indicated below; [99, 100]) or by mutation in another gene from this regulatory pathway located in the pseudoautosomal region. In this second case, the gene should be haplo-insufficient or the hermaphroditism is caused by the loss of sensitivity of the gynoecium forming gene(s) to the inhibitory effect of the gynoecium suppressor. Existence of somatic hermaphrodites could be easily explained by the presence of an X-linked (epi-) mutant allele. The study of female plants from the progeny of the somatic hermaphrodites was, however, not performed in this case [194, 195].

Such results were later obtained by P. Hertwig and G. Hertwig [90] as in their populations the hermaphroditism was transmitted through the female offspring. Based on the current knowledge of the basic mechanism of sex determination in *S. latifolia*, it is possible to interpret the observed mode of inheritance by the following scheme: Hermaphroditism was in this case caused by a mutated (or epimutated) allele of an X-linked gene, which had incomplete penetrance and variable expressivity. Another case of the X-linked inheritance of hermaphroditism was described by Winge [233] in hermaphrodites, which appeared in  $F_2$  generation after crossing of *S. dioica* female with *S. latifolia* male.

Conclusions of Winge suggest the existence of Y-chromosome constitution, which can lead to male plants even in the presence of the X-linked allele controlling hermaphroditism, and possible role of an autosomal inhibitor of hermaphroditism. Role of autosomes in the control of androhermaphroditism in *S. dioica* was also postulated by Van Nigtevecht [160], as the inheritance of androhermaphroditism in his populations could not be easily explained either with X- or Y-chromosome

linkage. Hermaphrodite and androhermaphrodite plants were observed only in wild type populations of *S. dioica* (3 per 10,000 plants) but not in *S. latifolia*.

Modification of the sex expression in *S. latifolia* was, however, possible using the DNA hypomethylation drug 5-azacytidine [99]. Upon the application of the drug on the germinating seeds, a significant number of males (21%) were androhermaphrodites while no apparent phenotypic effect was observed in females. The trait was transmitted to further generations even if the male flower of the androhermaphrodite plant was used as a donor of pollen in the cross with wild-type female (perhaps suggesting that the mosaic pattern of the sex expression does not mean mosaic character of plants from the genetic point of view). Androhermaphroditism was, however, not inherited, if the hermaphrodite plant was used as a seed parent. The next research [100] showed that androhermaphroditism is inherited as an allele closely linked with the Y-chromosome and that the Y-chromosome was not inherited through the female gametic line in the androhermaphrodite plants. From these results, it can be concluded that the most probable mechanism of the 5-azacytidine action on the sex expression in *S. latifolia* is an epigenetic inactivation of the gynoecium suppressor. An analogous example can be found in *Arabidopsis thaliana*, where over-all hypomethylation of the genome leads to the hypermethylation and transcriptional inactivation of the SUPERMAN gene [98], which is involved in the delineation of boundaries between the third and fourth flower whorl, and its mutations and epimutations usually lead to the increased number of stamens. The other epimutants display increased number of carpels [183]. In some cases the boundary can be completely abolished which leads to a mixed identity of male and females organs. According to the observed phenotypes, it is not likely that SUPERMAN could play the role of the gynoecium suppression. However, the fact that penetrance and expressivity of androhermaphroditism were increased in the subsequent generations supports the view that gynoecium suppressor is influenced by a pathway similar to the SUPERMAN gene.

Another case of the modulation of the gynoecium suppressor action was described by Lardon et al. [119]. Apart from a large number of the hermaphrodites carrying deletion of the gynoecium suppressor, they isolated in their mutation screen the plant which was androhermaphrodite and in which the androhermaphroditism was caused by a defect in a certain autosomal region. The locus behaved as haplo-

insufficient, as ability of the suppression of gynoecium development was concerned because heterozygosity was sufficient to promote androhermaphroditism.

The role of gene dosage in the modulation of gene expression remains an open question. Both Warmke [230] and Westergaard [231] support a hypothesis that the X/Y-chromosome ratio can modulate the sex expression in *S. latifolia*. However, the differences between experimental populations probably caused different conclusion concerning the role of autosomes. While in the population of Warmke, the increased X/Y ratio itself was able to promote development of gynoecium in the plants possessing the Y-chromosome, in the population of Westergaard, a change in the internal balance of autosomal sets was necessary to promote hermaphroditism [229]. Warmke [224] found a plant which possessed, in addition to the intact X-and Y-chromosomes, an X-chromosome with one deleted arm (the one with the pseudoautosomal region). He concluded that the differential arm of the X-chromosome carries a female promoting function, which, if present in an increased dosage, can escape the inhibitory signal of the gynoecium suppressor. It is possible to speculate that this (these) gene(s) diverged during the evolution of the *S. dioica* and *S. latifolia* species and so the Y-chromosome coming from *S. latifolia* in the population of Winge was not able to inhibit this X-linked gene (or genes) coming from *S. dioica*. A quantitative nature of the control of the gynoecium development in the Y possessing plants can also explain variable expressivity and incomplete penetrance of androhermaphroditism in some experiments [90, 100]. Surprisingly, no hormonal treatments could provoke gynoecium development in males indicating that plant hormones do not play a key role in the process of gynoecium suppression.

Since the filament, which appears in the middle of the male flower, does not resemble gynoecium, it was hypothesized that the filament lacks the gynoecium identity. This hypothesis was tested by Hardenack et al. [87] who cloned orthologues of floral homeotic genes from *S. latifolia*. RNA/RNA *in situ* hybridization analyses revealed expression patterns similar to those observed in hermaphroditic species. The gynoecium suppression sex determining gene(s) thus act downstream from or independently on the action of the floral homeotic genes.

A study of the cell division pattern during the early stages of gynoecium primordia formation revealed that the onset of cell division in the fourth floral whorl is significantly delayed in males when compared to

females [141]. It suggests the necessity of the existence of a mechanism delaying cell division onset in the gynoecium primordium in males; this mechanism might be caused by a transient expression of the gene *SICUC* in the centre of the floral meristem in male flower buds before the onset of cell division [240]. The gene *SICUC* is a *S. latifolia* orthologue of the genes NO APICAL MERISTEM in *Petunia hybrida* [203], CUP-SHAPED COTYLEDONS 1 and 2 in *Arabidopsis thaliana* [3] and CUPULIFORMIS in *Antirrhinum majus* [225], which are known to be involved in the establishment and maintenance of organ boundaries in the shoot apical meristem, the inflorescence meristem and the floral meristem [2, 205]. Since in *Arabidopsis thaliana*, the expression of *CUC2* is correlated with loss of the cell division activity [29], the temporal expression of *SICUC* in the fourth whorl in males can cause the observed delay of the cell division onset.

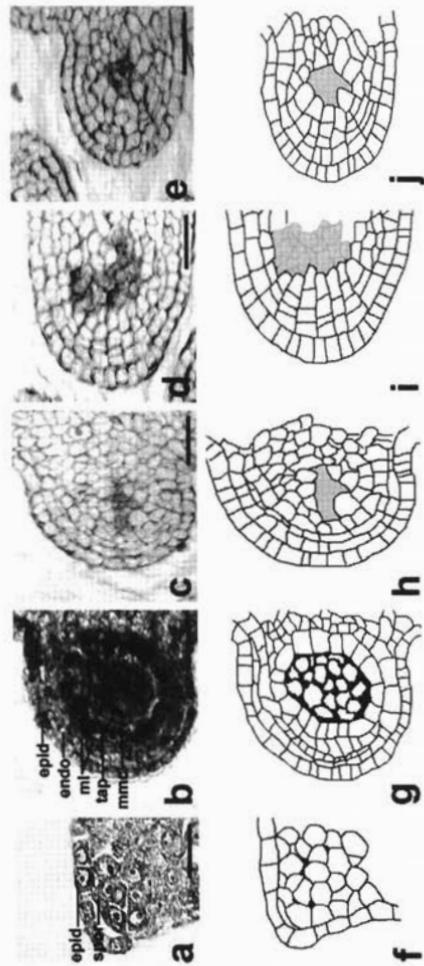
With an increased number of known mechanisms involved in the gynoecium formation, a new theory concerning the formation of filament in the place of the gynoecium has appeared: the filament can appear by a premature loss of meristem identity in the centre of the flower meristem. This can be caused, for example, by a premature loss of expression of the gene SHOOTMERISTEMLESS, which is necessary for the maintenance of undifferentiated cells in the shoot meristem and for proper proliferation of cells in the floral meristem [17, 48, 68, 130]. This hypothesis was tested by Zluvova et al. [240] who found a premature disappearance of *S. latifolia* orthologues of SHOOTMERISTEMLESS (*SISTM1, 2*) in male flower buds. This result conforms to the observed pattern of *SICUC* expression, and it is likely that the disappearance of *SISTM1, 2* expression leads to the expansion of *SICUC* in the centre of the male flower meristem. Although neither *SICUC* nor *SISTM1, 2* are located on the Y-chromosome, these results identified a regulatory pathway controlled by the gynoecium suppressing sex-determining gene.

### ***Control of the Stamen Development and Pollen Fertility***

In contrast to “active” gynoecium suppression in male plants, the suppression of the anther development in female plants is “passive”—based on the absence of stamen promoting genes in females. This model is supported by results of the experiments carried out on the hybrid plants obtained by cross of *S. latifolia* female with hermaphrodite species *S. viscosa*. Autosome coming from the hermaphrodite species *S. viscosa* was

able to compensate the lack of the stamen promoting region(s) in the hybrid plants [239]. The exact number of genes lacking in females is not known, but it is certain that at least in some of them their absence can also be compensated by external stimuli. Although stamen development in *S. latifolia* females is insensitive to exogenous application of phytohormones [91, 184], there exists a natural mechanism modulating the sex expression. When female plants are infected with the parasitic smut *Microbotryum violaceum* (Pers) G. Deml and Oberwinkler (formerly *Ustilago violacea*), the rudimentary stamens (Fig. 3a, schematically Fig. 3f) develop into stamens morphologically resembling mature stamens in males [77, 217]. However, the induced anthers harbour teliospores instead of pollen grains [6]. Histological analysis revealed that anther development of healthy male plants is characterized by periclinal division followed by anticlinal cell division in each layer. This process results in regular anther layer formation ([69]; Fig. 3b, schematically Fig. 3g). In the *M. violaceum*-infected *S. latifolia* male plants, a similar pattern of cell division was occasionally observed ([217]; Fig. 3c, schematically Fig. 3h) and these anthers are thus able to form tapetum and pollen grains. However, most of the anther samples displayed only multiplied periclinal cell divisions, resulting in centrifugally arranged chains of cells (Fig. 3d, schematically Fig. 3i). A similar periclinal cell division pattern (Fig. 3e, schematically Fig. 3j) was also observed in *M. violaceum*-induced anthers in female plants. The resulting female anthers thus contain little or no tapetal tissue and no pollen grains [9, 217, 218]. Thus, the stamen formation in the infected females is restored only partially. Several genes that normally have male-specific expression are induced in infected females [191], whereas tapetum- and pollen grains-specific genes are not induced [1].

Development of stamens in female plants is also promoted by the  $\text{Ag}_2\text{S}_2\text{O}_3$  and  $\text{AgNO}_3$  treatment [121]. Although the induced stamens are not as well-developed as those of male flowers, anther development proceeds further beyond the anthers induced by *M. violaceum* infection, and tapetum and non-viable pollen grains are formed. As the treatment of Y-chromosome deletion mutants lacking the stamen-promoting sex determining gene did not lead to the restoration of pollen fertility, it is possible that either the treatment stimulates anther development by a parallel pathway that is independent of the Y-linked sex-determination mechanism, or that other Y-linked genes necessary for the anther development were co-deleted with the sex-determining gene in the mutants.



**Fig. 3.** Histological analysis of anther development in healthy *S. latifolia* female plant (**a**, **f**), male plant (**b**, **g**), *M. violaceum* infected male plant (**c**, **d**, **h**, **i**), and *M. violaceum* infected female plant (**e**, **j**). Light microscopic pictures are given in the first line (**a** - **e**), schematic representations according to the pictures are given in the second line (**f** - **j**). The shaded areas illustrate the regions damaged by the infection. Bars represent 10  $\mu\text{m}$  (**a**) and 50  $\mu\text{m}$  (**b** - **e**). (**a**) In the *S. latifolia* healthy female plants, cell division stops before the formation of the parietal layers, just the epidermis (epid) and sponogenous cells (spor) are formed. (**b**) In the *S. latifolia* healthy male plants, all other layers are regularly formed: epidermis (epid), endothecium (endo), two middle layers (ml), and tapetum (tap). The figure shows the anther at the time of microspore mother cells (mmc) formation. (**c**) The *M. violaceum*-infected *S. latifolia* male; anther showing a healthy male-like pattern of cell division. (**d**) The *M. violaceum*-infected *S. latifolia* male; anther not displaying regular anther layer formation. (**e**) Schematic outline of the figure (**a**): anther rudiments in the female plants stop their development at very early stage. (**g**) Schematic outline of the figure (**b**): anthers in the healthy male plants form five easily distinguishable layers. (**h**) Schematic outline of the figure (**c**): anthers in the infected male plants form cell layers similar to the healthy plants. (**i**) Schematic outline of the figure (**d**): anthers in the infected male plants form centrifugally-arranged chains of cells. (**j**) Schematic outline of the figure (**e**): anthers in the infected female plants form centrifugally-arranged chains of cells.

The fact that the restored stamen development was never perfect, strongly suggests that some specific genes are present on the Y-chromosome, which is absent in females. In the hybrid *S. latifolia* x *S. viscosa* (lacking the Y-chromosome), the defects in the tapetum development were observed [239], which seems to reminiscence the cases of CMS found in crosses between different lines or species [190]. Moreover, the nucleo-cytoplasmic male sterility coupled with gynodioecy, seems to be relatively common in the genus *Silene*. Histological defects observed in the interspecific hybrid are very similar to the defects observed in Y-chromosome deletion mutants [239]. It is thus possible to speculate that the Y-chromosome evolution in the section *Elisanthe* started by the evolution of a restorer of fertility on a progenitor of sex chromosomes and that the current sex determining mechanism leading to the earlier anther arrest appeared later.

The better understanding of the mechanisms controlling stamen development in *S. latifolia* can be enabled by candidate gene approach. The anther rudiments in females strongly resemble anthers in the *Arabidopsis thaliana* mutant *sporocyteless/nozzle* [189, 235], which are arrested at the stage of parietal layer formation, but the sporogenous tissue is formed. The gene codes for a transcription factor involved in sporocyte formation in both anthers and ovules (mutants are both male- and female-sterile), and thus the gene cannot be the male-promoting sex determining gene. However, the similarity of phenotypes suggests that the SPOROCYTELESS/NOZZLE signalling pathway can be involved in the sex determination in *S. latifolia* females.

## ***SILENE GENOMICS***

### **General Information**

Karyotypes of most of the diploid *Silene* species contain 12 chromosome pairs ([22, 54]; Fig. 2). In most of the *Silene* species, chromosomes in a karyotype look very similar, and it is thus not possible to distinguish chromosome pairs solely on the basis of arm ratios or centromeric indexes [54]. However, the interspecific differences in the chromosome length and genome size are relatively large [54, 200]. The differences in the contents of nuclear DNA are probably caused by accumulation of repetitive sequences in bigger genomes. An example of this phenomenon was demonstrated in the section *Elisanthe*: the contents of nuclear DNA

of dioecious species *S. latifolia*, *S. dioica* and *S. diclinis* are approximately 1.2-fold bigger than the nuclear DNA content of a closely related hermaphroditic species *S. viscosa* [137]. The dioecious species differ from *S. viscosa* by the presence of a subtelomeric repetitive sequence X43.1, which is among the most abundant repetitive sequences in the genome [32, 80]. Although the repetitive sequence is present in the dioecious species and forms subtelomeric blocks of heterochromatin [32], it is absent from the genome of *S. viscosa* [137]. Interestingly, the nuclear DNA content has been also negatively correlated with flower size in *S. latifolia* [150].

As karyotypes in many *Silene* species are rather homogenous (for example in *S. pendula* or *S. vulgaris*; [54]), it is necessary to find out some markers, to distinguish chromosome pairs. These analyses have already been started in several species using 5S and 45S ribosomal DNA clusters [200]. The preliminary results show that *Silene* species differ considerably both in numbers and localization of rDNA clusters [200].

### ***Silene latifolia* as a Source of Anther-Specifically Expressed Genes**

The arrest of stamen or gynoecium development occurs in *S. latifolia* during very early stages and it is, therefore, a useful tool to be used as a ‘natural mutant’, giving access to molecular events downstream of organ identity genes. The possibility to search for anther-specifically expressed genes was used in several large-scale experiments based on differential screening of cDNA library [138], chemical cross-linking subtraction [14], high stringency subtraction [191] and fluorescent differential display [193], and yielded approximately 50 genes with anther-specific or anther-predominant expression.

### ***Silene latifolia* – Model for Sex Chromosome Evolution**

The most widely studied *Silene* species is *S. latifolia* possessing sex chromosomes [158, 222]. The main advantage of dioecious plants as models for sex chromosome evolution is the recent evolution of dioecy in angiosperms when compared to mammals [44]. *S. latifolia* is used as a model because there are many related *Silene* species without sex chromosomes [61] and thus gene orders in closely related species with and without sex chromosomes can be compared to track the origin of the

sex chromosomes. *S. latifolia* sex chromosomes are estimated to have evolved recently [159]; the degree of Y-chromosome degeneration is not yet known, but it may not be highly degenerated because it is not DNA hypermethylated [198] and histone H4 hypoacetylated [221] when compared to the X-chromosome. Moreover, many sex-linked markers are already known [8, 55, 74, 94, 123, 140, 153, 159, 237].

### **Overview of Sex Chromosome Evolution**

In many taxa, karyotype determines sex. To mediate this developmental decision, sex chromosome pairs have arisen independently among such lineages from separate pairs of ordinary autosomes [45]. The sex chromosomes of one taxon can, therefore, differ phylogenetically and structurally from those of another. Although the sex chromosomes evolved independently many times, the mechanisms of sex chromosome evolution appear to share common features [136].

Sex chromosomes are supposed to evolve from a regular pair of autosomes [163]; for example, traces of the common origin of the human X-and Y-chromosomes can be observed not only throughout their shared pseudoautosomal regions, but also in defined non-recombining regions [39]. The process of sex chromosome evolution started with the appearance of sex-determining gene(s). The presence of sex-determining locus (loci) on the Y-chromosome in taxa with male heterogamety (or Z in taxa with female heterogamety) is expected to serve as an “attractant” for sexually antagonistic genes benefit for the heterogametic sex, which were probably preferentially recruited to the proximity of the sex-determining gene(s) [78]. In the evolutionary old human Y-chromosome, male-specifically-expressed genes were found to be translocated, retrotransposed [117], or present in multiple copies in “ampliconic regions”, mostly with no homology with either X-linked or autosomal genes [201]. To strengthen the linkage disequilibrium between the sex-determining and benefit loci, a recombination arrest has probably occurred (reviewed in [44]).

Due to the recruitment of novel genes to the proximity of the non-recombining region, the non-recombining region is expected to expand along the sex chromosomes [180, 181]. This hypothesis is supported by the fact that the evolutionary divergence of homologous gene pairs on mammalian sex chromosomes increases with distance from the pseudoautosomal region, which suggests that recombination between the

X and nonrecombining parts of the Y-chromosomes was suppressed successively. This pattern has been termed “evolutionary strata” [116, 201].

In parallel with the expected process of accumulation of genes benefit for the heterogametic sex, the heterochromosome probably also underwent genetic erosion [38] as a result of its permanent haploid state. The erosion is expected to lead to the degeneration and loss of genes, which are dispensable or disadvantageous for reproduction of the heterogametic sex [45]. Some of the degenerated copies of proto-sex-chromosome-localized genes are still present in the human Y-chromosome [4]. Combination of both the recruiting of alleles/genes benefit for the heterogametic sex and degeneration of genes dispensable for male reproduction most likely leads to highly increased proportion of testis-specifically expressed genes on the human Y-chromosome [45]. Many of these genes are indispensable for male reproduction as their loss leads to oligospermia or azoospermia, disorders which disallow reproduction under natural conditions [109]. Degenerative processes on the Y chromosome subsequently lead to the loss of a subpopulation of the Y-chromosome and thus to a lowered nucleotide variability of Y-chromosomes [38].

As a consequence of the loss of crossing over on the nonrecombining part of the Y-chromosome, accumulation of both tandem arrays of repetitive sequences and transposable elements is expected [37]. This process seems to be very fast [45]. For example, on the evolutionary very young *Drosophila miranda* neo-Y-chromosome, several families of transposable elements are present much more abundantly than elsewhere in the genome [203].

An evolutionary consequence of Y-chromosome degeneration is dosage compensation of X-linked copies of Y-linked degenerated genes in mammals, which ensures that both sexes are provided with similar levels of gene products [88]. One of the best-known examples of the inactivation of one of the two X-chromosomes in mammalian females is a phenomenon known as Barr body formation or Lyonization [120].

### ***Sex Chromosome Evolution in S. latifolia***

The studied aspects of the sex chromosome evolution in *S. latifolia*, include search for sex chromosome progenitors [75], estimation of X-Y sequence divergence [74, 159], construction of both the recombination

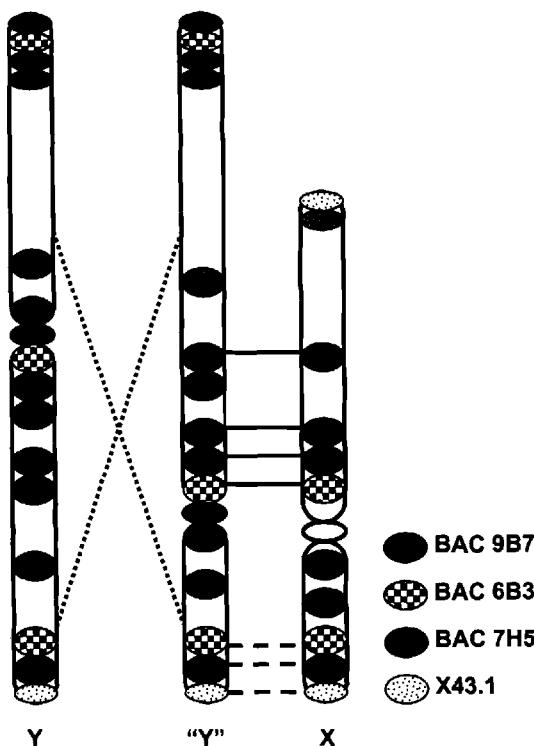
map of the X [75, 159] and deletion map of the Y and comparison of the gene order [123, 238], study of the variability of Y-linked genes [8, 71, 72, 73, 74, 118, 140], search for accumulation of genes with male function on the Y-chromosome [140], study of accumulation of repetitive sequences on the Y-chromosome [95, 107], as well as dosage compensation [198, 199, 221].

### ***Progenitors of S. latifolia Sex Chromosomes***

As nearly all Y-linked genes are also present on the X-chromosome and *vice versa* (the only exception is *SIAP3Y*, which probably originated due to a translocation from an autosome to the Y-chromosome [140]), it is highly probable that the sex chromosomes underwent mostly inversions, but not huge addition of genetic material enlarging their non-homologous parts. This idea was further supported by the comparison of the X-chromosome recombination map [75, 159] with the Y-chromosome deletion map [153, 238], which revealed different gene orders most probably arising from a single chromosomal inversion [238]. This result is further supported by multicolour FISH (fluorescence *in situ* hybridization) mapping of four repetitive markers on the sex chromosomes [125], as it is possible to localize a region on the Y-chromosome which, when inverted, would lead to similar patterns of repetitive sequences on both the X-and Y-chromosomes (Fig. 4). All the genes present in *S. latifolia* X-chromosome are also localized to one linkage group in non-dioecious *S. vulgaris* [75], which further supports the idea that *S. latifolia* sex chromosomes evolved mostly from a single pair of autosomes.

### ***Study of the Y-chromosome Evolution Using the Analysis of the Variability and Dynamics of the Sex linked DNA Sequences***

First research work used the gene *SLX1* and its Y-linked homologue as a model to compare variability of the X and Y-linked sequences [71]. The *SLX1/SIY1* genes encode WD-repeat proteins likely to be involved in cell proliferation [55]. The authors found significantly lower diversity in the *SIY1* in comparison with *SLX1* and tried to understand the possible mechanism causing this lack of diversity. Even the lower effective population size of Y-chromosome linked sequences (about one-third of X-linked sequences) would be sufficient to reduce variability on the Y-chromosome. However, the observed variation in the Y-linked sequence



**Fig. 4.** Schematic presentation of *S. latifolia* sex chromosome FISH pattern according to Lengerova et al. [125]. The current Y- and X-chromosomes are on the left and right, respectively. The "Y" chromosome in the centre stands for a hypothetical Y chromosome before its inversion. The dotted lines connect the hypothetical inversion sites. The solid lines connect the FISH signals where the homology between the sex chromosomes seems to be restored. The dashed lines connect the FISH signals in the parts, which are not positioned within the hypothetical inverted region.

was about 20 times lower than the variability in the X-linked homologue. This difference is significantly higher than expected. Sexual selection was supposed to be improbable to cause such a great effect in dioecious plant. Bottleneck effect (connected with X- and Y- recombination separation) was ruled out because the time (deduced from the divergence of X and Y-copies) should be sufficient to recover the SIY1 diversity. Recent selective sweep eliminating the diversity throughout the species (by strong selection in the neighbouring locus) was also improbable because the existing polymorphisms should then be variants accumulated since the selective sweep and should thus mostly be single singletons. The

favourite explanation of the observed reduction of diversity in *SlY1* was background selection (a selection against deleterious alleles maintained by recurrent mutation). This selection causes a reduction in the amount of linked genetic variability at linked neutral sites [36]. Muller's ratchet (stepwise accumulation of deleterious mutations in an irreversible manner as a result of the absence of recombination) is also considered as a putative mechanism of the difference in divergence between the *SlX1* and *SlY1*, but there is lack of theoretical data concerning this alternative. Gordo et al. [83] found that if only slightly deleterious mutations occur near the studied locus, Muller's ratchet can play an important role in the decrease of the variability of this locus. The next study [8] used *SlX1/SlY1* gene pair and yet another sex linked gene pair (*SlX4/SlY4*) to analyze possible selective constraints. The *SlX4/Y4* genes potentially encode fructose-2,6-bisphosphatases. The comparative molecular analysis of these two gene-pairs suggested selective constraint on both X- and Y-linked genes and, therefore, that both copies are functional. Divergence between *SlY4* and *SlX4* was much higher than that between the *SlY1* and *SlX1* genes. As for human XY-linked genes, the plant sex-linked loci ceased recombining at different times and reveal distinct events in the evolutionary history of the sex chromosomes. Filatov et al. [72] reported a more extensive study of nucleotide diversity in the *SlX1/SlY1* and *SlX4/SlY4* genes, including a larger *S. latifolia* sample and a sample from the closely related species *S. dioica*. They also included the study of the diversity of an autosomal gene, *CCLS37.1* and confirmed that nucleotide diversity in the Y-linked genes of both *S. latifolia* and *S. dioica* is very low compared with that of the X-linked genes. However, the autosomal gene also had low DNA polymorphism. The favourite explanation was a selective sweep hypothesis. Using single individual of the related hermaphrodite species *S. conica*, as an outgroup, they showed that the low *SlY1* diversity is not due to a lower mutation rate than that for the X-linked gene. The authors also tested several other hypotheses for explaining low *SlY1* diversity, including differential gene flow between the two species for Y-linked, X-linked, and autosomal genes. The frequency spectrum of nucleotide polymorphism on the Y-chromosome deviated significantly from that expected under the selective-sweep model. However, they detected population subdivision in both *S. latifolia* and *S. dioica*. A further gene important for evolutionary studies in *Silene* is *DD44* [153], which is homologous to the oligomycin sensitivity-conferring protein, an essential component of the mitochondrial ATP synthase, and

is ubiquitously expressed in both sexes. This gene was found to be under selective constraint in both the X-and Y-linked copy and, therefore, it is expected that both copies are functional. When another sex linked gene (*SlX3/SlY3*, putative calcium-dependent protein kinase) in *S. latifolia* was described, the number of the known genes became sufficient to test the hypothesis that the arrest of recombination in *S. latifolia* is progressive, similarly to mammals [116, 201] in spite of the fact that the sex chromosomes in *S. latifolia* are much younger.

Nicolas et al. [159] used a similar approach as Lahn and Page [116], i.e., they studied the relation between the distance of the given X-linked gene from the pseudoautosomal region and its level of synonymous substitution. In such experiments, the level of the silent-site divergence between X and Y serves as a measure of time from the period when the genes were not yet sex linked. Although the X-Y system in *S. latifolia* evolved much earlier than mammals, it was possible to find gradient in the level of synonymous substitutions in *S. latifolia* and two related species *S. dioica* and *S. diclinis* for the genes *SlX1/SlY1*, *DD44X/DD44Y*, *SlX3/SlY3* and *SlX4/SlY4* (and their homologues in other species, respectively), i.e., the larger the genetic distance of the studied gene from the pseudoautosomal marker, the higher was the level of the synonymous substitutions. The only exception was the gene pair *SlX4/SlY4*, which showed comparable level of synonymous substitutions with gene pair *SlX3/SlY3* (this fact suggests that in these two genes the recombination ceased with the X-linked homologues in the same period). A slightly different gene order was found in *S. dioica* for the gene, *X1* and *DD44X*, but comparison with the other species suggests that it was a recent inversion, hence for the evaluations of the data, the order of loci found in other two species *S. latifolia* and *S. diclinis* was used. These data show that the process of recombination arrest between the X and Y was in *S. latifolia* stepwise. While in most genes the X copies and Y copies formed separate groups in the phylogenetic analysis the *SlX1/SlY1* showed grouping of the sequences according to the species. The two possible explanations are: either *SlX1/SlY1* stopped recombining after the species split, or they stopped recombining shortly before the dioecious species split, and hence some species specific polymorphisms sites are common both to the X-and Y-chromosomes. Some data indicate incipient degeneration of the Y-chromosome obtained by the analysis of the larger parts of the three previously studied genes and analysis of the large part of the gene *SlY3/SlX3*. The rate of non-synonymous substitution was

about three times higher for the Y copies of two genes from the oldest “stratum” (*SlY3* and *SlY4*) in comparison with the X- copies of these genes. The difference was not caused by the higher level of the mutation on the Y-chromosome as the dS ratio for the X-and Y-linked allele was close to 1 in both these genes.

The gene *DD44* was also studied in relation to the diversity of the X and Y linked copy [118] and supported theories predicting that natural selection reduces the effective population size of the non-recombining Y-chromosome beyond the effect of low actual population size because there is only one Y per three X chromosomes or four autosomes. With these data, there are now four Y-linked loci, whose diversity was compared with X-linked or autosomal counterparts, and in each case diversity is reduced more than expected, based on the ploidy difference [71, 72, 140, 159]. The authors also discussed the possibility that higher divergence of the X-chromosome linked sequences is caused by a different ability of the X and Y-chromosome to introgress to the related species. The fact that the different divergence can influence the estimates of gene flow is taken into consideration. The final conclusion was that despite the striking difference in their divergence patterns, there is evidence for not more than minor differences in interspecific gene flow between the sex chromosomes. This is consistent with a high fertility of hybrid plants of both sexes [207]. The study of the other Y-linked gene *SlssY* and its X-linked copy *SlssX* tested the possible elevated synonymous rate in the *SlssY* gene [74]. The results suggest that some genes on the *S. latifolia* Y-chromosome can indeed have a higher mutation rate, compared with the X-chromosome [73]. High mutation rate variability has been already reported on the bird sex chromosomes [19]. When differences in silent substitution rate are taken into account, the Y-linked gene still demonstrates significantly faster accumulation of non-synonymous substitutions, consistent with the theoretical prediction of relaxed purifying selection in Y-linked genes, leading to the accumulation of non-synonymous substitutions and genetic degeneration of the Y-linked genes. Degeneration processes were also suggested as an explanation of the low diversity of the *SlAP3Y* (a copy of the gene which was duplicatively transferred on the Y-chromosome from an autosome) in comparison with its autosomal paralog [140]. The next research was based on the study of the diversity of noncoding region of the homologous genes *DD44Y* and *DD44X* [97] from *S. latifolia* population from different geographical areas and various populations of related species *S. dioica*, *S. diclinis* and *S.*

*heuffelii*. On the *S. latifolia* Y-chromosome, the authors found a substantial level of the DNA diversity [75]. The geographical population structure of the Y-linked copy was far higher than that of the X copy. Differentiation between the species was also higher for the Y- than for the X-chromosome. These findings indicate that the loss of genetic diversity on the Y-chromosome in *Silene* (described earlier by authors [72, 159]) occurs within local populations rather than within entire species. These results are compatible with the background selection, Muller's ratchet, and local selective sweeps, but not with species-wide selective sweeps. The higher interspecific divergence of *DD44Y*, compared to *DD44X*, supports the hypothesis that the Y-chromosome differentiation between incipient species precedes reproductive isolation of the entire genome, forming an early stage in the process of speciation. The mechanism that causes the decreased ability of the Y-chromosomes to introgress to the related species is not yet clear. As interspecific hybrids have a relatively good fertility, the most probable explanation can be that it is certation which prevents the spread of the introgressed Y-chromosome in the population of the related species. This would be in accordance with the data of Taylor [208] showing a higher sex bias in the interspecific crosses *S. latifolia* x *S. dioica*.

A mosaic of the data collected a step by step approach to the interesting question of the sex chromosome evolution and mechanisms of the recombination arrest. A typical example of such a synthesis [238] showed the presence of the large inversion distinguishing gene order on the X- and Y-chromosomes. In the context of the previously published data based on the synonymous site divergence, it was possible to deduce that this large inversion (including genes *SIY4*, *SIY3*, *DD44*, *SlssY* but not *SIY1*) occurred after the arrest of recombination between loci of the “oldest stratum” (*SIY4* and *SIY3*) and the gene *DD44Y*.

### **Male Functions of Y-Linked Genes**

As anthers in silver-treated female plants are not as developed as those of male flowers [121], it is possible to expect that there are gene(s) on the Y-chromosome which do not possess any functional homologues within the genome. Similar results were obtained in the *S. latifolia* x *S. viscosa* interspecific hybrid, where anther formation is not fully restored by the addition of the genome from hermaphroditic *S. viscosa* to the *S. latifolia* female genome [239]. One of the genes, which acquired on the Y-

chromosome a new function in the anther and pollen grain maturation, is *SLAP3Y* [140]. This gene was duplicated from an autosome to the Y-chromosome soon after the evolution of the sex chromosomes. The *SLAP3* copies underwent after the duplication onto the Y-chromosome subfunctionalization – the temporal expression of the autosomal copy was probably changed, hence the expression of *SLAP3Y* has changed from petal-specific to largely anther expression. However, this gene does not belong to the Y-linked genes without functional homologues on the X-chromosome or autosomes, because deletion of this gene does not lead to male sterility of the mutants [238].

### **Dosage Compensation of X-Linked Genes**

On the Y-chromosome, the reverse process of preservation of genes with functions in male reproduction is degeneration of genes, which are dispensable for male reproduction. When the degenerated gene has a functional X-linked copy, the gene expression in females is two-fold higher than in males, and this expression imbalance may be somehow compensated [88].

In *S. latifolia*, all known X-linked genes have their Y-linked copies [8, 55, 74, 153, 159]. So far, Y-linked copies of all studied gene pairs are expressed [8, 55, 153]. From the available information on the known X- and Y-linked genes, it is not possible to arrive at a conclusion regarding the degree of the Y-chromosome degeneration. Therefore, this information cannot be conclusive even for the presence of dosage compensation of X-linked genes.

Dosage compensation of X-linked genes was analyzed by global studies of DNA replication kinetics and DNA methylation patterns. The first experiments based on autoradiographic detection of DNA methylation [220] and replication [197] patterns indicated differences between the two X-chromosomes in female cells. Subsequent studies using more sensitive antibody labelling enabled much clearer conclusions [198, 199, 221]. The antibody-based study of DNA methylation patterns revealed DNA hypermethylation of the short arm in the X-chromosome in males, whereas the Y-chromosome remained lowly DNA methylated [198]. However, the DNA methylation pattern of the X-chromosomes in females was asymmetric: DNA methylation pattern of one of the X resembled that observed in males, whereas the other X-chromosome was DNA hypermethylated. Similar results were observed in tetraploid hairy-root cultures [199]. These results are further supported by the fact that

in *S. dioica*, a closely related species, different quinacrine-mustard staining patterns of the X-chromosome arms were observed [106]. Another piece of evidence indicating the presence of dosage compensation of X-linked genes comes from the study of DNA replication patterns [221] – the replication of one X-chromosome in females finishes later than the other X-chromosome. The analysis of tetraploid females shows a similar pattern of DNA replication – two X-chromosomes displayed a prominent delay in DNA replication [199].

Therefore, DNA methylation and replication patterns of sex chromosomes may indicate the existence of dosage compensation of X-linked genes in *S. latifolia*. As there are prominent differences in DNA replication timing and DNA methylation pattern between the two X-chromosomes in female cells, the mechanism is probably based on inactivation of one X-chromosome as in mammals [88], and not on the increase of sex-linked gene expression in males as in *Drosophila*, or on the partial down-regulation of both X-chromosomes as in *Caenorhabditis* [132].

### ***Accumulation of Repetitive Sequences***

The main advantage of *S. latifolia* as a model object for the study of sex chromosome evolution, is the presence of heteromorphic sex chromosomes [158, 222]. From cytogenetic point of view, it is easy to recognize both the X-chromosome from the Y, and the sex chromosomes from the autosomes. This feature makes *S. latifolia* a suitable model for direct dissection of sex chromosomes by both manual and laser beam-based microdissection and by flow sorting. The search for sex specific molecular markers using separated sex chromosomes in *S. latifolia* started at the beginning of last decade of the last century. The initial experiments were conducted by Grant et al. [84] and Veuskens et al. [219]. The latter developed a protocol to isolate the Y-chromosome by flow sorting approach. The procedure commences with *Agrobacterium rhizogenes* based transformation of *S. latifolia* plants (preparation of hairy-root cultures), cultivation of roots *in vitro*, synchronization of dividing cells and blocking mitosis, reducing cells to protoplasts, and lysis protoplasts to release chromosomes. Although the purity of the Y-chromosome fraction after sorting (90%) was not suitable enough for most of the subsequent applications (direct PCR with primers for a gene of interest, non-specific amplification before chromosome-specific library construction), the

protocol which utilizes hairy cultures and “protoplast step” during chromosome preparation is currently the most feasible method for making high quality preparations for cytogenetic experiments. One of the first sequences characterized in *S. latifolia* directly derived from the sex chromosomes was X43.1 [32]. To obtain a DNA library composed of a sex-chromosome specific DNA, the X-chromosomes were manually microdissected using an inverted microscope and micromanipulator. Chromosomes were amplified using DOP-PCR and PCR products were subsequently cloned. One of the cloned sequences called X43.1 revealed after Southern hybridization a strong signal and tandem-repeat like hybridization pattern. Subsequent FISH analysis revealed that this repetitive sequence is localized in subtelomeric regions of most chromosomes but is absent in the non-homologous arm of the Y-chromosome [32]. Since the *S. latifolia* Y-chromosome is metacentric, FISH localization of this repetition enables routine distinction between its chromosome arms (used, for example, by Lengerova et al. [124]). This repetitive sequence was later also isolated independently by others [80]. The subsequent attempt at microdissection of sex chromosomes was aimed at finding out if there exist Y-chromosome-enriched or specific sequences, and to what extent X-and Y-chromosomes diverged from each other in *S. latifolia*. Scutt et al. [192] used mitotic metaphase chromosome preparations on polyester membranes for microdissection. The manual dissection of sex chromosomes (by recovering excised fragments of polyester membrane) was carried out after ablation of autosomes using an argon ion laser microbeam. Dissected DNA was amplified by DOP-PCR, labelled and used as a complex cytogenetic probe for FISH. Both X-and Y-chromosome derived probes revealed signals homogenous to all chromosomes and the conclusion of the study was that the X and Y sex chromosomes of *S. latifolia* are of a very similar DNA composition and also that they share a significant repetitive DNA content with the autosomes. The same results were later achieved by Matsunaga et al. [139]. Subsequently, a new, improved FISH protocol devised by Hobza et al. [93] helped in revising the conclusions made by Scutt et al. [192] and Matsunaga et al. [139]. In this paper [93], for the purpose of microdissection of both the X-and Y-chromosomes, the authors used direct laser microdissection that minimizes potential contamination by extraneous DNA, since only a chromosome of interest is selected, no ablation step is needed and all manipulations are carried out using only a laser beam. The protocol follows the standard

amplification procedure using optimized DOP-PCR. By using a small amount of complex hybridization probe derived from a particular sex chromosome and short hybridization time, a specific “painting” of both X- and Y-chromosomes was achieved. This experiment revealed for the first time that differences exist between both the X-and Y-chromosomes, and between the sex chromosomes and autosomes. It was hypothesized that during the short hybridization period only repetitive DNA in the probe mixture could re-associate and that is why DNA repeats play a major role in diversification of sex chromosomes even at the beginning of their evolution. The next attempts to isolate sequences specifically accumulated on the sex chromosomes (especially the Y-chromosome) utilized a construction of a sample BAC library of *S. latifolia* [125]. To get suitable cytogenetic probes and exclude them from the BAC set clones containing widespread repetitive DNA that complicate hybridizations and disable achievement of specific chromosomal signal(s), the BAC library was hybridized with female genomic probe. Weakly hybridizing BACs then contained either low copy sequences from either autosome and sex chromosomes or repetitive DNA specific for the Y-chromosome. Finally, 25 BAC clones displaying discrete signals on chromosomes were found. Of these BAC clones, two gave discrete signals on the autosomes, 15 hybridized to subtelomeric regions of all chromosomes, and five clones hybridized to 45S rDNA clusters. Three of the BACs revealed strong signals on the sex chromosomes with tendency to accumulate on the Y-chromosome (BAC9B7, BAC7H5 and BAC6B3). Subsequent experiments by Kejnovsky et al. [107] have focused on characterization of BAC7H5 that is preferentially localized on the Y-chromosome. Sequencing of the BAC showed a strong homology to the chloroplast genome of some species and it was thereafter proved that this BAC clone is of chloroplast origin. To isolate DNA that is present on the sex chromosomes and is highly similar to the chloroplast DNA, both X- and Y-chromosome specific libraries were constructed. Screening of the libraries with the BAC7H5 probe led to isolation of five Y-chromosome-linked clones and twelve X-chromosome derived positive clones that revealed a high degree of similarity to chloroplast genes. Out of all the clones, only one Y-derived clone exhibited a higher level of divergence from its chloroplast homologue. In general, degeneration of genic regions on the Y-chromosome was higher than that on the X-chromosome. Three out of four chloroplast sequences located on the Y-chromosome contained stop codons and did not present functional copies of the genes.

The authors concluded that similar to the human Y-chromosome, which is more susceptible to transfer of organellar (mitochondrial) DNA, *S. latifolia* Y-chromosome accumulated chloroplast DNA that moreover contributes to the large size of the Y-chromosome.

Recently, studies that focused on characterization of another cytogenetic marker BAC9B7, previously published by Lengerova et al. [125], revealed a role of tandem repeats in the Y-chromosome evolution in *S. latifolia* [95]. The original BAC9B7 was subcloned and hybridized with genomic DNA to obtain clones that could contain repetitive elements responsible for the accumulated signal on the Y-chromosome. The sequence of several sub-clones revealed a tandem repetitive DNA motif with monomer unit 172 bp. The tandem arrangement of the sequence was subsequently proved by Southern hybridization. FISH experiments with labelled monomer of the repeat named TRAYC (Tandem Repeat Accumulated on the Y Chromosome) gave the same signal pattern as was originally achieved with the entire BAC9B7. An interesting feature of genomic distribution of TRAYC is that it is present in centromeric regions of some autosomes, but strongly accumulated on the Y and partially on the X chromosomes. That could be a consequence of cessation of recombination within the centromeric region that enabled the spread of centromeric repetitive DNA in non-recombining region(s). The concept that the centromeric region could play a key role both in functional and structural evolution of sex chromosomes is supported by the fact that TRAYC has accumulated in the closely related species with sex chromosomes (*S. dioica* and *S. diclinis*) in a similar pattern, but not in closely related non-dioecious species (*S. vulgaris* and *S. viscosa*). This indicates that accumulation of the TRAYC occurred after the sex chromosomes in *Elisanthe* section have evolved, which took place before speciation of the mentioned dioecious species from a common ancestor. An interesting feature of the TRAYC is a palindromic character of the monomer unit, which could play a role in the quick spread of the sequence during Y-chromosome evolution.

## PERSPECTIVES

The plant genus *Silene* has become an emerging model system for plant ecology, genetics, evolution and development. Many testable conclusions arise from the data summarized in this chapter. The first step is to obtain as many sequencing data as possible. It would be very interesting to reach

complete sequence of the sex chromosomes in dioecious *Silene* species, as well as sequences of homologous chromosomes in species without sex chromosomes. Similarly, very useful information would be obtained from sequencing of CMS factors bearing mitochondria in gynodioecious species, especially in *S. vulgaris*. The currently published *S. vulgaris* AFLP map [28] will serve to map and clone fertility restorer(s) in this species. To prove the identity of the obtained genes (fertility restorers in *S. vulgaris* and sex determining genes in *S. latifolia*), it will be necessary to construct transgenic plants. However, at present, there is no efficient transformation protocol enabling to carry out large-scale preparation of transgenic plants. A promising fact is that *Agrobacterium rhizogenes* is able to induce formation of hairy roots in many *Silene* species, thus *Agrobacterium*-based transformation in *Silene* species should be, in principle, feasible.

## Acknowledgment

The authors are grateful for the support extended by: (1) Institutional Research Plan (AVOZ50040507), (2) Grant Agency of the Czech Republic (204/05/P505 and 521/05/2076), and (3) Czech Ministry of Education.

## References

- [1] Ageez A, Kazama Y, Sugiyama R, Kawano S. Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. *Genes and Genetic Systems* 2005; 80: 403-413.
- [2] Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M. Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. *Plant Cell* 1997; 9: 841-857.
- [3] Aida M, Ishida T, Tasaka M. Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development* 1999; 126: 1563-1570.
- [4] Ali S, Hasnain SE. Genomics of the human Y-chromosome. 1. Association with male infertility. *Gene* 2003; 321: 25-37.
- [5] Andersson H. Female and hermaphrodite flowers on a chimeric gynomonoecious *Silene vulgaris* plant produce offspring with different genders: a case of heteroplasmic sex-determination? *Journal of Heredity* 1999; 90: 563-565.
- [6] Antonovics J, Alexander HM. Epidemiology of anther-smut infection of *Silene alba* (= *S. latifolia*) caused by *Ustilago violacea*: patterns of spore deposition in experimental populations. *Proceedings of the Royal Society of London Series B-Biological Sciences* 1992; 250: 157-163.

- [7] Assuncao AGL, Martins PDC, Folter SD, Vooijs R, Schat H, Aarts MGM. Elevated expression of metal transporter genes in three accessions of the metal hyperaccumulator *Thlaspi caerulescens*. *Plant Cell and Environment* 2001; 24: 217-226.
- [8] Atanassov I, Delichere C, Filatov DA, Charlesworth D, Negrutiu I, Moneger F. Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. *Mol Biol Evol* 2001; 18: 2162-2168.
- [9] Audran JC, Batcho M. Microsporogenesis and pollen grains in *Silene dioica* (L.) Cl. and alterations in its anthers parasitized by *Ustilago violacea* (Pers.) Rouss. (Ustilaginales). *Acta Societatis Botanicorum Poloniae* 1981; 50: 29-33.
- [10] Bailey MF. A cost of restoration of male fertility in a gynodioecious species, *Lobelia siphilitica*. *Evolution* 2002; 56: 2178-2186.
- [11] Bailey MF, Delph LF, Lively CM. Modeling gynodioecy: novel scenarios for maintaining polymorphism. *American Naturalist* 2003; 161: 762-776.
- [12] Bailey MF, McCauley DE. The effects of inbreeding, outbreeding and long-distance gene flow on survivorship in North American populations of *Silene vulgaris*. *Journal of Ecology* 2006; 94: 98-109.
- [13] Baltisberger M, Widmer A. Chromosome numbers of plant species from the Canary Islands. *Botanica Helvetica* 2006; 116: 9-30.
- [14] Barbacar N, Hinnidaels S, Farbos I, Moneger F, Lardon A, Delichere C, Mouras A, Negrutiu I. Isolation of early genes expressed in reproductive organs of the dioecious white campion (*Silene latifolia*) by subtraction cloning using an asexual mutant. *Plant Journal* 1997; 12: 805-817.
- [15] Barrett SCH, Case AL, Peters GB. Gender modification and resource allocation in subdioecious *Wurmbea dioica* (Colchicaceae). *Journal of Ecology* 1999; 87: 123-137.
- [16] Barrett SCH. The evolution of plant sexual diversity. *Nature Reviews Genetics* 2002; 3: 274-284.
- [17] Barton M, Poethig RS. Formation of the shoot apical meristem in *Arabidopsis thaliana*: An analysis of development in the wild type and in the shoot meristemless mutant. *Development* 1993; 119: 823-831.
- [18] Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 2002; 99: 10887-10892.
- [19] Berlin S, Brandstrom M, Backstrom N, Axelsson E, Smith NG, Ellegren H. Substitution rate heterogeneity and the male mutation bias. *Journal of Molecular Evolution* 2006; 62: 226-233.
- [20] Biere A, Elzinga JA, Honders SC, Harvey JA. A plant pathogen reduces the enemy-free space of an insect herbivore on a shared host plant. *Proceedings of the Royal Society of London Series B-Biological Sciences* 2002; 269: 2197-2204.
- [21] Biere A, Honders SC. Coping with third parties in a nursery pollination mutualism: *Hadena bicruris* avoids oviposition on pathogen-infected, less rewarding *Silene latifolia*. *New Phytologist* 2006; 169: 719-727.
- [22] Bittrich V. Caryophyllaceae. In: Kubitzki K, Rohwer JG, Bittrich V, eds. *The families and genera of vascular plants*. Berlin: Springer, 1993.
- [23] Blackburn KB. Sex chromosomes in plants. *Nature* 1923; 112: 687-688.
- [24] Blackburn KB. On the relation between geographical races and polyploidy in *Silene ciliata*. *Genetica* 1933; 15: 49-66.

- [25] Brady KU, Kruckeberg AR, Bradshaw HD. Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology Evolution and Systematics* 2005; 36: 243-266.
- [26] Bratteler M, Baltisberger M, Widmer A. QTL analysis of intraspecific differences between two *Silene vulgaris* ecotypes. *Annals of Botany* 2006; 98: 411-419.
- [27] Bratteler M, Lexer C, Widmer A. Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. *Journal of Evolutionary Biology* 2006; 19: 1149-1156.
- [28] Bratteler M, Lexer C, Widmer A. A genetic linkage map of *Silene vulgaris* based on AFLP markers. *Genome* 2006; 49: 320-327.
- [29] Breuil-Broyer S, Morel P, de Almeida-Engler J, Coustham V, Negruiti I, Trehin C. High-resolution boundary analysis during *Arabidopsis thaliana* flower development. *Plant Journal* 2004; 38: 182-192.
- [30] Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS. The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant Journal* 2003; 35: 262-272.
- [31] Budar F, Touzet P, De Paepe R. The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. *Genetica* 2003; 117: 3-16.
- [32] Buzek J, Koutnikova H, Houben A, Riha K, Janousek B, Siroky J, Grant S, Vyskot B. Isolation and characterization of X chromosome-derived DNA sequences from a dioecious plant *Melandrium album*. *Chromosome Research* 1997; 5: 57-65.
- [33] Carroll SB, Mulcahy DL. The relationship between pollen grain size and progeny gender in dioecious *Silene latifolia* (Caryophyllaceae). *Sexual Plant Reproduction* 1991; 4: 203-207.
- [34] Carroll SB, Mulcahy DL. The effect of limited pollen on progeny sex ratios in dioecious *Silene latifolia* (Caryophyllaceae). *Amer J Bot* 1993; 80: 551-556.
- [35] Charlesworth B, Charlesworth D. A model for the evolution of dioecy and gynodioecy. *American Naturalist* 1978; 112: 975-997.
- [36] Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics* 1993; 134: 1289-1303.
- [37] Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 1994; 371: 215-220.
- [38] Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 2000; 355: 1563-1572.
- [39] Charlesworth B. The organization and evolution of the human Y chromosome. *Genome Biology* 2003; 4: 226.
- [40] Charlesworth D. A further study of the problem of the maintenance of females in gynodioecious species. *Heredity* 1981; 46: 27-39.
- [41] Charlesworth D. Androdioecy and the evolution of dioecy. *Biological Journal of the Linnean Society* 1984; 22: 333-348.
- [42] Charlesworth D, Laporte V. The male-sterility polymorphism of *Silene vulgaris*: analysis of genetic data from two populations and comparison with *Thymus vulgaris*. *Genetics* 1998; 150: 1267-1282.
- [43] Charlesworth D. Evolution: an exception that proves the rule. *Current Biology* 2001; 11: R13-15.

- [44] Charlesworth D. Plant sex determination and sex chromosomes. *Heredity* 2002; 88: 94-101.
- [45] Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 2005; 95: 118-128.
- [46] Charlesworth D. Evolution of plant breeding systems. *Current Biology* 2006; 16: R726-735.
- [47] Chowdhuri P. Studies in the genus *Silene*. Notes from the Royal Botanic Garden Edinburgh 1957; 22: 221-278.
- [48] Clark SE, Jacobsen SE, Levin JZ, Meyerowitz EM. The CLAVATA and SHOOT MERISTEMLESS loci competitively regulate meristem activity in *Arabidopsis*. *Development* 1996; 122: 1567-1575.
- [49] Collin CL, Pennings PS, Rueffler C, Widmer A, Shykoff JA. Natural enemies and sex: how seed predators and pathogens contribute to sex-differential reproductive success in a gynodioecious plant. *Oecologia* 2002; 131: 94-102.
- [50] Correns C. Bestimmung, Vererbung und Verteilung des Geschlechtes bei den höheren Pflanzen. Handbuch der Vererbungswissenschaft Band II. Berlin: Verlag von Gebruder Borntraeger, 1928.
- [51] Curtsinger JW. Components of selection in X chromosome lines of *Drosophila melanogaster*: sex ratio modification by meiotic drive and viability selection. *Genetics* 1984; 108: 941-952.
- [52] Cutler M. An account of some of the vegetable productions, naturally growing in this part of America, botanically arranged. *Memoirs from the American Academy of Arts Science* 1785; 1: 396-493.
- [53] Damboldt J, Phitos D. Zur Cytotaxonomie einiger Arten der Gattung *Silene* L. (Caryophyllaceae). *Verhandlungen des botanischen Vereins der Provinz Brandenburg* 1968; I05: 44-51.
- [54] Degraeve N. Étude de diverses particularités caryotypiques des genres *Silene*, *Lychnis* et *Melandrium*. *Boletim Sociedade Broteriana* 1980; 53: 595-643.
- [55] Delichere C, Veuskens J, Hernould M, Barbacar N, Mouras A, Negruiti I, Moneger F. SY1, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. *EMBO Journal* 1999; 18: 4169-4179.
- [56] Delph LF. Sex-ratio variation in the gynodioecious shrub *Hebe strictissima* (Scrophulariaceae). *Evolution* 1990; 44: 134-142.
- [58] Delph LF. Sexual dimorphism in gender plasticity and its consequences for breeding system evolution. *Evolution and Development* 2003; 5: 34-39.
- [57] Delph LF, Carroll SB. Factors affecting relative seed fitness and female frequency in a gynodioecious species, *Silene acaulis*. *Evolutionary Ecology Research* 2001; 3: 487-505.
- [59] Delph LF, Wolf DE. Evolutionary consequences of gender plasticity in genetically dimorphic breeding systems. *New Phytologist* 2005; 166: 119-128.
- [60] Desfeux C, Lejeune B. Systematics of euromediterranean *Silene* (Caryophyllaceae): Evidence from a phylogenetic analysis using ITS sequences. *Comptes Rendus de l'Academie des Sciences Serie III-Sciences de la Vie - Life Sciences* 1996; 319: 351-358.
- [61] Desfeux C, Maurice S, Henry JP, Lejeune B, Gouyon PH. Evolution of reproductive systems in the genus *Silene*. *Proceedings of the Royal Society of London Series B-Biological Sciences* 1996; 263: 409-414.

- [62] Donnison IS, Siroky J, Vyskot B, Saedler H, Grant SR. Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. *Genetics* 1996; 144: 1893-1901.
- [63] Dotterl S, Wolfe LM, Jurgens A. Qualitative and quantitative analyses of flower scent in *Silene latifolia*. *Phytochemistry* 2005; 66: 203-213.
- [64] Dotterl S, Burkhardt D, Weissbecker B, Jurgens A, Schutz S, Mosandl A. Linalool and lilac aldehyde/alcohol in flower scents - Electrophysiological detection of lilac aldehyde stereoisomers by a moth. *Journal of Chromatography A* 2006; 1113: 231-238.
- [65] Dotterl S, Jurgens A, Seifert K, Laube T, Weissbecker B, Schutz S. Nursery pollination by a moth in *Silene latifolia*: the role of odours in eliciting antennal and behavioural responses. *New Phytologist* 2006; 169: 707-718.
- [66] Durand B, Durand R. Sex determination and reproductive organ differentiation in *Mercurialis*. *Plant Sci* 1991; 80: 49-65.
- [67] Ehlers BK, Maurice S, Bataillon T. Sex inheritance in gynodioecious species: a polygenic view. *Proceedings of the Royal Society of London Series B-Biological Sciences* 2005; 272: 1795-1802.
- [68] Endrizzi K, Moussian B, Haecker A, Levin JZ, Laux T. The SHOOT MERISTEMLESS gene is required for maintenance of undifferentiated cells in *Arabidopsis* shoot and floral meristems and acts at a different regulatory level than the meristem genes WUSCHEL and ZWILLE. *Plant Journal* 1996; 10: 967-979.
- [69] Farbos I, Oliveira M, Negrutiu I, Mouras A. Sex organ determination and differentiation in the dioecious plant *Melandrium album* (*Silene latifolia*): a cytological and histological analysis. *Sexual Plant Reproduction* 1997; 10: 155-167.
- [70] Farbos I, Veuskens J, Vyskot B, Oliveira M, Hinnidaels S, Aghmir A, Mouras A, Negrutiu I. Sexual dimorphism in white campion: deletion on the Y chromosome results in a floral asexual phenotype. *Genetics* 1999; 151: 1187-1196.
- [71] Filatov DA, Moneger F, Negrutiu I, Charlesworth D. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* 2000; 404: 388-390.
- [72] Filatov DA. Substitution rates in a new *Silene latifolia* sex-linked gene, SlssX/Y. *Mol Biol Evol* 2005; 22: 402-408.
- [73] Filatov DA, Charlesworth D. Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol Biol Evol* 2002; 19: 898-907.
- [74] Filatov DA, Laporte V, Vitte C, Charlesworth D. DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Mol Biol Evol* 2001; 18: 1442-1454.
- [75] Filatov DA. Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic Omapping of four genes. *Genetics* 2005; 170: 975-979.
- [76] Fior S, Karis PO, Casazza G, Minuto L, Sala F. Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast MATK and nuclear rDNA its sequences. *Amer J Bot* 2006; 93: 399-411.
- [77] Fischer GW, Holton CS. Biology and control of the smut fungi. New York: Ronald Press, 1957.
- [78] Fisher RA. The evolution of dominance. *Biological Reviews* 1931; 6: 345-368.
- [79] Galan R. Teoría genética del sexo zigótico en el caso de *Echium elatieri*. *Revista de Biología* 1964; 1: 187-220.

- [80] Garrido-Ramos MA, de la Herran R, Ruiz Rejon M, Ruiz Rejon C. A subtelomeric satellite DNA family isolated from the genome of the dioecious plant *Silene latifolia*. *Genome* 1999; 42: 442-446.
- [81] Glaettli M, Goudet J. Inbreeding effects on progeny sex ratio and gender variation in the gynodioecious *Silene vulgaris* (Caryophyllaceae). *New Phytologist* 2006; 172: 763-773.
- [82] Gleiser G, Verdu M. Repeated evolution of dioecy from androdioecy in *Acer*. *New Phytologist* 2005; 165: 633-640.
- [83] Gordo I, Navarro A, Charlesworth B. Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* 2002; 161: 835-848.
- [84] Grant S, Houben A, Vyskot B, Siroky J, Pan W-H, Macas J, Sadler H. Genetics of sex determination in flowering plants. *Developmental Genetics* 1994; 15: 214-230.
- [85] Greuter W. *Silene* (Caryophyllaceae) in Greece - a Subgeneric and Sectional Classification. *Taxon* 1995; 44: 543-581.
- [86] de Haan AA, Koelewijn HP, Hundscheid MP, Van Damme JM. The dynamics of gynodioecy in *Plantago lanceolata* L. II. Mode of action and frequencies of restorer alleles. *Genetics* 1997; 147: 1317-1328.
- [87] Hardenack S, Ye D, Saedler H, Grant S. Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. *Plant Cell* 1994; 6: 1775-1787.
- [88] Heard E, Distech CM. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes and Development* 2006; 20: 1848-1867.
- [89] Hermanutz LA, Innes DJ. Gender variation in *Silene acaulis* (Caryophyllaceae). *Plant Sys Evol* 1994; 191: 69-81.
- [90] Hertwig G, Hertwig P. Die Vererbung des hermaphroditismus bei *Melandrium*. *Zeitschrift fur induktive Abstammung und Vererbungslehre* 1922; 28: 259-294.
- [91] Heslop-Harrison J. Sex expression in flowering plants. *Brookhaven Symposium of Quantitative Biology* 1963; 16: 109-125.
- [92] Hickey WA, Craig GB. Genetic distortion of sex ratio in a mosquito, *Aedes aegypti*. *Genetics* 1966; 53: 1177-1196.
- [93] Hobza R, Lengerova M, Cernohorska H, Rubes J, Vyskot B. FAST-FISH with laser beam microdissected DOP-PCR probe distinguishes the sex chromosomes of *Silene latifolia*. *Chromosome Research* 2004; 12: 245-250.
- [94] Hobza R, Hrusakova P, Safar J, Bartos J, Janousek B, Zluvova J, Michu E, Dolezel J, Vyskot B. MK17, a specific marker closely linked to the gynoecium suppression region on the Y chromosome in *Silene latifolia*. *Theo Appl Genet* 2006; 113: 280-287.
- [95] Hobza R, Lengerova M, Svoboda J, Kubekova H, Kejnovsky E, Vyskot B. An accumulation of tandem DNA repeats on the Y chromosome in *Silene latifolia* during early stages of sex chromosome evolution. *Chromosoma* 2006; 115: 376-382.
- [96] Houliston GJ, Olson MS. Non-neutral evolution of organelle genes in *Silene vulgaris*. *Genetics* 2006; 174: 1983-1994.
- [97] Ironside JE, Filatov DA. Extreme population structure and high interspecific divergence of the *Silene* Y chromosome. *Genetics* 2005; 171: 705-713.
- [98] Jacobsen SE, Meyerowitz EM. Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis*. *Science* 1997; 277: 1100-1103.
- [99] Janousek B, Siroky J, Vyskot B. Epigenetic control of sexual phenotype in a dioecious plant, *Melandrium album*. *Molecular and General Genetics* 1996; 250: 483-490.

- [100] Janousek B, Grant SR, Vyskot B. Non-transmissibility of the Y chromosome through the female line in androhermaphrodite plants of *Melandrium album*. *Heredity* 1998; 80: 576-583.
- [101] Jurgens A, Witt T, Gottsberger G. Flower scent composition in night-flowering *Silene* species (Caryophyllaceae). *Biochem Syst Ecol* 2002; 30: 383-397.
- [102] Jurgens A, Witt T, Gottsberger G. Pollen grain numbers, ovule numbers and pollen-ovule ratios in Caryophylloideae: correlation with breeding systems, pollination, life form, style number, and sexual system. *Sexual Plant Reproduction* 2002; 14: 279-289.
- [103] Jurgens A. Flower scent composition in diurnal *Silene* species (Caryophyllaceae): phylogenetic constraints or adaption to flower visitors? *Biochem Syst Ecol* 2004; 32: 841-859.
- [104] Jurgens A. Nectar sugar composition and floral scent compounds of diurnal and nocturnal *Conophytum* species (Aizoaceae). *South African Journal of Botany* 2004; 70: 191-205.
- [105] Jurgens A. Comparative floral morphometrics in day-flowering, night-flowering and self-pollinated Caryophylloideae (*Agrostemma*, *Dianthus*, *Saponaria*, *Silene*, and *Vaccaria*). *Plant Syst Evol* 2006; 257: 233-250.
- [106] Kampmeijer P. Fluorescence pattern of the sex-chromosomes of *Melandrium dioicum*, stained with quinacrine-mustard. *Genetica* 1972; 43:201-206.
- [107] Kejnovsky E, Kubat Z, Hobza R, Lengerova M, Sato S, Tabata S, Fukui K, Matsunaga S, Vyskot B. Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. *Genetica* 2006; 128: 167-175.
- [108] Keller SR, Schwaegerle KE. Maternal sex and mate relatedness affect offspring quality in the gynodioecious *Silene acaulis*. *Journal of Evolutionary Biology* 2006; 19: 1128-1138.
- [109] Kent-First M. The Y chromosome and its role in testis differentiation and spermatogenesis. *Seminars in Reproductive Medicine* 2000, 18: 67-80.
- [110] Khoshoo TN. Basic chromosome numbers in *Silene* and *Saponaria*. *Nature* 1960; 186: 412-413.
- [111] Khoshoo TN, Bhatia SK. Cytogenetical approach to the taxonomy of *Silene conica conoidea*-Complex. *Proceedings of the Indian Academy of Sciences Section B* 1963; 57: 368-378.
- [112] Knuth P. *Handbook of flower pollination*. Vol. 2. Oxford: Clarendon Press, 1908.
- [113] Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant Journal* 2004; 37: 315-325.
- [114] Krahenbuhl M, Yuan YM, Kupfer P. Chromosome and breeding system evolution of the genus *Mercurialis* (Euphorbiaceae): implications of ITS molecular phylogeny. *Plant Syst Evol* 2002; 234: 155-170.
- [115] Kruckeberg AR. Interspecific hybridizations of *Silene*. *American Journal of Botany* 1955; 42: 373-378.
- [116] Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. *Science* 1999; 286: 964-967.
- [117] Lahn BT, Pearson NM, Jegalian K. The human Y chromosome, in the light of evolution. *Nature Reviews Genetics* 2001; 2: 207-216.
- [118] Laporte V, Filatov DA, Kamau E, Charlesworth D. Indirect evidence from DNA sequence diversity for genetic degeneration of the Y-chromosome in dioecious species

- of the plant *Silene*: the *SIY4/SIX4* and *DD44-X/DD44-Y* gene pairs. *Journal of Evolutionary Biology* 2005; 18: 337-347.
- [119] Lardon A, Georgiev S, Aghmir A, Le Merrer G, Negruiti I. Sexual dimorphism in white campion: complex control of carpel number is revealed by Y chromosome deletions. *Genetics* 1999; 151: 1173-1185.
- [120] Larsson J, Meller VH. Dosage compensation, the origin and the afterlife of sex chromosomes. *Chromosome Research* 2006; 14: 417-431.
- [121] Law TF, Lebel-Hardenack S, Grant SR. Silver enhances stamen development in female white campion (*Silene latifolia* [Caryophyllaceae]). *Amer J Bot* 2002; 89: 1014-1020.
- [122] Lebel-Hardenack S, Grant SR. Genetics of sex determination in flowering plants. *Trends in Plant Science* 1997; 2: 130-136.
- [123] Lebel-Hardenack S, Hauser E, Law TF, Schmid J, Grant SR. Mapping of sex determination loci on the white campion (*Silene latifolia*) Y chromosome using amplified fragment length polymorphism. *Genetics* 2002; 160: 717-725.
- [124] Lengerova M, Kejnovsky E, Hobza R, Macas J, Grant SR, Vyskot B. Multicolor FISH mapping of the dioecious model plant, *Silene latifolia*. *Theor Appl Genet* 2004; 108: 1193-1199.
- [125] Lengerova M, Moore RC, Grant SR, Vyskot B. The sex chromosomes of *Silene latifolia* revisited and revised. *Genetics* 2003; 165: 935-938.
- [126] Levin SA, Muller-Landau HC, Nathan R, Chave J. The ecology and evolution of seed dispersal: a theoretical perspective. *Annual Review of Ecology and Systematics* 2003; 34: 575-604.
- [127] Lewis D. Male sterility in natural populations of hermaphrodite plants. *New Phytologist* 1941; 40: 56-63.
- [128] Lloyd DG. Breeding systems in *Cotula* L. (Compositae, Anthemideae). *New Phytologist* 1972; 71: 1181-1194.
- [129] Lloyd DG. The maintenance of gynodioecy and androdioecy in angiosperms. *Genetica* 1975; 45: 325-339.
- [130] Long JA, Moan EI, Medford JI, Barton MK. A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of *Arabidopsis*. *Nature* 1996; 379: 66-69.
- [131] Lovett Doust J, O'Brian G, Lovett Doust I. Effect of density on secondary sex characteristics and sex ratio in *Silene alba* (Caryophyllaceae). *Amer J Bot* 1987; 74: 40-46.
- [132] Lucchesi JC, Kelly WG, Panning B. Chromatin remodeling in dosage compensation. *Annual Review of Genetics* 2005; 39: 615-651.
- [133] Lyons EE, Shah-Mahoney N, Lombard LA. Evolutionary dynamics of sex ratio and gender dimorphism in *Silene latifolia*: II. Sex ratio and flowering status in a potentially male-biased population. *The Journal of Heredity* 1995; 86: 107-113.
- [134] Manicacci D, Couvet D, Belhassen E, Gouyon PH, Atlan A. Founder effects and sex ratio in the gynodioecious *Thymus vulgaris* L. *Mol Ecol* 1996; 5: 63-72.
- [135] Marienfeld JR, Unseld M, Brandt P, Brennicke A. Mosaic open reading frames in the *Arabidopsis thaliana* mitochondrial genome. *Biological Chemistry* 1997; 378: 859-862.
- [136] Marin I, Baker BS. The evolutionary dynamics of sex determination. *Science* 1998; 28: 1990-1994.

- [137] Markova M, Lengerova M, Zluvova J, Janousek B, Vyskot B. Karyological analysis of an interspecific hybrid between the dioecious *Silene latifolia* and the hermaphroditic *Silene viscosa*. *Genome* 2006; 49: 373-379.
- [138] Matsunaga S, Kawano S, Takano H, Uchida H, Sakai A, Kuroiwa T. Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, *Melandrium album* (*Silene latifolia*). *Plant Journal* 1996; 10: 679-689.
- [139] Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S, Sakai A, Kuroiwa T. Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. *Plant and Cell Physiology* 1999; 40: 60-68.
- [140] Matsunaga S, Isono E, Kejnovsky E, Vyskot B, Dolezel J, Kawano S, Charlesworth D. Duplicative transfer of a MADS box gene to a plant Y chromosome. *Mol Biol Evol* 2003; 20: 1062-1069.
- [141] Matsunaga S, Uchida W, Kawano S. Sex-specific cell division during development of unisexual flowers in the dioecious plant *Silene latifolia*. *Plant and Cell Physiology* 2004; 45: 795-802.
- [142] Maurice S, Desfeux C, Mignot A, Henry J. Is *Silene acaulis* (Caryophyllaceae) a trioeorous species? Reproductive biology of two subspecies. *Canadian Journal of Botany* 1998; 76: 478-485.
- [143] McArthur ED, Freeman DC, Luckinbill LS, Sanderson SC, Noller GL. Are trioeocy and sexual lability in *Atriplex canescens* genetically based? Evidence from clonal studies. *Evolution* 1992; 46: 1708-1721.
- [144] McCauley DE, Brock MT. Frequency-dependent fitness in *Silene vulgaris*, a gynodioecious plant. *Evolution* 1998; 52: 30-36.
- [145] McCauley DE, Olson MS, Emery SN, Taylor DR. Population structure influences sex ratio evolution in a gynodioecious plant. *American Naturalist* 2000; 155: 814-819.
- [146] McCauley DE, Olson MS. Associations among cytoplasmic molecular markers, gender, and components of fitness in *Silene vulgaris*, a gynodioecious plant. *Molecular Ecology* 2003; 12: 777-787.
- [147] McCauley DE, Smith RA, Lisenby JD, Hsieh C. The hierarchical spatial distribution of chloroplast DNA polymorphism across the introduced range of *Silene vulgaris*. *Mol Ecol* 2003; 12: 3227-3235.
- [148] McCauley DE, Bailey MF, Sherman NA, Darnell MZ. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of *Silene vulgaris*, a gynodioecious plant. *Heredity* 2005; 95: 50-58.
- [149] McNeill J. *Silene alba* and *Silene dioica* in North-America and generic delimitation of *Lychnis*, *Melandrium*, and *Silene* (Caryophyllaceae). *Canadian Journal of Botany - Revue Canadienne de Botanique* 1978; 56: 297-308.
- [150] Meagher TR, Vassiliadis C. Phenotypic impacts of repetitive DNA in flowering plants. *New Phytologist* 2005; 168: 71-80.
- [151] Miller JS, Venable DL. Polyploidy and the evolution of gender dimorphism in plants. *Science* 2000; 289: 2335-2338.
- [152] Mogensen HL. The hows and whys of cytoplasmic inheritance in seed plants. *Amer J Bot* 1996; 83: 383-404.
- [153] Moore RC, Kozyreva O, Lebel-Hardenack S, Siroky J, Hobza R, Vyskot B, Grant SR. Genetic and functional analysis of DD44, a sex-linked gene from the dioecious plant

- Silene latifolia*, provides clues to early events in sex chromosome evolution. *Genetics* 2003; 163: 321-334.
- [154] Morris WF, Doak DF. Life history of the long-lived gynodioecious cushion plant *Silene acaulis* (Caryophyllaceae), inferred from size-based population projection matrices. *Amer J Bot* 1998; 85: 784-793.
- [155] Moyle LC, Olson MS, Tiffin P. Patterns of reproductive isolation in three angiosperm genera. *Evolution* 2004; 58: 1195-1208.
- [156] Muhlemann JK, Waelti MO, Widmer A, Schiestl FP. Postpollination changes in floral odor in *Silene latifolia*: Adaptive mechanisms for seed-predator avoidance? *Journal of Chemical Ecology* 2006; 32: 1855-1860.
- [157] Nakao S, Matsunaga S, Sakai A, Kuroiwa T, Kawano S. RAPD isolation of a Y chromosome specific ORF in a dioecious plant, *Silene latifolia*. *Genome* 2002; 45: 413-420.
- [158] Negrutiu I, Vyskot B, Barbacar N, Georgiev S, Moneger F. Dioecious plants. A key to the early events of sex chromosome evolution. *Plant Physiology* 2001; 127: 1418-1424.
- [159] Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I, Charlesworth D, Moneger F. A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *Public Library of Science Biology* 2005; 3: e4.
- [160] Nigtervecht van G. Genetic studies in dioecious *Melandrium* II. – sex determination in *Melandrium album* and *Melandrium dioicum*. *Genetica* 1966; 37: 307-344.
- [161] Nordborg M. A model of genetic modification in gynodioecious plants. *Proceedings of the Royal Society of London Series B-Biological Sciences* 1994; 257: 149-154.
- [162] Obara M, Matsunaga S, Nakao S, Kawano S. A plant Y chromosome-STS marker encoding a degenerate retrotransposon. *Genes and Genetic Systems* 2002; 77: 393-398.
- [163] Ohno S. Sex Chromosomes and Sex-Linked Genes. Berlin: Springer-Verlag, 1967.
- [164] Olson MS, McCauley DE. Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proceedings in Biological Sciences / The Royal Society* 2000; 267: 1801-1808.
- [165] Olson MS, McCauley DE. Mitochondrial DNA diversity, population structure, and gender association in the gynodioecious plant *Silene vulgaris*. *Evolution* 2002; 56: 253-262.
- [166] Olson MS, McCauley DE, Taylor D. Genetics and adaptation in structured populations: sex ratio evolution in *Silene vulgaris*. *Genetica* 2005; 123: 49-62.
- [167] Olson MS, Graf AV, Niles KR. Fine scale spatial structuring of sex and mitochondria in *Silene vulgaris*. *Journal of Evolutionary Biology* 2006; 19: 1190-1201.
- [168] Oxelman B, Liden M. Generic boundaries in the tribe Sileneae (Caryophyllaceae) as inferred from nuclear rDNA sequences. *Taxon* 1995; 44: 525-542.
- [169] Oxelman B, Liden M, Berglund D. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Syst Evol* 1997; 206: 393-410.
- [170] Oxelman B, Liden M, Rabeler RK, Popp M. A revised generic classification of the tribe Sileneae (Caryophyllaceae). *Nordic Journal of Botany* 2001; 20: 743-748.
- [171] Pannell JR. The evolution and maintenance of androdioecy. *Ann Rev Ecol Syst* 2002; 33: 397-425.
- [172] Parker JS, Clark MS. Dosage sex-chromosomes systems in plants. *Plant Science* 1991; 80: 79-92.

- [173] Philipp M. Genetic diversity, breeding system, and population structure in *Silene acaulis* (Caryophyllaceae) in West Greenland. *Opera Bot.* 1997; 132: 89-100.
- [174] Popp M, Oxelman B. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homologous nuclear DNA sequences. *Mol Phylogen Evol* 2001; 20: 474-481.
- [175] Popp M. Disentangling the reticulate history of polyploids in *Silene* (Caryophyllaceae). Uppsala: Acta Universitatis Upsaliensis, 2004.
- [176] Popp M, Erixon P, Eggens F, Oxelman B. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. *Systematic Botany* 2005; 30: 302-313.
- [177] Prentice HC. Experimental taxonomy of *Silene* section Elisabethae (Caryophyllaceae) Crossing experiments. *Bot J Lin Soc* 1978; 77: 203-216.
- [178] Reboud X, Zeyl C. Organelle inheritance in plants. *Heredity* 1994; 72: 132-140.
- [179] Renner SS, Ricklefs RE. Dioecy and its correlates in the flowering plants. *Amer J Bot* 1995; 82: 596-606.
- [180] Rice WR. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 1987; 41: 911-914.
- [181] Rice WR. Evolution of the Y sex chromosome in animals. *BioScience* 1996; 46: 331-343.
- [182] Richards AJ. Plant breeding systems. London: George Allen & Unwin, 1986.
- [183] Rohde A, Grunau C, De Beck L, Van Montagu M, Rosenthal A, Boerjan W. Carpel, a new *Arabidopsis* epi-mutant of the SUPERMAN gene: phenotypic analysis and DNA methylation status. *Plant and Cell Physiology* 1999; 40: 961-972.
- [184] Ruddat M, Kokontis J, Birch L, Garber ED, Chiang KS, Campanella J, Dai H. Interactions of *Microbotryum violaceum* (*Ustilago violacea*) with its host plant *Silene alba*. *Plant Sci* 1991; 80: 157-165.
- [185] Saur Jacobs M, Wade MJ. A synthetic review of the theory of gynodioecy. *American Naturalist* 2003; 161: 837-851.
- [186] Schat H, Tenbookum WM. Genetic control of copper tolerance in *Silene vulgaris*. *Heredity* 1992; 68: 219-229.
- [187] Schat H, Vooijs R, Kuiper E. Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of *Silene vulgaris*. *Evolution* 1996; 50: 1888-1895.
- [188] Schat H, Vooijs R. Multiple tolerance and co-tolerance to heavy metals in *Silene vulgaris*: a co-segregation analysis. *New Phytologist* 1997; 136: 489-496.
- [189] Schieftaler U, Balasubramanian S, Sieber P, Chevalier D, Wisman E, Schneitz K. Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 1999; 96: 11664-11669.
- [190] Schnable PS, Wise RP. The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends in Plant Science* 1998; 3: 175-180.
- [191] Scutt CP, Li T, Robertson SE, Willis ME, Gilmartin PM. Sex determination in dioecious *Silene latifolia*. Effects of the Y chromosome and the parasitic smut fungus (*Ustilago violacea*) on gene expression during flower development. *Plant Physiology* 1997; 114: 969-979.

- [192] Scutt CP, Kamisugi Y, Sakai F, Gilmartin PM. Laser isolation of plant sex chromosomes: studies on the DNA composition of the X and Y sex chromosomes of *Silene latifolia*. *Genome* 1997; 40: 705-715.
- [193] Scutt CP, Jenkins T, Furuya M, Gilmartin PM. Male specific genes from dioecious white campion identified by fluorescent differential display. *Plant and Cell Physiology* 2002; 43: 563-572.
- [194] Shull GH. Inheritance of sex in *Lychnis*. *Botanical Gazette* 1910; 49: 110-125.
- [195] Shull GH. Reversible sex-mutants in *Lychnis dioica*. *Botanical Gazette* 1911; 52: 329-368.
- [196] Shykoff JA. Maintenance of gynodioecy in *Silene acaulis* (Caryophyllaceae): stage-specific fecundity and viability selection. *Amer J Bot* 1988; 75: 844-850.
- [197] Siroky J, Janousek B, Mouras A, Vyskot B. Replication patterns of sex chromosomes in *Melandrium album* female cells. *Hereditas* 1994; 120: 175-181.
- [198] Siroky J, Castiglione MR, Vyskot B. DNA methylation patterns of *Melandrium album* chromosomes. *Chromosome Research* 1998; 6: 441-446.
- [199] Siroky J, Hodurkova J, Negruiti I, Vyskot B. Functional and structural chromosome analyses in autotetraploid *Silene latifolia*. *Annals of Botany* 1999; 84: 633-638.
- [200] Siroky J, Lysak MA, Dolezel J, Kejnovsky E, Vyskot B. Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Research* 2001; 9: 387-393.
- [201] Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003; 423: 825-837.
- [202] Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The no apical meristem gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* 1996; 85: 159-170.
- [203] Steinemann M, Steinemann S. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica* 1998; 102-103: 409-420.
- [204] Storchova H, Olson MS. Comparison between mitochondrial and chloroplast DNA variation in the native range of *Silene vulgaris*. *Mol Ecol* 2004; 13: 2909-2919.
- [205] Takada S, Hibara K, Ishida T, Tasaka M. The CUP-SHAPED COTYLEDON1 gene of *Arabidopsis* regulates shoot apical meristem formation. *Development* 2001; 128: 1127-1135.
- [206] Taylor DR. The genetic basis of sex ratio in *Silene alba* (= *S. latifolia*). *Genetics*. 1994; 136: 641-651.
- [207] Taylor DR. Sex ratio in hybrids between *Silene alba* and *Silene dioica*: evidence for Y-linked restorers. *Heredity* 1994; 73: 518-526.
- [208] Taylor DR. Parental investment and offspring sex ratios in the dioecious plant *Silene alba* (= *S. latifolia*). *American Naturalist* 1996; 147: 870-879.
- [209] Taylor DR. Genetics of sex ratio variation among natural populations of a dioecious plant. *Evolution*. 1999; 53: 55-62.

- [210] Taylor DR, McCauley D, Trimble S. Colonization success of females and hermaphrodites in the gynodioecious plant, *Silene vulgaris*. *Evolution* 1999; 53: 745-751.
- [211] Taylor DR, Saur MJ, Adams E. Variation in pollen performance and its consequences for sex ratio evolution in a dioecious plant. *Evolution* 1999; 53: 1028-1036.
- [212] Taylor DR, Olson MS, McCauley DE. A quantitative genetic analysis of nuclear-cytoplasmic male sterility in structured populations of *Silene vulgaris*. *Genetics* 2001; 158: 833-841.
- [213] Taylor DR, Ingvarsson PK. Common features of segregation distortion in plants and animals. *Generica*, 2003; 117: 27-35.
- [214] Testolin R, Cipriani G, Costa G. Sex segregation ratio and gender expression in the genus *Actinidia*. *Sexual Plant Reproduction* 1995; 8: 129-132.
- [215] Thompson JD, Tarayre M. Exploring the genetic basis and proximate causes of female fertility advantage in gynodioecious *Thymus vulgaris*. *Evolution* 2000; 54: 1510-1520.
- [216] Thrall PH, Biere A, Antonovics J. Plant life history and disease susceptibility - the occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *Journal of Ecology* 1993; 81: 489-498.
- [217] Uchida W, Matsunaga S, Sugiyama R, Kazama Y, Kawano S. Morphological development of anthers induced by the dimorphic smut fungus *Microbotryum violaceum* in female flowers of the dioecious plant *Silene latifolia*. *Planta* 2003; 218: 240-248.
- [218] Uchida W, Matsunaga S, Kawano S. Ultrastructural analysis of the behavior of the dimorphic fungus *Microbotryum violaceum* in fungus-induced anthers of female *Silene latifolia* flowers. *Protoplasma* 2005; 226: 207-216.
- [219] Veuskens J, Marie D, Brown SC, Jacobs M, Negrutiu I. Flow sorting of the Y sex-chromosome in the dioecious plant *Melandrium album*. *Cytometry* 1995; 21: 363-373.
- [220] Vyskot B, Araya A, Veuskens J, Negrutiu I, Mouras A. DNA methylation of sex chromosomes in a dioecious plant, *Melandrium album*. *Mol Gen Genet* 1993; 239: 219-224.
- [221] Vyskot B, Siroky J, Hladilova R, Belyaev ND, Turner BM. Euchromatic domains in plant chromosomes as revealed by H4 histone acetylation and early DNA replication. *Genome* 1999; 42: 343-350.
- [222] Vyskot B, Hobza R. Gender in plants: sex chromosomes are emerging from the fog. *Trends in Genetics* 2004; 20: 432-438.
- [223] Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 2006; 18: 676-687.
- [224] Warmke HE. Sex determination and sex balance in *Melandrium*. *Amer J Bot* 1946; 33: 648-660.
- [225] Weir I, Lu J, Cook H, Causier B, Schwarz-Sommer Z, Davies B. CUPULIFORMIS establishes lateral organ boundaries in *Antirrhinum*. *Development* 2004; 131: 915-922.
- [226] Welch ME, Darnell MZ, McCauley DE. Variable populations within variable populations: quantifying mitochondrial heteroplasmy in natural populations of the gynodioecious plant *Silene vulgaris*. *Genetics* 2006; 174: 829-837.
- [227] Weller SG, Sakai AK. Using phylogenetic approaches for the analysis of plant breeding system evolution. *Ann Rev Ecol Syst* 1999; 30: 167-199.

- [228] Westerbergh A. Serpentine and non-serpentine *Silene dioica* plants do not differ in nickel tolerance. *Plant and Soil* 1994; 167: 297-303.
- [229] Westergaard M. Aberrant Y chromosomes and sex expression in *Melandrium album*. *Hereditas* 1946; 32: 419-443.
- [230] Westergaard M. The relation between chromosome constitution and sex in the offspring of triploid *Melandrium*. *Hereditas* 1948; 34: 257-279.
- [231] Westergaard M. The mechanism of sex determination in dioecious plants. *Advances in Genetics* 1958; 9: 217-281.
- [232] Winge O. On sex chromosomes, sex determination and preponderance of females in some dioecious plants. *C. R. Trav. Lab. Carlsberg* 1923; 15: 1-26.
- [233] Winge O. X-and Y-linked inheritance in *Melandrium*. *Hereditas* 1931; 15: 127-165.
- [234] Wolf DE, Takebayashi N. Pollen limitation and the evolution of androdioecy from dioecy. *American Naturalist* 2004; 163: 122-137.
- [235] Yang WC, Ye D, Xu J, Sundaresan V. The SPOROCYTELESS gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes and Development* 1999; 13: 2108-2117.
- [236] Zhang LB, Simmons MP, Kocyan A, Renner SS. Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: implications for morphological and sexual system evolution. *Mol Phylogen Evol* 2006; 39: 305-322.
- [237] Zhang YH, Stilio VS, Rehman F, Avery A, Mulcahy D, Kesseli R. Y chromosome specific markers and the evolution of dioecy in the genus *Silene*. *Genome* 1998; 41: 141-147.
- [238] Zluvova J, Janousek B, Negruiti I, Vyskot B. Comparison of the X and Y chromosome organization in *Silene latifolia*. *Genetics* 2005; 170: 1431-1434.
- [239] Zluvova J, Lengerova M, Markova M, Hobza R, Nicolas M, Vyskot B, Charlesworth D, Negruiti I, Janousek B. The inter-specific hybrid *Silene latifolia* x *S. viscosa* reveals early events of sex chromosome evolution. *Evolution and Development* 2005; 7: 327-336.
- [240] Zluvova J, Nicolas M, Berger A, Negruiti I, Moneger F. Premature arrest of the male flower meristem precedes sexual dimorphism in the dioecious plant *Silene latifolia*. *Proc Natl Acad Sci USA* 2006; 103: 18854-18859.



# Molecular Phylogenetics of Hamamelidaceae: Evidence from DNA Sequences of Nuclear and Chloroplast Genomes

JIANHUA LI<sup>1, 2, 3</sup>

<sup>1</sup>Arnold Arboretum of Harvard University, Cambridge, MA, USA

<sup>2</sup>College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

## ABSTRACT

Sequences of chloroplast gene *matK* and noncoding *trnL*-F region were used to elucidate phylogenetic relationships of Hamamelidaceae in a broad context of the Saxifragales. The results show that Hamamelidaceae forms a robust clade with the inclusion of *Exbucklandia* and *Rhodoleia*, and that *Chunia* and *Mytilaria* are closely related to each other. Sequences from nuclear rDNA ITS and *GBSSI* gene strongly support the sister relationship of *Shaniodendron* and *Parrotia*, which agrees with morphology. However, chloroplast data suggest that *Shaniodendron* forms a clade with the *Sycopsis* lineage containing *Distyliopsis*, *Distylium*, and *Sycopsis*. This indicates that there may be a chloroplast capture of *Shaniodendron* from ancestral populations of the *Sycopsis* lineage. The rDNA ITS data recognize two major clades in the eastern Asian genus *Corylopsis*; one consists of *C. multiflora*, *C. pauciflora*, *C. glabrescens*, *C. willmottiae*, and *C. oemeiensis* and the other contains *C. sinensis* (and varieties), *C. hypoglauca*, and *C. spicata*. Series *Spicatae* is polyphyletic.

<sup>3</sup>Address for Correspondence: Arnold Arboretum and Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA 02138, USA. Tele: +617-496-6429; Fax: +617-495-9484; E-mail: jli@oeb.harvard.edu

**Key Words:** Saxifragales, *matK*, *trnL-F*, rDNA ITS, *GBSSI*, Fothergilleae, *Corylopsis*

Hamamelidaceae (Witch-hazel family) consists of 28 genera and ca. 120 species, which are distributed in Africa, southeast and central Asia, North, Central, and North America, and northern South America, with the centers of diversity in the tropics and subtropics of eastern Asia (53 species) and Indo-China (24 species) [18, 60]. Many genera are regional endemics (e.g., *Sinowilsonia* Hemsl. and *Fortunearia* Rehd. & Wils. in central to northern China, and *Noahdendron*, Endress, Hyland & Tracey, *Ostreaaria* Baill., and *Neostreaaria* Smith in northeastern Queensland, Australia), while *Hamamelis* L. is an intercontinental disjunct genus between eastern Asia and eastern North America [4, 25, 37, 58, 59]. The earliest reliable fossils of Hamamelidaceae date back to the late Cretaceous [19, 39, 40]. Other fossil reports of Hamamelidaceae indicate that the family had become widely distributed by the early Tertiary in Europe, Asia, and North America [42, 55, 56].

Most traditional classification systems have placed Hamamelidaceae (including Altingiaceae) in the “Lower” Hamamelidae, which includes Cercidiphyllaceae, Tetracentraceae, Trochodendraceae, Daphniphyllaceae, Platanaceae, Myrothamnaceae, and Eupteleaceae [10, 54]. Cladistic analyses of morphological characters have suggested that Hamamelidaceae are a natural group closely linked with hamamelids and some elements of rosids and asterids [27, 28]. Based on molecular data [9, 26], APG [1] separated Altingiaceae from Hamamelidaceae and placed both families in Saxifragales, which also included Cercidiphyllaceae, Crassulaceae, Daphniphyllaceae, Grossulariaceae, Haloragaceae, Iteaceae, Paeoniaceae, Penthoraceae, Pterostemonaceae, Saxifragaceae, and Tetracarpaeaceae. Although it is unclear which family or families are sister to Hamamelidaceae, the latter forms a weakly-supported clade with the inclusion of *Exbucklandia* R. W. Brown and *Rhodoleia* Champ. [21, 22, 49], which some have considered as a separate family [14].

Within Hamamelidaceae, many classification systems have been proposed [2, 3, 5, 8, 17, 24, 43, 44, 46, 57]. The first comprehensive classification system divided the family into four subfamilies, the largest of which (Hamamelidoideae) was further divided into five tribes [24]. Chang [7, 8] erected a new subfamily Mytilarioideae for *Mytilaria* Lec. and *Chunia* Chang [6]. Endress [17] recognized three subfamilies: Exbucklandioideae, Rhodoleioideae, and Hamamelidoideae. His

Exbucklandioideae was a combination of Exbucklandioideae Harms, Disanthoideae Harms, and the Mytilarioideae Chang. Within Hamamelidoideae, Endress [16, 17] united the tribes Fothergilleae *sensu* Harms and Distylieae *sensu* Harms, forming the tribe Fothergilleae s.l. The most recent classification system proposed was based on phylogenetic relationships, recognizing five monophyletic subfamilies in Hamamelidaceae [32] and six monophyletic tribes in the subfamily Hamamelidoideae [31]. This phylogenetically based classification system redefined the tribes; for example, Corylopsideae comprised a single genus to the exclusion of *Fortunearia* and *Sinowilsonia* [31, 38]. *Matudaea* Lundell and *Molinadendron* Endress were excluded from the apetalous tribe Fothergilleae [30].

Within Hamamelidaceae, Exbucklandioideae and Rhodoleioideae form a clade that is sister to the clade of Mytilarioideae, Disanthoideae, and Hamamelidoideae, the latter two are sister to each other [34, 47]. Within Hamamelidoideae, Loropetaleae and Corylopsideae form a clade that is sister to the clade containing Eustigmatae + Dicorypheae + Hamamelideae + Fothergilleae [31]. *Chunia*, a genus endemic to a small area of Hainan Island of China [6], has not been included in a broad scale phylogenetic study [48].

Sequences of the internal transcribed spacers of (ITS) nuclear rDNA showed that *Shaniodendron* Deng, Wei and Wang, a segregate genus from *Hamamelis* [12, 13], was most closely related to *Parrotia* Mey., a western Asian monotypic genus [23, 30]. Based on this, Hao and Wei [23] merged *Shaniodendron* into *Parrotia*. In the chloroplast gene tree, however, *Shaniodendron* was not clustered with *Parrotia* [35]. Nonetheless, neither relationship was well supported. Additional studies are warranted to examine generic relationships of the tribe Fothergilleae.

*Corylopsis* Sieb. and Zucc. is a monophyletic genus characterized by the clawed, circular petals (versus spathulate, strap-shaped, reduced, or apetalous in other Hamamelidaceae). More than 36 species have been published since the establishment of the genus in 1835. Morley and Chao [41], however, recognized only nine. Chang divided the species of *Corylopsis* into two sections: sect. *Henryana* and sect. *Corylopsis*, which included three series (*Multiflorae*, *Pauciflora*e, and *Spicatae*). Li et al. [33] conducted a taxonomic revision of *Corylopsis*, recognizing 11 species. A phylogenetic framework is needed to test previous taxonomic treatments of the genus, and examine evolutionary patterns of morphological characters (e.g., staminodes).

In this chapter, molecular phylogenetics of Hamamelidaceae has been discussed based on sequences of nuclear and chloroplast genes, focusing on the following questions:

- (1) Are Hamamelidaceae monophyletic with the inclusion of *Exbucklandia* and *Rhodoleia* in a phylogenetic context of Saxifragales?
- (2) Should *Shaniodendron* and *Parrotia* be merged?
- (3) What are the interspecific relationships within *Corylopsis*?

## MATERIALS AND METHODS

Fifty-five species were included in this study (Table 1), representing all genera of Hamamelidaceae, except for *Embolanthera* Merr., (for which DNA material was not available), all families of Saxifragales and outgroups [22], and nine species of *Corylopsis* [33, 41]. DNAs were extracted from fresh or silica gel dried leaf tissue using either standard CTAB method [15] or a Qiagen DNeasy Plant Mini Kit (Qiagen, CA). Protocols for PCR (polymerase chain reaction) and sequencing for rDNA ITS, chloroplast noncoding region *trnL-F*, and GBSSI were as described by Li et al. [35], Li et al. [36], and Li et al. [37], respectively. Sequences of the chloroplast regions were obtained via direct sequencing of the PCR products. However, for nuclear regions (rDNA ITS and GBSSI), cloning was carried out to detect sequence polymorphisms using the standard T-A tail method with either Promega pGEM®-T Easy Vector Systems (Madison, WI) and Stratagene *E. Coli* XL1-Blue competent cells, or a StrataClone PCR Cloning Kit (Cedar Creek, TX). Five to 10 clones were sequenced for each sample to detect allelic variation.

Sequences of all regions were aligned using Clustal W in MEGA software package (version 3.1, [29], and then adjusted manually. Phylogenetic analyses were conducted using both maximum parsimony (MP) and maximum likelihood (ML) methods, as implemented in PAUP\* (version 4.0b, [53]). Characters were equally weighted and their state transformations were ordered. Gaps were treated as missing data or coded as presence (1) or absence (0) and added to the data set using Gapcoder program (Rick Ree, available at [http://www.maen.huh.harvard.edu:8080/services/gap\\_recoder](http://www.maen.huh.harvard.edu:8080/services/gap_recoder)). MP analyses were conducted for the chloroplast *matK* and *trnL-F* data set of Saxifragales using a heuristic search with random sequence addition of 1,000

**Table 1.** Samples and their DNA sequence accessions used in this study. AA, Arnold Arboretum; Sequences obtained in this study are in boldface

Taxa	Voucher and source	matK	t <sub>rnl</sub> -F	ITS	GBSSI
<b>Hamamelidaceae</b>					
<i>Corylopsis pauciflora</i> S.& Z.	AA 101-56 (A); Japan			U65462	
<i>C. sinensis</i> Hemsl. 1	AA 642-81B (A); China	AF013038	AF14765	U65461	
<i>C. sinensis</i> 2	AA 642-81C (A); China			EF456711	
<i>C. sinensis</i> var. <i>glandulifera</i> Rehd. & Wils.	AA 267-77A (A); China			a. EF456718	
<i>C. sinensis</i> var. <i>veitchiana</i> Morley & Chao	AA 391-65 (A); China			b. EF456719	
<i>C. sinensis</i> var. <i>calvescens</i> Rehd. & Wils.	AA 641-81B (A); China			EF456710	
<i>C. spicata</i> S.& Z. 1	AA 898-81B (A); Japan			EF456713	
<i>C. spicata</i> 2	AA 10544 (A); Japan	AF013039		U65463	
<i>C. spicata</i> 3	AA 157-81A (A); Japan			a. EF456716	
<i>C. glabrescens</i> Franchet & Savatier 1	AA 898-81A (A); Japan			b. EF456717	
<i>C. glabrescens</i> 2	AA 1030-90B (A); Japan			EF456714	
<i>C. hypoglauca</i> Cheng 1	AA 557-88A (A); China			EF456715	
<i>C. hypoglauca</i> 2	AA 557-88B (A); China			EF456708	
<i>C. willmottiae</i> Rehd. & Wils.	Li 3297 (SZ), Mt. Emei, China			EF456709	
				EF456720	

Table 1 contd.

Table 1 contd.

<i>C. omeiensis</i> Yang	Li 3298 (SZ), Mt. Emei, China			EF456721
<i>C. multiflora</i> Hance 1	Donglin Zhang s.n., Hunan, China			AB236983 EF456722
<i>C. multiflora</i> 2				
<i>C. yunnanensis</i> Diels = <i>C. himalayana</i> Griffith	Zhong Yi, s.n., Hainan, China	AF108466	EF456729	AB236985
<i>Chunia bucklandioides</i>		AF013040	AF14766	AF015419-20
<i>Dicoryphe stipulacea</i> Janum St. Hil.		U77091	AF14767	AF015421-22
<i>Disanthus cecropioides</i> Max.	Li 1101, Woodlanders, Inc. SC.	AF013042	AF14770	AF019231-2
<i>Ditsylophopsis tuckeri</i> Endress				EF456723
<i>Distylium myricoides</i> Hemsl.	Li 80, Woodlanders, Inc. SC.	AF013041	AF14768	U65464
<i>Distylium racemosum</i> S. & Z.		AF14769	U65465	EF456725
<i>Eustigma oblongifolium</i> Gardn. & Champ.		AF013043	AF14771	U65466
<i>Exbucklandia populnea</i> (R.Br.) R.W.Bt.				AF014772 AF14773 U65467
<i>Fortunearia sinensis</i> R. & W.		AF013044	AF14774	AF015425-26
<i>Fothergilla major</i> Lodd.	Li 1161, UNH Campus	AF013045	AF14775	AF015656-57
<i>Hamamelis virginiana</i> L.	Li 126, UNH Campus	AF013046	AF14776	AF015427-28
<i>Loropetalum chinense</i> (R.Br) Oliv.		AF013059	AF14776	

Table 1 contd.

Table 1 contd.

<i>Mangyea malayana</i> Oliv.		AF025393	AF14777	AF022241		
<i>Matudaea trimervia</i> Lund.		AF013048	AF14778	AF015437		
<i>Molinadendron guatemalense</i> Endress		AF013049	AF14779	AF015438		
<i>Mytilaria laosensis</i> Lec.		U77093	<b>EF456731</b>			
<i>Neostrearia fleckeri</i> Smith		AF013050	AF14780	AF015439		
<i>Noahdendron nicholasi</i> Endress, Hyland & Tracey		AF013051	AF14781	AF015440		
<i>Ostreastralia</i> Baill.		AF013052	AF14782	AF015441		
<i>Parrotia persica</i> C.A.Mey		AF013053	AF14783	AF015443	<b>BF456726</b>	
<i>Parrotiopsis jacquemontiana</i> Rehd.		AF013054	AF14784	AF015442	<b>EF456727</b>	
<i>Rhodoleia championii</i> Hook. f.		U77094	AF14785	AF015429-30		
<i>Shaniodendron subaequale</i> Deng, Wei & Wang		AF013055	AF14786	AF015431-32	<b>EF456728</b>	
<i>Sinowilsonia henryi</i> Hemsl.		AF013056	AF14787	U65468		
<i>Sycopsis sinensis</i> Oliv.		AF013057	AF14788	AF015433-34	<b>EF456724</b>	
<i>Tetrathyrium subcordatum</i> Benth.			AF14789	AF022242		
<i>Trichocladus crinitus</i> Pers.			AF14790	AF019233		
<b>Altingiaceae</b>						
<i>Altinigia excelsa</i>		AF013037	DQ352225			
<i>Liquidambar styraciflua</i>		AF015652	DQ352217			
<b>Cercidiphyllaceae</b>						
<i>Cercidiphyllum japonicum</i>		AM1396508	AF200928			

Table 1 contd.

Table 1 contd.

Daphniphyllaceae				
<i>Daphniphyllum macropodium</i>			AF200935	
Saxifragaceae and alliance				
<i>Itea virginica</i>	<b>EF456732</b>	AF374818		
<i>Sullivantia oregana</i>	L34113	AF374811		
<i>Ribes aureum</i>	L34153	AF374816		
<i>Heuchera rubescens</i>	L34127	AF374806		
Haloragiaceae				
<i>Haloragis aspera</i>	AF274616			
<i>Pennhorum sedoides</i>	AF274628			
Pterostemonaceae				
<i>Pterostemon rotundifolius</i>	AF274630	AF37481		
Crassulaceae				
<i>Crassula rupestris</i>	AF115602	X71984		
<i>Kalanchoe uniflora</i>	AF115623	X71986		
<i>Tylecodon wallichii</i>	AF115590	AY692310		
Paeoniaceae				
<i>Paonia suffruticosa</i>	AF033593	<b>EF456730</b>		
<i>Paonia lutea</i>	AF033590	AY328201		
Outgroups				
<i>Tetracentron sinense</i>	AM396504	AF200961		
<i>Vitis vinifera</i>	AJ429274	AF300295		

replicates with 10 trees held per replicate, TBR (tree-bisection-reconnection) branch swapping, and steepest descent option off. For the Fothergilleae and *Corylopsis* data sets, exhaustive or branch and bound tree searches were performed with default options in PAUP\*. Bootstrap analyses of 500 replicates were done to estimate support for individual clades and the tree search options were as MP analyses, except for simple sequence addition [20]. Modeltest (version 3, [45]) was used to determine the best model for the data sets, and the estimated parameters were then used in ML analyses. Heuristic tree search was used in the ML analyses with options as in the MP analyses, except for simple sequence addition.

## RESULTS

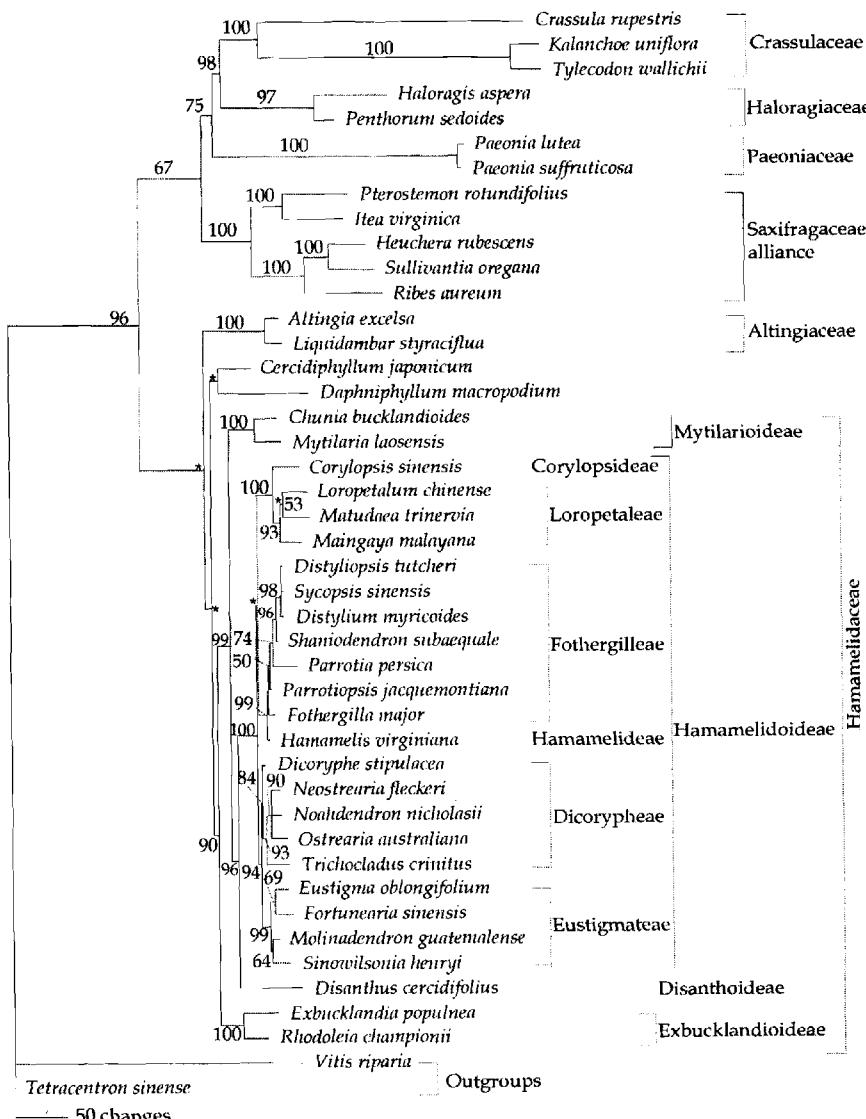
While most sequences used in this study were retrieved from the previous work as well as the GenBank, 25 sequences were newly obtained and their accession numbers are given in Table 1.

Sequence alignment of the *matK* and *trnL-F* regions in Saxifragales generated a data set of 2,772 characters, 523 of which were parsimony informative. When gaps were treated as missing data, MP analyses generated 24 trees of 2,242 steps on one island, a consistency index (CI) of 0.71, and retention index (RI) of 0.71. One of the trees is presented here (Fig. 1). There were five major clades in Saxifragales:

- (1) Daphniphyllaceae,
- (2) Cercidiphyllaceae,
- (3) Altingiaceae (bootstrap support, bs = 100%),
- (4) Hamamelidaceae (bs = 90%), and
- (5) Crassulaceae + Haloragiaceae + Paeoniaceae + Saxifragaceae alliance (bs = 67%).

Relationships among the clades, however, were not resolved in the consensus tree. Within Hamamelidaceae, *Exbucklandia* and *Rhodoleia*, *Mytilaria* and *Chunia*, *Disanthus* Maxim. and Hamamelidoideae formed consecutive clades with strong support (bs = 96-100%). Three major clades were recognized in Hamamelidoideae:

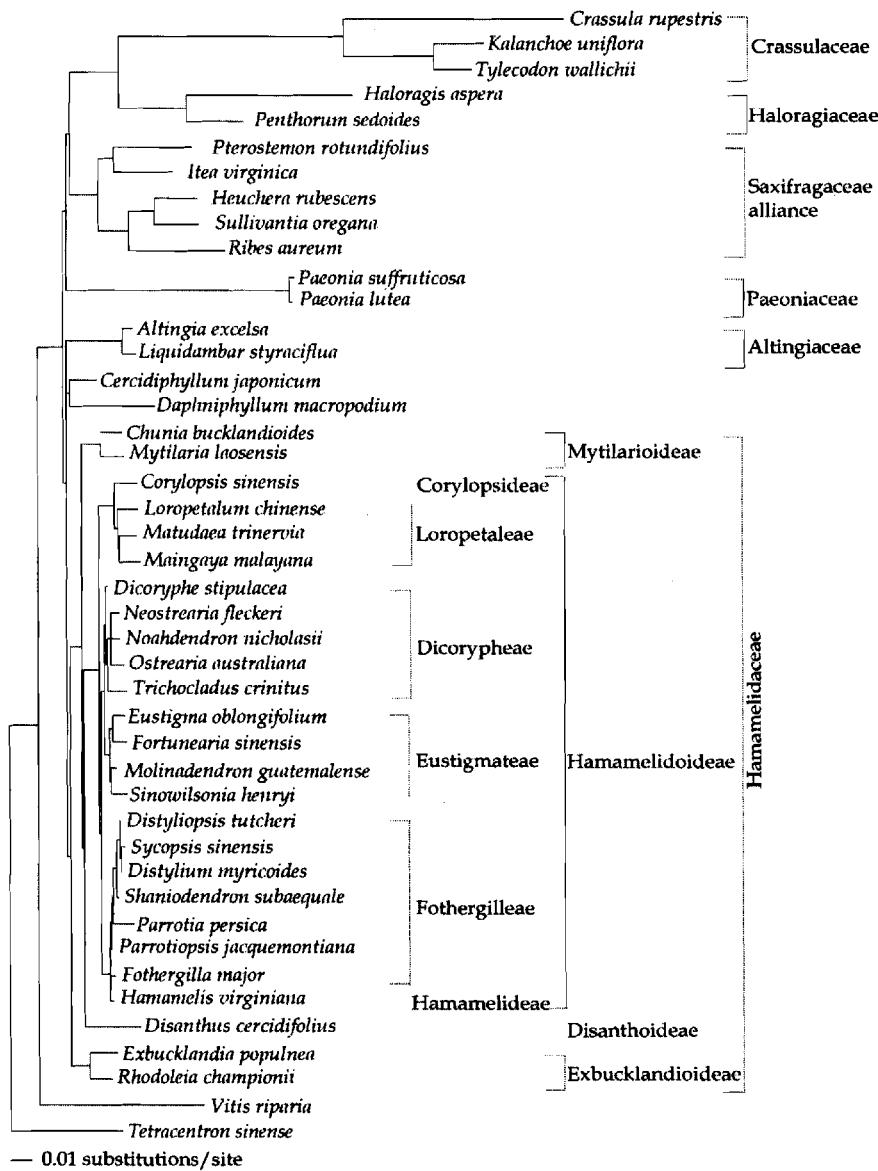
- (1) *Hamamelis* + Fothergilleae (bs = 99%),
- (2) Eustigmataeae + Dicorypheae (bs = 94%), and
- (3) *Corylopsis* + Loropetaleae (bs = 100%).



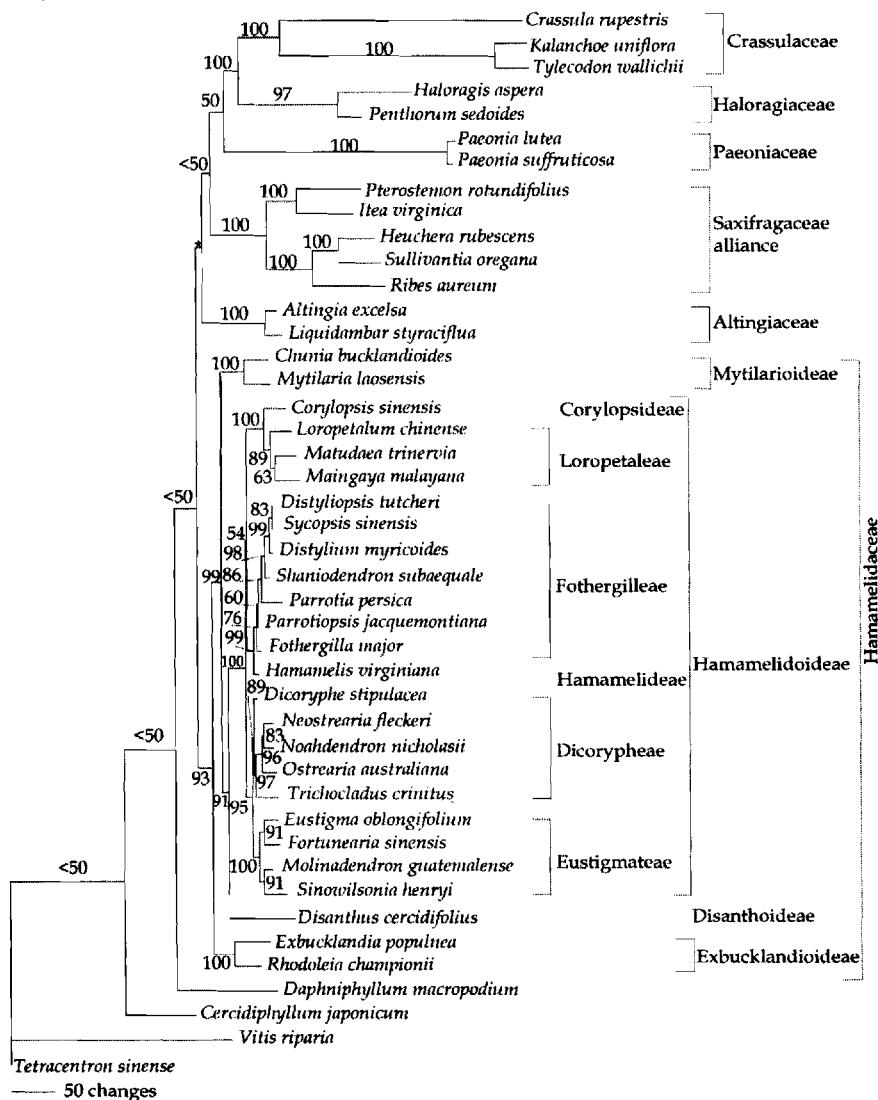
**Fig. 1.** One of 24 parsimonious trees of 2,242 steps based on sequences of matK and trnL-F with gaps as missing data. Numbers at branches are bootstrap support > 50%, otherwise not shown. CI = RI = 0.71. Asterisks indicate nodes absent in the strict consensus tree.

*Penthorum* L. and *Haloragis* Forster & Forster f. formed a clade (bs = 97%) sister to Crassulaceae (bs = 98%), together they were sister to Paeoniaceae (bs = 75%). The Saxifragaceae alliance formed a well-supported clade (bs = 100%). Hierarchical likelihood ratio test (hLRT) indicated that the best fit model for *matK* and *trnL-F* was the GTR+Gamma model, and the estimated rates of substitutions were A to C = 1.35, A to G = 1.90, A to T = 0.28, C to G = 0.81, C to T = 1.90, and G to T = 1.00, and Gamma shape parameter ( $\alpha$ ) = 1.07. ML analyses with the estimated parameters produced a single tree (Fig. 2), whose topology was congruent with Fig. 1, except that *Cercidiphyllum* Sieb. & Zucc. and *Daphniphyllum* Blume formed a clade, and together with Altingiaceae, they formed a clade with Hamamelidaceae, and that Paeoniaceae was sister to the clade containing the Saxifragaceae alliance + Haloragiaceae + Crassulaceae. When gaps were coded as presence (1) or absence (0) and were included in the analyses, MP analyses produced two trees of 2,700 steps, the first of which was shown in Fig. 3 (CI = 0.70, RI = 0.71). The tree topology was similar to Fig. 1, with the exceptions that Cercidiphyllaceae was sister to the remaining Saxifragales, and Altingiaceae sister to Saxifragaceae alliance + Paeoniaceae + Haloragiaceae + Crassulaceae. However, the relationships were weakly supported. The second tree (not shown) differed from the first (Fig. 3), in that *Daphniphyllum* was sister to the clade containing Altingiaceae, Saxifragaceae alliance, Paeoniaceae, Haloragiaceae, and Crassulaceae.

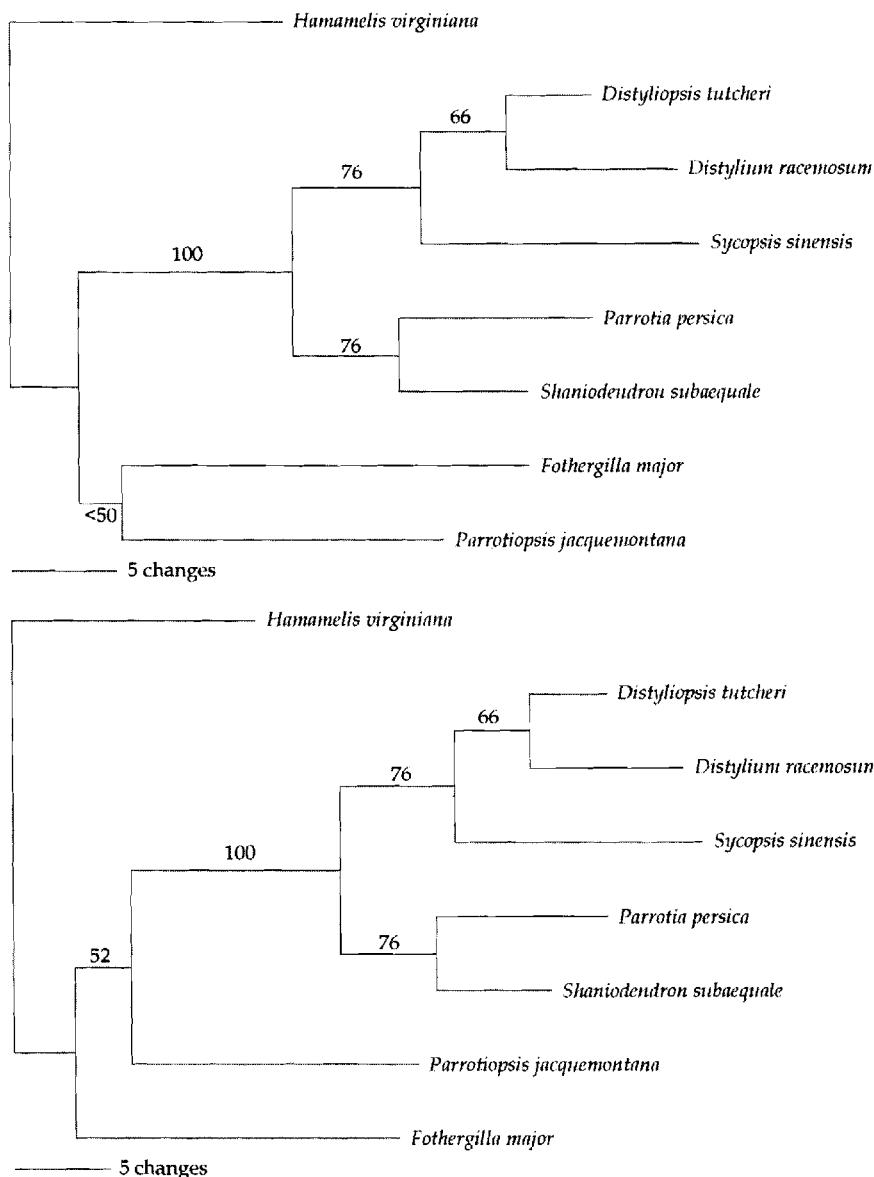
Within Fothergilleae, the chloroplast sequence data suggested identical relationships regardless of how gaps were treated (Figs. 1 and 2). *Fothergilla* Murray diverged first and was followed by *Parrotiopsis* Schneider, *Parrotia*, *Shaniodendron*, *Distylium* Sieb. & Zucc., and *Sycopsis* Oliv. + *Distyliopsis* Endress. For nuclear GBSSI, a single sequence was used for each species because there was little sequence variation among clones of the gene within a species. A partition homogeneity test performed in PAUP\* indicated that sequences of GBSSI and rDNA ITS were congruent ( $P = 0.41$ ). The combined nuclear data set had 8 species and 1,241 characters, 27 of which were parsimony informative. MP analyses with gaps treated as missing data and exhaustive tree search generated two trees of 235 steps (Fig. 4), a CI = 0.91, and RI = 0.76. *Fothergilla* and *Parrotiopsis* formed a clade in one tree (Fig. 4a), but not in the other (Fig. 4b). *Distyliopsis*, *Distylium*, and *Sycopsis* formed a clade, as also *Shaniodendron* and *Parrotia*. The clades received only moderate



**Fig. 2.** ML tree based on sequences of matK and trnL-F and GTR+G model with parameters estimated using Modeltest.



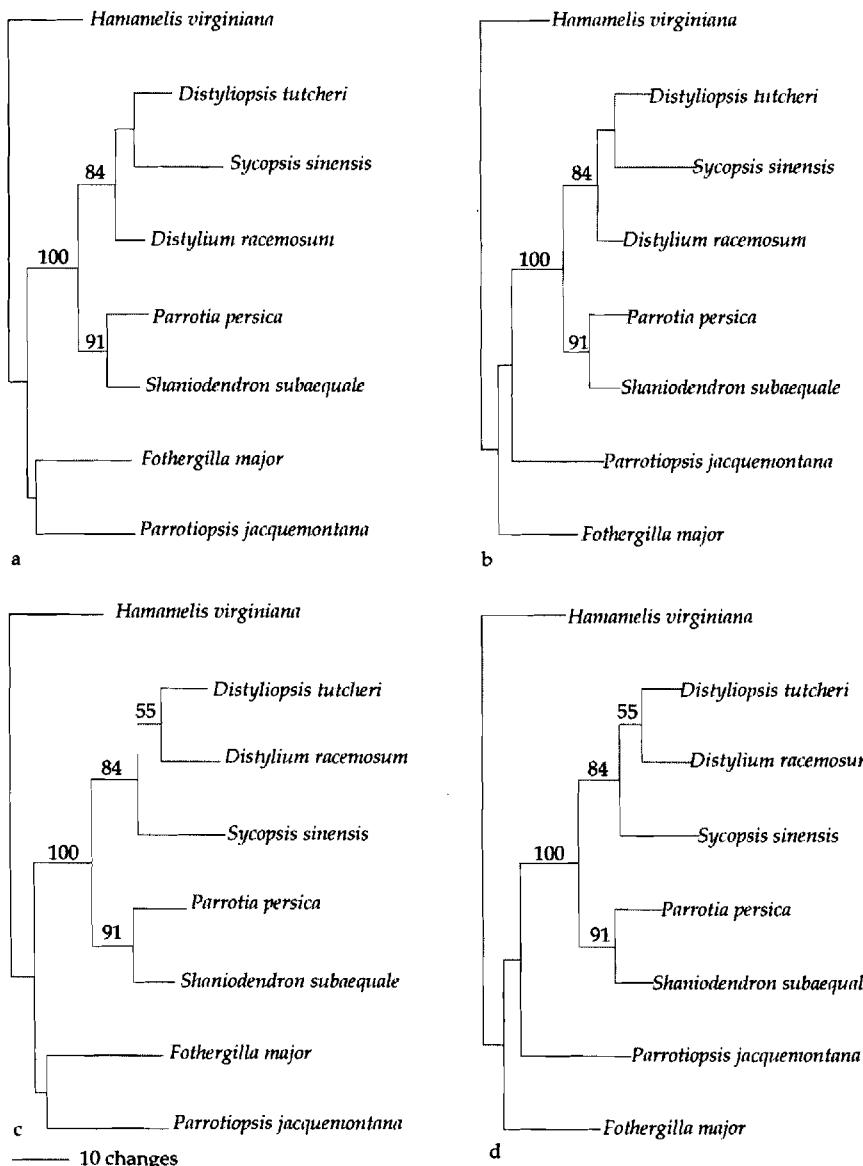
**Fig. 3.** One of two parsimonious trees of 2,700 steps based on matK and trnL-F regions with gaps coded as presence (1) or absence (0) and included in the data set. Numbers at branches are bootstrap support > 50%, otherwise not shown. CI = 0.70, RI = 0.71. Asterisk indicates the branch to which *Daphniphyllum macropodium* is attached in the second tree (not shown).



**Fig. 4.** Parsimonious trees of 235 steps based on sequences of rDNA ITS and GBSSI. Numbers at branches are bootstrap percentages ( $CI = 0.91$ ,  $RI = 0.76$ ).

support ( $bs = 76\%$ ). However, together the two clades formed a strongly supported group ( $bs = 100\%$ ). The best evolutionary model for the nuclear regions was the K2P + Gamma model, as determined by the hLRT test. The estimated parameters were as follows: transition to transversion ratio = 1.880 and Gamma shape parameter = 0.015. ML analyses using the estimated parameters generated a single tree with a  $-Ln$  likelihood = 2,472.72, and the tree topology was congruent with Fig. 4b, but had higher bootstrap support ( $bs = 81\%$ ) for the sister relationship of *Parrotia* and *Shaniodendron* (tree not shown). When gaps were coded as presence (1) and absence (0) and appended to the data set, 31 characters were parsimony informative. MP analyses produced 4 trees of 141 steps (Fig. 5, CI = 0.91, RI = 0.76). These trees differed in the systematic positions of *Parrotiopsis* and *Fothergilla*, and in relationships among *Distyliopsis*, *Distylium*, and *Sycopsis*. Nevertheless, the close relationship of *Shaniodendron* and *Parrotia* was strongly supported ( $bs = 91\%$ ).

The rDNA ITS data set of *Corylopsis* contained 21 samples representing nine species of *Corylopsis* and *Loropetalum chinense* as the outgroup. *Corylopsis henryi* Hemsley and *C. yui* Hu & Cheng were not available for the study. Sequence polymorphisms of the ITS region were not observed in any of the samples except for *C. sinensis* var. *glandulifera*, which had two different sequence types. There were 131 variable characters, 50 of which were parsimony informative. MP analyses of the ITS data produced 39 trees of 154 steps, one of which is shown here (Fig. 6). Species of *Corylopsis* formed two major clades, one consisting of *C. sinensis* + *C. spicata* + *C. hypoglauca* and the other containing *C. pauciflora*, *C. glabrescens*, *C. multiflora*, *C. willmottiae*, *C. omeiensis*, and *C. himalayana*. In *Corylopsis sinensis* var. *glandulifera*, one sequence type of the rDNA ITS was grouped with the other samples of the same species, while the other type was allied with two samples of *C. hypoglauca*. The optimal model of sequence evolution was GTG + Gamma model as determined by the Modeltest and the estimated parameters were as follows: Gamma shape parameter = 0.7; substitution rates A to C = 1.0 A to G = 1.5, A to T = 1.0, C to G = 1.0, C to T = 3.3, and G to T = 1.0. ML analyses based on the estimated parameters produced a single tree with  $-Ln$  Likelihood = 1,759.85. The tree topology was congruent with Fig. 6, except for the collapse of the clade containing *C. pauciflora* and *C. glabrescens* (tree not shown).



**Fig. 5.** Parsimonious trees of 141 steps produced from the combined data set of rDNA ITS and GBSSI with gaps coded and added to the data set. Numbers at branches are bootstrap percentages ( $CI = 0.91$ ,  $RI = 0.76$ ).

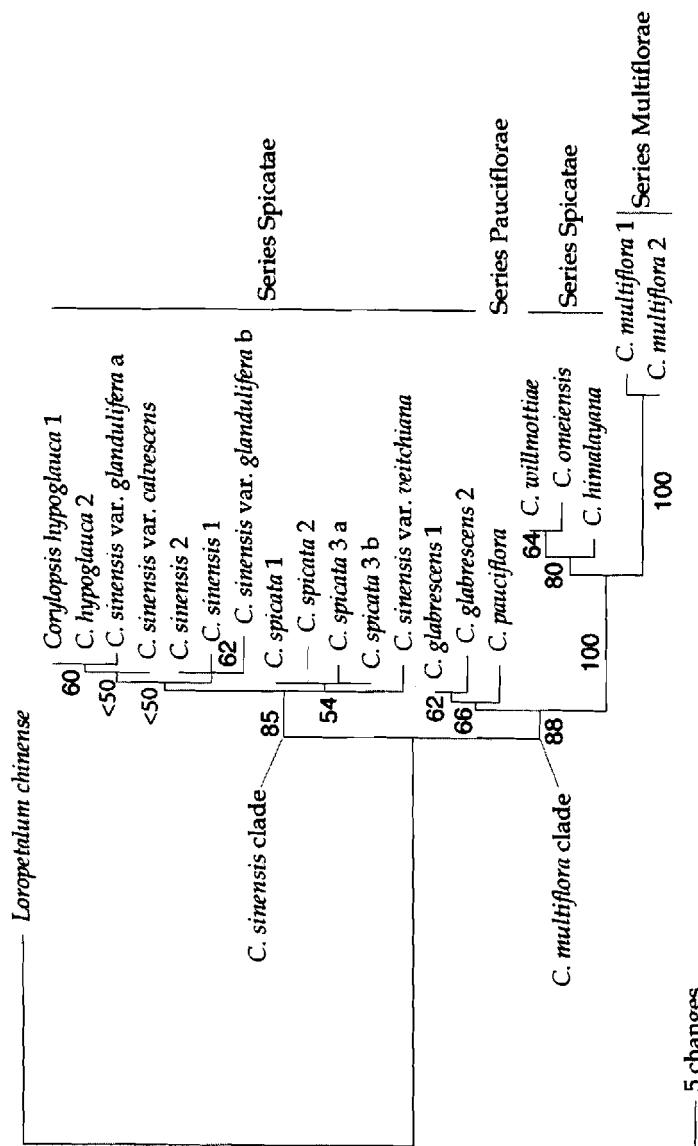


Fig. 6. One of 39 trees of 154 steps based on parsimony analyses of rDNA ITS data of *Corylopsis*. Numbers at branches are bootstrap percentages ( $CI = RI = 0.92$ ). Letters after species name represent different sequence types.

## DISCUSSION

### Relationships of Saxifragales and the Monophyly of Hamamelidaceae

Molecular systematics has revolutionized the understanding of the evolution of organisms. One of the most surprising results in molecular systematics of angiosperms has been the composition of Saxifragales [9, 11, 22, 26, 51, 52], which includes many taxa that have not been considered as closely related. They have been placed in three subclasses in recent classifications: Altingiaceae, Hamamelidaceae, Cercidiphyllaceae, and Daphniphyllaceae in subclass Hamamelidae [10, 54], Saxifragaceae, Iteaceae, Pterostemonaceae, Peridiscaceae, Grossulariaceae, Crassulaceae, Penthoraceae in subclass Rosidae, and Paeoniaceae in Magnoliidae or Dilleniidae [10]. Molecular phylogenetic analyses have recognized a few robust clades in Saxifragales, including Haloragiaceae and Crassulaceae, Pterostemonaceae and Iteaceae, and Grossulariaceae and Saxifragaceae. The latter four families as a robust clade comprise the Saxifragaceae alliance [21, 22, 50]. However, relationships among the lineages within Saxifragales have not been well-resolved, the lack of a satisfactory resolution has been attributed to a simultaneous diversification of lineages [22]. Nevertheless, rigorous model-based phylogenetic analyses of the same data sets appear to reveal more phylogenetic structures within the group. For example, Paeoniaceae is sister to the clade consisting of Saxifragaceae alliance, Iteaceae, Pterostemonaceae, Haloragiaceae, and Crassulaceae. Altingiaceae, Cercidiphyllaceae, and Daphniphyllaceae may be most closely related to Hamamelidaceae [21]. The MP analyses of the *matK* + *trnL-F* data set support the clades within Saxifragales (Figs. 1 and 2), and the close relationship of Hamamelidaceae with Altingiaceae, Cercidiphyllaceae and Daphniphyllaceae. However, Paeoniaceae are sister to the clade containing Crassulaceae and Haloragiaceae (Fig. 1, bs = 75%). The ML analyses of the data set, nevertheless, produce results that agree with Fishbein and Soltis [21].

*Exbucklandia* and *Rhodoleia* have sometimes been excluded from Hamamelidaceae [14]. Previous phylogenetic analyses provide weak support for the inclusion of the two genera in Hamamelidaceae [22]. In the *matK* and *trnL-F* phylogenies (Figs. 1 and 2), *Exbucklandia* and *Rhodoleia* form a group and together they are in a robust clade with the

remaining Hamamelidaceae, supporting their inclusion in Hamamelidaceae.

## Relationships Within Hamamelidaceae

When rooted with Altingiaceae, the rDNA ITS phylogeny shows that *Exbucklandia*, *Rhodoleia*, and *Mytilaria* form a weakly supported clade [36]. In the *matK* tree, however, *Exbucklandia* and *Rhodoleia* form a well-supported clade, whereas *Mytilaria* is positioned between the *Exbucklandia*-*Rhodoleia* clade and *Disanthus* + Hamamelidoideae [35]. Nevertheless, relationships are only moderately supported (bs = ca. 75%). In the *matK* + *trnL-F* trees, *Exbucklandia* and *Rhodoleia* form a clade, followed by consecutive clades of *Chunia* + *Mytilaria*, *Disanthus*, and Hamamelidoideae. All relationships are well supported (bs > 90%). *Chunia* and *Mytilaria*, two endemic, monotypic genera from China, have been considered as sole members of Mytilarioideae [7, 8]. The *matK* and *trnL-F* data support their close relationship (Fig. 1), and the conclusion from rDNA ITS sequences [49]. *Disanthus* and Hamamelidoideae are sister taxa, which is consistent with Endress [16]. Within Hamamelidoideae, the previous analyses [31] based on the combined data of morphology and DNA sequences of *matK*, *trnL-F*, and rDNA ITS, produced a robust phylogenetic tree containing six clades (or tribes): *Corylopsideae*, *Loropetaleae*, *Eustigmataeae*, *Dicorypheae*, *Hamamelideae*, and *Fothergilleae*. The clades hold well in this study where the family was analyzed in the broad context of Saxifragales and the rDNA ITS data were not included (Figs. 1, and 2). However, relationships among the three major clades (*Corylopsideae* + *Loropetaleae*, *Eustigmataeae* + *Dicorypheae*, and *Hamamelideae* + *Fothergilleae*) were not resolved.

*Shaniodendron*, a segregate genus from *Hamamelis*, was first described by Deng et al. [12]. The rDNA ITS data suggest that the segregate genus is most closely related to *Parrotia* [31]. Chloroplast data, however, indicate that *Shaniodendron* is more closely related to the clade of *Distyliopsis* + *Distylium* + *Sycopsis* than to *Parrotia* [35, 36]. The nuclear GBSSI data provide further support for the sister relationship of *Parrotia* and *Shaniodendron* (Fig. 5), and consequent merger of the two genera [23]. Morphologically, the two genera share many characters (e.g., exfoliating bark and deciduous leaves). Therefore, it is possible that *Shaniodendron* has captured the chloroplast genomes of the ancestral lineages of *Sycopsis*, *Distylium*, and *Distyliopsis* via ancient hybridization.

However, it warrants a genome level study of the genera to test the hypothesis.

*Corylopsis* is an eastern Asian genus with 11 species, three in Japan and adjacent islands, and eight on continental Asia [33]. Following Harms [24], Chang (1979) recognized three series in *Corylopsis*: *Multiflorae*, *Pauciflorae*, and *Spicatae*. In the ITS tree (Fig. 6), species of series *Spicatae* are distributed on three separate branches, suggesting that the intrageneric classification of *Corylopsis* needs reconsideration. Staminodes are diverse in number and morphology in *Corylopsis* and may have important systematic value [33]. In *C. multiflora* and *C. pauciflora*, there are five columnar, truncate staminodes, while in flowers of *C. himalayana*, *C. willmottiae*, and *C. omeiensis* there exist ten columnar staminodes. In the ITS tree, these species form a robust clade including *C. glabrescens* (Fig. 6), which has five scale-like, bifid staminodes, a character seen in the remaining species of *Corylopsis* (*C. sinensis*, *C. hypoglauca*, and *C. spicata*). Therefore, staminodes appear to have a complex evolutionary history in *Corylopsis*. Relationships among species of the *C. multiflora* clade are better-resolved than those in the *C. sinensis* clade. One possible explanation is that these species may have recently diverged from one another. All species other than *C. spicata* have been generally considered as infraspecific taxa of *C. sinensis*. These results are consistent with the taxonomic treatment but a more comprehensive taxon sampling is needed at the population level to further test the hypothesis.

The phylogenetic analyses in the broad context of Saxifragales recognize the monophyly of Hamamelidaceae with the inclusion of *Exbucklandia* and *Rhodoleia*, and further support the close relationship of *Chunia* and *Mytilaria*. Within the tribe Fothergilleae, in the chloroplast gene tree *Shaniodendron* appears to be more closely related to *Distylium* + *Distyliopsis* + *Sycopsis* than *Parrotia*. In contrast, in the nuclear gene trees, *Shaniodendron* and *Parrotia* form a robust clade, which agrees with morphology. This suggests that there might be chloroplast capture in *Shaniodendron* from the ancestral population of the *Sycopsis* clade. The rDNA ITS data resolve two clades in *Corylopsis*, the first being species with columnar staminodes, except for *C. glabrescens*, and the other consisting of all species with bilobed, scale-like staminodes.

## Acknowledgment

The author thanks Drs. Arun Kumar Sharma and Archana Sharma for inviting him to contribute to this volume. The author is also grateful to Suzanne Shoup for assistance in lab work, and Dr. A. Linn Bogle for reading the initial manuscript and providing valuable comments.

## References

- [1] APG. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc* 2003; 141: 339-436.
- [2] Bentham G' Hooker JD. *Gen. Pl. Hamamelideae*. Vol. 1. London, UK: 1865: 664-669.
- [3] Bogle AL. Floral vascular anatomy and the nature of the hamamelidaceous flower. Minneapolis, MN: University of Minnesota. 1968. Available from Proquest/UMI Dissertation Publishing, Ann Arbor, MI. Publication No. AAT 6812241. 335 p. in *Department of Botany and Plant Pathology*. 1968, University of Minnesota: Minn.
- [4] Boufford DE, Spongberg SA. Eastern Asian-Eastern North American phytogeographical relationships - a history from the time of Linnaeus to the twentieth century. *Ann Missouri Bot Gard* 1983; 70: 423-439.
- [5] Candolle, APD. *Prodr. Hamamelideae*. Vol. 4. Paris. p: 1830: 267-270.
- [6] Chang HT. A new genus of Hamamelidoideae in the flora of China. *Sunyatsen Univ Bull Nat Sci* 1962; 1: 34-44.
- [7] —. A revision of the Hamamelidaceous flora of China. *Sunyatsen Univ Bull*. 1973; 1: 54-71.
- [8] —. *Florae Reipublicae Popularis Sinicae*. Vol. 35 (2). 1979. Beijing: Science Press, 1979; 36-116.
- [9] Change M, Soltis DE, Olmstead RG, Morgan D, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KJ, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CR, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quin CJ, Egliate LE, Golenberg E, Learn GHJ, Graham SW, Barrett SCH, Dayanandan S, Albert VA. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Missouri Bot Gard* 1993; 80: 528-580.
- [10] Cronquist A. An integrated system of classification of flowering plants. New York: Columbia University Press, 1981.
- [11] Davis CC, Chase MW. Elatinaceae are sister to Malpighiaceae; Peridiscaceae belong to Saxifragales. *Amer J Bot* 2004; 91: 262-273.
- [12] Deng MB, Wei HT, Wang XQ. *Shaniodendron*, a new genus of Hamamelidoideae from China. *Acta Phytotax. Sin.* 1992; 30: 57-61.
- [13] —, —, —, Su P, Xing JY. On the significance of the discovery of Fothergilleae in China. *J. Plant Resour Environ* 1992; 1: 30-35.
- [14] Doweld AB. Carpology, seed anatomy and taxonomic relationships of *Tetracentron* (Tetracentraceae) and *Trochodendron* (Trochodendraceae). *Annals of Botany* 1998; 82: 413-443.

- [15] Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 1987; 19: 1-15.
- [16] Endress PK. Phylogenetic relationships in the Hamamelidoideae. In: Crane PR, Blackmore S, eds. *Evolution, Systematics and Fossil History of the Hamamelidae*. Vol. 1. Introduction and "Lower" Hamamelidae, Oxford University Press: New York, 1989: 227-248.
- [17] —. A suprageneric taxonomic classification of the Hamamelidoideae. *Taxon* 1989; 38: 371-376.
- [18] —. Hamamelidaceae. In: Kubitzki K, Rohwer JG, Bittrich V, eds. *The Families and Genera of Vascular Plants. Vol. 2. Flowering Plants. Dicotyledons*, Springer-Verlag: Berlin, 1993: 322-331.
- [19] — Friis EM. *Archamamelis*, hamamelidaceous flowers from the Upper Cretaceous of Sweden. *Plant Syst Evol.* 1991; 175: 101-114.
- [20] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; 39: 783-791.
- [21] Fishbein M, Soltis DE. Further resolution of the rapid radiation of Saxifragales (angiosperms, Eudicots) supported by mixed model Bayesian analysis. *Syst Bot* 2004; 29: 883-991.
- [22] —, Hibsch-Jetter C, Soltis DE, Hufford L. *Phylogeny of Saxifragales (angiosperms, Eudicots): analysis of a rapid, ancient radiation*. *Syst Biol* 2001; 50: 817-847.
- [23] Hao RM, Wei HT. A new combination in Hamamelidaceae. *Acta Phytotax Sin* 1998; 36: 80.
- [24] Harms H. *Hamamelidoideae*. In: Engler A, Prantl K, eds. *Nat. Pflanzenfam.* 2<sup>nd</sup> ed. Engelmann: Leipzig. Vol. 18a. 1930: 330-343.
- [25] Hong DY. Eastern Asian-North American disjunctions and their biological significance. *Cathaya* 1993; 5: 1-39.
- [26] Hoot SB, Magallon S, Crane PR. Phylogeny of basal Eudicots based on three molecular data sets: *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequences. *Ann Missouri Bot Gard* 1999; 86: 1-32.
- [27] Hufford L. Rosidae and their relationships to other nonmagnoliid dicotyledons: A phylogenetic analysis using morphological and chemical data. *Ann Missouri Bot Gard* 1992; 79: 218-248.
- [28] — Crane PR. A preliminary phylogenetic analysis of the "Lower Hamamelidae". In: Crane PR, Blackmore S, eds. *Evolution, Systematics and Fossil History of the Hamamelidae*. Vol. 1, Introduction and "Lower" Hamamelidae, Oxford University Press: New York, 1989: 175-193.
- [29] Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 2004; 5: 150-163.
- [30] Li J, Bogle AL, Klein AS, Pan K. Close relationship between *Shaniodendron* and *Parrotia* (Hamamelidaceae), evidence from ITS sequences of nuclear ribosomal DNA. *Acta Phytotax Sin* 1997; 35: 481-493.
- [31] — —. A new suprageneric classification system of the Hamamelidoideae based on morphology and sequences of nuclear and chloroplast DNA. *Harvard Pap Bot* 2001; 5: 499-515.

- [32] —. A systematic study of the Hamamelidaceae based on morphological and molecular data. Durham, NH. University of New Hampshire. 1997. Available from Proquest/ UMI Dissertation Publishing, Ann Arbor, MI. Publication no. AAT 9819680. p. 291
- [33] —, Pan KY, Lu AM. A revision of *Corylopsis* (Hamamelidoideae). In Proceedings of the 60th Anniversary of Chinese Botanical Society, Beijing. 1993.
- [34] —, Bogle AL Klein AS. Phylogenetic relationships in the Hamamelidaceae: evidence from the nucleotide sequences of the plastid gene *matK*. Plant Syst Evol 1999; 218: 205-219.
- [35] —, —, —. Phylogenetic relationships of the Hamamelidaceae inferred from sequences of internal transcribed spacers (ITS) of nuclear ribosomal DNA. Amer J Bot 1999; 86: 1027-1037.
- [36] —, —, Donoghue MJ. Phylogenetic relationships of the Hamamelidoideae inferred from sequences of trn non-coding regions of chloroplast DNA. Harvard Pap Bot 1999; 4: 343-356.
- [37] —, —, Klein AS, Donoghue MJ. Phylogeny and biogeography of *Hamamelis* (Hamamelidaceae). Harvard Pap Bot 2000. 5: 171-178.
- [38] —, —, —. Phylogenetic Relationships in the *Corylopsis* Complex (Hamamelidaceae) Evidence From Sequences of the Internal Transcribed Spacers of Nuclear Ribosomal DNA and Morphology. Rhodora 1997; 99: 302-318.
- [39] Magallon S, Herendeen PS, Crane PR. *Androdecidua* Endressii gen. et sp. nov. from the Late Cretaceous of Georgia (United States): Further floral diversity in Hamamelidoideae (Hamamelidaceae). Int J Plant Sci 2001; 162: 963-983.
- [40] —, —, Endress PK. *Allonia decandra*: florae remains of the tribe Hamamelideae (Hamamelidaceae) from Campanian strata of southeastern USA. Plant Syst Evol 1996; 202: 177-198.
- [41] Morley B Chao JM. A review of *Corylopsis* (Hamamelidoideae). J Arnold Arb, 1977; 58: 382-414.
- [42] Muller J. Fossil pollen records of extant angiosperms. Bot Rev 1981; 47: 1-142.
- [43] Niedenzu F. *Hamamelidoideae*. In: Engler A, Prantl K, eds. in Nat. Pflanzenfam. III 2a, Engelmann: Leipzig, 1891: 115-130.
- [44] Oliver D. Description of three new genera of plants in the Malayan herbarium of the late Dr. A. C. Maingay. Trans Linn Soc Lond. 1872; 28: 517. T44.
- [45] Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics application note 1998; 14: 817-818.
- [46] Reinsch A. Ueber die anatomischen Verhaltnisse der Hamamelidoideae mit Rucksicht auf ihre systematische Gruppierung. Bot Jahrb 1889. 11: 347-395. Pl. 8.
- [47] Shi SH, Chang HT, Chen YQ, Qu LH, Wen J. Phylogeny of the Hamamelidaceae based on the ITS sequences of nuclear ribosomal DNA. Biochem Syst Ecol 1998; 26: 55-69.
- [48] —, Huang YL, Zhang Q, Qu LH, Zhang HD. The phylogenetic relationships between Mytilarioideae and the other subfamilies of Hamamelidaceae inferred from the ITS sequences of nuclear ribosomal DNA. Sunyatsen Act. Sci Nat Univ 1999; 38: 34-38.
- [49] Soltis DE, Soltis PS, Zanis MJ. Phylogeny of seed plants based on evidence from eight genes. Amer, J Bot 2002; 89: 1670-1681.
- [50] —, —, Endress PK, Chase MW. Phylogeny and evolution of angiosperms. Sunderland Massachusetts: Sinauer Associates, 2005.

- [51] —, —, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA, Kuzoff RK, Kron KA, Chase MW, Swensen SM, Zimmer EA, Chaw SM, Gillespie LJ, Kress WJ, Sytsma KJ. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann Missouri Bot Gard* 1997; 84: 1-49.
- [52] —, —, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Bot J Linn Soc* 2000; 133: 381-461.
- [53] Swofford DL. PAUP\* (version 4.010b). Phylogenetic analysis using parsimony (\* and other methods) Sunderland: Sinauer Associates 2002.
- [54] Takhtajan A. Diversity and classification of flowering plants. New York: Columbia Univ. Press, 1997.
- [55] Taylor DW. Paleobiogeographic relationships of angiosperms from the Cretaceous and early Tertiary of the North America area. *Bot, Rev*, 1990; 56: 279-320.
- [56] Tiffney BH. Fruits and seed dispersal and the evolution of the Hamamelidae. *Ann, Missouri Bot, Gard*, 1986; 73: 394-416.
- [57] Tong K. Studien über die Familie der Hamamelidaceae, mit besonderer Berücksichtigung der systematik und entwicklungsgeschichte von *Corylopsis*. SunYatsen Univ, Bull Biol Dept Sci Coll 1930; 2: 1-72.
- [58] Wen J, Shi S. A phylogenetic and biogeographic study of *Hamamelis* (Hamamelidaceae), an eastern Asian and eastern North American disjunct genus. *Biochem, Syst, Ecol*, 1999; 27: 55-66.
- [59] Wu CI. On the significance of Pacific intercontinental discontinuity. *Ann Missouri Bot Gard*, 1983; 70: 577-590.
- [60] Zhang ZY, Lu AM. Hamamelidaceae: Geographic distribution, fossil history and origin. *Acta Phytotax Sin* 1995; 33: 313-339.

# The Potential of Molecular Analysis and Interspecific Hybridization for Azalea Phylogenetic Research

JAN DE RIEK<sup>1</sup>, VALENTINA SCARIOT<sup>2</sup>, TOM EECKHAUT<sup>1</sup>,  
ELLEN DE KEYSER<sup>1</sup>, NOBUO KOBAYASHI<sup>3</sup> and TAKASHI  
HANDA<sup>4</sup>

<sup>1</sup>Institute for Agricultural and Fisheries Research (ILVO), Plant Unit,  
Caritasstraat, Melle, Belgium

<sup>2</sup>Department of Agronomy, Forest and Land Management – University of  
Turin, via Leonardo da Vinci 44, 10095 Grugliasco (TO), Italy

<sup>3</sup>Laboratory of Plant Breeding, Faculty of Life and Environmental  
Science, Shimane University, Nishikawatsu, Matsue, Shimane 690-8504,  
Japan

<sup>4</sup>Graduate School of Life and Environmental Sciences, University of  
Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan

## ABSTRACT

Plants belonging to either the subgenus *Pentanthera* or the subgenera *Tsutsusi* and *Azaleastum* of the genus *Rhododendron* are called respectively deciduous and evergreen azaleas. Taking into account their mutual phylogenetic positions, the subgenus *Pentanthera* is closer to evergreen rhododendrons (subgenera *Rhododendron* and *Hymenanthes*) than to the *Tsutsusi* subgenus. Both azalea types are important ornamentals with a long breeding tradition. Different hybrid groups are often named after the supposed principal ancestor species. Apart

from uses in taxonomy and reconstruction of phylogenetic and cross-hybridization processes taking place in natural populations, most of the genetic research has been initiated keeping practical breeding applications in view. Interspecific hybridization in *Rhododendron* breeding is used as a tool to delineate species. The capability to cross species and the occurrence of fertilization barriers can give insight into speciation. Molecular techniques for phylogenetic and kinship research have been evaluated to a great extent. Comparative gene sequencing was then widened to the use of molecular markers to reveal more detailed genetic relationships. Finally, the use of candidate genes as functional markers for the assessment of genetic diversity is presented, opening up new research lines to the genetic mapping of plant traits and azalea genomics.

**Key Words:** Azalea, phylogeny, molecular markers, genetic diversity, interspecific hybridization

## INTRODUCTION

Although taxonomy and reconstruction of phylogenetic and cross-hybridization processes are scientifically sound functions, plant breeders are applied evolutionists. Most of the research reported here has been initiated with practical breeding applications in mind.

Pot azalea production in the Ghent area (Belgium) accounts for approximately 40 million plants per year, generating a value of 45 million Euro. Very little was known about the origin and the ancestor species used for this important ornamental crop. These facts were, however, later revealed by an AFLP (amplified fragment length polymorphism) study was confirmed by the comparative gene sequencing of *matK*. Although commonly referred to as *R. simsii* hybrids, many more species from the Tsutsusi subgenus have contributed in the past and still continue to be interesting progenitors for the introduction of new characteristics. Consequent to a collaboration with the Chinese Kunming Institute, wild accessions from indigenous *R. simsii* populations could be obtained together with other members of the Tsutsusi subgenus. The tradition of azalea breeding in Japan has resulted in several other important horticultural hybrid groups from species other than the Belgian pot azalea hybrids. In Europe, due to massive introduction of several types of evergreen azaleas in historical gardens, breeders have created interesting new genetic variation. The best characterized examples are the gardens and nurseries around the Northern Italian lakes.

However, a major hurdle in the breeding of pot azalea is the lack of yellow, orange and blue flower color in the rather closely related Tsutsusi species. The Ghent horticulturists in the past (from 1800 onwards) have

been working on the basic knowledge of the well-known Hardy Ghent hybrids, deciduous azaleas related to *R. luteum* (Europe), *R. molle* (Asia) and many North American species belonging to the *Pentanthera* subgenus. These hybrids are characterized by many cultivars with appealing yellow and orange flowers. Phylogenetic studies by Kron [55, 56] on the *Rhododendron* genus and the *Pentanthera* subgenus reveal an important feature: the *Pentanthera* subgenus seems to be a very old lineage that has relatively recently diversified, especially in North America. This *Pentanthera* subgenus (deciduous azaleas) is closer to evergreen rhododendrons (subgenera *Rhododendron* and *Hymenanthes*) than to the *Tsutsusi* subgenus (evergreen azaleas). Kron states that the deciduous azaleas represent the older more primitive lineages in the genus *Rhododendron*, whereas the evergreen azaleas are much more evolved and more (55, 56) closely related to the evergreen rhododendrons.

The experiments on intersubgeneric hybridization, comprising a range of genotypes covering all economically important groups within the *Rhododendron* genus, indicate that aiming for hybridization between *Tsutsusi* and *Pentanthera* genotypes holds very little promise towards the introgression of new flower petals in pot azalea. Contrary, promising results were obtained after pollination of *Hymenanthes*, *Rhododendron* and *Vireya* genotypes with pot azalea. The objective of this series of experiments was to select possible parent plants for the breeding of pot azaleas with new petal colors (yellow, orange or blue).

## **THE TAXONOMICAL POSITION OF AZALEAS WITHIN THE GENUS RHODODENDRON**

### **Taxonomy [7]**

Division: *Magnoliophyta*

Class: *Dicotyledonae*

Subclass: *Dilleniidae*

Order: *Ericales*

Family: *Ericaceae*

### **Plant Description**

The five-lobed structure of a typical *Rhododendron* L. flower is shown in Fig. 1. A single ovary contains several hundreds of ovules in different developmental stages. *R. simsii* Planch. is the main ancestor of Belgian pot

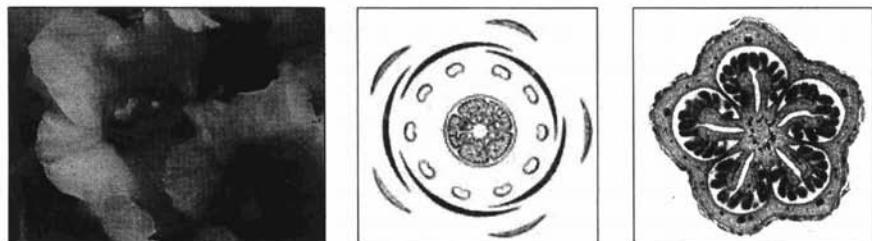


Fig. 1. The Rhododendron flower: *R. simsii* 'Rolinda', flower diagram, ovary structure [97].

azaleas (*R. simsii* hybrids, formerly known as *Azalea indica*) and is used here for typifying the genus. Distinct *R. simsii* features are the wide funnel-shaped large corolla and the inflorescence, in clusters of two-three or sometimes up to five-six flowers. Leaves are evergreen or dimorphic (spring leaves often deciduous, summer leaves persistent). Inflorescences are terminal; the calyx is five-lobed; the corolla and (usually) 10 stamens are shorter than the gynoecium [13].

### The Genus *Rhododendron*

The genus *Rhododendron* of the heather family (Ericaceae) is well-known for the beauty and diversity of floral and vegetative form of its more than 1,000 species. Due to the many species within this genus, taxonomists have made several classifications based mainly on morphological data. The genus is divided into eight subgenera (Table 1), the four most important being *Tsutsusi* (evergreen azaleas except *Brachycalyx* section), *Pentanthera* (deciduous azaleas), *Rhododendron* (lepidote rhododendrons) and *Hymenanthes* (elepidote rhododendrons) [7]. *Hymenanthes* species are evergreen and have large non-scaly leaves; *Rhododendron* (subgenus) species have smaller scaly leaves, are usually evergreen, but occasionally semi-deciduous. These latter two subgenera are commonly referred to as "rhododendrons", while *Pentanthera* and *Tsutsusi*, together with the *Azaleastrum* subgenus (evergreen), comprise the "azaleas". Due to their ornamental importance, all these subgenera have been discussed, revised and/or classified by several authors [6, 8, 9, 11, 83]. Rhododendrons have been in existence over 50 million years [44]. Their greatest natural gene center, with more than 300 species, is in Asia in an area ranging from Nepal along the line of the Himalayas into northern Myanmar and the provinces of Yunnan and Szechwan in Southwest China [60]. At present

**Table 1.** The genus *Rhododendron* [7]

Subgenus	Section	Comments
<i>Rhododendron</i>	<i>Rhododendron</i>	27 subsections, 211 species, scaled leaf
	<i>Vireya</i>	7 subsections, 310 species, 'Malaysian' species
	<i>Pogonanthum</i>	21 species
<i>Hymenanthes</i>	<i>Ponticum</i>	24 subsections, 302 species, unscaled leaves, garden hybrids with large flowers
<i>Pentanthera</i>	<i>Pentanthera</i>	23 species, deciduous azaleas
	<i>Rhodora</i>	2 species
	<i>Viscidula</i>	1 species
	<i>Sciadorhodion</i>	4 species
<i>Tsutsusi</i>	<i>Tsutsusi</i>	94 species, evergreen azaleas
	<i>Brachycalyx</i>	23 species, deciduous azaleas
<i>Azaleastrum</i>	<i>Azaleastrum</i>	11 species
	<i>Choniastrum</i>	19 species
<i>Therorhodion</i>	-	2 species
<i>Mumeazalea</i>	-	1 species
<i>Candidastrum</i>	-	1 species

rhododendrons occur in all continents of the Northern Hemisphere. Most species are found in China and the Himalayas, while tropical species (*Vireya*) grow mainly in Malaysia and Indonesia. The genus also occurs in northeastern Asia, Caucasus and North America. Their present distribution is mainly affected by three physical factors: global climate change, the drift of continents, and the rise of mountains.

Diverse morphological characteristics are encountered in the various species of *Rhododendron*. The flowers may be rotate-campanulate, campanulate, funnel-campanulate, funnel-shaped, or tubular-campanulate, and the number of stamens is usually five to ten.

An azalea can be a shrub or a small tree. It may be evergreen or deciduous, and the flowers may be single or grouped in clusters. Galle [28] states that azaleas perform best when the pH of the medium is between 4.5 and 6.0 (acidic). A wide range of growth rates is encountered, partly reflecting the genetic make-up of the plants, and partly depending on climate, soil, availability of water, etc.

Deciduous azaleas are generally taller, upright plants with less branching than evergreen species. The height ranges from about 1 to 10

m and the leaf-blade length from 3.4 to 12 cm. A flower bud may produce as few as 3 or up to 24 flowers, depending on the species. Leaves are often much larger than those of the evergreen species, and they are usually not glossy. The flowers of most species have relatively long, narrow tubes. Colors include white and a variety of tones, ranging from pale to deep, in the spectrum of yellows, oranges, yellowish pinks, reds, pinks, purples, and purplish pinks [96].

Typically, evergreen azaleas are much-branched shrubs, as broad as high, and entirely covered with medium- to dark-green, somewhat glossy, entire (non-serrate or lobed) leaves. In winter, the foliage of some varieties takes on a bronze or deep purplish coloration. In cultivated azaleas, the flower form ranges from single to fully double, and color patterns may be solid, flecked, striped, sectored, or picotee. Colors include white, yellowish pinks, reds, pinks, purples, and purplish pinks. Evergreen azaleas are not truly 'evergreen' in the usual sense of the word. Their foliage is dimorphic. In the wild, the height of familiar species of evergreen azaleas range from prostrate to 3 m; their leaf-blade lengths from 0.3 cm to 8 cm; and their corolla length from 0.6 to 6 cm. Flower buds commonly produce one to four flowers, depending on the species.

Within the evergreen azaleas, at least four more or less well-defined groups of cultivars are distinguished according to their traditionally recognized most important ancestor: pot or Belgian azaleas (*R. simsii* putative hybrids, formerly erroneously called *Azalea indica* referring to *R. indicum* (L.) Sweet), Hirado azaleas (*R. scabrum* G. Don putative hybrids), Kurume azaleas (putative hybrids of *R. kiusianum* Makino var. *kiusianum* and var. *sataense* (Nakai) D.F. Chamb. and *R. kaempferi* Planch.) and Satsuki azaleas (putative mutants or hybrids of *R. indicum* and *R. eriocarpum* (Hayata) Nakai). The group of Hirado, Kurume and Satsuki azaleas is often referred to as Japanese azaleas because of their geographical origin; a distinction is made with regard to hardiness. Belgian pot azaleas have been created from a relatively narrow genetic basis of collectors' material, introduced in botanical gardens and private collections from the far east. The Ghent region in Flanders (Belgium) is now the main European production area of pot azaleas. *R. simsii*, the species that is accepted to be the main ancestor, originates from hilly areas in China (Chang Jiang valley), Thailand, Laos and Burma. Apart from *R. simsii*, at least three other species from the *Tsutsusi* subgenus, from South Asia and Japan, might have contributed in locating the ancestors: *R. indicum*, *R. X mucronatum* (Blume) G. Don and *R. scabrum* [38].

## **INTERSPECIFIC HYBRIDIZATION IN RHODODENDRON: BREEDING AS A TOOL TO DELINEATE SPECIES?**

In biology, a species is one of the basic units of biodiversity. Traditionally, researchers relied on observations of anatomical differences and on observations of whether different populations were able to interbreed successfully, to distinguish species; both anatomy and breeding behavior are still important to assign species status. Some crossbreeding experiments are reported in this chapter. When possible, pollination was performed using pot azaleas both as pollen or seed parent, since interspecific incongruity might be unilateral or bilateral.

Pollen tube staining was performed to detect possible prefertilization barriers. Ovules were put *in vitro* to prevent possible abortion and to allow germination of a larger number of seeds. Microsatellite markers were analyzed for the final screening of the seedlings.

### **"Capability to Cross" as a Criterion for Speciation**

#### ***Fertilization Barriers and How They Can Be Overcome***

Interspecific hybridization is a means of transferring genes within related species; Van Tuyl [93] considered it to be the most important source of genetic variation in ornamentals. The difficulty in creating interspecific hybrids increases relatively with the phylogenetic distance between the parents [82]. As suggested by Hogenboom [41, 42], breeding incompatibility between plants may be controlled by either of two major mechanisms, i.e., incompatibility or incongruity.

Incompatibility is typically involved in crosses between closely related genotypes that possess at least some elements of an S-gene controlled self- or inter-incompatibility system. Incongruity on the other hand is typically seen in crosses where separate evolution of the partners (being different species) has led to a breakdown in control of pollen tube behavior in the pistil. It may, therefore, be considered as an isolating mechanism developed as a by-product of evolutionary divergence [42]. The development of the hybrid genome may also be hampered or even arrested by multiple barriers at a later stage. Incongruity can be caused by both prefertilization and postfertilization barriers; incompatibility mainly occurs before gamete fusion [73]. Whereas incompatibility promotes outbreeding, incongruity limits the possibilities for species hybridization

[64]. Neither incompatibility nor incongruity is caused by sterility (non-viability of the gametes [51].

Six main hybridization barriers may be distinguished in incongruous crosses (Table 2): (1) pollen tube growth inhibition, (2) absence of seed germination (most probably due to lack of endosperm), (3) hybrid albinism, (4) limited growth vigor, (5) selective elimination of the paternal genome, and (6) sterility of the seedling [24]. Growth and development of "alien" pollen tubes can be impeded in the female style (prezygotic barriers); after fertilization embryo development can be arrested by malformation of nurse tissue, mostly endosperm (postzygotic barriers). Several techniques have been used to overcome fertilization barriers [84]. *In vitro* methods assist in the development of an integrated procedure for overcoming fertilization barriers, since environmental conditions can be controlled and optimized. Well-known examples are the application of exogenous growth substances and immunosuppressants, embryo rescue, intraovarian pollination *in vivo* and *in vitro*, *in vitro* fertilization of ovules, and somatic cell hybridization [102]. Hermsen [39] quotes "the statement, that two species are not crossable, is controversial unless a broad genetic variation of the parental species has been used and the cross combinations have been carried out on a large scale under a wide range of environmental conditions".

Experience on interspecific breeding with *Tsutsusi* hybrids has mainly been gathered by Preil [72] and Rouse [74]. Rouse concentrated on *Vireya* crosses, whereas Preil used genotypes from various subgenera in unilateral crosses. Neither of them was successful in obtaining yellow flowering hybrids. Ureshino [87] succeeded in obtaining yellow hybrids after three-way *Tsutsusi* x *Pentanthera* crosses, but all seedlings turned out

**Table 2.** Different manifestations of incongruity after interspecific pollination

	Incongruity type	Possible solution
Prezygotic	Pollen tube growth inhibition	Prefertilization treatments
Postzygotic	Absence of endosperm, no seed germination	Embryo rescue
	Hybrid albinism	Bilateral crosses; 'bridge' plants
	Lack of growth vigor ('hybrid vigor')	Other crosses
	Hybrid breakdown	Other crosses
	Sterility	Allopolyploidization

to be (semi)deciduous and not vigorous. Color synthesis and inheritance within *R. simsii* hybrids has mainly been described by Heursel [36, 38] and De Loose [16, 17]. Especially *R. keiskei* Miq., *R. lutescens* Franch. and *R. burmanicum* Hutch. were used as breeding partners in earlier experiments [62]. In *Rhododendron*, prezygotic barriers are present in many parent combinations [23, 68, 72, 74] and pollen tube growth is significantly influenced by temperature treatment [49]. Probably, heat sensitive inhibitors of pollen tube growth are inactivated.

Ploidy differences between parent plants might be one of the main incongruity barriers according to Badger [2] by causing a 'triploid block' resulting in the malformation of endosperm and the inhibition of germination. Endosperm absence or retarded development is often the cause of spontaneous abortion of hybrid embryos after interspecific pollinations [71, 82, 92]. Johnston [46] proposed the endosperm balance number theory to explain the basis for normal seed production, but currently there is no consensus on the requirements for the formation of hybrid endosperm. In Table 3, ploidy levels of different species and hybrids are listed according to literature and flow cytometric measurements.

Genome sizes of plants are often linked with phylogenetic evolution [47, 67]. They are calculated compared to the known standards. As an internal standard for flow cytometric analysis of *Rhododendron* genome sizes, soybean (*Glycine max* Merr. 'Polanka') was used (1.25 pg for 1C, [3]. The genome size of some measured diploid *Rhododendron* cultivars is about half the size of soybean; those of the species analyzed is higher (Table 4). These results show that there is variation in genome sizes in the genus *Rhododendron*.

*In vitro* embryo rescue experiments on *Rhododendron* [98] revealed that although more seedlings were obtained, after crosses like *Psutsusi Pentathera* the relative number of albinos showed an increase. During interspecific *Rhododendron* breeding, albinism occurs rather frequently, even after intrasubgeneric pollinations [65]. In Japan, evergreen *R. indicum* hybrids could be crossed with deciduous azaleas by embryo rescue techniques provided the evergreen azalea was used as the seed parent [66, 87, 88]. After a three-way cross, more seedlings inherited plastid-DNA from the pollen donor than after a control cross. The reason for this phenomenon remains unclear; however the inheritance of plastid-DNA from pollen donors has been described before [31, 100]. Using a rapid screening method through DAPI fluorescence, Corriveau [12] detected putative plastid-DNA in the generative and/or sperm cells of pollen from

**Table 3.** Ploidy levels of *Hymenanthes* (H), *Pentanthera* (P), *Rhododendron* (R) and *Vireya* (V) species and hybrids

SG	Tested species or hybrid	2n (measured)	2n (literature)	SG	Tested species or hybrid	2n (measured)	2n (literature)
H	R. <i>wardii</i>	2x	2x	R (Cont.)	R. 'Blaufeder'	4x	*
	R. 'Nancy Evans'	2x	2x		R. 'Habashan'	2x	2x
	R. <i>dichroanthum scyphocalyx</i>	-	-		R. 'Hillier's form'	4x	*
	R. 'Goldsworth Orange'	2x	2x		R. 'Yellow Hammer'	7x - 8x	2x/6x
	R. 'Jingle Bells'	2x	2x		R. <i>leptocladon</i>	2x	*
	R. 'Kupferberg'	2x	2x		R. <i>burmamicum</i>	2x	*
	R. <i>citriniflorum</i>	-	-		R. <i>valentinianum</i>	2x	2x
	R. <i>sanguineum haematum</i>	-	-		R. <i>xanthostephianum</i>	2x	2x
	R. 'Viscy'	2x	2x		R. 'Euan Cox'	2x	2x
	R. <i>laetum</i>	> 4x	2x		R. 'Shamrock'	2x	2x
P	R. 'Hollandia'	> 4x	2x/4x	R V	R. 'Golden Bee'	2x	2x
	R. 'Nancy Waterer'	> 4x	2x/4x		R. <i>lutescens</i>	2x	2x
	R. 'Unique'	> 4x	2x/4x		R. 'Curlew'	2x	2x
	R. <i>chrysodon</i>	2x	2x		R. 'Chikor'	2x	2x
R	R. <i>megeratum</i>	2x	2x	V	R. <i>kawakamii</i>	2x	*
	R. <i>sulfureum</i>	2x	*		R. <i>laetum</i>	4x	*
	R. <i>cinnabarinum concatenans</i>	-	-		R. <i>macgregoriae</i>	4x	*
	R. <i>cinnabarinum xanthocodon</i>	-	-		R. <i>aequabile</i>	-	*
	R. 'Azurwolke'	4x	4x		R. 'Sunny'	4x	*

SG = subgenus, \* not described by Janaki [45]

**Table 4.** Genome sizes of some species and cultivars in the genus *Rhododendron* experimentally determined by propidium iodide staining and flow cytometry (Leen Leus, unpublished results)

	Ploidy level (x)	1C DNA content (pg)
<i>R. simsii</i> hybrid 'Mistral'	2	0.62
<i>R. simsii</i> hybrid 'Nordlight'	2	0.61
<i>R. simsii</i> hybrid 'White Hexe'	2	0.56
<i>R. lutescens</i>	2	0.75
<i>R. spiciferum</i>	2	0.76
<i>R. megeratum</i>	2	0.86

43 species (out of 235 species tested). One of these 43 species was *Rhododendron maximum* L.

### **Plant Material, Pollination and Pollen Tube Staining**

Altogether, 19 botanical species and 19 cultivated hybrids were used in hybridization experiments. Different pot azalea cultivars and seedlings were used. Yellow or orange rhododendrons were crossed with white pot azaleas; purple pot azaleas were used in breeding experiments with blue or lavender rhododendrons (Table 5).

The seed parent was emasculated by removal of the anthers. When possible, every genotype was used both as seed parent and as pollen parent. Due to male sterility or plant disease, reciprocal pollination was impossible in some cases. The risk of self-pollination appeared high on *R. valentinianum* Forrest ex Hutch., *R. 'Goldsworth Orange'*, *R. 'Hillier's form'* (high pollen production) and *R. 'Jingle Bells'* (cleistogamy), which were emasculated before the opening of the flower. *R. 'Jingle Bells'* and *R. 'Nancy Evans'* have colored sepals, which in *R. simsii* hybrids is linked to female sterility ('hose-in-hose'), but were nevertheless used as seed parents. A total number of 2,361 flowers was pollinated.

Pollen tube staining was performed as described by Cuevas [10], Ureshino [89], and Vervaeke [94]. Four pistils in each treatment were harvested four days and four weeks after pollination and aniline blue stained after fixation. The stained materials were then squashed on a glass slide and observed with a fluorescent photomicroscope (LEICA DM IRB/MPS 52). Eight combinations of subgenera (*T* x *H*, *H* x *T*, *T* x *P*, *P* x *T*, *T* x *R*, *R* x *T*, *T* x *V* and *V* x *T*) were established (*T* = *Tsutsusi*; *H* = *Hymenanthes*; *P* = *Pentanthera*; *V* = *Vireya*). For every subgenus

**Table 5.** Wild species and cultivated hybrids from the *Hymenanthes*, *Pentanthera* and *Rhododendron* subgenera used in a breeding program with pot azalea (\*male sterile, °only pollen available) (after Chamberlain, 1996 [7])

Subgenus	Subsection		Species and hybrids
Hymenanthes, Section <i>Ponticum</i>	Campylocarpa	<i>R. wardii</i> Smith	* 'Nancy Evans'
	Neriflora	° <i>R. dichroanthum</i>	'Goldsworth Orange'
		scyphocalyx Balf	'Jingle Bells'
		° <i>R. sanguineum</i>	
		haematum Balf	'Kupferberg'
		'Viscy'	
<i>Pentanthera</i> , Section <i>Pentanthera</i>	<i>Pentanthera</i>	<i>R. luteum</i> Sweet	'Hollandia'
			'Unique'
<i>Rhododendron</i> , Section <i>Rhododendron</i>	Boothia	<i>R. chrysodorum</i> Tagg	<i>R. megeratum</i> Balf
	Cinnabarinum	° <i>R. cinnabarinum</i>	° <i>R. cinnabarinum</i>
		<i>xanthocodon</i> Hutch	concatenans Hutch
		'Azurwolke'	'Blaufeder'
	Laponica	'Hillier's Form'	'Yellow Hammer'
Maddenia Tephropepla Triflora Umiflora		* <i>R. burmanicum</i> Hutch	<i>R. leptoclada</i> Dop
		**Euan Cox'	
		**'Shamrock'	
		*'Chikor'	
<i>Rhododendron</i> , Section <i>Vireya</i>	Pseudovireya	<i>R. kauakamii</i> Hayata	
	Euvireya	<i>R. macgregoriae</i> Muell	<i>R. laetum</i> Smith
		° <i>R. aequabile</i> Blume	'Sunny'

combination, three parent combinations were chosen (depending on flower availability).

### **Pollen Tube Staining**

Lack of pollen tube growth, combined with an early abortion, was observed after P x T and T x V crosses, similar as Rouse [74], Ureshino [89]. H pollen tubes grow slowly in T styles, whereas T pollen tubes reach the ovary of H flowers quickly. Growth of pollen tubes after interspecific T x R and R x T pollinations seems delayed, but mostly not inhibited. In *R. kawakamii* Hayata, which belongs to the *Pseudovireya* section, T pollen tube growth is inhibited, whereas *Euvireya* styles ('Sunny') are able to guide the same pollen until the ovary.

The fruits, which aborted after four weeks, showed no or weakened pollen tube growth after four days. Nevertheless, not every parent combination that showed weak pollen tube growth after four days resulted in spontaneous abortion. Based on this data, bilateral pollination seems absolutely imperative to get a thorough understanding of crossing congruity.

### **Ovule Culture**

From 2,361 pollinated flowers, 349 (14.78%) did not abort before *in vitro* initiation. From the 13,549 ovules cultured *in vitro*, 1488 (10.98%) germinated, 319 (21.44%) developed into green seedlings (Table 6).

After T x H pollination, only two out of seven parent combinations yielded green seedlings; four out of seven did not result in fruit setting. After the reciprocal pollination type, three out of 10 parent combinations resulted in green seedlings. Non-aborted fruits were empty in four combinations. Although a majority of T x P pollinations yielded seedlings, this pollination type was not successful since all seedlings were albinos. Due to the tetraploid DNA level of *R. luteum* Sweet and its hybrids, the hybrid character of the seedlings (having an intermediate ploidy level) could be confirmed through flow cytometry. The reversed subgenus combination was incongruent due to pollen tube growth inhibition and, therefore, did not yield fruits. Out of 31 T x R parent combinations, six provided fertilized ovules that could be put *in vitro*, five of which resulted in the germination of a considerable number of green plantlets. Especially upon pollination by *R. lutescens*, the number of seedlings was high. For the establishment of R x T crosses, 38 parent combinations were performed; nine combinations yielded green seedlings. The three most successful

**Table 6.** Efficiency of intersubgeneric *Rhododendron* pollinations after ovule culture

Pollination type	# Flowers pollinated	# Fruits not aborted	# Ovules initiated	# Ovules germinated	# Green seedlings	# Vigorous survivors
T x H	114	18	1,179	15	5	0
H x T	179	58	756	270	161	32
T x P	509	160	5,146	142	0	-
P x T	44	0	-	-	-	-
T x R	279	20	989	250	51	16
R x T	836	76	4,841	787	98	66
T x V	120	0	-	-	-	-
V x T	280	17	640	25	25	8
Total	2,361	349	13,551	1,489	340	122

combinations, all having *R. burmanicum* as seed parent, suffered from severe chlorosis once the first 'true' leaves appeared. T x V combinations were incongruent due to pollen tube growth inhibition. Of eleven different pollinations only the use of *R. 'Sunny'* as seed parent enabled the fertilization of ovules that germinated into green seedlings. 'Sunny' is tetraploid, which allowed selection of hybrids through flow cytometry. All flow cytometrically tested plantlets were true hybrids.

### **SSR Molecular Analysis**

Altogether, SSR analysis was performed on plantlets originating from 10 different parent combinations. Microsatellite analysis using two SSR primer sets yielded conclusive evidence for the interspecific character of half of the seedlings that were tested. Mendelian inheritance of uniquely maternal alleles due to self-pollination was observed in pot azalea x *R. chrysodonon* Tagg ex Hutch., *R. chrysodonon* x pot azalea, *R. xanthostephnum* Merr. x pot azalea, *R. 'Azurwolke'* x pot azalea and *R. 'Goldsworth Orange'* x pot azalea. On the other hand, microsatellite gel patterns yielded obvious proof for the inheritance of paternal alleles in *R. 'Jingle Bells'* x pot azalea, *R. 'Nancy Evans'* x pot azalea, *R. burmanicum* x pot azalea, *R. 'Euan Cox'* x pot azalea and *R. 'Sunny'* x pot azalea (exemplified in Fig. 2). This verifies their interspecific status. The molecular testing revealed a final efficiency of four vigorously growing and non-chlorophyll deficient intersubgeneric hybrids per 179 H x T pollinations (2 *R. 'Nancy Evans'* x pot azalea + 2 *R. 'Jingle Bells'* x pot

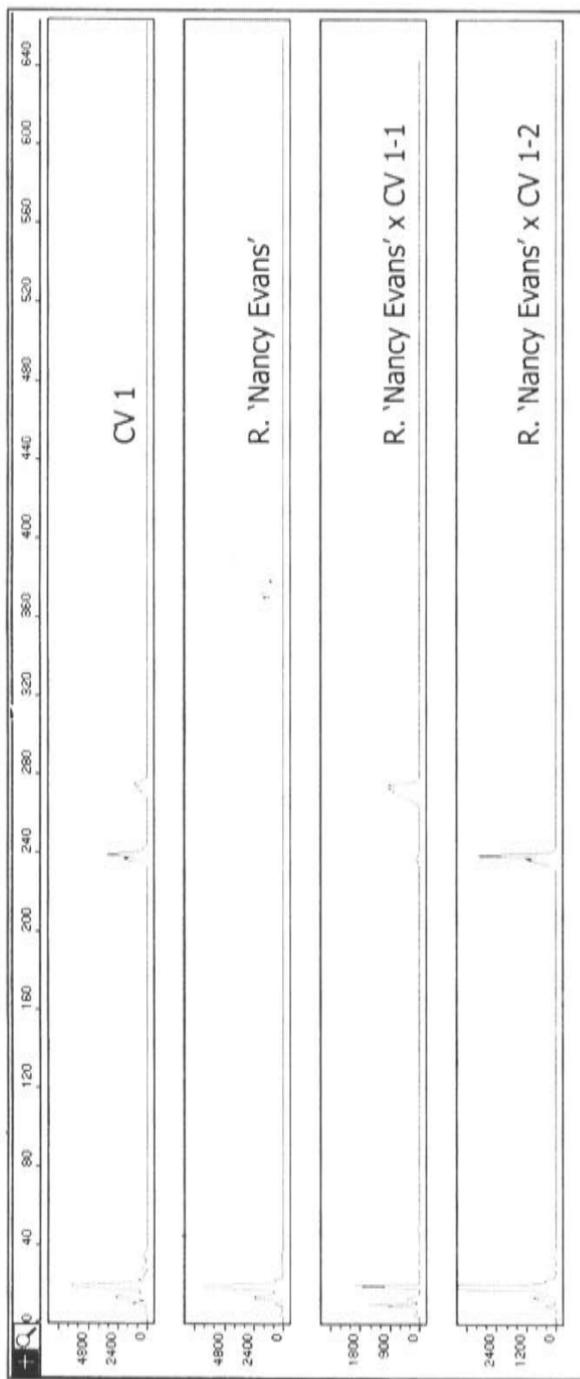


Fig. 2. DNA fragments obtained after applying the SSR primer sets A2A 2.40 on total genomic DNA of *R. 'Nancy Evans'* x CV 1-1 and *R. 'Nancy Evans'* x CV 1-2 (lower 2 electrophograms) and their respective seed and pollen parents (X-axis: bp length; Y-axis: fluorescence).

azalea). For R x T and V x T, these efficiencies amounted to 19 hybrids/836 pollinations (18 *R. burmanicum* x pot azalea + 1 *R. 'Euan Cox'* x pot azalea) and 8 hybrids/280 pollinations (8 *R. 'Sunny'* x pot azalea), respectively. All other subgenus combinations showed a zero efficiency. Finally, 31 green and vigorous hybrid seedlings were obtained from 2,361 interspecific pollinations.

Six major incongruity barriers can be distinguished after interspecific pollinations within the *Rhododendron* genus, which are usually successful only when both parents belong to the same subgenus. After crosses in the *Rhododendron* genus between T genotypes and plants of different subgenera, severe pollen tube growth inhibition was observed only after P x T and T x V pollinations. Seed germination after other combinations is low but can be optimized *in vitro* by adding 50 mg/l GA<sub>3</sub>. Total albinism only occurred in T x P seedlings; however, heavy chlorosis was observed in many seedlings after different subgenus combinations, thus causing low growth vigor. Hybrid breakdown was not observed and haploids were not found. A matching ploidy level between parent plants is no absolute requirement for the creation of hybrids; for this reason ploidy analysis could be used to reveal the hybrid character of V x T and T x P seedlings. Plants with decreased growth vigor, are very likely to be hybrids. Due to low growth vigor of most seedlings acclimatization, flowering and consequent fertility research will probably be delayed.

Numeric data (Table 6) very clearly show that both the subgenus applied and the direction of the cross is of utmost importance for abortion, germination and albinism. H x T and T x P crosses yield a high number of fruits, whereas P x T and T x V do not yield fruits at all, the other combinations having an intermediate efficiency. *In vitro* germination ( $\pm$  65% after control pollinations) was fairly high after H x T, T x R and R x T combinations, and low in T x H, T x P and V x T combinations. Albinism seldom occurs in crosses with V or R, but all T x P seedlings are white. T x R hybrids, however, often suffer from heavy chlorosis when forming the first 'true' leaves. Albino quantities after T x H combinations are relatively large, but the absolute number of green seedlings remains high after H x T pollinations. When expressed as number of green seedlings obtained per 100 flowers pollinated, the pollination efficiencies of H x T, T x R and R x T combinations are much higher than after T x H and V x T pollinations, whereas T x P, P x T and T x V combinations show a zero efficiency. In Table 6 also the number of initiated ovules/flowers pollinated is mentioned. This number, which is about 150-200 after a fully congruent pollination, gives an idea of the

seed development inside the fruit; however, as the table indicates, there is no correlation with the ultimate pollination efficiency. This table is not accurate since it remains unclear as to what extent hybrid vigor will inhibit plant growth and development, and which combinations will be fertile.

SSR analysis (Fig. 2) was performed on samples from 10 different origins, five of which yielded unambiguous negative results (no hybrids) and five of which yielded unambiguous positive results (hybrids). Next to the possible occurrence of "zero alleles", the drawbacks of SSR are: high developmental costs, species specificity and stutter bands (artifacts produced by DNA polymerase slippage, mostly +1 or -1 repeat), interfering with the results. However, these weaknesses are out of proportion to its benefits: a relatively simple PCR (polymerase chain reaction) assay, easy interpretation of the results and a co-dominant character since the observed peaks are directly coherent with allele presence and reflect their length (determined by the number of repeats between the primer annealing sites).

### **INTERGENERIC HYBRIDIZATION BETWEEN RHODODENDRON AND MENZIESIA (ERICACEAE)**

*Menziesia* Sm. is a small genus in the *Ericaceae* family, which includes nine species composed of two North American and seven Japanese species. Flowers are small and bell-shaped, with white, yellow, pink or brick red colors. They are relatively dwarfish, and have soft branches, and are cold and shade tolerant. Amongst seven species of Japanese *Menziesia*, *M. ciliocalyx* Maxim. var. *multiflora* (Maxim.) Makino is most popular. This species is widely distributed and has attractive pendulous flowers.

Interspecific pollinations within the genus *Rhododendron* are usually successful only when both parents belong to the same subgenus; however, several inter-subgeneric hybrids were successfully obtained. The molecular phylogenetic studies revealed the nested position of the genus *Menziesia* in the genus *Rhododendron*. Therefore, there is a possibility to create new intergeneric hybrids between these two genera. If *Menziesia* could be hybridized with *Rhododendron*, hybrid progenies will have novel characters. Kita [50] tried to make intergeneric hybrids by hand pollination between *M. multiflora* and seven *Rhododendron* species (Table 7).

Some of the intergeneric combinations between the two species were reciprocally crossed. Capsule sets were observed 60 days after pollination,

**Table 7.** Frequency of capsule set and number of immature seeds per capsule in reciprocal crosses between *Menziesia multiflora* and *Rhododendron* species

Cross combination (female x male)		Number of capsule set/number of crosses (% of capsule set <sup>(x)</sup> )	Average number of immature seeds per capsule
M. multiflora x Rhododendron spp.			
M. multiflora	R. kaempferi	26/132 (19.7)	3.46
M. multiflora	R. kiusianum	32/165 (19.4)	7.81
M. multiflora	R. tashiroi	60/79 (75.9)	34.55
M. multiflora	R. wadanum	5/22 (22.7)	0
M. multiflora	R. pentaphyllum	15/27 (55.6)	0
M. multiflora	R. quinquefolium	4/16 (25.0)	0
M. multiflora	R. ovatum	5/12 (41.7)	0
Rhododendron spp. x M. multiflora			
R. kaempferi	M. multiflora	0/23	-
R. kiusianum	M. multiflora	18/26 (69.2)	10.44
R. tashiroi	M. multiflora	0/30	-
R. wadanum	M. multiflora	0/17	-
R. pentaphyllum	M. multiflora	0/20	-
R. quinquefolium	M. multiflora	0/10	-
R. ovatum	M. multiflora	0/12	-

(x): Number of capsule set/number of crosses × 100

collected 120 days after pollination, and sterilized in 1% sodium hypochlorite solution for 10 minutes followed by three washes in sterile distilled water. When *Rhododendron* species were used as maternal parents for crossing, capsule set was observed only in a cross of *R. kiusianum* × *M. multiflora* (Table 7). When *M. multiflora* was used as a female parent, capsule sets were observed in all seven cross combinations. Immature seeds were obtained in crosses of *R. kiusianum* × *M. multiflora*, *M. multiflora* × *R. kaempferi*, *M. multiflora* × *R. kiusianum* and *M. multiflora* × *R. tashiroi* Maxim. The greatest number of immature seeds per capsule was obtained in a cross of *M. multiflora* × *R. tashiroi*.

After sterilization, immature seeds were dissected from capsules, and then cultured on Anderson's rhododendron medium [1] supplemented with 50mg/l gibberellic acid ( $GA_3$ ), 30 g/l sucrose, and solidified with 3g/l gellan gum after adjusting at pH 5.0. Obtained immature seeds could germinate in all cross-combinations. Germinated seedlings were then transferred to the same medium with 10mg/l N6-[2-Isopentenyl] adenine

(2ip), 30g/l sucrose and 3g/l gellan gum, pH 5.0, to induce multiple shoots. Germination ratio was calculated one month after cultivation on the medium. Vigorously grown seedlings were transferred to the same medium with 30g/l sucrose and 3g/l gellan gum, without phytohormone, pH 5.0, to induce roots under the same condition. Plants were potted in soil and acclimatized.

Leaf color of the seedlings was classified into four classes: green, pale-green, variegated (sectional chimeras) with both green and albino leaf segments, and albino (Table 8). Most of the seedlings from any cross showed pale-green or albino leaf colors. Green seedlings could only be obtained when *M. multiflora* was used as a maternal parent, although their frequency was still low.

To confirm the intergeneric origin of the seedlings, PCR-RFLP (restriction fragment length polymorphism) analysis was conducted. Total DNA was extracted from fresh leaf tissues following the methods of Kobayashi [52]. The ITS region of nrDNA was amplified by PCR. The PCR products were digested by 0.1-0.5 units of appropriate restriction enzymes for each parental combination at 37°C for 2 hours. Banding patterns of restricted fragments in ITS region between *M. multiflora* and *Rhododendron* spp. used in this study were different. The ITS region of *M. multiflora* was digested into two fragments (400 and 500 bp) by *Nsp*V restriction enzyme, whereas those of *Rhododendron* species were digested into 750 bp and 150 bp bands by *Afl* II enzyme. Therefore, hybrid seedlings possessed 4 bands (750, 500, 400 and 150 bp) after double digestion with *Nsp*V and *Afl* II enzymes. All of the investigated seedlings had both specific restricted fragments, confirming they were intergeneric hybrids (Fig. 3). The success of intergeneric hybridization of *Menziesia* with *Rhododendron* reflects a close relationship between these two genera. This result also indicates that molecular phylogenetic data can be used as a guide for breeding program.

## MOLECULAR TECHNIQUES FOR PHYLOGENETIC AND KINSHIP RESEARCH

Morphological data to support taxonomical classifications are relatively easy to obtain, but only a small number of traits can be evaluated. Published research focusing on evergreen azalea morphology concerns diversity of flower colors [35], the inheritance of the hose-in-hose character (sepals forming a corolla like structure; [33], content and composition of flavonoids which determine flower color [38], corolla size,

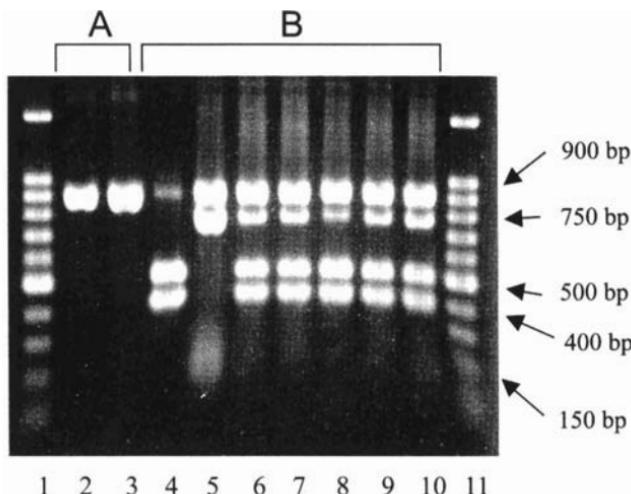
**Table 8.** Germination rate of immature seeds and leaf colors of progenies obtained by crosses between *Menziesia multiflora* and three *Rhododendron* species

♀	♂	Number of immature seeds cultured	% of germination (x)	Number of seedlings (% of seedlings <sup>(y)</sup> )		
				G <sup>(z)</sup>	PG	Sec
<i>M. multiflora</i>	<i>R. kaempferi</i>	90	34.4	7 (17.1)	19 (46.3)	2 (4.9)
<i>M. multiflora</i>	<i>R. kiusianum</i>	250	47.2	10 (8.5)	8 (6.8)	0
<i>M. multiflora</i>	<i>R. tashiroi</i>	2073	33.6	96 (13.8)	77 (11.0)	107 (15.4)
<i>R. kiusianum</i>	<i>M. multiflora</i>	188	17.6	0	9 (27.3)	0
						24 (72.7)

(x): Number of germinated seeds/number of cultured seeds × 100

(y): leaf colors are G = green, PG = pale green, Sec = sectional chimera of green and white, and albino: Number of seedlings in each color/total number of seedlings × 100

(z): leaf colors are G = green, PG = pale green, Sec = sectional chimera of green and white, and albino



**Fig. 3.** Profiles of PCR products (A) and restriction fragments of the products digested with *Afl*III and *Nsp*V (B). Lane 2&4, *M. multiflora*; Lane 3 & 5, *R. tashiroi*; Lane 6-10, hybrid from *M. multiflora* × *R. tashiroi*; Lane 6 & 7, Green hybrid; Lane 8, Pale green hybrid; Lane 9 & 10, Albino hybrid; Lane 1 & 11, 100 bp ladder.

number of stamens and percentage of plants with petaloid stamens [37] and other taxonomic characters [6]. An alternative classification may be based on molecular data which provide the possibility of studying a tremendous number of genetic loci in three distinct genomes (nuclear, mitochondrial, and chloroplast DNA). In plant systematics, the chloroplast genome has proved to be very informative. Among many genes studied for phylogenetic reasons, two that have been very useful in constructing evolutionary trees in the *Ericaceae* are *rbcL* (rubisco large subunit) and *matK* (maturase K). The *matK* gene is known to evolve approximately three times faster than *rbcL*, and is, therefore, a powerful tool for phylogenetic reconstruction within angiosperm families and genera [40]. On the other hand, the use of molecular markers can be very effective for studying diversity within species or the relationships between closely related species.

*Rhododendron* taxonomy has been tested in recent times by molecular phylogenies based on several DNA regions. Most of these studies have aimed at higher-level relationships. Since the reference papers on the phylogenetic relationships by Kron [54, 55], integrating previous molecular studies using *rbcL* and *nr18s* sequences and *matK* sequences of 42 taxa from traditional *Rhododendroideae* and potentially related clades, several authors have used these techniques to clarify taxonomical positions of specific genera. So far, classification of *Rhododendron* species

based on morphology has led to a consensus taxonomy recognizing the major subgenera *Azaleastrum*, *Hymenanthes*, *Pentanthera*, *Rhododendron*, *Tsutsusi*, and three minor ones. To determine whether these subgenera are monophyletic and to infer phylogenetic relationships between sections and species, Goetsch [30] carried out a cladistic analysis using molecular data, including all groups within the genus. They introduced the comparison of the nuclear gene RPB2-I, encoding a major RNA Polymerase II subunit. A phylogeny based on data analysis of 87 species by means of maximum parsimony, maximum likelihood, and Bayesian methods showed subgenera *Azaleastrum* and *Pentanthera* to be polyphyletic and grouped all *Rhododendron* species (except the two in section *Therorhodion*) into three large clades. Based on these results, modifications in *Rhododendron* classification were proposed, which consolidate minor subgenera and recognize monophyletic subgenera and sections.

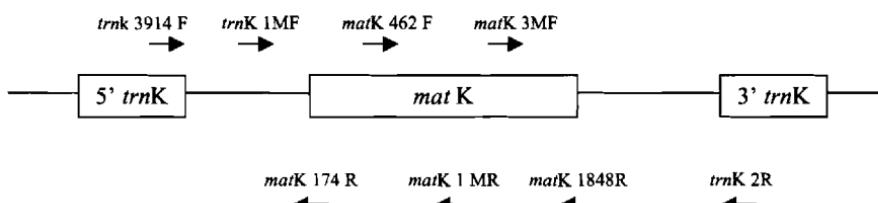
Two interesting studies, because of their link to azaleas as an ornamental group, relate to the genetic relationships of section *Pentanthera* and *Vireya*. Scheiber [80] unraveled the genetic relationships within *Rhododendron* section *Pentanthera* based on sequences of the ITS region. For specimens of the 15 currently recognized species in *Rhododendron* section *Pentanthera*, sequences of the entire ITS region including ITS1, ITS2, and the 5.8S subunit were generated by direct sequencing of PCR amplified fragments. *Rhododendron vaseyi* A. Gray section *Rhodora*, was used as an outgroup. Aligned sequences of the 16 taxa resulted in 688 characters. The region contained 38 variable sites and eight phylogenetically informative characters. A bootstrap analysis was performed and a dendrogram was constructed. Divergence values among the taxa were extremely low ranging from 0.00 to 3.51%, providing support to traditional views of section *Pentanthera* as a group of very closely related species. Brown [4, 5] studied the phylogenetic relationships of *Rhododendron* section *Vireya* inferred from the ITS nrDNA region and based on two noncoding regions of cpDNA (chloroplast DNA). Section *Vireya* is one of the most morphologically diverse groups of the genus *Rhododendron*. *Vireyas* have a unique distribution for the genus, being predominantly found throughout the Malaysian Archipelago. The alpha taxonomy of section *Vireya* is relatively well-understood and taxa are easily distinguished from other rhododendrons by their general appearance. Results of phylogenetic analyses of the ITS region (ITS-1, 5.8S and ITS-2) and of two cpDNA

regions, *psbA-trnH* and *trnT-trnL* intergenic spacers for the genus *Rhododendron*, with sampling concentrated on section *Vireya*, were presented. Relationships between the species of section *Vireya* do not correspond to the traditional classification based on morphology, instead correlating strongly with geographic areas, with a disjunction between an Australian-New Guinea clade and clades of west and middle Malaysian taxa. The phylogeny also indicates that the ITS region may not undergo complete homogenization in all species of *Rhododendron*.

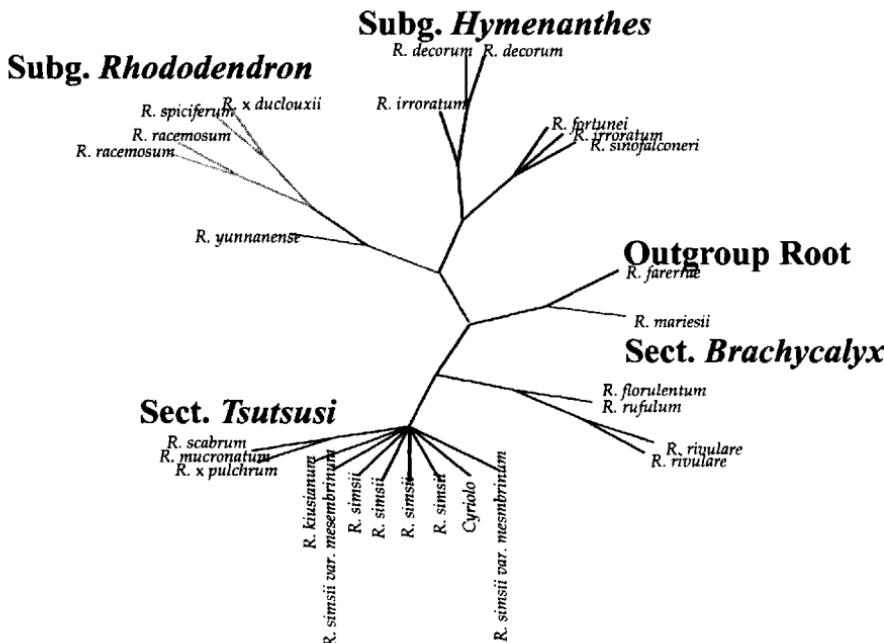
Our own research focused especially on section *Tsutsusi*. First, some studies using comparative gene sequencing were presented; this approach was then extended to the use of molecular markers to reveal further genetic relationships.

### **Phylogenetic Analysis Based on *MatK* Sequencing for Different Species from the *Tsutsusi* Subgenus**

Phylogenetic analysis based on the *matK* sequences was used to confirm the taxonomic position of the different species of the *Tsutsusi* subgenus. Some additional species belonging to the subgenera *Hymenanthes* and *Rhododendron* were also analyzed. A specific PCR to amplify the *matK* gene was performed on 33 Chinese *Rhododendron* species (Fig. 4) using four primer couples [58]. The amplified PCR fragments were directly used for sequencing, using the PCR primers as sequencing primers (forward and reverse). After sequencing and alignment of the different contig, a fragment of 2,529 bp was obtained covering *matK* and its flanking *trnK* sequences. Analyzing these 2,529 characters by means of PAUP, 2,438 characters were constant, 30 variable and parsimony uninformative, and 61 variable and parsimony informative. Trees were generated using 100 resampled bootstrap datasets, stepwise addition, and tree bisection recognition as branch swapping algorithm. The nodes that separated the different subgenera and sections all received high bootstrap values. In



**Fig. 4.** *MatK* locus showing the position of the primers used for amplification of the partial sequences [58].



**Fig. 5.** Consensus parsimony tree based on *matK* data for a selection of Chinese *Rhododendron* species and cultivated *R. simsii* hybrids; length = 98; consistency index (excluding uninformative characters) = 0.9403; retention index = 0.9987.

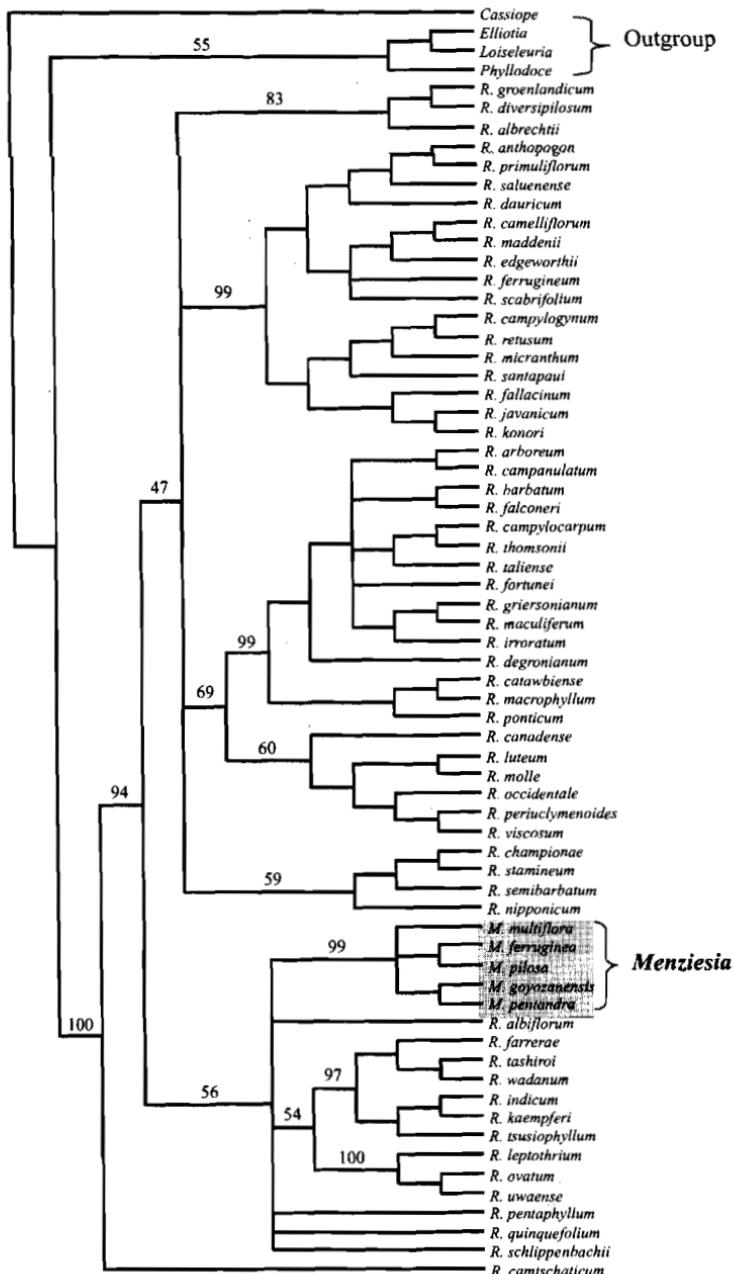
Fig. 5, a parsimony cladogram shows the genetic similarity between the assayed Chinese *Rhododendron* species and some *R. simsii* hybrids.

Five major clusters can be distinguished. The first group contains *R. mariesii* Hemsl. & E. H. Wilson and *R. farrerae* Tate ex Sweet; the latter was used as an outgroup species to root the dendrogram. Both species belong to the section *Brachycalyx* of the subgenus *Tsutsusi*. The second cluster contains *R. yunnanense* Franch., *R. spiciferum* Franch., *R. Xduclouxii* Levl., and two populations of *R. racemosum* Franch. All these species belong to the subgenus *Rhododendron*. A third cluster contains *R. fortunei* Lindl., *R. sinofalconeri* Balf., *R. argyrophyllum* Franch., *R. irroratum* Franch. and two populations of *R. decorum* Franch. belonging to the subgenus *Hymenanthes*. The fourth cluster contains *R. florulentum* P.C. Tam, *R. rufulum* P.C. Tam and two populations of *R. rivulare* Hand.-Mazz, all from the subgenus *Tsutsusi*. The last and largest cluster contains species all belonging to the section *Tsutsusi* four *R. simsii* populations, two *R. simsii* var. *mesembrinum* Rehder populations, *R. scabrum*, *R. kiusianum*, the Kurume hybrid "Cupido", *R. Xpulchrum* Sweet and *R. Xmucronatum*.

Kurashige [58, 59] also analyzed *matK* with *trnK* intron sequences of chloroplast (cp) genome of the genus *Rhododendron* and its closely related genera, and elucidated a substantial part of the phylogenetic relationships (Fig. 6). Total DNA was extracted from fresh leaf tissues following the methods of Kobayashi [52] or Yukawa [101]. All parsimony analyses were conducted with PAUP version 3.1 [86]. The heuristic search option with 100 random replicates [63] was used to perform Fitch parsimony analyses [27]. Branch lengths for trees were calculated by ACCTRAN optimization [85]. For assessment of the relative robustness for clades found in each Fitch parsimony analysis, the bootstrap method [26] was used. All the species of genus *Menziesia* were nested within *Rhododendron* species.

Apart from the above studies, some other authors also took up the section of *Tsutsusi* as a specific subject for their research. Hwang-ShihYing [43] investigated the molecular phylogeny of eight Taiwanese *Rhododendron* species based on sequence comparisons of the chloroplast *trnF-trnL* intergenic spacer region. Aligned sequences of the eight *Rhododendron* taxa were composed by 460 bp. The region contained 16 variable sites, of which 10 were phylogenetically informative. Neighbor-joining and parsimony analyses were conducted. Identical topology with three major clades was obtained from parsimony and neighbor-joining trees. Clade 1 consisted of *R. pseudochrysanthum* Hayata, *R. rubropunctatum* Hayata, *R. morii* Hayata, *R. hyperythrum* Hayata and *R. formosanum* Hemsl. The latter was sister of the other four species in the clade 1. Clade 2 consisted of *R. oldhamii* Maxim. and *R. kanehirae* E.H. Wilson. *R. ellipticum* Maxim. formed clade 3 and was basal to all other *Rhododendron* species. Close phylogenetic relationships among *R. pseudochrysanthum*, *R. morii*, *R. rubropunctatum* and *R. hyperythrum* based on the cpDNA sequences agreed with those derived from morphological characters.

Gao-LianMing [29] compared the sequences of nuclear ribosomal DNA internal ITS (including 5.8S rDNA) of 12 species representing two sections of subgenus *Tsutsusi* and four species representing three sections of subgenus *Pentanthera*. To reconstruct the phylogeny of these two subgenera using parsimony methods, *R. redowskianum* (Maxim.) Hutch. of subgenus *Therorhodion* was included in the analysis as an outgroup. The size of ITS region within the three subgenera ranges from 642 to 645 bp. The aligned length is 653 positions, which provide 6.58% variable sites and 3.68% phylogenetic informative sites when gaps are treated as



**Fig. 6.** Strict consensus of most parsimonious Fitch trees based on *matK* and *trnK* intron sequences for *Rhododendron* and *Menziesia* [59]. Numbers above internodes indicate bootstrap values from 1,000 replicates.

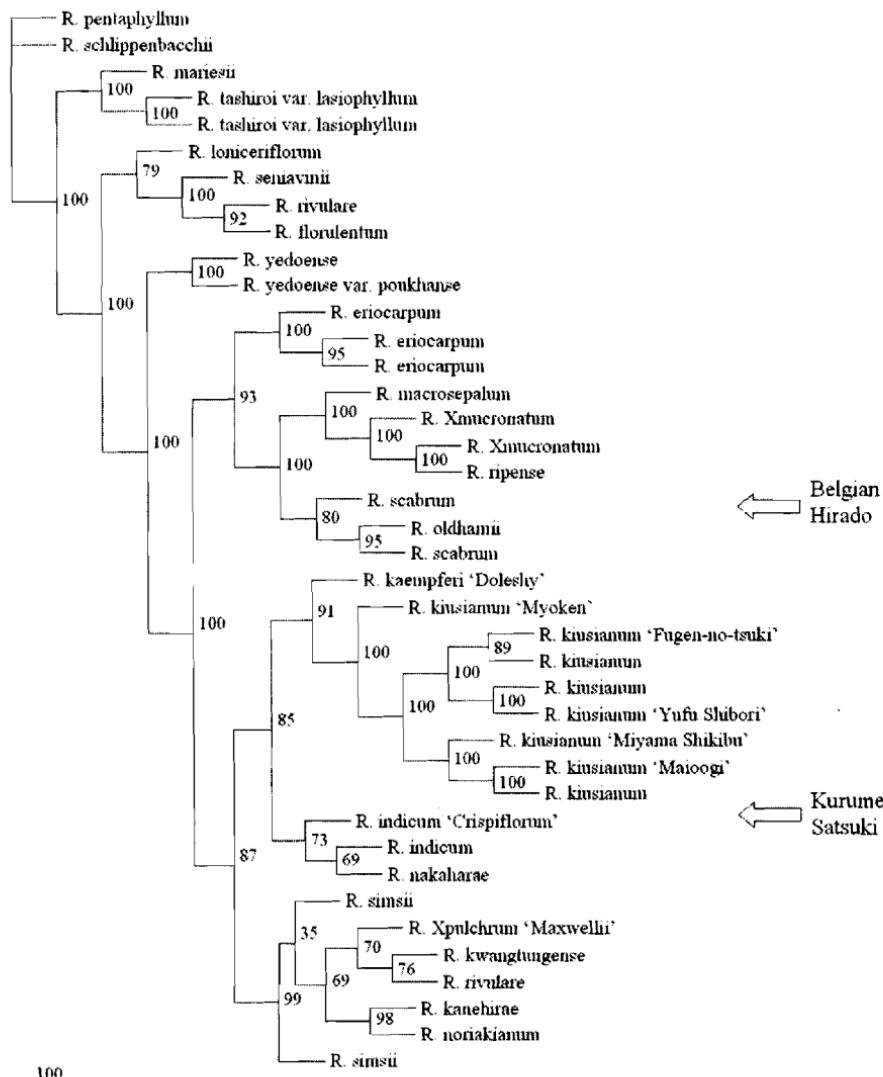
missing. Fifteen most parsimonious trees, with a length of 75 steps, consistency index of 0.9333 and retention index of 0.9515, were obtained. The topology of the strict consensus tree shows that: (1) subgenus *Tsutsusi* is a monophyletic group, the bootstrap value is 81%; (2) *R. tashiroi* is neither a member of section *Tsusiopsis* nor of section *Tsutsusi*, but it should be a member of section *Brachycalyx*; (3) *R. tsusiophyllum* Sugim. is a member of section *Tsutsusi*; and (4) the phylogenetic position of *R. schlippenbachii* Maxim. is not resolved in this study and needs further research.

### **Characterization of Natural Hybrids between *R. kiusianum* and *R. kaempferi* by RFLP-PCR**

Especially in Kyushu, the south main island of Japan, wild evergreen azaleas of section *Tsutsusi* distribute widely and abundantly. *R. kiusianum* is endemic on the top area of volcanic mountains in Kyushu island while *R. kaempferi* most commonly occurs at the bottom of these mountains [99]. These azaleas have been used as gene source for azalea breeding and hundreds of azalea cultivars were produced since the Edo era (1603~1867) [57]. Studying the introgressive hybridization between *R. kiusianum* and *R. kaempferi* of the Kirishima mountains, the PCR-RFLP analysis of cpDNA could detect specific bands for the two species in the 16S rDNA region when digested with *Hha*I restriction enzyme [53]. Populations of interspecific hybrids were composed of individuals that had a banding pattern of *R. kiusianum* or *R. kaempferi*. This indicates *R. kiusianum* and *R. kaempferi* are clearly distinct species and natural hybrid populations consist of individuals which have one of the two cpDNA.

### **The Usefulness of AFLP Markers for the Establishment of Phylogenetic Relationships in Evergreen Azaleas**

To evaluate if the AFLP technique can be applied for the establishment of phylogenetic relationships and yields comparable results to common taxonomy and gene sequencing, 84 accessions of evergreen azaleas, including cultivars (19 Belgian azaleas, 6 Hirado azaleas, 14 Kurume azaleas, and 5 Satsuki azaleas) as well as the species and varieties that were most likely to have been involved in their origin, were analyzed [78]. Based on the pair wise Jaccard similarity matrix, cluster analysis with bootstrapping was applied on this dataset. The phenogram (Fig. 7)



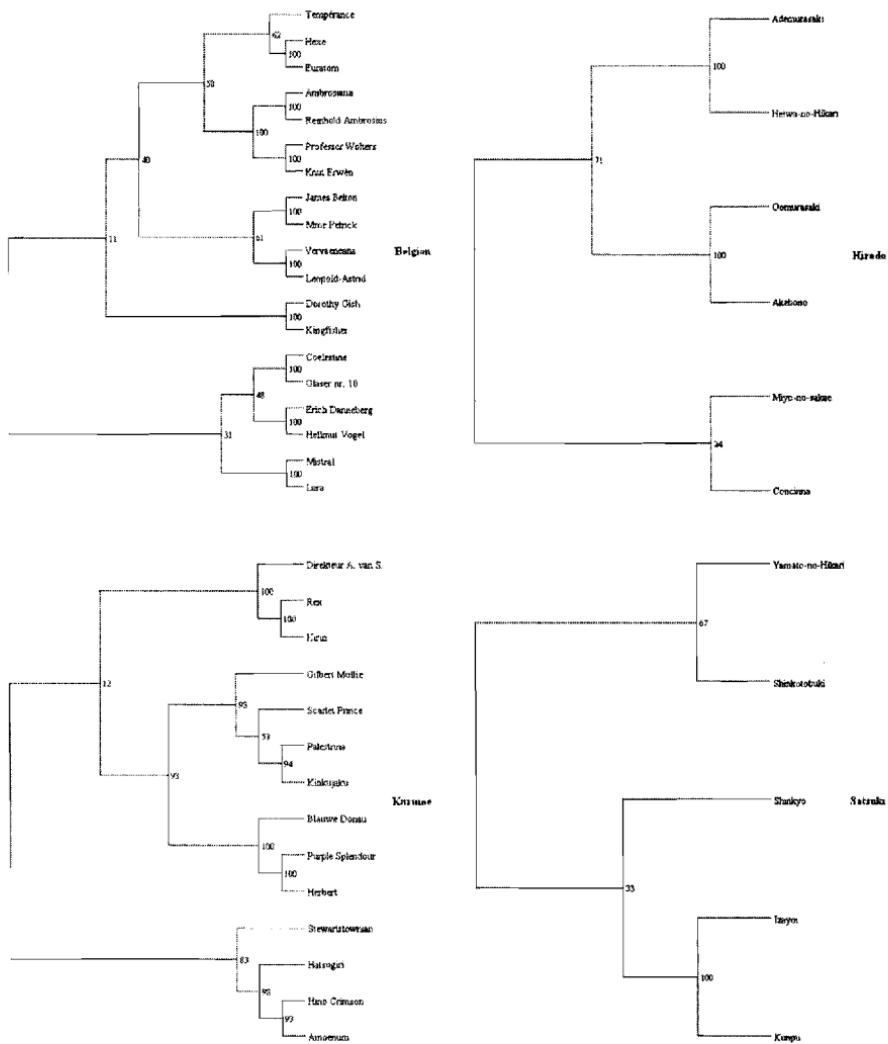
**Fig. 7.** Consensus tree between the species and varieties that were most likely to have been involved in the origin of the horticultural groups for the presence/absence of AFLP fragments. Numbers at the internodes indicate bootstrap values from 100 re-sampling cycles. The amount of replacement was fixed at 1% of the fragments per bootstrap. The arrows indicate the collocation of the four horticultural groups obtained by means of the same method [78].

was compared to the taxonomic classification [6]; for the Japanese species, at subsections and series level: Yamazaki, [99] and previous phylogenetic studies based both on PCR-RFLP analysis of cpDNA [53] and on *matK* and *trnK* intron sequences [58, 59]. Subgenus *Pentanthera* was included as an outgroup.

Six distinct groups were supported by high bootstrap values (Fig. 7). The first group included the Japanese species classified by Yamazaki [99] in the subsection *Tsutsusi* and it is divided into two subclusters: one contains *R. kaempferi* 'Doleshy' and *R. kiusianum* with its cultivars (series *Kaempferia*); the other includes *R. indicum*, *R. indicum* 'Crispiflorum' and *R. nakaharae* Hayata (series *Tsutsusi*). The second group consists of *R. simsii*, *R. Xpulchrum* 'Maxwellii' and some Chinese species. The third group is formed by a subcluster enclosing *R. eriocarpum* belonging to the subsection *Tsutsusi*, series *Tsutsusi*, and a subcluster containing all the species belonging to the subsection *Scabra*, series *Scabra*, considered in this study, except for *R. yedoense* Maxim. and *R. yedoense* var. *poukhanense* (H. Lév.) Nakai which have a Korean origin as the fourth group. The fifth and sixth groups include *R. tashiroi* Maxim. var. *lasiophyllum* Hatus. ex T. Yamaz and Chinese species that are not or poorly involved in the origin of the Belgian, Hirado, Kurume and Satsuki azaleas.

The well-characterized horticultural groups were then brought into the cluster analysis (results not shown). As indicated in Fig. 7, the Hirado and Belgian azaleas, morphologically characterized by large flowers, grouped together with the species belonging to the subsection *Scabra*, series *Scabra*; the Kurume and the Satsuki azaleas, both with small flowers, clustered with the species belonging to the subsection *Tsutsusi* (series *Kaempferia* and *Tsutsusi*).

Relationships between cultivars with a known pedigree may be verified in more detail in Fig. 8. Inside the Kurume group: 'Kirin' and 'Directeur Van Slycken' are clustered together with 'Rex' from which they derive, while 'Herbert' is grouped together with its sister 'Purple Splendour'. Referring to the Hirado group, 'Akebono' is clustered with its sport 'Oomurasaki'. 'Miyo-no-sakae', which is a hybrid between the Hirado 'Shirotae' and the Belgian azalea 'Albert Elizabeth', is clustered with 'Concinnina' differs and apart from other four cultivars. Among the Belgian azaleas: 'Reinhold Ambrosius' and 'Ambrosiana' with unknown parentage, but obtained from the same breeder, are clustered together. 'Knut Erwèn' is grouped with 'Professor Wolters', one of its indirect parentages; 'Glaser nr. 10', 'Euratom' and 'Hellmut Vogel' are clustered



**Fig. 8.** Consensus trees of the four reference horticultural groups based on AFLP data. Numbers at the internodes indicate bootstrap values from 100 re-sampling cycles. The amount of replacement was fixed at 1% of the fragments per bootstrap. Branch lengths are not on scale [78].

respectively with 'Coelestine', 'Hexe' and 'Erich Danneberg' from which they derive. 'James Belton', 'Mistral' and 'Lara' are clustered together with their common parentage 'Heiwa-no-Hikari', a Hirado azalea (data not shown). Among the Satsuki azaleas, no cultivars with known relationships were included in this study.

Results generated by AFLP fingerprinting are generally in agreement with other phylogenetic studies, geographical distribution and horticultural classification. Deviations, if observed, are in line with open issues concerning the taxonomical position of certain species or hybrids. The close relationship of *R. kaempferi* and *R. kiusianum* revealed by the present analysis was previously established in their cpDNA [52]. Likewise, the genetic closeness between *R. indicum* and *R. kaempferi* was formerly shown by Kurashige [58]. The origin of *R. Xpulchrum* is unclear. According to Heursel [36], this might be a Japanese natural garden hybrid. On the other hand, Chamberlain [7] considers *R. Xpulchrum* originated from *R. indicum* x 'Ledifolia'. The present study seems to suggest that Chinese azaleas, including *R. simsii*, could be involved in its origin. Concerning *R. Xmucronatum*, the hypothesis that it is an artificial hybrid of *R. ripense* Makino and its close ally *R. macrosepalum* Maxim. [syn. *R. stenopetalum* (Hogg) Mabb.] [6] seems to be confirmed. Based on morphological information, the taxonomic position of *R. tashiroi* has not been stabilized. It is considered an anomalous species that lies between section *Tsutsusi* Sweet and *Brachycalix* Sweet [6]. In the present analysis it was nested within a clade with a species of subgenus *Tsutsusi* section *Brachycalyx* as in Kurashige [58, 59], supporting Yamazaki's [99] view.

Regarding the horticultural groups, Hirado and Belgian azaleas share at least one important common ancestor, *R. scabrum*, originating from some small southern Japanese islands. *R. simsii* is generally considered to be the most important ancestor for Belgian azaleas due to the type of the flowers and the only ancestor exclusively distributed to China. In the present analysis *R. simsii* seems to be different, in agreement with a previous research [21]. Kurume azaleas are considered to be related to *R. kiusianum* and *R. kaempferi*, which are species native to Kyushu Island in Japan. The Satsuki azaleas are believed to have derived from *R. indicum* and *R. eriocarpum*.

## **Some Applied AFLP Studies for Characterization of Genetic Resources**

In Europe the origin of azaleas as a cultivated ornamental plant dates back to the XIX century, when some exotic species arrived from the American and Asian continents. Since then, several generations of flower growers have worked with the objective of creating new cultivars. Especially in Belgium and in the United Kingdom, and also in France and Germany, the hybridization activity reached remarkable levels.

From the 1970s onwards, Heursel searched for new introduction of Japanese and Chinese cultivars and wild species in order to broaden the breeders gene pool available in Western Europe. He has become famous for his Belgium bred Japanese hybrids and for creating the first hybrids between Japanese azaleas (typically outdoor grown medium sized shrubs) and pot azaleas for indoor use. Later on, through a bilateral collaboration, he obtained different seed lots from wild populations in mountain areas with an altitude ranging from 250 to 3,500 m from the Kunming Institute of Botany (China). The majority of these *Rhododendron* species belong to the *Tsutsusi* subgenus; eight are *R. simsii*.

An interesting link was established when researchers of the University of Turin started to characterize the cultural heritage of old ornamental cultivars located in historical gardens in the north of Italy. It appeared that from different European countries azaleas arrived in Italy, especially in the Lake Maggiore area (Piedmont and Lombardy districts – Northern Italy), where they have been hybridized and selected by local flower growers. The result is the creation of a significant gene pool, locally referred to as belonging to three different groups, according to the phenotype: (1) Indica, including plants characterized by large flowers, (2) Amoena, formed by azaleas with very small, purple flowers and (3) Japonica, a morphologically intermediate group of Indica and Amoena [77]. These group names seem to be related to an old and inaccurate nomenclature system.

The name Indica seems to be derived from a group also known as Southern Indian Hybrids having their origin in the evergreen azalea hybrids developed in Belgium around 1840 under another botanical name [61]. The azaleas were first classified by Linneaus in *Species Plantarum* (1753) as *Azalea indica*. The origin of the name Japonica probably lies in one of the several synonyms of *R. kiusianum* var. *kiusianum* {syn.: *R. indicum* var. *amoenum* (Lindley) Maxim. forma *japonicum* Maxim., *R. kaempferi* var. *japonicum* (Maxim.) Rehder, *R. indicum* var. *japonicum* (Maxim.) Makino, *R. obtusum* var. *japonicum* (Maxim.) Kitamura; Chamberlain and Rae [6]}. The group Amoena should be called in this way because of its relationship with the cultivar ‘Amoenum’ of the Kurume group.

Starting with material from these three “origins” (breeding gene pool, Chinese accessions, the Italian cultivars), an elaborated genetic characterization was performed by the use of AFLP markers. Similarities

of the fingerprint patterns were evaluated as an estimate for genetic conformity and for relatedness in view of: (1) use of the new material as a potential genetic resource for breeding, and (2) taxonomic classification of the cultivated azaleas. Plant material was harvested in the wild, from specimen in (botanical) gardens or from potted plants; leaves were immediately immersed in liquid nitrogen and stored at -80°C or were lyophilized for storage. DNA isolation was performed as in De Riek [18]. AFLP-reactions were run on an ABI Prism 377 DNA Sequencer using the commercially available kit for fluorescent fragment detection [70]. EcoRI and MseI were used for DNA digestion. Selective amplification was done using three fluorescent labelled EcoRI-MseI primer combinations with six selective bases: EcoRI-ACT/MseI-CTA, EcoRI-ACT/MseI-CAT and EcoRI-AAG/MseI-CTA [18].

### **Classification of a Breeders' Gene Pool**

Starting from a genetic similarity matrix based on the Jaccard measure, principal co-ordinate analysis was used to produce a 2-dimensional ordination for the breeders' gene pool (Fig. 9). The current pot azalea assortment was clearly separated from the wild *Rhododendron* species. The latter clustered together with the Kurume azaleas. Hirado azaleas were

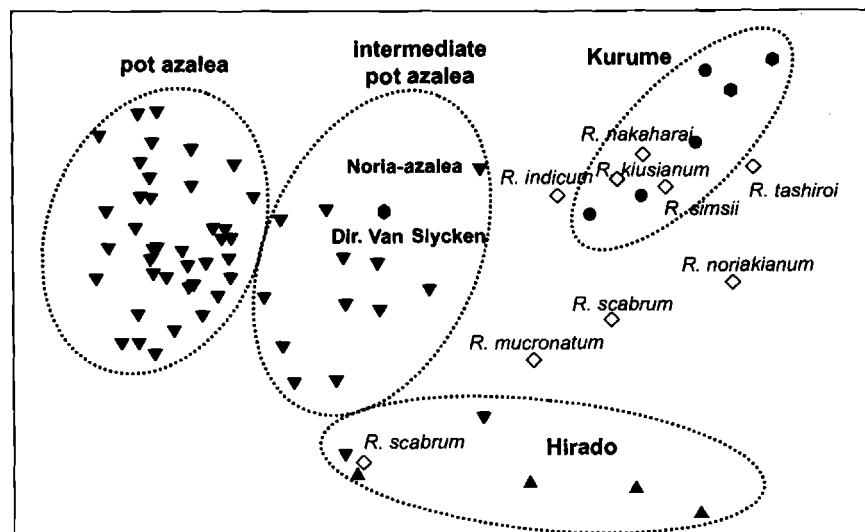


Fig. 9. PCO-plot of the breeders' gene pool based on AFLP data [18].

separately grouped together with *R. scabrum*. Both Kurume and Hirado azaleas were grouped with their most important ancestor. Within the *R. simsii* hybrids, two subgroups may be distinguished. The first group contains the archetype of a Belgian pot azalea, – globular shape, dark green leaves, early flowering and carmine red double flowers – although it has never been created so far in all its details as a single genotype. It is best characterized by its members as 'Madame Petrick' (1880), 'Paul Schaeeme' (1890), 'Ambrosiana' (1948) and 'Reinhold Ambrosius' (1930), which were before 'Hellmut Vogel' (1967) the most cultivated commercial pot azaleas. The second subgroup is a looser group of cultivars, intermediate to the other groups. They can be generally typified as flowering late and having single or half double flowers. A lot of them are older cultivars e.g., 'Coelestine' (unknown), 'Professor Wolters' (1871), 'Leopold-Astrid' (1923) or 'Tempérance' (1890). Some are cultivars that originate from intermediate crosses: 'Cheops', a new pyramidal azalea created from *R. noriakianum*, 'Lara' and 'Mistral', crosses between Hirado azaleas and *R. simsii* hybrids and 'Directeur Van Slycken', a cross between a Kurume azalea and a *R. simsii* hybrid. 'Coelestine', a very old cultivar with unknown origin, is probably also an intermediate form. It shares its globular shape, single flowers and late flowering date with *R. kiusianum* and the Kurume azaleas. Part of this character is also present in its daughter 'Glaser Nr. 10'. 'Dorothy Gish', 'Kingfisher', 'Hexe' and 'Euratom' show a hose-in-hose type of flower. This trait is typical for Kurume azaleas as 'Kirin' and 'Rex'. It is believed to be a mutation. The distinctive features of 'Tempérance' and 'James Belton' are very atypical for *R. simsii* hybrids: both have late flowering and are very woody and erect, 'Tempérance' shows very uncommon lilac to blue flowers, 'James Belton' has tiny green, hairy leaves.

### ***Position of Wild Material Collected from China Towards the Breeders' Gene Pool***

To evaluate the genetic conformity between and within Chinese populations and the breeders' gene pool, a pair wise similarity matrix was calculated based on the frequencies of the markers in each population (Euclidean distance). In Fig. 10, a cladogram shows the genetic similarity between assayed populations. The Belgian, Hirado and Kurume azaleas were considered to be separate and well-defined populations to be able to calculate marker frequencies. Bootstrap analysis was used to evaluate the

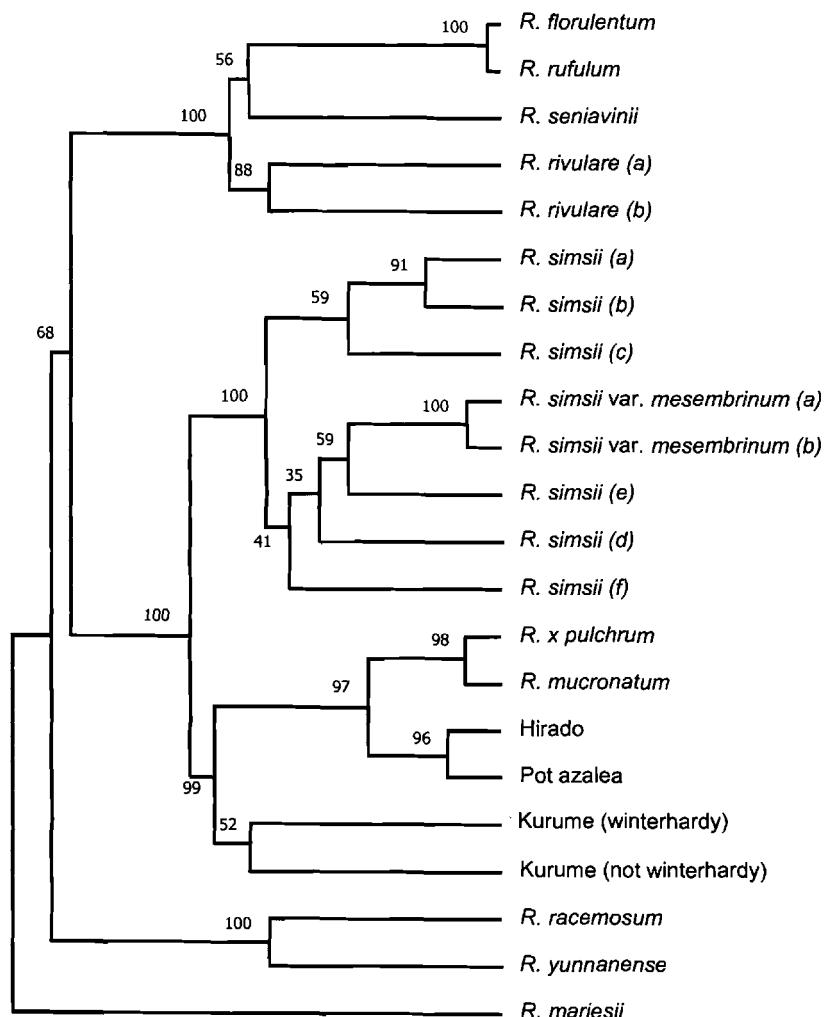


Fig. 10. UPGMA cladogram based on Euclidean distance between different populations or groups for AFLP marker frequencies.

reproducibility of the obtained groupings. Being supported by high bootstrap values, three major clusters and some outlying species can be distinguished. The first group contains *R. florulentum*, *R. rufulum*, *R. seniavini* Maxim. and two populations of *R. rivulare*. The second group solely holds *R. simsii* populations clustered with its subspecies *mesembrinum*. The third group consists of the cultivated species *R. Xpulchrum* and *R. Xmucronatum* and the different groups of azalea

hybrids. The remaining species (*R. racemosum*, *R. yunnanense* and *R. mariesii*) are clearly separated based on AFLP data. The *matK* data, previously obtained, confirmed the AFLP data. By combining both approaches, a complete picture emerged. Four distinct groups and some outlying species can be distinguished. The first group contains *R. florulentum*, *R. rufulum*, *R. seniavini* (no *matK* data) and two populations of *R. rivulare*. All these species belong to the *Tsutsusi* section of the *Tsutsusi* subgenus. The marker data for *R. florulentum* and *R. rufulum* appeared to be very similar. These species are also phenotypically very similar; taxonomic difference is based on the size of the flowers. The second group contains the *R. simsii* populations and cultivated *R. simsii* hybrids. *R. simsii*, generally considered to be the most important ancestor for pot azaleas due to type of flowers and the only ancestor exclusively distributed from China, seemed to be different. However, because a predominant number of *R. simsii* populations is included in this study, this might somewhat distort the AFLP ordination. The *R. simsii* populations are apart from the cultivated *R. simsii* hybrids; the latter are clustering with *R. Xpulchrum*, *R. scabrum*, and *R. Xmucronatum*. This was confirmed by the *matK* sequencing data. The third group consists of species belonging to subgenus *Rhododendron*, these are lepidote (scaly) species, namely, *R. yunnanense*, *R. racemosum*, *R. spiciferum* (no AFLP data) and *R. x duclouxii* (no AFLP data). The last group consists of *R. mariesii* and *R. farrerae* (outgroup for *matK* analysis) both belonging to the Section *Brachycalyx* of the Subgenus *Tsutsusi*.

Regarding the parentage of the cultivated hybrids, a closer clustering of *R. Xpulchrum* and *R. Xmucronatum* with Hirado (AFLP data) and *R. simsii* hybrids or pot azalea (AFLP data and *matK* data) was observed. Kurume azaleas (AFLP and *matK* data) were more distant. *R. Xpulchrum* is considered to be a hybrid between *R. indicum* and *R. Xmucronatum*. Both species are also related to pot azaleas. According to Sweet, who described the species *R. Xpulchrum* first, this taxon may be loosely grouped with the *Azalea indica* hybrids (a former erroneous name for *R. simsii* hybrids but also a term to describe the Satsuki azaleas) although it is not certain whether both type of hybrids have the same parentage [6]. *R. Xmucronatum* has been cultivated at least for 300 years in Japan and China. The original wild form, native to Japan, is var. *ripense* with rose pink flowers but widely cultivated together with var. *mucronatum*, an albino form. Hirado and pot azaleas share at least one important common ancestor, *R. scabrum*, originating from some small southern Japanese

islands. 'Phoeniceum', one of the early crossing parents with supposed *R. simsii* plants from Shanghai that lead to the current *R. simsii* hybrids, is accepted to be a cross between *R. Xmucronatum* and *R. scabrum* [28]. Kurume azaleas are related to *R. kiusianum* and *R. kaempferi*, species native to Japan but with a different geographic origin.

As stated by Galle [28], "the development of the Belgian Indian (synonym for pot azaleas) is shrouded in confusion beginning in the early 1800s. These first accessions gave rise to the early English Indian hybrids that became popular indoors and greenhouse plants in Europe around 1840. However according to Galle [28], "the major parents of the Belgian Indians were three forms of *R. simsii* collected by Robert Fortune in a Shanghai nursery and sent to England in 1851, but again under the incorrect name of 'Indica'. By 1854 these three varieties reached Belgium, launching an enormous breeding and growing program". Due to the historical confusion on correct classification of the primary accessions brought to Europe and on the unknown but certainly not negligible breeding input in Japan and China before introduction, the origin of Belgian pot azaleas will remain covered with eastern mystery and enigma. Based on the genetic conformity as revealed by AFLPs, a genetic continuum must be accepted, spanning a lot of species of the *Tsutsusi* subgenus. This continuum has been exploited in the past by both Asian and European breeders and still proves to be a good source for genetic variation and interesting forms.

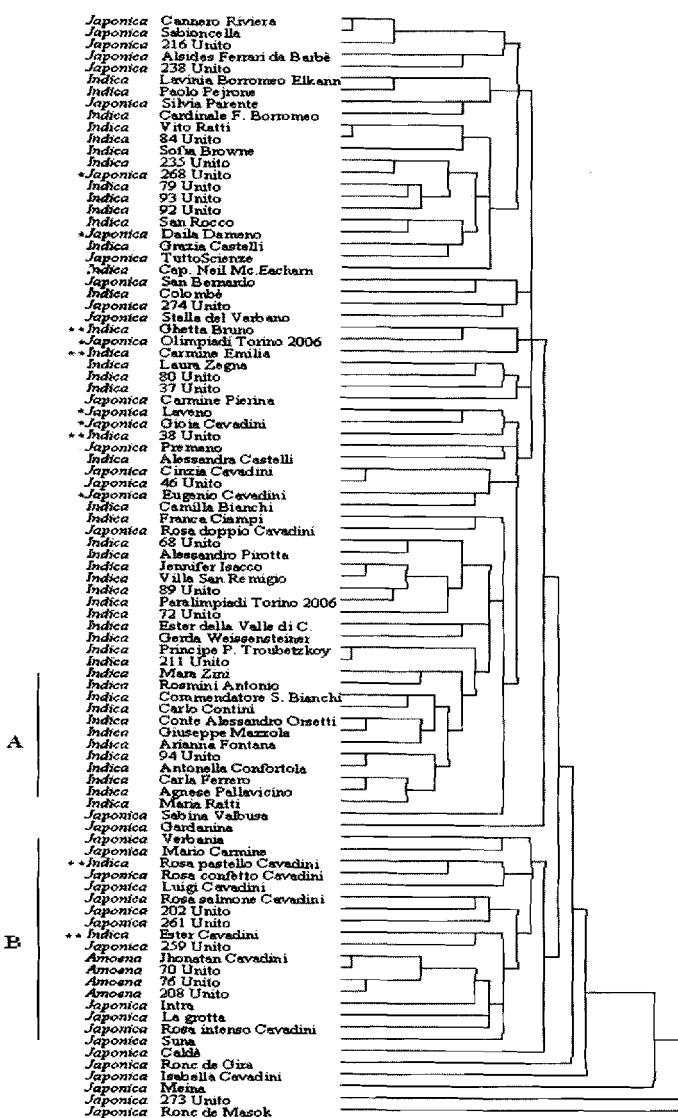
### ***AFLP Markers for Classifying an Italian Gene Pool of Evergreen Azaleas***

Although the Italian cultivars are recognized to represent an outstanding flower germplasm, no information about their genetic diversity has been available. A characterization based on morphological and molecular data could help to clarify their origins, classification and diversity [75]. The AFLP fingerprinting technique was employed to define the genetic profiles of 93 azalea accessions (48 Indica, 41 Japonica, 4 Amoena) grown in historical gardens and nurseries located in the Lake Maggiore area (Northern Italy) [76]. Aiming to provide information useful for their classification and pedigree reconstruction, a set of reference azaleas [78] was included in the analyses. At first, ordination analyses were used to investigate the DNA-based relatedness within the Italian gene pool and between the latter and the one used as a reference.

With some exceptions, in the cluster analysis performed on AFLP data (Fig. 11) the Indica accessions, mainly located in historical gardens and parks, clustered separately from the Japonica ones, mainly present in nurseries. The third minor local group (Amoena) formed a unique branch very close to the Japonica azaleas [76]. Continuations performed on both AFLP and morphological data set yielded groupings similar to the ones obtained with cluster analysis [78]. The pole of the Indica group is characterized by azaleas with large, simple, purple or, less frequently, white flowers and nine stamens; the pole of the Japonica group is formed by azaleas with white to pink bluish flowers and five stamens.

To estimate the genetic conformity between the cultivars of the Lake Maggiore area and the reference accessions, cluster analysis, average Jaccard similarities and assignment tests were evaluated. The dendrogram (results not shown) enabled to show how the Italian azaleas are genetically spread within the reference germplasm. With some exceptions, the Indica azaleas tended to cluster together with the Hirado and the Belgian azaleas; the Japonica and Amoena clustered with the Kurume and Satsuki ones. The average Jaccard similarity taken over all pair wise comparisons between all plants of two horticultural groups was calculated [78]. Here, (1) the Indica group showed to be clearly related to the Belgian and Hirado hybrids; (2) the Amoena group was probably derived from the cv 'Amoenum', being in this way more related to the Kurume hybrids; and (3) the Japonica group, although diverse and non-homogeneous, showed higher similarities to the Amoena, Kurume and Satsuki hybrids.

Assignment tests, both on the level of the groups as on individual plant level for the reference and Italian genotypes, were applied to further evaluate the relatedness of the groups or individual accessions. In the assignment tests on the level of the groups, Amoena is confirmed to be a close and unique group also showing some affinity to Kurume hybrids; Indica appears as a typical group too, highly distinctive from all other hybrids but more related to Belgian and Hirado hybrids; Japonica includes fairly diverse azaleas, also relating back to the Indica hybrids. Performing the assignment of the individual Italian accessions, the definition of the Italian groups as separate and distinct gene pools seems to be justified. In fact, the highest numbers of accessions were assigned in their putative group [78]. Their genetic closeness to the supposed ancestor species included in this study is also revealed.



**Fig. 11.** Ordination of azalea cultivars from the Lake Maggiore area (Northern Italy) based on AFLP data, with indication of their belonging to the local horticultural groups [77]. The pole of the Indica (A) and Japonica (B) groups are indicated. Italian cultivars have been registered in the International Rhododendron Register of the Royal Horticultural Society. The cultivars which are still without a registered name are indicated with the label "Unito" (= Università di Torino)

\* Cultivar included in the Indica group by the assignment test (software Doh) performed on AFLP data

\*\* Cultivar included in the Japonica group by the assignment test (software Doh) performed on AFLP data

Based on the degree of similarity between each accession belonging to a horticultural group and its most related plant, some considerations about the origin of several Italian accessions could be drawn, supported by morphological data. Italian accessions with high Jaccard coefficients (= 0.7) should be sports (e.g., 'Carla Ferrero'/'Agnese Pallavicino' and 69 Unito/'Villa San Remigio') or seedlings (e.g., 'Cannero Riviera'/'Sabioncella' and '207 Unito/94 Unito). Likewise, Jaccard coefficient values included between 0.5 and 0.7 should be indicative of direct or indirect parentages [78].

Results demonstrated that the AFLP technique together with morphological characterization could be a useful tool for clarifying the origin and classification of evergreen azalea gene pools. The Indica azaleas, genetically close to the Hirado and Belgian group, were derived especially from the Belgian ones. The Japonica revealed to be much more heterogeneous, related to the other hybrid groups. All this leads to suggestion that Japonica azaleas should not be considered as a unique distinct group. No doubt exists about the origin of the Italian Amoena group that lies in the Kurume one. A new classification could thus be drawn and be the basis for future strategies of germplasm conservation and breeding programs.

## CANDIDATE GENES AS FUNCTIONAL MARKERS FOR THE ASSESSMENT OF GENETIC DIVERSITY AND MAPPING OF PLANT QUALITY TRAITS

### Development of EST Markers

Molecular markers can be very effective for studying diversity within species or the relationships between closely related species and hybrids. DNA fingerprinting is usually based on techniques amplifying arbitrary sequences or specific loci. Among the arbitrary DNA fingerprinting systems, the multilocus AFLP technique [95] has already proved to be successful in assessing the genetic variation of a breeders' collection of evergreen azaleas [18] and for classifying an Italian gene pool of evergreen azaleas [78]. Among the DNA fingerprinting techniques based on specific loci, microsatellite markers have the potential to provide reliable and highly informative genetic data. Molecular markers from the transcribed region of the genome also have great potential for applications in plant genotyping. Expressed Sequence Tag (EST) markers

have already shown a high level of polymorphism in various plants but they were not previously developed in *Rhododendron*.

Starting from two cDNA libraries made from flowers of the *R. simsii* hybrid 'Flamenco' and the species *R. luteum*, 323 cDNA fragments were randomly picked and sequenced. The putative functions of the cDNA fragments were determined by comparison of the sequences with EMBL accessions using FASTA33 [69]. Reliable homologies were found for 31% of all fragments. Primers were developed on 127 cDNA fragments and used for PCR amplification on six different azalea cultivars and species. PAGE was used to separate the fragments and the presence of polymorphic bands was evaluated. In the end, this resulted in 45 polymorphic EST markers (De Keyser [15], 32 of which originated from 'Flamenco' cDNA. Although markers were developed in less related species, markers from *R. luteum* could be applied in *R. simsii* hybrids and also 'Flamenco' EST markers generated polymorphisms in *R. luteum*. Therefore, the use of the markers will not be limited to the species in which they were developed, but they have a good chance to be useful in the whole *Rhododendron* genus as functional markers.

### **Comparison of AFLP, STMS and EST Markers for the Assessment of Genetic Diversity**

This study refers to the use of these EST markers in comparison to AFLP and STMS (sequence tagged microsatellite site) markers to define genetic differentiation among a set of cultivars. The material is representative of the four more or less well-defined groups of evergreen azalea cultivars (Belgian, Hirado, Kurume and Satsuki azaleas) and of the species and varieties that were most likely to have been involved in their origin. These techniques have shown different capacities in discriminating and assessing genetic relationships in evergreen azaleas [79]. AFLPs revealed the highest polymorphism detection capacity while STMS and EST markers showed the highest discrimination capacity. This study evaluated the usefulness of AFLP, STMS and EST markers in: (1) evaluation of the transferability of the different markers across the subgenus *Tsutsusi*; (2) differentiating between horticultural hybrid groups of evergreen azaleas; (3) assessing the extent of genetic variation within horticultural groups; and (4) revealing the genetic relationships between species and cultivars.

## **Marker Transferability Across Species and Horticultural Groups**

At first, the transferability of 10 EST markers (developed from the 'Flamenco' cDNA library) and the 10 STMS markers developed by Dendauw [20] across the species and cultivars genotyped were investigated. Regarding the STMS loci, four of ten proved to be not fully transferable across the sample set: AZA-005 and AZA-006 did not amplify consistently across the broad range of accessions included here; the loci AZA-004 and AZA-007 produced profiles with multiple stutter bands, very unreliable to score. These four loci were, therefore, omitted from any further analyses. Band profiles obtained using the other STMS loci were easily scored, except for AZA-011 that produced profiles with ambiguous stutter bands. However, allele calling still was feasible. A total of 99 alleles were detected at six STMS loci. The number of alleles per locus ranged from 12 to 29, with an average of 16.5 alleles per locus. The six STMS primer sets amplified across all the horticultural groups (except for AZA-003 that did not amplify any allele in the Satsuki cultivar 'Yamato-no-Hikari') and almost all the different species included in this study, exhibiting a great transferability. In particular, AZA-008, AZA-009 and AZA-011 primer pairs amplified in all the species; AZA-010 did not amplify in *R. kanehirae*, neither AZA-003 in *R. simsii*, *R. eriocarpum* and *R. seniavini* nor AZA-002 in *R. indicum*, *R. kiusianum* 'Myoken', *R. Xmucronatum*, *R. scabrum*, *R. eriocarpum* and *R. florulentum*.

Regarding the EST loci, one primer combination, EST-03, produced too many aspecific bands. Therefore, it was excluded from further analysis. At these nine loci, a total of 38 alleles could be scored. All primer sets allowed amplification across the Hirado and Belgian azaleas (except for EST-02 that did not amplify any allele in the cultivar 'Leopold Astrid'). Across the Kurume cultivars, no alleles were found in 'Rex' and 'Hino Crimson' by EST-04 neither in 'Hatsugiri' by EST-10 nor in 'Herbert' by both EST-04 and EST-10. No EST primer pair could amplify across all the Satsuki azaleas; only the cultivar 'Shinkyo' could be typed by all of them. Also, in several accessions of *R. kiusianum*, 'Myoken', 'Fugen-no-tsuki', 'Miyama Shikibu', few primer sets amplified successfully; only by means of EST-02 and EST-06 all the *R. kiusianum* accessions showed amplified alleles. Apart from the *R. kiusianum* accessions, the nine EST primer sets exhibited a great transferability across the species. Few cases of no amplifications were found (*R. macrosepalum* by EST-01, *R.*

*yedoense* by EST-07, *R. simsii* and *R. schlippenbachii* by EST-09, *R. rivulare* and *R. loniceriflorum* PC. Tam by EST-10). Primer EST-05 amplified only one allele in all but one sample; in *R. loniceriflorum* no allele was obtained.

### **Genetic Variation Within Horticultural Groups**

Knowledge of genetic diversity and relationships between genotypes is required for efficient rationalization and utilization of germplasm resources. Furthermore, it is important for planning future breeding programs. To attribute the distribution of the genetic variation in AFLP, STMS and EST patterns among and within horticultural groups (populations), AMOVA [25] was computed at one level from squared Euclidean distances with the number of permutations set at 1000, using the Arlequin software [81]. The Species group was not included in this analysis being, by its nature, artificial and comprising very different genotypes. For all marker techniques used, most of the genetic diversity was attributable to differences among cultivars within horticultural groups (89.08, 89.09 and 85.56% for AFLP, STMS and EST markers, respectively, Table 9). The precise origin of horticultural evergreen azalea cultivars is still not well established. Recently, several studies have been undertaken on this subject [18, 21, 32, 53]. Based on the genetic conformity as revealed by AFLP markers [78], a genetic continuum must be accepted, spanning a lot of species of the *Tsutsusi* subgenus. Common parental species could explain the low genetic diversity attributable to differences among horticultural groups. Besides, intensive breeding activities within groups, where seed is commonly mass selected on the basis of phenotypic characteristics, have yielded highly significant differentiation among cultivars within each horticultural group. Comparing the marker techniques, EST markers outperformed AFLP and STMS markers concerning  $F_{st}$  values (approx. 0.15 versus 0.11) indicating a low but significant differentiation among horticultural groups. This might be due to the more conservative nature of the polymorphism which EST markers can detect, that could assist in better emphasizing differences among parental species.

### **Exploration of Segregation of Plant Quality Traits and the Construction of Genetic Maps**

When breeding new Belgian pot azaleas, seedlings are at first selected based on their flower characteristics, it is only at a later stage that other

**Table 9.** AMOVA (Analysis of Molecular Variance) Significance tests (1,023 permutations) at one level based on AFLP, STMS and EST data sets (binary Euclidian distance coefficient). The Species group was not included in this analysis being, by its nature, artificial and comprising very different genotypes

	Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation
<b>AFLP</b>	Among hybrid groups	3	33.900	0.623	10.92**
	Within hybrid groups	40	203.235	5.081	89.08
	Total	43	237.135	5.704	
	Fixation Index $F_{st}$ : 0.10920 <sup>a</sup>				
<b>STMS</b>	Among hybrid groups	3	11.942	0.219	10.91**
	Within hybrid groups	40	71.643	1.791	89.09
	Total	43	83.585	2.010	
	Fixation Index $F_{st}$ : 0.10908 <sup>a</sup>				
<b>EST</b>	Among hybrid groups	3	10.804	0.226	14.44**
	Within hybrid groups	40	53.651	1.341	85.56
	Total	43	64.455	1.567	
	Fixation Index $F_{st}$ : 0.14438 <sup>a</sup>				

<sup>a</sup> Significant with  $p < 0.001$  as tested by 1,023 permutations

important plant quality traits such as leaf shape, leaf color, growth vigor, compactness, etc., are evaluated. Attractive flowering plants are still too often rejected at a later stage of breeding because the plant as a whole doesn't comply with the present cultivation standard. The inheritance of flower color has been well-studied by Heursel [38], but information on the heritability of other plant quality traits is still lacking and a more specific and consequently a more efficient selection is required when new crosses are made. For this purpose, different crosses are made for each trait between parents that reveal the extreme phenotypes of the range of the examined trait. All the siblings of a cross segregating for one specific trait are then scored individually. To assure that these plants are evaluated under natural growth conditions, they receive no treatments with growth inhibitors or other sprays.

The phenotype of these (semi)-quantitative traits will be integrated on a genetic map of azalea by QTL-mapping. Flower color is inherited as a qualitative trait, hence mapping is quite straightforward. For the construction of the genetic map of Belgian pot azalea, AFLP and STMS markers are generated on four populations that are segregated for flower

color and several plant quality traits under investigation. Since azalea has a rather high chromosome number ( $2n=26$ ), integration of the separate population maps into a consensus genetic map of azalea will only be successful when there is at least one segregating co-dominant locus for each chromatide. STMS markers [20, 22], and also EST markers [15], will be used for this purpose. Following the candidate gene approach, EST markers will also be developed for genes that are known to be involved in flower color biosynthesis, or which, according to literature, are reported to have a major impact on the generation of other complex plant quality traits. In this way, phenotypic and genetic data (candidate genes) will both be integrated on the genetic map of azalea. In case both are located at the same mapping position, these genes are proven to be directly involved in the creation of the phenotypic variation of the trait. Nevertheless, it is very likely that it is not the genes themselves, but it is the transcription factors which are the switches that regulate the phenotype of the trait. In that case, phenotype and genotype will be mapped at different positions, but phenotype is then expected to be mapped together with the true regulators, the transcription factors. To confirm this theory, gene expression profiles of the candidate genes will be generated in several populations using real-time RT-PCR [14] and QTL mapping will integrate these data with the genetic map (Fig. 12).

When successful, the strategy can be expanded to search for genes involved in traits such as lime, cold or drought tolerance in other species and cultivars from the *Rhododendron* genus. Keeping this in view, the breeding objectives (Table 10) and the potential approach to reach them (N. Kobayashi, Shimane Univ.) are very illustrative.

## Acknowledgment

The authors wish to thank Romain Uytterhaegen for his hands-on experience in azalea breeding and cultivation; Veerle Buysens, Veerle Cools and Laurence Desmet for their sincere efforts in making DNA preparations and AFLP, STMS and EST-reactions; and Hilde Carlier and Leen Leus for ploidy analysis. Deep gratitude is also expressed to all the staff of the Biotech Laboratory for extending their support and expertise. The authors are especially indebted to the Chinese Kunming Institute for Botany for their excellent cooperation and exchange of material and trainees; and to Dr. David Chamberlain of the Royal Botanical Garden in Edinburgh (UK) who unraveled the *Tsutsusi* taxonomy for this study and was of great assistance in determination of the Chinese accessions.

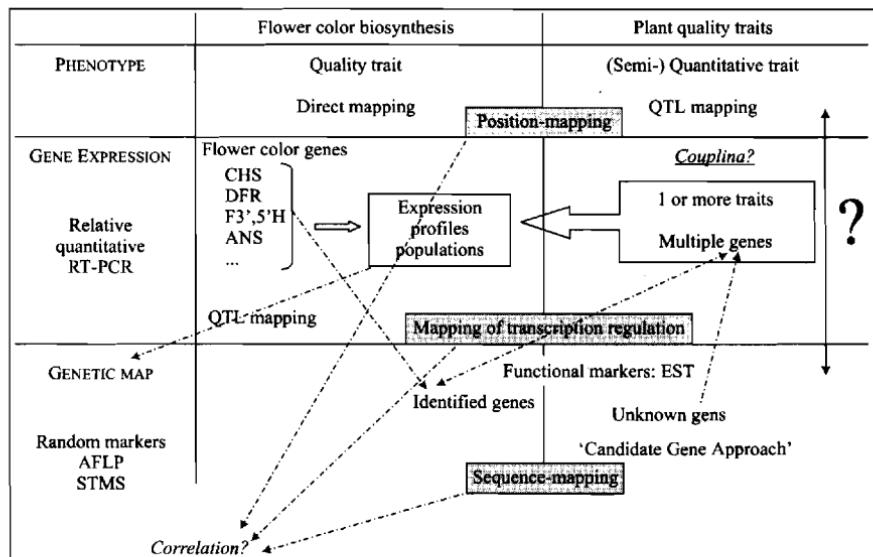


Fig. 12. Strategy followed for mapping of plant quality traits in Belgian pot azalea.

**Table 10** List of different breeding objectives for evergreen azalea hybrids, available genetic resources and possible approaches to reach these goals (as formulated by N. Kobayashi).

Breeding objective	Available genetic resources	Possible approach
1. Adaptation to extreme conditions - Lime tolerance - Strong acid tolerance - SOx Gas tolerance - Cold, frost tolerance - Drought, heat tolerance	Japanese wild azalea Chinese wild azalea Japanese old cultivar	Exploration of wild habitat Root system research Crossbreeding program Screening of tolerance
2. New flower forms - Everlasting form - Spider form - Hose-in-hose form - Double form	Japanese old cultivar Pot azalea cultivar	Crossbreeding program Cytological observation Inheritance analysis MADS gene analysis
3. New flower color - Blue flower - Green flower	Japanese old cultivar Pot azalea cultivar	Anthocyanin gene isolation Trans formation (cyanidin block)

Table 10 contd.

Table 10 contd.

- Yellow flower	Chinese wild azalea Deciduous azalea Scale azalea	Mutation breeding Crossbreeding program
4. New flowering cycle - Ever-flowering (three seasons) - Spring and autumn flowering	Chinese wild azalea Japanese wild azalea Japanese old cultivar	Crossbreeding program Ever-flowering gene isolation
5. Addition of other new characters - Fragrance  - Creeping type  - Brightness of leaf surface	Chinese wild azalea Deciduous azalea Scale azalea Taiwanese wild azalea Japanese wild azalea Japanese old cultivar Japanese wild azalea Japanese cultivar Taiwanese wild azalea	Crossbreeding program Inheritance analysis Fragrance gene isolation Crossbreeding program Inheritance analysis Evaluation as cover plant Crossbreeding program Inheritance analysis

## References

- [1] Anderson W. A revised tissue culture medium for shoot multiplication of Rhododendron. J Am Soc Hort Sci 1984; 109: 343-347.
- [2] Badger B. In search of a yellow evergreen azalea (how to hybridize for a yellow evergreen azalea). J Am Rhododendron Soc 1988; 42: 74-79.
- [3] Bennett MD, Leitch IJ. Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 2005; 95: 45-90.
- [4] Brown GK, Craven LA, Udoovicic F, Ladiges PY. Phylogeny of Rhododendron section Vireya (Ericaceae) based on two non-coding regions of cpDNA. Plant Syst Evol 2006a; 257: 57-93.
- [5] Brown GK, Craven LA, Udoovicic F, Ladiges PY. Phylogenetic relationships of Rhododendron section Vireya (Ericaceae) inferred from the ITS nrDNA region. Aust Syst Bot 2006b; 19: 329-342.
- [6] Chamberlain DF, Hyam G, Argent G, Fairweather G, Walter K. The genus *Rhododendron*. Its classification & synonymy. Edinburgh: Royal Botanical Garden, 1996.
- [7] Chamberlain DF, Rae SJ. A revision of Rhododendron IV Subgenus Tsutsusi. Edinburgh J Bot 1990; 47: 89-200.
- [8] Cox P, Cox K. Encyclopedia of Rhododendron hybrids. London: Batsford Ltd, 1995.
- [9] Cox P, Cox K. Encyclopedia of Rhododendron species. Perth: Glendoick Publishing, 1997.

- [10] Cuevas J, Rallo L, Rapoport H, Lavee S, Klein I. Staining procedure for the observation of olive pollen tube behaviour. *Acta Hort* 1994; 356: 264-267.
- [11] Cullen J. The logic of *Rhododendron* classification. *Rhododendrons* (1992). London: The Royal Horticultural Society, 1991; 44: 14-19.
- [12] Corriveau J, Coleman A. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Amer J Bot* 1988; 75: 1443-1458.
- [13] Davidian H. The *Rhododendron* species, Vol. IV. Azaleas. Portland, Oregon: Timber Press, 1995.
- [14] De Keyser E, De Riek J, Van Bockstaele E. Gene expression profiling of key enzymes in azalea flower colour biosynthesis. *Acta Hort*, Proceedings of the 22nd Int. EUCARPIA Symposium Section Comamentols 2007; 743: 115-120..
- [15] De Keyser E, Scariot V, De Riek J. Development of EST Markers and evaluation of their use in evergreen azalea analysis. *Acta Hort*, Proceedings of the 22nd Int. EUCARPIA Symposium Section Comamentols 2007; 743: 107-113.
- [16] De Loose R. The flower pigments of the Belgian hybrids of *Rhododendron simsii* and other species and varieties from *Rhododendron* subseries *obtusum*. *Phytochem* 1969; 8: 253-259.
- [17] De Loose R. Characterisation of *Rhododendron simsii* Planch. Cultivars by flavonoid & isozyme markers. *Sci Hort* 1970; 11: 175-182.
- [18] De Riek J, Dendauw J, Mertens M, De Loose M, Heursel J, Van Bockstaele E. Validation of criteria for the selection of AFLP markers to assess the genetic variation of a breeders' collection of evergreen azaleas. *Theor Appl Genet* 1999; 99: 1155-1165.
- [19] De Schepper S, Leus L, Mertens M, Van Bockstaele E, De Loose M. Flow cytometric analysis of ploidy in *Rhododendron* (subgenus *Tsutsusi*). *Hort Sci* 2001; 36: 125-127.
- [20] Dendauw J, De Riek J, Arens P, Van Bockstaele E, Vosman B, De Loose M. Development of sequenced tagged microsatellite site (STMS) markers in azalea. *Acta Hort* 2001; 546: 193-197.
- [21] Dendauw J, De Riek J, De Loose M, Van Bockstaele E. Identification of 33 Chinese *Rhododendron* species using matK sequences and AFLP data. *Acta Hort* 2002; 572: 169-177.
- [22] Dunemann F, Kahnau R, Stange I. Analysis of complex leaf and flower characters in *Rhododendron* using a molecular linkage map. *Theor Appl Genet* 1999; 98: 1146-1155.
- [23] Eeckhaut T, Samyn G, Van Bockstaele E. Interspecific breeding in the *Rhododendron* genus involving *R. simsii* hybrids. *Acta Hort* 2003; 612: 165-172.
- [24] Eeckhaut T, Van Laere K, De Riek J, Van Huylenbroeck J. Overcoming interspecific barriers in ornamental plant breeding. In: Texeira da Silva J, ed. *Floriculture, Ornamental and Plant Biotechnology: advances and topical issues* (1st Edition). London, Global Science Books 2006; 540-551.
- [25] Excoffier L, Smouse P, Quattro J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genet* 1992; 131: 479-491.
- [26] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evol* 1985;39: 783-791.
- [27] Fitch WM. Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst Zool* 1971; 20: 406-416.
- [28] Galle F. Azaleas. Portland, Oregon: Timber Press, 1987.

- [29] Gao-LianMing, Yang-JunBo, Zhang-ChangQin, Li-DeZhu. Phylogenetic relationship of subgenus Tsutsusi (*Rhododendron*) based on ITS sequences. *Acta Bot Yunnanica* 2002; 24: 313-320.
- [30] Goetsch L, Eckert AJ, Hall BD. The molecular systematics of *Rhododendron* (Ericaceae): a phylogeny based upon PRB2 gene sequences. *Syst Bot* 2005; 30: 616-626.
- [31] Hagemann R. Plastid genetics in higher plants. In: Herrman R, ed. *Cell organelles*. New York: Springer, 1992: 65-96.
- [32] Handa T, Eto J, Kita K, Kobayashi N. Genetic diversity of Japanese wild evergreen azaleas in Kyushu (south main island of Japan) characterized by AFLP. *Acta Hort* 2002; 572: 159-161.
- [33] Heursel J. Die Vererbung des Merkmals "hose in hose" (Doppelkronigkeit) bei Azaleen. *Gartenwelt* 1976; 76: 111-113.
- [36] Heursel J. Inheritance of flower colours and breeding of evergreen azaleas. *J Am Rhododendron Soc* 1987; 41: 141-145.
- [38] Heursel J. Azalea's: oorsprong, veredeling en cultivars. Tielt, Belgium: Lannoo, 1999.
- [35] Heursel J, De Roo R. Polyploidy in evergreen azaleas. *Hort Sci* 1981; 16: 765-766.
- [37] Heursel J, Garretsen F. Inheritance of corolla size, number of stamens & percentage of plants with petaloid stamens in evergreen azaleas (*Rhododendron* Subsect. *obtusa*). *Plant Breeding* 1989; 103: 304-309.
- [34] Heursel J, Horn W. A hypothesis on the inheritance of flower colours and flavonoids in *Rhododendron simsii* Planch. *Z Pflanzenzüchtung* 1977; 79: 238-249.
- [39] Hermsen J. Nature, evolution, and breeding of polyploids. *IOWA State J Res* 1984; 58: 461-474.
- [40] Hilu KW, Liang H. The matK gene: sequence variation and application in plant systematics. *Am J Bot* 1997; 84: 830-839.
- [41] Hogenboom N. A model for incongruity in intimate partner relationships. *Euphytica* 1973; 22: 219-233.
- [42] Hogenboom N. Incompatibility and incongruity: two different mechanisms for the non-functioning of intimate partner relationships. *Proc Royal Soc London* 1975; 188: 361-375.
- [43] Hwang-ShihYing, Hsu-KuoKai. Molecular phylogeny of eight Taiwanese *Rhododendron* species based on chloroplast *trnF-trnL* DNA sequences. *Taiwan J For Sci* 2001; 16: 153-160.
- [44] Irving E. Concerning the origin and distribution of *Rhododendrons*. *J Am Rhododendron Soc* 1993; 47: 139-145.
- [45] Janaki AE, Enoch I, Bridgewater M. Chromosome numbers in species of *Rhododendron*. *The Rhododendron Yearbook*. Royal Hort Soc 1950; 78-91.
- [46] Johnston S, Den Nijs T, Peloquin S, Haneman R. The significance of gene balance to endosperm development in interspecific crosses. *Theor Appl Genet* 1980; 57: 5-9.
- [47] Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ. Evolution of genome size in *Brassicaceae*. *Ann Bot* 2005; 95: 229-235.
- [48] Kehr A. Hybridizing azaleas. In: Galle F, ed. *Azaleas*. Portland: Oregon, Timber Press, 1987.
- [49] Kho Y, Baer J. Improving the cross *Rhododendron impeditum* x *Rhododendron 'Elizabeth'* by temperature treatment. *Euphytica* 1973; 22: 234-238.

- [50] Kita K, Kurashige Y, Yukawa T, Nishimura S, Handa T. Intergeneric hybridization between Menziesia and Rhododendron based on molecular phylogenetic data. *J Jpn Soc Hort Sci* 2005; 74: 51-56.
- [51] Knox R, Williams E, Dumas C. Pollen, pistil and reproductive function in crop plants. In: Janick J, ed. *Plant Breeding Reviews*, Vol. 4., Westport, AVI Publishing Co, 1984: 9-79.
- [53] Kobayashi N, Handa T, Yoshimura K, Tsumura Y, Arisumi K and Takayanagi K. Evidence for introgressive hybridization based on chloroplast DNA polymorphisms and morphological variation in wild evergreen azalea populations in the Kirishima Mountains, Japan. *Edinburgh J Bot* 2000; 57: 209-219.
- [52] Kobayashi N, Kawashima S, Handa T, Takayanagi K, Arisumi K. Introgression in Japanese evergreen azaleas (*Rhododendron kiusianum* and *R. kameapheri*). *Acta Hort* 1998; 454: 325-328.
- [54] Kron KA. A revision of Rhododendron section Pentanthera. *Edinburgh J Bot* 1993; 50: 249-363.
- [55] Kron KA. Phylogenetic relationships of Rhododendroideae (Ericaceae). *Amer J Bot*. 1997; 84: 973-980.
- [56] Kron KA. Phylogenetic relationships and major clades of Rhododendron (Rhodoreae, Ericoideae, Ericaceae) In: Argent G, McFarlane M, eds. *Rhododendrons in Horticulture and Science*. Edinburgh: Royal Botanic Garden, 2002: 79-85.
- [57] Kunishige M, Kobayashi Y. Chromatographic identification of Japanese azalea species and their hybrids. In : Luteyn JL, O'Brien ME (eds.). *Contributions toward a classification of Rhododendron*. New York, New York Bot. Garden, 1980.
- [58] Kurashige Y, Mine M, Kobayashi N, Handa T, Takayanagi K, Yukawa T. Investigation of sectional relationships in the genus Rhododendron (Ericaceae) based on matK sequences. *J Jpn Bot* 1998; 73: 143-154.
- [59] Kurashige Y, Etoh JI, Handa T, Takayanagi K, Yukawa T. Sectional relationships in the genus Rhododendron (Ericaceae): evidence from matK and trnK sequences. *Plant Syst Evol* 2001; 228: 1-14.
- [60] Leach D. *Rhododendrons of the world*. New York: Charles Scribner's Sons, 1961.
- [61] Lee FP. *The Azalea Book*. Van Nostrand Company Inc., Am Hort Soc 1958.
- [62] Longly B. Centre d' etude de la reproduction vegetale. Rapport 1992-1994, Convention D1/4-3309/5489A, 1994.
- [63] Maddison DR. The discovery and importance of multiple islands of most-parsimonious trees. *Syst Zool* 1991; 40: 315-328.
- [64] McCubbin A, Kao T. Self-incompatibility and pollen rejection in angiosperms. In: Mohan S, Sopory S, Veilleux R, eds. *Current plant science and biotechnology in agriculture: In vitro haploid production in higher plants*. Dordrecht, Kluwer Academic Publishers, 1996.
- [65] Micishita A, Ureshino K, Miyajima I. Plastome-genome incompatibility of Rhododendron serpyllifolium (A. Gray) Miq. to evergreen azalea species belonging to series Kaempferia. *J Jpn Soc Hort Sci* 2002; 71: 375-381.
- [66] Miyajima I, Ureshino K, Kobayashi N, Akabane M. Flower color and pigments of intersubgeneric hybrid between white-flowered evergreen and yellow-flowered deciduous azaleas. *J Jpn Soc Hort Sci* 2000; 69: 280-282.
- [67] Ohri D. Genome size variation and plant systematics. *Ann Bot* 1998; 82: 75-83.

- [68] Okonkwo C, Campbell R. Pollen tube growth in pistils of *Rhododendron* following intra- and interspecific pollinations. *Genet Breeding* 1990; 23: 9-15.
- [69] Pearson WR, Lipman DJ. PNAS 1988; 85: 2444-2448.
- [70] Perkin-Elmer. AFLP™ Plant Mapping Kit: Protocol 1995.
- [71] Pickersgill B, Hayward M, Bosemark N, Romagosa I. Interspecific hybridization by sexual means. In: Hayward M, Bosemark N, Romagosa I, eds. *Plant Breeding: principles and prospects*. (Plant Breeding Series; Vol. 1) Dordrecht, Kluwer Academic Publishers, 1993: 63-78.
- [72] Preil W, Ebbinghaus R. Bastardierungen von Topfazaleen (*Rhododendron simsii*) mit anderen *Rhododendron*-Arten. *Rhododendron und immergrüne Laubgehölze*, Jahrbuch 1985: 85-92.
- [73] Raghavan V. Molecular embryology of flowering plants. New York: Cambridge University Press, 1997.
- [74] Rouse J. Inter- and intraspecific pollinations involving *Rhododendron* species. *J Am Rhododendron Soc* 1993; 47: 23-45.
- [75] Scariot V. The DNA-typing of Ornamental Plants: Evergreen Azaleas (*Rhododendron* spp.) and Old Garden Roses (*Rosa* spp.). University of Torino, PhD. thesis 2006a.
- [76] Scariot V, Handa T, De Riek J. Morphological characteristics and AFLP markers for classifying an Italian gene pool of evergreen azaleas. *Acta Hort* 2006b; 714:221-225.
- [77] Scariot V, Remotti D, Merlo F. Le Azalee Sempreverdi del Lago Maggiore. Supplemento al n.50 dei "Quaderni della Regione Piemonte - Agricoltura" 2006c.
- [78] Scariot V, Handa T, De Riek J. A contribution to the classification of evergreen azalea cultivars located in the Lake Maggiore area (Italy) by means of AFLP markers. *Euphytica*, in press.
- [79] Scariot V, De Keyser E, Handa T, De Riek J. Comparative study of the discriminating capacity and effectiveness of AFLP, STMS and EST markers in assessing genetic relationships among evergreen azaleas. *Plant Breeding*, 2007; 126: 207-212.
- [80] Scheiber SM, Jarret RL, Robacter CD, Newman M. Genetic relationships within *Rhododendron* L. section *Pentanthera* G. Don based on sequences of the internal transcribed spacer (ITS) region. *Sci Hort* 2000; 85: 123-135.
- [81] Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis, Version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland, 2000.
- [82] Sharma H. How wide can a wide cross be? *Euphytica* 1995; 82: 43-64.
- [83] Spethmann W. A new infrageneric classification and phylogenetic trends in the genus *Rhododendron* (Ericaceae). *Plant Syst Evol* 1987; 157: 9-31.
- [84] Stebbins G. The inviability, weakness, and sterility of interspecific hybrids. *Adv Genet* 1958; 9: 147-215.
- [85] Swofford DL, Maddison WP. Reconstructing ancestral character states under Wagner parsimony. *Math Biosci* 1987; 87: 199-299.
- [86] Swofford DL. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, IL, 1993.
- [87] Ureshino K, Miyajima I, Akabane M. Effectiveness of three-way crossing for the breeding of yellow-flowered evergreen azalea. *Euphytica* 1998; 104: 113-118.
- [88] Ureshino K, Miyajima I, Ozaki Y, Kobayashi N, Michishita A, Akabane M. Appearance of albino seedlings and ptDNA inheritance in interspecific hybrids of azalea. *Euphytica* 1999; 110: 61-66.

- [89] Ureshino K, Kawai M, Miyajima I. Factors of unilateral cross incompatibility between several evergreen azalea species and *Rhododendron japonicum* *flavum*. J Jpn Soc Hort Sci 2000; 69: 261-265.
- [90] Väinölä A. Polyploidization and early screening of *Rhododendron* hybrids. Euphytica 2000; 112: 239-244.
- [91] Vanoverschelde M. Essais de polyploidisation dans le cadre de la sterilité dans le genre *Rhododendron*. Mémoire en vue de l'obtention du grade d'ingenieur agronome, UCL, 1989.
- [92] Van Tuyl J, Van Dien M, Van Creij M, Van Kleinwee T, Franken J, Bino R. Application of in vitro pollination, ovary culture, ovule culture and embryo rescue for overcoming incongruity barriers in interspecific *Lilium* crosses. Plant Sci 1991; 74: 115-116.
- [93] Van Tuyl J, De Jeu M. Methods for overcoming interspecific crossing barriers. In: Shivanna K, Sawhney V, eds. Pollen biotechnology for crop production and improvement. Cambridge University Press, 1997: 273-292.
- [94] Vervaeke I, Parton E, Maene L, Deroose R, De Proft M. Prefertilization barriers between different Bromeliaceae. Euphytica 2001; 118: 91-97.
- [95] Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Horres M, Fijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 1995; 21: 4407-4414.
- [96] Voss D. What is an azalea? J Am Rhododendron Soci 2001; 55: 188-192.
- [97] Williams E, Rouse J, Palser B, Knox R. Reproductive biology of *Rhododendron*. Hort Rev 1990; 12: 1-67.
- [98] Yamaguchi S. In vitro culture of remote hybrid seedlings aiming to breed new yellow flowered evergreen azalea. Plant Cell Incompatibility Newsletter 1986; 18: 50-51.
- [99] Yamazaki T. A revision of the genus *Rhododendron* in Japan, Taiwan, Korea and Sakhalin. Tokyo, Tsumura Laboratory, 1996: 1-125.
- [100] Yao J, Cohen D, Rowland R. Plastid DNA inheritance and plastome-genome incompatibility in interspecific hybrids of *Zantedeschia* (Araceae). Theor Appl Genet 1994; 88: 255-260.
- [101] Yukawa T, Koga S, Handa, T. DNA uncovers paraphyly of *Dendrobium* (Orchidaceae). In: Andrews S *et al.*, eds. Taxonomy of Cultivated Plants. London: Royal Botanic Gardens, Kew, 2000: 351-354.
- [102] Zenkteler M. In vitro fertilization and wide hybridization in higher plants. Crit Rev Plant Sci 1990; 9: 267-279.

# Phylogeny, Diversity and Evolution of Eucalypts

MARGARET BYRNE

Science Division, Department of Environment and Conservation, WA 6983, Australia

## ABSTRACT

Eucalypts are a large group in the family Myrtaceae endemic to the Australian region. Intensive research is being carried out in eucalypt genomics worldwide providing molecular tools that can readily be used in biodiversity studies. This review highlights the contribution of molecular genetics in further understanding the phylogeny, diversity and evolution of eucalypts. Molecular phylogenies have supported the hypothesis from the limited fossil record that the eucalypt group dates from the Late Cretaceous and became more dominant during the Cenozoic with increased aridity. Molecular studies indicate recent speciation of eucalypts generally through allopatric speciation processes. Phylogenetics of populations have been useful for clarification of taxonomy at lower hierarchical levels where morphology can be misleading. Molecular population genetics indicates evolutionary processes associated with high levels of genetic diversity that are maintained through a mixed mating system with long distance pollen dispersal. Further development of eucalypt genomics will enable assessment of the level of variation in functional genes, leading to greater understanding of adaptation and ecological determinants of genetic variation.

**Key Words:** Phylogeny, evolution, genetic diversity, eucalypts

---

*Address for correspondence:* Dr Margaret Byrne, Science Division, Department of Environment and Conservation, Locked Bag 104, Bentley Delivery Centre, WA 6983, Australia. Tel: + 618 9334 0503, Fax + 618 93340327, E-mail: margaret.byrne@dec.wa.gov.au

## INTRODUCTION

Eucalypts are a very large group in the family Myrtaceae with 783 taxa recognized in EUCLID, the latest electronic treatment of the group [161]. Most eucalypts are endemic to Australia although ten Australian species are also found in southern New Guinea, and five are endemic to islands off the north coast of Australia, including New Guinea, Timor, Indonesia and the southern Philippines [63, 149]. Within Australia eucalypts are a dominant component of the flora except for the most arid areas, although they may still be locally dominant in areas of greater water availability [148]. They grow in a range of vegetation communities including tall forests, open woodlands and mallee shrublands.

The distinctive features of eucalypts, with a few exceptions, are possession of one or two opercula covering the flower buds and lack of free petals and sepals [10], heterophylly and development of ovulodes (sterile ovules or chaff). Beyond these common traits, eucalypts display a wide range of form and are highly morphologically diverse. They can be multi-stemmed shrubs (known as mallees) or single stemmed trees, and have a varied range of bark types, leaf shapes and sizes. The majority of eucalypts produce a large number of small flowers, with numerous stamens generally white, cream or pale in colour, but a number of species produce fewer large flowers, often red or yellow, that are attractive to birds [72]. Most eucalypts are entomophilous or zoomophilous, i.e., pollinated by insects, birds, possums and bats [72].

Many eucalypt species have commercial value, particularly in forestry, hence there is considerable research in eucalypt genomics worldwide, and the status of genomic research in eucalypts has recently been reviewed by Poke et al. [138]. Despite its commercial forestry perspective, progress in genomics provides valuable research tools, such as genetic resources, gene discovery and identification of gene function, that enable further investigation and understanding of evolutionary processes and ecological responses in eucalypts. Molecular analysis, through use of gene sequences and other molecular markers that have become available in the course of genomic research and development, has enabled increased insights into the evolution, phylogeny and diversity of eucalypts beyond that which can be derived solely from morphological data. In this review of the evolution of eucalypts, the contribution of molecular genetics to the understanding of the phylogeny, diversity and evolution of eucalypts has been highlighted.

## GENOME STRUCTURE

Eucalypts are diploid with a haploid chromosome number of 11; although  $n=12$  has been observed in 10 species, no polyploidy has been documented [153]. Thus, eucalypts have maintained the base chromosome number and high seed production, which are considered to be primitive conditions within the Myrtaceae [154]. The size of the eucalypt genome is small, estimated as between 370-700 million base pairs in various species using flow cytometry [54]. This is comparable with the size of rice (420-466 Mbp [53, 186]) and poplar (473 Mpb, <http://www.genome.jgi-psf.org/Poptr1/Poptr1.info.html>) genomes, and much smaller than the estimates of genome size in pines (20,830 - 26,920 Mbp [11]). The organization of the eucalypt genome appears to be conserved across species, as genetic mapping has demonstrated high synteny of markers across linkage groups at a range of taxonomic levels; within subgenus *Sympyomyrtus* between *Eucalyptus urophylla* S.T. Blake and *E. grandis* Hill ex Maiden [14], between *E. grandis* and *E. globulus* Labill. [126], and between *E. grandis*, *E. urophylla*, *E. tereticornis* and *E. globulus* [106]; across subgenera between *E. nitens* (Deane and Maiden) Maiden (subgenus *Sympyomyrtus*) and *E. marginata* Donn. Ex Smith (subgenus *Eucalyptus*) (Moran and Byrne unpublished data); and between the genera *Eucalyptus* and *Corymbia* [159].

Organelle genome structure has also been investigated. As in most angiosperms, the chloroplast genome in eucalypts is maternally inherited [23, 115], and has the same structure with two inverted repeats separating large and small single copy regions. Sequencing shows that the chloroplast genome has a size of 160,286 bp and high homology with coding regions in the chloroplast sequence of other species (e.g., *Nicotiana* and *Oenothera*) but high divergence in intragenic regions [169]. Maternal inheritance of the mitochondrial genome has been determined in *E. globulus* [177], but the size and structure of the genome have not been investigated.

## EVOLUTIONARY HISTORY AND PHYLOGENY

### Classification and Distribution

The genus *Eucalyptus* L'Hér. sensu stricto is currently recognized as a lineage within a wider 'eucalypt group' that includes six other genera: *Angophora* Cav., *Corymbia* Hill and Johnson, *Eucalyptopsis* C.T. White,

*Allosyncarpia* S.T. Blake, *Stockwellia* Carr, Carr & Hyland and *Arillastrum* Pancher ex Baill. [97]. The last four mentioned are small, monotypic or bitypic rainforest genera; *Eucalyptopsis* (two species) is found in New Guinea, the Moluccan Archipelago and Woodlark Island, *Allosyncarpia* is endemic to the Arnhem Plateau of northern Australia, *Stockwellia* is endemic to Atherton in the wet tropics of northeast Queensland, and *Arillastrum* grows in New Caledonia. *Angophora* and *Corymbia* are sclerophyllous trees that occur in the tropical and warm temperate regions of northern and southern Australia, but are largely absent from the cooler regions in the south (e.g., Victoria and Tasmania) [96].

The genera *Eucalyptus*, *Angophora* and *Corymbia* have generally been referred to as eucalypts in the broad sense in various classifications [19, 96] but that of Pryor and Johnson [147] has been most widely used. This classification recognized two genera, *Angophora* and *Eucalyptus*, with seven informal subgenera within *Eucalyptus*: *Corymbia*, *Blakella*, *Eudesmia*, *Gaubea*, *Idiogenes*, *Sympphyomyrtus* and *Monocalyptus*. Johnson [77] subsequently raised a group of four tropical box species within *Sympphyomyrtus* to subgenus level, named *Telocalyptus*, and Hill and Johnson [64] formally recognized subgenus *Eudesmia*. In 1995, Hill and Johnson [63] recognized the similarity between subgenera *Corymbia* and *Blakella* and combined them into a single genus, *Corymbia*. Brooker [19] disputed this revision and included both *Angophora* and *Corymbia* into *Eucalyptus*. He recognized 13 subgenera within the genus (six monotypic), including *Corymbia*, *Blakella* and *Angophora*; he accepted *Eudesmia* and *Idiogenes* (*E. cloeziana* F. Muell.), split subgenus *Gaubea* into *Acerosa* (*E. curtisii* F. Muell.) and *Cuboidea* (*E. tenuipes* (Maiden and Blakely) Blakely and C.T. White), raised *Primitiva* (*E. rubiginosa* Brooker) from *Monocalyptus* (renamed *Eucalyptus* as the type subgenus), raised *Cruciformes* (*E. guilfoylei* Maiden) and *Alveolata* (*E. microcorys* F. Muell.) from *Sympphyomyrtus*, and renamed *Telocalyptus* as *Minutifructus*. A summary of the distribution and classification of the major groups of eucalypts is given in Table 1.

## Morphology

Historically, classifications of the eucalypts have been based on morphological characters and are essentially phenetic, although both Pryor and Johnson [147] and Brooker [19] recognized that classification should reflect phylogeny. Pryor and Johnson [148], along with Briggs and Johnson [13], considered that the major groups within *Eucalyptus*

**Table 1.** Summary of the distribution and classification of groups within the eucalypts according to Pryor and Johnson [147], Hill and Johnson [63] and Brooker [18]

Group	Distribution	Pryor and Johnson	Hill and Johnson	Brooker
<i>Angophora</i>	Eastern mainland Australia	Genus <i>Angophora</i>	Genus <i>Angophora</i>	Subgenus <i>Angophora</i>
<i>Corymbia</i>	Mainland Australia, New Guinea	Subgenus <i>Corymbia</i>	Genus <i>Corymbia</i>	Subgenus <i>Corymbia</i>
<i>Blakella</i>	Tropical, subtropical, arid Australia, New Guinea	Subgenus <i>Blakella</i>	Genus <i>Corymbia</i>	Subgenus <i>Blakella</i>
<i>Symphyomyrtus</i>	Australia, some species in Timor, adjoining islands and New Guinea	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>
<i>E. deguptia</i> , <i>E. brachyantha</i> , <i>E. howittiana</i> , <i>E. raveretiana</i>	Northern Australia, <i>E. deguptia</i> endemic to New Guinea, New Britain, Ceram, Sulawesi and Mindanao	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>
<i>E. microcorys</i>	Southeast Queensland, New South Wales	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>
<i>E. gulffoylei</i>	Southwest Australia	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>	Subgenus <i>Sympphyomyrtus</i>
<i>Monocalyptus</i>	Southern Australia, some species to the north of Queensland	Subgenus <i>Monocalyptus</i>	Subgenus <i>Monocalyptus</i>	Subgenus <i>Monocalyptus</i>
<i>E. rubiginosa</i>	Southeast Queensland	Subgenus <i>Monocalyptus</i>	Subgenus <i>Monocalyptus</i>	Subgenus <i>Monocalyptus</i>
<i>Eudesmia</i>	Mainland Australia except for Victoria	Subgenus <i>Eudesmia</i>	Subgenus <i>Eudesmia</i>	Subgenus <i>Eudesmia</i>
<i>E. cloeziana</i>	Northeast Australia	Subgenus <i>Idiogenes</i>	Subgenus <i>Idiogenes</i>	Subgenus <i>Idiogenes</i>
<i>E. curtisii</i>	Northeast Australia	Subgenus <i>Gaubea</i>	Subgenus <i>Gaubea</i>	Subgenus <i>Gaubea</i>
<i>E. tenuipes</i>	Northeast Australia	Subgenus <i>Gaubea</i>	Subgenus <i>Gaubea</i>	Subgenus <i>Gaubea</i>

represented a number of independent lines of evolution that were equivalent to other genera within Myrtaceae.

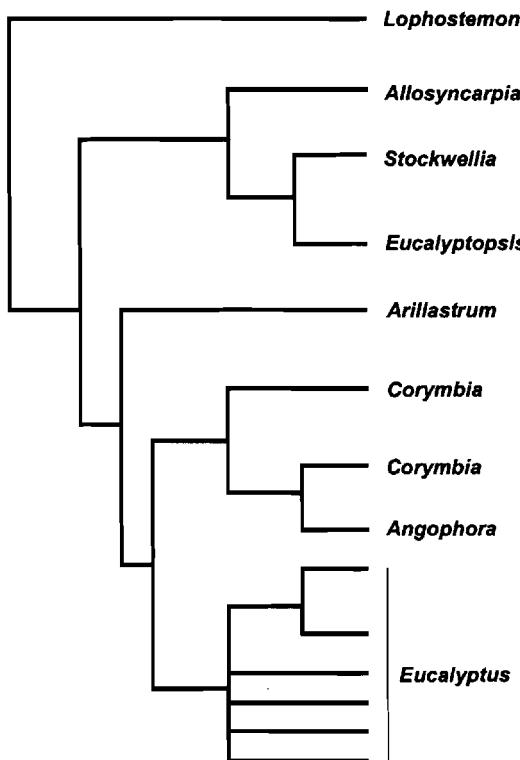
Morphological cladistic analyses of relationships within eucalypts were undertaken by Ladiges and Humphries [89] and Ladiges et al. [95]. The latter study supported a sister relationship between *Angophora* and *Corymbia*, which share a number of features that distinguish them from *Eucalyptus*: pinnate leaf venation, raised oil glands in seedlings and cap cells of oil glands with micropapillae [90]. *Angophora* is restricted to eastern Australia, whereas *Corymbia* is more widespread with several sublineages recognized (subgenera in Hill and Johnson's [63] classification) that are also supported by morphological characters (red bloodwoods, yellow bloodwoods and spotted gums, and ghost gums or paper fruited bloodwoods).

In *Eucalyptus* sensu stricto, Ladiges et al. [95] identified three major lineages characterized by unique morphological features, with subgenus *Eudesmia* sister to the *Sympyomyrtus* and *Monocalyptus* clades. Within the *Monocalyptus* lineage, subg. *Idiogenes* was sister to *Monocalyptus*, and *Gaugea* was sister to the *Monocalyptus*/*Idiogenes* clade. The *Sympyomyrtus* clade included *Telocalyptus* as sister to the rest of *Sympyomyrtus*. *Eudesmia* is a small subgenus with 11 species endemic to southwest Western Australia, and the other 12 scattered in central and northern Australia. In contrast, *Sympyomyrtus* is the largest subgenus and is widely distributed throughout Australia. *Sympyomyrtus* comprises six major sections that tend to be dominant in different regions: *Latoangulatae* (previously *Transversaria*) in eastern Australia; *Bisectae* and *Dumaria* in southwestern Australia; *Exsertaria* in northern and eastern Australia; *Maidenaria* in southeastern Australia; and *Adnataria* throughout Australia [96]. The east-west division of major lineages within *Sympyomyrtus* indicates that the subgenus is relatively old within *Eucalyptus*. Subgenus *Eucalyptus* (*Monocalyptus*) is a well-defined relatively large group characterized by a single operculum [96] and is predominantly distributed in southern Australia, but extends to northeast Queensland. Ladiges et al. [92] recognized three major lineages within *Monocalyptus* as sections *Rubiginia* (raised to subgenus *Primitiva* in Brooker [19]), *Eucalyptus* and *Diversifolia* (Brooker [19] recognized 11 sections within subgenus *Eucalyptus*). Ladiges et al. [92] recognized that the species restricted to southwestern Australia are not monophyletic and eastwest sister-species relationships indicate that the evolution of lineages within groups is relatively old.

## Molecular Phylogeny

Due to lack of clarity over the phylogeny and classification of eucalypts, numerous molecular phylogenetic studies have been carried out using the nuclear genes 5S rDNA [175, 176], ITS (internal transcribed spacer) [166, 168, 176] and ETS (external transcribed spacer) [133]; and chloroplast genes, *psbA-tmH* spacer, *tmL* intron and *tmL-tmF* spacer [176], and *tmL – tmF*, *atpB – rbcL* and *psbA-tmK* spacers in addition to partial sequences from *matK* and *ndhF* [183]. These studies have generally focused on higher-level relationships amongst the genera at various hierarchical taxonomic levels. The ITS data set of Steane et al. [168] is the most comprehensive in terms of genera, subgenera and sections sampled, and the number of species included. All molecular studies identified three major clades, the rainforest genera (*Eucalyptopsis*, *Stockwellia* and *Allosyncarpia*), the *Angophora* and *Corymbia* clade, and the *Eucalyptus* clade. The position of the small rainforest genus, *Arillastrum*, was unresolved being placed either at a polytomy with the other rainforest genera or as sister to *Angophora/Corymbia* (combined analysis of *tmL – tmF*, *atpB – rbcL* and *psbA-tmK* spacers and partial sequences from *matK* and *ndhF*), sister to *Angophora/Corymbia/Eucalyptus* (*psbA-trnH*) or basal to the group (ITS, 5S, ETS). All molecular studies supported the sister relationship between *Angophora/Corymbia* and *Eucalyptus*. The monophyly of *Corymbia* identified in morphological analysis is supported by the *tmL-tmF* spacer data sets and more recently by ETS [133] using a larger sample of species, but in the other studies *Corymbia* was paraphyletic. All data sets support the subgeneric groupings within *Corymbia*. Ochieng et al. [132] have also used microsatellite loci to investigate the relationships between *Corymbia* and *Angophora*. Their phylogenetic analysis also support the monophyly of *Corymbia* and the sister group relationship with *Angophora*. A summary of the relationships between genera based on molecular phylogenies is presented in Fig. 1.

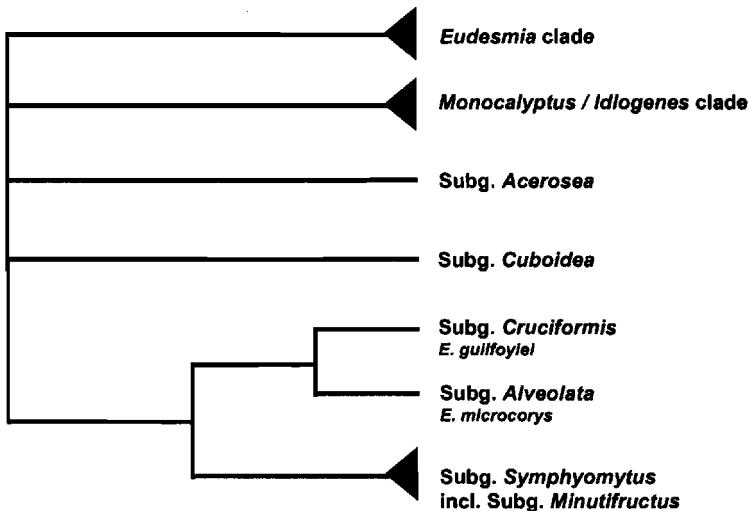
Most of the molecular phylogenies have generally insufficient sampling within *Eucalyptus*, or the DNA regions used were not sufficiently variable, to allow definite evaluation of relationships between subgenera. The major clades derived from morphological analysis, *Eudesmia*, *Monocalyptus/Idiogenes/Gaubea* and *Symphyomyrtus/Telocalyptus* [95], are also identified in the ITS data set [168] and the combined chloroplast sequence analysis of Whittlock et al. [183]. The phylogenies indicated that *E. curtisii* (subg. *Acerosea*), which lacks opercula, may be basal in *Eucalyptus*; *E. guilfoylei* (Cruciformes) and *E. microcorys*



**Fig. 1.** Summary of the molecular phylogeny of the eucalypt group based on nuclear (ITS, ETS and 5S rDNA) and chloroplast (trnL intron, trnL-trnF, atpB – rbcL, psbA-trnH and psbA-trnK spacers, matK and ndhF) DNA sequence data. *Lophostemon* is an outgroup taxon. (Modified from Ladiges et al. [97]).

(*Alveolata*) are possibly relictual lineages related to the *Sympyomyrtus* clade, the position of *E. tenuipes* (subgenus *Cuboidea*) was unresolved, and the tropical boxes (subg. *Minutifructus*) were nested within subgenus *Sympyomyrtus*. Whittlock et al. [183] suggested that the subgenus rank of the tropical boxes should be revised due to their lack of monophyly. A summary of the relationships within *Eucalyptus* ss. based on molecular phylogenies is presented in Fig. 2.

The ITS study of Steane et al. [168] is the only molecular study with sufficient sampling to evaluate relationships within subgenus *Sympyomyrtus*. Sections *Adnataria* and *Maidenaria* were shown as monophyletic; *Bisectae* and *Exsertaria* polyphyletic or paraphyletic, and *Dumaria* was unresolved. Some caution may need to be exercised when



**Fig. 2.** Summary of the molecular phylogeny of major clades within *Eucalyptus* *sensu stricto* based on nuclear (ITS) and chloroplast (trnL-trnF, atpB-rbcL and psbA-trnK spacers and partial sequences from matK and ndhF) sequence data. (Modified from Ladiges et al. [97]).

interpreting phylogenies based on ITS sequences because a recent study has identified divergent paralogues in eucalypt ITS sequences [5]. Bayly and Ladiges [5] amplified parologue ITS sequences from eight species of subgenus *Eucalyptus* and showed that in all species the paralogues were highly divergent from the orthologous sequence. They have since confirmed the presence of paralogues in *Eudesmia*, *Corymbia*, *Angophora*, *Stockwellia* and *Arillastrum* [6]. Unrecognized inclusion of paralogues could explain unexpected divergences in ITS sequences in previous phylogenies, for example the long branch length of *E. pleurocarpa* Schauer (formerly erroneously known as *E. tetragona* {R.Br.} F. Muell.) and its sister relationship to other sequences in the ITS phylogeny of Udovicic and Ladiges [176]. The ITS sequences used in Steane et al. [168] all represent the orthologues identified by Bayly and Ladiges [5] (Steane and Bayly unpublished data). Ochieng et al. [131] have also identified ITS paralogues in *Corymbia* and *Angophora*. Different paralogues within individuals showed greater divergence than when the same parologue was compared across species. Different topologies were obtained for phylogenies of *Corymbia* and *Angophora* constructed with the putative orthologue, or the putative parologue [131] with the orthologue showing

*Corymbia* as paraphyletic with *Angophora*, and the parologue showing *Corymbia* as monophyletic and sister to *Angophora*.

Relationships between sections *Exsertaria*, *Latoangulatae* (*Transversaria* of Pryor and Johnson [147]) and *Maidenaria* were investigated using a single copy nuclear gene, cinnamoyl CoA reductase, but the presence of intragenic recombination made it unsuitable for phylogenetic reconstruction [139]. The utility of amplified fragment length polymorphisms (AFLPs) for resolving relationships in section *Maidenaria* is being examined and preliminary results show resolution of species into clades that are largely concordant with Brooker's [19] taxonomic series (McKinnon unpublished data).

## **Evolutionary History**

### **Fossil Record**

The fossil record of eucalypts is limited compared to that of other taxa [65]. Fossil records indicate the development of eucalypt type morphology within the family Myrtaceae during the late Cretaceous (100-65 Myr) and Palaeogene (65-26 Myr). Fossil pollen attributed to bloodwood/*Angophora* (*Myrtaceidites*) has been recorded in Gabon, Colombo, Borneo and the Antarctic Peninsula from around 86 to 65 Myr [124] and the first recorded eucalyptoid leaves and fruits date from the middle Eocene (around 48 Myr) and were located in central Australia [55].

Macrofossils attributed to *Eucalyptus* have been found in the middle Miocene (approximately 21 Myr) deposits in New South Wales [69] and Victoria [43, 69, 140]. Well-searched deposits in coastal areas from the preceding Eocene (54-38 Myr) and Oligocene (38-26 Myr) eras failed to reveal eucalyptus-like macro fossils [101] in these mesic areas. In more arid areas of central Australia, eucalypt-like fruits have been described [100] near Woomera, and these are likely to be Miocene in age [1].

Macrofossils from the Warrambungle Mountains in New South Wales reveal two forms, one with similarities to *Angophora/Corymbia*, and the other more typical of *Eucalyptus* [69]. Holmes et al. [69] suggested divergence of these two lineages prior to the middle Miocene. Flower buds and fruits from late Miocene (10-5 Myr) deposits in Bacchus Marsh, Victoria [8] are similar to present day subgenus *Eucalyptus* species and are evidence of the differentiation between this subgenus and *Sympyomyrtus* at that time [96]. Leaves of *E. pluti* McCoy, very similar to the extant *E. globulus*, have been identified in Pliocene (5-2 Myr) deposits in Victoria

[109], and eucalypt pollen described as *E. spathulata* Hook. has been identified in Pliocene sites in southwest and southeast Australia [107]. Eucalypt pollen becomes more frequent in recent sediments from 120,000 years to the present and is often associated with elevated levels of charcoal particles [160].

### ***Biogeography and Phylogeography***

Phylogeny of the Myrtaceae family showed eucalypts as an early diverging lineage within the Leptospermoideae subfamily in the late Cretaceous, and supports an Australasian origin for extant lineages of Myrtaceae [173]. The four genera, *Arillastrum*, *Eucalyptopsis*, *Allosyncarpia* and *Stockwellia*, are considered to be relicts of ancient Gondwanic rainforest [152], and their association in the eucalypt group implies a rainforest origin for eucalypts. Ladiges et al. [97] reviewed the biogeographical connections of *Arillastrum* and the *Eucalyptopsis* group and concluded that separation of *Arillastrum* on New Caledonia dated to the rifting of New Caledonia from eastern Gondwana in the late Cretaceous, 65-70 Myr. The divergence of genera in the *Eucalyptopsis* group was related to progressive contraction of rainforest and development of monsoonal climate through the Miocene (26.5 Myr). Other authors have identified similar northern tracks and dispersal routes [20, 41, 42]. Molecular dating of the ITS phylogeny by Crisp et al. [39], using the divergence of *Arillastrum* from eucalypts of 70 Myr, led to estimates of divergence in the *Eucalyptopsis* group of 37-35 Myr, similar to divergence estimates of 30-25 Myr by Ladiges et al. [97]. Thus, the biogeographic patterns reviewed by Ladiges et al. [97] imply that the eucalypts are an old group with the two major lineages of *Angophora/Corymbia* and *Eucalyptus* diverging after 65 Myr.

Ladiges et al. [97] also reviewed the biogeography of *E. deglupta* Blume, a species endemic to New Guinea, New Britain, Ceram, Sulawesi and Mindanao in the Philippines, and concluded that its distribution in southeast Asia may be related to accretion of terranes between 10-2 Myr. The position of *E. deglupta* as either sister to or basal in the *Sympphyomyrtus* clade implies that the divergence of major groups in *Sympphyomyrtus* is of similar age with later species radiations. The distribution of three other species, *E. urophylla*, *E. wetarensis* L.D. Pryor, and *E. orophila* L.D. Pryor, in the Lesser Sunda Islands, suggests divergence from the Australian taxa in the Pliocene (5-2 Myr) with compression and uplift of Timor. Biogeography implies divergence of

major clades in *Sympphyomyrtus* between 15-2 Myr with later radiation of species through the Pleistocene (1.8-0.01 Myr). Crisp et al. [39] estimated divergence events based on molecular dating the ITS phylogeny and they were all much earlier than those estimated from biogeography; 38-26 Myr for divergence of *E. deglupta*, 30-13 Myr for divergence of sections within *Sympphyomyrtus*, and 17-7 Myr for divergence of *E. urophylla*. The date estimated for divergence of *E. deglupta* and *E. urophylla*, appear to be related to spurious accessions that show much greater divergence than other accessions [98]. These differences highlight the high degree of uncertainty associated with dating evolutionary events.

Other estimates of dating in the evolutionary history of eucalypts include phylogeographic analysis of chloroplast DNA (cpDNA) variation in series *Subulatae* across southern Australia (Nicolle unpublished) and in section *Maidenaria* in Tasmania [114]. Both studies implied relatively recent speciation within *Sympphyomyrtus*, similar to the estimates of Ladiges et al. [97]. *Subulatae* is a series in section *Bisectae* that is widely distributed across southern Australia and phylogeography revealed major geographic patterning in chloroplast haplotypes with times of divergence between various clades estimated between 1.35 and 0.49 Myr (Nicolle unpublished). The dating in the *Subulatae* study used generic divergence rates for the chloroplast genome, which are uncalibrated in eucalypts. However, in another study, dating of divergence between two lineages of *E. marginata* was consistent with the influence of a geomorphological barrier indicating reasonable calibration of molecular dating with this approach [181]. The geographic pattern between clades in series *Subulatae* was consistent with other known patterns of divergence in the flora identified using track analysis [71]. The *Subulatae* study suggested that the chloroplast variation in this group represents ancestral variation at the series level that has arisen across geographic regions and then been retained through more recent speciation and radiation at lower taxonomic levels. This implies that speciation has occurred relatively recently, within the last 500,000 and is consistent with the large increase in pollen records in fossil deposits spanning the last 200,000 years [160].

Chloroplast haplotype variation in Tasmanian species of section *Maidenaria* revealed three major lineages, i.e., a widespread central lineage and two lineages restricted to southern and eastern Tasmania [114]. The central lineage was present in species from both series *Viminales* and *Foveolatae* (previously *Ovatae*) in Tasmania and southern

Victoria, suggesting that this lineage predates speciation at the series level. The central and southern lineages were estimated to have diverged 3-0.8 Myr and their distribution suggests divergence through separation by alpine regions during glacial maxima. Both the southern and eastern lineages have centres of distribution correlated with proposed refugial areas based on phytogeography and climate reconstruction. Species that are endemic to the refugial areas carry only the southern cpDNA lineage, implying that they may have arisen from an isolated local gene pool containing this lineage following its divergence from the central lineage before the middle Pleistocene [117].

## SPECIES AND SPECIATION

Evidence from the fossil record, biogeography, phylogeography and phylogeny indicate increased radiation and speciation of eucalypts in the Pleistocene. Increased speciation maybe related to expansion and contraction of ranges due to cycles of mesic and arid conditions that increased in amplitude from the mid-Pleistocene [12]. Increased abundance of eucalypts also appears to be related to an increase in fire frequency, possibly associated with the arrival of humans in Australia [160]. Vicariance arising from climatic fluctuations would lead to allopatric speciation processes. Cladistic biogeographic studies of eucalypt species, particularly within subgenus *Eucalyptus* [89, 91, 92, 93, 94], show spatial separation of sister species, implying allopatric speciation. Similarly, replacement series of taxa across the landscape due to allopatric speciation occurs in western species of subgenus *Sympyomyrtus* [16, 17, 18] and is commonly observed in other flora in Western Australia in association with climatic gradients and complex edaphic patterns [70]. The yellow gum complex (*E. viminalis* J.D. Hook, *E. subcrenulata* Maiden and Blakely, and *E. johnstonii* Maiden) in Tasmania consists of three species that occur in different niches in alpine and subalpine habitat. Analysis of genetic diversity patterns supported allopatric speciation for the origin of *E. viminalis* rather than parapatric speciation from *E. subcrenulata* [111].

## Species Recognition

The absence of clear boundaries at lower taxonomic levels with the recognition of many subspecies is likely to be due to recent, incomplete and ongoing speciation. Identification of relationships among related

species based on morphology can be difficult. Systematic relationships based on molecular evidence are also particularly difficult to establish due to lack of resolution in the classically used gene sequences, e.g., ITS, *trnL-trnF* spacer. Diversity assessed by anonymous genetic markers (e.g., allozymes, RFLPs, and microsatellites) at the population level is generally high in eucalypts. Population based assessment of phylogenetic relationships at lower taxonomic levels can be useful in delineation of species relationships and taxonomic status [29, 163], i.e., at the boundary of reticulate relationships between individuals, and divergent relationships between species [2]. At this level, the benefits of both the biological and phylogenetic species concepts can be combined, thus enabling taxonomy to reflect evolutionary process [2].

Genetic relationships have been investigated in several species complexes in *Sympyomyrtus* where resolution of morphological variation is incomplete, leading to confusion among closely related taxonomic units. Taxa in the oil mallee *E. kochii* group (*E. kochii* Maiden & Blakely subsp. *kochii*, subsp. *plenissima* (Gardner) Brooker (= *E. plenissima*), subsp. *borealis* (Gardner) Nicolle (formerly erroneously known as *E. horistes* Johnson & Hill)) are defined mainly by slight differences in operculum shape. Phylogenetic analysis with nuclear Restriction Fragment Length Polymorphism (RFLP) loci revealed high genetic identity among the taxa [26] and led to a taxonomic reversal of all taxa to subspecific rank [129]. Another oil mallee species complex, the *E. angustissima* F. Muell. group (*E. angustissima*, *E. quaerenda* (Johnson and Hill) Byrne (= *E. angustissima* subsp. *quaerenda* Johnson and Hill), *E. misella* Johnson and Hill, *E. foliosa* Johnson and Hill). differ in the size of the leaves, buds and fruits, but they also occupy distinct niches in the landscape and occur in different habitats [62]. Phylogenetic analysis of the *E. angustissima* group showed differentiation between taxa and the two subspecies of *E. angustissima* were the most differentiated [47], leading to a taxonomic revision in which *E. quaerenda* (Johnson & Hill) Byrne was raised to species rank [32]. The phylogenetic relationships between taxa in the *E. angustissima* complex based on nuclear RFLP analysis are shown in Fig. 3. Speciation in this group has occurred in association with niche differentiation and differing eco-physiological preferences.

Habit is a morphological character that has been used extensively in the classification and taxonomy of eucalypt species [18]. Eucalypts have both tree and multi-stemmed “mallee” forms where several sparingly branched stems arise from a main underground stem structure, the

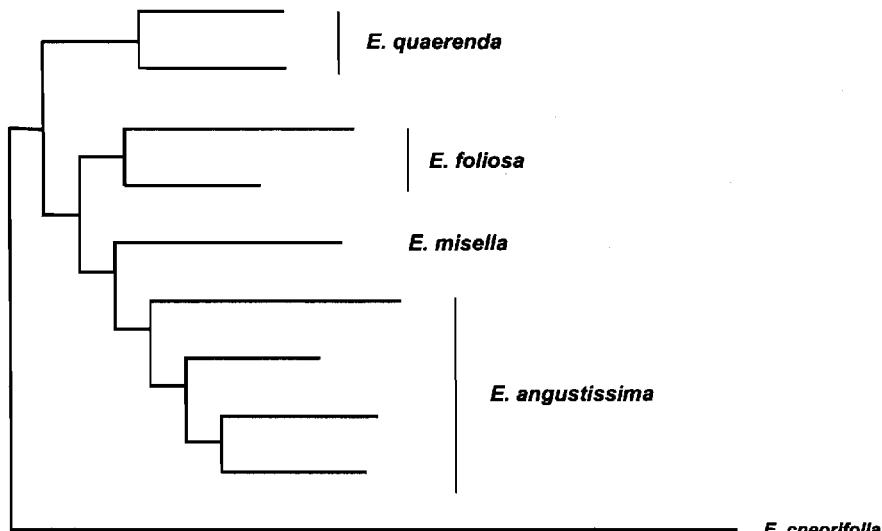
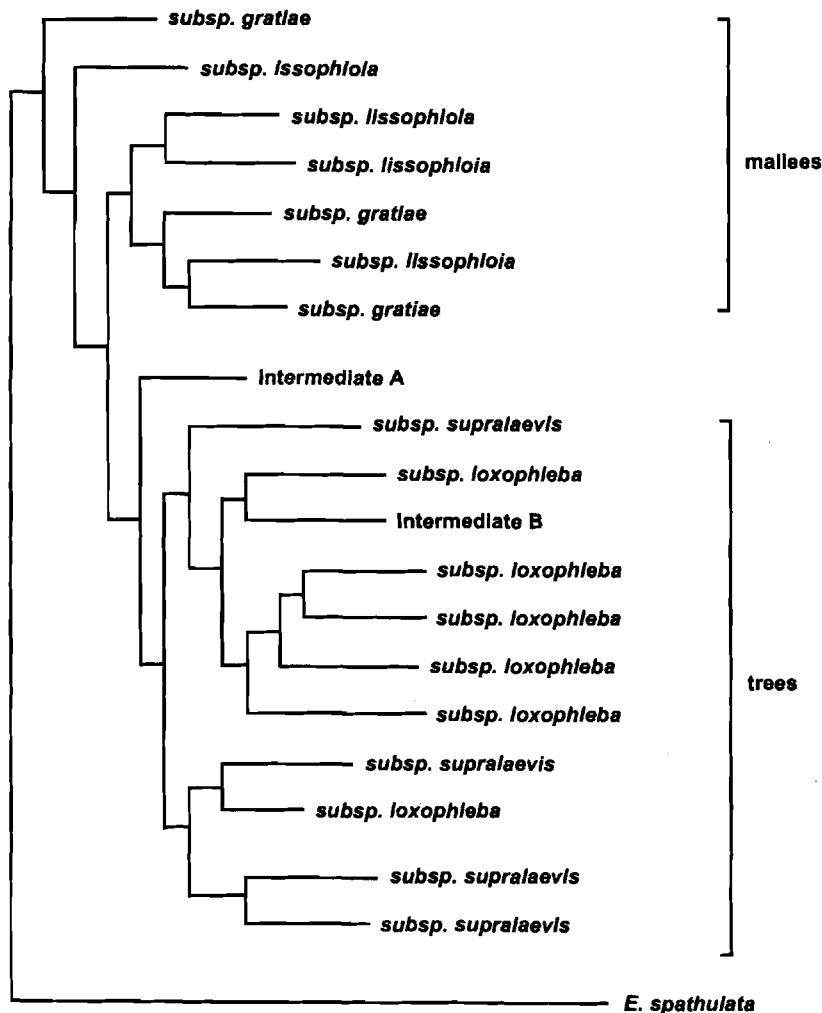


Fig. 3. Phylogenetic relationships between populations of the *Eucalyptus angustissima* complex with *E. cneorifolia* as the outgroup. Continuous character maximum likelihood analysis of nuclear variation using RFLP markers (Modified from Elliott and Byrne [47]).

lignotuber [130]. This habit occurs throughout the genus and is hypothesized to have evolved numerous times [61]. Mallees are dominant in the semi-arid regions and may be an adaptation to stress resulting from climatic conditions [108] or nutrient deficient soils [105]. The presence or absence of a lignotuber appears to be a heritable character [88, 130, 146, 184], although presence of a lignotuber does not necessarily mean that the plant will express the mallee habit. Habit in eucalypt species is generally consistent within taxa, although it can be modified in response to disturbance [86, 87].

Habit is one of the main characters used to define taxa in the *E. loxophleba* Benth. group; subspecies *loxophleba* and subsp. *supralaevis* Johnson and Hill are trees, subspecies *lissophloia* Johnson and Hill and subspecies *gratiae* Brooker are mallees [62]. Molecular analysis of the *E. loxophleba* complex identified some genetic differentiation between the mallee and tree forms, but little between subspecies with the same habit [66], as shown by the phylogenetic relationships between populations in Fig. 4. Two intermediate populations with features of more than one taxon were included in the study. One population had a mallee habit but a basal stocking of rough bark characteristic of subsp. *loxophleba* (tree



**Fig. 4.** Phylogenetic relationships between populations of the *Eucalyptus loxophleba* complex with *E. spathulata* as the outgroup. Continuous character maximum likelihood analysis of nuclear variation using RFLP markers. Populations are identified as subsp. *loxophleba*, *supralaevis*, *lissophloia* or *gratiae* based on morphological identification. Two populations with features of more than one taxon are represented as Intermediate A and Intermediate B (Modified from Hines and Byrne [66]).

habit, Intermediate A in Fig. 4). It was positioned between the mallee and tree groups, indicating that it may be a population in which the differentiation of morphological characters is incomplete. The second intermediate population was morphologically subspecies *loxophleba*, but

had the oil composition characteristics of subspecies *lissophloia* and occurs within the range of subspecies *lissophloia* (Intermediate B in Fig. 4). The clustering of this population with the tree forms is consistent with its morphological identification as subspecies *loxophleba*. Genetic differentiation based on habit reflects greater taxonomic differentiation than other morphological characters used in taxon identification in this complex. Evolution of the mallee habit in this species may be recent in evolutionary history as common ancestry is still evident in the overall genetic similarity between taxa [66].

## Hybridization

### Recent Hybridization

There is evidence of recent occurrence of hybridization in eucalypts but there is some debate over the evolutionary significance of hybridization in the genus. In a review of natural and manipulated hybridization in eucalypts, Griffin et al. [57] concluded that the former is relatively rare. Rate of hybridization varied among sections in subgenus *Sympyomyrtus* with the predominantly eastern sections, *Latoangulatae* (previously *Transversaria*), *Exsertaria* and *Maidenaria*, having higher rates (average 1.63%) than the western sections, *Bisectae* and *Dumaria* (0.23%). Thus, hybridization is rare at higher taxonomic levels and the higher rates reported among lower taxonomic levels are caused by taxonomic anomalies in past and current classification, misinterpretation of phenotypic characters, sampling bias, and a large proportion of single records [57].

The majority of hybrid records are localized and infrequent, and Griffin et al. [57] found 37% of hybrids to be reported from a single herbarium record. Molecular investigation due to the presence of a few intermediate trees in sympatric populations has confirmed the hybrid (usually  $F_1$ ) status of the intermediate trees. Hybridization between *E. acmenoides* Schauer and *E. cloeziana* in southeast Queensland was verified with genetic analysis using microsatellites and variation in cpDNA by Stokoe et al. [171]. Genetic evaluation of the critically endangered *E. bennettiae* DJ Carr and SGM Carr, which occurs as a single multi-stemmed mallee in each of two populations in southwest Western Australia, confirmed a hybrid origin between *E. lehmannii* (Schauer) Benth. and *E. sporadica* Brooker and Hopper (Byrne unpublished data). In both cases the parental populations were genetically differentiated and the hybrids were genetically intermediate between the two parents.

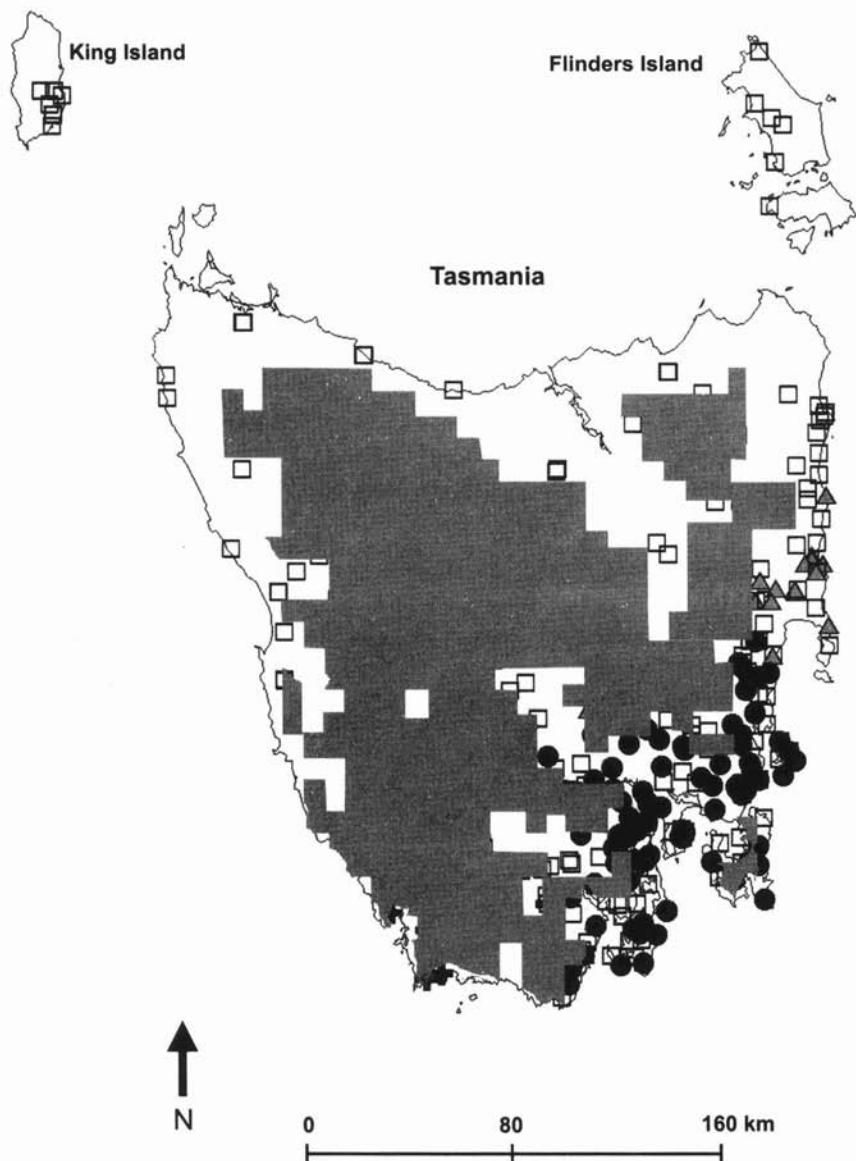
Hybridization is also recognized on a larger scale where species form clines and the intermediate morphology is assumed to be due to a large hybrid zone between two species at either end of the cline. However, close inspection throughout the distribution often fails to identify a clear distinction between the taxa and they intergrade across wide geographic zones. In investigation of a morphological cline between *E. populnea* F. Muell. and *E. brownii* Maiden and Cambage, which occur in Queensland (*E. brownii*) and northern New South Wales, both molecular and morphological analysis failed to identify major differentiation between the taxa or across the cline [68]. The authors concluded that the cline represented a single evolutionary lineage that has undergone directional diversification as a consequence of a selection gradient. Similar results were also obtained from a study of a second morphological cline between *E. melanophloia* F. Muell. and *E. whitei* Maiden and Blakely (Holman unpublished data). In the case of morphological clines, careful evaluation of the taxonomic delineation is warranted due to the taxonomic confusion that exists at species level in eucalypts. Investigation of a morphologically intermediate population of the yellow gum complex (*E. viminalis*, *E. subcrenulata* and *E. johnstonii*) on Mt Arrowsmith in Tasmania found no evidence for either hybridization or lack of primary differentiation between *E. viminalis* and *E. subcrenulata* [111]. The authors concluded that this case of clinal variation along an altitudinal gradient represented convergence of similar morphology between two species at their upper and lower altitudinal limits. Molecular evaluation of genetic diversity is essential where morphological variation within eucalypts confounds lower level taxonomy and species concepts based primarily on phenotypic characters.

### ***Historical Hybridization***

In contrast to Griffin et al. [57], studies of diversity in cpDNA in Tasmanian eucalypts have led to suggestions that hybridization has been a significant component of the evolution of eucalypts, with reticulate evolution of chloroplast lineages [73, 113, 114, 165]. This conclusion is based on wide scale sharing of haplotypes between species with a significant geographical pattern in the haplotype distributions. This level of chloroplast sharing has been found in species of both subgenus *Sympyomyrtus* and subgenus *Eucalyptus* in Tasmania [113, 114]. Sharing of chloroplast haplotypes between species can occur through convergent mutation, shared retention of ancestral polymorphism through

incomplete lineage sorting, and chloroplast capture through introgression. The latter two processes leave a similar pattern of incongruence between species [179]. Incomplete lineage sorting complicates the use of cpDNA in phylogenies as it can easily be misinterpreted as evidence of interspecific gene flow [156]. The distribution pattern of shared haplotypes can give some clues to the cause of the incongruence between morphological and molecular phylogenies [179]. In general, the pattern of chloroplast diversity in the Tasmanian eucalypts appears to be sharing of haplotypes within the same major clades except in areas of putative glacial refugia where sharing of identical haplotypes between species is common. [113, 114]. In the *Sympyomyrtus* species, widespread sharing of haplotypes from the central lineage in Victoria and northern Tasmania may reflect common ancestry. However, the sharing of haplotypes from the restricted southern and eastern lineages between many species that otherwise have haplotypes from the central lineage, is strongly indicative of introgressive hybridization [116]. The southern and eastern lineages are common in areas that have been recognized as putative glacial refugia and these areas were separated from more mesic northern areas by alpine habitat (Fig. 5). Contraction of species to refugia would provide opportunities for hybridization when reproductive barriers may have been weaker. Further detailed investigation of chloroplast diversity within and between *E. globulus* and the rare *E. cordata* Labill. suggested that *E. globulus* has acquired southern lineage haplotypes from *E. cordata* during contraction to glacial refugia [116]. In subgenus *Eucalyptus* species, sharing of identical haplotypes also occurs between species from different series in the refugial areas in southern Tasmania [113], suggesting introgressive hybridization during contraction to refugia.

In contrast to the case in Tasmania, studies of chloroplast cpDNA in eucalypts in Western Australia do not reveal extensive sharing of haplotypes, indicating that hybridisation may not be a major evolutionary factor [31, 33, 181]. Phylogeography of *E. loxophleba* [31] revealed patterns that were congruent with those in two other widely distributed species from two other families, i.e., Mimosoidaceae and Santalaceae [28, 30], and the dating of lineage divergence consistent with the influence of major climatic fluctuations from the mid Pleistocene (1-0.7 Myr). Congruence of phylogeography across plant families indicates that patterns of chloroplast variation in *E. loxophleba* have been influenced by major historical events not species specific factors. A large-scale study of cpDNA variation in the series *Subulatae* (subg. *Sympyomyrtus*) across



**Fig. 5.** Distribution of chloroplast haplotypes in Tasmanian species of *Eucalyptus* subgenus *Sympphyomyrtus* showing the restriction of the southern and eastern lineages to the southeast region where glacial refugia were located. •-southern lineage; □-central lineage; ▲-eastern lineage; + -intermediate lineage. Areas shaded grey represent modelled alpine vegetation at the height of the last glacial maximum [84] (Modified from McKinnon et al. [114]).

southern Australia revealed major geographic patterning in chloroplast haplotypes and little association with morphologically based taxonomic relationships within the series (Nicolle et al. unpublished). In 24 taxa sampled at single locations and in all except one case, the taxa did not share the same haplotypes. In addition, the geographic pattern of divergence between lineages was consistent with other known patterns of divergence in the flora identified using track analysis [71]. The *Subulatae* study suggested that the high level of chloroplast variation represents ancestral variation at higher taxonomic levels that has arisen across geographic regions and then been retained at lower taxonomic levels. This is consistent with the evolutionary history of eucalypts identified above where the major lineages have evolved through the Miocene and Pliocene with more rapid speciation and radiation in the last 200,000 years. Influences of glaciation and contraction to major refugia in Tasmania, and the subsequent expansion and recolonization of glaciated areas, may have led to hybridization and chloroplast capture in the Tasmanian eucalypts, similar to those hypothesized to have occurred in the European oaks [7, 136]. Such factors would not have contributed to major evolutionary processes in the main eucalypt distributions through southern Australia where glaciation was not experienced.

## **EVOLUTIONARY PROCESSES**

Evolutionary change at the level of species and genera is fundamentally driven by processes occurring at the gene, genome and population level [128]. Genetic diversity and selection drive the evolutionary processes of adaptation and speciation [128]. Mutation, migration and random drift influence the level and pattern of genetic diversity, and together with natural selection, contribute to the cumulative change in genetic composition that underpins evolution [60]. Population genetic studies in eucalypts reveal generally high levels of genetic diversity that are supported by sexual reproduction and mixed mating but secondarily outcrossed breeding systems, although clonality is common in mallee species.

### **Genetic Diversity**

Eucalypts show generally high levels of genetic diversity at the population level. Analysis of diversity parameters in 29 species assayed with allozymes [141] showed diversity levels similar to those of gymnosperms

[58], which are considered to have high genetic diversity compared to angiosperms [125]. There are some species where diversity is low, indicating specific factors involved in the evolutionary history of these species, e.g., small historically isolated populations in *E. caesia* Benth. [33, 121]; small disjunct populations in *E. pulverulenta* Sims [135]; very small populations in *E. johnsoniana* Brooker and Blaxell [122]. Most of the diversity in eucalypts is maintained within populations and genetic differentiation is generally very low. The average amount of diversity maintained between populations ( $G_{ST}$ ) for 29 species was estimated as 16.1%, although species with regionally disjunct distributions showed higher levels of differentiation across the regions [141]. In general, widespread species maintain greater levels of diversity than regional and localized species.

Analysis of genetic variation with more variable DNA markers (e.g., microsatellites, RFLPs) has also shown high levels of diversity within eucalypt species (Tables 2, and 3) similar to isozymes. In a few species genetic diversity has been estimated using multiple marker systems either as population studies or as specific comparisons of the same individuals (Table 4). Direct comparisons of diversity in the same individuals show an approximate twofold increase in diversity detected with microsatellites compared to RFLPs (allelic diversity – 90 to 118% increase; heterozygosity – 58 to 114% increase; Table 4). Comparison of population studies conducted in the same species (although not necessarily assessing the same populations) with different marker systems show approximately two-fold increase in diversity assayed with RFLPs compared to allozymes (allelic diversity – 0 to 115% increase; heterozygosity – 198 to 247% increase; Table 4) and two to four-fold increase in diversity assayed with microsatellites compared to allozymes (allelic diversity – 173 to 520% increase; heterozygosity – 302% increase; Table 4). The exception to this is *E. caesia* where diversity is low for RFLP loci as well as for allozyme loci, and this is believed to be due to inbreeding within small isolated populations that are confined to granite outcrops [33]. The same level of difference in diversity is also evident in comparison of mean diversity values for all studies using the various markers (Table 5). Therefore, this approximate two-fold difference in diversity between isozymes and RFLPs, and between RFLPs and microsatellites appears consistent across direct comparisons and general comparisons, and can be taken as a generalization across the genus.

**Table 2.** Genetic diversity parameters for eucalypt species assayed using nuclear RFLP analysis  
 Distribution as in Moran [123]: widespread – range of 600 km in at least one direction; regional – range between 150 and 600 km; localized – geographic area of less than 100 km<sup>2</sup>. A-number of alleles per locus; He-expected heterozygosity; H<sub>T</sub>-total heterozygosity; F<sub>ST</sub>-population differentiation.

Species	Distribution	A	He	H <sub>T</sub>	F <sub>ST</sub>	Reference
<i>E. camaldulensis</i>	Widespread	3.850	0.490	0.530	0.078 <sup>a</sup>	[21]
<i>E. kochii</i>	Widespread	5.900	0.489	0.514	0.045	[26]
<i>E. loxophleba</i>	Widespread	3.400	0.377	0.418	0.089	[66]
<i>E. niitens</i>	Widespread	3.400	0.373	0.445	0.162 <sup>b</sup>	[25]
<i>E. marginata</i>	Widespread	2.700	0.332	0.345	0.034 <sup>a</sup>	[180]
Mean	Widespread	3.850	0.412	0.450	0.082	
<i>E. occidentalis</i>	Regional	3.000	0.351	0.373	0.065 <sup>a</sup>	[46]
<i>E. angustissima</i>	Regional	2.500	0.330	0.371	0.136 <sup>a</sup>	[47]
<i>E. caesia</i>	Regional	1.310	0.068	0.250	0.527 <sup>a</sup>	[33]
<i>E. rhodantha</i>	Regional	1.950	0.590	0.280	0.099 <sup>a</sup>	Byrne unpublished
Mean	Regional	2.190	0.335	0.319	0.207	

<sup>a</sup>  $F_{ST} = \theta$   
<sup>b</sup> G<sub>ST</sub> value

**Table 3.** Genetic diversity parameters for eucalypt species assayed using microsatellites

Distribution as in Moran [123]: widespread – range of 600 km in at least one direction; regional – range between 150 and 600 km; localized – geographic area of less than 100 km<sup>2</sup>. Sampling indicates number of populations sampled and number of loci assayed. A-number of alleles per locus; He-expected heterozygosity;  $F_{ST}$ -population differentiation.

Species	Distribution	Sampling	A	He	$F_{ST}$	Reference
<i>E. globulus</i>	Widespread	34 populations /8 loci	9.500	0.750	0.090	[170]
<i>E. acmenoides</i>	Widespread	2 populations /6 loci	4.080			[171]
<i>E. cloeziana</i>	Widespread	2 populations /6 loci	4.170			[171]
<i>E. loxophleba</i>	Widespread	1 population /5 loci	10.200	0.741		Byrne unpublished
<i>E. brownii/E. populnea</i>	Widespread	10 populations /5 loci	16.100	0.880	0.021	[68]
Mean	Widespread		8.810	0.790	0.055	
<i>E. cladocalyx</i>	Regional	7 populations /8 loci	4.640	0.595	0.194	Byrne unpublished
<i>E. sporadica</i>	Regional	1 population /5 loci	6.600	0.763		Byrne unpublished
<i>E. lehmannii</i>	Regional	1 population /5 loci	5.600	0.745		Byrne unpublished
<i>E. curtisii</i>	Regional	12 populations /4 loci	4.960	0.540	0.300	[162]
<i>E. vernicosa/E. subcrenulata</i>	Regional	Species level (11 populations) /4 loci	22.500	0.860	0.120	[111]
<i>E. johnstonii</i>	Regional		5.450	0.661	0.247	
Mean <sup>a</sup>	Localised	4 populations /22 loci	6.660	0.696	0.105	[22]
<i>E. benthamii</i>	Localised	2 populations /6 loci	6.300	0.640	0.190	[81]
<i>E. morrisbyi</i>	Localized	14 populations /5 loci	6.000	0.545		Byrne unpublished
<i>E. absita</i>	Localized		6.320	0.627	0.147	
Mean	Localized					

<sup>a</sup> excluding species level values for *E. vernicosa/E. subcrenulata/E. johnstonii*

<sup>b</sup> all populations except one represent large clonal clumps

**Table 4.** Comparison of genetic diversity parameters for eucalypt species assayed using different markers

Sampling indicates number of populations sampled and number of loci assayed. A-number of alleles per locus;  $H_0$ -observed heterozygosity;  $H_e$ -expected heterozygosity;  $H_T$ -total gene diversity;  $F_{ST}$ -population differentiation.

Species	Marker	Sampling	A	$H_0$	$H_e$	$H_T$	$F_{ST}$	Reference
<i>E. nitens</i>	Isozymes	13 populations/17 loci	1.580	0.125				[38]
	RFLPs	8 populations/40 loci	3.400	0.342	0.373	0.445	0.162 <sup>b</sup>	[25]
	RFLPs	20 individuals <sup>a</sup> /40 loci	4.500	0.330				[24]
<i>E. caesia</i>	Microsatellites	20 individuals <sup>a</sup> /4 loci	9.800	0.575				[24]
	Isozymes	13 populations/8 loci	1.310	0.075	0.068	0.176	0.614 <sup>b</sup>	[121]
	RFLPs	14 populations/20 loci	1.190	0.075	0.072	0.250	0.527	[33]
<i>E. rhodantha</i>	Isozymes	6 populations/12 loci	1.980	0.100	0.170	0.265	0.101 <sup>b</sup>	[155]
	RFLPs	6 populations/19 loci	1.950	0.267	0.590	0.280	0.099	Byrne unpublished
	Isozymes	8 populations/20 loci	1.700	0.085	0.148	0.262		[110]
<i>E. cladocalyx</i>	Microsatellites	7 populations/8 loci	4.640	0.485	0.595	0.716	0.194	Byrne unpublished
	RFLPs	610 trees <sup>a</sup> /10 loci	8.400			0.404		[50]
	Microsatellites	610 trees <sup>a</sup> /2 loci	18.000		0.864			[50]
<i>E. consideriana</i>	RFLPs	4 populations <sup>a</sup> /18 loci	5.820		0.499			[52]
	Microsatellites	4 populations <sup>a</sup> /11 loci	11.070		0.790			[52]
	Microsatellites	34 populations/8 loci	9.500		0.750		0.090	[170]
<i>E. globulus</i>	Cinnamoyl CoA reductase gene	10 populations	4.900	0.640	0.800		0.088	[118]
	RAPDs <sup>c</sup>	31 populations/149 loci					0.1600	[127]

<sup>a</sup>direct comparison of each marker system in the same individuals in each species.

<sup>b</sup> $G_{ST}$  value

<sup>c</sup>Random Amplified Polymorphic DNA

**Table 5.** Mean genetic diversity parameters for eucalypt species assayed using allozyme, RFLP and microsatellite markers  
Distribution as in Moran [123]: widespread – range of 600 km in at least one direction; regional – range between 150 and  
600 km; localized – geographic area of less than 100 km<sup>2</sup>. A-number of alleles per locus;  $H_T$ -expected heterozygosity;  $H_e$ -expected heterozygosity;  $F_{ST}$ -population differentiation.

Species	Distribution	A	$H_e$	$H_T$	$F_{ST}$	Reference
Allozymes – mean 11 species	Widespread	2.330	0.240	0.270	0.137 <sup>a</sup>	[141]
RFLP – mean 5 species	Widespread	3.850	0.412	0.452	0.062	Table 2
Microsatellites – mean 5 species	Widespread	7.540	0.790		0.055	Table 3
Allozymes – mean 10 species	Regional	1.980	0.170	0.231	0.205 <sup>a</sup>	[141]
RFLP – mean 4 species	Regional	2.190	0.335	0.319	0.207	Table 2
Microsatellites – mean 4 species	Regional	5.450	0.661		0.205	Table 3
Allozymes – mean 8 species	Localized	1.870	0.182	0.207	0.140 <sup>a</sup>	[141]
Microsatellites – mean 3 species	Localized	6.320	0.627		0.145	Table 3
Allozymes – mean 29 species	All	2.070	0.198	0.238	0.161 <sup>a</sup>	[141]
RFLP – mean 9 species	All	3.110	0.380	0.390	0.130	Table 2
Microsatellites – mean 12 species	All	6.540	0.690		0.190	Table 3

<sup>a</sup>  $F_{ST}$  value

The greater level of diversity detected with microsatellites is useful for situations where fine-scale differentiation is required. Glaubitz et al. [52] found that microsatellites provided sufficient power to detect impacts of silvicultural treatments on diversity levels within a logging coupe, whereas the impact was not evident with RFLP loci. The high diversity detected with microsatellites provided high exclusion probabilities (0.997 to 0.99997 with 6 loci) for assignment of paternity in *E. grandis* and *E. wandoo* Blakely [34, 36, 85]. In eucalypts, microsatellites have been developed in species of *Sympphyomyrtus* section *Maidenaria*, *E. nitens* [24] and *E. globulus* [167, 174], and section *Latoangulatae*, *E. urophylla* and *E. grandis* [14, 15]; in subgenus *Eucalyptus*, *E. sieberi* L. Johnson [51]; and in *Corymbia variegata* (F. Muell.) Hill and Johnson [80, 159]. Use of these microsatellite loci in other species indicates reasonably high conservation of loci across sections and subgenera. However, null alleles can cause some problems when using heterologous microsatellite loci and should be taken into consideration when trialling microsatellites in other species. In population diversity studies, high frequency of null alleles may be detected through assessment of Hardy-Weinberg equilibrium. High frequency of a null allele has been observed in a heterologous microsatellite locus in *E. loxophleba* and was detected through segregation of progeny arrays.

## **Selection and Drift**

The high genetic variation in eucalypts that is detected with molecular markers is also expressed in quantitative traits [44, 141]. It is generally assumed that most molecular marker loci are selectively neutral, although exceptions to this have been observed [120], whereas quantitative variation may be subject to selection and local adaptation to diverse ecological challenges [102, 104, 150]. Therefore, comparison of genetic variation assayed with molecular markers and quantitative traits may provide insights into the relative importance of selection versus drift as evolutionary forces driving diversification [102, 112, 120, 150].

A comparison of quantitative variation in morphological traits and molecular variation in microsatellites in the forest tree *E. globulus* showed correlation at the regional level to be highly influenced by divergence of the mainland and island populations, but with little correlation within regions [170]. The authors concluded that neutral microsatellite loci were useful for identifying evolutionary affinities, and comparison with

morphological traits suggest that at least five traits that had previously been used to identify races in *E. globulus* were influenced by natural selection. In *E. marginata* there is evidence for local adaptation in Coastal Plain populations and Darling Plateau populations because the Coastal populations perform poorly in growth trials on the Darling plateau (Alcoa World Alumina Australia, unpublished data). This adaptation is correlated with divergent chloroplast lineages between the Coastal Plain and the Darling Plateau [181], suggesting that it has arisen in association with historical isolation, but is maintained in the face of gene flow since there was little divergence between Coastal Plain and Darling Plateau populations for nuclear molecular markers [180]. Latta [102] identified co-variance of allele frequencies across loci as a mechanism by which adaptive divergence of polygenic traits can be maintained despite the homogenizing effect of gene flow. The high population differentiation associated with regional disjunctions in distributions of *E. nitens* and *E. delegatensis* R.T. Baker corresponds to differences in growth and morphology between the regions [123] suggesting the effects of genetic drift rather than selection.

## BREEDING SYSTEM

### ***Mating System***

The high levels of diversity in eucalypts are generally maintained through a mixed mating system where outcrossing predominates but there is also a significant fraction of inbreeding. Although eucalypt flowers are protandrous, geitenogamous pollination is readily facilitated in most species by mass flowering of a large number of small flowers with asynchrony of flower development throughout the canopy [72]. Potts and Wiltshire [141] recorded mating system parameters for 18 species estimated with allozyme markers and these studies showed a mean population rate of outcrossing of 0.75, with a range from 0.51 to 0.96. Since then, mating system parameters have been estimated in five more species [37, 83, 110, 119, Coates unpublished] and the mean outcrossing rate over the 23 species is 0.74, indicating that this value is relatively representative of outcrossing in eucalypts in general.

Several mating system studies have now been carried out with more variable microsatellite markers. The high level of polymorphism detected with microsatellite loci theoretically provides greater power in detection of selfed versus outcrossed mating events. In addition, the genotype of the

maternal tree can be readily determined when using microsatellites, providing greater resolution of non-maternal alleles. Outcrossing rates were high in two species, 0.94 in *E. morrisbyi* Brett [81] and 0.79 in *E. wandoo* [34], but a relatively low outcrossing rate of 0.62 was found in *E. benthamii* Maiden and Cambage [22]. In only one species analysis with both allozyme and microsatellite markers has been carried out in the same populations. The average outcrossing rate in two populations of *E. wandoo* using allozymes was 0.68 (Coates unpublished), whereas with microsatellites, and identification of the maternal genotype, the outcrossing rate was 0.79 [34]. This indicates that mating system analysis of progeny arrays with allozymes may underestimate the true outcrossing rate by approximately 10%. While outcrossing rates are generally high across populations and species, most studies have identified high variability among mothers where seed crops range from completely selfed to completely outcrossed (e.g., *E. wandoo*, [34]). Outcrossing can also vary across the canopy within individual trees, and investigation in *E. globulus* showed greater outcrossing in the upper canopy than in the lower canopy [134], probably due to the foraging behaviour of bird pollinators [67].

### ***Post-zygotic Incompatibility***

The maintenance of an outcrossed breeding system in eucalypts whilst capable of self pollination, is facilitated by post-zygotic incompatibility mechanisms. This post-zygotic incompatibility is generally expressed as seed abortion, and seed set following self-pollination is severely reduced compared with outcrossing. Controlled crosses showed that the average reduction in seed set across 11 species was 64% [141]. Pre-zygotic self-incompatibility has been investigated but rarely observed, although it was documented as reduced pollen tube penetration in the style and ovules in *E. spathulata* Hook. and *E. platypus* Hook. [158] and *E. woodwardii* Maiden [157]. Most investigations into impacts of selfing in eucalypts have shown late acting post-zygotic effects as reduced numbers of viable seed per capsule induced by seed abortion [45, 59, 142, 144, 145]. James [76] considered seed abortion in eucalypts to be a highly evolved component of the genetic system. From his investigations he concluded that the capsule provides a mechanism for efficient selection of the fittest (outcrossed) seed [74] through the effects of intense competition for space in the tightly packed capsules [75]. Griffin et al. [56] also

concluded that competition occurred between embryo genotypes and selection appeared to be dependent on embryo genotype and maternal resource allocation. Similarly, Ellis and Sedgley [45] considered that maternal selection of embryo genotypes is a major determinant of seed quality. Late acting self-incompatibility is believed to be due to the expression of lethal recessives during seed development [56, 74]. Thus, the genetic load associated with inbreeding in eucalypts is reduced through competition and expression of lethal mutations of minor effect, rather than coalescence of chromosome numbers generating lethal mutations of major effect, as is apparent in other Australian genera such as *Isotoma* (R.Br.) Lindl., *Laxmanni* R.Br., *Stylium* Sw. ex Willd. and *Drosera* L. [75, 76]. The reduction of genetic load enables maintenance of high levels of diversity in lineages, thus allowing evolutionary progress, adaptation and diversification in the face of increasing aridity through the evolutionary history of eucalypts.

### **Gene Flow**

Migration of genes among populations maintains genetic connectivity within species, thus limiting the genetic divergence that ultimately leads to speciation. The inverse relationship between gene flow and population divergence means that indirect estimates of historical gene flow can be derived from genetic differentiation in neutral markers under the assumptions of an island model of migration [185]. For most eucalypt species, except those with significant range disjunctions, genetic differentiation is low to moderate ( $G_{ST}$  or  $F_{ST}$  values of 0.034 to 0.18) and gives estimates of gene flow ( $Nm$ ) in the range of 1.1 to 7.1. Gene flow values greater than 1.0 are considered high and sufficient to prevent genetic differentiation over time. In eucalypts, gene flow by seed dispersal is quite limited as the seed has no adaptations for dispersal and a sizeable quantity of the seed is deposited within twice the height of the canopy [9, 141]. Gene flow in eucalypts is more of a function of pollen dispersal that is generally mediated by non-specific biotic vectors such as insects and birds [72].

Comparison of genetic differentiation estimates made with nuclear markers versus chloroplast markers gives some indication of the relative influence of seed and pollen mediated gene flow [49] since the nuclear genome is transmitted via both the pollen and seed, whereas the chloroplast genome is transmitted only via the seed. The ratio of gene

flow by pollen to seed was high (7.2) in *E. nitens* [25] indicating that pollen dispersal is seven times as strong as seed dispersal, but low in *E. grandis* (1.2, [79]) suggesting that pollen dispersal is similar to seed dispersal. Calculation of pollen to seed flow ratios for other eucalypts where both nuclear and chloroplast studies have been conducted show values of 32 to 581 (Table 6), similar to those of eight other tree species that ranged from 18 to 196 [49], and the mean value of 17 that was obtained from 93 plant species [137]. The lower pollen to seed flow ratio identified in *E. nitens* is likely to be influenced by significant disjunction in the species range compared to *E. kochii*, *E. loxophleba* and *E. marginata* that are all widespread species with no major geographic disjunctions. The generally high pollen to seed dispersal ratios indicate much greater contribution of pollen dispersal to gene flow and are consistent with observations of limited seed dispersal in eucalypts. The low pollen to seed flow ratio in *E. grandis* is anomalous and warrants further investigation to determine whether it reflects greater seed dispersal distances or more restricted pollen dispersal than in other eucalypt species.

Pollen dispersal is idiosyncratic and greatly influenced by landscape context and ecological variables [164], particularly factors that influence the behaviour of pollen vectors, such as the spatial structure of populations [48]. Most evidence for pollen dispersal in natural populations has come from morphological detection of hybrids in seed crops [143]. A more comprehensive study of the pollen dispersal curve from a planted *E. nitens* population into a natural *E. ovata* Labill. population showed an average of 7% hybrid seed within 100 m of the *E. nitens* plantation boundary, dropping to 0.7% by 200-300 m, although occasional hybrids were still detected at 1.6 km from the plantation boundary [3]. These data have been interpreted as indicating low levels of pollen dispersal in eucalypts and a highly leptokurtic distribution where most dispersal occurs at short distances with a long tail of low level pollen dispersal over larger distances [103].

However, a recent study of direct estimates of pollen dispersal in *E. wandoo* using microsatellites and paternity assignment [34] has shown much higher levels of pollen dispersal over large distances than the previous evidence of limited dispersal. In *E. wandoo*, 30% to 65% of pollination originated from outside the study populations, and came from pollen sources a minimum of 0.6-1.08 km away. Pollen immigration of 46% was detected in a planted population of *E. grandis* that was not isolated from surrounding natural stands [79]. Estimates of pollen

**Table 6.** Estimates of the ratio of pollen to seed flow for eucalypt species

$F_{ST}$  (nuc)-population differentiation based on nuclear marker,  $F_{ST}$  (cp)-population differentiation based on chloroplast marker,  $Fis$ -inbreeding coefficient. Pollen flow/seed flow ratio =  $[(1/F_{ST}(\text{nuc}) - 1)(1 + Fis) \cdot 2(1/F_{ST}(\text{cp}) - 1)]/(1/F_{ST}(\text{cp}) - 1)$  according to equation 5a of Ennos [49]. In all species, except *E. grandis*, the nuclear and chloroplast diversity studies were conducted with the same individuals.

Species	$F_{ST}$ (nuc)	$F_{ST}$ (cp)	$Fis$	Pollen flow/seed flow	Reference
<i>E. nitens</i>	0.162	0.620	0.085	7.2	[25]
<i>E. kochii</i>	0.045	0.612	0.080	34.2	[26, 27]
<i>E. loxophleba</i>	0.089	0.930	0.046	140.3	[31, 66]
<i>E. marginata</i>	0.031	0.944	0.107	581.4	[180, 181]
<i>E. angustissima complex</i>	0.191 <sup>a</sup>	0.980 <sup>a</sup>	0.031	212.0	[47]
<i>E. brownii/E. populnea</i>	0.020	0.350 <sup>b</sup>	0.300	32.3	[68]
<i>E. grandis</i>	0.120	0.300	0.000 <sup>c</sup>	1.2	[78]

<sup>a</sup>values calculated from raw data not presented in reference

<sup>b</sup>value taken from supplementary data in Petit et al. [137]

<sup>c</sup>assumed to be zero as no value given in reference

immigration into exotic seed orchards in Brazil and Malagasy using allozyme or microsatellite markers also showed a high level of pollen dispersal. In *E. grandis/E. urophylla* orchards in Brazil, 14% pollen immigration was detected over 400 m [35] and 2.8% over 800 m [82], and in a *E. grandis* orchard in Malagasy 40% pollen immigration was observed over 100 m [36]. The few molecular investigations to date suggest that pollen dispersal is much more extensive than previously suspected. Long distance pollen dispersal identified using molecular markers is consistent with the high levels of historical gene flow estimated from genetic differentiation values. In *E. wandoo*, comparisons of gene flow using direct and indirect estimates were similar [34] and indicate that gene flow has been maintained despite extensive anthropogenic fragmentation due land clearing for agricultural production.

### Clonality

Clonality in mallee eucalypts is relatively common as they resprout from the lignotuber following disturbance. Clonality often occurs as small clusters of stems that are clearly part of the same clump, e.g., *E. obtusiflora*

DC. [83] and *E. caesia* [33] but can also occur as separate stems across large areas. In these cases molecular markers are a practical way of determining clone size and population structure. Genetic analysis in several species has determined very large mallee clones spanning several hundred square metres, e.g., 900 m<sup>2</sup> in *E. phylacis* Johnson and Hill [151], and 300 m<sup>2</sup> in *E. argutifolia* Grayling and Brooker [83]. However, even when clonal clumps are relatively small, some populations may contain only one clone, eg., *E. caesia* [33]. In *E. absita* Grayling and Brooker, the 14 "populations" present in a very localized area (approximately 30 km<sup>2</sup>) were shown to be large multiple stemmed single clones, except for one population where two genetically different individuals were present in one clump (Byrne unpublished). Clonality may exacerbate inbreeding in small populations, and some species with very large clones have become infertile and produce no viable seed, e.g., *E. phylacis* [151]. In *E. caesia*, where clones are relatively small, the populations exhibit very low genetic diversity but maintain adequate seed production suggesting that the genetic load may have been purged [33], possibly through repeated inbreeding or through selective sweeps generating highly adapted genotypes and reducing diversity [4, 99]. Lignotubers are usually very long-lived and age estimates for some lignotubers range from 330 to 6,380 years [130]. Clonality results in very long generation times and thus decreases the speed of evolutionary processes.

## CONCLUSION

Molecular genetic analysis has provided significant insights into the phylogeny and diversity of eucalypts and the evolutionary processes that have led to the variability evident in this large genus that is a dominant component of Australian ecosystems and of major importance to international forestry. Further development of genomics in eucalypts will enable assessment of the level of variation in functional genes leading to greater understanding of adaptation and ecological determinants of genetic variation. Variation in functional gene regions underpins the characteristics that enable a species to occupy a certain geographical range or ecological niche [178]. Marker systems targeting genes that exhibit ecologically relevant variation are being developed where sequence information is available [178]. The study of variation in a lignin biosynthesis gene (cinnamoyl CoA reductase) by McKinnon et al. [118] represents the first use of single copy nuclear gene sequence to investigate variation in natural populations in eucalypts. The presence of intragenic

recombination is a factor that can complicate use of gene sequence variation and has been identified in the cinnamoyl CoA reductase gene in eucalypts [139].

Studies of gene expression through microarray technology at the population and species level will provide insights into the relative influences of drift and selection, and identify genes involved in adaptive differences and speciation [182]. Genome wide scans of DNA polymorphism should provide a means of detecting the signature of positive directional selection leading to different adaptive peaks among populations [172].

There is enormous potential for such genomic approaches in biodiversity studies of eucalypts due to the great interest in eucalypt genomics for breeding and improvement of commercial species that will lead to development of these technologies. Sequencing of the eucalypt genome, as proposed by the Eucalypt Genome Initiative (<http://www.ieugc.up.ac.za/>), will provide the genomic resources (e.g., sequence data, and EST libraries) that will allow significant progress in gene-targeting, microarrays and genome scans. Understanding variation in functional gene regions may provide indepth knowledge of ecological processes leading to greater options in ameliorating the negative environmental impacts currently threatening many species.

## Acknowledgment

The author thanks Michael Bayly, David Coates, Gay Mckinnon, Dean Nicolle, Dorothy Steane and Alcoa World Alumina for access to unpublished data, and Michael Bayly, Joel Ochieng and Carlos Parra-O for providing manuscripts prior to publication. The author is also indebted to Pauline Ladiges, Dean Nicolle, Maurice McDonald, Brad Potts, Kevin Thiele and Dorothy Steane for valuable comments on the chapter.

## References

- [1] Ambrose GJ, Callen RA, Flint RB, Lange RT. *Eucalyptus* fruits in stratigraphic context in Australia. *Nature* 1979; 280: 387-389.
- [2] Avise JC, Wollenberg K. Phylogenetics and the origin of species. *Proc Natl Acad Sci USA* 1997; 94: 7748-7755.
- [3] Barbour RC, Potts BM, Vaillancourt RE. Pollen dispersal from exotic eucalypt plantations. *Cons Genet* 2005; 6: 253-257.

- [4] Barrett SCH, Charlesworth D. Effect of a change in the level of inbreeding on the genetic load. *Nature* 1991; 352: 522-524.
- [5] Bayly MJ, Ladiges PY. Divergent paralogues of ribosomal DNA in eucalypts (Myrtaceae). *Mol Phylogen Evol*; 2007; 44: 346-356.
- [6] Bayly MJ, Udovicic F, Gibbs AK, Parra-O C, Ladiges PY. Ribosomal DNA pseudogenes are widespread in the eucalypt group (Myrtaceae): implications for phylogenetic analysis. *Cladistics* 2007; in press.
- [7] Belahbib N, Pemonge M-H, Ouassou A, Sbay H, Kremer A, Petit RJ. Frequent chloroplast exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Mol Ecol* 2001; 10: 2003-2012.
- [8] Blazey J. A late Miocene flora from Bacchus Marsh. Honours Thesis 1994; School of Botany, The University of Melbourne.
- [9] Boland DJ, Brooker MIH, Turnbull JW, eds. *Eucalyptus seed*. Canberra: CSIRO, 1980.
- [10] Boland DJ, Brooker MIH, Chippindale GM, Hall N, Hyland BPM, Johnston RD, Kleinig DA, Turner JD, eds. *Forest trees of Australia*. Melbourne: CSIRO, 1984.
- [11] Bogunic F, Muratovic E, Brown SC, Siljak-Yakovlev S. Genome size and base composition of five *Pinus* species from the Balkan region. *Plant Cell Rep* 2003; 22: 59-63.
- [12] Bowler JM. Aridity in the late Tertiary and Quaternary of Australia. In: Barker WR, Greenslade PJM, eds. *Evolution of the flora and fauna of arid Australia*. Frewville: Peacock Publications, 1982: 35-45.
- [13] Briggs BG, Johnson LAS. Evolution in the Myrtaceae - evidence from inflorescence structure. *Proc Linn Soc NSW* 1979; 102: 157-256.
- [14] Brondani RPV, Brondani C, Grattapaglia D. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Mol Genet Genom* 2002; 267: 338-347.
- [15] Brondani R, Brondani C, Tarchini R, Grattapaglia D. Development, characterisation and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor Appl Genet* 1998; 97: 816-827.
- [16] Brooker MIH. A revision of the informal series *Foecundae* Pryor and Johnson of the genus *Eucalyptus* L'Herit and notes on variation in the genus. *Brunonia* 1979; 2: 125-170.
- [17] Brooker MIH. *Eucalyptus foecunda* revisited and six related new species (Myrtaceae). *Nuytsia* 1988; 6: 325-334.
- [18] Brooker MIH, Hopper SD. A taxonomic revision of *Eucalyptus wandoo*, *E. redunca* and allied species (*Eucalyptus* series *Levispermae* Maiden – Myrtaceae) in Western Australia. *Nuytsia* 1991; 8: 1-189.
- [19] Brooker MIH. A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Aust Syst Bot* 2000; 13: 79-148.
- [20] Burbidge N. The phytogeography of Australia. *Aust J Bot* 1991; 8: 75-211.
- [21] Butcher PA, Otero A, McDonald MW, Moran GF. Nuclear RFLP variation in *Eucalyptus camaldulensis* Dehnh. from northern Australia. *Heredity* 2002; 88: 402-412.
- [22] Butcher PA, Skinner AK, Gardiner CA. Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii*. *Cons Genet* 2005; 6: 213-226.

- [23] Byrne M, Moran GF, Tibbets WN. Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. J Heredity 1993; 84: 218-220.
- [24] Byrne M, Marquez-Garcia MI, Uren T, Smith DS, Moran GF. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. Aust J Bot 1996; 44: 331-341.
- [25] Byrne M, Parrish TL, Moran GF. Nuclear RFLP diversity in *Eucalyptus nitens*. Heredity 1998; 81: 225-232.
- [26] Byrne M. High genetic identities between three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii*, ssp. *plenissima* and *E. horistes* based on nuclear RFLP analysis. Heredity 1999; 82: 205-211.
- [27] Byrne M, Macdonald B. Phylogeography and conservation of three oil mallee taxa, *Eucalyptus kochii* ssp. *kochii*, ssp. *plenissima* and *E. horistes*. Aust J Bot 2000; 48: 305-312.
- [28] Byrne M, Macdonald B, Coates DJ. Phylogeographical patterns in chloroplast DNA variation within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. J Evol Biol 2002; 15: 576-587.
- [29] Byrne M. Phylogenetics and the conservation of a diverse and ancient flora. Comptes Rendus Biologies 2003; 326: S73-S79.
- [30] Byrne M, Macdonald B, Brand J. Phylogeography and divergence in the chloroplast genome of Western Australian Sandalwood (*Santalum spicatum*). Heredity 2003a; 91: 389-395.
- [31] Byrne M, Hines B. Phylogeographic analysis of cpDNA variation in *Eucalyptus loxophleba* (Myrtaceae). Aust J Bot 2004; 52: 459-470.
- [32] Byrne M. Recognition of *Eucalyptus quaerenda* (Myrtaceae) at specific rank. Nuytsia 2004a; 15: 321-323.
- [33] Byrne M, Hopper SH. Granite outcrops as ancient islands in old landscapes: evidence from the phylogeography and population genetics of *Eucalyptus caesia* in Western Australia. Biol J Linn Soc 2007; in press.
- [34] Byrne M, Elliot C, Yates CJ, Coates DJ. Extensive pollen dispersal in *Eucalyptus wandoo*, a dominant tree of the fragmented agricultural region in Western Australia. Cons Genet 2007; in press DOI 10.1007/s10592-007-9311-5.
- [35] Campinhos EN, Peters-Robinson I, Bertolucci FL, Alfenas AC. Interspecific hybridisation and inbreeding effect in seed from a *Eucalyptus grandis* x *E. urophylla* clonal seed orchard in Brazil. Genet Mol Biol 1998; 21: 369-374.
- [36] Chaix G, Gerber S, Razafimaharo V, Vigneron P, Verhaegen D, Hamon S. Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. Theor Appl Genet 2003; 107: 705-712.
- [37] Coates DJ, Keighery GJ, Broadhurst L. Genetic and morphological variation, and the mating system in *Eucalyptus gomphocephala* DC (Tuart). In: Keighery BJ, Longman VM, eds. Tuart and Tuart communities. Nedlands: Perth Branch Wildflower Society of Western Australia, 2002: 89-107.
- [38] Cook IO, Ladiges PY. Isozyme variation in *Eucalyptus nitens* and *E. denticulata*. Aust J Bot 1998; 46: 35-44.
- [39] Crisp M, Cook L, Steane D. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities. Philos Trans R Soc Lond B 2004; 359: 1551-1571.

- [40] Crisp M, Cook L, Steane D. Molecular dating and eucalypts: reply to Ladiges and Udovicic. *Aust Syst Bot* 2005; 18: 295-296.
- [41] Crisp MD, Linder HP, Weston PH. Cladistic biogeography of plants in Australia and New Guinea: congruent pattern reveals two endemic tropical tracks. *Syst Biol* 1995; 4: 457-473.
- [42] Croizat L. Space, time and form: the biological synthesis. Leon Croizat. Caracas, 1962.
- [43] Deane H. Notes on the fossil record of Berwick. Records of the Geological Survey of Victoria 1902; 1: 21-32.
- [44] Dutkowski GW, Potts BM. Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. *Aust J Bot* 1999, 7: 237-263.
- [45] Ellis MF, Sedgley M. Floral morphology and breeding system of three species of *Eucalyptus*, section *Bisectaria* (Myrtaceae). *Aust J Bot* 1992; 40: 249-262.
- [46] Elliott CP, Byrne M. Genetic diversity within and between natural populations of *Eucalyptus occidentalis* (Myrtaceae). *Silv Genet* 2003; 52: 169-173.
- [47] Elliott CP, Byrne M. Phylogenetics and the conservation of rare taxa in the *Eucalyptus angustissima* complex in Western Australia. *Cons Genet* 2004; 5: 39-47.
- [48] Ellstrand NC. Gene flow by pollen: implications for plant conservation genetics. *Oikos* 1992; 63: 77-86.
- [49] Ennos RA. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 1994; 72: 250-259.
- [50] Glaubitz JC, Strk J, Moran GF. Genetic impacts of different silvicultural practices in native eucalypt forests. In: Matyas C, ed. *Forest genetics and Sustainability*. Dordrecht: Kluwer Academic Publishers, 1999: 183-195.
- [51] Glaubitz JC, Emebiri LC, Moran GF. Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. *Genome* 2001; 44: 1041-1045.
- [52] Glaubitz JC, Murrell JC, Moran GF. Effects of native forest regeneration practices on genetic diversity in *Eucalyptus consideniana*. *Theor Appl Genet* 2003; 107: 422-431.
- [53] Goff SA, Ricke D, Lan TH, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 2002; 296: 92-100.
- [54] Grattapaglia D, Bradshaw HD. Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Can J For Res* 1994; 24: 1074-1078.
- [55] Greenwood DR. Middle Eocene megafloras from central Australia: earliest evidence for Australian sclerophyllous vegetation. *Am J Bot* 1991; 78: 114-115.
- [56] Griffin AR, Moran GF, Fripp YJ. Preferential outcrossing in *Eucalyptus regnans* F Muell. *Aust J Bot* 1987; 35: 465-475.
- [57] Griffin AR, Burgess IP, Wolf L. Patterns of natural and manipulated hybridisation in the genus *Eucalyptus* L'Herit – a review. *Aust J Bot* 1988; 36: 41-66.
- [58] Hamrick JL, Godt MJW. Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS, eds. *Plant Population Genetics, Breeding and Genetic Resources*. Sunderland, Massachusetts: Sinauer Associates, 1990: 43-63.
- [59] Hardner CM, Potts BM. Inbreeding depression and changes in variation after selfing in *Eucalyptus globulus* ssp. *globulus*. *Silv Genet* 1995; 44: 46-54.
- [60] Hartl DL, ed. *A primer of population genetics*. Sunderland, Massachusetts: Sinauer Associates, 2000.

- [61] Hill KD. Mallee Eucalypt communities: their classification and biogeography. In: Noble JC, Bradstock RA, eds. Mediterranean Landscapes in Australia, Mallee Ecosystems and their Management. Melbourne: CSIRO, 1989: 93-108.
- [62] Hill KD, Johnson LAS. Systematic studies in the eucalypts. 5. New taxa and combinations in *Eucalyptus* (Myrtaceae) in Western Australia. *Telopea* 1992; 4: 561-634.
- [63] Hill KD, Johnson LAS. Systematic studies in the eucalypts. 7. A revision of the bloodwoods, genus *Corymbia* (Myrtaceae). *Telopea* 1995; 6: 184-504.
- [64] Hill KD, Johnson LAS. Systematic studies in the eucalypts. 8. A review of the eudesmoid eucalypts, *Eucalyptus* subgenus *Eudesmia*. *Telopea* 1998; 7: 375-414.
- [65] Hill RS. The history of selected Australian taxa. In: Hill RS, ed. History of the Australian vegetation: Cretaceous to Recent. Cambridge: Cambridge University Press, 1994: 390-419.
- [66] Hines B, Byrne M. Genetic differentiation between mallee and tree forms in the *Eucalyptus loxophleba* complex. *Heredity* 2001; 87: 566-572.
- [67] Hingston AB, Potts BM. Pollinator activity can explain variation in outcrossing rates within individual trees. *Aust Ecol* 2005; 30: 319-324.
- [68] Holman JE, Hughes JM, Fensham RJ. A morphological cline in *Eucalyptus*: a genetic perspective. *Mol Ecol* 2003; 12: 3013-3025.
- [69] Holmes WBK, Holmes FM, Martin HA. Fossil *Eucalyptus* remains from the Middle Miocene Chalk Mountain Formation, Warrambungle Mountains, New South Wales. *Proc Linn Soc NSW* 1982; 106: 299-310.
- [70] Hopper SD. Patterns of plant diversity at the population and species level in south-west Australian Mediterranean ecosystems. In: Hobbs RJ, ed. Biodiversity of Mediterranean Ecosystems in Australia. Chipping Norton, NSW: Surrey Beatty and Sons, 1992: 27-46.
- [71] Hopper SD, Gioia P. The southwest Australian floristic region: Evolution and conservation of a global hotspot of biodiversity. *Ann Rev Ecol Syst* 2004; 35: 623-650.
- [72] House SM. Reproductive biology of eucalypts. In: Williams J, Woinarski J, eds. *Eucalypt Ecology. Individuals to ecosystems*. Cambridge: Cambridge University Press, 1997: 30-55.
- [73] Jackson HD, Steane DA, Potts BM, Vaillancourt RE. Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). *Mol Ecol* 1999; 8: 739-751.
- [74] James SH, Kennington WJ. Selection against homozygotes and resource allocation in the mating system of *Eucalyptus camaldulensis* Dehnh. *Aust J Bot* 1993; 41: 381-391.
- [75] James SH. Seed abortion and the evolution of genetic systems in some Australian native plant groups. In: Hopper SD, Chappill JA, Harvey MS, George AS, eds. *Gondwanan Heritage. Past present and future of the Western Australian Biota*. Sydney: Surrey Beatty and Sons, 1996: 90-99.
- [76] James SH. Genetics systems in the south-west flora: implications for conservation strategies for Australian plant species. *Aust J Bot* 2000; 48: 341-347.
- [77] Johnson LAS. Problems of species and genera in *Eucalyptus* (Myrtaceae). *Plant Syst Evol* 1976; 125: 155-167.
- [78] Jones M, Shepherd M, Henry R, Delves A. Chloroplast DNA variation and population structure in the widespread forest tree, *Eucalyptus grandis*. *Cons Genet*, 2006; 7: 691-703.

- [79] Jones M, Shepherd M, Henry R, Delves A. Pollen flow in *Eucalyptus grandis* determined by paternity analysis using microsatellite markers. *Tree Genet Genom* 2007; in press DOI 10.1007/s11295-007-0086-0.
- [80] Jones ME, Stokoe RL, Cross MJ, Scott LJ, Maguire TL, Shepherd M. Isolation of microsatellite loci from spotted gum (*Corymbia variegata*), and cross species amplification in *Corymbia* and *Eucalyptus*. *Mol Ecol Notes* 2001; 1: 276-278.
- [81] Jones RC, McKinnon GE, Potts BM, Vaillancourt RE. Genetic diversity and mating system of an endangered tree *Eucalyptus morrisbyi*. *Aust J Bot* 2005; 3: 367-377.
- [82] Jungans TG, Peters-Robinson I, Bertolucci FL, Alfenas AC. The use of self-incompatibility in the production of hybrid eucalyptus seed by 'Aracruz Celulose' in Brazil. *Genet Mol Biol* 1998; 21: 375-379.
- [83] Kennington WJ, James SH. Contrasting patterns of clonality in two closely related mallee species from Western Australia, *Eucalyptus argutifolia* and *E. obtusiflora* (Myrtaceae). *Aust J Bot* 1997; 45: 679-689.
- [84] Kirkpatrick JB, Fowler M. Locating likely glacial forest refugia in Tasmania using palynological and ecological information to test alternative climatic models. *Biol Cons* 1998; 85: 171-182.
- [85] Kirst M, Cordeiro CM, Rezende GDSP, Grattapaglia D. Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. *J Hered* 2005; 96: 161-166.
- [86] Lacey CJ. Development of large plate-like lignotubers in *Eucalyptus botryoides* Sm. in relation to environmental factors. *Aust J Bot* 1983; 31: 105-118.
- [87] Lacey CJ, Johnston RD. Woody clumps and clumpwoods. *Aust J Bot* 1990; 38: 299-334.
- [88] Ladiges PY, Ashton DH. Variation in central Victorian populations of *Eucalyptus viminalis* Labill. *Aust J Bot* 1974; 22: 81-102.
- [89] Ladiges PY, Humphries CJ, Brooker MIH. Cladistic relationships and biogeographic patterns in the peppermint group of *Eucalyptus* (informal subseries *Amygdalininae*, subgenus *Monocalyptus*) and the description of a new species, *E. willisii*. *Aust J Bot* 1983; 31: 565-584.
- [90] Ladiges PY. A comparative study of trichomes in *Angophora* Cav. and *Eucalyptus* L'Her – a question of homology. *Aust J Bot* 1984; 32: 561-574.
- [91] Ladiges PY, Humphries CJ. Relationships in the Stringybarks, *Eucalyptus* L'Herit. informal subgenus *Monocalyptus* Series *Capitellatae* and *Olseniana*: Phylogenetic hypotheses, biogeography and classification. *Aust J Bot* 1986; 34: 603-632.
- [92] Ladiges PY, Humphries CJ, Brooker MIH. Cladistic and biogeographic analysis of Western Australian species of *Eucalyptus* L'Herit., informal subgenus *Monocalyptus* Pryor & Johnson. *Aust J Bot* 1987; 35: 251-281.
- [93] Ladiges PY, Newnham MR, Humphries CJ. Systematics and biogeography of the Australian 'green ash' eucalyptus (*Monocalyptus*). *Cladistics* 1989; 5: 345-364.
- [94] Ladiges PY, Prober S, Nelson G. Cladistic and biogeographic analysis of the 'blue ash' eucalypts. *Cladistics* 1992; 8: 103-124.
- [95] Ladiges PY, Udrovicic F, Drinnan AN. Eucalypt phylogeny – molecules and morphology. *Aust Syst Bot* 1995; 8: 483-497.
- [96] Ladiges PY. Phylogenetic history and classification of eucalypts. In: Williams J, Woinarski J, eds. *Eucalypt Ecology. Individuals to ecosystems*. Cambridge: Cambridge University Press, 1997: 16-29.

- [97] Ladiges PY, Udovicic F, Nelson G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *J Biogeog* 2003; 30: 989-998.
- [98] Ladiges PY, Udovicic F. Comment on molecular dating of the age of eucalypts. *Aust Syst Bot* 2005; 18: 291-293.
- [99] Lande R, Schemske DW. The evolution of self-fertilisation and inbreeding depression in plants I. Genetic models. *Evol* 1985; 39: 24-40.
- [100] Lange RT. Carpological evidence for fossil *Eucalyptus* and other *Leptospermae* (subfamily Leptospermoideae of Myrtaceae) from a tertiary deposit in the South Australian arid zone. *Aust J Bot* 1978; 26: 221-223.
- [101] Lange RT. Evidence for lid-cells and host specific microfungi in the search for the Tertiary *Eucalyptus*. *Review of Palaeobotany and Palynology* 1980; 29: 29-33.
- [102] Latta RG. Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytol* 2004; 161: 51-58.
- [103] Levin DA, Kerster HW. Gene flow in seed plants. *Evol Biol* 1974; 7: 139-220.
- [104] Lynch M, Pfrender M, Spitz K, Lehman N, Hicks J, Allen D, Latta L, Ottene M, Bogue F, Colbourne J. The quantitative and molecular genetic architecture of a subdivided species. *Evol* 1999; 53: 100-110.
- [105] Main AR. Ghosts of the past: Where does environmental history begin? *Environ History* 1996; 2, 97-114.
- [106] Marques CM, Brondani RPV, Grattapaglia D, Sederoff R. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theor Appl Genet* 2002; 6: 727-737.
- [107] Martin HA, Gadek PA. Identification of *Eucalyptus spathulata* pollen and its presence in the fossil record. *Memoirs of the Association of Australasian Palaeontologists* 1988; 5: 311-327.
- [108] Martin HA. Evolution of mallee and its environment. In: Noble JC, Bradstock RA, eds. *Mediterranean Landscapes in Australia, Mallee Ecosystems and their Management*. Melbourne: CSIRO, 1989: 83-92.
- [109] McCoy F. Prodromus of the palaeontology of Victoria. *Geological Survey of Australia, Special Publications, Decade IV* 1876.
- [110] McDonald MW, Rawlings M, Butcher PA, Bell JC. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust J Bot* 2003; 51: 393-403.
- [111] McGowen MH, Wiltshire RJE, Potts BM, Vaillancourt RE. The origin of *Eucalyptus vernicosa*, a unique shrub eucalypt. *Biol J Linn Soc* 2001; 74: 397-405.
- [112] McKay JK, Latta RG. Adaptive population divergence: markers, QTL and traits. *Trend Ecol Evol* 2002; 17: 285-291.
- [113] McKinnon GE, Steane DA, Potts BM, Vaillancourt RE. Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). *Am J Bot* 1999; 86: 1038-1046.
- [114] McKinnon GE, Vaillancourt RE, Jackson HD, Potts BM. Chloroplast sharing in the Tasmanian eucalypts. *Evol* 2001; 55: 703-711.
- [115] McKinnon GE, Vaillancourt RE, Tilyard PA, Potts BM. Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. *Genome* 2001a; 44: 831-835.
- [116] McKinnon GE, Vaillancourt RE, Steane DA, Potts BM. The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. *Mol Ecol* 2004; 13: 3751-3762.

- [117] McKinnon GE, Jordan GJ, Vaillancourt RE, Steane DA, Potts BM. Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. *Phil Trans R Soc Lond B* 2004a; 359: 275-284.
- [118] McKinnon GE, Potts BM, Steane DA, Vaillancourt RE. Population and phylogenetic analysis of the cinnamoyl coA reductase gene in *Eucalyptus globulus* (Myrtaceae). *Aust J Bot* 2005; 48: 827-838.
- [119] Millar MA, Byrne M, Coates DJ, Stukely MJC, McComb JA. Mating system studies in jarrah, *Eucalyptus marginata* (Myrtaceae). *Aust J Bot* 2000; 48: 475-479.
- [120] Merila J, Crnokrak P. Comparison of genetic differentiation at marker loci and quantitative traits. *J Evol Biol* 2001; 14: 892-903.
- [121] Moran GF, Hopper SD. Genetic diversity and the insular population structure of the rare granite rock species *Eucalyptus caesia* Benth. *Aust J Bot* 1983; 31: 161-172.
- [122] Moran GF, Hopper SD. Conservation of the genetic resources of rare and widespread eucalypts in remnant vegetation. In: Saunders DA, Arnold GW, Burbidge AA, Hopkins AJM, eds. *Nature Conservation: The role of remnants of native vegetation*. Chipping Norton, NSW: Surrey Beatty and Sons, 1987: 151-162.
- [123] Moran GF. Patterns of genetic diversity in Australian tree species. *New For* 1992; 6: 49-66.
- [124] Muller J. Fossil pollen records of extant angiosperms. *Bot Rev* 1981; 47: 1-42.
- [125] Muona O. Population genetics in forest tree improvement. In: Brown AHD, Clegg MT, Kahler AL, Weir BS, eds. *Plant Population Genetics, Breeding and Genetic Resources*. Sunderland, Massachusetts: Sinauer Associates, 1990: 282-298.
- [126] Myburg AA, Griffin AR, Sederoff RR, Whetten RW. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F<sub>1</sub> hybrid based on a double pseudo-backcross mapping approach. *Theor Appl Genet* 2003; 107: 1028-1042.
- [127] Nesbitt KA, Potts BM, Vaillancourt RE, West AK, Reid JB. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* 1995; 74: 628-637.
- [128] Nevo E. Genetic diversity in nature and domestication. In: Henry RJ, ed. *Plant diversity and evolution: Genotypic and phenotypic variation in higher plants*. Wellingford: CABI, 2005: 287-315.
- [129] Nicolle D, Byrne M, Whalen M. A taxonomic revision and morphological variation within *Eucalyptus* series *Subulatae* subseries *Oleagineae* (Myrtaceae), including the oil mallee complex, of south-western Australia. *Aust Syst Bot* 2005; 18: 525-553.
- [130] Nicolle D. A classification and census of regenerative strategies in the eucalypt (*Angophora*, *Corymbia* and *Eucalyptus* – Myrtaceae), with special reference to the obligate seeders. *Aust. J. Bot* 2006; 54: 391-407.
- [131] Ochieng JW, Henry RJ, Baverstock PR, Steane DA, Shepherd M. Nuclear ribosomal pseudogenes resolve a corroborated monophyly of the eucalypt genus *Corymbia* despite misleading hypotheses at functional ITS paralogs. *Mol Phylogen Evol* 2007; 44: 751-764.
- [132] Ochieng JW, Steane DA, Ladiges PY, Baverstock PR, Henry RJ, Shepherd M. Microsatellites retain phylogenetic signals across genera in eucalypts (Myrtaceae). *General and Molecular Biology* 2007: in press.
- [133] Parra-O C, Bayly M, Udovicic F, Ladiges P. ETS sequences support the monophyly of the eucalypt genus *Corymbia* (Myrtaceae). *Taxon* 2006; 55: 653-663.
- [134] Patterson B, Vaillancourt RE, Pilbeam DJ, Potts BM. Factors affecting variation in outcrossing rate in *Eucalyptus globulus*. *Aust J Bot* 2004; 52: 773-780.

- [135] Peters GB, Lonie JS, Moran GF. The breeding system, genetic diversity and pollen sterility in *Eucalyptus pulverulenta*, a rare species with small disjunct populations. Aust J Bot 1990; 38: 559-570.
- [136] Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousoo A, Kremer A. Chloroplast DNA footprints of postglacial recolonisation by oaks. Proc Natl Acad Sci USA 1997; 94: 9996-10001.
- [137] Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG. Comparative organisation of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 2005; 14: 689-701.
- [138] Poke FS, Vaillancourt RE, Potts BM, Reid JB. Genomic research in *Eucalyptus*. Genetica 2005; 125: 79-101.
- [139] Poke FS, Martin DP, Steane DA, Vaillancourt RE, Reid JB. The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). Mol Phylogen Evol 2006; 39: 160-170.
- [140] Pole MS, Hill RS, Green N, Macphail MK. The Oligocene Berwick quarry flora – rainforest in a drying environment. Aust Syst Bot 1993; 6: 399-427.
- [141] Potts BM, Wiltshire RJE. Eucalypt genetics and genecology. In: Williams JE, Woinarski JCZ, eds. Eucalypt ecology. Individuals to ecosystems. Cambridge: Cambridge University Press, 1997: 56-91.
- [142] Potts BM, Savva M. Self-incompatibility in eucalypts. In: Knox RB, Sing MB, Troiani LF, eds. Pollination '88. Melbourne: Plant Cell Biology Research Centre, University of Melbourne, 1988: 165-170.
- [143] Potts BM, Barbour RC, Hingston AB, Vaillancourt RE. Genetic pollution of native eucalypt gene pools – identifying the risks. Aust J Bot 2003; 51: 1-25.
- [144] Pound LM, Wallwork MAB, Potts BM, Sedgley M. Self-incompatibility in *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). Aust J Bot 2002; 50: 365-372.
- [145] Pound LM, Patterson B, Wallwork MAB, Potts BM, Sedgley M. Pollen competition does not affect the success of self-pollination in *Eucalyptus globulus* (Myrtaceae). Aust J Bot 2003; 51: 189-195.
- [146] Pryor LD. The inheritance of some characters in *Eucalyptus*. Proc Linn Soc NSW. 1957; 82: 199-200.
- [147] Pryor LD, Johnson LAS. A classification of the Eucalypts. Canberra: Australian National University Press, 1971.
- [148] Pryor LD, Johnson LAS. *Eucalyptus* the universal Australian. In: Keast A, ed. Ecological Biogeography of Australia. Dr W Junk bv Publishers: The Hague 1981: 500-536.
- [149] Pryor LD, Williams ER, Gunn BV. A morphometric analysis of *Eucalyptus urophylla* and related taxa with descriptions of two new species. Aust Syst Bot 1995; 8: 57-70.
- [150] Reed DH, Frankham R. Correlation between fitness and genetic diversity. Cons Biol 2002; 17: 230-237.
- [151] Rossetto M, Jezierski G, Hopper SD, Dixon KW. Conservation genetics and clonality in two critically endangered eucalypts from the highly endemic south-western Australian flora. Biol Cons 1999; 88: 321-331.
- [152] Russell-Smith J, Lucas DE, Brock J, Bowman DMJS. Allosyncarpia-dominated rainforest in monsoonal northern Australia. J Veget Sc 1983; 4: 67-82.

- [153] Rye BL. Chromosome number variation in the Myrtaceae and its taxonomic implications. *Aust J Bot* 1979; 27: 547-573.
- [154] Rye BL, James SH. The relationship between dysploidy and reproductive capacity in Myrtaceae. *Aust J Bot* 1992; 40: 829-848.
- [155] Sampson JF. The population genetic structure of *Eucalyptus rhodantha* Blakely and Steedman and its allies *Eucalyptus crucis* Maiden and *Eucalyptus lane-poolei* Maiden. PhD Thesis. Nedlands: The University of Western Australia 1988.
- [156] Schaal BA, Olsen KM. Gene genealogies and population variation in plants. *Proc Natl Acad Sci USA*. 2000; 97: 7024-7029.
- [157] Sedgley M, Smith RM. Pistil receptivity and pollen tube growth in relation to the breeding system of *Eucalyptus woodwardii* (Sympyomyrtus: Myrtaceae). *Annal Bot* 1989; 64, 21-31.
- [158] Sedgley M, Granger L. Embryology of *Eucalyptus spathulata* and *E. platypus* (Myrtaceae) following selfing, crossing and reciprocal interspecific pollination. *Aust J Bot* 1996; 4: 661-671.
- [159] Shepherd M, Kasem S, Lee D, Henry R. Construction of microsatellite genetic linkage maps for *Corymbia*. *Silv Genet* 2006; 55: 228-238.
- [160] Singh G, Kershaw AP, Clark R. Quaternary vegetation and fire history in Australia. In: Gill AM, Groves RH, Noble IR, eds. *Fire and the Australian biota*. Canberra: Australian Academy of Science, 1980: 23-54.
- [161] Slee AV, Brooker MIH, Duffy SM, West JG. EUCLID. Eucalypts of Australia. 3rd ed. Canberra: CSIRO, 2006.
- [162] Smith S, Hughes J, Wardell-Johnson G. High population differentiation and extensive clonality in a rare mallee eucalypt: *Eucalyptus curtisii*. *Cons Genet* 2003; 4: 289-300.
- [163] Soltis PS, Gitzendanner MA. Molecular systematics and the conservation of rare species. *Cons Biol* 1999; 3: 471-483.
- [164] Sork VL, Nason J, Campbell DR, Fernandez JF. Landscape approaches to historical and contemporary gene flow in plants. *Trend Evol Ecol* 1999; 14: 219-224.
- [165] Steane DA, Byrne M, Vaillancourt RE, Potts BM. Chloroplast DNA polymorphism signal complex interspecific interactions in *Eucalyptus* (Myrtaceae). *Aust Syst Bot* 1998; 11: 25-40.
- [166] Steane DA, McKinnon GE, Vaillancourt RE, Potts BM. ITS sequence data resolve higher level relationships among the eucalypts. *Mol Phylogen Evol* 1999; 12: 215-223.
- [167] Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM. Development and characterisation of microsatellite loci in *Eucalyptus globulus* (Myrtaceae). *Silv Genet* 2001; 50: 89-91.
- [168] Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM. Higher-level relationships among the eucalypts are resolved by ITS-sequence data. *Aust Syst Bot* 2002; 15: 49-62.
- [169] Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus*. *DNA Res* 2005; 12: 215-220.
- [170] Steane DA, Conrod N, Jones RC, Vaillancourt RE, Potts BM. A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using microsatellite markers and quantitative traits. *Tree Genet Genom* 2006; 2: 30-38.
- [171] Stokoe RL, Shepherd M, Lee DJ, Nikles G, Henry RJ. Natural inter-subgeneric hybridisation between *Eucalyptus acmenoides* Schauer and *Eucalyptus cloeziana* F. Muell (Myrtaceae) in southeast Queensland. *Annal Bot* 2001; 88 563-570.

- [172] Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 2005; 14: 671-688.
- [173] Systsma KJ, Litt A, Zjhra ML, Pires C, Nepokroeff M, Conti E, Walker J, Wilson PG. Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern hemisphere. *Int J Pl Sci* 2004; 164: S85-105.
- [174] Thamarus KA, Groom K, Murrell J, Byrne M, Moran GF. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre and floral traits. *Theor Appl Genet* 2002; 104: 379-387.
- [175] Uдовичиц F, McFadden G, Ladiges PY. Phylogeny of *Eucalyptus* and *Angophora* based on rDNA spacer sequence data. *Mol Phylogen Evol* 1995; 4: 247-256.
- [176] Uдовичиц F, Ladiges PY. Informativeness of nuclear and chloroplast DNA regions and the phylogeny of the eucalypts and related genera (Myrtaceae). *Kew Bull* 2000; 55: 633-645.
- [177] Vaillancourt RE, Petty A, McKinnon GE. Maternal inheritance of mitochondria in *Eucalyptus globulus*. *J Hered* 2004; 95: 353-355.
- [178] Van Tienderen PH, de Haan AA, van der Linden CG, Vosman B. Biodiversity assessment using markers for ecologically important traits. *Trend Ecol Evol* 2002; 17: 577- 582.
- [179] Wendel JF, Doyle JJ. Phylogenetic incongruence, window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ, eds. Molecular systematics of plants II, DNA sequencing. Norwell: Kluwer Academic Publishers, 1998: 265-296.
- [180] Wheeler M, Byrne M, McComb JA. Little genetic differentiation within the dominant forest tree, *Eucalyptus marginata* (Myrtaceae) of south-western Australia. *Silv Genet* 2003; 52: 254-259.
- [181] Wheeler M, Byrne M. Congruence between phylogeographic patterns in cpDNA variation in *Eucalyptus marginata* (Myrtaceae) and geomorphology of the Darling Plateau. *Aust J Bot* 2006; 4: 17-26.
- [182] Whitehead A, Crawford DL. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol* 2006; 15: 1197-1211.
- [183] Whittlock S, Steane DA, Vaillancourt RE, Potts BM. Molecular evidence shows that the tropical boxes (*Eucalyptus* subgenus *Minutifructus*) are over-ranked. *Trans R Soc SA* 2003; 127: 27-32.
- [184] Whittlock SP, Apiolaza LA, Kelly CM, Potts BM. Genetic control of coppice and lignotuber development in *Eucalyptus globulus*. *Aust J Bot* 2003; 51:57-67.
- [185] Wright S. The general structure of populations. *Ann Eugen* 1951; 15: 323-354.
- [186] Yu J, Hu SN, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 2002; 296: 79-92.

## Dahlia: Cytogenetics and Evolution

BRIAN GRANT MURRAY

School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland, New Zealand

### ABSTRACT

The genus *Dahlia* Cav. (Asteraceae, Coreopsideae) contains 36 species plus the well-known garden plant *D. variabilis*. Taxonomic treatments over the years have recognized increasing numbers of species and have subdivided the genus into several different groupings of species into sections. With the exception of one of the sections, *Entemophyllum*, the molecular phylogenetic studies do not support these subdivisions and instead divide the genus into three major, well-supported clades. The genus is cytologically variable with three basic numbers,  $x=16$ , 17 and 18, and the  $x=16$  group contains species with  $2n=32$  and/or 64 chromosomes. The chromosome numbers are distributed into specific clades and the results of the phylogenetic studies suggest that  $x=17$  is ancestral,  $x=16$  and 18 having arisen via aneuploid loss or gain, and that the polyploids have arisen independently several times. Using GISH on interspecific hybrids, it has been demonstrated that meiotic chromosome pairing occurs both within and between the chromosomes derived from the parental species, suggesting that the genus is paleopolyploid. Thus, the  $2n=32$ , 34 and 36 chromosome species are tetraploid and the  $2n=64$  species are octoploid.

**Key Words:** *Dahlia*, FISH, GISH, molecular cytogenetics, phylogenetics

---

*Address for correspondence:* School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland 1001, New Zealand. Tel: +64 9 373 7599 x 88334, Fax: +64 9 373 7416, E-mail: b.murray@auckland.ac.nz

## INTRODUCTION

The genus *Dahlia* Cav. (Asteraceae, Coreopsideae) provides an excellent example of the success of applying a variety of molecular techniques to elucidate evolutionary and systematic problems. The genus consists of 36 species, several of which have been described only in the past decade [19, 37, 38, 39, 40, 41, 42]. Most of the species are endemic to Mexico but some grow further south in Central America and northern South America [41] and the cultigen, the garden dahlia, usually referred to as *D. variabilis* is a cosmopolitan garden plant with many thousands of named cultivars [52].

### Brief Taxonomic History of Dahlia

*Dahlia* species were sent from Mexico to the Royal Botanic Garden in Madrid in 1789 where they were cultivated and described by Cavanilles in his *Icônes* during the years 1791–1801 [2, 3]. In the first modern taxonomic treatment, Sherff [43] recognised 18 species, with eight varieties and three forms, which he placed in three sections, *Pseudodendron*, *Epiphytum* and *Dahlia*. The subsequent monograph by Sørensen [46], which was based on extensive new collections, field observation and chromosome counts, contained 29 species and six varieties that were placed in four sections. Sørensen's sections *Pseudodendron* and *Epiphytum* contained the same species as those of Sherff, but the species grouped by Sherff in section *Dahlia* were divided by Sørensen into two sections, *Entemophyllum* and *Dahlia*. Section *Dahlia* was further divided into two subsections, one of which, *Merckii*, contained just the single species, *D. merckii*. The third major study was by Giannasi [11, 12] who carried out an in depth analysis of flavonoid profiles of 19 species. His conclusions largely agree with those of Sorensen, though the flavonoid profiles indicated that there were two major chemically-based groups in the genus; sections *Dahlia* and *Pseudodendron* were characterized by the production of flavonols in all taxa examined, whereas species in section *Entemophyllum* lacked flavonols and had 6-methoxy and 6,4'-dimethoxy flavones. He suggested that the genus could be divided into just two sections, *Entemophyllum* and *Dahlia*, though he recognized that there were morphological, chromosomal and chemical variants in the two groups. Since then 10 additional species have been described [19, 37, 38, 39, 40, 41, 42] Thus, there are now 36

species plus the cultigen *D. variabilis*, whose nomenclatural status is somewhat controversial [18], that have been placed in two, three or four sections by various authors.

## Early Chromosome Studies

The first chromosome counts were made by Ishikawa [20] who found  $2n=32$  in *D. coccinea* (as *D. coronata*) and  $2n=64$  in *D. variabilis*. Thus, some cytological complexity with a polyploid series was apparent from this early stage. Many other species have now been shown to have either, or both, of these chromosome numbers (Table 1), and two other numbers  $2n=34$  and  $2n=36$  have also been observed [25, 46]. *Dahlia*, therefore, has three basic numbers,  $x=16$ , 17 or 18, which Johnson and Brandham [21] have proposed represent an aneuploid series.

There are some notable discrepancies in some of the published counts for several species, some of which have become more apparent following the phylogenetic studies described below. A count of  $2n=72$  by Sundberg et al. [51] for *D. apiculata* contrasts with  $2n=32$  obtained by Hansen and Hjerting [18]. Sundberg et al. [51] got  $2n=48$  for *D. australis* var. *australis* but three other studies got  $2n=32$  or 64 (Table 1). In *D. dissecta* Turner et al. [54] got  $2n=36$  compared to  $2n=34$  by Sørensen [46]; in *D. excelsa* Mehra and Remanandan [32] reported  $2n=26$ , but in a subsequent paper (Mehra and Remanandan [33]) got  $2n=32$ . Koul and Gohil [23] counted  $2n=36$  in *D. pinnata*, but all other counts are  $2n=64$  for this species. In *D. scapigera*, Turner et al. [54] found  $2n=36$ , but Sørensen [46] reported  $2n=32$ . It seems most likely that the deviant numbers have arisen from observations on incorrectly identified material, but without a careful check of any available herbarium voucher specimens for the chromosome counts this cannot be verified. However, given the relatively large size and ease of preparation of *Dahlia* chromosomes, it seems most unlikely that these aberrant chromosome numbers are correct.

## Karyotype Studies

Karyotype information is available for about one-third of *Dahlia* species [7, 8]. Conventional karyotyping has shown that the chromosome complements of the species are very similar, with the majority of the chromosomes being metacentric and the remainder being acrocentric. However, the proportions of these two groups differ depending on basic number. In the  $x=16$  species, at both ploidy levels, there are 14

**Table 1.** Published chromosome numbers of *Dahlia* species. Where gametic numbers (n) have been reported, these have been converted to somatic numbers (2n) for ease of comparison.

Species		Chromosome number (2n)	Reference
<i>D. apiculata</i> (Sheff) P.D. Sørensen		32	[7, 18]
<i>D. cf. apiculata</i>		72	[51]
<i>D. atropurpurea</i> P.D. Sørensen		64	[46]
<i>D. australis</i> var. <i>australis</i> P.D. Sørensen		32, 64	[7, 11, 46]
<i>D. australis</i> var. <i>serratior</i> (Sheff) P.D. Sørensen		48	[51]
<i>D. australis</i> P.D. Sørensen		32	[11]
<i>D. australis</i> var. <i>barkeriae</i> Knowles and Westc.		32, 64	[18]
<i>D. brevis</i> Sørensen		64	[46]
<i>D. campanulata</i> Saar, P.D. Sørensen and Hjert.		32	[18, 46]
<i>D. cardiophylla</i> S.F. Blake and Sheriff		32	[40]
<i>D. coccinea</i> Cav.		32	[46]
<i>D. coccinea</i> Cav.		32, 64	[20, 24, 25, 26, 27, 30, 31]
<i>D. coccinea</i> Cav.		64	[7, 11, 18, 21, 46]
<i>D. dissecta</i> S. Watson		36	[23, 51, 53]
<i>D. dissecta</i> S. Watson var. <i>dissecta</i>		34	[54]
<i>D. excelsa</i> Bentham		36	[8, 10, 46]
<i>D. excelsa</i> Bentham		34	[32]
<i>D. hertingii</i> H.V. Hansen and P.D. Sørensen		32	[33]
<i>D. imperialis</i> Rozel		32	[19]
			[1, 7, 11, 22, 25, 26, 35, 45, 46]

Table 1 contd.

Table 1 contd.

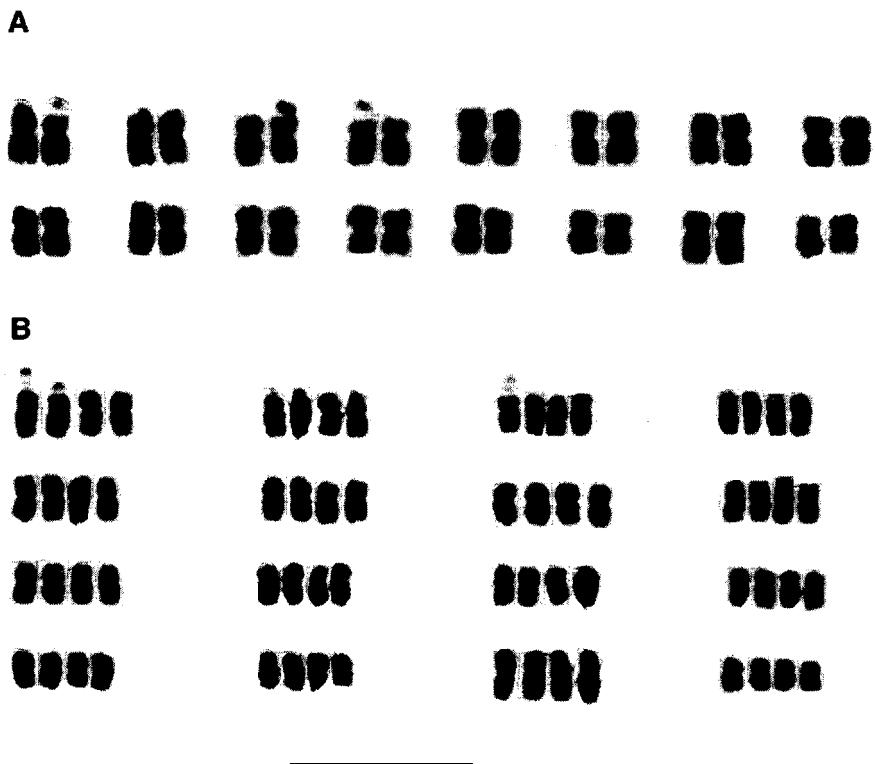
<i>D. linearis</i> Scherff	34	[11, 46]
<i>D. merckii</i> Lehman	36	[8, 11, 18, 21, 25, 26, 27, 46]
<i>D. mollis</i> P.D. Sørensen	32	[46]
<i>D. parvibracteata</i> Saar and P.D. Sørensen	32	[37]
<i>D. pinnata</i> Cav.	64	[7, 15, 16, 17, 46, 55, 56]
	36	[23]
<i>D. pteropoda</i> Sherff	64	[50]
<i>D. rufidis</i> P.D. Sørensen	32	[7, 18]
<i>D. ru picola</i> P.D. Sørensen	34	[46]
<i>D. scapigera</i> Knowles and Westc.	36	[54]
<i>D. scapigeroides</i> Sherff	32	[46]
<i>D. sheffii</i> P.D. Sørensen	34	[7, 21]
	32	[21]
	64	[46]
	32, 64	[7, 18]
<i>D. sorense nii</i> H.V. Hansen and J.P. Hjerting	64	[18]
<i>D. subligosa</i> (P.D. Sørensen) Saar and P.D. Sørensen	34	[42]
<i>D. tenuis</i> B.I. Rob. and Greenm.	32	[46]
<i>D. tenuicaulis</i> P.D. Sørensen	32	[7, 18, 46, 51]
<i>D. tubulata</i> P.D. Sørensen	32	[18, 47]
<i>D. variabilis</i> Desf.	64	[7, 20, 22, 25, 26, 27, 36]

metacentrics and two acrocentrics per chromosome set or genome (Fig. 1). These observations support an autotetraploid origin for the  $2n=64$  species and races within species. In the  $x=17$  and  $x=18$  species there are 11 metacentrics plus six acrocentrics, and 16 metacentrics and two acrocentrics, respectively [8] (Fig. 2).

There are distinct differences in the number of nucleolar organizer (NOR) chromosomes in the different basic number groupings. Using fluorescent in situ hybridization (FISH) with a 45S rDNA probe Gatt et al. [7, 8] showed that the  $2n=32$  species have eight hybridization sites, and the  $2n=64$  species have 16. However, not all these sites were transcriptionally active since staining with silver nitrate, which can be used to identify NORs that were transcriptionally active in the preceding interphase [13], usually identified fewer than the maximum number of four per genome. In the two species with  $2n=34$  that were studied, i.e., *D. dissecta* and *D. scapigeroides*, there were two and four 45S rDNA hybridization sites, respectively, and in *D. merckii*, with  $2n=36$ , there were eight hybridization sites.

## Molecular Systematic Studies

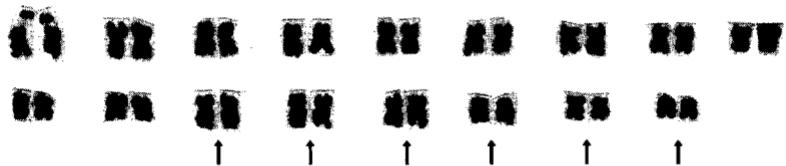
Two studies have used DNA sequence information to examine taxonomic affinities and evolutionary relationships in *Dahlia*. Gatt et al. [9] used sequence data from the internal transcribed spacers (ITS) of nuclear ribosomal DNA (nrDNA) to construct phylogenies of 13 species from three of Sørensen's four sections; material from section *Epiphytum* was not available to them for analysis. Phylogenetic reconstructions from both parsimony analysis and pairwise distance data produced a single tree with four clades. One clade contained the two species with  $2n=34$  that had been placed in Sørensen's section *Entemophyllum*. A second branch, which was divided into two clades, included all the species with chromosome numbers based on  $x=16$ . One clade contained species from section *Dahlia*, whereas the other included species from both sections *Dahlia* and *Pseudodendron*. Sequence variation between the species based on  $x=16$  was small with several polytomies. The fourth clade contained just *D. merckii* with  $2n=36$  chromosomes. Thus, this study did not support the separation of the  $x=16$  taxa into two sections. Additional support for this conclusion comes from earlier artificial hybridization studies that showed that hybrids between species in the two sections exhibited regular meiotic chromosome pairing and high fertility, and is detailed below [9,10].



**Fig. 1.** Karyotypes of *D. coccinea* ( $2n=32$ ) (A) and *D. variabilis* 'Pineholt Princess' ( $2n=64$ ) (B). Scale bar =  $10\mu\text{m}$ . (source: Gatt et al.)[7].

The second study by Saar et al. [41] used more than twice as many species as that of Gatt et al. [9], and used the external transcribed spacer of nuclear ribosomal DNA (ETS) in addition to ITS. Thus, there were also more than twice as many bases available for phylogenetic reconstruction in this study. The phylogenies produced using the two regions separately were very similar, with just a couple of differences in the ordering of species, and the placement of the species analyzed by Gatt et al. [9] was essentially similar in the two studies. Consequently, Saar et al. [41] combined their two data sets and this combined tree is shown in Fig. 3. One branch of the tree forms a well supported clade containing five species, all of which have been placed in section *Entemophyllum* and all have the same chromosome number,  $2n=34$  (although there is a report of  $2n=36$  for *D. dissecta* – referred above). *Dahlia merckii* and *D.*

A



B

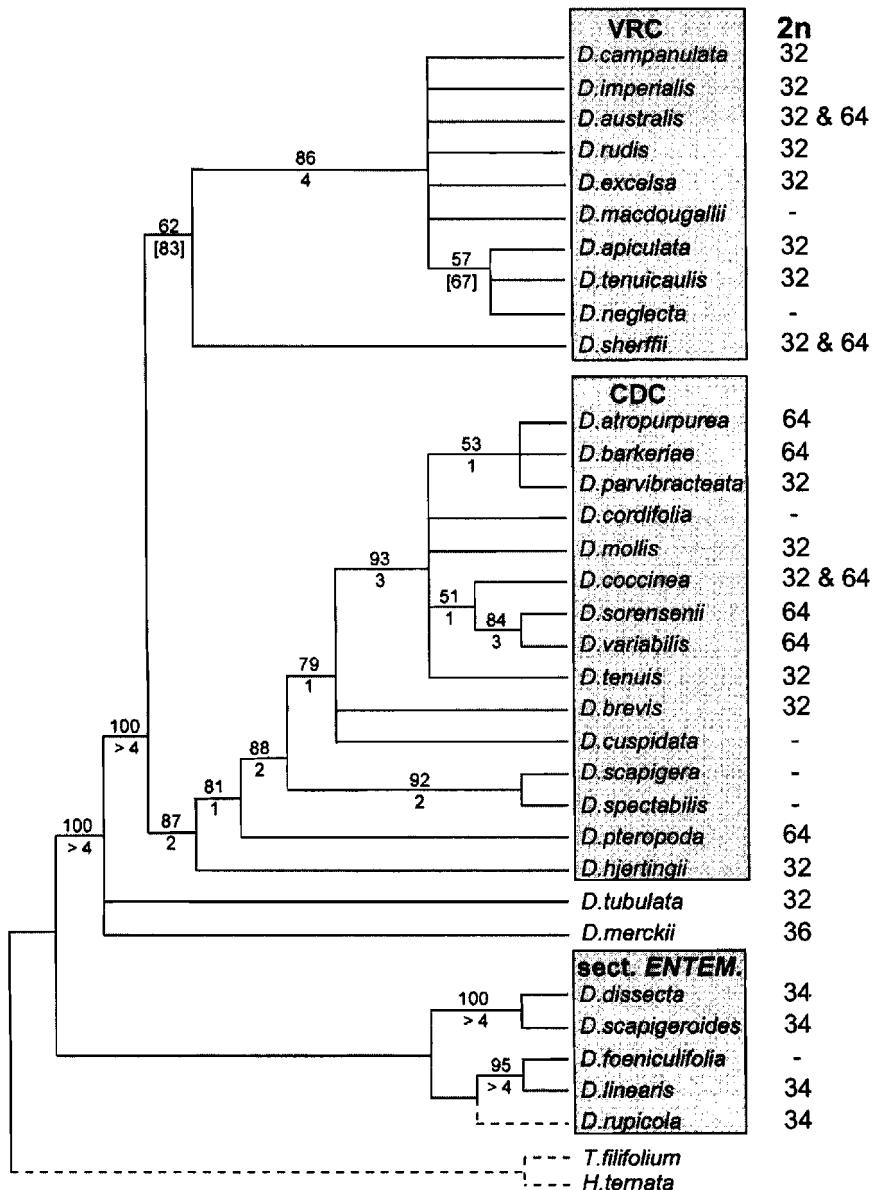


C



**Fig. 2.** Karyotypes of *D. dissecta*  $2n=34$  (A), *D. scapigeroides*  $2n=34$  (B) and *D. merckii*  $2n=36$  (C). Arrows indicate the acrocentric chromosomes. Scale bar =  $10\mu\text{m}$ . (source: Gatt et al.)[8].

*tubulata* form a polytomy but are well separated from the rest of the species in the second major branch of the tree. *Dahlia merckii* is the only species of the genus that has reliably been reported to have  $2n=36$  and it is noteworthy that it shows such a close relationship to *D. tubulata*, which has  $2n=32$ . The remaining species, all of which are based on  $x=16$ , fall into two clades that Saar et al. [41] have called “the variable root clade” (VRC), because of variation in their root systems, and “the core Dahlia clade” (CDC), which includes the cultigen. Thus, there appears to be little support for the other sections of the genus as the VRC clade contains species previously placed in *Epiphytum*, *Pseudodendron* and



**Fig. 3.** Phylogenetic relationships in *Dahlia*, bootstrap tree for combined ITS and ETS sequences. VRC = variable root clade, CDC = core *Dahlia* clade, Sect. Entem = section *Entemophyllum*, 2n = chromosome numbers taken from Table 1. source: Figure modified from Saar et al. [41] with permission.

*Dahlia*; the CDC clade exclusively contains species from section *Dahlia*. Since the publication of their phylogenetic tree, Saar and Sørensen [42] have described an additional species, *D. sublignosa*. Without reproducing their complete tree from the earlier phylogenetic study, they have placed it in the clade that contains all the species previously placed in section *Entemophyllum*, as it is most closely related to *D. foeniculifolia* or *D. linearis*. *Dahlia sublignosa* was also found to have  $2n=34$ , as do all the other species in this clade (Fig. 3).

These phylogenetic studies suggest that the ancestral chromosome number for the genus is  $x=17$ , i.e., the number found in the species in the most basal clade and, interestingly, also in one of the outgroup species, *Hidalgoa ternata* Llave and Lex [49]. The two other basic numbers in the genus,  $x=16$  and  $x=18$ , would appear to have evolved from the loss of two chromosomes, i.e., in the  $x=16$  group; and the gain of two chromosomes, i.e., in the  $x=18$  species. The karyotypes of *Dahlia* species are comprised of metacentric and acrocentric chromosomes (Figs. 1 and 2) that give no indication of Robertsonian fusion or fission events [7, 8] so it would appear that these changes are probably examples of aneuploid loss and gain. The polyploid nature of the genus (referred below) and the likely polyploid origin of the entire angiosperm phylum [6] mean that there is a greater buffering capacity of initial genetic imbalances created by the loss and gain of chromosomes than previously thought. Also, there are now well-documented examples, including several in the Asteraceae, where different basic numbers have arisen via aneuploid change both within a species as in *Rutidosis leptorrhynchoides* [34] and between species as in *Hypochaeris* [4]. Thus, in *Dahlia* a straightforward aneuploid series is not present, as suggested by Johnson and Brandham [21], but two apparently independent lineages exist, one associated with chromosome loss and the other with chromosome gain.

One final obvious conclusion that can be drawn from the *Dahlia* phylogeny is that chromosome doubling has occurred independently a number of times in both the VRC and CDC clades, where the species are based on  $x=16$ .

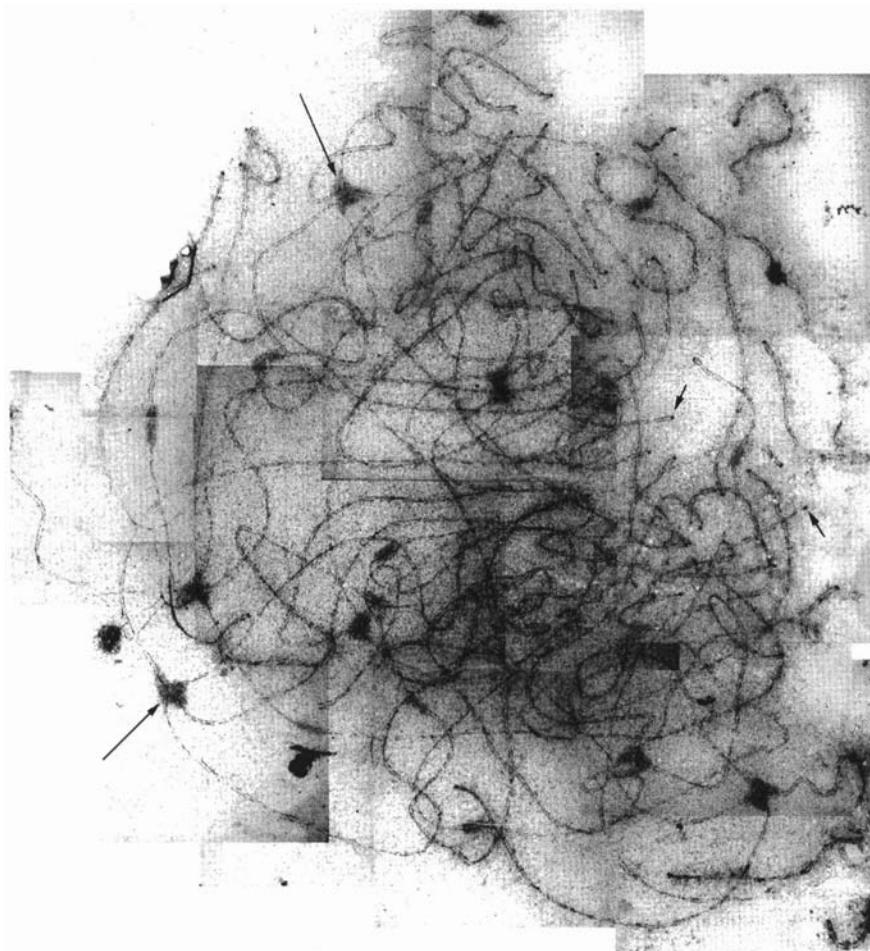
### **Pachytene Pairing, Synaptonemal Complexes and Secondary Associations At Metaphase I**

Lawrence [26, 27] observed that at metaphase I in *D. coccinea* ( $2n=32$ ) and *D. merckii* ( $2n=36$ ) there were secondary associations of bivalents,

and in *D. variabilis* ( $2n=64$ ) secondary associations of bivalents, quadrivalents and hexavalents. These associations were not chiasmate and he proposed that they were a consequence of post-synaptic phenomena that were a reflection of homologous relationships in hybrid polyploids. A noteworthy feature of the observations of synaptonemal complexes at pachytene in several *Dahlia* species by Gatt et al. [7, 8] was the association of centromeres in clusters involving from two to four pairs of chromosomes (Fig. 4). The composition of the centromeric material that was observed is unknown but this clear association of sets of chromosomes may to some extent explain the secondary associations observed by Lawrence [26, 27]. Association of centromeres at pachytene may well result in the close alignment of metaphase I bivalents, but whether these metaphase associations involve homologous chromosome sets is also unknown at this stage. The spread synaptonemal complexes also showed a clustering of telomeres in a restricted area despite the dispersal of the chromatin by the surface spreading technique. Telomeres were readily identified by the presence of a dark staining spherical granule associated with the ends of the lateral elements of the synaptonemal complexes (Fig. 4).

### **Are Dahlias Diploids or Ancient Polyploids?**

The high basic numbers found in the genus would suggest that species with 32, 34 or 36 chromosomes are ancient polyploids, following the suggestion that basic numbers higher than  $x=12$  [48] or 14 [14] are indicative of polyploidy. The earliest suggestion that *Dahlia* species were all polyploid followed the genetic studies carried out by Lawrence more than 70 years ago [25]. He summarized the evidence for this as follows [29]: the basic numbers in *Dahlia* are high for species in the Asteraceae,  $x=9$  being the most common number in the family [45]; four genes were identified that governed flower colour in *D. variabilis*, each of which showed tetrasomic inheritance associated with the random pairing of four homologous chromosomes, and at diakinesis and metaphase I frequent secondary associations [5] were observed between bivalents. In addition, Lawrence [27] showed that the cultigen was the only species that contained both pigment groups found in the genus, yellow, orange and red pigments that he proposed came from *D. coccinea*, and pink and lilac from *D. pinnata*. These species were the ones that formed the initial introductions to Europe, and Lawrence [25, 26] and Lawrence and Scott-



**Fig. 4.** Electron micrograph of surface-spread synaptonemal complexes of *D. pinnata* ( $2n=64$ ). Long arrows indicate centromere associations, short arrows indicate telomeres. Scale bar = 10  $\mu\text{m}$ . (source: Gatt et al.)[7].

Moncrieff [28] proposed that *D. variabilis* was an autoallopolyploid that had arisen following chromosome doubling of a hybrid between *D. coccinea* and *D. pinnata*. A contrary view put forward by Sørensen [46] was that the 32 chromosome species were diploids, presumably because this was the lowest chromosome number found in the genus.

Gatt et al. [7] examined meiotic pairing in a range of *Dahlia* species and observed exclusive bivalent formation in seven species with  $2n=32$  and a very low frequency of quadrivalent formation in five species with  $2n=64$ . They interpreted this to suggest that there is a bivalent-promoting mechanism that inhibits multivalent formation at both ploidy levels. A major insight into the problem has come from the application of genomic in situ hybridization (GISH) to the analysis of chromosome pairing in interspecific hybrids between *Dahlia* species [8]. Hybrids were made between species with different basic numbers, *D. sherffii* ( $2n=32$ ) with *D. dissecta* ( $2n=34$ ), and *D. sherffii* with *D. merckii* ( $2n=36$ ). The chromosomes of the parental species in these hybrids can be readily distinguished in both mitosis and meiosis with GISH (Fig. 5) and consequently it was possible to analyze the pairing behaviour of the hybrids knowing the origins of the chromosomes involved in the different configurations. Chromosome pairing was somewhat irregular with univalents formed in appreciable numbers, but bivalents, trivalents and quadrivalents were also observed (Fig. 5; Table 2). The key point obtained from meiotic studies of these hybrids was that pairing occurred both between chromosomes from different parental genomes as well as between chromosomes within parental genomes. In other words, there was sufficient homology for pairing to occur between chromosomes from the same parent, suggesting that genome duplication had occurred in the ancestors of these *Dahlia* species. Trivalents involving two chromosomes from one parent and one from the other were observed in both hybrid combinations (Fig. 5) but no trivalents involving three chromosomes from the one parental genome were seen. In the *D. dissecta* x *D. merckii* hybrid, quadrivalents involving two chromosomes from each parent were also observed to form at low frequency. Gatt et al. [8] concluded that *Dahlia* species must be paleopolyploid. Gatt et al. [10] have also produced hybrids between eight different species with  $2n=32$  and five with  $2n=64$ . In these combinations, chromosome pairing was essentially similar to that seen in the parental species and using GISH, they showed that pairing in the 32 chromosome hybrids was always between genomes from different species. Thus, the  $x=16$  species would all appear to have similar genomes and obviously similar karyotypes (referred above) despite the fact that the genomes can be differentiated by GISH.

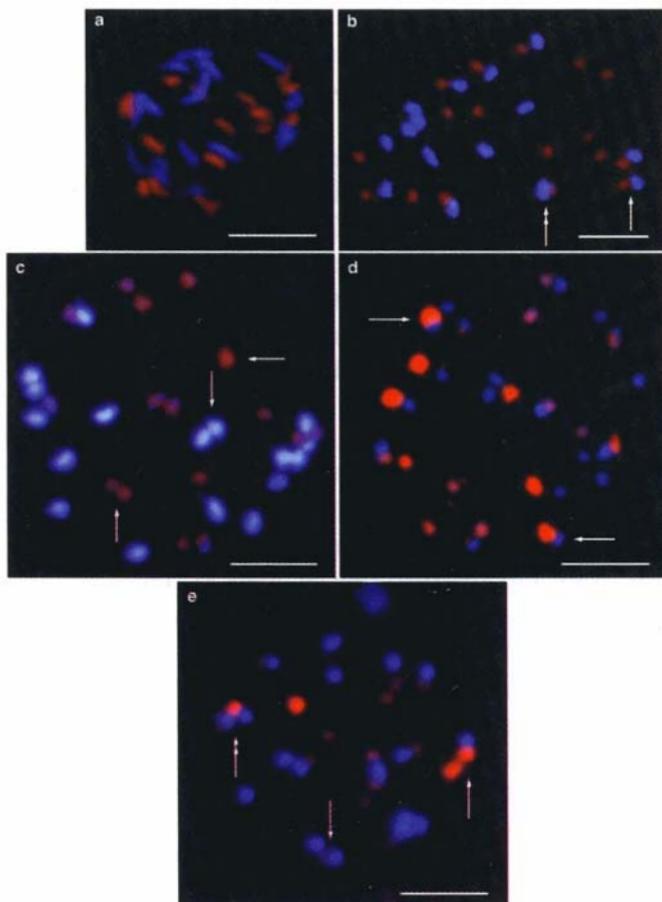


Fig. 5. Analysis of mitotic and meiotic metaphase I chromosome pairing in *Dahlia* hybrids using GISH.

- (a) *D. sherffii* x *D. dissecta* mitotic chromosomes showing 16 *dissecta* chromosomes (red) and 17 *sherffii* chromosomes (blue).
- (b) meiotic metaphase I in the same hybrid showing bivalent (single arrow head) and trivalent (double arrow heads) pairing between chromosomes from the different genomes.
- (c) meiotic metaphase I in the same hybrid showing bivalent formation between chromosomes from the same parental genome (arrows).
- (d) and (e) meiotic metaphase I in *D. dissecta* x *D. merckii* showing in (d) bivalent pairing between different genomes (arrows) and in (e) trivalents involving two chromosomes from one genome with one from the other (arrows with single heads) and a quadrivalent (arrow with double head). Scale bar = 10 $\mu$ m. (source: Gatt et al.) [8].

**Table 2.** Mean meiotic metaphase I chromosome configurations in interspecific hybrids following GISH showing the occurrence of pairing both between parental genomes as well as within parental genomes (modified from Gatt et al. 1999)[8].

Hybrid	No. of cells	I	I	II	II	II	III	III	IV
<i>D. sherffii</i> x <i>dissecta</i>	26	●	○	●-●	○-○	●-○	●-○	●-○-○	●-○-○
<i>D. dissecta</i> x <i>merckii</i>	29	10.2	8.5	1.0	1.2	3.6	0.3	0.6	0
		7.0	7.6	0.6	0.9	5.9	1.0	0.7	0.2

● = chromosome of *D. dissecta* (red in Fig. 5)

○ = chromosome of *D. sherffii* or *D. merckii* (blue in Fig. 5)

Note: The order of chromosomes for the trivalent and quadrivalent columns is not known, as it was often impossible to determine order, especially in V-shaped trivalents.

## CONCLUSION

A combination of phylogenetic and molecular cytogenetic studies has clarified many aspects of the evolution and systematics in *Dahlia*. The phylogenetic studies give clear insight into the major subdivisions of the genus bringing together groups of closely related species that have the same basic chromosome number. The relationships of some species, such as *D. tubulata*, remain unclear. The major problem of the ploidy status of *Dahlia* species has also been satisfactorily resolved with the  $2n=32$  species being shown to be tetraploid, and the  $2n=64$  species being octoploid. The different basic numbers are suggested to be a consequence of both loss and gain of chromosomes.

## Acknowledgment

The author would like to thank his partners Dr. Melanie Gatt and Dr. Keith Hammett for a fruitful collaboration on *Dahlia* cytogenetics and evolution, and Dr. Keith Hawwett for his valuable comments on this manuscript.

## References

- [1] Belling J. Homologous and similar chromosomes in diploid and triploid hyacinths. *Genetics* 1925; 10: 59-71.
- [2] Cavanilles AJ. *Icones et Descriptiones Plantarum* 1791-1801 Vol 1, 57.
- [3] Cavanilles AJ. *Icones et Descriptiones Plantarum* 1791-1801 Vol 3, 33.
- [4] Cerbah M, Souza-Chies T, Jubier MF, Lejeune B, Siljak-Yakovlev S. Molecular phylogeny of the genus *Hypochaeris* using internal transcribed spacer of nuclear rDNA: Inference for chromosomal evolution. *Mol. Biol. Evol* 1998; 15: 345-354.
- [5] Darlington CD. Studies in *Prunus* I and II. *J Genet* 1928; 19: 213-256.
- [6] de Bodt S, Maere S, de Peer Y. Genome duplication and the origin of angiosperms. *TREE* 2005; 20: 591-597.
- [7] Gatt M, Ding H, Hammett K, Murray B. Polyploidy and evolution in wild and cultivated *Dahlia* species. *Ann Bot* 1998; 81: 647-656.
- [8] Gatt M, Hammett K, Murray B. Confirmation of ancient polyploidy in *Dahlia* (Asteraceae) species using genomic *in situ* hybridization. *Ann Bot* 1999; 84: 39-48.
- [9] Gatt MK, Hammett KRW, Murray BG. Molecular phylogeny of the genus *Dahlia* Cav. (Asteraceae, Heliantheae-Coreopsidinae) using sequences derived from the internal transcribed spacers of nuclear ribosomal DNA. *Bot J Linn Soc* 2000a; 133: 229-239.
- [10] Gatt M, Hammett K, Murray B. Interspecific hybridization and the analysis of meiotic chromosome pairing in *Dahlia* (Asteraceae – Heliantheae) species with  $x=16$ . *Plant Syst Evol* 2000b; 221: 25-33.
- [11] Giannasi DE. The flavonoid systematics of the genus *Dahlia* (Compositae). *Mem New York Bot Gdn* 1975a; 26: 1-125.

- [12] Giannasi DE. Flavonoid chemistry and evolution in *Dahlia* (Compositae). Bul Torrey Bot Club 1975b; 102: 404-412.
- [13] Goodpasture C, Bloom SE. Visualization of NOR in mammalian chromosomes using silver staining. Chromosoma 1975; 53: 37-50.
- [14] Grant V. Genetics of Flowering Plants. New York: Columbia University Press, 1963.
- [15] Gupta RC, Gill BS. IOPB chromosome number reports LXXI. Taxon 1981; 30: 508-519.
- [16] Gupta RC, Gill BS. Cytology of family Compositae of the Punjab plains. Proc Indian Natl Acad Sci 1983; 49: 359-370.
- [17] Gupta RC, Gill BS. Cytopalynology of north and central Indian Compositae. J Cytol Genet 1989; 24: 96-105.
- [18] Hansen HV, Hjerting JP. Observations on chromosome numbers and biosystematics on *Dahlia* (Asteraceae, Heliantheae) with an account on the identity of *D. pinnata*, *D. rosea* and *D. coccinea*. Nord J Bot 1996; 16: 445-455.
- [19] Hansen HV, Sørensen PD. A new species of *Dahlia* (Asteraceae, Coreopsideae) from Hidalgo State, Mexico. Rhodora 2003; 105: 101-105.
- [20] Ishikawa M. Cytologische Studien von Dahlien. Bot Mag Tokyo 1911; 25: 1-8.
- [21] Johnson MAT, Brandham PE. New chromosome numbers in petaloid monocotyledons and in miscellaneous angiosperms. Kew Bul 1997; 52: 121-138.
- [22] Jose JC, Mathew PM. Chromosome numbers in the south Indian Heliantheae (Compositae). Comp News 1995; 27: 7-10.
- [23] Koul AK, Gohil RN. Cytotaxonomic conspectus of the flora of Kashmir (1) Chromosome numbers of some common plants. Phyton (Horn) 1973; 15: 57-66.
- [24] Koul AK, Wakhlu AL, Karihaloo JL. Chromosome numbers of some flowering plants of Jammu (Western Himalayas). II. CIS Chromosome Inform Serv 1976; 20: 32-33.
- [25] Lawrence WJC. The genetics and cytology of *Dahlia* species. J Genet 1929; 21: 125-159.
- [26] Lawrence WJC. The genetics and cytology of *Dahlia variabilis*. J Genet 1931a; 24: 257-306.
- [27] Lawrence WJC. The secondary association of chromosomes. Cytologia 1931b; 2: 352-384.
- [28] Lawrence WJC, Scott-Moncrieff R. The genetics and chemistry of flower colour in *Dahlia*: A new theory of specific pigmentation. J Genet 1935; 30: 155-228.
- [29] Lawrence WJC. The evolution and history of the dahlia. The Dahlia Annual. Leamington: National Dahlia Society, 1970: 16-24.
- [30] Matthew A, Mathew PM. IOPB chromosome number reports XLIX. Taxon 1975; 24: 501-516.
- [31] Matthew A, Mathew PM. Cytological Studies on the south Indian Compositae. Glimpes Pl Res 1988; 8: 1-177.
- [32] Mehra PN, Remanandan P. IOPB chromosome number reports XXII. Taxon 1969; 18: 433-442.
- [33] Mehra PN, Remanandan P. Cytological investigations on the Indian Compositae. II. Astereae, Heliantheae, Helenieae and Anthemideae. Caryologia 1974; 27: 255-284.
- [34] Murray BG, Young AG. Widespread chromosome variation in the endangered grassland forb *Rutidosis leptorrhynchoides* F. Muell. (Asteraceae: Gnaphalieae). Ann Bot 2001; 87: 83-90.

- [35] Powell AM, King RM. Chromosome numbers in the Compositae: Colombian species. Amer J Bot 1969; 56: 116-121.
- [36] Razaq ZA, Vahidy AA, Ali SI. Chromosome numbers in Compositae from Pakistan. Ann Missouri Bot Gard 1994; 81: 800-808.
- [37] Saar DE, Sørensen PD. *Dahlia parvibracteata* (Asteraceae, Coreopsidæ), a new species from Guerro, Mexico. Novon 2000; 10: 407-410.
- [38] Saar DE, Sørensen PD, Hjerting JP. *Dahlia spectabilis* (Asteraceae, Coreopsidæ), a new species from San Luis Potosí, Mexico. Brittonia 2002a; 54: 116-119.
- [39] Saar DE. *Dahlia neglecta* (Asteraceae, Coreopsidæ), a new species from Sierra Madre Oriental, Mexico. Sida 2002b; 20: 593-596.
- [40] Saar DE, Sørensen PD, Hjerting JP. *Dahlia campanulata* and *D. cuspidata* (Asteraceae, Coreopsidæ): two new species from Mexico. Acta Bot Mex 2003a; 64: 19-24.
- [41] Saar DE, Polans NO, Sørensen PD. A phylogenetic analysis of the genus *Dahlia* (Asteraceae) based on internal and external transcribed spacer regions of nuclear ribosomal DNA. Syst Bot 2003b; 28: 627-639.
- [42] Saar DE, Sørensen PD. *Dahlia sublignosa* (Asteraceae): a species in its own right. Sida 2005; 21: 2161-2167.
- [43] Sheriff EE. *Dahlia*. North American Flora II. New York: New York Bot Gard 1955; 45-59.
- [44] Solbrig OT, Kyhos DW, Powell M, Raven PH. Chromosome numbers in Compositae VIII: Heliantheae. Amer J Bot 1972; 59: 869-878.
- [45] Solbrig OT. Chromosomal cytology and evolution in the family Compositae. In: Heywood VH, Harborne JB, Turner BL, eds. The Biology and Chemistry of the Compositae. Vol. 1. London: Academic Press, 1977: 267-281.
- [46] Sørensen PD. Revision of the genus *Dahlia* (Compositae, Heliantheae–Coreopsidinae). Rhodora 1969; 71: 309-416.
- [47] Sørensen PD. New taxa in the genus *Dahlia* (Asteraceae, Heliantheae–Coreopsidinae). Rhodora 1980; 80: 353-360.
- [48] Stebbins GL. Chromosomal Evolution in Higher Plants. London: Arnold, 1971.
- [49] Strother JL. More chromosome studies in Compositae. Amer J Bot 1983; 70: 1217-1224.
- [50] Strother JL, Panero JL. Chromosome studies: Mexican Compositae. Amer J Bot 2001; 88: 499-502.
- [51] Sundberg S, Cowan CP, Turner BL. Chromosome counts of Latin American Compositae. Amer J Bot 1986; 73: 33-38.
- [52] The International Register of *Dahlia* Names. London: Royal Horticultural Society 1969 et seq.
- [53] Turner BL, Powell M, King RM. Chromosome numbers in the Compositae. IV. North American species, with phyletic interpretations. Amer J Bot 1961; 48: 216-223.
- [54] Turner BL, Powell M, King RM. Chromosome numbers in the Compositae. VI. Additional Mexican and Guatemalan species. Rhodora 1962; 64: 251-271.
- [55] Wang X-L, Li M-X. Observation of chromosomes on 10 composite species. J Wuhan Bot Res. 1987; 5: 111-117.
- [56] Zhang C-S. A preliminary study on making plant chromosomal specimens using peppermint oil compound as pretreatment agent. J Wuhan Bot Res 1998; 16: 280-282.

## **Author Index**

---

- Aarts MGM 2132  
Aas G 70  
Abbott RJ 75  
Abundance 38  
Adams E 224  
Adham Y 148, 149  
Agarwal I 150  
Agarwal SC 148  
Ageez A 212  
Aghmir A 216, 219  
Aguinagalde I 99  
Ahmed M 151  
Ahn SN 41  
Ahuja MR 75  
Aida M 212  
Akabane M 300, 301  
Albach DC 250  
Albar L 40  
Alexander HM 212  
Alfenas AC 338, 341  
Alfonso AA 213  
Ali A 151  
Ali J 223  
Ali S 212  
Ali SI 364  
Allen D 342  
Almeida J 11  
Alsaghir M 134  
Alvarez I 70  
AM 99  
Ambrose GJ 336  
Amin PW 128  
Ammiraju JS 35, 40  
Amores A 36  
Anamthawat-Jónsson K 14, 72, 78  
Anderson KH 97  
Anderson W 297  
Andersson H 212  
Angelici C 128  
Antonovics J 212, 224  
Aoki T 12  
APG 247  
Apiolaza LA 346  
Araki H 98  
Aravind L 37  
Araya A 224  
Arens P 298  
Argent G 297  
Arias DM 73  
Arisumi K 300  
Arumuganathan K 35, 148  
Ashley MV 72  
Ashton DH 341  
Ashton P 75  
Assuncao AGL 213  
Atanassov I 213  
Atlan A 219

- Aubourg S 37  
Audran JC 213  
Auger DL 38  
Auger JA 13, 76  
Ausubel FM 74  
Avery A 225  
Avise JC 96, 336  
Avramova Z 35  
Axelrod DI 70  
Axelsson E 213  
Aydogan A 150
- Baba T 39  
Bachmann K 13  
Bacilieri R 70, 74, 77, 98, 344  
Backstrom N 213  
Badeau V 73  
Badger B 297  
Baer J 299  
Baetcke KP 10  
Bahrman N 71  
Baidya BN 148  
Bailey MF 213, 220  
Baker BS 220  
Bakr MA 148  
Balasubramanian S 222  
Baltisberger M 213, 214  
Bancroft I 40  
Bao Z 37  
Barbacar N 213, 215, 221  
Barbosa W 134  
Barbour RC 336  
Barreneche T 71, 76  
Barrett SCH 213, 337  
Barry GF 35  
Barton M 213  
Barton MK 219  
Bartos J 217  
Barulina HI 148  
Bataillon T 216
- Batcho M 213  
Bauer N 78  
Baum M 148, 149, 151  
Baverstock PR 343  
Bayaa B 148, 149, 150  
Bayly MJ 337, 343  
Beales J 36  
Beaulieu JL 71  
Beck JT 172  
Bedi YS 71  
Beharav A 14  
Bejiga G 148  
Beko-Iseppon AM 151  
Belahbib N 337  
Belhassen E 219  
Bell JC 342  
Bellarosa R 71  
Belling J 362  
Belyaev ND 224  
Benedetteli S 99  
Bennett KD 96  
Bennett MD 10, 11, 35, 71, 297  
Bennetzen J 11  
Bennetzen JL 35, 37, 39, 71  
Bentham G 247  
Bentolila S 213  
Berger A 225  
Berglund D 221  
Berlin S 213  
Bertioli DJ 132  
Bertolucci FL 338, 341  
Besendorfer V 75, 78  
Bestimmung 215  
Betran E 37  
Bevan MW 40  
Bhandari TMS 151  
Bhandol P 11  
Bharathi M 130  
Bharti AK 35, 38

- Bhatia SK 218  
Bianchi-Hall CM 128  
Bianco P 72  
Biere A 213, 224  
Bieri T 223  
Bino R 302  
Bir SS 71  
Birch L 222  
Birchler JA 38  
Birks HJB 74  
Birus I 78  
Bittkau C 99  
Bittrich V 213, 248  
Bjerknes CA 13, 14  
Blackburn KB 213  
Blair MW 150  
Blanc G 35  
Blazey J 337  
Bleeker M 151, 302  
Bloom SE 363  
Bodenes C 71, 76  
Boecklen WJ 74  
Boerjan W 222  
Bogle AL 247, 248, 249  
Bogunic F 337  
Bohlmann J 39  
Bohorques A 73  
Boland DJ 337  
Bonavia D 128  
Bordacs S 76, 98  
Borisjuk NV 71  
Borrill M 11  
Borzan ă 71  
Bosemark N 301  
Bossema I 71  
Boudet N 37  
Boufford DE 96, 247  
Bousquet J 96  
Bouvet J 172  
Bowers JE 35, 38  
Bowler JM 337  
Bowman DMJS 345  
Bradshaw HD 214, 339  
Bradshaw HDH 71  
Brady KU 214  
Bramel P 130  
Bramel PJ 128  
Branch WD 130  
Brand J 338  
Brandes A 73  
Brandham PE 78  
Brandstrom M 213  
Brandt P 219  
Bratteler M 214  
Bravo J 128  
Brennicke A 219  
Bretag T 150  
Breuil-Broyer S 214  
Brewer S 71, 76, 99  
Bridgewater M 299  
Briggs BG 337  
Broadhurst L 338  
Brock J 345  
Brock MT 220  
Broecker WS 100  
Brondani C 337  
Brondani R 337  
Brondani RPV 337, 342  
Brooker MIH 337, 341, 345  
Brosius J 35  
Brown DG 40  
Brown GG 214  
Brown GK 297  
Brown S 73  
Brown SC 78, 224, 337  
Broyles SB 96  
Bruggmann R 35  
Brummer EC 39  
Brummitt RK 171  
Brunner S 35

- Budar F 214  
Buell CR 38  
Bult C 171  
Burbidge N 337  
Burczyk J 73  
Bureau T 39  
Bureau TE 14, 37  
Burg K 76, 98  
Burgess IP 339  
Burkhardt D 216  
Burow MD 128  
Burrus M 73  
Busso CS 36  
Butcher PA 337, 342  
Butorina AK 71  
Buzek J 214  
Byrne M 338-340, 343, 345, 346  
  
C 214  
Cabaña Fader A 131  
Caccia R 12  
Caceres ME 11  
Cai Q 72  
Cai Z 40  
Callen RA 336  
Campanella J 222  
Campbell BD 11  
Campbell DR 345  
Campbell R 301  
Campinhos EN 338  
Camus A 72  
Candoll 247  
Cannon CH 75, 98  
Cao M 72  
Capesius I 13  
Caradus JR 11  
Carey WA 97  
Carnevale S 97  
Caron H 98  
Carpenter R 11  
  
Carroll SB 214, 215  
Casa AM 130  
Casasoli M 71  
Casazza G 216  
Case AL 213  
Castiglione MR 223  
Causier B 224  
Cavalier-Smith T 11  
Cavallini A 11, 13  
Cavanilles AJ 362  
Ceccarelli M 11-13  
Cerbah M 72, 78, 362  
Cernohorska H 217  
Cervera MT 72  
Chaix G 338  
Chamberlain DF 297  
Chambers KL 13  
Chandra S 149  
Chang K 40  
Chang SB 41  
Chapman BA 35, 38  
Chappill JA 340  
Charlesworth B 12, 36, 214, 215, 217  
Charlesworth D 213-216, 219, 220, 225,  
    337  
Chase MW 171, 247, 249, 250  
Chaudhury M 149  
Chave J 219  
Chaw SM 250  
Cheddadi R 71, 99  
Chen J 37  
Chen L 224  
Chen M 35, 36  
Chen Y 224  
Chen YQ 249  
Chen ZJ 38, 299  
Cheng Z 38  
Cherubini M 99  
Chestnut 97  
Cheung WY 214

- Chevalier D 222  
Chiang KS 222  
Chiang TY 37  
Chiang YC 37  
Chinwalla A 223  
Chippendale GM 337  
Chokchaichamnankit P 72  
Choudhary MA 148  
Choumane W 148, 149  
Chowdhuri P 215  
Chulalaksananukul W 72  
Chybicki I 73  
Cihalinkova J 74  
Cionini G 11, 13  
Cionini PG 11, 12, 13  
CIPAL 148  
Cipriani G 224  
Clark K 40  
Clark MS 222  
Clark R 345  
Clark SE 215  
Cleary W 12  
Clegg MT 152  
Coart E 72, 98  
Coates DJ 338, 343  
Cohen D 302  
Cohen E 11  
Coleman A 298  
Collin CL 215  
Colonna N 12  
Comps B 98  
Conagin CHTM 128  
Conedera M 97  
Conrod N 345  
Conti E 346  
Cook BG 128  
Cook H 224  
Cook IO 338  
Cooke HJ 36  
Cooke R 38  
Copeland SC 134  
Cordeiro CM 341  
Cordum HS 223  
Corre L 73  
Correns C 215  
Corriveau J 298  
Costa G 224  
Costich DE 12, 13  
Cottrell J 75, 98  
Coulaud J 72  
Coustham V 214  
Couvet D 219  
Cowan CP 364  
Cox K 297  
Cox P 297  
Craft KJ 72  
Craig GB 217  
Crandall KA 249  
Crane PR 248, 249  
Craven LA 297  
Crawford D 172  
Crawford DJ 98  
Crawford DL 346  
Cremonini R 11, 12  
Crepet WL 72, 96  
Creste S 128  
Crick FHC 13  
Crisp M 338, 339  
Crisp MD 339  
Crnokrak P 343  
Croizat L 339  
Cronquist A 247  
Cross MJ 341  
Crosthwaite IC 128  
Csaikl U 76, 98  
Csaikl UM 75  
Cuadrado A 72  
Cubero JI 148  
Cuesta PAM 133  
Cuevas J 298

- Cullen J 298  
Cullis CA 12, 13  
Cummings MP 74  
Curtsinger JW 215  
Cutler M 215  
  
D'Cruz R 131  
Dai H 222  
Daighlian CP 96  
Damboldt J 215  
Dane F 96, 97, 98  
Daniels MJ 38  
Darnell MZ 220, 224  
Darwin C 72  
Davidian H 298  
Davidjuk YM 71  
Davies B 224  
Davis CC 247  
Davis MB 96  
Davy AJ 73  
Dawe DC 37  
Dawson IK 151  
de Almeida-Engler J 214  
De Beaulieu J-L 98, 99  
De Beck L 222  
de Bodt S 362  
de Candolle A 72  
de Haan AA 217, 346  
De Jeu M 302  
de Jong H 41  
De Keyser E 298, 301  
de la Herran R 217  
De Loose M 72, 298  
De Loose R 298  
de Oliveira AC 36  
De Pace C 11, 12  
De Paepe 214  
de Peer Y 362  
De Proft M 302  
De Riek 298, 301  
  
De Roo R 299  
De Schepper S 298  
Dean R 130  
Deane H 339  
Deans JD 74, 98  
Degraeve N 215  
DeGuilloux M-F 97  
Delcourt HR 96  
Delehaunty A 223  
Delehaunty K 223  
Delichere C 213, 215  
Delph LF 213, 215  
Delre V 12, 71  
Delseny M 36, 38  
Delves A 340, 341  
D'Emerico S 72  
Demesure B 73, 96, 98, 171, 344  
Demesure-Musch B 99  
Den Nijs T 299  
Dendauw J 298  
Dendy C 214  
Deng MB 247  
Denk T 96  
Deroose R 302  
Desfeux C 215, 220  
Devos KM 11, 36-38  
Dhesi JS 133  
Difazio S 39  
Ding H 362  
Dirlewanger E 72  
Disteche CM 217  
Dixon KW 344  
Doak DF 221  
Dodd RS 72  
Doebley JF 36  
Dolezel J 74, 217, 220, 223  
Donnison IS 216  
Donoghue MJ 249  
Doolittle WF 12

- Dotterl S 216  
Douglas GC 74  
D'Ovidio R 12  
Dow BD 72  
Doweld AB 247  
Doyle JJ 75, 98, 171, 248, 346  
Drabek J 299  
Drinnan AN 341  
Du C 36  
Du H 223  
Dubcovsky J 36  
Ducasse D 133  
Ducousoo A 70, 72, 76, 77, 98, 344  
Duffy SM 345  
Dumas C 300  
Duminil J 99, 344  
Dumolin-Lapegue S 73, 96, 98  
Dunavin LS 132  
Dunemann F 298  
Dunn M 36  
Dupanloup I 98  
Dupouey JL 73, 74  
Durán Y 148  
Durand B 216  
Durand R 216  
Durante M 11, 13  
Durrant A 12  
Dutkowski GW 339  
Duvall MR 247  
Dwivedi SL 128, 150  
Dzialuk A 73  
  
Earle ED 35, 148  
Ebbinghaus R 301  
Eckert AJ 299  
Eddy SR 37  
Eeckhaut T 298  
Eggens F 222  
Ehlers BK 216  
Ehrendorfer F 129  
  
Ehrlich SD 12  
Eichler EE 73  
El-Hassan H 150  
Ellegren H 213  
Elliot C 338  
Elliott CP 339  
Ellis MF 339  
Ellis RH 148  
Ellis THN 151  
Ellstrand NC 339  
Elzinga JA 213  
Emebiri LC 339  
Emery SN 220  
Endress PK 248, 249  
Endrizzi K 216  
Ennos R 99  
Ennos RA 339  
Enoch I 299  
Erixon P 222  
Ershire W 149  
Erskine 149  
Erskine W 148-151  
Esposto MC 11  
Estill JC 38  
Etoh JI 299, 300  
EUCLID 345  
Eujayl I 149  
Eunus AN 148  
Evans GM 12  
Excoffier L 298, 301  
  
Fabregat I 12  
Fairweather G 297  
Falcinelli M 11  
Falistocco E 11  
Fan C 40  
Farbos I 213, 216  
Faries MJ 133  
Farmer SB 172

- Farris JS 171, 250  
Fàvero AP 128, 134  
Felber M 97  
Felsenstein J 171, 248, 298  
Fenchel S 74  
Feng Q 36  
Fennell S 131  
Fensham RJ 340  
Ferguson CJ 171, 172  
Ferguson J 133  
Ferguson ME 128, 130, 149  
Fernández A 129-131, 133  
Fernandez JF 73  
Ferreira ME 132  
Ferris C 73, 97  
Festuca L 12  
Feuillet C 35, 36, 40  
Fewell G 223  
Filatov DA 213, 216, 217, 219  
Fineschi S 73, 96-99  
Fineschi S 344  
Finkeldey R 98  
Fior S 216  
Fischer GW 216  
Fishbein M 248  
Fisher RA 216  
Fitch WM 298  
Fladung M 77  
Flavell AJ 36  
Fleming CC 75  
Flint RB 336  
Flintham JE 38  
Fluch S 77  
Folter SD 213  
Force A 36  
Ford R 149  
Formanova N 214  
Forster BP 131  
Fowler M 341  
Francki M 35  
Franken J 302  
Frankham R 344  
Fransz P 12  
Frascaria N 96  
Fratini R 148  
Frediani M 11, 12  
Freeling M 35  
Freeman DC 220  
French EC 129  
Friend SA 134  
Frigerio JM 72  
Friis EM 248  
Fripp YJ 339  
Fu G 36  
Fu Y 96  
Fuchs J 73  
Fujisawa H 212  
Fujisawa M 38, 41  
Fukaki H 212  
Fukui K 40, 218  
Fulton L 223  
Fulton R 223  
Furman BJ 149  
Furuya M 223  
Gadek PA 342  
Galan R 217  
Galasso I 12  
Gale MD 36, 38  
Galgaro L 130  
Galle F 298  
Gallego G 73  
Gallois A 73  
Gao G 129  
Gao-LianMing 299  
Garber ED 222  
Garcia GM 129  
Garcia P 148  
Gardiner CA 337  
Garg R 37

- Garnatje T 77  
Garrido-Ramos MA 217  
Gatt M 362  
Gatt MK 362  
Gaur A 40  
Gaut BS 36, 38, 39  
Ge S 36, 40  
Gelati MT 11, 12  
Gennai D 11  
Gennon RJ 132  
Georgiev S 219, 221  
Gerber S 338  
Ghimpu V 129  
Giannasi DE 362, 363  
Gielly L 73, 96, 172  
Gill BS 71, 363  
Gillespie LJ 250  
Gillies A 98  
Gilmartin PM 222, 223  
Gimenes MA 128-130, 132  
Gioia P 340  
Giordani T 11  
Gitzendanner MA 345  
Glaettli M 217  
Glaubitz JC 339  
Glaz I 98  
Gleiser G 217  
Gloessl J 74, 77  
Godt MJW 339  
Goetsch L 299  
Goff SA 36, 339  
Gohil RN 363  
Goicoechea JL 35  
Goicoechea PG 98  
Gomory D 73, 98  
Gonzalez-Rodriguez A 73  
Goodpasture C 363  
Gordo I 217  
Gotttsberger G 218  
Goudet J 217  
Gouy M 40  
Gouyon PH 215, 219  
Graf AV 221  
Graham A 96, 97  
Grandbastien MA 73  
Granger L 345  
Grant KA 172  
Grant S 214, 217  
Grant SR 216, 218, 219, 221  
Grant V 172, 363  
Grattapaglia D 73, 132, 337, 339, 341, 342  
Graves AH 97  
Graves T 223  
Greenslade PJM 337  
Greenwood DR 339  
Gregory MP 129, 133  
Gregory WC 129, 130, 133  
Greilhuber J 12, 73, 129  
Greishammer U 129  
Greuter W 217  
Griffin AR 339, 343  
Grigoriev I 39  
Grimm G 96  
Grimm GW 96  
Grivet D 97, 99  
Grof B 129  
Groom K 346  
Grosso NR 129  
Grunau C 222  
Guandt D 134  
Guarino L 130  
Guberek T 98  
Guimaraes PM 132  
Gundlach H 35, 38  
Gunn BV 344  
Guo J 224  
Guo Q 97  
Gupta RC 363  
Gupta V 40

- Gutman GA 149  
Guy CL 72  
Guye A 72  
Guyot R 36  
Guzmán CA 129  
Gy I 37  
  
Haecker A 216  
Hagemann R 299  
Hahn JH 133  
Hahn WJ 250  
Hall AE 299  
Hall CR 40  
Halward T 130  
Halward TM 129  
Hamdi A 148, 149  
Hamilton MB 172  
Hamilton W 97  
Hammett K 362  
Hammett KRW 362  
Hammons RO 129  
Hamon S 338  
Hampe A 99, 344  
Hamrick JL 339  
Hamwieg A 149  
Han B 36  
Han F 37  
Hancock JF 74  
Handa T 299-302  
Haneman R 299  
Hansen BCS 99  
Hansen HV 363  
Hansen J 149  
Hanson MR 213  
Hao B 40  
Hao P 36  
Hao RM 248  
Harden R 99  
Hardenack S 217  
Hardner CM 339  
  
Hardy B 37  
Harms H 248  
Harrington SE 40  
Harris S 75  
Harrison PM 37  
Hartl DL 339  
Hartley NM 38  
Harvest Plus 149  
Harvey JA 213  
Harvey MS 340  
Hasnain SE 212  
Hassan MS 150  
Hattemer HH 78  
Hattori AK 74  
Hauser E 219  
Hausman JF 97  
Havey MH 149  
Hawkins LK 96  
Hawtin G 151  
Hayashi-Tsugane M 41  
Hayward M 301  
He G 129  
Heard E 217  
Heinze B 74  
Hellsten U 39  
Hemleben V 77, 96  
Hemmat M 73  
Henry J 220  
Henry JP 215  
Henry R 340, 341, 345  
Henry RJ 343, 345  
Herendeen PS 249  
Hermanutz LA 217  
Hermsen J 299  
Hernould M 215  
Herselman L 129  
Hertwig G 217  
Hertwig P 217  
Herzog S 75  
Heslop-Harrison J 217

- Heslop-Harrison JS 12, 78  
Hettel GP 37  
Heuertz MS 97  
Heursel J 298, 299  
Hewitt GM 73, 97  
Hibara K 223  
Hibsch-Jetter C 248  
Hickey WA 217  
Hicks J 342  
Hiei Y 218  
Higashiyama T 220  
Hijishita S 41  
Hill KD 340  
Hill RS 340, 344  
Hillier L 223  
Hills HG 247  
Hilu KW 129, 134, 299  
Hines B 340  
Hingston AB 340  
Hinnidaels S 213, 216  
Hjerting JP 363, 364  
Hladilova R 224  
Hobza R 217-219, 221, 224, 225  
Hodgkin J 36  
Hodkinson TR 74  
Hodnett G 299  
Hodurkova J 223  
Hoentjes DR 37  
Hogenboom N 299  
Hogers R 151, 302  
Hoisington DA 40  
Hokanson SC 74  
Holbrook CC 129  
Holker S 149  
Holly L 148  
Holman JE 340  
Holmes FM 340  
Holmes WBK 340  
Holton CS 216  
Honders SC 213  
Hong DY 36, 248  
Hooker JD 247  
Hoot SB 248, 250  
Hopf M 152  
Hopkins MS 130, 132  
Hopper SD 337, 340, 343, 344  
Hopper SH 338  
Horn W 299  
Horvat T 78  
Hoshino A 128  
Hou D 74  
Hou SF 223  
Houben A 214, 217  
Houliston GJ 217  
House SM 340  
Howard DJ 74  
Hrusakova P 217  
Hsieh C 220  
Hsu-KuoKai 299  
Hu SN 41, 346  
Huang H 96, 97  
Huang Y 36, 41  
Huang YL 249  
Hudson A 11  
Hufford L 248  
Hughes J 345  
Hughes JM 340  
Humphries CJ 341  
Hundscheid MP 217  
Hung KH 37  
Hunt CL 11  
Huntley B 74  
Hussain A 149  
Husted L 130  
Hüttel B 149, 151  
Huynh ML 37  
Hwang-ShihYing 299  
Hyam G 297  
Hykelova V 221

- Hyland BPM 337  
Ikehashi H 40  
Imaseki H 218  
Immonen S 12, 74  
Ingvarsson PK 224  
Innes DJ 217  
Ironside JE 217  
Irving E 299  
Isebrands JG 74  
Ishida T 212, 223  
Ishida TA 74  
Ishikawa M 363  
Isleib TG 129, 130, 132  
Isono E 220  
Ito H 41  
  
Jackson HD 340, 342  
Jackson S 36  
Jackson ST 97  
Jacobs M 224  
Jacobsen SE 215, 217  
Jain AC 149  
Jain M 37  
Jaiswal P 40  
James SH 340, 341, 345  
Janaki AE 299  
Janousek B 214, 217, 218, 220, 221, 223,  
    225  
Jansen RK 171, 172  
Jansson S 39  
Jarret RL 301  
Jarvis A 130  
Jaynes R 97  
Jays and oaks 71  
Jedinakova J 73  
Jegalian K 218  
Jenkins T 223  
Jensen J 74  
Jensen JS 98  
Jezierski G 344  
  
Jiang N 37  
Jin H 214  
Jin W 41  
Jin YK 39, 77  
Joandet S 71  
Johnk N 74  
Johnson GP 97, 98  
Johnson H 74  
Johnson LA 172, 250  
Johnson LAS 337, 340, 344  
Johnson MAT 363  
Johnson SJ 40  
Johnston JS 13, 299  
Johnston RD 337, 341  
Johnston S 299  
Jones KM 38  
Jones M 340, 341  
Jones ME 341  
Jones PG 130  
Jones RC 341, 345  
Jones RN 76  
Jordan GJ 343  
Jose JC 363  
Jr 71, 131  
Jr, Wilson TC 131  
Jubier MF 72, 362  
Jungans TG 341  
Juretic N 37  
Jurgens A 216, 218  
  
Kahl G 148  
Kahnau R 298  
Kalender R 12, 74  
Källersjö M 171  
Kamau E 219  
Kamisugi Y 223  
Kampfer S 74  
Kampmeijer P 218  
Kanamori H 38  
Kaneko S 98  
Kao T 300

- Kapur A 40  
Karihaloo JL 363  
Karis PO 216  
Karłowski WM 38, 39  
Kasem S 345  
Kashani N 72  
Katayose Y 39  
Kaul AK 151  
Kaur N 37  
Kawai M 302  
Kawakami J 130  
Kawano S 212, 220, 221, 224  
Kawashima S 300  
Kazama Y 212, 224  
Kehr A 299  
Keighery GJ 338  
Kejnovsky E 217-220, 223  
Kelleher CT 74  
Keller B 35, 36, 40  
Keller SR 218  
Kelly CM 346  
Kelly DL 74  
Kelly WG 219  
Kennington WJ 340, 341  
Kent-First M 218  
Kershaw AP 345  
Kerster HW 342  
Kesseli R 225  
Keys RD 128  
Khare MN 149  
Kho Y 299  
Khoshoo TN 218  
Khurana JP 37, 40  
Khush GS 37  
Kilian A 37  
Kim H 40  
Kim HR 38  
Kim K-J 172  
Kim M 38  
Kimura MT 74  
King 97  
King RA 73, 97  
King RM 364  
Kirkpatrick JB 341  
Kirst M 341  
Kirti PB 130  
Kita K 299, 300  
Klein AS 248, 249  
Klein I 298  
Klein PE 37  
Klein RR 37  
Kleinhofs A 37  
Kleinschmit J 74, 75  
Kleinschmit JRG 74  
Kloos D 223  
Klosová E 130  
Kluge AG 171  
Knauft DA 132  
Knox MR 151  
Knox R 300, 302  
Knuth P 218  
Kobayashi N 299-301  
Kobayashi Y 300  
Koch K 12  
Kochert GD 129, 130, 134  
Kocyan A 225  
Koelewijn HP 217  
Koenig WD 72  
Koes R 223  
Koga S 302  
Kojima T 132  
Kokontis J 222  
Komori T 218  
Kondrashov FA 37  
Konieczny A 74  
Konig A 99  
Koonin E 37  
Koonin EV 37

- Kostishin SS 71  
Kotsonis P 11  
Koul AK 363  
Koutnikova H 214  
Kozyreva O 221  
Krahenbuhl M 218  
Krämer F 171  
Krapovickas A 129, 130, 133  
Krebs P 97  
Kreis M 37  
Kremer A 70, 71, 73-78, 98, 99, 337, 344  
Kresovich S 128, 130, 132  
Kress WJ 250  
Kretschmer AE 131  
Kron KA 247, 250, 300  
Kruckeberg AR 214, 218  
Kubalkova M 74  
Kubat Z 218  
Kubekova H 217  
Kubisiak TL 97, 98  
Kubitzki K 248  
Kudrna D 35, 40  
Kuiper E 222  
Kulmala J 13  
Kumar A 36  
Kumar D 40  
Kumar LSS 131  
Kumar S 248  
Kumari S 150  
Kunishige M 300  
Kupfer P 218  
Kurashige Y 300  
Kuraya Y 218  
Kuroda-Kawaguchi T 223  
Kuroiwa T 220, 221  
Kusmenoglu I 150  
Küsmeno™lu I 149  
Kuzoff RK 250  
Kyhos DW 364  
Labate J 39  
Labbe T 71, 77  
Lacey CJ 341  
Lacks G 131  
Ladiges PY 297, 337, 338, 341-343, 346  
Laforest M 214  
Lagercrantz U 37  
Lahn BT 218  
Lai J 35, 37, 39  
Laigret F 71  
Lakhani JP 149  
Lal SS 148  
Lamote V 72  
Lan TH 36, 339  
Lande R 342  
Landry BS 214  
Lang P 96-98  
Lange RT 342  
Langer M 96  
Lanham PG 131  
Laporte V 214, 216, 219, 221  
Lardon A 213, 219  
Larsson J 219  
Larue EA 129  
Lascano CE 133  
Lascoux M 99  
Latalowa M 98  
Latrille P 223  
Latta RG 342  
Laube T 216  
Laux T 216  
Lavee S 298  
Lavia GI 130, 131, 133  
Law TF 219  
Lawrence WJC 37, 363  
Lawson DM 73  
Le Corre V 96  
Le Merrer G 219

- Leach D 300  
Leal-Bertioli SCM 132  
Lebel-Hardenack S 219, 221  
Lecharny A 37  
Lee DJ 345  
Lee FP 300  
Lee HS 38  
Lee M 40  
Lehman N 342  
Leitch IJ 11, 35, 297  
Lejeune B 72, 215, 362  
Lendvai G 172  
Lengerova M 217-220, 225  
Lentil 150  
Leoi L 132  
Leonard S 223  
Les Chenes 72  
Lespinet O 37  
Leus L 298  
Levin DA 171, 172, 342  
Levin JZ 215, 216  
Levin SA 219  
Levinson G 149  
Lewis D 219  
Lewis PO 98  
Lewis R 78  
Lexer C 71, 74, 75, 77, 214  
Li H 41  
Li J 40, 248  
Li S 41  
Li T 222  
Li WH 39, 40  
Li X 224  
Li Y 41  
Li-DeZhu 299  
Liang H 299  
Lickey EB 172  
Liden M 221  
Lin W 41  
Linde-Laursen I 76  
Linder HP 339  
Linton E 37, 39  
Lipari SE 98  
Lipman DJ 301  
Lisenby JD 220  
Litt A 346  
Litt M 149  
Litt T 98  
Liu B 41  
Liu N 40  
Liu W 172  
Liu YG 224  
Lively CM 213  
Llaca V 37, 39  
Lloyd DG 219  
Londo JP 37  
Long JA 219  
Long M 37  
Long Y 224  
Longly B 300  
Lonie JS 344  
Loockerman D 172  
Lootens P 72  
Lopes CR 128, 129, 132  
Lopez J 73  
Lopez R 299  
Los Gavilanes 128  
Lovett Doust J 219  
Lowe A 75  
Lowe AJ 75, 98  
Loyd-Jones M 11  
Lu AM 249, 250  
Lu BR 36  
Lu J 131, 224  
Lu T 41  
Lu Y 41  
Lucas DE 345  
Lucchesi JC 219  
Luckinbill LS 220

- Lumaret R 72  
Luo D 224  
Luo J 40  
Luo M 35, 40  
Luty JA 149  
Lydiate DJ 37  
Lynch M 36, 342  
Lyons EE 219  
Lysak MA 223  
  
Maurice M 216  
Ma J 11, 37, 39  
Macas J 217, 219  
Macdonald B 338  
Machray GC 150  
Maclean JL 37  
Macphail MK 344  
Maddison DR 300  
Maddison WP 301  
Madsen SF 99  
Maene L 302  
Maere S 362  
Magallon S 249  
Maggia L 96  
Maggini F 11-13, 71  
Magni CR 98  
Magri D 98  
Maguire TL 341  
Main AR 342  
Makkouk K 150  
Malhotra RS 150  
Mallikarjuna N 131  
Manchester SR 98  
Manganaris AG 73  
Manicacci D 219  
Manos PS 75, 98, 99  
Marais G 215, 221  
Mardis E 223  
Marie D 224  
Marienfeld JR 219  
  
Mariette S 75, 77  
Marin I 220  
Markgraf-Dannenberg I 12  
Markova M 220, 225  
Marques CM 342  
Marquez-Garcia MI 338  
Marshall D 345  
Martin C 11  
Martin DP 344  
Martin HA 340, 342  
Martin JP 99  
Martins PDC 213  
Martins WS 132  
Matsumoto T 39  
Matsunaga S 218, 220, 221, 224  
Matthew A 363  
Matthew PM 363  
Matyas G 99  
Maupin R 223  
Maurice S 215, 216, 220  
Maxted N 149  
Mayer E 75  
Mayer K 38  
Mayer KF 39  
McArthur ED 220  
McCauley DE 213, 220, 221, 224  
McComb JA 343, 346  
McCouch SR 40, 41  
McCoy F 342  
McCubbin A 300  
McDonald MW 337, 342  
McFadden G 346  
McGowen MH 342  
McKay JK 342  
McKinnon GE 341-343, 345, 346  
McLachlan JS 99  
Mcneill J 220  
McNicol P 131  
McPherson J 223

- Meagher TR 12, 13, 220  
Medagli P 72  
Medford JI 219  
Mehan VK 131  
Mehra PN 363  
Melake-Berhan A 39  
Meller VH 219  
Mendes AJT 131  
Meng R 129  
Mengesha MH 133  
Merila J 343  
Merlo F 301  
Mertens M 298  
Messing J 36, 38, 39, 41  
Mewes HW 38  
Meyerowitz EM 215  
Meyers BC 38  
Michaelson M 13  
Michaud H 72  
Michaud M 96  
Michelmore RW 38  
Michimoto T 220  
Michishita A 301  
Michu E 217  
Micishita A 300  
Mignot A 220  
Milgroom MG 98  
Milla SR 132, 134  
Millar MA 343  
Miller G 98  
Miller J 172  
Miller JS 220  
Millie D 172  
Mine M 300  
Minelli S 11-13  
Miner T 223  
Minuto L 216  
Minx PJ 223  
Mishler BD 247  
Mishra HO 151  
Misra R 149  
Mitchell SE 128, 130, 132  
Miyajima I 300-302  
Mizuno H 38  
Mlinarec J 75  
Moan EI 219  
Mogensen HL 220  
Mohanty A 99  
Mol J 223  
Molecular markers 96  
Moneger F 213, 215, 216, 221, 225  
Moner R 72  
Moore G 38  
Moore GA 72  
Moore JE 132  
Moore K 130  
Moore RC 219, 221  
Moores TL 38  
Moran GF 337-339, 343, 344, 346  
Moreau F 75  
Morel P 214  
Moretzshon MC 132  
Morgan D 247  
Morgan MT 214  
Morgante M 38, 150  
Morimoto Y 98  
Morley B 249  
Morris AB 99  
Morris WF 221  
Mort ME 250  
Mosandl A 216  
Moscariello P 13  
Moscone EA 133  
Moss JP 131, 133  
Motchoulskaia N 39  
Mottley J 78  
Mouchiroud D 221  
Mouras A 213, 215, 216, 223  
Moussian B 216

- Moyle LC 221  
Muehlbauer FJ 149, 150  
Mueller C 35  
Mueller-Starck G 75, 78  
Muhlemann JK 221  
Muir G 75  
Mukai Y 132  
Mulcahy D 225  
Mulcahy DL 214  
Muller J 249, 343  
Muller-Landau HC 219  
Muller-Starck G 99  
Mullet JE 37  
Munro RC 99  
Muona O 343  
Murai N 218  
Muratovic E 337  
Murphy G 38  
Murphy GP 38  
Murphy JP 128, 134  
Murray BG 362, 363  
Murrell JC 339, 346  
Murty UR 130  
Myburg AA 343  
Myveci K and Izgin N 148
- Nagaki K 38  
Nagamura Y 36  
Nagi W 13  
Nakajima Y 39  
Nakao S 220, 221  
Nakhoul H 148  
Namiki N 38  
Nash W 223  
Nasir M 150  
Nason J 345  
Natali L 11, 13  
Nathan R 219  
Nauman CH 10
- Naumann A 13  
Navarro A 217  
Nayar NM 38  
Negrutiu I 213-216, 219, 221, 223-225  
Negrutiu N 216  
Nei M 248  
Neilson RP 77  
Nelson G 341, 342  
Nepokroeff M 346  
Nesbitt KA 343  
Nevo E 12-14, 74, 343  
Newman M 129, 301  
Newnham MR 341  
Nguyen C 223  
Ni J 40  
Nickrent DL 250  
Nicolas M 221, 225  
Nicolle D 343, 345  
Niedenzu F 249  
Nigam SN 130  
Nigtevecht van G 221  
Nikles G 345  
Niles KR 221  
Nishimura S 300  
Nitta N 218  
Nixon KC 72, 75, 98, 250  
Nóbile PM 132  
Nobuta K 41  
Noel L 38  
Noller GL 220  
Nomura K 98  
Nordborg M 221  
Norton JD 97
- Obara M 221  
O'Brian G 219  
Ochieng JW 343  
Ocumphaugh WR 129  
Ogihara Y 132

- Oh CS 41  
Ohmido N 40  
Ohno S 221  
Ohri D 75, 300  
Ohta S 218  
Ohtsubo E 40  
Ohtsubo H 40  
Oke JG 131  
Okonkwo C 301  
Olalde M 99  
Oliveira M 216  
Oliver D 249  
Oliver RP 73  
Olivieri AM 150  
Olmstead RG 171, 172, 247  
Olsen KM 345  
Olsewska MJ 13  
Olson MS 217, 220, 221, 223, 224  
Olson T 39  
Olszewska MJ 76  
Orgel LE 13  
Origin, 148  
Orr AH 76  
Ortiz AM 131, 132  
Osborn TC 38  
Osiecka R 13, 76  
Otero A 337  
Otto KG 71  
Otto SP 38  
Ouassou A 337  
Ouyang S 38, 41  
Overpeck JT 97  
Oxelman B 221, 222  
Oyama K 73, 77  
Ozaki T 98  
Ozaki Y 301  
Ozersky P 223  
  
Page DC 218  
Paik-Ro OG 132  
  
Paillet FL 98  
Paliwal KK 150  
Palme A 99  
Palme AE 98  
Palmer JD 172  
Palser B 302  
Pan KY 248, 249  
Pan R 39  
Pan W-H 217  
Pan X 40  
Panaud O 78  
Panero JL 364  
Pang ECK 149  
Pannell JR 221  
Panning B 219  
Papeš D 75  
Papes D 78  
Papini A 71  
Parker JE 38  
Parker JS 222  
Parks CR 99  
Parniske M 38  
Parra-O C 337, 343  
Parrish TL 338  
Parry DC 133  
Parton E 302  
Parvery C 72  
Paterson AH 35, 38, 128  
Patil P 214  
Patterson B 344  
Patthy L 38  
Paule L 73, 97, 98  
Pautou G 172  
Payne JA 98  
Pearce SR 36  
Pearson NM 218  
Pearson WR 301  
Pedersen C 76  
Pellegrino RM 11

- Peloquin S 299  
Pemonge M-H 73, 99, 337  
Peñaloza APS 132  
Peng J 38  
Peng K 40  
Pennings PS 215  
Pepin K 223  
Pepper AE 299  
Pérez de la Vega M 148  
Pérez RB 133  
Perkin-Elmer 301  
Pernes J 76  
Peters GB 213, 344  
Peters-Robinson I 338, 341  
Peterson DG 38  
Petit RJ 70, 73-76, 96, 98, 99, 171, 344  
Petty A 346  
Pfaff T 151  
Pfrender M 342  
Philipp M 222  
Phillips TD 134  
Phitos D 215  
Pickersgill B 131, 301  
Pickett FB 36  
Piegu B 38  
Pietrarelli JR 129  
Pigliucci M 99  
Pilbeam DJ 344  
Pineau E 98, 344  
Pires C 346  
Pires JC 38  
Pitterová K 130  
Pittman RN 129, 130  
Pizarro EA 132  
Plomion C 76  
Poethig RS 213  
Pohl T 38  
Poke FS 344  
Polans NO 364  
Pole MS 344  
Polzin K 39  
Ponger L 38  
Pontaroli AC 35  
Popescu F 99  
Popp M 221, 222  
Porter HC 13  
Porter JM 172  
Porth I 76, 77  
Posada D 249  
Postlethwait J 36  
Pot D 76  
Potts BM 336, 339-346  
Pound LM 344  
Powell AM 364  
Powell M 364  
Powell W 131, 149, 150, 345  
Pozzobon MT 132  
Pradella M 97  
Pradere S 71  
Prakash CS 129  
Prather LA 172  
Precerámico peruano 128  
Preil W 301  
Prentice HC 222  
Presting G 36  
Preszler RW 74  
Price 13  
Price HJ 13  
Prince LM 250  
Prine GM 129, 132  
Prober S 341  
Proite K 132  
Proner V 72  
Provan J 150  
Pryor LD 344  
Purugganan M 36  
Pyntikova T 223  
Qian H 99  
Qiu YL 247

- Qu LH 249  
Quader M 151  
Quattro J 298  
Queiroz-Voltan RB 134
- Rabeler RK 221  
Rae SJ 297  
Raes J 39  
Raghavan V 301  
Raghuvanshi S 40  
Raina SN 132  
Rallo L 298  
Ramaiah CD 133  
Ramakrishna W 35-37, 39  
Raman VS 132  
Rambaut A 172  
Ramgiry SR 150  
Rani V 132  
Rao NGP 130  
Rapoport H 298  
Rau TPS 133  
Raven PH 364  
Rawlings M 342  
Rayburn AL 13, 76  
Razafimaharo V 338  
Razaq ZA 364  
Reboud X 222  
Reed DH 344  
Rees H 12, 76  
Rehman F 225  
Reid JB 343, 344  
Reijans M 302  
Reille M 71  
Reinsch A 249  
Remanandan P 363  
Remington D 72  
Remotti D 301  
Remy RJ 97, 98  
Rendell S 99  
Renfrew JM 149
- Renner SS 222, 225  
Rensink WA 38  
Repping S 223  
Rezende GDSP 341  
Rice RW 129  
Rice WR 222  
Richards AJ 222  
Richards DE 38  
Ricke D 36, 339  
Ricklefs R 97  
Ricklefs RE 99, 222  
Rigoni VA 130  
Riha K 214  
Rincón A 132  
Rincón CA 133  
Rizzon C 38  
Robacker CD 301  
Roberts AV 78  
Robertson LD 149-151  
Robertson SE 222  
Robins T 11  
Robledo G 133  
Rock S 223  
Roessli D 301  
Rogozin IB 37  
Rohde A 222  
Rohlfing T 223  
Rohwer JG 248  
Roldan-Ruiz I 72  
Romagosa I 301  
Romero Lopes C 128, 130  
Rong J 35  
Roscini C 11  
Rosenthal A 222  
Rossetto M 344  
Rothn C 72  
Rounseley S 36  
Roure JM 98  
Rouse J 301, 302  
Roush RD 132

- Roussel G 76, 78  
Rowland R 302  
Rubeena Ford R 150  
Rubes J 217  
Ruddat M 222  
Rueffler C 215  
Ruiz Rejon C 217  
Ruiz Rejon M 217  
Rushton BS 76  
Russell J 345  
Russell K 71  
Russell-Smith J 345  
Rye BL 345  
  
Saar DE 364  
Sadler H 217  
Saedler H 216, 217  
Saeys Y 39  
Safar J 217  
Sahai JP 150  
Saintagne C 76  
Saji S 38  
Sakai A 220, 221  
Sakai AK 225  
Sakai F 223  
Sakata K 39  
Salse J 38  
Salvini D 99, 344  
Sampson JF 345  
Samyn G 298  
Sanderson SC 220  
Sang T 36, 172  
Sangam L 150  
Sankoff D 73  
SanMiguel P 35, 36, 39, 77  
SanMiguel PJ 36  
Santini S 13  
Sapienza C 12  
Sarker A 150, 151  
Sarr A 78  
  
Sarri V 11  
Sarwar DM 151  
Sasaki T 36, 39  
Sassoli O 11, 13  
Sastri DC 131  
Sato H 74  
Sato S 218  
Saucier JR 99  
Saur Jacobs M 222  
Saur MJ 224  
Savolainen V 250  
Savva M 344  
Sawhney V 302  
Saxena MC 149-151  
Saxena NP 149  
Sbay H 337  
Scarascia-Mugnozza GT 11  
Scariot V 298, 301  
Schaal BA 37, 78, 172, 345  
Schat H 213, 222  
Scheiber SM 301  
Schemske DW 342  
Schiebel K 77  
Schieftaler U 222  
Schiestl FP 221  
Schirone B 71, 72, 77  
Schlagenhauf E 40  
Schlotterer C 75  
Schmid J 219  
Schmidt R 39  
Schnable PS 222  
Schneeberger R 13  
Schneider S 301  
Schnitzler J-P 77  
Schubert I 12  
Schulman AH 12, 14, 74  
Schulze SR 128  
Schutz S 216  
Schwaegerle KE 218  
Schwarz O 77

- Schwarz-Sommer Z 224  
Schwarzacher T 72, 78  
Schwemmer SS 10  
Scott K 223  
Scott LJ 341  
Scott-Moncrieff R 363  
Scotti-Saintagne C 76, 77  
Scutt CP 222, 223  
Sebastiani F 97  
Sederoff RR 73, 342, 343  
Sedgley M 339, 344, 345  
Seifert K 216  
Seijo JG 131, 133  
Selfish genes 12  
Sen S 148  
Senter SD 98  
Sequence 39  
Shah-Mahoney N 219  
Sharma H 301  
Sharma SK 151  
Sharp PM 40  
Shaw J 172  
Shepherd M 340, 341, 343, 345  
Sherff EE 364  
Sherman NA 220  
Shi J 40  
Shi S 250  
Shi SH 249  
Shi X 40  
Shibata M 41  
Shih MC 39  
Shiloff BA 36  
Shimada T 98  
Shiu SH 39  
Shivanna K 302  
Shoemaker RC 39  
Shull GH 223  
Shykoff JA 215, 223  
Sieber P 222  
Siegmund HR 74  
Siljak-Yakovlev S 72, 78, 337, 362  
Simeone MC 71  
Simillion C 39, 40  
Simmons MP 225  
Simpson CE 128, 130, 133, 134  
Sindhu JS 151  
Singh AK 133  
Singh G 345  
Singh J 150  
Singh K 148  
Singh KB 150  
Singh KN 128  
Singh KP 132  
Singh M 151  
Siripun KC 172  
Siroky J 214, 216-218, 221, 223, 224  
Sivaramakrishnan S 133  
Skaletsky H 223  
Skinner AK 337  
Slade D 99  
Slee AV 345  
Slinkard AE 150, 151  
Smartt JP 130, 133  
Smith DS 338  
Smith JB 71  
Smith NG 213  
Smith RA 220  
Smith RL 132  
Smith RM 345  
Smith S 345  
Smouse P 298  
Sniegowski P 12, 36, 214  
Sodzi N 171  
Solbrig OT 364  
Soltis DE 99, 171, 247, 248, 249  
Soltis PS 99, 249, 345  
Song R 39  
Soppe W 12

- Sørensen PD 363, 364  
Sork VL 73, 97, 345  
Sorrells ME 40  
Souer E 223  
Souza-Chies T 72, 362  
Spada F 77  
Sparrow AM 10, 13  
Specht J 39  
Spellenberg R 77  
Spethmann W 301  
Spitze K 342  
Spongberg SA 96, 247  
Srinivasan S 39  
Srivastava SP 151  
Stabentheiner E 71  
Stalker HT 128-134  
Stanford A 99  
Strange I 298  
Starr JL 128  
Steane D 338, 339  
Steane DA 340, 342-346  
Stebbins CD 134  
Stebbins G 301  
Stebbins GL 77, 364  
Steele KP 75  
Steffenson B 37  
Steigner D 77  
Stein N 40  
Steinbrecher R 77  
Steinemann M 223  
Steinemann S 223  
Steinhoff S 77  
Steinkellner H 74, 77  
Stephan W 12, 36, 214  
Stephens SG 39  
Stettler RF 71  
Stewart S 71  
Stilio VS 225  
Stögerer K 96  
Stokoe RL 341, 345  
Stokstad E 39  
Storchova H 223  
Storz JF 346  
Strehl S 73  
Streiff R 77  
Strk J 339  
Strother JL 364  
Stuessy TF 172  
Stukely MJC 343  
Su D 224  
Sugiyama R 212, 224  
Summerfield RJ 148  
Sundaresan V 225  
Sundberg S 364  
Suoniemi A 14, 78  
Suzuki M 99  
Suzuki S 218  
Svachulová J 130  
Svoboda J 217  
Sweere JA 250  
Swensen SM 250  
Swigonova Z 36, 37, 39  
Swofford DL 172, 250, 301  
Systsma KJ 346  
Tabata S 218  
Tabbener H 99  
Taberlet P 73, 96, 172  
Tadesse N 148  
Tahir M 149  
Takada S 223  
Takakura Y 218  
Takano H 220  
Takayanagi K 300  
Takebayashi N 225  
Takhtajan A 250  
Talbert PB 38  
Tallury SP 134

- Tamura K 248  
Tanaka T 99  
Tanksley SD 40  
Tanskanen J 12, 14, 74  
Tantau I 98  
Tarayre M 224  
Tarchini R 38, 337  
Tasaka M 212, 223  
Taurchini D 96, 99  
Taylor DR 220, 221, 223, 224  
Taylor DW 250  
Taylor JS 39  
Taylor PWJ 149, 150  
Tenbookum WM 222  
Testolin R 224  
Thakur JK 37  
Thamarus KA 346  
Thompson JD 224  
Thornton K 37  
Thrall PH 224  
Tian S 41  
Tibbits WN 338  
Tidwell WD 99  
Tiffin P 221  
Tiffney 99  
Tiffney BH 99, 250  
Tikhonov A 35, 39, 77  
Tilyard PA 342  
Tin-Wollam A 223  
Tingey SV 38  
Tinner W 97  
Tohme J 73  
Tomar SK 150  
Tong K 250  
Torriani D 97  
Touzet P 214  
Tovar-Sanchez E 77  
Trehin C 214  
Trimble S 224  
Tsai SM 128  
Tsumura Y 300  
Tucker JM 77  
Tufail M 151  
Turcotte K 39  
Turetschek E 77  
Turkova V 130  
Turnbull JW 337  
Turner BL 364  
Turpeinen T 13  
Tuskan GA 39  
Tutin TG 12  
Tyagi AK 37, 40  
Tyler B 11  
Tzeng YH 39  
Uchida H 220  
Uchida W 220, 224  
Udovicic F 297, 337, 341-343, 346  
Udupa SM 149, 151  
Unfried K 77  
Unseld M 219  
Uozu S 40  
Uren T 338  
Ureshino K 300-302  
Vahidy AA 364  
Vaillancourt RE 151, 336, 341-346  
Väinölä A 302  
Vainola R 73  
Valencia S 73  
Valles J 77  
Valls JFM 128-130, 132, 134  
Van Bockstaele E 72, 298  
Van Creij M 302  
Van Dam B 98  
Van Damme JM 217  
Van de Peer Y 39, 40  
van Der Biezen EA 38

- van der Knaap WO 98  
van der Linden CG 346  
Van Dien M 302  
van Dodewaerd AM 40  
van Houwelingen A 223  
Van Huylenbroeck J 298  
Van Kleinwee T 302  
Van Laere K 298  
Van Montagu MC 39, 222  
Van Slageren M 149  
Van Tienderen PH 346  
Van Tuyl J 302  
Vandepoele K 39, 40  
Vanoverschelde M 302  
Van't Hof J 13, 14  
VanValen L 78  
Varshney RK 40  
Vassiliadis C 13, 220  
Veiga RFA 134  
Veitia RA 40  
Vello NA 128  
Vendramin GG 96-99, 344  
Venkateshwar A 133  
Verdu M 217  
Verhaegen D 338  
Vershinin AV 78  
Vervaeke I 302  
Veuskens J 215, 216, 224  
Vicient CM 14, 78  
Vigneron P 338  
Vij S 40  
Villani F 96, 99  
Villar M 71  
Vinogradov AE 14  
Vision TJ 40  
Vitte C 216  
Volckaert G 38  
Vooijs R 213, 222  
Vos P 151  
Vosman B 298, 346  
Voss D 302  
Voytas DF 74  
Vrebalov J 37  
Vydianathan R 40  
Vyskot B 214, 216-221, 223-225  
Wade MJ 222  
Waelti MO 221  
Wakhlu AL 363  
Walker J 346  
Wallwork MAB 344  
Walter K 297  
Wambutt R 38  
Wang J 41, 346  
Wang K 98  
Wang R 36  
Wang S 36, 40  
Wang T 130  
Wang W 35, 37, 40  
Wang X 40  
Wang XQ 247  
Wang Z 38, 224  
Wardell-Johnson G 345  
Ware D 40  
Wargachuk R 214  
Warmke HE 224  
Watson BD 71  
Watts WA 99  
Webb C 151  
Webb III T 97  
Webb RS 97  
Weber JL 151  
Wee JL 172  
Weeden NF 73, 134; 150  
Wei HT 247, 248  
Weigand F 151  
Weir I 224

- Weising K 149  
Weissbecker B 216  
Welch ME 224  
Weller SG 225  
Wen J 99, 249, 250  
Wendel JF 40, 70, 346  
Wessler SR 14, 37  
West AK 343  
West JG 345  
Westerbergh A 225  
Westergaard M 225  
Weston PH 339  
Wheeler M 346  
Wherry ET 172  
Whetten RW 343  
White MJD 78  
White SE 14  
Whitehead A 346  
Whittemore AT 78  
Whittock SP 346  
Whitton J 38  
Wicker T 40  
Widmer A 213-215, 221  
Wielgorskaya T 78  
Wilbur RL 172  
Williams DE 130  
Williams E 300, 302  
Williams ER 344  
Williams J 74  
Williams PC 148  
Willis ME 222  
Wilson WA 40  
Wiltshire RJE 342, 344  
Winder CT 172  
Wing RA 36, 40  
Winge O 225  
Winter P 148, 149, 151  
Wise RP 222  
Wisman E 222  
Witcombe, JR 148  
Witt T 218  
Wolf DE 215, 225  
Wolf L 339  
Wolf YA 37  
Wolf YI 37  
Wolfe JA 99, 100  
Wolfe KH 35, 40  
Wolfe LM 216  
Wollenberg K 336  
Wong C 151  
Wong GK 41  
Woo SS 36  
Woodman WL 40  
Wright S 346  
Wu CI 78, 250  
Wu H 224  
Wu J 38, 41  
Wullstein LH 77  
Wynne JC 129, 130, 133  
Xing JY 247  
Xu J 225  
Xu J-H 41  
Xu Y 41  
Xue Y 36  
Yadav CR 151  
Yakovlev I 73  
Yamagata H 41  
Yamamoto K 39  
Yamamoto T 99  
Yamazaki T 302  
Yan H 41  
Yan L 36  
Yan YL 36  
Yang TJ 41  
Yang WC 225  
Yang YW 40  
Yang-JunBo 299

- Yao J 302  
Yates CJ 338  
Yau SK 151  
Ye D 217, 225  
Yokoka K 78  
Yoshida M 98  
Yoshimura K 300  
Young AG 363  
Young ND 134  
Young S 35  
Youngman VE 152  
Yu J 41, 346  
Yu Y 38, 40, 41  
Yuan YM 218  
Yukawa T 300, 302
- Zabeau M 39  
Zakharov D 39  
Zanetto A 75, 78  
Zanis M 250  
Zanis MJ 249  
Zenkteler M 302  
Zeyl C 222  
Zhang C-S 364  
Zhang H 36  
Zhang HD 249  
Zhang J 214
- Zhang L 41  
Zhang LB 225  
Zhang Q 40, 41, 224, 249  
Zhang X 37  
Zhang Y 36, 41  
Zhang YH 225  
Zhang ZY 250  
Zhang-ChangQin 299  
Zhelev P 73  
Zheng H 40  
Zhong Y 224  
Zhou J 41  
Zhou Y 98  
Zhou, Z 100  
Zhou Z-K 72, 75  
Zhou ZK 98  
Ziegenhagen B 74, 99  
Zimmer EA 250  
Zimmer I 77  
Zjhra ML 346  
Zluvova J 217, 220, 225  
Zohary D 152  
Zoldos V 78  
Zolfino C 11  
Zou Y 224  
Zuckerkandl E 14  
Zurawski G 152

# **Detailed Contents**

---

<i>Preface to the Series</i>	v
<i>Preface to this Volume</i>	vii
<i>List of Contributors</i>	xi
<b>1. Repetitive DNA and Plant Evolution</b>	1
Introduction	2
Other Plant Species	7
Conclusion	8
References	10
<b>2. Evolution and Phylogenetic Relationship of the Rice Genome</b>	15
Introduction	16
Features of the Rice Genome	18
Centromeres and Telomeres	19
Organelle Genome Insertions	20
Gene Density	20
Duplications	21
Comparison of Rice Genomes	21
The Impact of Duplication on Evolution of the Rice Genome	22
Role of Transposable Elements in the Process of Evolution	24
Evolution of Gene Families	25
Comparison of Rice with Cereal Genomes	27
Comparison of Model Plant Genomes - Rice versus <i>Arabidopsis</i>	30
Conclusion	33
Acknowledgments	35
References	35

<b>3. Genome Organization and Evolution in Genus <i>Quercus</i> (Fagaceae): Special Attention to Two European White Oaks <i>Quercus petraea</i> (Matt.) Liebl. and <i>Q. robur</i> L.</b>	<b>43</b>
Introduction	44
Systematics, Phylogeny and History of the Genus <i>Quercus</i>	45
Ecology of Oaks	47
Interspecific Hybridization in Oaks: An Example of European White Oaks <i>Q. petraea</i> and <i>Q. robur</i>	49
Karyotype Evolution in Family Fagaceae (and in Genus <i>Quercus</i> in Particular): Chromosome Number, Karyotypic Features and Genome Size	54
Molecular Organization of the Oak Genomes	62
Genome/Karyotype Evolution and Speciation in <i>Quercus</i>	67
Conclusion	69
Acknowledgment	70
References	70
<b>4. Biodiversity and Evolution of <i>Castanea</i></b>	<b>79</b>
Phylogeography of American <i>Castanea</i> Species	88
References	96
<b>5. Cytogenetic and Molecular Evidences on the Evolutionary Relationships Among <i>Arachis</i> Species</b>	<b>101</b>
Introduction	102
Species Relationships within <i>Arachis</i> Section	104
Genetic Origin of <i>A. hypogaea</i>	108
Single vs. Multiple Origin	109
Wild Progenitors	112
Is <i>A. monticola</i> the Wild Tetraploid Ancestor of <i>A. hypogaea</i> ?	115
Origin of x=9 Basic Chromosome Number	116
Origin of Polyploids in Rhizomatosa Section	122
Origin of the Triploid Cytotype of <i>A. Pintoi</i> (Section Caulorrhizae)	125
Conclusion	126
Acknowledgment	128
References	128

<b>6. Biodiversity, Genetic Enhancement and Molecular Approaches in Lentil</b>	<b>135</b>
Introduction	136
Biodiversity and Evolution in Lentil	136
Variability in Key Traits	137
Major Accomplishments in Genetic Enhancement	139
a. Resistance to Biotic Factors	139
b. Improving Tolerance to Abiotic Stresses	140
c. Construction of Plant Type Suitable for Machine Harvest	140
d. Genetic Base of Lentil in South Asia has been Widened	141
e. Varietal Development	141
f. Development of Iron and Zinc-rich Lentil Varieties	141
Genome Evolution in Lentils	142
Molecular Approaches Used for Lentil	144
Acknowledgment	146
References	148
<b>7. Phylogenetic Relationships of the Genus <i>Phlox</i> (Polemoniaceae): Comparing and Combining Data Sets from the Chloroplast Genome</b>	<b>153</b>
Introduction	154
Materials and Methods	156
Taxon Sampling, DNA Extraction	156
Restriction Site Data	156
Sequence Data	156
Phylogenetic Analyses	162
Results	163
Conclusion	165
Acknowledgment	171
References	171
<b>8. Genus <i>Silene</i> (Caryophyllaceae) – Evolutionary Diversification and Sex Chromosome Formation</b>	<b>173</b>
Introduction	174
<i>Silene</i> – An Emerging Model System for Plant Ecology, Evolution and Development	176

Reproductive Systems within the Genus <i>Silene</i>	178
Overview of Reproductive Systems in Plants	178
Reproductive Systems in the Genus <i>Silene</i>	183
<i>Silene latifolia</i> as a Model Plant	187
<i>Silene</i> Genomics	197
General Information	197
<i>Silene latifolia</i> as a Source of Anther-Specifically Expressed Genes	198
<i>Silene latifolia</i> – Model for Sex Chromosome Evolution	198
Perspectives	211
Acknowledgment	212
References	212
<b>9. Molecular Phylogenetics of Hamamelidaceae: Evidence from DNA Sequences of Nuclear and Chloroplast Genomes</b>	<b>227</b>
Materials and Methods	230
Results	235
Discussion	244
Relationships of Saxifragales and the Monophyly of Hamamelidaceae	244
Relationships Within Hamamelidaceae	245
Acknowledgment	247
References	247
<b>10. The Potential of Molecular Analysis and Interspecific Hybridization for Azalea Phylogenetic Research</b>	<b>251</b>
Introduction	252
The Taxonomical Position of Azaleas within the Genus <i>Rhododendron</i>	253
Taxonomy	253
Plant Description	253
The Genus <i>Rhododendron</i>	254
Interspecific Hybridization in <i>Rhododendron</i> :	
Breeding as a Tool to Delineate Species?	257
“Capability to Cross” as a Criterion for Speciation	257

Intergeneric Hybridization between <i>Rhododendron</i> and <i>Menziesia</i> (Ericaceae)	267
Molecular Techniques for Phylogenetic and Kinship Research	269
Phylogenetic Analysis Based on <i>MatK</i> Sequencing for Different Species from the <i>Tsutsusi</i> Subgenus	273
Characterization of Natural Hybrids between <i>R. kiusianum</i> and <i>R. kaempferi</i> by RFLP-PCR	277
The Usefulness of AFLP Markers for the Establishment of Phylogenetic Relationships in Evergreen Azaleas	277
Some Applied AFLP Studies for Characterization of Genetic Resources	281
Candidate Genes as Functional Markers for the Assessment of Genetic Diversity and Mapping of Plant Quality Traits	290
Development of EST Markers	290
Comparison of AFLP, STMS and EST Markers for the Assessment of Genetic Diversity	291
Exploration of Segregation of Plant Quality Traits and the Construction of Genetic Maps	293
Acknowledgment	295
References	297
<b>11. Phylogeny, Diversity and Evolution of Eucalypts</b>	<b>303</b>
Introduction	304
Genome Structure	305
Evolutionary History and Phylogeny	305
Classification and Distribution	305
Morphology	306
Molecular Phylogeny	309
Evolutionary History	312
Species and Speciation	315
Species Recognition	315
Hybridization	319
Evolutionary Processes	323
Genetic Diversity	323
Selection and Drift	329

Breeding System	330
Conclusion	335
Acknowledgment	336
References	336
<b>12. Dahlia: Cytogenetics and Evolution</b>	<b>347</b>
Introduction	348
Brief Taxonomic History of Dahlia	348
Early Chromosome Studies	349
Karyotype Studies	349
Molecular Systematic Studies	352
Pachytene Pairing, Synaptonemal Complexes and Secondary Associations At Metaphase I	356
Are Dahlias Diploids or Ancient Polyploids?	357
Conclusion	362
Acknowledgment	362
References	362
<i>Author Index</i>	365
<i>Detailed Contents</i>	393