University of New Hampshire
University of New Hampshire Scholars' Repository

Doctoral Dissertations Student Scholarship

Spring 2003

Nuclear and cytoplasmic genetic variation in Picea: DNA markers for evaluating past migration, introgression and evolutionary history

Joselle Germano-Presby
University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation
Germano-Presby, Joselle, "Nuclear and cytoplasmic genetic variation in Picea: DNA markers for evaluating past migration, introgression and evolutionary history" (2003). *Doctoral Dissertations*. 125.
https://scholars.unh.edu/dissertation/125

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

NUCLEAR AND CYTOPLASMIC GENETIC VARIATION IN *PICEA*: DNA MARKERS FOR EVALUATING PAST MIGRATION, INTROGRESSION AND EVOLUTIONARY HISTORY

BY

Joselle Germano-Presby

B.S. Eckerd College, 1995

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in

Biochemistry

May, 2003

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3083730

UMI®

UMI Microform 3083730

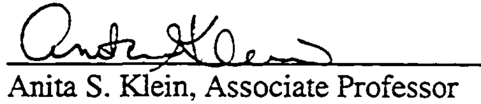
Copyright 2003 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Ph.D. DISSERTATION

This dissertation has been examined and approved.

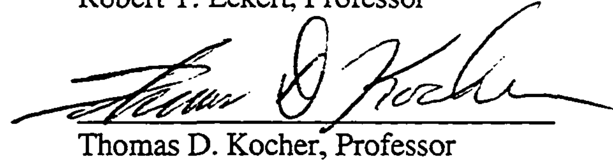


Anita S. Klein, Associate Professor

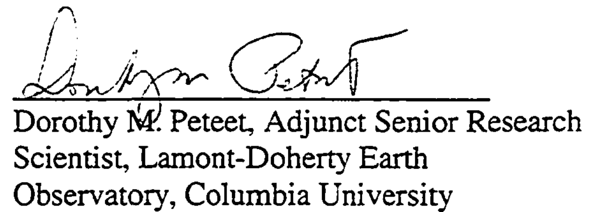
John J. Collins, Associate Professor

Richard H. Cote, Professor

Robert T. Eckert, Professor



Thomas D. Kocher, Professor



Dorothy M. Peteet, Adjunct Senior Research Scientist, Lamont-Doherty Earth Observatory, Columbia University

4·18·2003

Date

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ACKNOWLEDGEMENTS

For over seven years have I worked on the material presented in this dissertation. Those were some of the best years of my life. Not because the work was as fulfilling as it was, but because of all of the people with whom I worked and who supported me along the way. First and foremost, I must acknowledge my graduate advisor Anita Klein. Anita contacted me directly before my application to the program was even accepted to tell me about an exciting research opportunity in her lab. Since that day she has been a helpful resource, a knowledgeable advisor and a good friend. She has continually supported me; her high expectations have constantly challenged me. I have learned a tremendous amount from her and the content in this dissertation reflects that.

No graduate program is complete without one's graduate comrades. Lab-mates not only become some of your best friends, but they are your co-workers there to help you with a new protocol or aid you on a field trip, there to participate in intellectually stimulating conversations and keep your logic in check, there to share your troubled times and to celebrate your triumphs. They become your foundation, your circle; you depend on them. Danielle Cain, Drew West, Aaron Wallace, Brian Teasdale and Chery Whipple are among some of my closest comrades. Graduate life would not have been the same without you guys, you have helped and supported me tremendously. Also along the way were Jianhua Li, Helene Balkin, Heather Taylor, Chris Clouser, Aaron Thorner, Queta Boese, John Wool, Brad Anderson, Angie Norton and Karen Reed. Thank you for your help and your friendship.

Nothing is possible without family. Or at least I depend on mine so much that it seems like I couldn't accomplish anything without them behind me. No matter what you

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

do, your family backs you up, encourages you, and nourishes you. They are primarily interested in *your* well-being, what is best for *you*. I have thanked my Mom, Dad and sister Aleece countless times for always being there for me; I can't overemphasize my appreciation for them here. I am who I am because of your love, thank you once again.

Close friends are also like family. My extended family includes Katherine Jenkins, Sybil Rowe, Sarah Hoeksema and Courtney Daum. Good friends are there to help you move, there to help you plan your wedding, and always have a spare bed or couch when you need a place to crash. Thank you for being there.

There is always one person in your life on whom you depend the most, with whom you spend the most time, who knows you in and out and loves you no matter what state you're in. That person will do anything for you, be it driving long hours to see you each week, spending a long weekend to help you build a piece of equipment that you need for work, or accompanying you on long trips to the back woods of Virginia to help you collect samples or to other far away places to support you while you interview for a job. Even after you've soaked their shirt with your tears, yelled so loudly their ears ring, blamed them for half of your troubles and neglected to call them while spending two weeks in Europe, they still love you, asking for nothing in return. That person is your best friend; my best friend is my husband Thad Presby. Thanks for standing by me all these years. This work would truly not have been possible without you as my source of strength. You are my life.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS........iii
LIST OF FIGURES........viii
LIST OF TABLES........ix
ABSTRACT........x
INTRODUCTION........1
I. SPECIES SPECIFIC NUCLEAR AND CHLOROPLAST SINGLE NUCLEOTIDE POLYMORPHISMS TO DISTINGUISH *PICEA GLAUCA*, *P. MARIANA* AND *P. RUBENS*........3
Abstract........3
Introduction........3
Materials and Methods........5
  Plant Materials........5
  PCR Methods........5
  Cloning........5
  DNA Sequencing........5
  Sequence Analysis........5
  Population Studies........5
Results........6
  Chloroplast Gene Sequences........6
  Screening Interspecific *trnK* SNPs for Species-Specificity........6
  Screening Interspecific *rpl33-psaJ-trnP* SNPs for Species-Specificity........9
  ITS Sequences........9
  Screening Interspecific ITS SNP 1 for Species-Specificity........11
  Sequences of Cloned ITS Fragments........11
  Screening Interspecific ITS SNP 7 for Species-Specificity........11
  Testing Species-Specific SNPs as a Tool to Identify Species from Single Needles........12
Discussion........12
  Levels of Sequence Variation in Spruce........12
  Heterogeneity of ITS........12
  Species-Specific Markers to Distinguish between Spruce Species........13
Acknowledgements........14
References (for Chapter I only)........14
II. GEOGRAPHIC DISTRIBUTION OF MITOCHONDRIAL AND CHLOROPLAST HAPLOTYPES AMONG RANGE-WIDE POPULATIONS OF WHITE SPRUCE (*PICEA GLAUCA*) BLACK SPRUCE (*P. MARIANA*) AND RED SPRUCE (*P. RUBENS*): IMPLICATIONS OF SPECIATION, INTERSPECIFIC HYBRIDIZATION AND HISTORY OF THE WISCONSIN ICE AGE........16
Abstract........16
Introduction........17
  Driving Forces in Forest Species Migration........17
  Genetic Structure and Biogeography........19
  Migration History of White, Black and Red Spruce in North America........20
  Interspecific Hybridization........21

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Variable DNA Markers....23
Materials and Methods....24
Plant Materials and DNA Extraction....24
PCR Methods....26
DNA Sequencing....27
Screening Individuals for Polymorphisms....33
Phylogenetic Analysis....38
Estimation of Sequence Divergence Rates and Substitution to Indel Ratios....39
Population Genetic Analyses....40
Geographical Correlations....40
Results....41
Summary of Polymorphisms Initially Detected....42
Screening Populations for Identified Markers and Defining Haplotypes....42
Relationships and Frequency of Haplotypes....43
TrnK, Divergence Rates among Spruce....46
Genetic Differentiation and Geographical Distribution of Haplotypes....47
Discussion....53
Between- and Within-Species Variation....53
Nature of Mutations in the Cytoplasmic Genomes....54
Genetic Diversity, Structure and Gene Flow within Each Species....55
Interspecific Hybridization and Introgression....56
Introgression of White Spruce and Black Spruce....56
Introgression of Red Spruce and Black Spruce....57
The Speciation of Red and Black Spruce....58
Geographic Distribution of Haplotypes....60
Distribution of the Red Spruce *Mt-3* Haplotype in Allopatric Black Spruce....62
Distribution of Chloroplast Haplotypes....63
Summary....67
Acknowledgements....67
III. CHLOROPLAST AND MITOCHONDRIAL DNA EVIDENCE FOR THE EVOLUTION OF *PICEA*....69
Abstract....69
Introduction....69
Taxonomy of *Picea*....69
Phylogeny and Biogeography of *Picea*....72
Speciation and Inter-Species Hybridization....76
Materials and Methods....78
Results....80
Chloroplast *TrnK* Analyses....80
Mitochondrial *Nad1 and Nad7* Analyses....84
Comparison of Mitochondrial *Nad1* Trees to Chloroplast *TrnK* Trees....90
Discussion....91
Comparison of *TrnK* and *Nad1* Trees to Results of Other Systematic Studies..92
Position of the *P. glauca* Clade....94
Positions of Individual Species: Incongruencies between Chloroplast and Mitochondrial Trees....95

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Future Studies....97
Acknowledgements....98
IV. SPECIES-IDENTIFICATION OF LATE-PLEISTOCENE AND EARLY-HOLOCENE *PICEA* MACROFOSSILS FROM THE NORTHEASTERN UNITED STATES: USING ANCIENT DNA TO IMPROVE VEGETATION HISTORY AND PALEOCLIMATE RECONSTRUCTION....99
Abstract....99
Introduction....100
Importance of *Picea* in North American Quaternary Paleoclimate Reconstructions....101
The Last Glacial Maximum....106
Quantifying Climate Conditions from Fossil Pollen Assemblages....109
Estimates of Full- and Late-Glacial Climate Conditions in Eastern North America....112
Morphological Methods to Distinguish Fossil *P. glauca*, *P. mariana* and *P. rubens*....113
Molecular Methods to Distinguish Fossil *P. glauca*, *P. mariana* and *P. rubens*....116
Project Goals: Species Identification of *Picea* Macrofossils from Browns Pond, Virginia....117
Materials and Methods....118
Core Sample Collections....118
Precautions for Fossil Manipulation....121
Separating Macrofossils and Pollen from Core Sections....122
DNA Extraction from Macrofossils....123
Southern Blot and Hybridization Analysis....125
PCR Amplification....126
Results and Discussion....127
Core Samples: Effects of Handling and Time since Collection from the Ground....127
Problems Encountered with DNA Extraction of Spruce Macrofossils....132
PCR Amplification....133
Future Directions....143
Acknowledgements....148
LIST OF REFERENCES....150
APPENDIX A....166
APPENDIX B....168
APPENDIX C....169
APPENDIX D....177

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

CHAPTER I

Figure 1. Gene regions showing the relative positions of coding regions (boxed areas) and PCR amplification and sequencing primers.............................. 6

Figure 2. Representative results from screening provenance test samples............... 10

CHAPTER II

Figure 1. Phylograms showing relationships of haplotypes.................................... 44

Figure 2A. Geographic locations of populations and their included chloroplast haplotypes coded according to the key shown.. 49

Figure 2B. Geographic locations of populations and their included mitochondrial haplotypes coded according to the key shown.. 50

CHAPTER III

Figure 1. Maximum parsimony phylogenetic tree from chloroplast *trnK* intron sequence.. 82

Figure 2. Minimum evolution phylogenetic tree from chloroplast *trnK* intron sequence.. 83

Figure 3. Maximum parsimony phylogenetic tree from mitochondrial *nad1* intron 2 sequence depicted as a rectangular phylogram (no outgroup was defined; A) and as an unrooted phylogram (B) 87

Figure 4. Minimum evolution phylogenetic tree from mitochondrial *nad1* intron 2 sequence depicted as a rectangular phylogram (no outgroup was defined; A) and as an unrooted phylogram (B) 88

Figure 5. Maximum parsimony strict consensus cladogram from mitochondrial *nad1* intron 2 sequence.. 89

CHAPTER IV

Figure 1. Current ranges of *Picea rubens* (A), *P. glauca* (B), and *P. mariana* (C) (from Little, Jr., 1971) 105

Figure 2. Agarose gel showing DNA extracts (CTAB method) from BR92D *Picea* macrofossils.. 129

Figure 3. Hybridization of Southern blot with *Picea*-specific probe.. 130

Figure 4. Agarose gel showing PCR products using primers ITS1-5.8S-1 and ITS1-5.8S-1A.. 137

Figure 5. Agarose gel showing PCR products using primers ITS1-5.8S-1 and ITS1-5.8S-1A.. 139

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF TABLES

CHAPTER I

Table 1. PCR amplification and sequencing primers 7
Table 2. PCR profiles 8
Table 3. Samples used for sequencing 9
Table 4. Results of population studies 10

CHAPTER II

Table 1. Northern and western black spruce samples obtained from provenance test plantations 25
Table 2. Primers used for PCR and sequencing 28
Table 3. Conditions of individual polymerase chain reactions 30
Table 4. Geographic origin of samples used for primary sequencing and associated GenBank accession numbers 32
Table 5. Assays used to score chloroplast polymorphisms and corresponding haplotype definitions 34
Table 6. Assays used to score mitochondrial polymorphisms and corresponding haplotype definitions 36
Table 7. Genetic diversity and differentiation 48
Table 8. Results of analyses of molecular variance (AMOVA) 49

CHAPTER III

Table 1. *Picea* species included in the phylogenetic analysis 70
Table 2. List of constraints used for the Wilcoxon signed-ranks test 81

CHAPTER IV

Table 1. Climatic and ecological preferences, tolerances and life history characteristics of the three *Picea* species native to eastern North America 103
Table 2. Browns Pond cores used in this project 120

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

NUCLEAR AND CYTOPLASMIC GENETIC VARIATION IN *PICEA*: DNA MARKERS FOR EVALUATING PAST MIGRATION, INTROGRESSION AND EVOLUTIONARY HISTORY

by

Joselle Germano-Presby

University of New Hampshire, May, 2003

Mitochondrial and chloroplast haplotypes were identified in range-wide populations of white (*Picea glauca*), black (*P. mariana*) and red spruce (*P. rubens*). The chloroplast genome exhibited more intraspecific variation than the mitochondrial genome. Red spruce displayed the most total chloroplast genetic diversity ($H_T = 0.52$). Neighbor-joining analysis arranged the chloroplast haplotypes into three monophyletic groups that were nearly 100% species-specific. These results strongly refute a previously proposed progenitor/derivative relationship of black/red spruce. Red and black spruce were estimated to have diverged from their common ancestor ~0.6-3.5 million years ago. Mitochondrial diversity detected in black spruce was attributed to interspecific hybridization, estimated to have taken place during the Holocene epoch (≥4000 years ago). White spruce mitochondrial haplotypes detected in multiple black spruce populations indicated that unidirectional introgressive hybridization has occurred between these two species. An east-west divide and opposing clines of chloroplast haplotypes in black spruce are consistent with leptokurtic dispersal out of either 1) a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

southeast glacial refugium of North America or 2) a small hypothetical northwestern refugium.

Some individual markers within cytoplasmic and nuclear genomes are species-specific. Some of these single nucleotide polymorphisms are appropriate to identify spruce macrofossils recovered from eastern North American lake cores. Since white, black and red spruce have differing climate tolerances, this would enable paleoclimatologists to track the migration routes of the individual spruce species during the last ice age and infer more precise climate estimates for eastern North America. Agarose gel electrophoresis, Southern blot and hybridization suggested that authentically ancient DNA survived in spruce macrofossils from 10-20,000 year old sediments of Browns Pond, Virginia.

A robust species phylogeny of the *Picea* is desirable to answer multiple questions about the history of the genus' biogeography. A phylogenetic study of sixteen North American and Eurasian *Picea* species was conducted utilizing DNA sequences of the chloroplast *trnK* intron and the mitochondrial *nad1* intron 2. The topologies of inferred trees varied significantly, in particular to the placement of *P. omorika*, *P. mexicana* and *P. glauca*. Inter-species hybridization and introgression are discussed as possible reasons behind such incongruencies.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

INTRODUCTION

The organisms studied in this dissertation are species of the genus *Picea* (spruce). Most of the work deals with the three species native to eastern North America: *Picea glauca* (Moench) Voss (white spruce), *P. mariana* (Mill.) B.S.P. (black spruce) and *P. rubens* Sarg. (red spruce). In Chapter I, species-specific DNA markers in the nuclear and chloroplast genomes are identified that distinguish these three species. Population studies were conducted to assess the distribution of these markers in the eastern portion of the black spruce range, and the entire ranges of white and red spruce. This chapter is a publication in Theoretical and Applied Genetics, the International Journal of Plant Breeding Research, volume 99, 1999. The second chapter is a continuation of this work.

In Chapter II, the sampling of black spruce is extended to include the species' entire range, additional intra- and interspecific chloroplast variants are scored, and polymorphisms in the mitochondrial genome are identified. The focus of Chapter II is not on developing more species-specific markers to distinguish the species; however, many of the new markers can be used as such. Instead, the focus is drawn to the amount and distribution of intraspecific variation, which provides insights into the history of introgression and migration of these species since the last ice age.

Chapters I and II are a foundation for Chapter IV, which describes the application of species-specific single nucleotide polymorphisms to the identification of spruce macrofossils recovered from eastern North American lake cores dating 20,000-10,000 years ago. This approach would enable paleontologists to directly track the migration routes of the individual spruce species. Since white, black and red spruce have varying

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

climatic and environmental preferences, the information could be used to infer more precise climate estimates for eastern North America since the ice age.

Finally, Chapter III deals with the more distant evolutionary history of *Picea*. It is a phylogenetic study of spruce species representing North America and Eurasia. The evolutionary history of the genus would aid in answering questions about the biogeography of the spruce species. In this study, sequence from the chloroplast and mitochondrial genomes are compared, and incongruencies between the phylogenetic trees inferred from the two data sets are discussed. Recommendations are made that would facilitate generating a robust molecular phylogeny of the *Picea*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Theor Appl Genet (1999) 99: 37–49

© Springer-Verlag 1999

J. Germano · A. S. Klein

Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca, P. mariana* and *P. rubens*

Received: 16 December 1998 / Accepted: 5 January 1999

Abstract *Picea rubens* (red spruce) and *P. mariana* (black spruce) are closely related species which are difficult to differentiate morphologically. They are sympatric with *P. glauca* (white spruce) in the northern portion of their ranges. In order to identify potential interspecific polymorphisms, the chloroplast *trn*K intron and *rpl*33-*psa*J-*trn*P region were sequenced, and the nuclear-encoded ITS region of the rDNA repeat was partially sequenced. Thirteen chloroplast and 12 nuclear candidate interspecific single nucleotide polymorphisms (SNPs) were identified. The species-specificity of several SNPs was determined by surveying DNAs amplified from trees representing range-wide provenance tests; these included 46 red spruce from 11 provenances, 84 black spruce from 30 provenances and 90 white spruce from 22 provenances. Two SNPs (1 chloroplast and 1 nuclear), which distinguish black spruce from red and white spruce, were consistent among 96–100% of the trees surveyed. Five SNPs (4 chloroplast and 1 nuclear), which distinguish white spruce from red and black spruce, were consistent among 100% of surveyed trees. These species-specific SNPs were used to identify anonymous spruce samples in a blind test, and their utility for small amounts of tissue, as little as single needles, was demonstrated. Scoring these SNPs is much less labor intensive than previous molecular methods for taxa differentiation (restriction fragment length polymorphisms or random amplified polymorphic DNAs), therefore they can be applied to large population studies.

Key words *Picea glauca* · *Picea mariana* · *Picea rubens* · Single nucleotide polymorphisms (SNPs) · Spruce

This is scientific contribution 2002 of the New Hampshire Agricultural Experiment Station, Durham, N.H.

Communicated by G. Wenzel

J. Germano · A. S. Klein (✉)
Department of Biochemistry and Molecular Biology,
University of New Hampshire, 46 College Rd, Durham,
NH 03824-2617, USA
Fax: +1 603 862 4013
E-mail: anita.klein@unh.edu

Introduction

Red spruce (*Picea rubens* Sarg.), black spruce [*P. mariana* (Mill.) B.S.P.] and white spruce [*P. glauca* (Moench) Voss] are sympatric species of northeastern North America. Red spruce's current range is from the Appalachians in North Carolina to the Maritimes including New England, New York, southern Quebec and restricted areas of Ontario (Morgenstern and Farrar 1964; Little 1971). The ranges of black and white spruce extend west to Alaska and north to the tree-line, overlapping with red spruce primarily in northern New England and eastern Canada (Morgenstern and Farrar 1964; Little 1971; Fowler et al. 1988). Although their ranges overelap, each of these species has a distinctive ecological niche (Morgenstern and Farrar 1964; Gordon 1976).

Morphological classification of the closely related red spruce and black spruce is difficult and controversial (Morgenstern and Farrar 1964; Manley 1971; Gordon 1976; Fowler et al. 1988); many morphological characters are too variable within one species to reliably distinguish it from the other (Gordon 1976; Fowler et al. 1988). Some traits are phenotopically plastic, and other characteristics are only discernable during limited time periods (Morgenstern and Farrar 1964; Gordon 1976; Donoghue and Sanderson 1992). Gordon (1976) conducted factor analysis of 24 morphological characters to distinguish red and black spruce. White spruce is more distantly related to red and black spruce (Gordon 1976; Sigurgeirsson and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Szmidt 1993) and can be easily identified with morphological characters.

Attempts have been made to distinguish red and black spruce with molecular markers including isozymes (Eckert 1989), random amplified polymorphic DNAs (RAPDs) (Perron et al. 1995) and restriction fragment length polymorphisms (RFLPs) (Bobola et al. 1992a, b, 1996). Bobola et al. used Southern blot analysis to identify polymorphisms between red and black spruce including five RFLPs of the nuclear ribosomal DNA (rDNA) repeat (1992a, b), two RFLPs in the chloroplast genome and three RFLPs in the mitochondrial genome (1996). None of these markers were 100% species-specific, however a three-character index, using the organelle markers together with one of the nuclear markers, reliably distinguished red and black spruce and their hybrids (Bobola et al. 1996). Although these markers are reliable, conventional RFLPs are tedious and costly both in their characterization and application.

Perron et al. (1995) identified seven RAPDs distinguishing red and black spruce in individual trees from six black spruce provenances and three red spruce provenances outside the sympatric zone. Four of these markers were 100% species-specific (i.e. present in all surveyed trees of one species and absent in all surveyed trees of the other species). A disadvantage of RAPDs is that this type of polymerase chain reaction (PCR™) is very sensitive to varying reaction conditions, temperature profiles and DNA quality (Ellsworth et al. 1993; Muralidharan and Wakeland 1993; Micheli et al. 1994; Vos et al. 1995). Hence, fingerprints may be difficult to reproduce, particularly in different laboratories.

Single nucleotide polymorphisms (SNPs) provide an alternative form of molecular markers for the discrimination of spruce species. Unlike conventional RFLPs and RAPDs, SNPs are direct markers; the exact nature and location of the allelic variations are known. Another advantage of SNPs is that large numbers of samples can be screened for a marker using a variety of inexpensive, high-throughput techniques (reviewed by Landegren et al. 1998). These methods may include restriction digestion (Landegren et al. 1998), allele-specific PCR (ASPCR) (Okayama et al. 1989; Sommer et al. 1989; Wu et al. 1989), or single-strand conformation polymorphism (SSCP) (Sheffield et al. 1993). Since each of these screening methods is PCR-based, only a small amount of template DNA is required, allowing samples to be identified from DNA extracted from just a few needles.

SNPs are likely to occur at higher frequencies in variable, less conserved genes. Few gymnosperm genes have been sequenced, therefore the potential variability of candidate genes in spruce must be gauged according to (1) their variability within angiosperm genera and/or (2) the predicted degree of variability based on functional constraints of the gene.

The approximately 10530-bp *mat*K gene, encoding an RNA maturase, is located within the approximately 2500-bp chloroplast *trn*K (UUU) intron (Sugita et al. 1985). *Mat*K exhibits a relatively low percentage of amino acid similarity (59% between tobacco and rice) compared to other chloroplast genes (Olmsted and Palmer 1994). It has been used to resolve phylogenetic relationships within families (Johnson and Soltis 1994, 1995; Steele and Vilgalys 1994; Li et al. 1999), and it displays some variation at the intrageneric level (Johnson and Soltis 1995; Li et al. 1997).

In gymnosperms, *mat*K has an average of 3.4 times more nucleotide differences per site than *rbc*L (Johnson and Soltis 1995). *Pinus* is the sister genus to *Picea* (Chase et al. 1993; Chaw et al. 1997). Hilu and Liang (1997) reported 1.1% *mat*K nucleotide variation between *Pinus contorta* (lodgepole pine) and *P. thunbergii* (black pine).

Noncoding regions such as introns or intergenic spacers (IGS) of the chloroplast are expected to be more variable than coding regions (Taberlet et al. 1991; Gielly and Taberlet 1994; Demesure et al. 1995; Perez de la Rosa et al. 1995). Chloroplast gene order is relatively conserved throghout land plants (Olmstead and Palmer 1994). Gene organization of *Picea* is similar to that of *Pinus* and *Pseudotsuga* (White et al. 1993). In addition, the entire chloroplast genome of black pine has been sequenced (Wakasugi et al. 1994). The conservation of chloroplast gene order, and the knowledge of a complete conifer chloroplast sequence present an opportunity to design new PCR primers for the amplification of potentially variable noncoding regions in spruce.

In red and black spruce, the chloroplast is paternally inherited, and the mitochondria are maternally inherited (Bobola et al. 1996). Nuclear interspecific markers would complement chloroplast markers since nuclear genes represent both the maternal and paternal lineage (Soltis et al. 1992). The nuclear ribosomal DNA repeat in plants contains the 18S, 5.8S and 26S genes separated by internal transcribed spacers 1 and 2 (ITS1 and ITS2) (reviewed by Hillis and Dixon 1991; Hamby and Zimmer 1992).

The ITS region (encompassing ITS1, 5.8S and ITS2) evolves at a much faster rate than the 18S and 26S genes (Hamby and Zimmer 1992; Baldwin et al. 1995). Baldwin et al. (1995) concluded that the ITS region is phylogenetically useful in angiosperms at intrafamilial levels, resolving relationships between genera, between species and even, to some extent, within species. Liston et al. (1999) used the 3′ ITS sequence (5.8S, ITS2 and approx. 200 bp of ITS1) to infer phylogenetic relationships of 47 species of *Pinus*. ITS1 was more divergent than 5.8S and ITS2, therefore it was hypothesized that additional ITS sequences would be useful in resolving relationships between more closely related *Pinus* species (Liston et al. 1999).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ITS1 is unusually large in some gymnosperms, making the ITS region up to 4.5 times as long as in angiosperms. For example, the ITS regions, including 5.8S and ITS2, in red and black spruce (Bobola et al. 1996) and *Picea abies* (Liston et al. 1996) are approximately 3.1 kb. The potential variability and large size of the ITS region make it a candidate gene in which to find polymorphisms between spruce species.

We have identified SNPs in nuclear and organelle genes which distinguish between white, red and black spruce. The species-specificity of each SNP was verified by screening samples representing range-wide provenance tests. The species-specific SNPs were used to identify anonymous spruce samples in a blind test, and their utility for small amounts of tissue was demonstrated.

Materials and methods

Plant materials

Some of the tissue samples used in this study were from independent collections. Black spruce sample 63 was provided by Dr. Peter Garrett of the Northeastern Forest Experiment Station of the US Forest Service. White spruce sample 494 was provided by Dr. Gerald Rehfeldt of the Intermountain Research Station, US Forest Service. White spruce sample 64 was collected from a tree on the University of New Hampshire campus. DNA was extracted from foliage using a standard CTAB method (Doyle and Doyle 1987).

Anonymous white, black and red spruce samples were provided by Dr. Robert Eckert, Department of Natural Resources, University of New Hampshire. A scaled down version of the CTAB method (Doyle and Doyle 1987), employing a 1.5-ml micro-centrifuge tube and mini-pestle to grind and material, was used to extract DNA from one to seven needles (9–16 mg).

DNAs from samples representing provenance tests were utilized for population studies. Red spruce DNAs represent a range-wide provenance test located in Coleman State Forest, Stewartstown, New Hampshire (see Bobola et al. 1996). Black spruce DNAs represent a provenance test corresponding to the eastern half of the black spruce range. This site is maintained by the USDA Forest Service (Northeastern Forest Experiment Station) in the Massabesic Experimental Forest, Alfred, Maine (see Bobola et al. 1996). White spruce DNAs representing a range-wide provenance test at Grand Rapids, Minnesota was provided by Dr. Glenn Furnier, Departments of Forest Resources and Plant Biology, University of Minnesota (see Furnier et al. 1991; Furnier and Stine 1995).

PCR methods

Taq DNA Polymerase (Promega, Madison, Wis.) was used in all reactions; *Taq* Extender™ PCR Additive (Stratagene, La Jolla, Calif.) was added to reactions to amplify fragments longer than 2 kb. Reactions contained 1 × Magnesium Free Reaction Buffer B (Promega) or 1 × *Taq* Extender Reaction Buffer (Stratagene), 0.2 m*M* each dNTP (Promega), 0.4 μ*M* each primer (Fig. 1, Table 1) and 5–10 ng/μl whole genomic DNA. All amplifications were carried out in an MJ Research PTC-100 Programmable Thermal Controller. The denaturation temperature was 94°C, and the extension temperature was 72°C. Each profile had an initial denaturation step of 3 min and a final extension step of 10 min. The *trn*K intron, the *rpl*33-*psa*J-*trn*P region and the ITS region (Fig. 1) were amplified as indicated in Table 2. Amplification of the ITS region was preceded by a hot start (D'Aquila et al. 1991). Primers rpl33 and trnP (Fig. 1b, Table 1) were designed from a black pine chloroplast sequence (GenBank D17510; Wakasugi et al. 1994). Additional amplification and sequencing primers (Fig. 1, Table 1) were designed using the PrimerSelect algorithm, part of the Lasergene software package (DNASTAR, version 3.72, Madison, Wis.).

PCR products were purified via electrophoresis through low-melting-point agarose (Gibco BRL, Gaithersburg, Md.). Bands were excised and liquified by adding 5 U Agarase (Sigma®, St. Louis, Mo.) per 100 μl gel and incubating at 37°C for 1 h.

Cloning

Gel-purified ITS PCR product (approx. 3.1-kb fragment) was digested with *Sau*3AI (New England Biolabs (NEB), Beverly, Mass. or Stratagene) according to the manufacturers' specifications. The resulting fragments, ranging from 200 bp to 900 bp, were gel-purified as described above, precipitated and resuspended in water. pGEM®-3Z vector (Promega) was digested with *Bam*HI (Promega), dephosphorylated with Calf Intestinal Alkaline Phosphatase (Promega) and ligated with ITS fragments using T4DNA ligase (Promega). The recombinant plasmids were used to transform Epicurian Coli® XL1-Blue MRF' supercompetent cells (Stratagene) according to the manufacturer's directions. Transformants were grown up on LB medium containing 50 μg/ml ampicillin. Plasmids were isolated using Wizard® Minipreps kit (Promega).

DNA sequencing

The ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA Polymerase, FS (Perkin Elmer, Emeryville, Calif.) was used. Plasmids were sequenced in separate reactions (20 μl) containing 3.2 pmol pUC/M13 forward or reverse primers (Promega), 250–500 ng DNA and 8 μl Ready Reaction Mix. PCR-amplified fragments were sequenced using 10 pmol primer (Fig. 1, Table 1) and 30–90 ng template. Cycle Sequencing was carried out, and extension products were purified using Ethanol Precipitation Protocol 1 as described by the manufacturer. Extension products were separated and analyzed on an ABI PRISM 373 Automated Sequencer (UNH Sequencing Facility). Sequences were generated with ABI DNA Sequencing Software version 2.1.1, Base caller ABI50, and edited by eye using ABI SeqEd Software version 1.0.3.

Both strands of DNA were sequenced for *mat*K and the *rpl*33-*psa*J-*trn*P region. One strand of DNA was sequenced for the ITS region and for the ITS1 cloned fragment.

Sequence analysis

Contiguous sequences were assembled and aligned using SeqMan II and MegAlign (DNASTAR) algorithms. Percentage divergence between sequences was calculated with MegAlign (DNASTAR). Restriction maps of the sequences were generated using MapDraw (DNASTAR).

Population studies

In order to assess the distribution of SNPs across populations, we used several molecular screening methods. Restriction degests using *Dra*I, *Ssp*I, *Bst*UI, *Bsp*1286I (NEB) and *Msp*I (or isoschizomer *Hpa*II,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

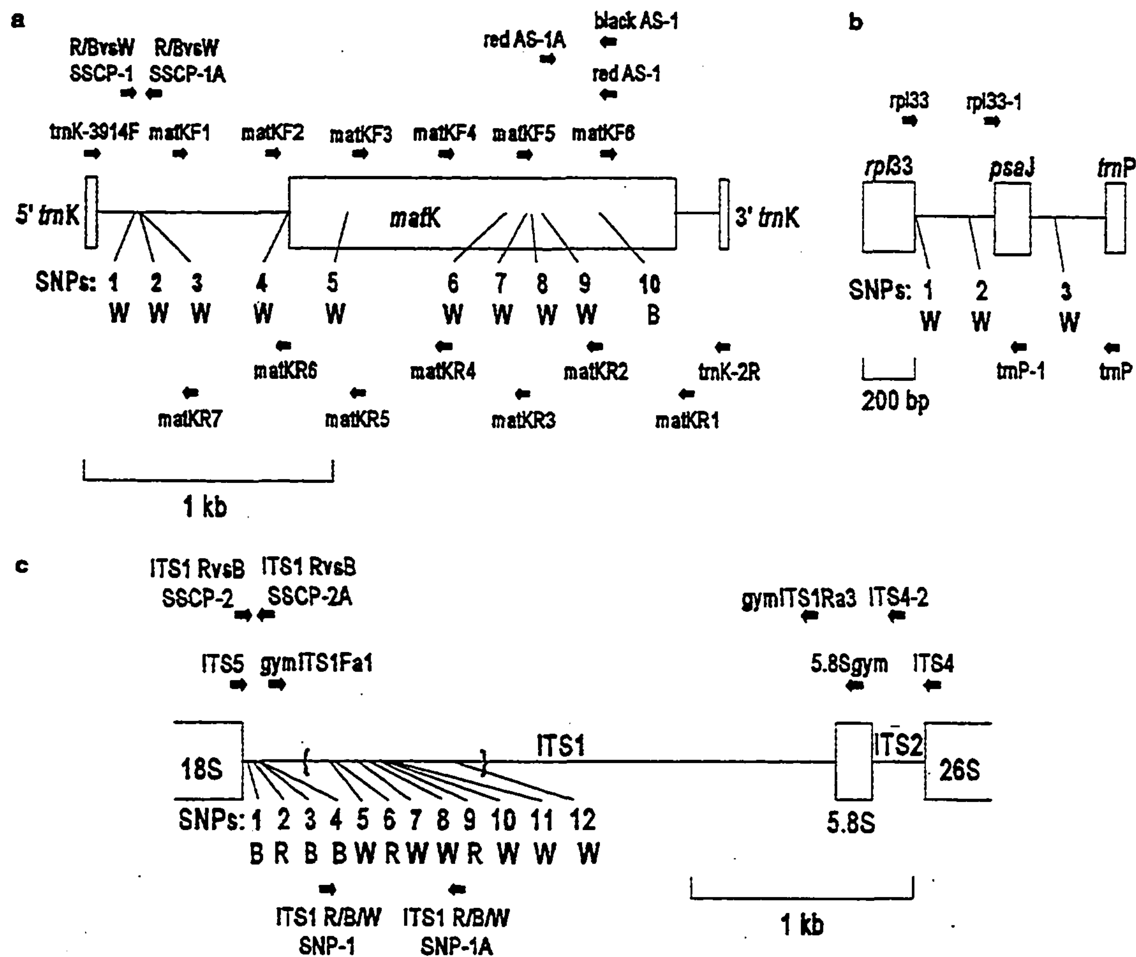


Fig. 1a–c Gene regions showing the relative positions of coding regions (*boxed areas*) and PCR amplification and sequencing primers. *Arrows* point in the 5' → 3' direction of primers. The relative positions of interspecific SNPs are shown. *Letters below* SNPs indicate the species that each SNP distinguishes: *W* distinguishes white spruce from red and black spruce, *B* distinguishes black spruce from white and red spruce, *R* distinguishes red spruce from white and black spruce. **a** *trn*K intron, **b** *rpl*33-*psaJ*-*trn*P region, **c** ITS region. *Brackets* indicate the cloned 778-bp fragment of ITS1

Promega) were carried out according to the manufacturers' specifications. Resulting fragments were separated on 1.4–3% agarose; fragments smaller than 200 bp were separated on 3–4% MetaPhor® agarose (FMC BioProducts, Rockland, Me.). Single-strand conformation polymorphism (SSCP) analysis was carried out according to the protocol for MDE™ Gel (FMC). Two microliters (3–6 ng) of PCR product (119-bp fragment) was mixed with 10 µl stop solution, and the entire denatured sample was loaded onto a 1 × gel. The gel was stained in 1 µg/ml ethidium bromide for 15 min and visualized under UV light. Allele-specific PCR (ASPCR) was carried out using primers which differentially amplify specific alleles (Tables 1 and 2). All reactions contained an additional primer set as a positive control (see Wu et al. 1989). PCR products were separated by electrophoresis through 1.4% agarose.

Results

Chloroplast gene sequences

In order to identify potential interspecific polymorphisms, we sequenced the chloroplast *trn*K intron and the *rpl*33-*psaJ*-*trn*P region from 4 individual trees representing geographically distant populations of each species (Table 3) as suggested by Baverstock and Moritz (1996). There was no intraspecific variation among these individuals in the *trn*K intron. Intraspecific variation was limited in the *rpl*33-*psaJ*-*trn*P region, consisting of three *rpl*33-*psaJ* IGS nucleotide substitutions among 2 of the 4 back spruce individuals, and one *psaJ*-*trn*P IGS nucleotide substitution in 1 of the 4 white spruce individuals.

Screening interspecific *trn*K SNPs for species-specificity

Ten candidate interspecific single nucleotide polymorphisms (SNPs) were identified in the *trn*K intron: 4 in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1 PCR amplification and sequencing primers (*Seq* sequencing)

Primer name	PCR/Seq[a]	5′ Sequence 3′
trnK-3914F[b]	PCR and seq	TGG GTT GCT AAC TCA ATG G
trnK-2R[b]	PCR and seq	AAC TAG TCG GAT GGA GTA G
matKF1	PCR and seq	TCG AAT GAG TCA ATG GAG AAA G
matKF2	Seq	CGC ACC ATG TAT TGT ATT ATC TCA
matKF3	Seq	TGA ATC GGT ACT AGA AGG ACT GAC
matKF4	Seq	AGA TTC TTC CTG TTC CTG TGG
matKF5	Seq	CCA TCT TTG GAA CGA ACC
matKF6	Seq	CCT TTG GTC GAG ATG GTT TAT
matKR1	PCR and seq	CTC GGA TGG CAA AAT AAA AAT G
matKR2	Seq	GGA TCC ACT GTA GTA ATG AAA AAT
matKR3	Seq	CTA TAA GGT TCG TTC CAA AGA TG
matKR4	Seq	CCA CAG GAA CAG GAA GAA TC
matKR5	Seq	TCA GTC CTT CTA GTA CCG ATT CAG
matKR6	Seq	CTC AGT TAT GGC CCT CGT TC
matKR7	Seq	GGG AAT TCC TCG CTC GTT
R/BvsW SSCP-1	PCR	ACA GCA TGT CGT TCC AAC A
R/BvsW SSCP-1A	PCR	ATA GAC ATT TCC CAC CCA TTT
red AS-1	PCR	GGA TCC ACT GTA GTA ATG AAA AAG
black AS-1	PCR	GGA TCC ACT GTA GTA ATG AAA AAT
red AS-1A	PCR	TTG GTC AAG ACT AAA ATG CTA GA
rpl33	PCR and seq	TGT TAC TCT TAT ATC TCC GCT CTT
trnP	PCR and seq	CAA AAC AAA CAC GCT ACC AA
rpl33-1	Seq	TGA CTT TAA GGG GAG GAC AAC
trnP-1	Seq	CTG TGT TAG CTA TTT CAT CGT TCA
ITS5[c]	PCR and seq	GGA AGT AAA AGT CGT AAC AAG G
ITS4[c]	PCR and seq	TCC TCC GCT TAT TGA TAT GC
gymITS1Fa1	Seq	TGT TGT CCT TGG CCT CCT
ITS4-2	Seq	GAC AAT ATC ACC GCT CGC C
5.8Sgym	Seq	GAT GAT TCA CGG GAT TCT G
gymITS1Ra3	Seq	CCA CAA GAC ATA TGC ACT C
ITS1 RvsB SSCP-2	PCR and seq	TGC GGT AGG ATC ATT GTC AGT
ITS1 RvsB SSCP-2A	PCR and seq	CGA TCA ACC CTC CAA AAG TG
ITS1 R/B/W SNP-1	PCR and seq	CTT CGT TTG AGT CTT TGT TTT TCG
ITS1 R/B/W SNP-1A	PCR and seq	GGG CCA CCG GAG CAT TG

[a] Indicates whether a primer was used for PCR, sequencing, or both
[b] Johnson and Soltis (1994)
[c] White et al. (1990)

the 5′ noncoding region and 6 in *matK* (Fig. 1a, Table 4). Among these candidate SNPs were eight transversions and two transitions; 4 of the 6 SNPs in *matK* are non-synonymous. One of the SNPs (trnK SNP 10) distinguishes black spruce, while the remaining 9 distinguish white spruce. TrnK SNP 10 is located at the first position of codon 410 in *matK*, encoding isoleucine in black spruce and leucine in red and white spruce.

Five of the *trnK* SNPs were tested for species-specificity by screening for the presence or absence of the particular nucleotides (Table 4) in DNAs amplified from trees representing range-wide provenance tests. A total of 46 red spruce (3–5 from each of 11 provenances), 84 black spruce (1–4 from 30 provenances plus 1 separate sample) and 90 white spruce (4 from 22 provenances plus 2 separate samples) samples were screened. Several diagnostic methods were used, including single-strand conformation polymorphism (SSCP), restriction analysis and allele-specific PCR (ASPCR).

SSCP was used to screen trnK SNPs 1, 2 and 3 (Fig. 1a) simultaneously. At these positions, the sequence of the 4 white spruce individuals had nucleotides G, A and T, whereas the red and black spruce had T, T and G (Table 4). TrnK SNP 2 and 3 are adjacent, and SNP 1 is located 19 bp upstream. The primers R/BvsW SSCP-1 and R/BvsW SSCP-1A (Fig. 1a, Tables 1 and 2) were designed to amplify a 119-bp fragment in which trnK SNPs 1, 2 and 3 are centrally located.

Upon denaturation of the amplified fragments and separation on 1 × MDE (FMC), the white spruce single-stranded DNAs travelled at different rates than those of red and black spruce. This difference in electrophoretic mobility resulted in a distinct banding pattern for white spruce and a different pattern for red and black spruce (Fig. 2a). All of the provenance test DNAs were surveyed for these banding patterns: 100% (90/90) white spruce displayed the "close" pattern, and 100% (46/46) red and 100% (84/84) black spruce displayed the "wide" pattern (Fig. 2a, Table 4).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2 PCR profiles

Gene or region	Primers	Fragment size	Reaction volume (µl)	*Taq* units	*Taq* extender units	$MgCl_2$ (m*M*)	Annealing temperature (°C)	Annealing time	Extension time	Cycles
*trn*K intron	trnK-3914F & trnK-2R[a]	~2.5 kb	50	2	2		48	1 min, 30 s	2 min, 30 s	30
*rpl*33-*trn*P	rpl33 & trnP	809 bp	50	2		2.5	47	1 min, 30 s	1 min	30
ITS	ITS5 & ITS4[b]	~3.1 kb	50	4	4		45	1 min, 45 s	3 min	29
*trn*K intron	matKF1 & matKR1	2072 bp	50	2	2		51	1 min	2 min, 15 s	30
ITS1	ITS1 RvsB SSCP-2 ITS1 RvsB SSCP-2A	121 bp	25	0.5		2.5	50	1 min	30 s	30
ITS1	ITS1 R/B/W SNP-1 ITS1 R/B/W SNP-1A	609 bp	25	1		1	51	1 min	45 s	30
*trn*K intron	R/BvsW SSCP-1 R/BvsW SSCP-1A	119 bp	25	0.5		2.5	47	1 min	30 s	30
*mat*K	red AS-1 & red AS-1A R/BvsW SSCP-1 R/BvsW SSCP-1A	231 bp & 119 bp	25	1		1	47	1 min	20 s	28
*mat*K	black AS-1 & red AS-1A R/BvsW SSCP-1 R/BvsW SSCP-1A	231 bp & 119 bp	25	1		1	47	1 min	20 s	28

[a] Johnson and Soltis (1994)
[b] White et al. (1990)

The restriction enzyme *Dra*I was used to screen trnK SNP 4 (Fig. 1a), which distinguishes white spruce from red and black spruce. At this position, the white spruce sequences have a G, whereas red and black spruce have a T (Table 4), part of a *Dra*I restriction site (TTTAAA). The 2072-bp fragment, amplified with primers matKF1 and matKR1 (Fig. 1a, Tables 1 and 2), also contains an invariant *Dra*I site. This restriction site acts as a positive control by generating an invariant 229-bp fragment. Therefore, *Dra*I was predicted to produce two fragments (229 bp and 1843 bp) for white spruce and three fragments (229 bp, 466 bp and 1377 bp) for red and black spruce. One hundred percent (90/90) of white spruce, 96% (81/84) of black spruce and 85% (39/46) of red spruce provenance test sample DNAs displayed the predicted fragment size patterns following *Dra*I digestion (Table 4).

The restriction enzyme *Ssp*I was used to screen trnK SNP 10 (Fig. 1a), which distinguishes black spruce from white and red spruce. At this position, the red and white spruce sequences have a C, and black spruce has an A (Table 4), part of a *Ssp*I restriction site (AATATT). The 2072-bp fragment, amplified with primers matKF1 and matKR1 (Fig. 1a), also contains three invariant *Ssp*I sites which generate three fragments (52 bp, 358 bp and 812 bp). Therefore, *Ssp*I was predicted to produce five fragments (52 bp, 358 bp, 376 bp, 474 bp and 812 bp) for black spruce and only four fragments (52 bp, 358 bp, 812 bp and 850 bp) for red and white spruce. *Ssp*I restriction analysis was used to screen a subset of the white spruce provenance test samples; 100% (43/43) of them displayed the predicted fragment size pattern (Table 4).

ASPCR was used to screen trnK SNP 10 (Fig. 1a) in the red and black spruce provenance test DNAs and those white spruce DNAs not screened with *Ssp*I. Allele-specific primers were designed whose 3′ nucleotide anneals at the position of trnK SNP 10. Primer black AS-1 (Fig. 1a) contains a 3′ T (Table 1), complementary to the A (Table 4) at this site in black spruce. Primer red AS-1 (Fig. 1a) contains a 3′ G (Table 1) that is complementary to the C (Table 4) at this site is red and white spruce. The generic upstream primer red AS-1A (Fig. 1a, Table 1) pairs with either allele-specific primer to amplify a 231-bp fragment (Table 2). An additional pair of primers, R/BvsW SSCP-1 and R/BvsW SSCP-1A (Fig. 1a, Tables 1 and 2), was added to every PCR to amplify a 119-bp fragment as a positive control on the success of the reaction (see Wu et al. 1989). Two PCRs were done for each individual sample: one containing primer red AS-1, the other containing primer black AS-1, and both containing the generic primer and the control primer pair (Table 2). Scoring amplification of the 231-bp allele-specific fragment was carried out only if the 119-bp positive control fragment successfully amplified in both reactions (Fig. 2b). In 100% (48/48) of white and 98% (45/46) of red spruce samples, the 231-bp fragment

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3 Samples used for sequencing

Species	Sample	Geographic location	GenBank accession					
			*trn*K intron	*rpl*33-*trn*P	ITS1[a]	ITS1[b]	ITS1[c]	ITS[d]
White spruce	64	Durham, NH	AF133923	AF133935	AF117916	AF140755	AF119377-AF119379	AF136610
	494	Black Hills, SD	AF133924	AF133936	AF117917	AF140756		AF136611
	S-A	Saskatchewan	AF133925	AF133937	AF117918	AF140757		AF136612
	V-A or V-B	Alaska	AF133926	AF133938	AF117919	AF140758		AF136613
Black spruce	63	Durham, NH	AF133919	AF133931	AF117912	AF140751	AF119374-AF119376	AF136614
	4274-1	New Hampshire	AF133920	AF133932	AF117913	AF140752		AF136615
	4962-1	Newfoundland	AF133921	AF133933	AF117914	AF140753		AF136616
	5004-1	New Brunswick	AF133922	AF133934	AF117915	AF140754		AF136617
Red spruce	2019 21-1	Indian Gap, NC	AF133915	AF133927	AF117908	AF140747	AF119371-AF119373	AF136618
	2027 29-4	Pillsbury, NH	AF133916	AF133928	AF117909	AF140748		AF136619
	2032 8-1	Valcartier, Quebec	AF133917	AF133929	AF117910	AF140749		AF136620
	2505 38-5	Acadia Forest Experiment Station, NB	AF133918	AF133930	AF117911	AF140750		AF136621

[a] Sequence of PCR product amplified with primers ITS1 RvsB SSCP-2 and ITS1 RvsB SSCP-2A
[b] Sequence of PCR product amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A
[c] Sequence of cloned fragment of PCR product amplified with primers ITS5 and ITS4
[d] Includes partial sequence of ITS1 (3′ end) and complete 5.8S and ITS2 sequences

amplified with the red AS-1 primer but not with the black AS-1 primer. In 98% (82/84) of the black spruce samples, the 231-bp fragment amplified with the black AS-1 primer but not with the red AS-1 primer (Fig. 2b, Table 4).

Screening interspecific *rpl*33-*psa*J-*trn*P SNPs for species-specificity

Three candidate interspecific SNPs were identified in the *rpl*33-*psaJ*-*trn*P region: 2 in the *rpl*33-*psaJ* IGS and 1 in the *psaJ*-*trn*P IGS (Fig. 1b). All 3 SNPs in this region are transitions, and all distinguish white spruce from red and black spruce.

The restriction enzyme *Msp*I was used to screen rpl33-trnP SNP 3 (Fig. 1b). At this position, the white spruce sequences have an A, whereas red and black spruce have a G (Table 4), part of a *Msp*I restriction site (CCGG). The 809-bp fragment, amplified with primers rpl33 and trnP (Fig. 1b, Tables 1 and 2), also contains an invariant *Msp*I site, which generates a 407-bp fragment. Therefore, *Msp*I was predicted to generate two nearly equal-sized fragments (407 bp and 402 bp, seen as one bright band on an agarose gel) for white spruce and three fragments (407 bp, 227 bp and 175 bp) for red and black spruce (Fig. 2c). One hundred percent (90/90) of white spruce, 100% (84/84) of black spruce and 100% (46/46) of red spruce provenance test sample DNAs displayed the predicted fragment size patterns following *Msp*I digestion (Table 4).

ITS sequences

An approximately 3.1-kb fragment encompassing the ITS region (ITS1, 5.8S and ITS2) was amplified and partially sequenced (Fig. 1c, Tables 1 and 2). Sequences of 4 individual trees of each species (Table 3) were obtained, including complete sequences for 5.8S and ITS2 plus 300–550 bp from the 5′ end of ITS1 and 300–650 bp from the 3′ end of ITS1 (Fig. 1c). The 5.8S gene is 162 bp and ITS2 is 236 bp in each of the species. The endpoints of spruce 5.8S and ITS2 were determined by comparison to the ITS region sequence of *Pinus pinea* (stone pine) (GenBank X87936; Marrocco et al. 1996).

The rDNA repeat is present in multiple copies in the nuclear genome; it is assumed that the PCR-amplified DNAs represent ratios similar to the genomic copies. The sequences were obtained using PCR product as template, and therefore they represent the main fraction of the template DNAs. The sequence chromatograms, especially for ITS1, contained a significant amount of background; in many locations, there were two distinct peaks of two different nucleotides, one on top of the other. This result can be explained by heterogeneity of the template DNAs. This background caused the sequence signals to deteriorate rapidly, sometimes limiting chromatogram analysis to only 150–200 bases per reaction.

Primers gymITS1Fa1 and gymITS1Ra3, which anneal to opposite ends of ITS1 (Fig. 1c), yielded limited sequence from PCR product as template. One explanation for this sequencing difficulty is the heterogeneity of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4 Results of population studies

Gene region	*trn*K		*trn*K		*trn*K		*rpl*33-*trn*P		ITS		ITS	
Location	5′ *trn*K intron		5′ *trn*K intron		*mat*K		*psa*J-*trn*P IGS		ITS1		ITS1	
Species	SNPs 1, 2 & 3	SSCP analysis	SNP 4	*Dra*I analysis	SNP 10	ASPCR or *Ssp*I analysis	SNP 3	*Msp*I analysis	SNP 1	*Bst*UI analysis	SNP 7	*Bsp*1286I analysis
White	G AT[a]	90/90[b] (100%)	G	90/90 (100%)	C	90/90 (100%)	A	90/90 (100%)	G	89/90 (99%)	C	89/89[c] (100%)
Black	T TG	84/84 (100%)	T	81/84 (96%)	A	82/84 (98%)	G	84/84 (100%)	A	84/84 (100%)	G	84/84 (100%)
Red	T TG	46/46 (100%)	T	39/46 (85%)	C	45/46 (98%)	G	46/46 (100%)	G	44/46 (96%)	G	46/46 (100%)

[a] Nucleotide positions of trnK SNP 2 and trnK SNP 3 are adjacent
[b] Proportion of provenance test samples whose DNAs scored positive for the corresponding SNP allele
[c] This region would not amplify with the DNA from one white spruce individual

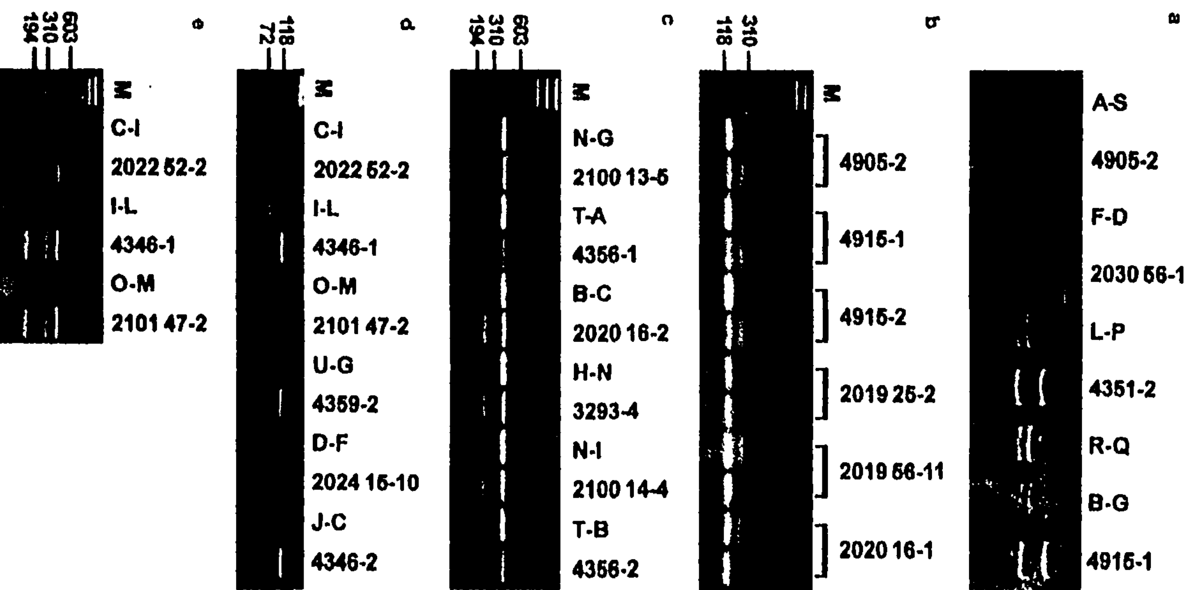


Fig. 2a–e Representative results from screening provenance test samples. *Alphabetic labels* represent white spruce samples, *Numeric labels* beginning with *2* represent red spruce samples and *numeric labels* beginning with *3 or 4* represent black spruce samples. The marker (*M*) is ϕX 174DNA/*Hae*III with the sizes of some fragments labeled in base pairs. **a** SSCP analysis of trnK SNPs 1, 2 and 3. **b** ASPCR analysis of trnK SNP 10. In the *left lane* for each sample are PCR products from primer red AS-1 plus the positive control primer pair. In the *right lane* for each sample are PCR products from primer black AS-1 plus the positive control primer pair. The 119-bp band (present in every lane) is the positive control PCR product. The 231-bp band is the product of the allele-specific primers. **c** *Msp*I analysis of rpl33-trnP SNP 3. **d** *Bst*UI analysis of ITS SNP 1. **e** *Bsp* 1286I analysis of ITS SNP 7

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the rDNA copies. Another possible explanation is the presence of subrepeats within ITS such as those reported in stone pine by Marrocco et al. (1996) and *Pseudotsuga* and *Larix* by Gernandt and Liston (1999). Preliminary results from sequencing multiple cloned ITS *Sau*3AI fragments of different lengths suggest that subrepeats also exist in red, black and white spruce ITS1 (data not shown). During cycle-sequencing using PCR product as template, if a primer anneals to subrepeats, then it would anneal to multiple locations of the template DNA. This would cause multiple sequences to be represented in the chromatograms, making the data increasingly difficult to analyze as the subrepeat sequences diverge.

There were no interspecific differences between white, black and red spruce in the 5.8S gene and ITS2, and in the final 300 bp of the 3′ end of ITS1. The 5′ end of ITS1 is more variable than its 3′ end. Four candidate interspecific SNPs were identified in the initial 100 bp from the 5′ end of ITS1 (Fig. 1c). They are all transitions, including three that distinguish black spruce and one that distinguishes red spruce.

Screening interspecific ITS SNP 1 for species-specificity

The restriction enzyme *Bst*U1 was used to screen ITS SNP 1, which distinguishes black spruce from red and white spruce. At this position, the black spruce sequences have an A, whereas red and white spruce have a G (Table 4), part of a *Bst*U1 restriction site (CGCG). This is the only *Bst*U1 site in the 121-bp fragment amplified with primers ITS1 RvsB SSCP-2 and ITS1 RvsB SSCP-2A (Fig. 1c, Tables 1 and 2). Therefore, *Bst*U1 was predicted to generate two fragments (44 bp and 77 bp) for red and white spruce but not to cleave the 121-bp fragment amplified from black spruce. When the 121-bp fragment amplified from DNAs of provenance test trees was digested with *Bst*U1 however, a fraction (0–20%) of the PCR product was cleaved in many of the back spruce samples, and a fraction (10–30%) of the PCR product was not cleaved in the red and white spruce samples (Fig. 2d). These results suggest that the copies of the rDNA repeat within an individual are heterogeneous for this *Bst*U1 site. The observed heterogeneity was considered for individual amplified DNAs that were scored for ITS SNP 1. The majority (⩾ 70%) of the PCR product was cleaved into two fragments in 99% (89/90) of white and 96% (44/46) of red spruce provenance test sample DNAs, and the majority (⩾ 80%) of the PCR product was not cleaved in 100% (84/84) of the black spruce provenance test sample DNAs (Fig. 2d, Table 4).

Sequences of cloned ITS fragments

In order to extend the sequence of ITS1, the approximately 3.1-kb PCR product encompassing ITS1, 5.8S and ITS2 was digested with *Sau*3AI. Some of the *Sau*3AI fragments were cloned and sequenced from 1 individual of each species (Table 3). The longest clone, 778 bp, is contiguous with the 5′ ITS1 sequence obtained with primers ITS5 and gymITS1Fa1 (Fig. 1c). Three 778-bp clones from each individual were sequenced. Alignment of these sequences revealed a significant amount of intra-individual variation, including single nucleotide transitions, transversions and insertion/deletions. There were 29 SNPs among the three clones from white spruce sample 64, 12 SNPs among the three clones from black spruce sample 63, and 16 SNPs among the three clones from red spruce sample 2019 21-1. These intra-individual polymorphisms further support the hypothesis that there is considerable heterogeneity among copies of ITS1.

Eight candidate interspecific SNPs were identified in the 778-bp *Sau*3AI clone, including three transversions, four transitions and one insertion/deletion (Fig. 1c). Two of the SNPs distinguish the red spruce sample, while the other 6 distinguish the white spruce sample. In order to confirm these candidate SNPs we sequenced a 609-bp fragment (amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A) from 4 individuals of each species (Table 3) using PCR product as template. This sequence confirmed the cloned fragment sequences at all positions for ITS SNPs 6-12.

Screening interspecific ITS SNP 7 for species-specificity

The restriction enzyme *Bsp*1286I was used to screen ITS SNP 7 (Fig. 1c), which distinguishes white spruce from red and black spruce. At this position, the white spruce sequences have a C, whereas the red and black spruce sequences have a G (Table 4), part of a *Bsp*1286I restriction site (GDGCHC). The 609-bp fragment, amplified with primers ITS1 R/B/W SNP-1 and ITS1 R/B/W SNP-1A (Fig. 1c, Tables 1 and 2), also contains another *Bsp*1286I site, generating a 322-bp positive control fragment. Therefore, *Bsp*1286I was predicted to generate three fragments (322 bp, 168 bp and 119 bp) from red and black spruce amplified DNAs and only two fragments (322 bp and 287 bp) from white spruce DNAs. Interestingly, the positive control *Bsp*1286I site coincides with an intra-individual polymorphic site, as evidenced by the alignment of cloned 778-bp fragment sequences from 1 individual. Hence, cleavage at the positive control site did not occur in a fraction (up to 35%) of the amplified DNAs from each individual sample. Therefore, some 609-bp fragments are observed in digests of white spruce DNAs, and some 441-bp fragments are observed in digests of red and black spruce DNAs (Fig. 2e). The nucleotide at the ITS SNP 7 position, however, appeared to be homogeneous within each individual. One hundred percent (89/89) of white spruce provenance test sample DNAs were not cleaved

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

at the ITS SNP 7 site; 100% (84/84) of black spruce and 100% (46/46) of red spruce sample DNAs were cleaved at the ITS SNP 7 *Bsp*1286I site (Table 4). Data for all of the SNPs listed in Table 4 from screening provenance test samples in population studies were analyzed with a χ^2 test of independence ($P < 0.0001$).

Testing species-specific SNPs as a tool to identify species from single needles

In order to use these molecular tools to identify species in a blind test, 45 anonymous white, black and red spruce samples were provided by Dr. Robert Eckert. DNA was extracted from as little as one needle per sample. SSCP analysis of trnK SNPs 1, 2 and 3, *Msp*I analysis of rpl33-trnP SNP3 and *Bsp*1286I analysis of ITS SNP 7 were carried out in order to identify the white spruce samples. Then ASPCR analysis of trnK SNP 10 and *Bst*UI analysis of ITS SNP 1 was used to distinguish the black spruce samples from the red and white spruce. One hundred percent (45/45) of the anonymous samples were identified correctly.

Discussion

We sequenced (1) the chloroplast *trn*K intron including *mat*K, (2) the region between chloroplast *rpl*33 and *trn*P including *psa*J and (3) nuclear rDNA 5.8S, ITS2, and portions of ITS1 in 4 individuals each of white, black and red spruce. We identified one nuclear and four chloroplast SNPs which distinguish white spruce from black and red spruce, plus one nuclear and one chloroplast SNP which distinguish black spruce from white and red spruce. These markers are strictly associated with species, as determined by surveying trees representing range-wide provenance tests of each species. In a blind test, the species-specific nuclear and chloroplast markers were used to correctly identify anonymous samples. These molecular markers can be used to identify the species of a white, black and red spruce tree from single needles.

Levels of sequence variation in spruce

The percentage divergence of all genes sequenced was remarkably low between the spruce species. For *mat*K, there was only 0.1% divergence between red and black spruce, 0.3% between white and red spruce and 0.4% between white and black spruce. This is markedly lower than the 1.1% *mat*K variation between lodgepole pine and black pine (Hilu and Liang 1997).

Between 5% and 15% sequence divergence between the taxa in question is thought to provide a sufficient number of characters for phylogenetic analysis (Olmstead and Palmer 1994). The relatively high level of variation between some angiosperm species makes *mat*K appropriate for resolving phylogenetic relationships (Johnson and Soltis 1995). Lodgepole pine and black pine are believed to be distantly related members of subgenus *Pinus* based on ITS sequences (Liston et al. 1999). White and black spruce are also thought to be distantly related based on RFLPs of chloroplast DNA (Sigurgeirsson and Szmidt 1993). The low levels of variation between distantly related species within these two genera suggest that *mat*K may not be sufficiently divergent to infer relationships among species of *Pinus* or among species of *Picea*.

The noncoding regions of the *trn*K intron displayed no more variation than *mat*K between spruce species. There was 0% divergence between red and black spruce, 0.4% between white and red spruce and 0.4% between white and black spruce. *Pinus trn*K intron noncoding regions also show slightly less variation than *mat*K having 1% divergence between lodgepole pine (GenBank X57097; Lidholm and Gustafsson 1991) and black pine (GenBank D17510; Tsudzuki et al. 1992). The combined noncoding intergenic spacers between *rpl*33 and *trn*P also had very low levels of variation between spruce species: 0% divergence between red and black spruce, 0.5% between white and red spruce and 0.5% between white and black spruce.

There was no variation in the nuclear rDNA ITS2, 5.8S gene or the 3′ end of ITS1 among these three spruce species. This is strikingly different than the level of 5.8S and ITS2 variation among species of *Pinus* (Liston et al. 1999).

Variation was present in the 5′ end of ITS1 (5′ 330 bp): approximately 0.9% divergence between red and black spruce, 2.2% between white and red spruce and 2.2% between white and black spruce. There was also variation in the cloned ITS1 fragment between species: as much as 1.4% divergence between red and black spruce, 2.4% between white and red spruce and 1.4% between white and black spruce. These values may be inflated, however, due to the observed heterogeneity of ITS within an individual.

Heterogeneity of ITS

Nuclear rDNA is present in multiple copies arranged in tandem repeats (Hamby and Zimmer 1992). Angiosperms have thousands of copies located at one or a few chromosomal loci (nucleolus organizer regions, NORs) (Hamby and Zimmer 1992). Gymnosperms have been shown to possess many more copies; in red and black spruce there are as many as 10^6 copies rDNA per nuclear genome (Bobola et al. 1992b). Using in situ hydridization Brown et al. (1993) identified 12-14 rDNA chromosomal loci in white spruce.

There have been previous reports of heterogeneity among rDNA repeats in conifers (Bobola et al. 1992b;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Beech and Strobeck 1993; Karvonen and Savolainen 1993; Liston et al. 1996) in both ITS and IGS. Here we present further evidence that the copies of ITS1 are heterogeneous within a spruce individual. (1) Sequence chromatograms from using PCR product as template showed multiple peaks at single positions. (2) Restriction endonucleases *Bst*U1 and *Bsp*1286I each cleaved PCR products differently within screened individuals. (3) There was significant variability between sequences of cloned ITS1 fragments within an individual. In fact, there was 0.8–1.7% divergence between the clones of a red spruce individual, 0.6–1.2% divergence between the clones of a black spruce individual and 1.3–3.2% divergence between the clones of a white spruce individual.

This heterogeneity may be explained by the large number of rDNA chromosomal loci in spruce. Liston et al. (1996, 1999) pointed out that the large number of rDNA loci might slow the process of concerted evolution which tends to homogenize the rDNA repeats within an individual. The high level of heterogeneity of ITS1 within an individual makes estimation of the levels of intra- and interspecific variation difficult. It also complicates the process of identifying species-specific SNPs in the ITS region as markers to distinguish between species. Furthermore, such high levels of intra-individual variation would complicate phylogenetic analysis using ITS1 sequences.

Species-specific markers to distinguish between spruce species

Perron et al. (1995) stressed the importance of identifying molecular markers from trees whose species has been carefully identified a priori. They chose populations from outside the sympatric zone, where the extent of natural hydridization of red and black spruce is probably low, and conducted a five-character morphological analysis on each tree. Only trees with a morphological composite index specific to either species were used in their study (Perron et al. 1995). In our study, each tree of the provenance tests was typed using six morphological characters (Gordon 1976; Eckert 1990). Additionally, any suspected hybrids and/or introgressed individuals from the provenance tests identified with nuclear and organelle RFLPs (Bobola et al. 1996) were excluded from the population studies.

Perron et al. (1995) identified RAPD markers which distinguish red and black spruce from only six black populations (12 trees) and three red spruce populations (12 trees), and verified them in ten black and nine red spruce F_1 progeny from interspecific crosses. They characterized four RAPD markers that were present in 100% of surveyed red spruce (21 total provenance and F_1 progeny trees) and absent in 100% of the surveyed black spruce (22 total trees). The three black-distinguishing markers were absent in 100% of the red spruce surveyed but only present in 75–92% of surveyed black spruce provenance trees and in 70–100% of black spruce F_1 progeny. Perron et al. (1995) referred to all of these markers as "species-specific".

In this study, we used SNPs as molecular markers to distinguish between white, black and red spruce. A total of 25 candidate interspecific SNPs were identified in the three sequenced gene regions. Of these, 18 distinguished white spruce, 4 distinguished black spruce and 3 distinguished red spruce. The abundance of white-distinguishing markers relative to the number of red- or black-distinguishing markers is not surprising because white spruce is not as closely related as red and black spruce (Sigurgeirsson and Szmidt 1993).

Eight of the candidate SNPs were screened in 11 red spruce populations (46 trees), 30 black spruce populations (84 trees) and 22 white spruce populations (90 trees) in order to assess their degree of species-specificity. These trees represent a considerable portion of the species' ranges.

During the Pleistocene Epoch, spruce populations were displaced in response to glaciation events. Repeated alterations in the geographic distribution of these trees influenced the amounts of genetic variability throughout species (Critchfield 1984). Therefore, range-wide sampling is important for this type of population study in order to include samples from both current and previously disjunct populations that may be experiencing or have undergone genetic bottlenecks.

The nuclear (ITS SNP 7) and chloroplast (trnK SNPs 1, 2 and 3, and rpl33-trnP SNP 3) SNPs which distinguish white spruce are 100% species-specific. The results of the indirect screening methods suggested that all of the white spruce trees surveyed had the white spruce-type nucleotide and that all of the black and red spruce trees surveyed had the black/red spruce-type nucleotide (Table 4). TrnK SNP 4, which also distinguishes white spruce, was 100% consistent in surveyed white spruce trees and 96% consistent in surveyed black spruce. As only 85% of the red spruce trees displayed the red/black spruce-type base, trnK SNP 4 was not considered "species-specific".

Chloroplast trnK SNP 10, which distinguishes black spruce from white and red spruce, is located at the first position of codon 410 in the *mat*K gene. This marker was consistent among 100% of the white spruce surveyed and among 98% of the red and black spruce. Only 1 red spruce tree (out of 46) appeared to have the black spruce-type base, and only 2 black spruce trees (out of 84) appeared to have the red/white spruce-type base.

Nuclear ITS SNP 1 has a more complicated species profile due to the heterogeneity of ITS1 within an individual. At this position, sequence chromatograms (using PCR as template) show a G in white and red spruce and an A in black spruce. *Bst*U1 digestion of PCR product, however, indicates that some of the white and red spruce DNAs have a G at this site and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that some of the black spruce DNAs do not have a G at this site. Although the rDNA copies are not homogeneous within an individual at this site, this marker is still useful for distinguishing between the spruce species because the majority of amplified DNAs have a G in white and red spruce individuals, and the majority of amplified DNAs have an A in black spruce individuals. These restriction profiles were consistent in 99% of surveyed white spruce, 100% of black spruce and 96% of red spruce. Therefore, only 1 white spruce tree (out of 90) and 2 red spruce trees (out of 46) had a majority of DNAs lacking a G at this site.

The validity of the nuclear and chloroplast markers was confirmed by using each of the species-specific SNPs to identify anonymous white, black and red spruce samples in a blind test. This test was carried out with DNA from as little as a single needle per sample, thereby demonstrating the utility of these markers when abundant tissue is not available.

We have identified seven species-specific SNP markers that can reliably identify white, black or red spruce from a single needle. SNPs are easier to identify and more robust than other molecular tools, such as isozymes (Eckert 1989), RFLPs (Bobola et al. 1992a, b, 1996) and RAPDs (Perron et al. 1995), previously used to distinguish spruce species. Furthermore, screening SNP markers with indirect PCR-based methods enables an efficient high throughput of samples.

Acknowledgements We wish to thank Dr. R.T. Eckert, Dr. G.R. Furnier and Dr. G.E. Rehfeldt for providing plant materials. We are grateful to Dr. C.D. Neefus for assisting with the statistical analysis and illustrations. We also thank Dr. J.J. Collins, Dr. R.H. Cote and Dr. T.D. Kocher for their critical review of the manuscript. This research was supported by a grant from the New Hampshire Agricultural Experiment Station (NHAES) MacIntire-Stennis Project MS27. The experiments carried out for this project comply with the current laws of the USA.

References

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann Mo Bot Gard 82:247–277

Baverstock PR, Moritz C (1996) Project design. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Assoc, Sunderland, Mass., pp 17–27

Beech RN, Strobeck C (1993) Structure of the intergenic spacer region from the ribosomal RNA gene family of white spruce (*Picea glauca*). Plant Mol Biol 22:887–892

Bobola MS, Eckert RT, Klein AS (1992a) Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana*. Can J For Res 22:255–263

Bobola MS, Smith DE, Klein AS (1992b) Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. Mol Biol Evol 9:125–137

Bobola MS, Eckert RT, Klein AS, Stapelfeldt K, Smith DE, Guenette D (1996) Using nuclear and organelle DNA markers to discriminate among *Picea rubens*, *Picea mariana*, and their hybrids. Can J For Res 26:433–443

Brown GR, Amarasinghe V, Kiss G, Carlson JE (1993) Preliminary karyotype and chromosomal localization of ribosomal DNA sites in white spruce using fluorescence in situ hybridization. Genome 36:310–316

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Retting JH, Conti E, Palmer JD, Manhart JR, Systma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang GY, Plunkett GM, Soltis PS, Swensen WM, Williams SE, Gadek PA, Quinn CJ, Equiarte LE, Golenberg E, Learn GHJ, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann Mo Bot Gard 80:528–580

Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14:56–68

Critchfield WB (1984) Impact of the Pleistocene on the genetic structure of North American conifers. In: Lanner RM (ed) Proc 8th N Am For Biol Workshop. International Union of Forest Research Organizations, Birmensdorf, Switzerland, pp 70–118

D'Aquila RT, Bechtel LJ, Videler JA, Eron JJ, Gorczyca P, Kaplan JC (1991) Maximizing sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Res 19:3749

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131

Donoghue MJ, Sanderson MJ (1992) The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 340–368

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Eckert RT (1989) Genetic variation in red spruce and its relation to forest decline in the Northeastern United States. In: Bucher JB, Bucher-Wallin I (eds) Air pollution and forest decline. (Proc 14th Int Meet Specialists Air Pollution Effects For Ecosystems). International Union of Forest Research Organizations P2.05, Birmensdorf, Switzerland, pp 319–324

Eckert RT (1990) Genetic variation in red spruce pollution response. Final report. Spruce-Fir Research Cooperative, Northeast Forest Experiment Station, Radnor, Pa.

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14:214–217

Fowler DP, Park YS, Gordon AG (1988) Genetic variation of red spruce in the Maritimes. Can J For Res 18:703–709

Furnier GR, Stine M (1995) Interpopulation differentiation of nuclear and chloroplast loci in white spruce. Can J For Res 25:736–742

Furnier GR, Stine M, Mohn CA, Clyde MA (1991) Geographic patterns of variation in allozymes and height growth in white spruce. Can J For Res 21:707–712

Gernandt DS, Liston A (1999) ITS region evolution in *Larix* and *Pseudotsuga*. Am J Bot (in press)

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. Mol Biol Evol 11:769–777

Gordon AG (1976) The taxonomy and genetics of *Picea rubens* and its relationship to *Picea mariana*. Can J Bot 54:781–813

Hamby RK, Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 50–91

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Quart Rev Biol 66:411–453

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hilu KW, Liang H (1997) The *mat*K gene: sequence variation and application in plant systematics. Am J Bot 84:830–839
Johnson LA, Soltis DE (1994) *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Syst Bot 19:143–156
Johnson LA, Soltis DE (1995) Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Ann Mo Bot Gard 82:149–175
Karvonen P, Savolainen O (1993) Variation and inheritance of ribosomal DNA in *Pinus sylvestris* L. (Scots pine). Heredity 71:614–622
Landegren U, Nilsson M, Kwok PY (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res 8:769–776
Li JH, Bogle AL, Klein AS (1997) Interspecific relationships and genetic divergence of the disjunct genus *Liquidambar* (Hamamelidaceae) Rhodora 99:229–240
Li JH, Bogle AL, Klein AS (1999) Phylogenetic relationships of the Hamamelidaceae: evidence from the nucleotide sequences of the plastid gene *mat*K. Plant Syst Evol (in press)
Lidholm J, Gustafsson P (1991) A three-step model for the rearrangement of the chloroplast *trn*K-*psb*A region of the gymnosperm *Pinus contorta*. Nucleic Acids Res 19:2881–2887
Liston A, Robinson WA, Oliphant JM, Alvarez-Buylla ER (1996) Length variation in the nuclear ribosomal DNA internal transcribed spacer regions of non-flowering seed plants. Syst Bot 21:109–120
Liston A, Robinson WA, Piñero D, Alvarez-Buylla ER (1999) Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phyl Evol 11:95–109
Little EL Jr (1971) Conifers and important hardwoods. (Atlas of United States trees). U.S. Department of Agriculture, Forest Service, Washington, D.C. Miscellaneous Publ No. 1146, vol 1
Manley SAM (1971) Identification of red, black and hybrid spruce. Can Dep Environ For Can For Serv, Ottawa, Canada
Marrocco R, Gelati MT, Maggini F (1996) Nucleotide sequence of the internal transcribed spacers and 5.8S region of ribosomal DNA in *Pinus pinea* L. DNA Seq 6:175–177
Micheli MR, Bova R, Pascale E, D'Ambrosio E (1994) Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. Nucleic Acids Res 22:1921–1922
Morgenstern EK, Farrar JL (1964) Introgressive hybridization in red spruce and black spruce. Technical Report No. 4. Faculty of Forestry, University of Toronto, Toronto, Ontario, pp 1–46
Muralidharan K, Wakeland EK (1993) Concentration of primer and template qualitatively affects products in random-amplified polymorphic DNA PCR. Biotechniques 14:362–364
Okayama H, Curiel DT, Brantly ML, Holmes MD, Crystal RG (1989) Rapid, nonradioactivite detection of mutations in the human genome by allele-specific amplification. J Lab Clin Med 114:105–113
Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205–1224
Perez de la Rosa J, Harris SA, Farjon A (1995) Noncoding chloroplast DNA variation in Mexican pines. Theor Appl Genet 91:1101–1106
Perron M, Gordon AG, Bousquet J (1995) Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. Theor Appl Genet 91:142–149
Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 16:325–332
Sigurgeirsson A, Szmidt AE (1993) Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. Nord J Bot 13:233–246
Soltis PS, Doyle JJ, Soltis DE (1992) Molecular data and polyploid evolution in plants. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 177–201
Sommer SS, Cassady JD, Sobell JL, Bottema CD (1989) A novel method for detecting point mutations of polymorphisms and its application to population screening for carriers of phenylketonuria. Mayo Clin Proc 64:1361–1372
Steele KP, Vilgalys R (1994) Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastis gene *mat*K. Syst Bot 19:126–142
Sugita M, Shinozaki K, Sugiura M (1985) Tobacco chloroplast tRNA Lys (UUU) gene contains a 2.5-kilobase-pair intron: an open reading frame and a conserved boundary sequence in the intron. Proc Natl Acad Sci USA 82:3557–3561
Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109
Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trn*Q, *trn*K, *psb*A, *trn*I and *trn*H and the absence of *rps*16. Mol Gen Genet 232:206–214
Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414
Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91:9794–9798
White EE, Watkins RF, Fowler DP (1993) Comparative restriction site maps of chloroplast DNA of *Picea abies*, *Picea glauca*, *Picea mariana*, and *Picea sitchensis*. Can J For Res 23:427–435
White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Inns M, Gelfand D, Sninsky J, White TJ (eds) PCR protocols: a guide to methods and applications. Academic Press, San Diego, pp 315–322
Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989) Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc Natl Acad Sci USA 86:2757–2760

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER II

GEOGRAPHIC DISTRIBUTION OF MITOCHONDRIAL AND CHLOROPLAST HAPLOTYPES AMONG RANGE-WIDE POPULATIONS OF WHITE SPRUCE (*PICEA GLAUCA*) BLACK SPRUCE (*P. MARIANA*) AND RED SPRUCE (*P. RUBENS*): IMPLICATIONS OF SPECIATION, INTERSPECIFIC HYBRIDIZATION AND HISTORY OF THE WISCONSIN ICE AGE

Abstract

A total of twenty-seven mitochondrial and chloroplast haplotypes were identified within and among range-wide populations of white spruce (*Picea glauca*), black spruce (*P. mariana*) and red spruce (*P. rubens*). The chloroplast genome exhibited more intraspecific variation than the mitochondrial genome. Among the species, red spruce displayed the most total chloroplast genetic diversity ($H_T = 0.52$). Neighbor-joining analysis arranged the chloroplast haplotypes into three monophyletic groups that were nearly 100% species-specific. The diversity of red spruce chloroplast haplotypes and their reciprocal monophyly with haplotypes detected in black spruce strongly refute a previously proposed progenitor/derivative relationship of black/red spruce. Red and black spruce were estimated to have diverged from their common ancestor approximately 0.6-3.5 million years ago. Mitochondrial diversity, detected primarily in black spruce, was attributed to interspecific hybridization. Hybridization between red and black spruce occurred during the Holocene epoch, at least 4000 years ago. White spruce mitochondrial haplotypes detected in multiple black spruce populations indicated that unidirectional introgressive hybridization has occurred between these two species. An east-west divide and opposing clines of chloroplast haplotypes in black spruce are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

consistent with leptokurtic dispersal either 1) out of a well-recognized southeast glacial refugium of North America or 2) out of a small hypothetical northwestern refugium in addition to the southeast.

Introduction

The current genetic structure of populations of forest trees reflects their recent history of migration and adaptation in response to climate change. During the Quaternary period (the last two million years), vast areas of North America, Europe and Asia were repeatedly covered by glacial ice due to oscillations in climate (Comes and Kadereit, 1998; Hewitt, 2000). At the onset of periods of glaciation, the ranges of forest trees shifted; some contracted and became fragmented as species retreated to southern refugia and isolated ice-free regions. Upon warming of the climate and shrinkage of the glaciers, ranges shifted again, expanding into newly unglaciated territories. Populations that spent glacial periods in geographic isolation may have reunited to form a continuous range (Critchfield, 1984). The spruce genus, *Picea* Dietr., of the Pinaceae comprises 35-50 species that are widespread across the northern hemisphere (Schmidt-Vogt, 1977; Vidakovic, 1991). Paleoclimate and vegetation history reconstructions have focused on the distribution of this important forest component during the Quaternary, with particular attention to the migration history of Norway spruce (*Picea abies*) in Europe, and white spruce (*Picea glauca* (Moench) Voss), black spruce (*P. mariana* (Mill.) B.S.P.) and red spruce (*P. rubens* Sarg.) in North America (e.g. Jackson *et al.*, 1997; Taberlet *et al.*, 1998).

Driving Forces in Forest Species Migration

Forests endured approximately twenty glacial-interglacial cycles during the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Quaternary period. During the most recent ice age (the Wisconsin period; ~100,000-10,000 years ago), a massive ice sheet covered most of Canada, extending as far south as 40°N over the northeastern United States by the Last Glacial Maximum (LGM; ~21,000 calendar years ago; Delcourt and Delcourt, 1987a; Hewitt, 2000). Pollen and plant macrofossil records from lake and bog sediments indicate that forest species, including spruce, had taken refuge in the southern United States (e.g. Delcourt and Delcourt, 1987a; Jackson *et al.*, 1997). Much of northern Europe was glaciated as well, with refugia located in the southern parts of the continent (Huntley and Birks, 1983). Following the LGM, glaciers melted as climate conditions warmed and the Holocene interglacial epoch ensued (~11,000 years ago to the present). Many species migrated from refugia to occupy their current positions. There is a growing amount of evidence that the present patterns of genetic variation in forest trees are largely the result of expansion/migration out of glacial refugia during the Holocene (reviewed in Comes and Kadereit, 1998; Newton *et al.*, 1999; Hewitt, 2000).

Historical expansions and contractions of individual species' ranges resulted from their inherent adaptability to climatic changes, involving genetic variability within species, fitness optima (Rehfeldt *et al.* 2001) and gene flow within and among populations (Davis and Shaw 2001). Furthermore, species migrations have had genetic consequences for the emergent population structure, which in turn affected the adaptation, and hence the evolution of that species (Critchfield 1984; Comes and Kadereit 1998; Newton *et al.* 1999; Hewitt 2000; Davis and Shaw 2001). During glacial periods reproductively isolated populations may have become differentiated due to mutation, selection and/or genetic drift. As a result of retreat and re-immigration, some

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

species experienced significantly reduced genetic variability via bottlenecks and founder events (Critchfield 1984; Comes and Kadereit 1998; Hewitt 2000; Davis and Shaw 2001).

Genetic Structure and Biogeography

A variety of molecular markers have been used to infer postglacial re-colonization patterns of forest trees since the LGM in North America and Europe. Chloroplast and mitochondrial genomes are well suited for the reconstruction of migration routes in forest trees because of their uniparental inheritance (Petit *et al.*, 1993). In spruce and most conifers, the chloroplast is paternally inherited while the mitochondrial genome is maternally inherited (Sutton *et al.*, 1991; David and Keathley, 1996; Bobola *et al.*, 1996b). Population subdivision of mitochondrial DNA markers is greater than that of chloroplast markers due to differential migration of pollen and seed, i.e. varying degrees of gene flow, within wind-pollinated conifer populations (Dong and Wagner, 1994; Latta and Mitton, 1997; Latta *et al.*, 1998). Therefore mitochondrial polymorphisms are popular for investigations of migration history, but pollen-dispersed markers have been informative as well (Scotti *et al.*, 2000; Vendramin *et al.*, 2000; Collignon and Favre, 2000).

Geographical patterns of genetic differentiation have been observed that point to routes of migration from putative glacial refugia. The molecular phylogeography of Norway spruce in Europe has been studied using a variety of markers including nuclear sequence-characterized amplified region (SCAR) markers (Scotti *et al.*, 2000), mitochondrial haplotypes (Sperisen *et al.*, 1998; Sperisen *et al.*, 2001), chloroplast restriction fragment length polymorphisms (RFLPs) (Sigurgeirsson, 1992), and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chloroplast microsatellites (Vendramin *et al.*, 2000). The spatial patterns of these genetic markers were associated with Norway spruce glacial refugia that were evidenced by pollen and macrofossil records. In North America, subdivided populations of limber pine (*Pinus flexilis*) were identified based on RFLP analysis of a mitochondrial gene intron (Mitton *et al.*, 2000a). The authors inferred directions of dispersal out of several different glacial refugia by analyzing the geographic patterns of the mitochondrial DNA markers.

Migration History of White, Black and Red Spruce in North America

White spruce, black spruce and red spruce are northern North American species that presently share a sympatric zone in New England and eastern Canada. The current range of red spruce extends north from the Appalachians in North Carolina to the Canadian Maritimes (Morgenstern and Farrar, 1964; Little, Jr., 1971). The ranges of black and white spruce reach west to Alaska and north to the tree-line (Morgenstern and Farrar, 1964; Little, Jr., 1971; Fowler *et al.*, 1988). Although their ranges overlap, each of these species has distinctive ecological, edaphic, mesic and climatological niches (Morgenstern and Farrar, 1964; Gordon, 1976; Nienstaedt and Zasada, 1990; Viereck and Johnston, 1990; Blum, 1990; Vann *et al.*, 1994; McLeod and MacDonald, 1997).

According to palynological (pollen) studies, spruce took refuge in a wide band between 40 and 35°N (in and around Virginia, Kentucky, Tennessee, etc.) across the unglaciated southeastern United States during the Wisconsin ice age (Davis, 1983; Ritchie and MacDonald, 1986; Delcourt and Delcourt, 1987a; Jackson *et al.*, 1997). White, black and red spruce are believed to have been present in this southern refugium; however the boundaries of individual species' ranges are unknown. An additional disjunct refugium of white spruce was hypothesized to have existed in the unglaciated

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

portion of northwestern North America (Yukon and Alaska [eastern Beringia]) during the Wisconsin period (Tsay and Taylor, 1978; Critchfield, 1984; Ritchie and MacDonald, 1986){Furnier & Stine 1995 50 /id}. This hypothesis is supported by pollen records and patterns of current genetic variation; however there is no radiocarbon-dated macrofossil evidence to confirm the existence of this putative refugium. As the climate warmed, the range of spruce expanded from the southeast refugium northeastward, reaching New England, Nova Scotia and Newfoundland between 12,700 and 8,000 yr BP (Delcourt and Delcourt, 1987a). Glacial retreat and repopulation by spruce occurred extremely rapidly in the western interior of Canada (Critchfield, 1984; Ritchie and MacDonald, 1986; McLeod and MacDonald, 1997).

Because of different ecological tolerances, individual species' ranges shifted at different times, rates and in varying directions during periods of climate change (Davis, 1983; Taberlet *et al.*, 1998; Comes and Kadereit, 1998; Davis and Shaw, 2001). For example, red spruce withstands much warmer summer temperatures than black and white spruce, while black and white spruce are more cold-hardy. These species also exhibit varying edaphic and mesic preferences. Considering such individual environmental tolerance limits, it is probable that white, black and red spruce migrated differentially since the LGM. Examining the distribution of DNA markers among extant spruce populations may improve our understanding of the postglacial migration histories of the individual species.

Interspecific Hybridization

Hybridization and introgression are common phenomena among spruce species and may contribute to adaptive variation in a species through gene enrichment. Under

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

extreme adaptive pressures of environmental change, populations that contain high levels of genetic variation would have an increased probability of giving rise to individuals capable of surviving such changes (Davis and Shaw, 2001). Pure and introgressed trees of white spruce, Engelmann spruce (*Picea engelmannii*) and Sitka spruce (*Picea sitchensis*) exhibit varying levels of adaptation to different ecological niches (Sigurgeirsson *et al.*, 1991; Sutton *et al.*, 1994). Extant hybrids and introgressants of red and black spruce occur naturally, however their persistence and abundance in the wild are debated (Manley, 1972; Gordon, 1976; Eckert, 1989; Perron *et al.*, 1995; Bobola *et al.*, 1996a; Bobola *et al.*, 1996b; Perron and Bousquet, 1997). It is also unclear whether red/black spruce introgressants display negative or positive heterosis with respect to environmental conditions (Morgenstern and Farrar, 1964; Fowler *et al.*, 1988; Bobola *et al.*, 1996a). Naturally occurring hybrids of white and black spruce are believed to be very rare (Wright, 1955; Hultén, 1968; Viereck and Little, Jr., 1972). Interspecific hybridization and introgression of spruce prior to or during the last ice age may have directly affected their response to the changing climate.

Much concern has been expressed over the decline of red spruce populations since the 1800's, with anthropogenic factors and climatic warming discussed as possible causes (Hamburg and Cogbill, 1988; Vann *et al.*, 1994). Analyses of isozymes and random amplified polymorphic DNAs (RAPDs) have suggested that red spruce harbors relatively low levels of allelic variation (Eckert, 1989; Hawley and DeHayes, 1994; Perron *et al.*, 1995). The lack of genetic diversity in red spruce may have resulted from genetic bottlenecks during the last glacial period, which in turn may have contributed to the species' recent decline (Eckert, 1989; Hawley and DeHayes, 1994). Based on a study

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

using sequence tagged site markers (STS), Perron *et al.* (2000) proposed that black and red spruce comprise a recently evolved progenitor-derivative species pair. The authors hypothesized that the lack of genetic variation in red spruce was a direct result of its allopatric speciation from a preexisting black spruce population, which may have been isolated and become differentiated during the Pleistocene epoch (~2 million to 10,000 years ago).

Variable DNA Markers

In order to investigate the distribution of genetic diversity in white, black and red spruce, potentially variable DNA markers located in the oppositely inherited chloroplast and mitochondrial genomes were chosen. We have previously identified intraspecific variation of the chloroplast *trnK* intron within black and red spruce, and of the chloroplast *rpl33-psaJ-trnP* region within black and white spruce (Germano and Klein, 1999). The *trnT-trnL-trnF* region of the chloroplast genome is phylogenetically informative for angiosperms at the species level (Gielly and Taberlet, 1994), and displays variation among species of pine and fir (*Abies*) (Perez de la Rosa *et al.*, 1995; Watano *et al.*, 1996; Isoda *et al.*, 2000). Universal primers have been used to amplify this region in spruce (Taberlet *et al.*, 1991; Perez de la Rosa *et al.*, 1995). Primers have also been developed to amplify non-coding regions of the mitochondrial genome (Demesure *et al.*, 1995; Dumolin-Lapegue *et al.*, 1997; Mitton *et al.*, 2000b). The second intron of *nad1* (*nad1 B/C*) displays variation within several species of pine and within Norway spruce (Latta and Mitton, 1997; Sperisen *et al.*, 1998; Grivet *et al.*, 1999; Mitton *et al.*, 2000a; Mitton *et al.*, 2000b; Sperisen *et al.*, 2001).

The goals of this project are 1) to identify chloroplast and mitochondrial variation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

within and among range-wide populations of white, black and red spruce, 2) to compare levels of genetic diversity and degrees of population subdivision between genomes and between species, and 3) to ascertain any geographical patterns of genetic differentiation. The results were used to address several questions. Do genetic diversity estimates from range-wide samplings of red and black spruce support the progenitor-derivative species pair hypothesis posed by Perron *et al.* (2000)? Are geographic patterns of genetic diversity consistent with white, black and red spruce's migration out of the well-documented southeastern refugium of the Wisconsin period? Do they support the hypothesis of a disjunct Alaskan glacial refugium for white spruce? Did introgressive hybridization play a role in red and black spruces' responses to the warming climate of the last 10,000 years?

Materials and Methods

Plant Materials and DNA Extraction

The acquisition of most provenance test sample DNAs used in this study has been previously described (Germano and Klein, 1999). White spruce samples were from a range-wide provenance test at Grand Rapids, Minnesota (see Furnier and Stine, 1995). Red spruce DNAs represent a range-wide provenance test located in Coleman State Forest, Stewartstown, New Hampshire. Black spruce DNAs, corresponding to the eastern portion of the range, are from a provenance test maintained by the USDA Forest Service (Northeastern Forest Experiment Station), in the Massabesic Experimental Forest, Alfred, Maine (Bobola *et al.*, 1992a; Bobola *et al.*, 1992b). Additional black spruce samples from northern and western Canada were collected as follows (see Table 1): samples representing the Alberta provenance were collected from Alberta Land and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Northern and western black spruce samples obtained from provenance test plantations.

Provenance[a]	Location	Sample size (n)	Latitude (N)	Longitude (W)
Research Plantation Exp G137A				
474	Slave Lake Forest, Alberta	3	54° 50'	115° 19'
510	Grande Prairie Forest, Alberta	3	54° 28'	119° 42'
1759	Lac La Biche Forest, Touchwood Lake, Alberta	3	54° 53'	111° 28'
1771	Athabasca Forest, Alberta	3	57° 07'	111° 40'
2000	Footner Lake Forest, Alberta	3	59° 30'	117° 30'
2240	Rocky/Clearwater, Alberta	3	52° 15'	115° 20'
2522	Peace River Forest, Alberta	3	56° 41'	116° 15'
Reno Trial				
6963	Duck Mountain, Manitoba	5	51° 38'	100° 47'
6964	Pulp River, Manitoba	5	51° 48'	100° 12'
6967	Point Lake, Manitoba	5	55° 30'	98° 04'
6971	Beaver River, Saskatchewan	5	54° 43'	107° 49'
6972	Nisbet Provincial Forest, Saskatchewan	5	53° 14'	105° 46'
6986	Fort St. John III, British Columbia	5	56° 37'	121° 28'
6987	Steamboat Creek, British Columbia	5	58° 47'	123° 36'
6988	Steamboat Mountain, British Columbia	5	58° 44'	123° 38'
7007	Bonanza Creek II, Alaska, USA	4	64° 44'	148° 18'
Research Plantation Experiment 353-H-5				
6850	Lebel sur Quevillon, Quebec	4	49° 07'	76° 57'
6855	Matagami, Quebec	4	49° 37'	77° 45'
6856	Manicouagan, Quebec	4	50° 40'	68° 46'
6859	Parc Mistassini, Quebec	4	50° 27'	73° 28'
6862	Murdochville, Quebec	4	48° 55'	65° 25'
6909	Otasawian River, Ontario	4	49° 45'	85° 05'
6920	Minchin Lake, Ontario	4	50° 44'	90° 34'
6924	Red Lake, Ontario	4	50° 53'	93° 44'
6930	Rainy Lake, Ontario	4	48° 48'	93° 40'
6932	Shebandowan, Ontario	3	48° 40'	90° 11'
6936	Moosonee I, Ontario	4	51° 16'	80° 46'
6961	Riding Mountain, Manitoba	5	54° 52'	95° 27'
6969	Jan Lake, Saskatchewan	5	54° 52'	102° 48'
7000	Mayo, Yukon	5	63° 34'	135° 55'

[a] Provenance designations and locations were provided by N. Dhir and L. Barnhardt, personal communication.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Forest Service Research Plantation Exp G137A, which is located near Smoky Lake, Alberta; samples representing northern Quebec, northern Ontario, Yukon, Saskatchewan, and Manitoba provenances were from Canadian Forest Service Research Plantation Exp 353-H-5 located in the Petawawa Research Forest, Chalk River, Ontario; samples representing British Columbia, Alaska, and additional regions of Yukon, Saskatchewan, and Manitoba were from the Canadian Forest Service Reno Range Wide Black Spruce Trial. One additional black spruce (collected from New Hampshire), two white spruce (collected from Durham, New Hampshire and from Black Hills, South Dakota) samples from independent collections (see Germano and Klein, 1999), and two red spruce samples (from Indian Gap and Newfound Gap in Great Smoky Mountain National Park, Tennessee) were also included in this study.

Red and black spruce trees from the Coleman State Forest and Massabesic Experimental Forest provenance tests were previously typed using six morphological characters (Gordon, 1976; Eckert, 1990) and nuclear and organelle RFLPs (Bobola *et al.*, 1996b). Individuals determined to be hybrid by these tests were excluded from the present study in order to rule out recent hybrids from the zone of sympatry (Bobola *et al.*, 1996b). DNAs were extracted from 5-10 g fresh or frozen foliage using a standard CTAB method (Doyle and Doyle, 1987).

PCR Methods

Taq DNA Polymerase (in Storage Buffer B, Promega, Madison, WI) was used in all reactions; Taq Extender™ PCR Additive (Stratagene, La Jolla, CA) was added to reactions to amplify fragments longer than 2 kb. Reactions contained final concentrations of 1X Reaction Buffer (without $MgCl_2$, Promega) or 1X Taq Extender

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reaction Buffer (Stratagene), 0.2 mM each dNTP (Promega), 0.4 µM each primer (Table 2) and 5-10 ng/µl whole genomic DNA. Reaction volume, amounts of Taq Polymerase and magnesium chloride are listed in Table 3 for individual reactions. Amplification and sequencing primers (Table 2) were designed using the PrimerSelect algorithm, part of the Lasergene software package (DNASTAR, version 3.72, Madison, WI).

Amplifications were carried out in a PTC-100 Programmable Thermal Controller (MJ Research Inc., Watertown, MA). All PCR profiles had an initial denaturation step of three minutes, a final extension step of ten minutes, and denaturation and extension temperatures of 94°C and 72°C, respectively. Annealing temperatures, annealing times and extension times for individual PCRs are summarized in Table 3. The *nad1 B/C* intron was amplified using a Touchdown PCR profile (Don *et al.*, 1991). Following a 3-4 minute 94°C 'Hot Start' (D'Aquila *et al.*, 1991), each cycle consisted of a 30-second denaturation step, a one-minute annealing step and a three-minute extension step. The first four cycles had an annealing temperature of 60°C; the annealing temperature decreased by 2°C every four cycles thereafter. The lowest annealing temperature of 46°C was used for the final 21 cycles.

PCR products to be sequenced were purified via electrophoresis through low melting point agarose (Gibco BRL, Gaithersburg, MD). Excised bands were liquefied by incubating at 37°C for one hour with 5 units Agarase (Sigma®, St. Louis, MO) per 100 µl gel.

DNA Sequencing

Samples were sequenced (Table 4) using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA Polymerase, FS (Perkin

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Primers used for PCR and sequencing

Primer name	PCR / Seq[a]	Sequence (5' → 3')
nad1 exon B[b]	PCR & Seq	GCA TTA CGA TCT GCA GCT CA
nad1 exon C[b]	PCR & Seq	GGA GCT CGA TTA GTT TCT GC
NAD1-B1	Seq	AGT GGG GAC CTT AAT GAT GAA
NAD1-C1	Seq	GGC GTA TAG GTT GTG TTG CT
NAD1-B2	PCR & Seq	AGC GGG TCC TGG TCA AA
NAD1-C2	Seq	GCA TAG CGA AGG GAA GGT C
NAD1-B3	Seq	CTC AAA GGG CTA GCT AAA AGG TA
NAD1-C3	Seq	GCG CGA ACG AAC ACT AAG A
NAD1-B4	PCR & Seq	AGC CCT TCC TGA TTG ACT CTA
NAD1-B5	Seq	AAT CCC TTT CTT TTG CGT AGC
NAD1-C4	Seq	CCC CCA CCG GAT AAG TT
NAD1-C5	Seq	CAA AAG ACG CCC ACT CCA
NAD1-C6	Seq	CCA AAA GCA TGG TTC CTG TC
NAD1-B Indel 2	PCR	AGC GGG TCC TGG TCA AAT C
NAD1-C Indel 2	PCR & Seq	AAA AGG GAG AGG GTA ATA AAC AA
NAD1-B Indel 3	PCR & Seq	AAT TAA TAT CGA ACG AAG AAA ATC
NAD1-C Indel 3	PCR & Seq	CCG AAG GGG TAG GGA GAT
NAD1-B Indel 4	PCR & Seq	AGA AAA TAG GCT ATA ACG AAA GT
NAD1-C Indel 4	PCR & Seq	GGA ACC GAA CCG CTC TA
NAD1-B Indel 5	PCR	CAA ATT TGA GGA TCT TGA TAC G
NAD1-C Indel 5	PCR	TAG TCC ACT CAT TGA AGG CTA AA
NAD1-B Indel 6	PCR & Seq	AGC GGT TCG GTT CCC TCA C
NAD1-C Indel 6	PCR	ATT TAA TGG CGG CCC AGA CTT T
NAD1-B Indel 7	PCR	CCA TCC TTA TTA ACC CTT ATC TGA
NAD1-C Indel 7	PCR	GAC AAT TTG AGT CGA TTT TCT TC
NAD1 MS2R	PCR & Seq	AAC CCT CCA TAA GAT CTC CTC TCA
NAD1 SNP B-G	PCR	TAC CTA CCC CTC GCT ACT ATC TC
NAD1 SNP B-T	PCR	TAC CTA CCC CTC GCT ACT ATC TA
nad7/1[c]	PCR & Seq	ACC TCA ACA TCC TGC TGC TC
nad7/2r[c]	PCR & Seq	CGA TCA GAA TAA GGT AAA GC
NAD7-1U	PCR & Seq	TTG GGA ATG AAC GGA GAA
NAD7-2L	PCR & Seq	CAC GCC ATT TTT GAC TTA CA
NAD7-1U2	Seq	GGC CTC GTT ATC CAC ACT G
NAD7-2L2	Seq	CCT CCC CTG GCT CTT TC
NAD7-1U3	Seq	CAG AAC GCA CGA GGG AAA AC
NAD7-2L3	Seq	CCT TGA CGG GCA CTC CA
NAD7-1 Indel 1	PCR	ACA CTG TAG GTA CGG GGA ATG G
NAD7-2 Indel 1	PCR	TGG GGT GCA GGT GGT GAT A
trnK SNP4/12 AS-GA	PCR	GGG CCA TAA CTG AGG TTT GA
trnK SNP4/12 AS-TC	PCR	GGG CCA TAA CTG AGG TTT TC
RPLTRNP 4/5/A-F	PCR	CCC ACG ATA AAA CAT TTG AAA TA
RPLTRNP 4/5/A-R	PCR	TCT AAG TTT GGA GAT GGA ATG G

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Continued.

Primer name	PCR / Seq[a]	Sequence (5' → 3')
RPLTRNP SNP6-F	PCR	TCT TGC ATT GTA AAT TCT CCT TAT
RPLTRNP SNP6-R	PCR	TTC ATC AGT TCC GAG TTT TTC CTA
a[d]	PCR & Seq	CAT TAC AAA TGC GAT GCT CT
b[d]	PCR & Seq	TCT ACC GAT TTC GCC ATA TC
c[d]	PCR & Seq	CGA AAT CGG TAG ACG CTA G
d[d]	PCR & Seq	GGG GAT AGA GGG ACT TGA C
e[d]	PCR & Seq	GGT TCA AGT CCC TCT ATC CC
f[d]	PCR & Seq	ATT TGA ACT GGT GAC ACG G
TRNT-A1	Seq	TCG ACT AGG GGA GGA TAA TAA CA
TRNL-B1	Seq	CCC CTA GTC GAT TTG GAA GA
TRNL-C1	Seq	ACC TAA AAA GTG GGA ATG TGA TA
TRNL-D1	Seq	ATA TCA CAT TCC CAC TTT TTA GG
TRNL-E1	Seq	TCG CAG TCC ATT TTT TCT CA
TRNL-F1	Seq	TGA GAA AAA ATG GAC TGC GA

[a] Indicates if the primer was used for PCR, sequencing or both.

[b] Universal primers (Demesure *et al.*, 1995).

[c] Consensus primers (Dumolin-Lapegue *et al.*, 1997).

[d] Universal primers (Taberlet *et al.*, 1991).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 3. Conditions of individual polymerase chain reactions.

PCR #	Region(s) amplified	Primer pair(s)	Fragment size[a] (bp)	Reaction volume	Taq units	$MgCl_2$ (mM)	Annealing temperature	Annealing time (sec)	Extension time (sec)	Cycles
1	*nad1 B/C*	nad1 exon B & nad1 exon C[b]	~3200	50 µl	4.0[c]	NA[d]	60-46°C[e]	60	180	50
2	*nad1 B/C*	NAD1-B Indel 2 & NAD1-C Indel 2	182-189	25 µl	1.0	2.5	50°C	30	10	30
3	*nad1 B/C*	NAD1-B Indel 3 & NAD1-C Indel 3	106-112	25 µl	1.0	2.5	47°C	30	10	30
4	*nad1 B/C*	NAD1-B Indel 4 & NAD1-C Indel 4	223-243	25 µl	1.0	2.5	50°C	30	15	31
5	*nad1 B/C*	NAD1-B Indel 5 & NAD1-C Indel 5	93-98	20 µl	1.0	2.5	50°C	30	10	31
6	*nad1 B/C*	NAD1-B Indel 6 & NAD1-C Indel 6	159-165	25 µl	2.0	2.0	55°C	30	15	31
7	*nad1 B/C*	NAD1-B Indel 7 & NAD1-C Indel 7	140-146	20 µl	1.0	2.5	50°C	30	10	31
8	*nad1 B/C*	NAD1-B4 & NAD1 MS2R	91-92	25 µl	0.75	2.0	49°C	60	10	31
9	*nad1 B/C*	NAD1-B2 & NAD1 SNP B-G	91-93	25 µl	1.0	1.5	61°C	25	6	30
	nad7 1/2	NAD7-1 Indel 1 & NAD7-2 Indel 1	148-153							
10	*nad1 B/C*	NAD1-B2 & NAD1 SNP B-T	91-93	25 µl	1.0	1.5	61°C	25	6	30
	nad7 1/2	NAD7-1 Indel 1 & NAD7-2 Indel 1	148-153							
11	*nad1 B/C*	nad1 exon B & NAD1 SNP B-G	~800	25 µl	2.0	2.5	53°C	30	45	31
12	*nad7 1/2*	nad7/1 & nad7/2r[f]	~1100	50 µl	4.0	1-2.0	46°C	60	90	30
13	*nad7 1/2*	NAD7-1U & NAD7-2L	~950	50 µl	4.0	2.0	52°C	60	60	31
14	*nad7 1/2*	NAD7-1 Indel 1 & NAD7-2 Indel 1	148-153	25 µl	1.0	2.5	54°C	30	10	31
15	*trnK*	trnK SNP4/12 AS-GA & matKR5[g]	326	25 µl	1.0	1.5	51°C	30	2	30
	trnK	R/BvsW SSCP-1 & R/BvsW SSCP-1A[g]	119							
16	*trnK*	trnK SNP4/12 AS-TC & matKR5[g]	326	25 µl	1.0	1.5	51°C	30	2	30
		R/BvsW SSCP-1 & R/BvsW SSCP-1A[g]	119							
17	*trnK*	matKF1 & matKR1	2072	25 µl	1.5[c]	NA[d]	51°C	60	135	31
18	*trnK*	matKF2 & matKR5	378	25 µl	1.0	2.5	50°C	60	30	30

19	*rpl33-trnP*	RPLTRNP 4/5/A-F & RPLTRNP 4/5/A-R	141	25 µl	1.0	2.5	49°C	30	10	50
20	*rpl33-trnP*	RPLTRNP SNP6-F & RPLTRNP SNP6-R	95	20 µl	0.8	2.5	49°C	30	10	50
21	*trnT-trnL5'*	a & b[h]	477-478	50 µl	2.0	2.0	50°C	60	30	30
22	*trnL* intron	c & d[h]	562	50 µl	2.0	2.0	53°C	60	30	30
23	*trnL3'-trnF*	e & f[h]	462	50 µl	2.0	2.0	50°C	60	30	30
24	*trnT-trnL5'*	a & b[h]	477-478	25 µl	1.0	2.0	50°C	60	30	30
25	*trnL5'-trnF*	c & f[h]	~1050	25 µl	2.0	2.0	52°C	60	60	30

[a] Amplicon sizes from white, black and red spruce.

[b] (Demesure *et al.*, 1995).

[c] An equal amount of Taq Extender units was also added (see text).

[d] Taq Extender Buffer (1X) contains 2 mM $MgSO_4$.

[e] See text for details of Touchdown PCR profile.

[f] (Dumolin-Lapegue *et al.*, 1997).

[g] See Germano and Klein (1999).

[h] (Taberlet *et al.*, 1991).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4. Geographic origin of samples used for primary sequencing and associated GenBank accession numbers.

Species	Sample	Geographic location	GenBank accession				
			trnK intron[a]	*rpl33-trnP*[a]	*trnT-L-F* spacers/intron	*nad1 B/C*	*nad7 1/2*
White Spruce	64	Durham, NH, USA	AF133923	AF133935		AY057958	AY057967
	494	Black Hills, SD, USA	AF133924	AF133936	AF156807-AF156809	AY057955	AY057968
	SA	Saskatchewan, Canada	AF133925	AF133937			
	VA	AK, USA	AF133926	AF133938			
Black Spruce	63	Durham, NH, USA	AF133919	AF133931			
	3293-2	WI, USA			AF156801-AF156803	AY057953	AY057963
	4274-1	NH, USA	AF133920	AF133932		AY057954	AY057964
	4962-1	Newfoundland, Canada	AF133921	AF133933		AY057956	AY057965
	5004-1	New Brunswick, Canada	AF133922	AF133934		AY057957	AY057966
Red Spruce	2019 21-1	Indian Gap, NC, USA	AF133915	AF133927		AY057949	AY057959
	2022 51-3	October Mtn, MA, USA			AF156798-AF156800	AY057950	AY057960
	2027 29-4	Pillsbury, NH	AF133916	AF133928			
	2032 8-1	Valcartier, Quebec, Canada	AF133917	AF133929		AY057951	AY057961
	2505 38-5	New Brunswick, Canada	AF133918	AF133930		AY057952	AY057962

[a] See Germano and Klein (1999).

Elmer, Emeryville, CA) or the DYEnamic™ ET terminator cycle sequencing premix kit (Amersham Pharmacia Biotech, Inc., Piscataway, NJ). PCR-amplified fragments (both DNA strands) were sequenced using 5-10 pmol primer (Table 2) and 30-90 ng template with 8 µl of the ABI PRISM Ready Reaction mix or the DYEnamic premix. Extension products were separated and analyzed on an ABI PRISM 373 or 377 Automated Sequencer (UNH Sequencing Facility). Sequences were generated with ABI DNA Sequencing Software version 2.1.1, Base caller ABI50, and edited by eye using ABI SeqEd Software version 1.0.3. Sequences were assembled, aligned and analyzed for restriction sites using SeqMan II, MegAlign and MapDraw (DNASTAR). Intron/exon endpoints were set as determined by Gugerli *et al.* (2001b) for *nad1* and by comparison to *Nicotiana sylvestris* for *nad7* (GenBank X86706; Pla *et al.*, 1995).

Screening Individuals for Polymorphisms

In order to screen a large number of individuals for the presence or absence of specific polymorphisms efficiently, several different PCR-based techniques were employed (Tables 5 and 6). Individual restriction digests using *Bsa*BI, *Bsm*AI, *Dra*I, *Hae*III, *Hin*fI, *Hpa*II, *Mse*I, *Sau*3AI (or isoschizomer *Dpn*II) and *Ssp*I (New England Biolabs, Beverly, MA or Promega) were carried out according to the manufacturers' specifications. Control DNA, with the cognate restriction site, was included in all diagnostic restriction digests as a positive control. Sequence-specific PCR or allele-specific PCR (ASPCR) was carried out using primers that differentially annealed to specific sequences (Tables 2 and 3) (Okayama *et al.*, 1989; Sommer *et al.*, 1989). All ASPCR reactions contained an additional non-specific pair of primers as a positive control (Wu *et al.*, 1989; Germano and Klein, 1999). PCR products and/or digested DNA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5. Assays used to score chloroplast polymorphisms and corresponding haplotype definitions.

Detected polymorphisms				Scoring method		Coding definitions	
Name	Location	Mutation[a]	Sequence	PCR #[b]	Assay	Possible results	Code
trnK SNP 1	*trnK* 5'[c]	tv	G / T	other[d]	SSCP	banding pattern 1 / 2[d]	0 / 1
trnK SNPs 2 & 3	*trnK* 5'	tv, tv	AT / TG	other[d]	SSCP	banding pattern 1 / 2[d]	0 / 1
trnK SNP 4	*trnK* 5'	tv	G / T	other[d]	*Dra*I	no cut (G) / cut (T)	0 / 1
trnK SNP 12	*trnK* 5'	tv	A / C	15 & 16	ASPCR	PCR specific for A / C	0 / 1
trnK SNP 10	*matK*	tv	C / A	other[d]	ASPCR	PCR specific for C / A	0 / 1
					*Ssp*I	no cut (C) / cut (A)	0 / 1
trnK SNP 14	*matK*	tv	A / C	17	Sequence	A / C	0 / 1
trnK SNP 11	*trnK* 5'	tv	G / T	17	*Ssp*I	no cut (G) / cut (T)	0 / 1
trnK SNP 13	*matK*	ts	T / C	18	*Hin*fI	no cut (T) / cut (C)	0 / 1
rpl-trnP SNP 3	*psaJ-trnP*	ts	A / G	other[d]	*Msp*I	no cut (A) / cut (G)	0 / 1
rpl-trnP SNPs 4 & 5	*rpl*33-*psaJ*	tv, tv	AG / TT	19	SSCP	banding pattern 3 / 4[e]	0 / 1
rpl-trnP SNP 6	*psaJ-trnP*	ts	C / T	20	SSCP	banding pattern 5 / 6[e]	0 / 1
rpl-trnP SNPs 7 & 8	*rpl*33-*psaJ*	tv, ts	AC / TT	19	SSCP	banding pattern 3 / 7[e]	0 / 1
rpl-trnP Indel A	*rpl*33-*psaJ*	SSR	C_n	19	length[f]	n = 8 / 9 / 10	0 / 1 / 2
trnT-L-F SNP 3	*trnT-trnL*	tv	T / G	24	*Bsa*BI	no cut (T) / cut (G)	0 / 1
trnT-L-F SNP 5	*trnL-trnF*	tv	G / T	25	*Mse*I	no cut (G) / cut (T)	0 / 1

Chloroplast haplotypes	*cp-1*	*cp-1a*	*cp-1b*	*cp-1c*	*cp-2*	*cp-2a*	*cp-2b*	*cp-2c*	*cp-2d*	*cp-2e*	*cp-2f*	*cp-2g*	*cp-3*	*cp-3a*	*cp-3b*	*cp-3c*	*cp-3d*	*cp-3e*	*cp-3f*
trnK SNP 1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
trnK SNPs 2 & 3	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
trnK SNP 4	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

trnK SNP 12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1
trnK SNP 10	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0
trnK SNP 14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
trnK SNP 11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
trnK SNP 13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1
rpl-trnP SNP 3	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
rpl-trnP SNPs 4 & 5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
rpl-trnP SNP 6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
rpl-trnP SNPs 7 & 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
rpl-trnP Indel A	1	1	0	2	1	1	1	1	1	2	0	0	1	1	1	1	1	1	2
trnT-L-F SNP 3	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
trnT-L-F SNP 5	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0

[a] Types of mutations detected: transversions (tv), transitions (ts) and simple sequence repeats (SSR). In the cases of trnK SNPs 2 & 3, rpl-trnP SNPs 4 & 5, and rpl-trnP SNPs 7 & 8, substitutions are directly adjacent to one another.

[b] Refer to Table 3.

[c] Located in the *trnK* intron 5' of *matK*.

[d] Described in Germano and Klein (1999).

[e] Clear differences in SSCP banding patterns were observed (data not shown) and assigned numbers 1-7.

[f] Length differences of one base pair were detected with high-resolution agarose gel electrophoresis after shortening the amplicon with *Dpn*II digestion; results were confirmed with SSCP.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 6. Assays used to score mitochondrial polymorphisms and corresponding haplotype definitions.

Detected polymorphisms				Scoring method		Coding definitions	
Name	Location	Mutation[a]	Sequence	PCR #[b]	Assay	Possible results	Code
nad7 Indel 1	*nad7 1/2*	VNTR	$GCCCG_n$	14	length	n = 1 / 2	0 / 1
nad1 Indel 2	*nad1 B/C*	indel	CGATA	2	length	absent / present	0 / 1
nad1 Indel 3	*nad1 B/C*	VNTR	$AATTGT_n$	3	length	n = 1 / 2	0 / 1
nad1 Indel 4	*nad1 B/C*	VNTR	$CCTTCAATGAGTGGACTAAA_n$	4	length	n = 1 / 2	0 / 1
nad1 Indel 5	*nad1 B/C*	indel	GACCA	5	length	absent / present	0 / 1
nad1 Indel 6	*nad1 B/C*	VNTR	$TATTTA_n$	6	length	n = 1 / 2	0 / 1
nad1 Indel 7	*nad1 B/C*	VNTR	$TAATAT_n$	7	length	n = 1 / 2	0 / 1
nad1 Indel 8	*nad1 B/C*	SSR	C_n	9 or 10	length[c]	n = 9 / 10 / 11	0 / 1 / 2
nad1 Indel 9	*nad1 B/C*	SSR	A_n	8	length[d]	n = 2 / 3	0 / 1
nad1 Indel 10	*nad1 B/C*	VNTR	$GATGCGGACGAGCCATCCTTCCTCGCCC_n$	4	length	n = 1 / 2 / 5 / 6 / 7	0/1/4/5/6
nad1 SNP A	*nad1 B*	tv	A / C	11	*Bsm*AI	no cut (A) / cut (C)	0 / 1
nad1 SNP B	*nad1 B/C*	tv	G / T	9 & 10	ASPCR	PCR specific for G/T	0 / 1
nad1 SNP C	*nad1 B/C*	tv	G / T	7	*Sau*3AI	no cut (G) / cut (T)	0 / 1
nad1 SNP D	*nad1 B/C*	ts	C / T	6	*Hpa*II	no cut (C) / cut (T)	0 / 1
nad1 SNP E	*nad1 B/C*	tv	A / C	6	*Dra*I	no cut (A) / cut (C)	0 / 1

Mitochondrial haplotype	*mt-1*	*mt-1a*	*mt-1b*	*mt-1c*	*mt-1d*	*mt-1e*	*mt-2*	*mt-3*
nad7 Indel 1	0	0	0	0	0	0	0	1
nad1 Indel 2	1	1	1	1	1	1	1	0
nad1 Indel 3	0	0	0	0	0	0	0	1
nad1 Indel 4	1	1	1	1	1	1	0	0
nad1 Indel 5	0	0	0	0	0	0	1	1
nad1 Indel 6	0	0	0	0	0	0	1	1

nad1 Indel 7	0	0	0	0	0	0	1	0
nad1 Indel 8	2	1	2	2	2	2	0	0
nad1 Indel 9	0	0	0	0	0	0	1	1
nad1 Indel 10	0	0	1	4	5	6	0	0
nad1 SNP A	0	0	0	0	0	0	1	1
nad1 SNP B	0	0	0	0	0	0	0	1
nad1 SNP C	1	1	1	1	1	1	0	0
nad1 SNP D	0	0	0	0	0	0	1	1
nad1 SNP E	1	1	1	1	1	1	0	0

[a] Types of mutations detected: variable number tandem repeats (VNTR), insertion/deletions (indel), simple sequence repeats (SSR), transversions (tv) and transitions (ts).

[b] Refer to Table 3.

[c] Amplicons were shortened by digestion with *Hpa* II in order to resolve fragments differing by one base pair.

[d] Amplicons were shortened by digestion with *Hae* III in order to resolve fragments differing by one base pair.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fragments were separated through 2% agarose or 4-5% MetaPhor® agarose (BioWhittaker Molecular Applications [BMA], Rockland, ME), depending on the degree of resolution required.

Single strand conformation polymorphism (SSCP) was used to survey populations for the presence or absence of some polymorphisms (Sheffield *et al.*, 1993). SSCP analysis was carried out according to the protocol for MDE™ Gel (BMA). One or two µl (3-6 ng) PCR product were combined with 10 µl stop solution, incubated at 95°C for 2 minutes and then chilled on ice for 5 minutes. The entire denatured sample was loaded onto a 0.75X MDE gel. Gels were stained with 2X GelStar® nucleic acid stain (BMA) in 50% glycerol for 15 minutes and visualized under UV light. Additional DNA sequencing was done as a control to determine the nature of novel SSCP banding patterns, and also as a check when restriction digestion was negative at potential single nucleotide polymorphism (SNP) sites.

Phylogenetic Analysis

Chloroplast and mitochondrial haplotypes were designated as each unique combination of polymorphisms detected within any one individual. Haplotypes were coded as series of ones and zeroes (with additional numerals in cases of polyallelic loci). Dinucleotide substitutions and insertion/deletions (indels) were considered as single mutational events (see Laroche *et al.*, 2000; Averof *et al.*, 2000). Phylograms were generated with Phylogenetic Analysis Using Parsimony (PAUP version 4.10b; Swofford, 1998) under the criterion of minimum evolution using neighbor-joining search with 1000 bootstrap replicates. Character types were unordered with the exception of the variable number tandem repeat (VNTR) nad1 Indel 10, assuming the stepwise mutation model (as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

reviewed in Jarne and Lagoda, 1996). Since simple sequence repeats (SSRs) are more likely to display convergent evolution (Jarne and Lagoda, 1996), the SSRs rpl-trnP Indel A and nad1 B/C Indel 8 were assigned weights equal to half those of all the other characters. *Picea schrenkiana* was used as an outgroup; the sample was collected from the arboretum of the Ministry of Forests, British Columbia, in Vernon B.C. The mitochondrial haplotype of *P. schrenkiana* was determined by directly sequencing the *nad1 B/C* and *nad7 1/2* introns (GenBank AY153782 and AY169718). Its chloroplast haplotype was determined by sequencing the *trnK* intron and the *rpl33-psaJ-trnP* region (GenBank AY035204 and AY115679) and by indirectly screening for the substitutions within the *trnT-trnL-trnF* regions by the methods described above.

Estimation of Sequence Divergence Rates and Substitution to Indel Ratios

An average number of substitutions per site was estimated for the chloroplast *trnK* intron (including *matK*) from five conifer pairs by dividing the number of pairwise substitutions by the number of compared nucleotides. The sequences used were those of white spruce (GenBank AF059341), Himalayan spruce (*Picea smithiana*; GenBank AF143429), jack pine (*Pinus banksiana*; GenBank AF143427) and Armand pine (*Pinus armandii*; GenBank AF143428), with pairwise comparisons between white spruce and each pine, between Himalayan spruce and each pine, and between the two pines. The sequence evolution rate was estimated by dividing the average per-site divergence by the time (geological date) since cladogenesis. The divergence of *Picea* from *Pinus* and the split between the *Pinus* subgenera, *Strobus* and *Pinus*, (each occurring 70-100 million years ago) were estimated from the fossil record (Miller, Jr., 1977; Aldén, 1987). The estimated range of time since the divergence of black and red spruce was interpolated

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

from the estimated *trnK* sequence evolution rates using the minimum and maximum number of *trnK* substitutions detected between any two individuals. The number of substitutions per site (K_o) and the number of indels per site (I; Laroche *et al.*, 2000) were estimated for *nad1 B/C* by counting polymorphisms between pairwise alignments of white, black and red spruce. Since multiple mitochondrial haplotypes were detected in white and black spruce, polymorphisms of the most common haplotypes were used for the calculations. The ratios of substitutions to indels (Ko/I; Laroche *et al.*, 2000) were subsequently estimated.

Population Genetic Analyses

Numbers of observed haplotypes were used to calculate haplotype frequencies, number of haplotypes per species (A_s) and average number of haplotypes per population (A_p; Hamrick and Godt, 1990). Analyses for genetic diversity and differentiation specific for haploid loci were carried out as specified by Pons and Petit (1995). Haplotypes were treated as alleles at a single genetic locus. Diversity within each population (h_k), average within-population diversity (h_S), total diversity (h_T) and degree of differentiation (G_{ST}) were calculated according to Pons and Petit (1995) using the authors' program Haplodiv for Windows. Data from populations represented by two or fewer individuals were excluded from these analyses.

Geographical Correlations

To test the significance of clinal patterns, linear regression analysis of geographical coordinates versus haplotype frequencies per population was carried out using SYSTAT Version 10 (SPSS Inc., 2000). Analysis of molecular variance (AMOVA) was conducted with Arlequin version 2.000 (Schneider *et al.*, 2000).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hierarchical groups for red spruce were designated geographically such that its northeastern populations were grouped separately from its southwestern populations. This was done in order to test for north/south trends in genetic diversity observed by Hawley and DeHayes (1994). The ranges of black and white spruce were divided at 95°W (see Wilkinson *et al.*, 1971; Chang and Hanover, 1991; Furnier and Stine, 1995). Black and white spruce populations were also divided among four of the climatic/forest regions of Canada defined by Rowe (1972). These included Acadia, Great Lakes-St. Lawrence, Boreal, and Grasslands; populations located in the United States were assigned to the nearest Canadian region (Morgenstern, 1978). Fixation indices (F_{ST}, F_{SC}, F_{CT}) for defined regions were tested for significance by running 3000 permutations. The fixation index F_{ST} (Weir and Cockerham, 1984) was calculated for each individual species in AMOVA by assigning all populations of a species to one group.

Results

The focus of this project was to assess the amount and geographic distribution of genetic variation within the oppositely inherited cytoplasmic genomes of white, black and red spruce. Since these genomes are haploid and uniparentally inherited, polymorphisms within an organelle genome are linked. Loci were chosen that were expected to display relatively high variability. Mitochondrial regions examined were the *nad1 B/C* and *nad7 1/2* introns. Chloroplast regions included the *trnK* intron (encompassing the protein-coding gene *matK*), spacers between *rpl33*, *psaJ* and *trnP*, spacers between *trnT*, *trnL* and *trnF*, and the *trnL* intron. *TrnK* and *rpl33-psaJ-trnP* sequences, and some of the variation found therein, have been previously described (Germano and Klein, 1999).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Summary of Polymorphisms Initially Detected

Sequencing of *trnT-trnL-trnF* from one individual per species revealed a total of one indel and seven base substitutions (or single nucleotide polymorphisms; SNPs) including one transition and six transversions. The *trnT-trnL* spacer was 414-415 bp and contained three polymorphisms that distinguished white spruce from the other two species and one polymorphism that distinguished black spruce. The 489-bp *trnL* intron contained one white spruce-distinguishing substitution, while the 380-bp *trnL-trnF* spacer revealed three base substitutions, two separating white and one differentiating red spruce.

The mitochondrial intron *nad1 B/C* ranged in size from 2983 - 3160 bp; comparison of the initial sequences revealed a total of nine indels and five base substitutions (1 transition, 4 transversions). Consensus primers (Dumolin-Lapegue *et al.*, 1997) worked poorly to amplify *nad7 I/2* in white and red spruce, therefore PCR primers NAD7-1U and NAD7-2L were designed that annealed just inside the exons (Table 2). The *nad7 I/2* intron was 945 bp in black spruce and only a 5-bp insertion was detected that separated red spruce from the other two species. The ratios of substitutions to indels (Ko/I; Laroche *et al.*, 2000) in the *nad1 B/C* intron (strictly) were 4/7 between white and red spruce, 3/6 between white and black spruce, and 1/3 between red and black spruce; giving a range of 0.33-0.57 among the species.

Screening Populations for Identified Markers and Defining Haplotypes

The presence/absence of every identified mitochondrial and selected chloroplast markers were scored in range-wide population samples using PCR-based techniques (Tables 5 and 6). The results of screening a subset of the populations for markers trnK SNP 1, trnK SNPs 2&3, trnK SNP 4, matK SNP 10 and rpl-trnP SNP 3 have previously

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

been reported (Germano and Klein, 1999). During the scoring process, additional intraspecific polymorphisms were detected as novel patterns when carrying out SSCP, restriction analysis or length assays; the molecular nature of these polymorphisms was determined by sequencing. These included a variable number tandem repeat (VNTR; nad1 Indel 10), three base substitutions (trnK SNPs 11, 12 and 13) and one dinucleotide substitution (rpl-trnP SNPs 7&8) (Tables 5 and 6).

Haplotypes were defined as each unique combination of polymorphisms found within an individual. A total of nineteen chloroplast haplotypes were defined based on the presence or absence of 15 polymorphisms: eight in the *trnK* intron, five in the *rpl33-psaJ-trnP* region, and two in the *trnT-trnL-trnF* region (Table 5). Eight mitochondrial haplotypes were identified by the presence/absence profiles of 15 markers: one in *nad7 1/2* and the rest in *nad1 B/C* (Table 6). The three most common mitochondrial (*mt-*) and chloroplast (*cp-*) haplotypes were numbered according to the species in which each was most frequently detected (1=white; 2=black; 3=red spruce). Variants of these haplotypes were designated by assigning letters (*a-g*) as suffixes to the numeric code.

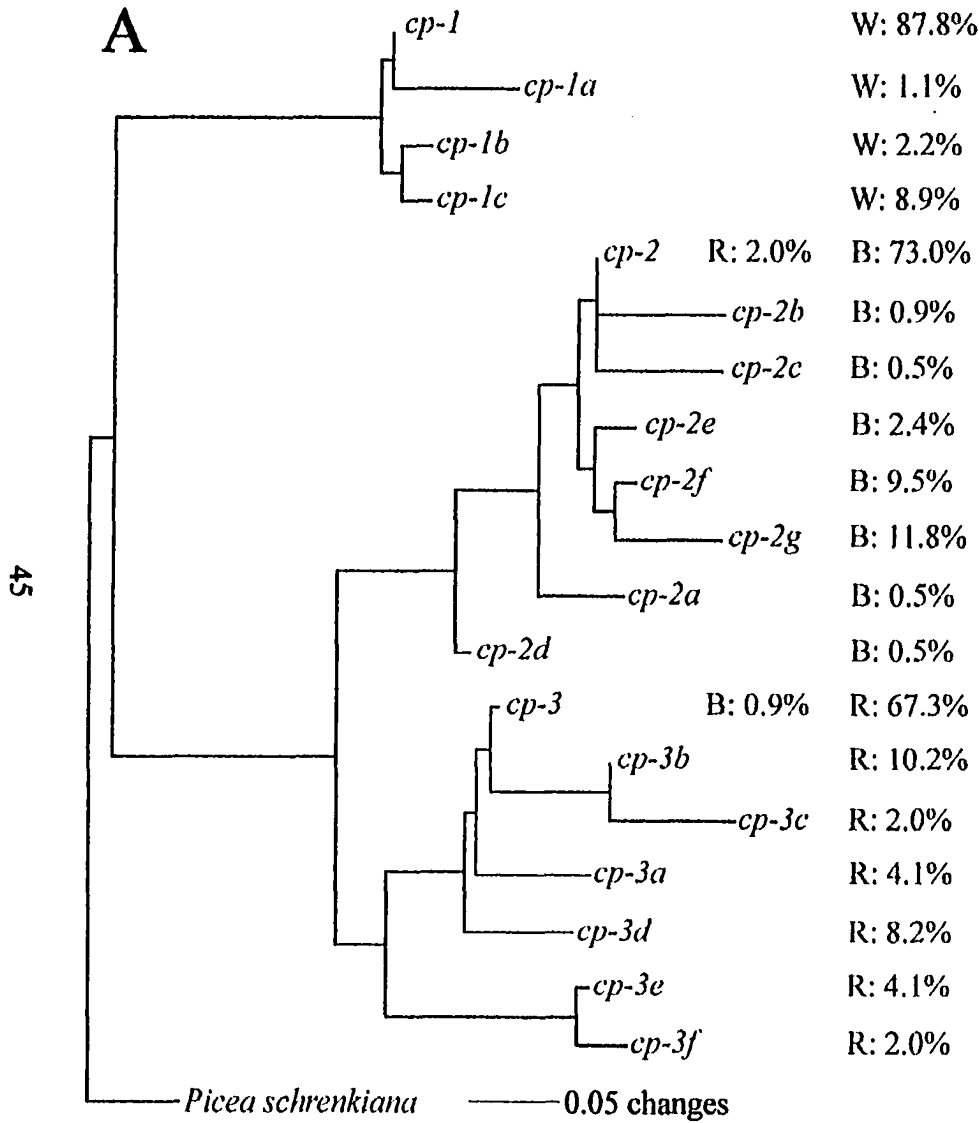
Relationships and Frequency of Haplotypes

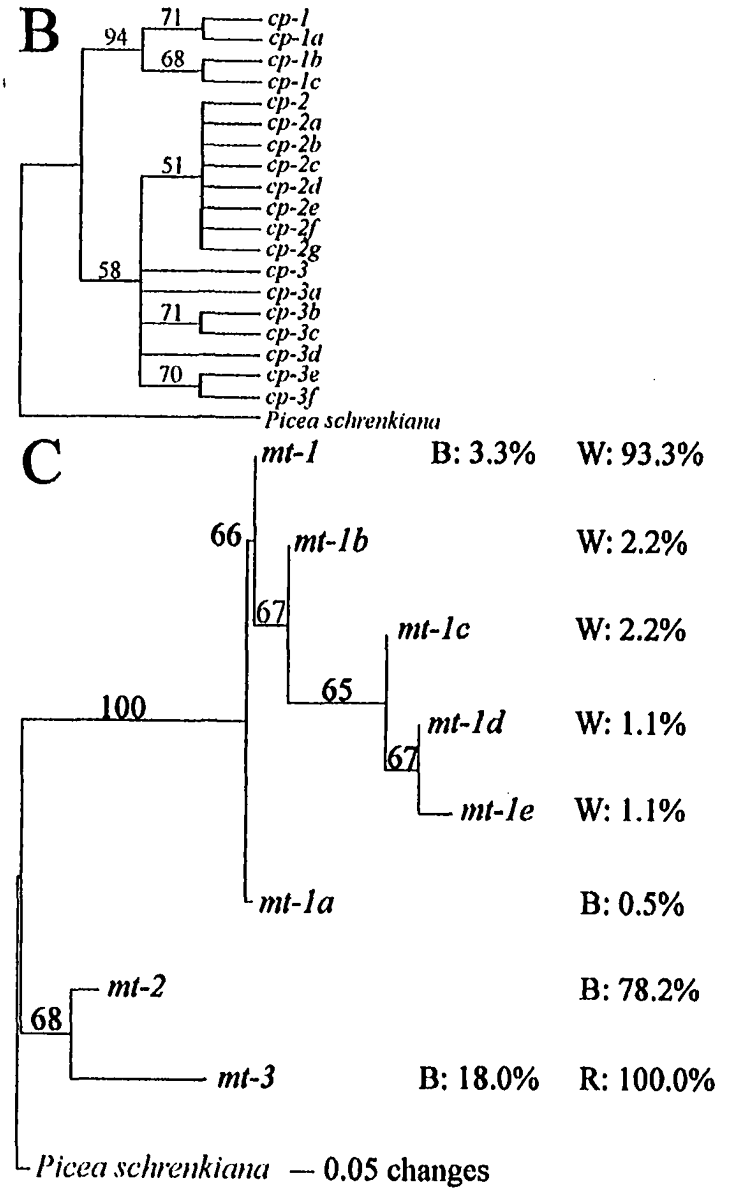
Phylogenetic relationships among the haplotypes and haplotype frequencies within each species are presented in Figure 1. Three lineages are evident as neighbor-joining analysis distributed the 19 chloroplast haplotypes among three monophyletic groups. Given the limited character set, much resolution was lost upon bootstrapping (Figure 1B), however the major clades remained intact. With the exception of three individuals, each species contained chloroplast haplotypes of only one lineage. White spruce contained only chloroplast haplotypes of the *cp-1* lineage, and these haplotypes

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1. Phylograms showing relationships of haplotypes. Gene trees were generated with PAUP using distance analysis and *Picea schrenkiana* as the outgroup. (A) Chloroplast haplotypes, neighbor joining search. (B) Chloroplast haplotypes, neighbor joining search with bootstrap. (C) Mitochondrial haplotypes, neighbor joining search with bootstrap. Each haplotype's frequency is listed to its right; frequencies are in white spruce (W:), black spruce (B:) and red spruce (R:).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.





Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were not detected in the other two species. *Cp-2* and its seven variants were found almost strictly in black spruce, with the exception of *cp-2* present in one red spruce. *Cp-3* and its six variants were typical of red spruce except that *cp-3* was also observed in two black spruce samples.

Ten, thirteen and five polymorphisms separate haplotypes *mt-1* and *mt-2*, *mt-1* and *mt-3*, and *mt-2* and *mt-3*, respectively. *Mt-2* and *mt-3* are more closely related to each other than they are to *mt-1*; *mt-1* and its variants (*mt-1a* through *e*) form a monophyletic group (Figure 1C). No variants of *mt-2* or *mt-3* were detected. White spruce contained only *mt-1* and four of its variants, with *mt-1* being the most common. *Mt-3* appeared to be fixed in red spruce as it was the only haplotype detected in the species. The most frequent mitochondrial haplotype in black spruce was *mt-2* (78.2%), and this haplotype was specific to black spruce. Interestingly, there were 38 black spruce individuals (18.0%) displaying *mt-3*, seven individuals (3.3%) with *mt-1*, and one individual with *mt-1a*.

***TrnK*, Divergence Rates Among Spruce**

Five pairwise comparisons of GenBank sequences were made to estimate *trnK* divergence. These included white spruce and jack pine, 119 substitutions/2431 bp; white spruce and Armand pine, 99 substitutions/2038 bp; Himalayan spruce and jack pine, 119 substitutions/2426 bp; Himalayan spruce and Armand pine, 98 substitutions/1976 bp; and jack pine and Armand pine, 75 substitutions/1987 bp. These species were chosen because they represent spruces from North America (white spruce) and Eurasia (Himalayan spruce) and pines from each of the two *Pinus* subgenera, *Strobus* (Armand pine) and *Pinus* (jack pine). The times since cladogenesis of *Picea* from *Pinus* and of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pinus subgenera were estimated from the fossil record to have both occurred during the late Cretaceous (100-70 million years ago; Miller, Jr., 1977; Aldén, 1987). Both ends of this time range (100 and 70 million years) were used to estimate average minimum and maximum rates of chloroplast *trnK* intron sequence evolution of 4.7×10^{-10} and 6.7×10^{-10} substitutions/(site·yr), respectively. Estimated times since the divergence of black and red spruce were then interpolated from the sequence evolution rates using the minimum and maximum number of *trnK* substitutions detected between any two individuals (e.g. 1 *trnK* substitution between *cp-2* and *cp-3*; 4 *trnK* substitutions between *cp-2a* and *cp-3f*) per number of compared nucleotides (2476 bp). Therefore four estimates of the time since the divergence of red and black spruce from their common ancestor were made; they ranged between approximately 0.6 and 3.5 million years ago.

Genetic Differentiation and Geographical Distribution of Haplotypes

Values of genetic diversity and differentiation are presented in Table 7 and the geographic distributions of haplotypes are displayed in Figure 2. All three species displayed more genetic variation in the chloroplast sequences than in the mitochondrial introns. Red spruce displayed the highest total chloroplast diversity ($h_T = 0.52$) of the three species, yet no mitochondrial variation was detected within it. White spruce exhibited lower diversity than black spruce in both genomes. Genetic structure among populations was not detected in any of the species for the chloroplast haplotypes (G_{ST} and F_{ST} were close to zero). Significant structure of mitochondrial haplotypes, however, was observed among populations of white and black spruce, which had F_{ST} values of 0.49 and 0.59, respectively ($P < 0.001$). Reasons for this structure are proposed in the Discussion. All of the mitochondrial variation identified within white spruce (haplotypes *mt-1b*, *-1c*,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 7. Genetic diversity and differentiation.

	White spruce	Black spruce	Red spruce
n^a	89	208	49
k	22	53	11
n/k	4-5	3-5	3-6
Mean n/k (SD)	4.05 (0.21)	3.92 (0.81)	4.45 (0.93)
Chloroplast			
A_S	4	9	8
A_P	1-3	1-4	1-3
Mean A_P (SD)	1.36 (0.58)	1.87 (0.76)	2.36 (0.81)
h_S (SE)	0.18 (0.06)	0.42 (0.04)	0.56 (0.09)
h_T (SE)	0.20 (0.06)	0.44 (0.04)	0.52 (0.06)
G_{ST} (SE)	0.10 (NC)	0.05 (0.05)	−0.06 (0.05)
F_{ST}	0.09	0.07	−0.02
Mitochondria			
A_S	5	4	1
A_P	1-3	1-2	1
Mean A_P (SD)	1.14 (0.47)	1.15 (0.36)	1.00 (0.00)
h_S (SE)	0.07 (0.05)	0.15 (0.04)	0.00
h_T (SE)	0.13 (0.09)	0.36 (0.06)	0.00
G_{ST} (SE)	0.49 (NC)	0.58 (0.09)	NA
F_{ST}	0.49***	0.59***	NA

Abbreviations: *n*, total number of trees used in diversity analyses; *k*, total number of populations they represent; *n/k*, number of trees per population; SD, standard deviation; SE, standard error; A_S, number of haplotypes detected per species (Hamrick and Godt, 1990); A_P, number of haplotypes detected per population (Hamrick and Godt, 1990); h_S, average within-population diversity; h_T, total within-species diversity; G_{ST}, differentiation parameter (Pons and Petit, 1995); F_{ST}, fixation index (Weir and Cockerham, 1984); NC, not calculated; NA, not applicable; *** $P < 0.001$.

[a] Data from populations with two or less individuals were excluded from the diversity analyses (one white spruce and three black spruce individuals).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

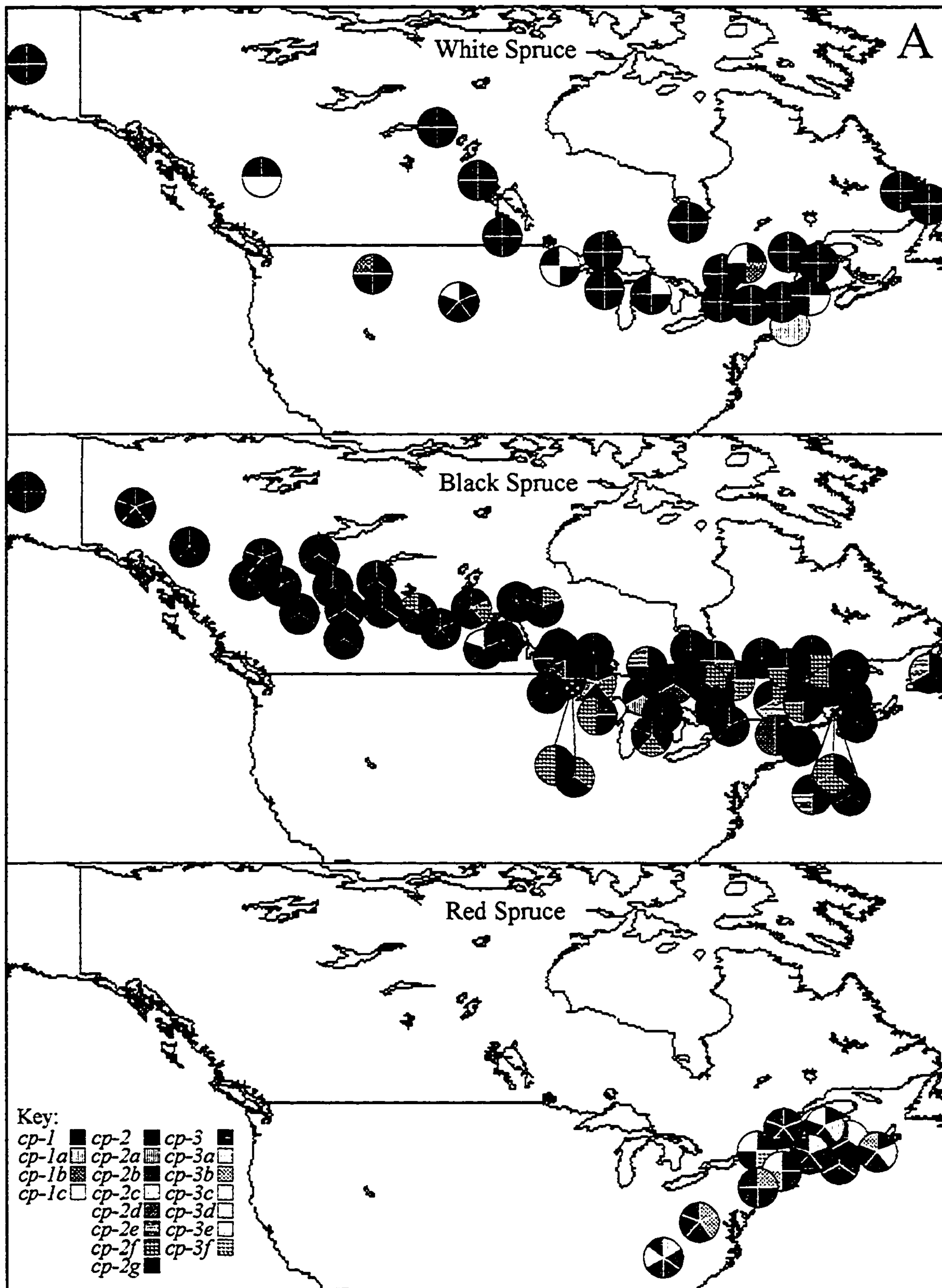


Figure 2A. Geographic locations of populations and their included chloroplast haplotypes coded according to the key shown. Each circle represents one population; circle divisions represent the number of individual trees scored in each population.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

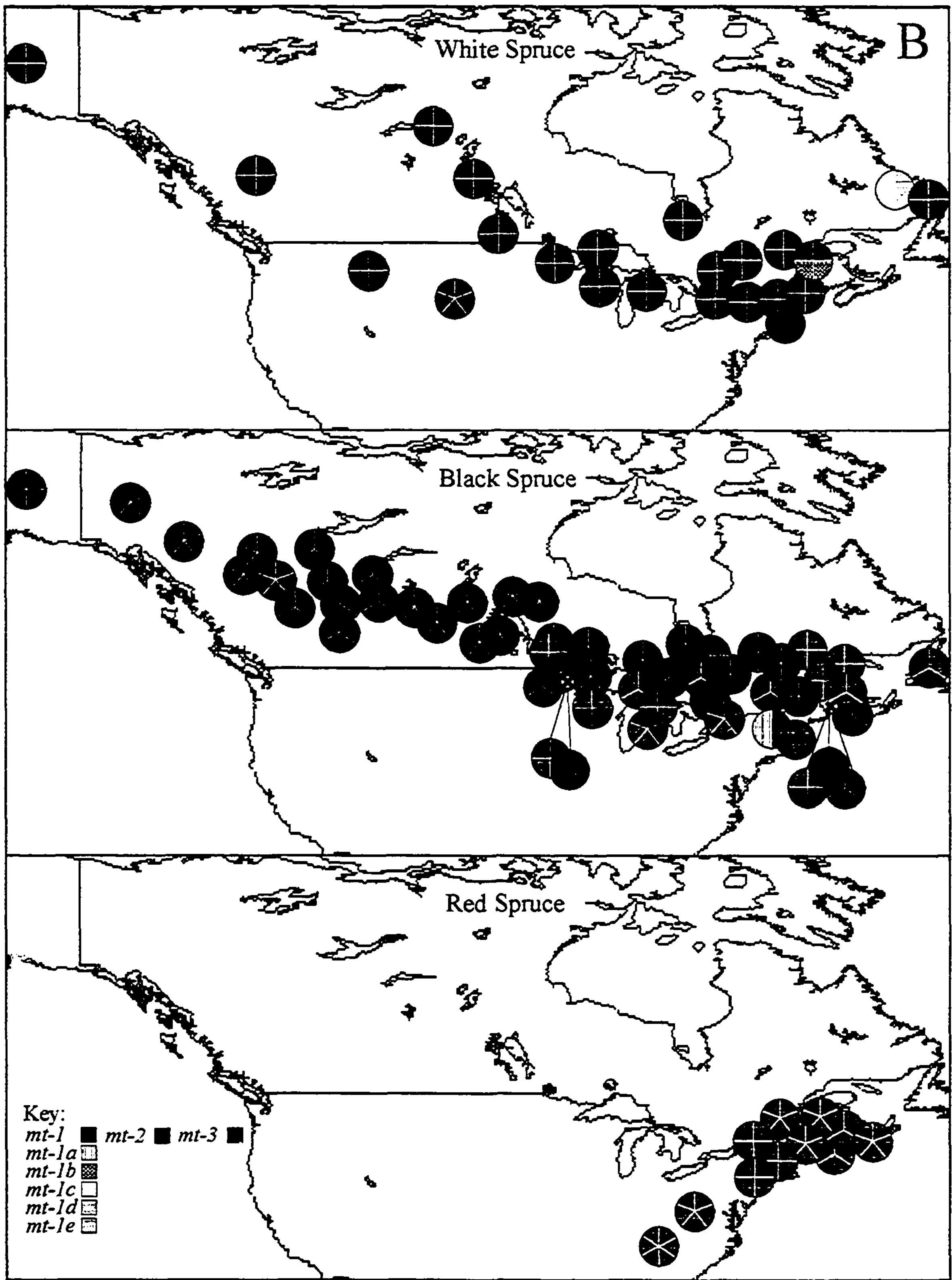


Figure 2B. Geographic locations of populations and their included mitochondrial haplotypes coded according to the key shown. The positions of individual trees within populations (i.e. divisions within circles) correspond between A and B.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

-1d and *-1e*) was confined to two of its northeastern populations (Figure 2B).

The chloroplast haplotypes in red and white spruce appeared to be randomly scattered across their ranges (Figure 2A); no significant differences were found between hierarchal groups tested with analysis of molecular variation (AMOVA) in either of these species (Table 8). Black spruce demonstrated geographic patterns of the distribution of its chloroplast and mitochondrial haplotypes, with much of its variation residing around the Great Lakes region and in the eastern half of its range (Figure 2B). This is supported by AMOVA, which indicated that 3.2% of the total chloroplast variance ($P < 0.04$) and 18.2% of the total mitochondrial variance ($P < 0.001$) in black spruce were due to variation between the eastern and western halves of its range. The chloroplast haplotypes *cp-2e* and *cp-2f* were principally confined to the east; *cp-2e* was absent and *cp-2f* was present in only two individuals in the western half the black spruce range. Linear regression analysis indicated that the frequency of *cp-2f* was clinal ($P < 0.04$), with a distribution decreasing from southeast to northwest. *Cp-2g* was absent from the eastern quarter of the black spruce range, and its distribution was clinal ($P < 0.001$) in the opposite direction. Mitochondrial haplotype *mt-3* was only present in the eastern half of the black spruce range. Its frequency was clinal ($P < 0.001$), decreasing from southeast to northwest. Furthermore, AMOVA indicated that 10.8% of the total mitochondrial variance in black spruce, which was due mostly to the presence of *mt-3*, was assigned to variation between climatic regions ($P < 0.02$; Table 8). The other mitochondrial haplotypes detected in black spruce, *mt-1* and *mt-1a*, were detected in four different populations dispersed across its range, with one population in the west appearing to be fixed for *mt-1* (Figure 2B).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 8. Results of analyses of molecular variance (AMOVA).

Groupings of regions	Chloroplast d.f.	Chloroplast SS	Chloroplast Variance (% of total)	Chloroplast F_{CT}	Mitochondria SS	Mitochondria Variance (% of total)	Mitochondria F_{CT}
White spruce							
east/west	1	0.23	0.00 (2.3%)	0.02	0.15	0.00 (0.0%)	0.01
Acadia/Boreal/Grasslands/ Great Lakes-St. Lawrence	3	0.52	0.00 (0.2%)	0.00	0.52	0.00 (0.2%)	0.00
Black spruce							
east/west	1	1.51	0.01 (3.2%)	0.03*	24.16	0.21 (18.2%)	0.18***
Acadia/Boreal/ Great Lakes-St. Lawrence	2	1.04	0.00 (0.7%)	0.01	17.4	0.12 (10.8%)	0.11*
Red spruce							
northeast/southwest	1	0.55	0.00 (0.9%)	0.07	NA		

Abbreviations: d.f., degrees of freedom; SS, sum of squares; F_{CT}, fixation index among groups (Schneider *et al.*, 2000); * $P < 0.05$; *** $P < 0.001$; NA, not applicable.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Discussion

Between- and Within-Species Variation

Chloroplast sequence divergences among white, black and red spruce ranged from 0.0 to 0.4 %, while the mitochondrial regions displayed levels of 0.0 to 0.2% divergences between species. The mitochondrial *nad1 B/C* intron, while containing many more polymorphisms than *nad7 1/2*, displayed low levels of variation between the three species. Such levels of divergence are remarkably low when compared to divergences among species of its sister genus *Pinus*. The *trnL* intron, for example, displays up to 10 times more variation between pine species than between spruce species. The low levels of sequence divergence in spruce are likely not due to slow rates of mutation (i.e. differences in DNA replication and repair mechanisms), because there are low levels of between-species variation (relative to *Pinus*) in both cytoplasmic genomes, and also in the ribosomal internal transcribed spacer DNA of the nuclear genome (Germano and Klein, 1999). They are more likely attributed to the relatively young age of the spruce genus. This is consistent with Sigurgeirsson (1992), who estimated from chloroplast RFLPs that the spruces diverged from their common ancestor long after the radiation of the pines took place.

The chloroplast genome displayed significantly more intraspecific variation and hence higher total chloroplast genetic diversity within species than the mitochondrial genome. Eleven of the chloroplast sites varied within species (Table 5) compared to only two of the screened mitochondrial sites (Table 6; assuming that the presence of *mt-1* and *mt-3* in black spruce is due to interspecific hybridization, see below). Although additional cryptic variation for these organelle regions may have gone undetected with

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

our screening methods, the pattern of greater variation of chloroplast DNA versus mitochondrial DNA is accurate since the same numbers of sites were screened using the same methods for both organelles. The difference in levels of observed intraspecific variation for each of the genomes in spruce is consistent with the slow rate of primary sequence changes of mitochondrial DNA relative to chloroplast DNA in plants (Palmer, 1992). This observation is also consistent with findings of higher evolutionary rates in gymnosperm organelle DNA that is inherited paternally than in DNA inherited maternally (Whittle and Johnston, 2002).

Nature of Mutations in the Cytoplasmic Genomes

The frequencies of different polymorphism types (e.g. indels, base substitutions, and microsatellites) varied significantly between the two organelle genomes. All of the detected chloroplast polymorphisms were base substitutions except for one single-nucleotide SSR (Table 5). Conversely, only a third of the mitochondrial markers were substitutions, the remaining two thirds being indels (Table 6). Most of the observed indels were repetitive in nature, an apparently common feature of *nad1 B/C* in the Pinaceae (Grivet *et al.*, 1999; Mitton *et al.*, 2000b; Gugerli *et al.*, 2001a). Interestingly however, the only VNTR that was present with greater than two repeat units long was nad1 Indel 10. It had a unit length of 28 bp and was repeated up to six times (Table 6) in two populations of white spruce (Figure 2B). Furthermore, the tandem repeat array of nad1 Indel 10 was flanked by short (8 bp) direct repeats, a pattern observed in the mitochondrial DNA of Norway spruce (Sperisen *et al.*, 2001), sugar beet, wheat and *Arabidopsis* (Nishizawa *et al.*, 2000). The ratios of *nad1 B/C* substitutions to indels (K_s/I) of 0.33-0.57 among white, black and red spruce were determined by counting the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

number of *nad1 B/C* polymorphisms between the species in a pairwise fashion, using the most common haplotypes within species. The ratios are comparable to Norway and Siberian spruce (*Picea obovata*) (*nad1 B/C* K_o/I = 1.0-2.0), however they are strikingly lower than *Pinus cembra* and *Pinus pumila* (K_o/I = 16) and certain angiosperms (K_o/I = 6.2-6.7) (Sperisen *et al.*, 1998; Laroche *et al.*, 2000; Sperisen *et al.*, 2001; Gugerli *et al.*, 2001a).

Genetic Diversity, Structure and Gene Flow within Each Species

Chloroplast haplotypes were distributed throughout each species' range whereas mitochondrial haplotypes appeared to be clustered together within populations (Figure 2). Therefore fixation indices G_{ST} and F_{ST} were effectively zero for the chloroplast yet significantly above zero for the mitochondria (Table 7). This is consistent with pollen being the main agent of gene flow among wind-pollinated plant populations (Ennos, 1994; Latta *et al.*, 1998). On average, wind-borne pollen can travel 10-100 km (Delcourt and Delcourt, 1987b), whereas spruce seeds are maximally dispersed 100 m from the maternal tree without physical assistance other than wind (Viereck and Johnston, 1990; Blum, 1990).

Detected levels of chloroplast genetic diversity were unexpectedly low in white spruce and unexpectedly high in red spruce (Table 7) compared to previous observations with other markers. For example, the average observed heterozygosity of nuclear STS markers in white spruce were reported to be 0.37 (Perry and Bousquet, 1998), which was over three times that of black spruce and over five times that of red spruce (Perron *et al.*, 2000). Furthermore, observed allozyme heterozygosities for red and white spruce were 0.075 (Hawley and DeHayes, 1994) and 0.306 (Furnier *et al.*, 1991), respectively. In

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fact, multiple studies have suggested that red spruce harbors less diversity than many other northeastern forest species (Eckert, 1989; Hawley and DeHayes, 1994). In the present study, not only did red spruce display significantly more chloroplast diversity than white spruce, but as much, if not more diversity than black spruce. The significance of this is magnified when considering the narrow geographic range of red spruce, which was represented by 49 individuals, equaling approximately one half the white spruce and one quarter of the black spruce trees sampled in this study.

Interspecific Hybridization and Introgression

The chloroplast haplotypes in each of the three lineages were nearly 100% specific to their respective species (Figure 1A). The only exceptions were one red spruce individual possessing *cp-2* and two black spruce individuals containing *cp-3*. These presumably represent introgressed individuals. The mitochondrial haplotypes on the other hand, were not as strictly species-specific. While white spruce displayed only the haplotypes of the *mt-1* lineage, and red spruce was monomorphic for *mt-3*, black spruce possessed one unique haplotype, *mt-2*, as well as haplotypes *mt-1* and *mt-3* (Figures 1 and 2).

Introgression of White Spruce and Black Spruce

The eight black spruce individuals (4%) from four populations that possessed *mt-1* or its close relative *mt-1a* (Figures 1 and 2) can only be explained by hybridization between white and black spruce and subsequent back crossing with black spruce. Since ten polymorphisms distinguish *mt-1* from *mt-2* (Table 6), it is highly unlikely that *mt-1* arose separately in these two species by convergent evolution. Nor were these white spruce trees mistakenly morphologically identified because they all contained nuclear

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(Germano and Klein, 1999) and chloroplast markers specific to black spruce (Figure 2).

Trees resulting from the hybridization of white spruce with Sitka or Engelmann spruce occur readily in nature (Sutton *et al.*, 1991; Sigurgeirsson *et al.*, 1991; Sutton *et al.*, 1994). However, natural hybrids between white and black spruce were believed to be extremely rare (Wright, 1955); only one such tree has been documented. Little and Pauley (1958) identified a tree in Cromwell, Minnesota, which they deemed *P. glauca* X *P. mariana*: variety Rosendahl. The intermediate nature of this tree was confirmed, and its progeny produced viable seed (Riemenschneider and Mohn, 1975). The introgressed black spruce populations I identified were dispersed, suggesting that hybridization between black and white spruce has happened on numerous occasions, and therefore is not as rare as previously thought. Although all of the populations are within the current range of white spruce, species-specific nuclear markers indicated that none of the eight trees were F_1 hybrids (data not shown). Since neither black spruce trees with white spruce chloroplast haplotypes, nor white spruce trees with black spruce haplotypes were detected, hybridization and subsequent introgression appears to have been unidirectional. This is consistent with the findings of Wright (1955) in which white (female) X black (male) crosses were deemed successful or probably successful, yet the reciprocal crosses were not.

<u>Introgression of Red Spruce and Black Spruce</u>

A considerable number of black spruce (18%) possessed *mt-3*, the only mitochondrial haplotype detected in red spruce (Figure 1C). Here again I hypothesize that interspecific hybridization accounts for the presence of this haplotype in black spruce, especially since hybridization between these two species is well known. These

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

samples were not misidentified red spruce trees because they contained chloroplast haplotypes of the *cp-2* lineage. The only exceptions to this, two black spruce trees possessing both *mt-3* and *cp-3*, could not have been misidentified red spruce either because 1) their provenances were west of the Great Lakes, far removed from the nearest red spruce populations, and 2) they both contained nuclear markers specific to black spruce (Germano and Klein, 1999). Five polymorphisms distinguished *mt-2* from *mt-3*. In order for convergent evolution to have produced the *mt-3* haplotype in black spruce, either five independent mutations in a single black spruce lineage, or recombination of haplotypes from separate lineages, would have had to have occurred.

Although the data appear to suggest that hybridization and subsequent introgression of red and black spruce were more successful in one direction, our sampling excluded some red and black spruce individuals that were previously classified as F_1 hybrids. Due to this selective sampling, the actual number of trees possessing organelle haplotypes of the opposite species is probably even greater. Significant asymmetric directionality of introgression was not observed in allopatric populations by Perron and Bousquet (1997), or in a coastal island population (Bobola *et al.*, 1996a).

<u>The Speciation of Red and Black Spruce</u>

It is broadly recognized that black and red spruce are closely related species. Perron *et al.* (2000) proposed that they comprise a recently evolved progenitor-derivative species pair with red spruce speciating from a once-isolated population of black spruce. The authors based their hypothesis on the following considerations: 1) the species can successfully hybridize, 2) red spruce has a narrow geographic range, 3) red spruce possessed fewer polymorphic STS loci, fewer alleles and less genetic diversity than black

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

spruce, and 4) red spruce displayed no unique alleles. Indeed, the range of red spruce currently is small, and the hybridization of the species is well documented. However, the reciprocal monophyletic relationship of the red spruce *cp-3* and black spruce *cp-2* haplotype lineages is not consistent with a progenitor-derivative relationship (Figure 1A). Furthermore, red spruce displayed as many polymorphic sites (Table 5), a higher ratio of haplotypes per sample size, and as much, if not more, genetic diversity as black spruce (Table 7). Observed heterozygosities of STS markers were 0.069 and 0.103, while chloroplast diversities were 0.52 and 0.44 (Table 7) in red and black spruce, respectively.

A key element of the progenitor-derivative hypothesis is that red spruce contained no unique STS alleles, therefore its genetic diversity appeared to be a subset of that of black spruce (Perron *et al.*, 2000). In the present study, six of the eight chloroplast haplotypes detected in red spruce were unique to it. In fact, a total of 30% of red spruce contained a red spruce-specific chloroplast haplotype (Figure 1A). This discrepancy may be due to a lack of sufficient sampling of black spruce in the STS study. Our data indicate that one of the three provenances used to represent allopatric ("pure") black spruce by Perron *et al.* (2000) was in fact introgressed: four out of four of the black spruce trees from the Manicouagan, Quebec population (sampled in both studies) possessed the mitochondrial haplotype *mt-3* (Figure 2B). It is conceivable that some of the "black spruce" STS alleles actually originated in red spruce and were inherited by black spruce through interspecific hybridization during the Wisconsin period. An expansion of the STS study to incorporate samples from the western portion of the black spruce range would have improved the data set, and possibly changed the conclusions drawn from it.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Perron *et al.* (2000) used genetic drift simulations to infer that 100,000 years would have been long enough for the level of genetic diversity in an isolated black spruce population to be reduced to the level of red spruce STS diversity. Although the authors proposed that the speciation of red spruce occurred some time during the Pleistocene, they use the genetic drift simulations to imply that the split could have happened as recently as the Wisconsin period. However, my analysis using differences in *trnK* intron sequences to estimate time since divergence from a common ancestor suggests the divergence of red and black spruce occurred approximately 0.6 to 3.5 million years ago. The estimated mutation rates are expected to be accurate because they were approximately two-fold higher than those reported for *rbcL* in conifers (Albert *et al.*, 1994). A divergence time of 600,000 years ago means that red and black spruce endured at least six ice ages as distinct species, which means they also underwent six range displacements and hence six alterations of their population genetic structure. The current geographic distribution of eastern spruce haplotypes is likely the result of their expansion from the southeast glacial refugium during the Holocene. The black spruce populations possessing haplotype *mt-3* were restricted to the east, nearer to the range of red spruce (Figure 2B). This observation is consistent with our hypothesis that black spruce obtained this haplotype through hybridization with red spruce; it cannot be attributed to a Wisconsin period speciation event.

Geographic Distribution of Haplotypes

European and North American pollen records indicate that forest trees migrated at rates between 150 and 500 meters per year during the late-Pleistocene and Holocene (Clark *et al.*, 1998). Spruce was estimated to have migrated across the eastern United

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

States at an average rate of 275 m/yr (Delcourt and Delcourt, 1987a). Such rates would be impossible under a stepping stone model of dispersal whereby seeds move unassisted (Clark *et al.*, 1998). Instead, a model of leptokurtic dispersal is assumed in which a fraction of seeds travel a long distance, via animals, wind across frozen landscapes or rushing water from melting glaciers, that results in the establishment of individuals far ahead of the migration front (Clark *et al.*, 1998).

Under the leptokurtic dispersal model, a fraction of seeds travel a long distance ahead of the migration front, and newly established populations are expected to possess less genetic diversity (loss of alleles) due to the founder effect (Hewitt, 1996; Comes and Kadereit, 1998). Consequently, populations nearest to the putative refugia display higher levels of genetic diversity (e.g. more haplotypes) than populations further away (Taberlet *et al.*, 1998; Comes and Kadereit, 1998; Newton *et al.*, 1999; Lowe *et al.*, 2000). The routes taken from a refugium determine the distribution patterns of intraspecific polymorphisms; specifically, a clinal pattern of variation follows the direction of migration (Taberlet *et al.*, 1998; Comes and Kadereit, 1998). Founder events also facilitate increased interpopulation and decreased intrapopulation variation of maternally inherited markers (Furnier and Stine, 1995). On the other hand, areas of merging populations that have expanded from different glacial refugia, i.e. zones of secondary contact or suture zones, are expected to contain increased genetic diversity (Hewitt, 1996; Taberlet *et al.*, 1998).

During the Holocene, the establishment of new territory by forest trees involved more than simple seed dispersal ahead of the advancing species front (Davis and Shaw, 2001). Not only were trees selected that could better invade the newly exposed yet still

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

harsh conditions of the north, but also that selection against phenotypes poorly adapted to the new (warm) local conditions occurred resulting in range contraction (Davis and Shaw, 2001). Therefore it is conceivable that hybrid and introgressed black/red spruce trees had a selective advantage over pure black spruce in the southern portions of its range during the rapid climate warming of the Holocene considering red spruce's adaptation to warmer conditions (Viereck and Johnston, 1990; Blum, 1990). Hybrid vigor has been reported under particular circumstances. For example Fowler *et al.* (1988) observed that red spruce with black spruce morphological characteristics outperformed pure red spruce under certain conditions. Gordon (1976) maintains that presently, natural hybrids are less fit, although he concedes that some hybrids have been quick to invade newly opened landscapes, e.g. clear cut land, and have become well established when not in direct competition with the parent species.

Distribution of the Red Spruce *Mt-3* Haplotype in Allopatric Black Spruce

Eleven of the seventeen *mt-3* containing black spruce populations were outside the natural range of red spruce (Little, Jr., 1971). The westernmost of these populations was located in the southwest corner of Ontario, over 1100 km from the closest region of sympatry with red spruce. Using the average spruce migration rate of 275 m/yr inferred from eastern North American pollen records (Delcourt and Delcourt, 1987a) and a generation time of 10-30 years (Fowells, 1965; Viereck and Johnston, 1990), I estimated that the original hybridizations occurred over 4000 years or 130-400 generations ago. Such a long time span can be attributed to the maternal inheritance of the *mt-3* haplotype and the relative short dispersal distances of spruce seed. The significant number of allopatric black spruce populations retaining the red spruce *mt-3* haplotype suggests that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

hybridization occurred frequently during the Holocene, possibly more so than it is believed to occur today. It is feasible that hybrid and introgressed black/red spruce trees had a selective advantage over pure black spruce in the southern portions of its range during the rapid climate warming of the Holocene considering red spruce's adaptation to warmer conditions.

Distribution of Chloroplast Haplotypes

No distributional patterns of the chloroplast haplotypes were detected within red spruce (Figure 2A). Hawley and DeHayes (1994) observed less allozyme variation in southern red spruce populations, which they attributed to genetic drift within small isolated populations. Khalil (1987) also detected clinal patterns of growth characters in red spruce. The apparent random spread of chloroplast haplotypes across the red spruce range is probably due to the increased amount of gene flow of chloroplast DNA with respect to nuclear genes in conjunction with the narrow range of the species.

Previously, east-west trends were reported for cpDNA RFLPs in white spruce (Furnier and Stine, 1995) and for monoterpene frequencies in both white and black spruce (Wilkinson *et al.*, 1971; Chang and Hanover, 1991). Although geographic patterns of haplotypes in white spruce were not observed (likely because little variation in the surveyed genes was detected in the species), there was an obvious difference in black spruce chloroplast haplotype frequencies between the eastern and western portions of its range (Figure 2A). The chloroplast haplotype *cp-2f* was detected primarily in the east with a significant clinal pattern decreasing from south to north and from east to west. The *cp-2g* haplotype however, was predominant in the west with a significant cline in the opposite direction. The westward decrease in overall chloroplast diversity is consistent

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

with a northward and westward expansion of black spruce from the well-documented southeastern glacial refugium. However the opposing clines of *cp-2f* and *cp-2g* frequencies are consistent with multiple scenarios. The prevalence of *cp-2g* in the west could be due to the founder effect from leptokurtic dispersal out of the southeastern refugium. Additionally, the large Glacial Lake Agassiz, present around 14,000-11,000 years BP, was probably a significant barrier between the Great Lakes region and the Great Plains of the west (Critchfield, 1984). Another interesting possibility was suggested by Furnier and Stine (Furnier and Stine, 1995) who attributed their observed east-west genetic divide among white spruce to a small glacial refugium in the west. It is possible that this hypothetical western refugium contained black spruce as well.

This disjunct refugium for white spruce has been hypothesized to have existed in the unglaciated portion of northwestern North America during the Wisconsin period (Tsay and Taylor, 1978; Critchfield, 1984; Ritchie and MacDonald, 1986), with several lines of evidence indirectly pointing to its existence. Firstly, the postglacial climatic warming, ice retreat (McLeod and MacDonald, 1997) and repopulation by spruce of the western interior of Canada occurred extremely rapidly (Critchfield, 1984; Ritchie and MacDonald, 1986; McLeod and MacDonald, 1997). McLeod and MacDonald (1997) demonstrated that black and white spruce pollen appeared on either side of a ~1900 km region (from central Saskatchewan to the northwest corner of the Northwest Territories) within 3000 years. Moreover, Ritchie and MacDonald (1986) concluded that white spruce spread across 2000 km of the same region in only 1000 years. Strong winds were suggested by Ritchie and MacDonald (1986) as the proponent for such long-range seed dispersal, however these migration rates are over 2-7 times the average rate estimated for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

re-invasion of more southern parts of North America (Delcourt and Delcourt, 1987a) and would have required seeds to travel about 20-60 km in a northwest direction from their parent tree every generation (Nienstaedt and Zasada, 1990).

Secondly, Ritchie (1980) cites pollen evidence of spruce forests in the northern Yukon Territory during the Middle Wisconsin, and macrofossil evidence of spruce in north central Alberta between ~43,000 and 27,000 yr BP. Spruce pollen was present throughout a >24,900 yr BP sediment record from northern Yukon, however the author did not conclude that it was present in the vicinity of the site due to its low pollen percentage (2-10%) and low maximum influx values. In fact, it is common for palynologists to reject the regional presence of a taxa despite low pollen percentages in sediment cores, but this conservative interpretational practice may lead to inaccurate inferences. For example, lodgepole pine (*Pinus contorta*) pollen percentages of less than 2% in southeast Alaska were originally attributed to long-range pollen dispersal until the definitive presence of the species at 10,000 yr BP was evidenced by macrofossils at the site (Peteet, 1991). Low pollen production would be expected of any tree species in the harsh conditions of a northern glacial refugium (Beiswenger, 1991), therefore low pollen percentages in sediments from such areas may indeed indicate the presence of the respective taxa. There is growing evidence that some pollen studies, which originally attributed the presence of pollen of unexpected taxa to long distance transport, should be reinterpreted to consider the possibility of cryptic northern refugia (Stewart and Lister, 2001).

A full-glacial refugium in northwestern North America has also been suggested for species other than spruce. Peteet (1991) suggested that a late Wisconsin refugium

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

may have existed for lodgepole pine in southeastern Alaska. Some molecular phylogeography studies have also suggested such a refugium (Comes and Kadereit, 1998). In a study of chloroplast RFLPs in perennial herbs and other species, Soltis *et al.* (1997) hypothesized that plants survived the glaciation in two widely separated refugia: south of the ice sheet and much further north perhaps in areas of central Alaska and coastal islands. Chloroplast DNA diversity supported previous fossil evidence of an eastern Beringian refugium for arctic plants *Saxifraga oppositifolia* (Abbott *et al.*, 2000) and *Dryas integrifolia* (Tremblay and Schoen, 1999).

If a northwest refugium existed for spruce, then one would expect to see relatively more genetic diversity in present populations of Alaska and decreasing diversity continuing eastward until the zone of secondary contact where they would have merged with populations migrating from the southeast refugium. However, as Furnier and Stine (Furnier and Stine, 1995) point out, such a western refugium could have been small enough to allow genetic drift to diminish the amount of diversity within it. Additionally, if chloroplast haplotype *cp-2g* truly was present in an Alaskan refugial population during the Wisconsin period, then one would expect to see it in relatively high frequencies in present-day populations in Alaska and western Canada. In this study, Alaska was represented by only one black spruce population. In order to effectively test this hypothesis, many more intraspecific chloroplast and mitochondrial polymorphisms would need to be identified and their distribution assessed in an extensive sampling of black and white spruce populations in Alaska and western Canada.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Summary

Range-wide distributions of mitochondrial and chloroplast haplotypes have shed some light on the recent history of white, black and red spruce. We have shown several lines of evidence that black and red spruce do not constitute a progenitor-derivative species pair, and that they have endured multiple ice ages as distinct species. Hybridization and introgression have played a role in the population genetics of red and black spruce and, unexpectedly, of black and white spruce during the Holocene. Detection of additional variation (especially in the mitochondrial genome) and increased sampling are necessary to further investigate the east-west divide among black spruce chloroplast haplotypes. This work provides a good foundation for future studies aimed at elucidating the post-glacial migration history of the eastern North American spruce species.

Acknowledgements

I wish to thank Dr. Narinder Dhir, Manager, Genetics and Tree Improvement, Land and Forest Service, Alberta Environment, Alberta, Canada, Mr. Peter Copis, Research Forest Technologist, Petawawa Research Forest, Canadian Forest Service, Chalk River, Ontario, Canada, Mr. Leonard Barnhardt, Site Manager, Alberta Tree Improvement and Seed Center, Alberta, Canada, Dr. Jerome I. Klein, Research Scientist, Canadian Forest Service (retired), Dr. Glenn R. Furnier, Dr. Peter Garrett, Northeastern Forest Experiment Station of the US Forest Service, Ms. Kristine Johnson, Supervisory Natural Resource Specialist, National Park Service, Dr. Gerald E. Rehfeldt and Mr. Gulya Kiss for providing plant materials. I am grateful to Dr. Chris D. Neefus and Dr. Remy J. Petit for assisting with statistical analyses. Much of the data presented in this

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

chapter was collected by Mr. Christopher Clouser and Ms. Helene Balkin. Support was received in part by the Conrad Research Award in Plant Biology for Undergraduates, University of New Hampshire, Durham, NH and a Pfizer Summer Undergraduate Research Fellowship, Pfizer Inc. This research was supported by a grant from the New Hampshire Agricultural Experiment Station (NHAES) MacIntire-Stennis Project MS27.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER III

CHLOROPLAST AND MITOCHONDRIAL DNA EVIDENCE FOR THE EVOLUTION OF *PICEA*

Abstract

A robust species phylogeny of *Picea* (spruce) is desirable in order to answer multiple questions about the history of the genus' biogeography. A phylogenetic study of sixteen North American and Eurasian *Picea* species was conducted utilizing DNA sequences of the chloroplast *trnK* intron and the second intron of the mitochondrial *nad1* gene. The topologies of the trees inferred from the chloroplast and mitochondrial data varied significantly, in particular to the placement of *P. omorika*, *P. mexicana* and *P. glauca*. These trees also differ from results of previous phylogenetic and systematic studies. Inter-species hybridization and introgression, events not uncommon in the *Picea*, are discussed as possible reasons behind such incongruencies.

Introduction

Taxonomy of *Picea*

The genus *Picea* Dietr. is comprised of approximately 35-50 species distributed across the Northern Hemisphere (Schmidt-Vogt, 1977; Vidakovic, 1991; Gordon, 1992). While most of the species occur in northern and east/central Asia, nine are indigenous to North America and two are native to Europe (Table 1). The *Picea* are distinct from other members of the Pinaceae family. The genus' monophyly is supported by morphological

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. *Picea* species included in the phylogenetic analysis.

Species	Current Distribution	Source[a]	*trnK* intron	*nad1* intron 2	*nad7* intron 1
Picea abies (L.) Karst.	Northern Eurasia	Min.	AY289610	AY289611	
P. asperata Mast.	Central China	Min.	AY035202	AY153790	AY169715
P. breweriana S. Wats.	Northwest U.S.A.	Min.	AY035197	AY153786	AY169711
P. chihuahuana Martinez	Mexico	Min.	AY035198	AY153784	AY169710
P. engelmannii (Parry) Engelm.	Western North America	Min.	AY035196	AY153788	AY169707
P. glauca (Moench) Voss	North America	SD	AF059341	AY057955	AY057968
P. jezoensis (Sieb. et Zucc.) Carr.	East Asian Islands	Min.	AY035201	AY153791	AY169716
P. mariana (Mill.) B.S.P.	North America	Mas.	AF059343	AY057953	AY057963
P. mexicana Martinez	Mexico	Min.	AY035194	AY153785	AY169709
P. omorika (Pancic) Purkyne	Yugoslavia	Min.	AY035200	AY153792	AY169713
P. pungens Engelm.	Western U.S.A.	Min.	AY035195	AY153783	AY169712
P. rubens Sarg.	Eastern North America	Col.	AF059342	AY057949	AY057959
P. schrenkiana Sarg.	Central Asia	Min.	AY035204	AY153782	AY169718
P. sitchensis (Bong.) Carr.	Western North America	Min.	AY035203	AY153787	AY169708
P. smithiana (Wall.) Bioss.	Central Asia	Min.	AY035199	AY153781	AY169717
P. wilsonii Mast.	Central China	Min.	AY035193	AY153789	AY169714

[a] Abbreviations: Min. = Ministry of Forests, British Columbia, in Vernon B.C. (courtesy of Mr. Gulya Kiss); SD = Black Hills, South Dakota (courtesy of Dr. Gerald Rehfeldt; Germano and Klein, 1999); Mas. = USDA Forest Service (Northeastern Forest Experiment Station), in the Massabesic Experimental Forest, Alfred, Maine (Germano and Klein, 1999); Col. = Coleman State Forest, Stewartstown, New Hampshire (Germano and Klein, 1999).

and chloroplast DNA (cpDNA) assessments (Wright, 1955; Schmidt-Vogt, 1977; Sigurgeirsson and Szmidt, 1993; Chapter II).

A systematic classification of the genus *Picea* has previously been examined by comparing morphology, anatomy, physiology and inter-species crossability, although a consensus has not been achieved. In 1887 Willkomm recognized two sections or subgenera using needle morphology. The Eupicea were defined as having 4-sided needles with stomata on all sides, like *P. abies*, *P. glauca* and *P. orientalis*, whereas species in the Omorika section have 2-sided needles with stomata only on the dorsal side like *P. omorika* and *P. sitchensis*. In 1890 Mayr added cone morphology to the system and divided the genus into three sections. The Morinda section roughly paralleled Eupicea Willk., and the Omorika section included *P. omorika*, *P. bicolor*, *P. breweriana* and *P. glehnii*. Mayr's new section Casicta had needles like those of the Omorika section, but had ripe cones with thin, flexible wavy cone scales; it included *P. jezoensis*, *P. engelmannii*, *P. pungens* and *P. sitchensis* (previously assigned to Omorika), (Aldén, 1987).

Aldén (1987) presents a "practical" system of three sections, borrowing from both Willkomm and Mayr. Section Picea (not Eupicea Willk.), having 4-angled needles with stomata on all sides, hard cone scales with rounded apices, included the majority of the species: *P. abies*, *P. obovata* (as a subspecies of *P. abies*), *P. koraiensis*, *P. pungsaniensis* (as a variety of *P. koraiensis*), *P. koyamai*, *P. polita*, *P. asperata*, *P. retroflexa* (as a variety of *P. asperata*) *P. gemmata*, *P. meyeri*, *P. neoveitchii*, *P. wilsonii*, *P. schrenkiana*, *P. crassifolia*, *P. smithiana*, *P. maximowiczii*, *P. glehnii*, *P. morrisonicola*, *P. bicolor*, *P. orientalis*, *P. mariana*, *P. rubens*, and *P. glauca*. Section Casicta Mayr, with 4-angled or flat needles, with stomata on all sides or fewer on the ventral side, and thin flexible cone

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

scales with wavy apices, included *P. engelmannii*, *P. mexicana*, *P. pungens*, *P. chihuahuana*, *P. likiangensis*, *P. hirtella*, *P. montigena*, *P. jezoensis* and *P. sitchensis*. Section Omorika Willk., having compressed needles with strictly dorsal stomata, included *P. brachytyla*, *P. farreri*, *P. spinulosa*, *P. omorika* and *P. breweriana*. Vidaković (1991) presents a slight variation of this system. He divides section Eupicea Willk. (not section Picea Aldén) into three subsections: the first containing *P. smithiana*, *P. schrenkiana*, *P. wilsonii*, *P. maximowiczii*, *P. polita*, *P. neoveitchii*, *P. asperata* , the second including *P. gemmata*, *P. meyeri*, *P. retroflexa* (as a species), *P. abies*, *P. obovata* (as a species), *P. orientalis* and *P. koyamai*, and the third containing *P. pungsaniensis* (as a species), *P. bicolor*, *P. glehnii*, *P. morrisonicola*, *P. mariana*, *P. rubens*, and *P. glauca*. Section Casicta Mayr is also divided into three subsections: *P. engelmannii*, *P. mexicana*, *P. pungens* and *P. chihuahuana* in the first, *P. likiangensis* in the second and *P. montigena*, *P. jezoensis* and *P. sitchensis* in the third group. Finally Vidaković divides section Omorika Willk. into two subsections: the first including *P. brachytyla* and *P. omorika* and the second including *P. breweriana* and *P. spinulosa*. Aldén (1987) notes that the section Casicta Mayr in particular is heterogeneous and that the sections are probably unnatural, i.e. related by appearance and not necessarily by descent. He concludes that a systematic treatment of the genus is warranted. A robust phylogeny of the *Picea* would facilitate an understanding of its biogeography and assist breeding and reforestation programs.

Phylogeny and Biogeography of *Picea*

The history of radiation and speciation in the genus is controversial due to the paucity and fragmented nature of the fossil record. Fossil wood from Manchuria China of the fossil genus *Protopiceoxylon* (Middle Jurassic; ~160 MYA), the earliest fossil

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

resembling *Picea*, suggests that *Picea* evolved in Asia. Fossil wood from England (Early Cretaceous; 145-100 MYA) and from Japan and North America (Late Cretaceous; 100-70 MYA) are described as *Picea* (Aldén, 1987). However, these records are not contained within a summary of the *Picea* fossil record compiled by LePage (2001), and fossil woods of similar age have been scrutinized by Miller, Jr., (1977; 1989). Although he cites *Picea* pollen from Paleocene coal deposits (66-58 MYA) in Montana, USA, LePage (2001) generally believes that *Picea* first appeared during the middle Eocene (58-37 MYA). In fact, more than a dozen distinct *Picea* species have been reported from the Eocene, mostly from North America, with a few from Asia (LePage, 2001). These are consistent with Miller's (1977) belief that the genus evolved no earlier than the very end of the Cretaceous (~65 MYA). *Picea* had further diversified and established itself in western North America and Asia by the Oligocene Epoch (Miller, Jr., 1989; LePage, 2001).

Wright (1955) compared results of artificial crossing experiments with geographic distribution and morphology to predict evolutionary relationships among the *Picea*. Although he recognized no natural breaks in the genus sufficient to divide it into sections, he discussed several phylogenetic groups of species, the members of each group he believed to share a common origin. What is noteworthy about this study is that geographically disjunct species appeared to be more closely related than other more proximal species. *Picea rubens* and *P. mariana* were considered closely related and comprised the eastern America group. *Picea glauca*, *P. engelmannii*, *P. pungens* and *P. sitchensis* comprised the northwest America group; *P. breweriana* and *P. chihuahuana* were excluded from this group because morphologically they are more similar to the Asian species. The southwestern Chinese and Formosan species (*P. brachytyla*, *P. purpurea*, *P.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

wilsonii and *P. morrisonicola*) formed another group with morphological and distributional evidence indicating a common origin. The Himalayan species *P. spinulosa* and *P. smithiana* were also considered to be closely related. Although distributed in the same geographic location, the Japanese species (*P. koyamai*, *P. jezoensis*, *P. polita*, *P. maximowiczii*, *P. bicolor*, *P. glehnii*), did not comprise a group because of their lack of crossability with each other. *Picea koyamai* was proposed to be the oldest of the *Picea* species because of its "generalized" morphological traits and crossability with multiple other species. Morphological and crossing data suggested that it belonged in a group of closely related species of north Eurasia and north China (*P. abies*, *P. asperata* and *P. likiangensis*). *Picea jezoensis* appeared more closely related to *P. brachytyla*, and was suggested to be a link to species of the northwest America group. *Picea omorika* is morphologically very similar to *P. rubens* (Wright, 1955), and it displays higher crossability with *P. rubens* and *P. mariana* than they do with one another (Gordon, 1976).

The evolutionary relationships of North American *Picea* species and their historical biogeography are disputed. Wright (1955) believed that the genus originated in northeast Asia because the largest number of extant species are present in the region and it is the location of the putative primitive species, *P. koyamai*. He hypothesized that radiation from its origin occurred in multiple waves: at least three migration events from Asia to North America would account for the ancestor of the eastern America group, the ancestor of the northwest America group and the ancestor of *P. breweriana* and *P. chihuahuana*. Although he hypothesized that east-west migrations occurred via both northern and southern routes, he did not discuss the specific locations of these postulated land bridges (Wright, 1955). In contrast to Wright's hypothesis, Fowler (1966) believed that a single

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ancestor gave rise to all of the North American species. Hills and Ogilvie (1970) built on Fowler's hypothesis, proposing that the migration event occurred via the Asian-Alaskan land bridge, giving rise to the progenitor of *P. rubens* and *P. mariana* and to the progenitor of *P. glauca* and the rest of the North American species.

In a molecular phylogeny based on cpDNA restriction fragment length polymorphisms (RFLPs), Sigurgeirsson and Szmidt (1993) identified several alliances that were in agreement with Wright's groupings. These included a *P. glauca* alliance (*P. glauca*, *P. engelmannii* and *P. mexicana*), a *P. abies* alliance (*P. abies*, *P. asperata*, *P. aurantiaca*, *P. glehnii*, *P. koraiensis*, *P. koyamai* and *P. meyeri*) and a *P. brachytyla* alliance (*P. brachytyla*, *P. bicolor*, *P. chihuahuana*, *P. maximowiczii*, *P. morrisonicola*, *P. neoveitchii*, *P. orientalis*, *P. polita*, *P. purpurea* and *P. wilsonii*). Furthermore, as LePage (2001) points out, species possessing thin and flexible cone scales were allied with each other, and species with thick and woody cone scales were clustered together in the cpDNA RFLP trees. Sigurgeirsson and Szmidt's results were also consistent with Wright in that there was a close relationship of *P. omorika* with *P. rubens* and *P. mariana*. The molecular data, however, did not support the putative ancient status of *P. koyamai* (Sigurgeirsson and Szmidt, 1993). In fact, some of the North American species (the *P. glauca* alliance, *P. sitchensis* and *P. breweriana*) were positioned near the root of the phylogenetic tree while others were tightly nested with Eurasian species among higher branches. This suggested that *Picea* originated in North America and that intercontinental migrations occurred multiple times following its initial radiation (Sigurgeirsson and Szmidt, 1993). Considering the current geographic locations of *P. omorika* (Europe) and *P. rubens* and *P. mariana* (eastern North America), it would seem plausible that these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

species' migrations occurred when the continents were adjoined. However, the supercontinent of Pangaea began to break up during the Triassic Period (~245-200 MYA), and the Atlantic Ocean formed in the Jurassic Period (~200-145 MYA). Since *Picea* did not evolve until millions of years later (during the end of the Cretaceous Period according to the fossil record), these migrations must have occurred between northwestern North America and northeastern Asia.

Speciation and Inter-Species Hybridization

Reproductive isolation is a hallmark of the Biological Species Concept (Mayr, 1992), yet this criterion is applied with difficulty to the systematics of *Picea*. Reproductive barriers among *Picea* are weak, and the results of artificial crossability experiments have found cross compatibility to be moderately high among many species. Introgressive hybridization has been documented for many of the species, including *P. rubens* and *P. mariana* (Morgenstern and Farrar, 1964; Manley, 1972; Gordon, 1976; Bobola *et al.*, 1996a; Bobola *et al.*, 1996b; Perron and Bousquet, 1997), *P. sitchensis*, *P. engelmannii* and *P. glauca* (Sutton *et al.*, 1991; Sutton *et al.*, 1994), *P. abies* and *P. obovata* (Krutovskii and Bergmann, 1995), and *P. glauca* and *P. mariana* (see Chapter II; Riemenschneider and Mohn, 1975). Critchfield (1984) hypothesized that introgressive hybridization played an important role in the evolution of conifers by resulting in the formation of new races or species. Transitory geographic races would have evolved during the Quaternary Period (2 MYA-present) as species were repeatedly displaced from common territories to geographic isolation in response to cycling climate change. This confounds a systematic treatment of the *Picea* as some contemporary species may actually be ancient hybrids.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Sigurgeirsson and Szmidt (1993) discuss the potential of cpDNA to cross species boundaries through introgressive hybridization as a possible drawback of using uniparentally-inherited genomes to infer organismal phylogenies. In *Picea*, as well as other conifers, the chloroplast DNA is paternally inherited whereas the mitochondrial genome is maternally inherited (Stine *et al.*, 1989; Stine and Keathley, 1990; Sutton *et al.*, 1991; David and Keathley, 1996; Bobola *et al.*, 1996b; Grivet *et al.*, 1999). This offers a unique opportunity to compare the evolution of two independent lineages, as well as to detect evidence of past introgressive hybridization events. The chloroplast *trnK* intron, encompassing the *matK* gene, has been used for multiple phylogenetic studies (Johnson and Soltis, 1994). *TrnK* intron sequences demonstrate low levels of variation among *Picea* species compared to *Pinus* species, however they are reasonably informative for phylogenetic study of the genus (Germano and Klein, 1999; Chapter II). Mitochondrial base substitutions occur at a relatively low frequency in plants (Palmer, 1992). However, short duplications and insertion/deletions in introns of the mitochondrial *nad1* and *nad7* genes were observed between closely related *Picea* species and may therefore be phylogenetically informative in an assessment of the genus (see Chapter II).

In this chapter, I compare phylogenetic gene trees from the chloroplast and mitochondrial genomes representing nine North American, two European and five Asian *Picea* species. The purpose of this preliminary study is not to present a robust phylogeny of the genus, but to investigate whether chloroplast and mitochondrial gene trees for *Picea* would be generally congruent.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Materials and Methods

Species included in this study, their current geographic distributions, and their corresponding GenBank accession numbers for the various genes sequenced are listed in Table 1. *Picea glauca*, *P. mariana* and *P. rubens* specimens were obtained from provenance tests and independent collections. They have associated range-wide population studies (Germano and Klein, 1999; Chapter II). The individuals of these three species used in this study had chloroplast and mitochondrial haplotypes that were the most frequent and hence representative of their respective species. *Picea breweriana* was collected from a stand identified as Little Grayback, in T18N R6E Section 4 HBM by Charles L. Frank, North Zone Genetic Resource Program, USDA Forest Service, Klamath National Forest. *Picea sitchensis* was collected near Arcata, California by Yan Linhart, University of Colorado. The remaining species were collected at the arboretum for the Ministry of Forests, British Columbia, in Vernon B.C. (courtesy of Mr. Gulya Kiss). Species designations at the arboretum were verified by Dr. Alan Gordon and voucher specimens have been deposited at the University of Maine herbarium. DNAs were extracted from 5-10 g tissue (buds and/or needles) using a standard CTAB method (Doyle and Doyle, 1987). The chloroplast *trnK* intron and mitochondrial *nad1* intron 2 and *nad7* intron 1 were amplified and sequenced as previously described (see Chapter II).

Sequences were assembled using SeqMan II and aligned with MegAlign (Lasergene software package; DNASTAR version 3.72, Madison, WI). Multiple long insertions and deletions in the *nad1* sequences necessitated manual editing of their alignment. Gaps were positioned parsimoniously; i.e. in a way such that a minimum number of mutations could account for the observed polymorphisms. Portions of the *nad1*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

intron that were unique to some *Picea* species were analyzed with BLASTN 2.2.4 (Altschul *et al.*, 1997) to check for possible shared ancestry with *Pinus*.

Phylogenetic trees were generated with Phylogenetic Analysis Using Parsimony (Swofford, 1998) using maximum parsimony and minimum evolution. For the *trnK* analyses, gaps were treated as missing; indels were coded as present or absent (1 or 0) at the end of the alignments (see Appendix B). Due to the complexity of the *nad1* data, all of the sequence polymorphisms, including base substitutions and indels, were converted to binary code (see Appendix C). In cases where taxa displayed polymorphisms within sequence regions that were otherwise deleted in other taxa, a question mark was used for that character in the taxa with the region deleted (question marks were defined as missing data in the nexus file). All characters were unordered and weighted equally. Settings were as follows: 1) parsimony optimality criterion, full heuristic search, random sequence addition (10 replicates) with 1000 bootstrap replicates, tree-bisection-reconnection (TBR) branch-swapping algorithm or 2) distance optimality criterion (minimum evolution), distance measure = mean character difference, neighbor-joining search with 1000 bootstrap replicates. A maximum parsimony analysis (parsimony optimality criterion, full heuristic search, random sequence addition [10 replicates]) of the *nad1* data set was also performed without bootstrapping. *Pinus thunbergii* (Tsudzuki *et al.*, 1992), *Pinus armandii* and *Pinus banksiana* (Wang *et al.*, 2000) served as the outgroup for the *trnK* analyses; no outgroup was defined for the *nad1* analyses. Maximum parsimony and minimum evolution trees generated from the same data set were compared and tested for length differences using the Kishino-Hasegawa (KH) test in PAUP (Goldman *et al.*, 2000). Trees for the KH test were generated with 1) parsimony optimality criterion, heuristic search, random sequence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

addition (10 replicates), TBR branch-swapping and 2) distance optimality criterion, distance measure = mean character difference, heuristic search, TBR branch-swapping. In order to test the significance of topological (tree length) differences between the chloroplast *trnK* and mitochondrial *nad1* gene trees, a maximum parsimony analysis (heuristic search, computing strict consensus) of the *nad1* data was carried out constraining trees in various ways (Table 2) that were consistent with the *trnK* tree topologies and vice versa. To determine if each constrained tree was significantly longer than the original non-constrained consensus tree, the non-parametric ranked-sign test of Templeton (Wilcoxon signed-ranks test) was applied using PAUP (Larson, 1994).

Results

Chloroplast *TrnK* Analyses

The chloroplast *trnK* intron ranged from 2481-2492 bp among *Picea* species. The alignment of the sequences can be viewed via links to the GenBank accessions in Table 1. Of the 2507 total characters scored by PAUP, 12 were indels, 126 were parsimony-informative and 73 were variable but parsimony-uninformative. All of the indels were small; none were over 7 bp. The number of polymorphisms between any two *Picea* species ranged from 0-22 differences. Species pairs that had identical sequences included *P. mexicana* and *P. glauca*, *P. jezoensis* and *P. wilsonii*, *P. rubens* and *P. omorika*, and *P. abies* and *P. asperata* (see Appendix D for the pairwise distance matrix). The maximum parsimony (MP) and minimum evolution (ME) *trnK* trees are displayed in Figures 1 and 2. Their topologies do not vary significantly, i.e. they are of the same tree length and therefore no differences were identified with the KH test ($P = 1$). The only differences between the trees were differences in resolution: 1) In the MP tree, the *P. breweriana*/*P.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. List of constraints used for the Wilcoxon signed-ranks test.

Constraints	Δ Tree length[a]	*P*
Constraints on the mitochondrial *nad1* trees		
Sister groups:		
P. engelmannii and *P. glauca* with *P. abies*, *P. asperata*, *P. jezoensis*, *P. wilsonii*	11	0.0050**
P. mexicana with *P. engelmannii* and *P. glauca*	2	0.3173
Monophyly of:		
P. rubens, *P. mariana* and *P. omorika*	14	0.0062**
Constraints on the chloroplast *trnK* trees		
Sister groups:		
P. mariana and *P. rubens* with *P. schrenkiana* and *P. smithiana*	3	0.0833
P. sitchensis with *P. engelmannii* and *P. glauca*	21	0.0010**

[a] Difference in tree length between the unconstrained and constrained maximum parsimony trees.

** $P < 0.01$.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

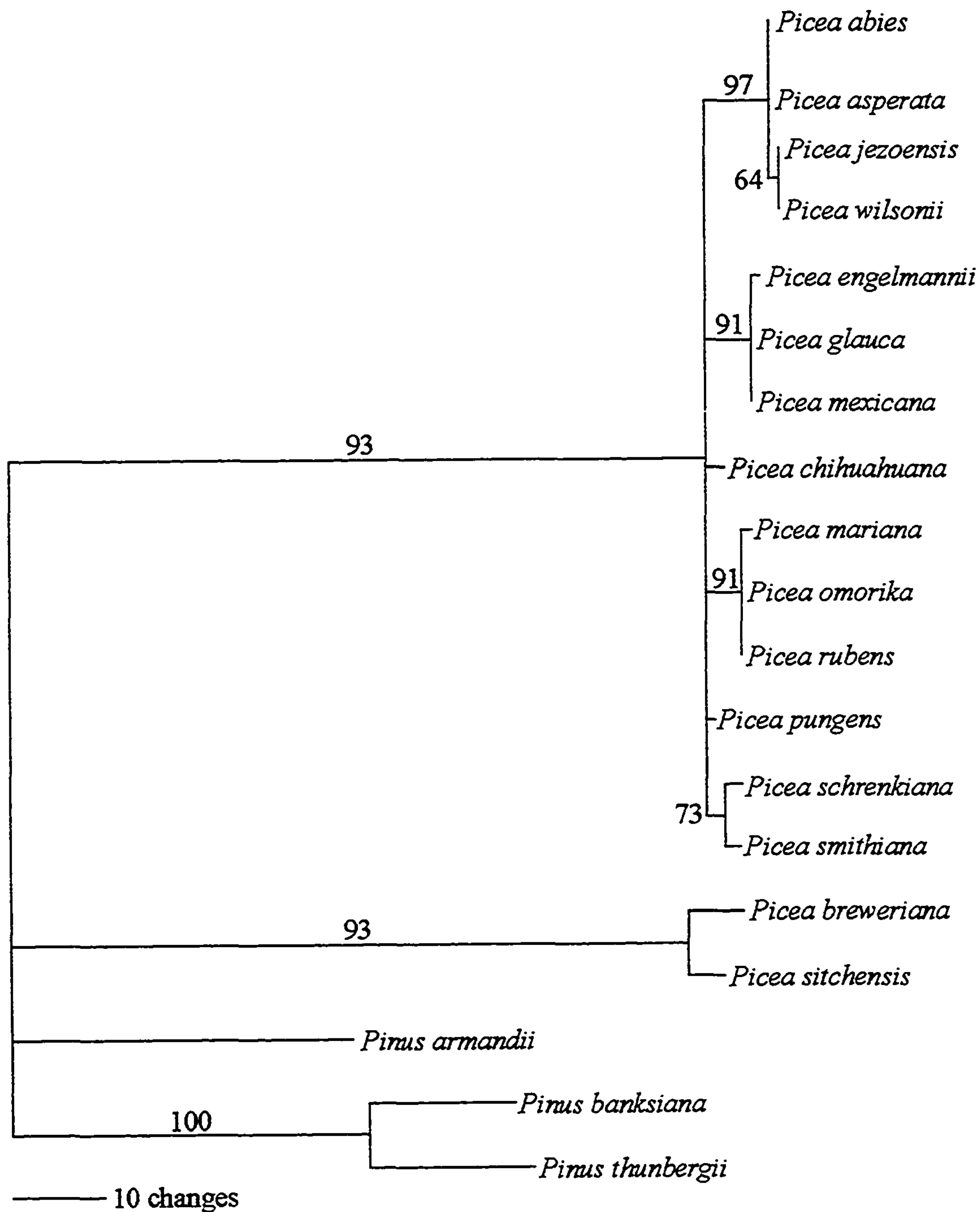


Figure 1. Maximum parsimony phylogenetic tree from chloroplast *trnK* intron sequence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

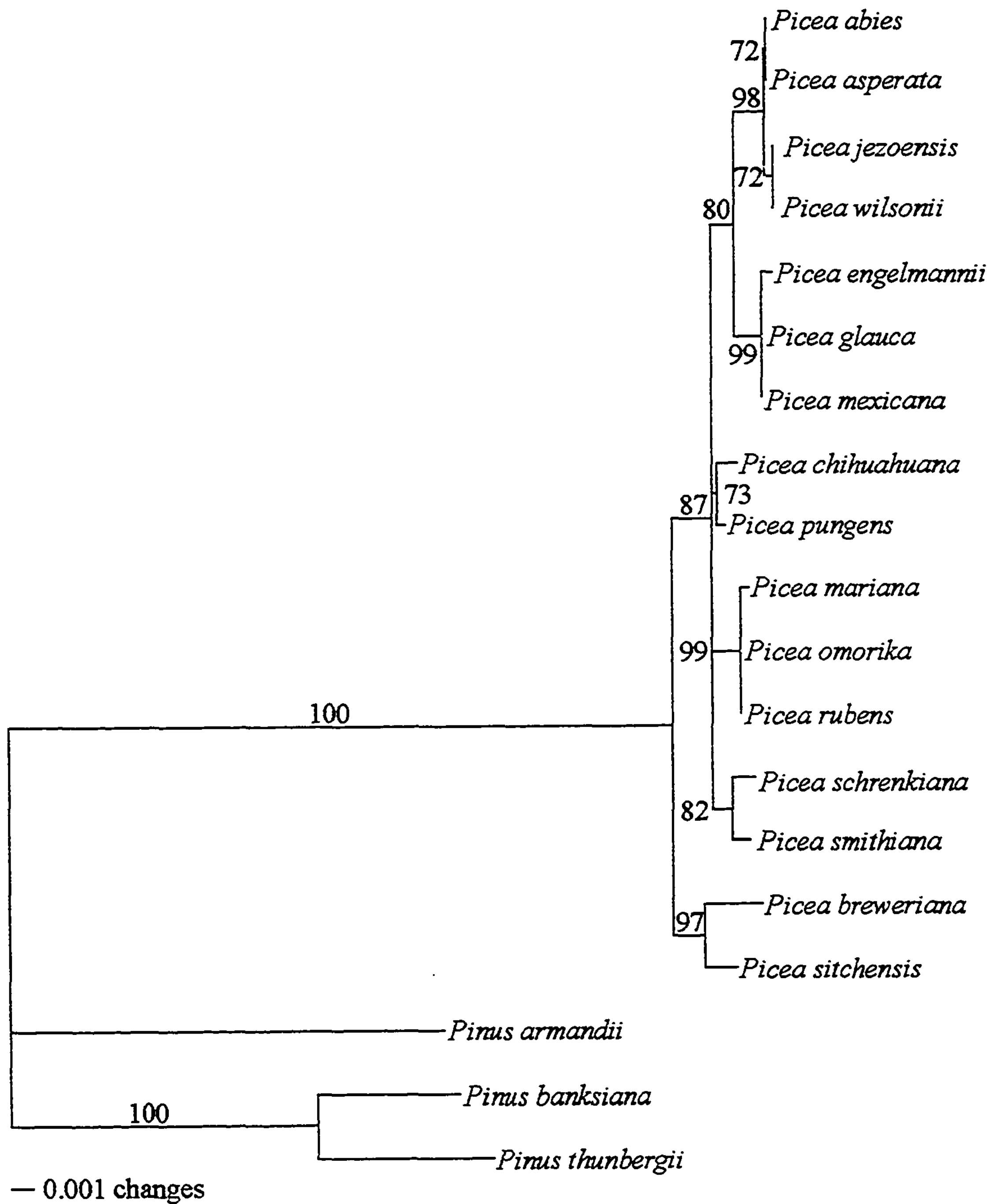


Figure 2. Minimum evolution phylogenetic tree from chloroplast *trnK* intron sequence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sitchensis clade is basal with respect to the *Picea* but it is part of an unresolved polytomy with the *Pinus* species. In the ME tree, this clade forms a monophyletic group with the *Picea* and it is basal within that group. 2) In the MP tree, the position of *P. pungens* is unresolved, whereas in the ME tree *P. pungens* is sister to *P. chihuahuana* (73% bootstrap support). 3) In the ME phylogram and the MP cladogram (not shown), the *P. abies*/*P. asperata*/*P. jezoensis*/*P. wilsonii* clade is sister to the *P. engelmannii*/*P. glauca*/*P. mexicana* clade (80% and 84% bootstrap support, respectively). In the MP phylogram (Figure 1) however, this node is depicted as collapsed and the two clades are part of an unresolved polytomy. Overall, bootstrap values were similar between the MP and ME *trnK* trees.

In the *trnK* trees overall, the *Picea* form a monophyletic group. Within this group, *P. breweriana* and *P. sitchensis* are sister taxa in a strongly supported (93-97%) basal clade. The remaining *Picea* species form a clade (87-93%) in which there are four relatively well-supported clades: *P. abies*/*P. asperata*/*P. jezoensis*/*P. wilsonii* (97-98%), *P. engelmannii*/*P. glauca*/*P. mexicana* (91-99%), *P. mariana*/*P. rubens*/*P. omorika* (91-99%) and *P. smithiana*/*P. schrenkiana* (73-82%). The only relationship resolved among these clades is the one between the *P. abies*/*P. asperata*/*P. jezoensis*/*P. wilsonii* and *P. engelmannii*/*P. glauca*/*P. mexicana* clades (discussed above). Relationships among this clade, the *P. smithiana*/*P. schrenkiana* clade, *P. pungens*, *P. chihuahuana* and the *P. mariana*/*P. rubens*/*P. omorika* clade are unresolved.

Mitochondrial *Nad1 and Nad7* Analyses

The *Picea* species surveyed in this study displayed virtually identical mitochondrial *nad7* intron 1 sequences. The only polymorphisms detected out of the 895-900 bp were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

one 5-bp tandem duplication distinguishing *P. rubens*, a base substitution distinguishing *P. sitchensis*, and a base substitution shared by *P. smithiana* and *P. schrenkiana*. Due to the paucity of sequence variation in this intron, it was excluded from the phylogenetic analyses.

Although less variation was observed in the mitochondrial *nad1* intron than in the *trnK* intron, the *nad1* sequence was still phylogenetically informative. The sequences ranged in size from 2091-3157 bp due to the presence of many large indels. The alignment of the sequences can be viewed via links to the GenBank accessions in Table 1. The largest indel was a >1 kb section in the middle of the intron that was present in all the North American species as well as in *P. schrenkiana* and *P. smithiana*, but missing from *P. abies*, *P. asperata*, *P. jezoensis*, *P. omorika* and *P. wilsonii*. A BLASTN analysis indicated that various sections (between 30 and 83 bp) of this region are present (86-100% identical) in multiple *Pinus* species, therefore its absence in the five aforementioned Eurasian *Picea* must be due to a deletion. Just 50 bp upstream of this deletion was a ~160-340 bp insertion unique to *P. abies*, *P. asperata*, *P. jezoensis*, *P. omorika* and *P. wilsonii*. Sequences contained within this insertion were found in no other database accessions by a BLASTN search. The variable size of this insertion was caused by large indels within it. The *P. abies* sample analyzed here contained the 68-bp insertion that was characterized as variation intraspecifically in *P. abies* by Grivet *et al.* (1999). Interestingly, the expansion of a 34-bp tandem repeat characterized in *P. abies* by Sperisen *et al.* (2001) was not present in the *P. abies* used here, however it was present in *P. omorika*. The 32-bp tandem repeat characterized in *P. abies* by Sperisen *et al.* (2001) was not present in any of the samples in this study. Other interesting indels included a 31-bp insertion caused by a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

tandem duplication event in *P. sitchensis. Picea smithiana* displayed a 255-bp deletion and a 29-bp insertion that was flanked by a 4-bp direct repeat but otherwise showed no signs of being a duplication event. *Picea breweriana* displayed a 113-bp deletion. There were many other small indels in the *nad1* intron, many of which appeared to have occurred via tandem duplication events.

There were a total of 53 base substitutions and 38 indels detected in *nad1* intron 2; 41 of these polymorphisms were parsimony-informative. Species that had identical *nad1* sequences included *P. chihuahuana* and *P. mexicana*, *P. engelmannii* and *P. glauca*, and *P. wilsonii*, *P. asperata*, and *P. jezoensis*. The MP (with and without bootstrapping) and ME *nad1* intron 2 trees are displayed in Figures 3, 4 and 5. The lengths of the MP and ME trees were not significantly different (KH test $P = 0.18$). Their overall topologies are similar; both demonstrate bootstrap support for the following clades: *P. schrenkiana* and *P. smithiana* (83-92%), *P. mariana* and *P. rubens* (70-80%), *P. chihuahuana* and *P. mexicana* (53-68%), *P. engelmannii* and *P. glauca* (81-98%), *P. engelmannii*, *P. glauca* and *P. sitchensis* (62-72%), and *P. asperata*, *P. jezoensis* and *P. wilsonii* (99-100%). It is surprising that the *P. engelmannii/P. glauca* and *P. chihuahuana/P. mexicana* clades did not have higher bootstrap scores since these pairs had identical sequences. It is possible that the bootstrap values were low because of the relatively small number of characters separating these clades from the rest of the *Picea* species.

In the MP tree, *P. abies* is contained within a clade with *P. asperata*, *P. jezoensis*, *P. wilsonii* and *P. omorika* (82%), whereas in the ME tree, this node is collapsed. The relationship between the *P. abies*, *P. asperata/P. jezoensis/P. wilsonii* and *P. omorika nad1* sequences is supported by the large >1 kb region deleted in these species and by the large

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

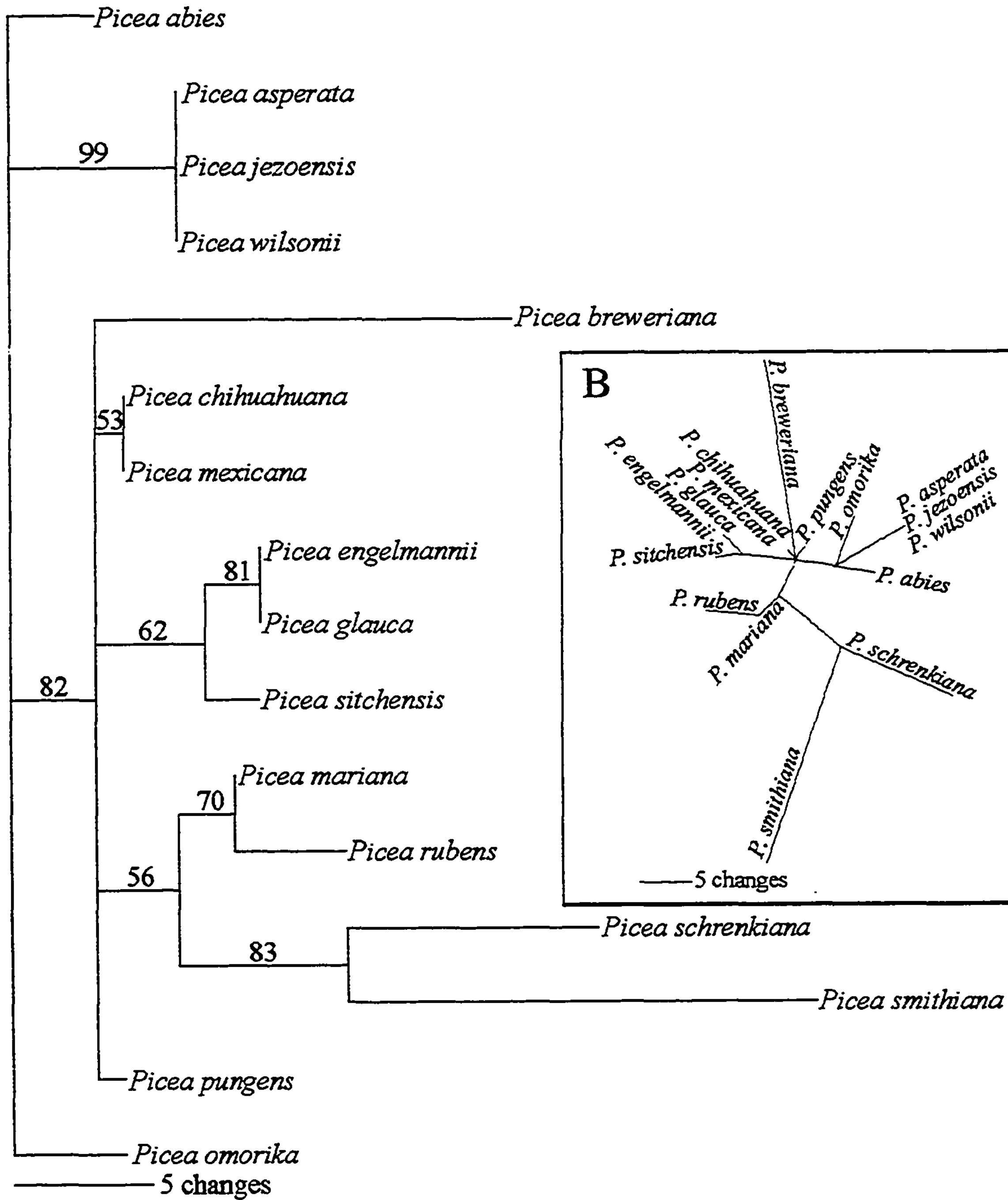


Figure 3. Maximum parsimony phylogenetic tree from mitochondrial *nad1* intron 2 sequence depicted as a rectangular phylogram (no outgroup was defined; A) and as an unrooted phylogram (B).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

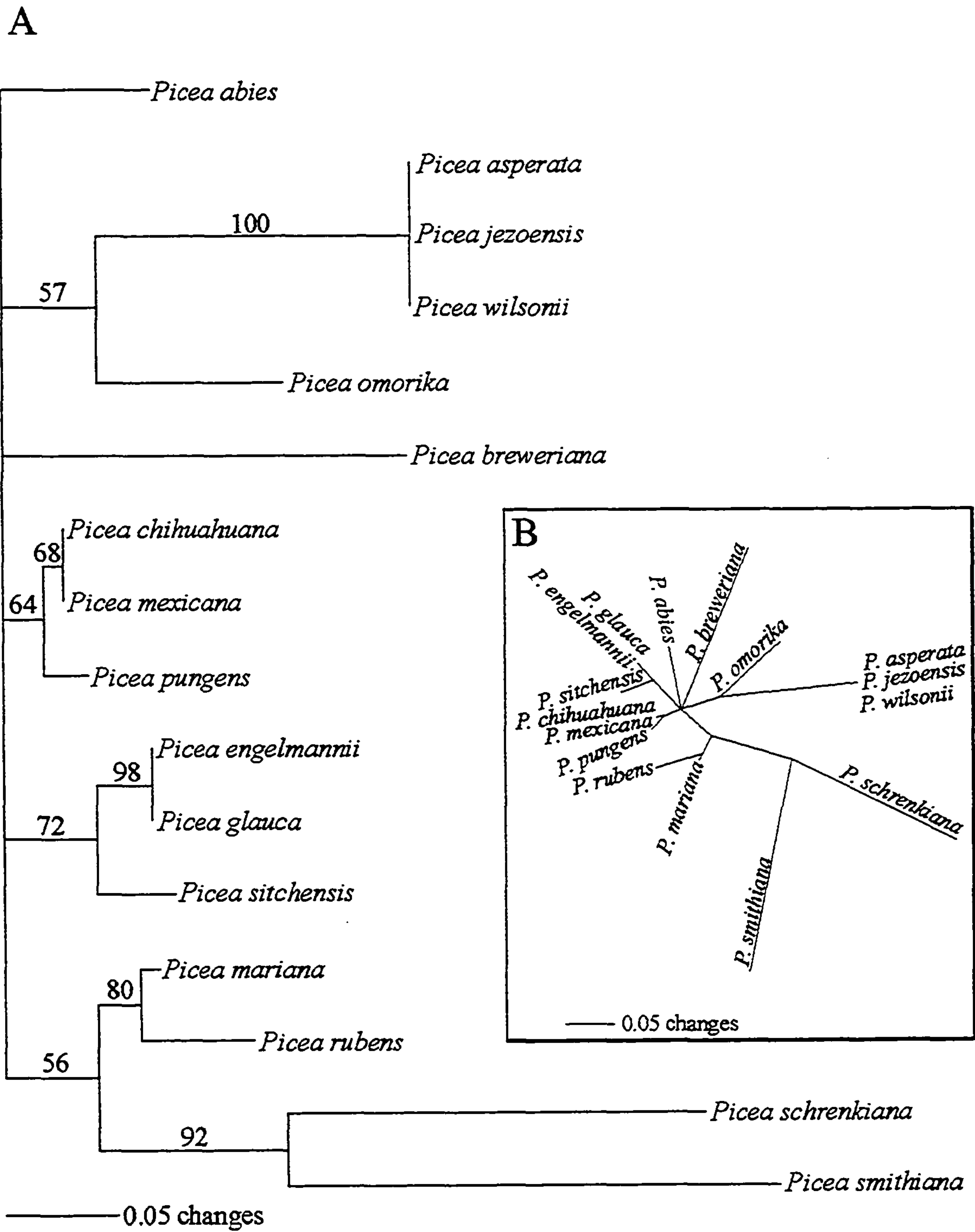


Figure 4. Minimum evolution phylogenetic tree from mitochondrial *nad1* intron 2 sequence depicted as a rectangular phylogram (no outgroup was defined; A) and as an unrooted phylogram (B).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

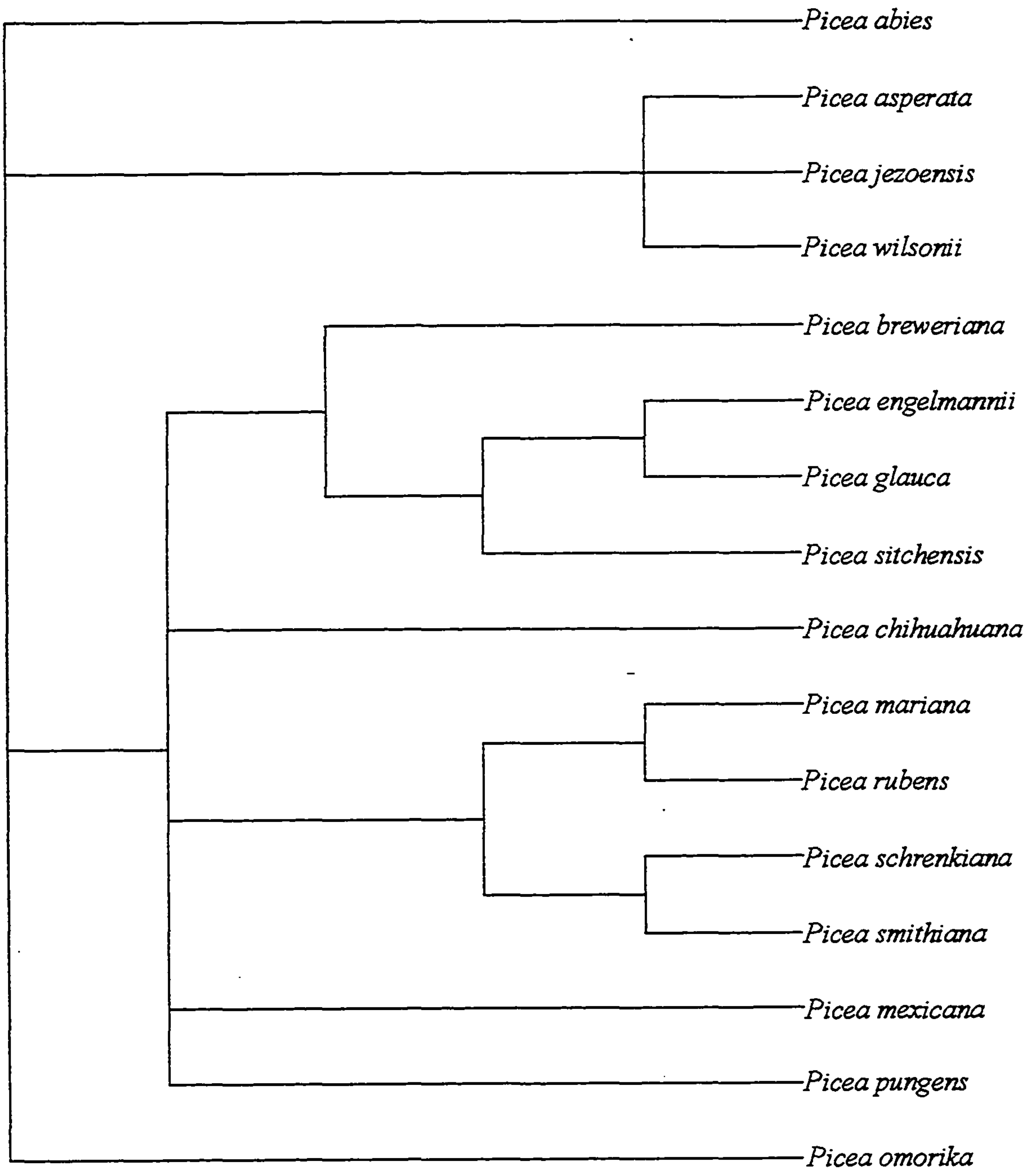


Figure 5. Maximum parsimony strict consensus cladogram from mitochondrial *nad1* intron 2 sequence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

~160-340 bp insertion unique to these species (see above). Furthermore, a relationship between *P. abies* and *P. omorika nad1* sequences is evidenced by the presence of the 34-bp tandem repeat (Grivet *et al.*, 1999) in *P. omorika* (see above). In the ME tree, *P. pungens* forms a clade with *P. chihuahuana* and *P. mexicana* (64%); this node is collapsed however in the MP tree. It is interesting to point out here that the relationship between *P. pungens* and *P. chihuahuana* was supported with a moderate bootstrap value in the chloroplast *trnK* ME tree, but not in the *trnK* MP tree. Finally, in the ME tree *P. breweriana* is part of a large unresolved polytomy. Its position is still largely unresolved within the bootstrap-supported MP tree; however in the strict consensus MP tree (no bootstrap; Figure 5) *P. breweriana* is grouped with *P. engelmannii*, *P. glauca* and *P. sitchensis*.

Comparison of Mitochondrial *Nad1* Trees to Chloroplast *TrnK* Trees

The topologies of the mitochondrial *nad1* and chloroplast *trnK* phylogenetic trees varied significantly. Primary differences included the positions of the *P. engelmannii*/*P. glauca* clade, *P. mexicana*, *P. omorika*, *P. sitchensis* and the *P. schrenkiana*/*P. smithiana* clade. In the chloroplast trees, *P. engelmannii* and *P. glauca* are grouped with *P. mexicana*, and this clade is sister to the *P. abies*/*P. asperata* /*P. jezoensis* /*P. wilsonii* clade in the chloroplast ME tree. In the mitochondrial trees however, *P. glauca* and *P. engelmannii* are sister to *P. sitchensis*, whereas *P. mexicana* is sister to *P. chihuahuana*. If the *nad1* parsimony analysis is carried out while constraining *P. mexicana* as a sister taxa to *P. engelmannii* and *P. glauca*, the tree is not significantly longer. However, if the constraint of *P. engelmannii* and *P. glauca* sister to the *P. abies*/*P. asperata* /*P. jezoensis*/*P. wilsonii* clade is imposed on the mitochondrial analysis, the tree is significantly longer ($P < 0.01$). Furthermore, when *P. sitchensis* is forced to be sister to *P.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

glauca and *P. engelmannii* in the chloroplast analysis, the resulting tree is also significantly longer ($P < 0.01$; Table 2).

In the chloroplast trees *P. omorika* is grouped with *P. mariana* and *P. rubens* with strong bootstrap support (91-99%). On the other hand in the mitochondrial trees, it is closely associated with *P. abies* and the Asian species *P. asperata*, *P. jezoensis* and *P. wilsonii*. If the monophyly of *P. rubens*, *P. mariana* and *P. omorika* is constrained, the mitochondrial *nad1* maximum parsimony trees are significantly longer ($P < 0.01$; Table 2). Finally, if *P. mariana* and *P. rubens* are forced to be sister to *P. schrenkiana* and *P. smithiana* in the chloroplast analysis, the trees are not significantly longer.

Discussion

The goal of this project was to compare phylogenetic gene trees from oppositely inherited cytoplasmic genomes of selected species of *Picea*. All of the North American species were included in this study because of the many questions regarding their phylogeography. *Picea abies* and *P. omorika* were included because they are the only two extant species native to Europe, and also to further investigate the relationship between *P. omorika* and the North American species *P. mariana* and *P. rubens*. The remaining Asian species were selected as widespread geographic representatives: northeastern Asia and Japan (*P. jezoensis*), China (*P. asperata* and *P. wilsonii*), central Asia (*P. schrenkiana*) and the western Himalayas (*P. smithiana*). *Picea asperata* and *P. wilsonii* also represent the *P. abies* and *P. brachytyla* alliances, respectively, in the phylogeny generated by Sigurgeirsson and Szmidt (1993). Furthermore, *P. jezoensis* has been postulated to be a link to the North American species (Wright, 1955) and it was closely associated with *P.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

pungens in Sigurgeirsson and Szmidt's chloroplast RFLP phylogenetic trees (Sigurgeirsson and Szmidt, 1993).

Species of *Pinus* were chosen as outgroups for the chloroplast DNA analysis because *Pinus* and *Picea* are considered to be sister genera (Chase *et al.*, 1993). The particular *Pinus* species were selected because 1) their complete (or nearly complete) *trnK* sequences were available in GenBank, 2) they represent the two subgenera of *Pinus* (*Pinus banksiana* and *Pinus thunbergii*: *Pinus*; *Pinus armandii*: *Strobus*), and 3) they represent North America (*Pinus banksiana*) and eastern Asia (*Pinus armandii* and *Pinus thunbergii*). *Pinus* could not be used as an outgroup for the mitochondrial analysis because it was not possible to align *Pinus* and *Picea nad1* intron 2 sequences. Therefore, no outgroup was defined for the mitochondrial analyses and trees were left unrooted. If a *Picea* species were to be chosen as an outgroup, either *P. breweriana*, *P. schrenkiana* or *P. smithiana* might be appropriate. These three species exhibit the most *nad1* sequence variation from the remainder of the *Picea* species (Appendix D). If a constant molecular clock model is assumed, these species would be the oldest and most closely related to the common ancestor of the *Picea*. Furthermore, all three of these species contain the large >1 kb region (deleted in some Asian species) that was determined by BLASTN to be present (in part) in the *nad1* intron 2 of *Pinus*.

Comparison of *TrnK* and *Nad1* Trees to Results of Other Systematic Studies

The topologies of the *trnK* trees presented here do not change upon the addition of sequence data from non-coding sections of the chloroplast *trnT-L-F* region (Germano *et al.*, 2002). However, there are several differences between these trees and the maximum parsimony tree generated from cpDNA RFLPs (Sigurgeirsson and Szmidt, 1993). In the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

RFLP tree, *P. pungens* was sister to *P. jezoensis*, *P. chihuahuana* was tightly nested in a group with *P. wilsonii* and *P. schrenkiana*, *P. smithiana* was clustered with *P. mariana*, *P. rubens* and *P. omorika*, and the *P. glauca* clade (*P. glauca*, *P. engelmannii* and *P. mexicana*) was basally located near the root of the tree. In the chloroplast *trnK* and mitochondrial *nad1* trees, the Asian species *P. asperata*, *P. jezoensis* and *P. wilsonii* form a single clade, *P. pungens* and *P. chihuahuana* are more closely related to each other than they are to the Asian species, and *P. schrenkiana* and *P. smithiana* are sister species (Figures 1-5).

In their article, Sigurgeirsson and Szmidt (1993) discuss some of the drawbacks to using RFLP data for phylogenetic analysis. Firstly, if some of the RFLPs were caused by rearrangements of the chloroplast genome, the interpretation of the data would be skewed. Secondly, there is a risk of homoplasy because restriction fragments of the same size could potentially be non-homologous. Furthermore, the presence of length mutations increases the level of potential homoplasy (Sigurgeirsson and Szmidt, 1993). It is for these reasons that the trees generated from sequence data are generally believed to be more robust. The species relationships in the *trnK* and *nad1* trees discussed above are more likely to be true than the ones depicted in the RFLP tree, especially the ones that are supported by both the mitochondrial and chloroplast sequence data.

In traditional systematic classifications of the *Picea*, morphological characters such as 4- or 2-angled needles and hard or flexible cone scales were weighted heavily (Aldén, 1987; Vidakovic, 1991). The species belonging to various groups based on these characters do not necessarily group together in the chloroplast and mitochondrial phylogenetic trees. For example, data from both organelles support the close association

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of *P. jezoensis* with *P. asperata* and *P. wilsonii*, however *P. jezoensis* possesses thin flexible cones scales whereas the other two species have hard cone scales. Likewise, *P. engelmannii* (thin cone scales) was consistently closely affiliated with *P. glauca* (hard cone scales). Another example involves two species possessing 2-angled or flat needles, *P. omorika* and *P. sitchensis*. According to the chloroplast and mitochondrial trees, these two species are much more closely related to other species than they are to each other. Although LePage (2001) discussed the relative congruency of cone scale morphology with the results of the cpDNA RFLP phylogeny (Sigurgeirsson and Szmidt, 1993), the important morphological characters do not seem to be consistent with the chloroplast and mitochondrial sequence data presented here.

Position of the *P. glauca* Clade

Interestingly, the topologies of the *nad1* and *trnK* phylogenetic trees differed from one another with respect to the position of the *P. glauca* clade. In the chloroplast trees, the clade is closely associated with the Eurasian species *P. abies*, *P. asperata*, *P. jezoensis* and *P. wilsonii*. In the mitochondrial trees, its position is relatively unresolved. The basal position of the *P. glauca* clade in Sigurgeirsson and Szmidt's cpDNA RFLP tree (Sigurgeirsson and Szmidt, 1993) led the authors (in part) to suggest a North American origin of the genus. Interpreting this result, however, should be approached with caution for several reasons. The separation of the *P. glauca* clade from the rest of the *Picea* species is only weakly supported by bootstrap (27% in the RFLP tree). The large (>1 kb) indel in the middle of the *nad1* intron is present in all the North American *Picea* species, including *P. glauca*. The fact that sections of this indel's sequence are present in multiple *Pinus* species suggests that the North American *Picea* species are more ancient in origin

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

than some of the Eurasian species; however the indel sequence is also present within the two Asian species *P. schrenkiana* and *P. smithiana*. The grouping of Eurasian species with North American species in both the chloroplast and mitochondrial trees suggests that *Picea* traveled cross-continentally more than once.

Positions of Individual Species: Incongruencies between Chloroplast and Mitochondrial Trees

Picea mexicana. The *trnK* sequence of the *P. mexicana* accession was identical to that of *P. glauca*, while its *nad1* intron sequence was identical to that of the *P. chihuahuana* accession. Therefore *P. mexicana* was sister to *P. glauca* in the *trnK* trees, yet sister to *P. chihuahuana* in the *nad1* trees. Morphological, phenolic and terpenoid analyses have indicated a close relationship between *P. mexicana* and *P. engelmannii*, so much so that it is considered a variety of *P. engelmannii* by some authors (Taylor *et al.*, 1994). *Picea engelmannii* and *P. glauca* are known to be closely related; there was only a single base substitution between their *trnK* sequences. *Picea glauca* and *P. engelmannii* are also known to readily introgress. In fact, trees of their sympatric zone are referred to as "interior spruce" (a complex of the two species) (Sutton *et al.*, 1994). *Picea mexicana* is also thought to cross with *P. engelmannii* (Gordon, 1992). Two scenarios could account for the *trnK* identity between *P. mexicana* and *P. glauca*: 1) *P. mexicana*, *P. glauca* and *P. engelmannii* are closely related species and there has been little chloroplast sequence divergence since their split from a common ancestor, or 2) the *P. mexicana* accession represents an introgressed population and its chloroplast sequence is a product of reticulate evolution. Similar scenarios (close phylogenetic relationship versus introgression) could be postulated to explain *P. mexicana*'s mitochondrial *nad1* identity with *P. chihuahuana*.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

These are perplexing in this case because *P. chihuahuana* is morphologically distinct and thought to be reproductively isolated (Taylor and Patterson, 1980; Gordon, 1992).

Picea omorika. The position of *P. omorika* with *P. rubens* and *P. mariana* in the *trnK* trees is consistent with the cpDNA RFLP trees (Sigurgeirsson and Szmidt, 1993). *Picea omorika*'s position on the mitochondrial *nad1* parsimony tree within the *P. abies/P. asperata/P. jezoensis/P. wilsonii* clade is interesting. In his inter-species crossability experiments, Wright (1955) deemed the artificial *P. omorika* X *P. abies* (female X male) cross to be unsuccessful, yet Hoffman and Kleinschmidt (1979) found the artificial cross of the two species to be successful in both directions. Although *P. omorika* was once widespread in Europe (before the Quaternary Period, 2 million years ago), it currently has a very narrow range that is sympatric with *P. abies* (Wright, 1955). Due to their geographic proximity, can the alliance between these two species' *nad1* sequences be explained by introgressive hybridization? After all, *P. abies* possesses much chloroplast (Sigurgeirsson, 1992) and mitochondrial (Sperisen *et al.*, 1998) variation that could be attributed to its introgressive hybridization with other neighboring *Picea* species such as *P. obovata* (Siberian spruce).

The true organismal phylogeny of species like *P. mexicana*, *P. engelmannii*, *P. chihuahuana* and *P. omorika* should not be determined without conducting range-wide population surveys of their chloroplast and mitochondrial haplotypes. Such studies would provide evidence of past introgressive hybridization and allow individuals with the most common organelle haplotypes to be selected to represent the species in a robust phylogeny. In a range-wide population study of *P. glauca*, *P. mariana* and *P. rubens*, three significantly divergent mitochondrial haplotypes were detected within *P. mariana* (Chapter

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

II). The most common of these haplotypes was unique to *P. mariana*, while the other two were concluded to be the result of organelle capture through introgression with *P. rubens* and with *P. glauca*, even though *P. mariana* and *P. glauca* are distantly related species and were previously thought not to cross hybridize (see Chapter II). These results, in addition to the vast knowledge of introgressive-hybridization among the *Picea*, highlight the danger of choosing a single individual to represent a species without prior population surveys. The incongruence of the chloroplast and mitochondrial phylogenetic trees presented here may be the result of hybrid or introgressed individuals representing particular species (e.g. *P. mexicana* or *P. omorika*).

Future Studies

The slow rate of mitochondrial primary sequence evolution combined with the complex indels of the *nad1* intron 2 makes this region less desirable for phylogenetic inquiries at the intra-genus level. Because of the opposite inheritance of the organellar genomes however, mitochondrial markers should be used in combination with chloroplast markers in order to prevent using putative hybrids/introgressants as species representatives. The nuclear genome of plants has been shown to undergo a higher frequency of base substitutions than the chloroplast genome (Palmer, 1990). The internal transcribed spacer (ITS) of the nuclear ribosomal DNA tandem repeat unit has been useful in phylogenetic studies of angiosperms (Baldwin *et al.*, 1995). However, ITS was determined to be a poor candidate for a phylogenetic assessment of *Picea* because it is significantly heterogeneous within single individuals of several *Picea* species (Germano and Klein, 1999; Wright *et al.*, 2001). The nuclear 4CL gene was used in a phylogenetic study of the members of the Pinaceae. Its exons were estimated to have diverged twice as fast as the chloroplast *matK*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

gene and five times as fast as the mitochondrial *nad5* gene. Sequence comparisons of multiple 4CL clones suggested that the gene was present as a single copy in *Picea smithiana* (Wang *et al.*, 2000). A nuclear single-copy gene such as 4CL in combination with non-coding chloroplast regions (such as *trnK* and *trnT-L-F*) and appropriate population surveys would make for a more robust species phylogeny of the *Picea*.

Acknowledgements

Much of the data presented in this chapter was collected by Mr. Aaron Thorner and Mr. Richard Roy. I would like to thank Dr. Marianne Litvaitis, Department of Zoology, UNH for her help with some of the phylogenetic analyses. Support was received in part by the Undergraduate Research Opportunities Program, University of New Hampshire, Durham, NH. This research was supported by a grant from the New Hampshire Agricultural Experiment Station (NHAES) MacIntire-Stennis Project MS27.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

CHAPTER IV

SPECIES-IDENTIFICATION OF LATE-PLEISTOCENE AND EARLY-HOLOCENE *PICEA* MACROFOSSILS FROM THE NORTHEASTERN UNITED STATES: USING ANCIENT DNA TO IMPROVE VEGETATION HISTORY AND PALEOCLIMATE RECONSTRUCTION

Abstract

The Quaternary period (the last 2 million years) is characterized by severe oscillations in climate that caused the repeated formation and retreat of glaciers in North America and Europe. These glacial cycles had drastic effects on flora and fauna, causing migrations of many species. The migration histories of white spruce (*Picea glauca*), black spruce (*P. mariana*) and red spruce (*P. rubens*) are of particular interest to paleoclimatologists because these species have differing climate tolerances and are therefore used to infer climate conditions that occurred during the last ice age. *Picea* macrofossils were recovered from sediment cores of Browns Pond, Virginia, dating approximately 20,000-10,000 years ago. DNA was extracted from macrofossils that 1) were separated from the sediment and stored in water at 4°C for 5-7 years, 2) were stored within the core sediment at 4°C for 5-7 years and 3) were stored within the core sediment at −20°C for a few months. Macrofossils that were stored at 4°C for long periods of time yielded high molecular weight DNA that was concluded to be contemporary in nature presumably due to bacterial and/or fungal contamination that occurred after the cores were removed from the ground. DNA extracts from some macrofossils yielded low molecular weight DNAs that were detected by agarose gel electrophoresis (100-1000 bp

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DNA smear from cores stored at −20°C) or by Southern blot and hybridization with *Picea*-specific DNA probes (~250 to 650 bp DNA from cores stored at 4°C). These results suggested that authentically ancient DNA survived in some of the10-20,000 year old Browns Pond *Picea* macrofossils. All attempts to amplify ancient DNA by PCR from macrofossils and from fossil pollen failed. The primary impediments included inhibition of PCR, non-specific amplification, contamination, lack of template DNA (regarding the pollen) and primer design.

Introduction

The Quaternary period (the last 2 million years) is characterized by climate oscillations between glacial and interglacial extremes. The overall warming trend that took place during the transition from the Last Glacial Maximum (LGM; ~21,000 calibrated calendar years ago) to the present interglacial or the Holocene (which began ~11,000 cal. years ago) was sporadically interrupted by pronounced and abrupt changes in climate. Paleoclimate research of this transitional period has revealed evidence from ice (Alley *et al.*, 1993; Mayewski *et al.*, 1993), marine (Lehman and Keigwin, 1992; Hughen *et al.*, 1996), and terrestrial sediment cores (Levesque *et al.*, 1993; MacDonald *et al.*, 1993; Björck *et al.*, 1996) strongly suggesting that dramatic changes in regional, and perhaps global climate can occur over as little time as a few decades.

A global array of lake and bog core sediments contains pollen and plant macrofossils. The sediment or macrofossils can be radiocarbon dated by Accelerator Mass Spectrometry (AMS)[1] and used to reconstruct vegetation histories. The histories of

---

[1] The radiocarbon time scale is believed to diverge from the calendar year scale. This is based on isotope dating of marine corals and annual layer counting in Greenland ice cores. Hence, 11,000 cal. yr is roughly equivalent to 10,000 ^{14}C yr BP, and 21,000 cal. yr

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

North America and Europe, from before the LGM through the Holocene, show great shifts in vegetation assemblages (Watts, 1983; Webb, III *et al.*, 1993). In the northern hemisphere, the range of forest trees changed as the ice sheets retreated during the late-glacial. Broad scale maps of vegetation inferred from pollen diagrams illustrate the magnitude of forest response to climatic changes (Delcourt and Delcourt, 1987a). Late-glacial and Early-Holocene regional climate oscillations have been inferred from shifts in vegetation assemblages (for example Mott *et al.*, 1986; Levesque *et al.*, 1993; MacDonald *et al.*, 1993; Björck *et al.*, 1996; Kneller and Peteet, 1999). Some of these oscillations are coincident with climate changes inferred from ice and marine cores (Mott *et al.*, 1986; Peteet *et al.*, 1990; Koç Karpuz and Jansen, 1992; Levesque *et al.*, 1993; Peteet *et al.*, 1993; Björck *et al.*, 1996). Vegetation histories have been extremely useful in quantifying past climate change at the regional level and the consequences of past climate change upon ecosystems. Knowledge of Late-Quaternary temperate vegetation history is important to our understanding of contemporary patterns and processes of temperate forest ecosystems. It also has implications for future environmental/climatic changes and their consequences (Delcourt and Delcourt, 1987a).

Importance of *Picea* in North American Quaternary Paleoclimate Reconstructions

In eastern North America, *Picea* (spruce) has been a critical taxon in defining previous temperature changes (Watts, 1983; Davis, 1983a; Jacobson *et al.*, 1987; Webb, III, 1988). For example, at 11,000 ^{14}C yr BP (radiocarbon years before present where present equals 1950 A.D.), the resurgence of *Picea* (along with increases in *Abies*, *Betula*

---

is roughly equivalent to 18,000 ^{14}C yr BP. In this chapter, all radiocarbon dates are denoted as ^{14}C yr BP. Dates lacking the ^{14}C notation can be assumed to be in calendar years.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

and *Alnus*) at sites in southern New England has been used as evidence for an estimated 3 to 4°C sudden cooling in mean July temperatures (Peteet *et al.*, 1994). This cooling occurred within the Younger Dryas Chronozone (11,000-10,000 years ago), a time period when several regions of the earth experienced severe and rapid cooling (Hansen *et al.*, 1984; Johnsen *et al.*, 1992). Several quantitative methods have been used to reconstruct vegetation from fossil pollen databases, and *Picea* is a key taxon in these reconstructions. These quantitative vegetation reconstructions have been used to propose reconstructions of regional climates of the last 21,000 years. The vegetation-based climate reconstructions have been compared to paleoclimate simulations from General Circulation Models (GCMs), and in some cases, they are very different.

Presently, three *Picea* species occupy eastern North America: *Picea glauca* (Monech) Voss (white spruce), *P. mariana* (Mill) B. S. P. (black spruce), and *P. rubens* Sarg. (red spruce). While their ranges overlap (Morgenstern and Farrar, 1964; Little, Jr., 1971), each species has distinctive climatic and ecological preferences and tolerances (Table 1; Fowells, 1965; Gordon, 1976; Nienstaedt and Zasada, 1990; Viereck and Johnston, 1990; Blum, 1990). Sediment cores from many lakes, bogs and fens in eastern North America have been analyzed for their pollen and plant macrofossils. *Picea* macrofossils (e.g. needles, cones, seeds) are a frequent component of many eastern United States sites older than 9,000 ^{14}C yr BP (Delcourt and Delcourt, 1985; Jackson *et al.*, 1997). Fossil *Picea* pollen and macrofossils can be differentiated morphologically from other conifers. Typically, needles and seeds of *P. glauca*, *P. mariana* and *P. rubens* are not morphologically distinctive, therefore fossil *Picea* needles and seeds cannot be identified to the species level. Species-identification of Late-Pleistocene and Early-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Climatic and ecological preferences, tolerances and life history characteristics of the three *Picea* species native to eastern North America.[a]

	Picea glauca	*Picea mariana*	*Picea rubens*
General Climate Conditions	cold, moist to dry	cold, humid to dry subhumid	cool, moist
Mean July Temperature	10-18°C (50-64°F) [43°C (110°F)][b]	10-21°C (51-70°F) [41°C (105°F)]	18-21°C (65-70°F) [32-35°C (90-95°F)]
Mean January Temperature	[54°C (65°F)]	30-6°C (22-21°F) [62°C (79°F)]	12-2°C (10-35°F) [34-17°C (30-0°F)]
Mean Annual Precipitation	250-1270 mm	380-760 mm [150-1520 mm]	910-1320 mm
Freezing/Winter Desiccation	resistant	resistant	prone
Shade Tolerance	moderate	low	high
Maximum Age	250-300 yr [1000 yr]	200-250 yr [280 yr]	>400 yr
Seed Production Age	60 yr [30 yr]	[10 yr]	45 yr [30 yr]
Good Seed Crop	every 2-6 yr	every 2-6 yr	every 3-8 yr
Elevation	0-1520 m	150-760 m [0-1830 m]	0-1370 m [1370-1520 m][c]

[a] (Fowells, 1965; Gordon, 1976; Nienstaedt and Zasada, 1990; Viereck and Johnston, 1990; Blum, 1990)

[b] Recorded extremes are indicated in [brackets].

[c] Elevations in the central Appalachian Mountains.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Holocene *Picea* macrofossils would enable more detailed vegetation history and hence more precise estimates of climate changes that occurred following the LGM in North America.

<u>*Picea* Species as Climatological/Ecological Indicators</u>. *Picea rubens*' current range (Figure 1A) is from the Appalachians in North Carolina to the Maritimes including New England, New York, southern Quebec and restricted areas of Ontario (Morgenstern and Farrar, 1964; Fowells, 1965; Little, Jr., 1971; Blum, 1990). In general, its range is associated with a cool moist climate regime with mean July temperatures of 18 to 21°C (65 to 70°F), mean January temperatures of −12 to 2°C (10 to 35°F) (Fowells, 1965) and mean annual precipitation of 910-1320 mm (Blum, 1990). *Picea rubens* grows in areas where average maximum summer temperatures of 32 to 35°C (90 to 95°F) and average minimum winter temperatures of −34 to −17°C (−30 to 0°F) have been reported (Fowells, 1965).

The current range of *P. glauca* (Figure 1B) overlaps with *P. rubens* primarily in northern New England and eastern Canada, and extends west to Alaska and north to the tree-line, (Morgenstern and Farrar, 1964; Fowells, 1965; Little, Jr., 1971; Nienstaedt and Zasada, 1990; Blum, 1990). *Picea glauca* is considered one of the most cold-hardy conifers of North America (Fowells, 1965). The northern limit of its range has a mean July temperature of 10°C (50°F), whereas its southern limit has a mean July temperature of 18°C (64°F). Extreme temperatures of −54°C (−65°F) and 43°C (110°F) have been recorded in its range.

The range of *P. mariana* (Figure 1C) extends from the eastern coast of Canada west to Alaska and north to the tree line. It grows south to central British Columbia,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

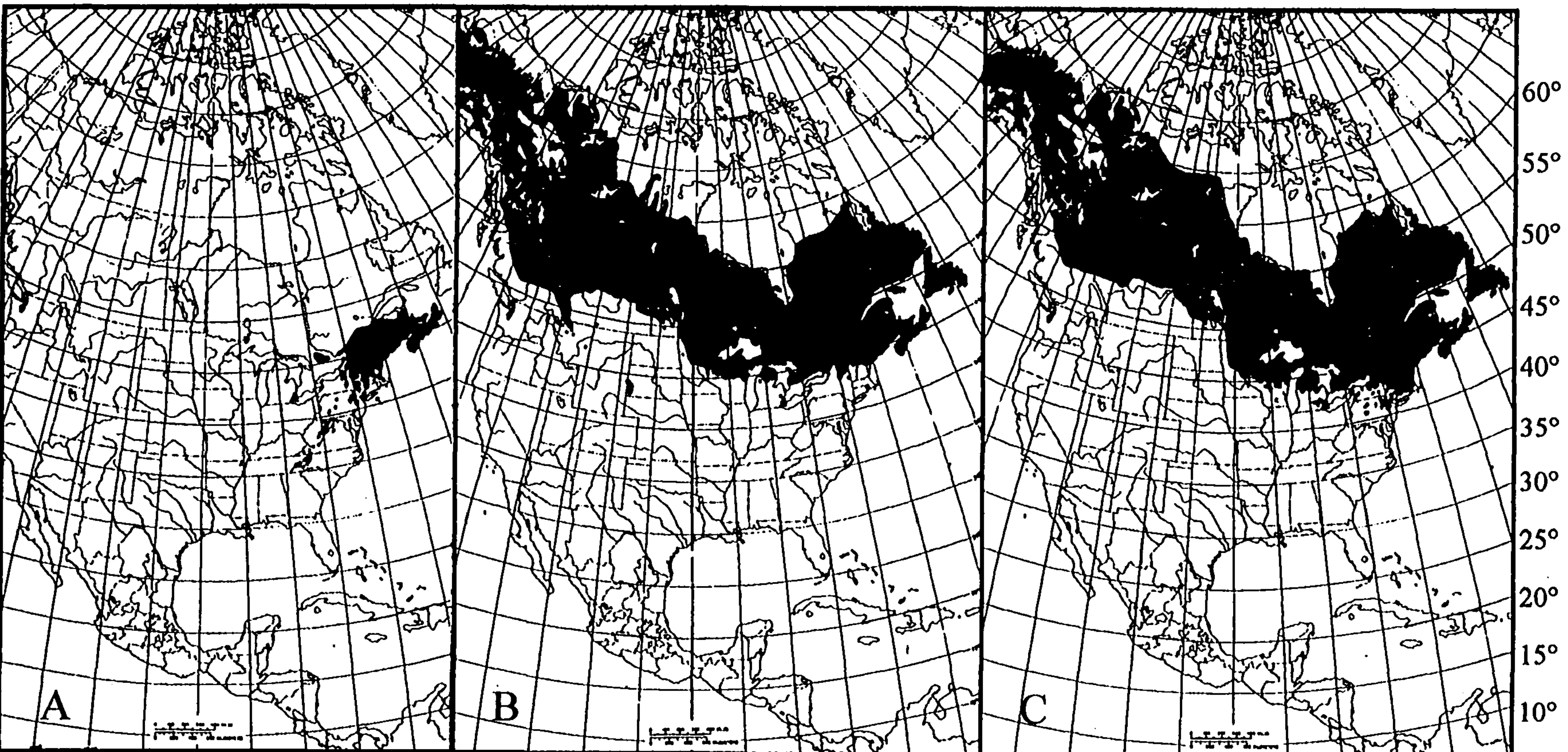


Figure 1. Current ranges of *Picea rubens* (A), *P. glauca* (B), and *P. mariana* (C) (from Little, Jr., 1971).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

southern Manitoba, central Minnesota and New York with isolated populations extending into Pennsylvania (Morgenstern and Farrar, 1964; Fowells, 1965; Little, Jr., 1971; Viereck and Johnston, 1990). *Picea mariana* is generally associated with cold humid to dry subhumid climate. The northern limit of its range experiences a mean January temperature of −30°C (−22°F) and a mean July temperature of 10°C (51°F). The southern limit of its range has a mean January temperature of −6°C (21°F) and a mean July temperature of 21°C (70°F). *Picea mariana* tolerates extreme temperatures of −62°C (−79°F) and 41°C (105°F) (Fowells, 1965; Viereck and Johnston, 1990). It is resistant to freezing and winter desiccation (Gordon, 1976) and can grow on permafrost due to its shallow rooting habit (Viereck and Johnston, 1990).

The Last Glacial Maximum

The Quaternary Period comprises the Pleistocene epoch followed by the Holocene epoch that began approximately 10,000 ^{14}C yr BP. The Quaternary is characterized by its many glacial-interglacial cycles. During the LGM (~18,000 ^{14}C yr BP) the Laurentide ice sheet of North America covered most of Canada and extended as far south as ~40°N over the northeastern United States. The southeastern portion of the United States remained unglaciated, supporting many flora and fauna of North America.

Analysis of ice cores, marine sediments and terrestrial records provide evidence of changing global climate conditions during the last glaciation. During the LGM, sea levels were at their lowest, and ocean surface temperatures, inferred from foraminiferal plankton assemblages preserved in marine sediments, reached a minimum. Atmospheric temperatures, inferred from isotopic oxygen levels in Greenland and Antarctic ice cores, were at a minimum during the LGM. Following the LGM, a general warming trend

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

corresponded with the retreat of the ice sheets. This warming trend was punctuated by distinct changes in climate (Dansgaard *et al.*, 1984; Lehman and Keigwin, 1992; Johnsen *et al.*, 1992; Alley *et al.*, 1993; Levesque *et al.*, 1993; MacDonald *et al.*, 1993; Mayewski *et al.*, 1993; Hughen *et al.*, 1996).

Migrational History of *Picea*. *Picea* is a climate-sensitive genus: the expansion and contraction of its range in the last 20,000 years raises several questions relevant to establishing Late-Quaternary temperature fluctuations. *Picea* pollen and macrofossils, such as needle fragments, seeds, sterigmata, twigs, buds or cones, are frequently identified in North American Quaternary sediments and are often given significant weight when making climatic interpretations (Watts, 1983; Davis, 1983a; Jacobson *et al.*, 1987; Webb, III, 1988).

The vegetation history of the unglaciated southeastern United States since the LGM has been reviewed by Whitehead (1973), Davis (1983b), Watts (Watts, 1980b; 1983), Delcourt and Delcourt (1985; 1987b), Jacobson *et al.* (1987), Webb (1988) and Jackson *et al.* (1997) among others. Using taxon calibrated data from 162 pollen records, Delcourt and Delcourt (1987a) presented quantitative reconstructions of past eastern North American forest composition at 500 and 1000-year intervals for the last 20,000 years. Overpeck *et al.* (1992) analyzed 11,700 fossil and 1744 modern pollen assemblages to map the different forest types that occurred since the LGM.

At the LGM, discontinuous patches of tundra were located just south of the glacial margin and at high elevations in the Appalachian Mountains (Watts, 1979; Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b). According to pollen records, boreal tree taxa including *Picea* and northern *Pinus* (usually thought to be *Pinus*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

banksiana) were distributed across the northern half of the unglaciated United States south of the ice sheet to approximately 33°N during the full-glacial (Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b; Overpeck *et al.*, 1992). *Picea* was the dominant taxon in this forest (Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b). *Picea* macrofossil needles dated 17,300 ^{14}C yr BP which were isolated from Browns Pond in western Virginia (~38°N) provide definitive proof of *Picea*'s existence at northern latitudes (Kneller and Peteet, 1993).

South of 33°N, the full-glacial range of *Picea* extended down the Mississippi valley (Delcourt and Delcourt, 1987a; 1987b). In Tunica Hills, Louisiana-Mississippi (~30°N), many *Picea* macrofossils were found which predated the LGM, and *Picea* composed 40-80% of the pollen assemblages from 25,000 to 17,500 ^{14}C yr BP (Givens and Givens, 1987; Jackson and Givens, 1994). Pollen also indicated the presence of *Picea* in northwest Florida around 14,330 ^{14}C yr BP, and in central Texas before 15,000 ^{14}C yr BP west (Holloway and Bryant, Jr., 1984; Watts *et al.*, 1992).

During the late-glacial (~16,000 to ~13,000 years ago), a northward expansion of the *Picea* range occurred as *Picea* trees successfully invaded tundra and other areas of deglaciated landscapes. *Picea* was dominant in the northern and western portions of the boreal forest as northern *Pinus* began to diminish (Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b). The arrival of *Picea* to southern Pennsylvania is evidenced by macrofossils dated 15,210 ^{14}C yr BP from Crider's Pond, just south of 40°N (Watts, 1979). Farther north, *Picea* macrofossils dating to 12,290 ^{14}C yr BP were collected from Alpine Swamp in northeastern New Jersey (north of 40°N) (Peteet *et al.*, 1990).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

By 12,000 yr BP, during the transition from Late-Pleistocene to Early-Holocene, *Picea* had advanced across the deglaciated landscape to ~45°N. The southern margin of its range, however, still remained at 33°N. It was dominant between 40°N and 42°N, but its numbers had diminished south of 40°N (Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b). Between 12,000 and 8,000 yr BP, *Picea* advanced across New England, Nova Scotia and Newfoundland and persisted at higher elevations of the central Appalachians. Its populations collapsed, however, in the forested Great Lakes region. By 4,000 yr BP, *Picea* became the dominant tree taxon in the northern boreal forest region in a band from 50°N to 58°N. The southern limit of its range had retreated to 42°N in the Great Lakes region, and it persisted at high elevations in the central and southern Appalachians (Delcourt and Delcourt, 1987a).

Quantifying Climate Conditions from Fossil Pollen Assemblages

Pollen is preserved in most terrestrial (and even marine) sediments deposited in aquatic environments. The structurally strong pollen exine helps preserve individual grains in sediments as old as 100 million years. Most pollen can be identified morphologically to the genus level with a standard optical light microscope. Arboreal taxa commonly analyzed in paleoecological studies of eastern North America include *Alnus* (alder), *Abies* (fir), *Acer* (maple), *Betula* (birch), *Castanea* (chestnut), *Fagus* (beech), *Picea*, *Pinus* (pine), *Quercus* (oak) and *Tsuga* (hemlock). A large data set of pollen stratigraphies (both fossil and modern) from all continents is available (http://www.ngdc.noaa.gov/paleo/pollen.html) along with plant macrofossil data (http://www.ngdc.noaa.gov/paleo/macrofossils.html).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

One limitation of using arboreal pollen in vegetative reconstructions is that many arboreal taxa have wind-dispersed pollen that can travel up to 10-100 km from their original source (Delcourt and Delcourt, 1987b). Furthermore, pollen grains are generally only identified to the genus level, and are often referred to as pollen "types" (Delcourt and Delcourt, 1987b). Unlike pollen, macrofossils are usually deposited near their point of origin. Therefore, macrofossils provide definitive proof of the local presence of particular plant taxa (Davis, 1983b). Plant macrofossils are often analyzed in conjunction with pollen studies (see for example Whitehead, 1973; Watts, 1979; Givens and Givens, 1987; Peteet *et al.*, 1990; Watts *et al.*, 1992; Overpeck *et al.*, 1992; Kneller and Peteet, 1993; Kneller and Peteet, 1999). Furthermore, some macrofossils can be identified morphologically to the species level thereby making vegetation history and ecological reconstructions more precise.

Several methods have been developed to quantify regional climate conditions, which are reflected by pollen records. Pollen response surfaces are functions that describe the way in which a single pollen taxon's abundance depends on the combined effects of two or more environmental factors (e.g. annual precipitation, mean July temperature, mean January temperature, etc.). Response surfaces for individual taxa are generated from contemporary pollen assemblages and modern climate variable measurements. Fossil pollen assemblages are then used to interpolate a range of paleoclimate conditions. The range of conditions is narrowed down for each different taxon that is used for interpolation (Bartlein *et al.*, 1986; Prentice *et al.*, 1991; Webb, III *et al.*, 1993).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The modern analogue method compares modern pollen assemblages to fossil pollen assemblages, employing a variety of dissimilarity measures. For example, Overpeck *et al.* (1992) used the squared-chord distance dissimilarity coefficient as a measure of pollen assemblage dissimilarity. Assemblages are considered analogous if the taxonomic make-up is the same and if abundances of pollen types are similar (i.e. below a given dissimilarity threshold). The climate conditions of the closest modern analogue are then assumed as the local (or regional, depending on the number of fossil sites analyzed) climate for the period from which the fossil pollen originated (Delcourt and Delcourt, 1987a; Delcourt and Delcourt, 1987b; Overpeck *et al.*, 1992).

A drawback of the modern analogue method is that many fossil pollen assemblages do not have modern analogues (Davis, 1983b; Delcourt and Delcourt, 1987b). In fact, much of the forests covering the eastern United States during the transition of the Pleistocene to the Holocene lacked good analogues to any modern-day forests (Delcourt and Delcourt, 1987a; Overpeck *et al.*, 1992). During periods of climate change, vegetation is thought to be in dynamic equilibrium (Prentice *et al.*, 1991). This means that vegetation changes in response to climate change occur continually over large scales of space (continental scale) and time (resolution of at most 1,500 years). Furthermore, different taxa respond individually to climate change and therefore migrate at different rates and in different directions (Davis, 1983b; Bartlein *et al.*, 1986; Delcourt and Delcourt, 1987a; Prentice *et al.*, 1991). Due to the dynamic nature of vegetation change during times of climate change, it is not surprising that many Late-Pleistocene and Holocene forests were very different from those that exist today. Macrofossil

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

assemblages can be used in conjunction with pollen records to help define analogous and non-analogous forest assemblages (Overpeck *et al.*, 1992).

Estimates of Full- and Late-Glacial Climate Conditions in Eastern North America

Few estimates of climate change during the LGM and late-glacial exist for unglaciated eastern North America, most of which are based on fossil pollen records. Watts (1980b) estimated LGM southeast United States climate by comparing modern and LGM distributions of *Pinus banksiana* (jack pine). Transferring modern central Maine climate to Columbia, South Carolina, he estimated a mean January temperature of −10.1°C ($\Delta T = -17.8$°C; where ΔT = past temperature − modern temperature), a mean July temperature of 19.9°C ($\Delta T = -7.3$°C), 114 frost free days (Δdays = −134) and 1050 mm annual precipitation ($\Delta P = -10$ mm). He believed this conservatively estimated the change since the climate is even colder and drier inland from the Maine coast. Whitehead (1981) found the closest modern analogue to the Rockyhock Bay, North Carolina LGM assemblage to be the boreal region from Ontario and Saskatchewan, Canada. This analogue corresponds to a slightly colder and more seasonal climate: for mean January temperature, $\Delta T = -25$°C, and for mean July temperature, $\Delta T = -10$°C.

Using pollen response surfaces (Bartlein *et al.*, 1986), the 18,000 ^{14}C yr BP climate inferred from six taxa yielded a 10°C cooling in mean January temperatures and 8°C cooling in mean July temperatures at approximately 35°N on the Atlantic Ocean coast (Prentice *et al.*, 1991; Webb, III *et al.*, 1993). From approximately 26 to 40°N, estimated precipitation ranged from 1200 to 600 mm per day which represents a 20 to 40% decrease from modern observed values.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Temperature and precipitation estimated from vegetation-based methods have been compared to climate simulations of General Circulation Models (GCMs). These 3-dimensional computer models can vary the prescribed boundary conditions, such as solar insulation, land surface elevation, continental ice distribution, sea level, land surface albedo (fraction of incident radiation that is reflected) and sea surface temperature, in order to simulate past climates (Broccoli and Manabe, 1987). GCMs can also be used to predict future climate changes pending a given change in a specific boundary condition.

Climate simulations of the past 21,000 years in eastern North America using version 1 of the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM1) (Kutzbach *et al.*, 1998) were summarized by Webb III *et al.* (1998). This GCM simulated LGM January temperatures 8-16°C lower and July temperatures 2-8°C lower than present. These simulated temperatures, especially those for July, are significantly higher than vegetation-based estimates by Watts (1980b), Whitehead (1981) and Bartlein *et al.* (1986). CCM1 simulations for 16,000 and 11,000 years ago were also markedly different from those inferred by from pollen data (Webb, III *et al.*, 1998).

Morphological Methods to Distinguish Fossil *P. glauca*, *P. mariana* and *P. rubens*

Positive identifications of *Picea* pollen and/or macrofossils to the level of species would make past forest reconstructions, and hence the ecological and climatic interpretations inferred from them, more detailed and precise. Measurements of grain size and distinguishing morphological characteristics have been used individually or in combination to separate modern *P. glauca*, *P. mariana* and *P. rubens* pollen (Davis, 1958; Birks and Peglar, 1980; Hansen and Engstrom, 1985). These pollen studies are often based on populations from small geographic regions, and hence are unlikely to have

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

encompassed the full range in pollen morphology (B. Hansen, pers. comm. to M. Kneller, 1992) and the whole phenotypic range of each species. Therefore identified pollen morphology characters do not reliably represent the individual species.

A few researchers have tentatively identified fossil *Picea* pollen to the species level. Watts (1980a) and Whitehead (1981) used grain size (Davis, 1958; Watts, 1979) to distinguish *Picea* species. Watts (1980a) identified primarily *P. rubens* pollen with some *P. glauca* pollen from White Pond, South Carolina in the central Appalachians during the late-glacial (from roughly 13,000 to 10,000 ^{14}C yr BP). Whitehead (1981) indicated the possible presence of two *Picea* species from Rockyhock Bay on the coast of North Carolina that dated from 21,000 to 10,000 ^{14}C yr BP. He believed one corresponded to *P. mariana* and did not identify the second. Further south in central Texas, *P. glauca* pollen was identified in sediments older than 15,000 ^{14}C yr BP (Holloway and Bryant, Jr., 1984).

Reliable species-identification of *Picea* macrofossil needles and seeds cannot be accomplished using traditional morphological characters. Identifications of macrofossil seed cones have been made to the species level (LePage, 2001), however very few sites in North America yield intact cones (Critchfield, 1984) and even these identifications are not always dependable. For example, fluvial terrace deposits of Tunica Hills, Louisiana-Mississippi have yielded a large number of intact *Picea* cones predating the LGM (Givens and Givens, 1987; Jackson and Givens, 1994). Givens and Givens (1987) assigned these cones to *P. glauca*, and radiocarbon dates for *Picea* wood from the same deposits ranged between 24,920 and 12,430 ^{14}C yr BP. Jackson and Givens (1994) questioned the taxonomic placement of *P. glauca* for the Tunica Hills cones.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subsequently, evidence from fossil cone morphology and fossil needle anatomy suggested that a now-extinct *Picea* species occupied the Mississippi Valley before the LGM (Jackson and Weng, 1999), a species likely misidentified as *P. glauca* in previous studies.

Resin cavity pattern in cross-sections of fresh *Picea* needles show some potential to distinguish *P. glauca* from *P. mariana* and *P. rubens* (Duman, 1957). *Picea mariana* and *P. rubens* have two continuous resin ducts running the entire length of needle, therefore every cross-section, from base to tip, reveals two resin ducts. *Picea glauca* resin canals are interrupted by transverse partitions of mesophyll, therefore cross-sections reveal two, one or no resin canals (Duman, 1957). Obviously, the whole needle is required for this technique, and macrofossil needles are usually in fragments. Furthermore, this technique is not always reliable (Kneller and Peteet, pers. comm.).

It is desirable to develop reliable methods for the species-identification of *Picea* macrofossil needles or seeds because these are commonly found at many levels throughout lake core sediments. Identifying which *Picea* species were present at several intervals throughout a core would enable a precise reconstruction of the paleovegetation over a long time span (e.g. 10,000 years). *Picea* needles and seeds are commonly found in lake sediments throughout eastern North America that date from the Late-Pleistocene through the Holocene. Species-identification of these macrofossils at multiple cross-continental sites would enable the identification of the range of each *Picea* species. This would provide a detailed regional vegetation history since the LGM, which could ultimately be used to make more accurate paleoclimate reconstructions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Molecular Methods to Distinguish Fossil *P. glauca, P. mariana* and *P. rubens*

New molecular biology techniques provide a potential means to identify macrofossils at the species level. DNA can be isolated from fossil tissues that have been well preserved. Pääbo (1989) demonstrated that only DNA fragments between 40 and 500 bp, mostly between 100 and 200 bp, could be isolated from dry remains of animal soft tissue ranging in age from 4 years to >50,000 years old. The yield of these DNAs varied between 1 and 200 μg per gram of dry tissue (Pääbo, 1989). DNA is a chemically unstable molecule that undergoes hydrolysis and oxidation spontaneously (Höss *et al.*, 1996b; Austin *et al.*, 1997). DNA extracted from ancient tissues is considerably decayed and fragmented (Pääbo, 1989; Pääbo, 1990; Höss *et al.*, 1996b; Yang, 1997). The polymerase chain reaction (PCR™) allows the *in vitro* amplification of trace amounts of DNA (Mullis and Faloona, 1987); even extensively fragmented or damaged DNA (such as ancient DNA) can be amplified (reviewed by Williams, 1995; Yang, 1997; Lindahl, 1997; Austin *et al.*, 1997).

Fragments as large as 91 to 377 bp have been PCR-amplified from animal soft tissue or bone dating between 9,000 and 50,000 years old (Yang, 1997). For example, Höss and Pääbo (1993) PCR-amplified and sequenced a 91 bp fragment of mitochondrial (mtDNA) 16S rDNA from a 25,000 year old *Equus* (horse) bone. Krings *et al.* (1997) amplified and sequenced several fragments of the mtDNA control region ranging in size between ~100-150 bp from a 30,000-100,000 year old Neandertal bone. Poinar *et al.* (1998) sequenced 183-bp chloroplast *rbc*L fragments and 153-bp mitochondrial 12S rDNA fragments from a ground sloth coprolite dating back to 19,875 ^{14}C yr BP. Höss *et al.* (1996a) amplified and sequenced ~200-340-bp mitochondrial 12S and 16S rDNA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fragments from a 13,000 year old ground sloth, and Höss *et al.* (1994) sequenced 93-bp mitochondrial 16S rDNA fragments from wooly mammoth. The ground sloth and mammoth sequences were reproduced in a different laboratory (Taylor, 1996).

Species-specific single nucleotide polymorphisms (SNPs) that distinguish between *P. glauca*, *P. mariana* and *P. rubens* have been identified (Germano and Klein, 1999). These markers would be appropriate for distinguishing *P. glauca*, *P. mariana* and *P. rubens* macrofossils. PCR primers can be designed to amplify short DNA fragments (50-200 bp including the primers) that contain the site of a species-specific SNP. The PCR products can then be purified and directly sequenced. Aligning the ancient DNA sequences with the analogous extant sequences would permit the identification of the macrofossil species. Although other DNA markers, including restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNAs (RAPDs), have been identified to distinguish these species (Bobola *et al.*, 1992a; Perron *et al.*, 1995; Bobola *et al.*, 1996), these types of markers are unsuitable for the identification of fossils because of the limited quantity and degraded nature of ancient DNA.

Project Goals: Species Identification of *Picea* Macrofossils from Browns Pond, Virginia

Picea macrofossils (needles, sterigmata, twigs, buds and seeds) were deposited in sediments of Browns Pond, Virginia, during times between the LGM and the Early-Holocene (Kneller and Peteet, 1993; Kneller and Peteet, 1999). The site is located approximately 50 km west of Staunton, Virginia (38°09' N, 79°37' W), about 300 km south of the LGM Laurentide margin. Browns Pond is one of only approximately twenty sites east of the Mississippi River with a pollen and macrofossil sequence extending up to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

17,000 ^{14}C yr BP. The sediment cores contain an unusually abundant number of well-preserved plant macrofossils (approximately 1-5 *Picea* needle fragments per cm^3) and an AMS radiocarbon date chronology (Kneller and Peteet, 1993; Kneller and Peteet, 1999). The goal of this project is to use PCR-amplification and sequence analysis of DNA fragments containing species-specific SNPs (Germano and Klein, 1999) to identify *Picea* macrofossils from Browns Pond to the species-level. This could be used in conjunction with future *Picea* species histories from other eastern North America sites to reconstruct a regional picture of *Picea* species' northward migration following the retreat of the ice sheet. Overall, the successful taxonomic identification of the Browns Pond *Picea* would serve a twofold purpose: 1) to increase the precision of vegetation-based paleoclimate reconstructions; and 2) to reconstruct the migration patterns of this major forest component in this region of North America.

Materials and Methods

Core Sample Collections

Browns Pond is a small pond, about 20 by 60 m, situated at 620 m elevation approximately 50 km west of Staunton, Virginia (38°09' N, 79°37' W). There are no surface inlets into the pond. The slopes surrounding the pond are forest-covered predominantly with *Quercus alba* and *Q. rubra* (white and red oak). Other trees growing nearby include *Pinus strobus* (white pine), *Acer rubrum* (red maple), *Carya* (hickory), *Betula lenta* (black birch) and *Castanea dentata* (chestnut). Currently, Virginia mid-elevations have a mean annual precipitation of 1,000 mm, mean January temperatures between −6°C and 6°C and mean July temperatures between 13°C and 27°C (Kneller and Peteet, 1993).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Picea micro- and macrofossil samples used in this project were obtained from three different sets of cores collected from Browns Pond at three different times (Table 2). The first set of samples consisted of *Picea* needle macrofossils that were separated from the sediment and sorted by Dr. Kneller. These macrofossils were from a core collected from Browns Pond in 1989 (Kneller and Peteet, 1993). In order to remove sediment from the fossils, Dr. Kneller subjected them to customary chemical treatments including soaking them in a solution of KOH (Kneller and Peteet, 1993). After sorting, the *Picea* needle macrofossils were stored in water and refrigerated until they were sent to us in 1998.

The second set of samples came from a core collected in 1992 (Kneller and Peteet, 1999). This core had been stored at 7°C until a portion of it was brought to UNH in January of 1999. *Picea* macrofossils were sorted from it as described below.

The final set of samples came from six sediment cores (designated 99-1 through 99-6) collected near the northwest perimeter of the Browns Pond in April of 1999 using a 5 cm diameter Livingston piston corer. These cores reached depths between about 250 and 700 cm, with organic material occurring mostly within the top 2-4 m, then changing over to all inorganic sand and clay. The sediments along the length of each core were described by Dr. Kneller. Organic sections of cores suspected to contain *Picea* macrofossils totaled approximately 690 cm. These were divided in half longitudinally, and one half of each core was cut into 1 cm lateral sections after its outer surface had been scraped clean of loose sediment. Each of these sections (~10 cm^3) was placed in a separate plastic bag and labeled. The cores remained at ambient temperature (~20°C) for no longer than one week before they were divided into sections. The sections were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Browns Pond cores used in this project.

Core Name	BR89W1	BR92D	99-1 through 99-6
Date Collected	1989	1992	1999
Location of Sediment Separation	LDEO[a]	Spaulding G07	Spaulding G07 or Rudman G30
Chemical Treatment of Macrofossils	Yes	No	No
Core Storage Conditions	Refrigerated (7°C)	Refrigerated (7°C)	Sections individually wrapped in plastic and frozen (−20°C)
Macrofossil Storage Conditions	Refrigerated (4°C) in water	Frozen (−20°C) individually	Frozen (−20°C) individually

[a] Lamont Doherty Earth Observatory, Columbia University, Palisades, NY

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

refrigerated (4°C) for a period no longer than one month, and then stored frozen at (-20°C).

Precautions for Fossil Manipulation

Laboratory surfaces and equipment were periodically wiped down with either RNase Away (Molecular BioProducts Inc., San Diego, CA) or 8% bleach (both solutions chemically destroy DNA). Plastic tube racks were soaked overnight in 10% RNase Away and rinsed with water between uses. New micropipets were purchased, and only aerosol-resistant filter pipet tips were used. An acid-treated glass distillation apparatus provided the distilled water. Water was autoclaved before use in solutions or for PCR. All glassware, microcentrifuge tubes, spatulas, mini-pestles, etc. were autoclaved (121°C for at least 20 min) between uses. All the supplies and chemicals used were purchased new. To avoid contamination from amplified *Picea* DNAs abundant in the main research laboratory of Dr. Anita Klein, all of the equipment (water bath, microcentrifuges, balance, etc.) was borrowed from either the Biochemistry or Microbiology teaching laboratories. Refrigerators and freezers used were located in the Biochemistry or Microbiology teaching labs or the basement of Rudman Hall.

The 1999 cores were sectioned and prepared for storage on the bench tops of Spaulding Life Sciences room G07. This laboratory is located in the corner of the ground floor of the building, adjacent to the Biochemistry Department teaching lab prep room. The Biochemistry teaching labs do not routinely do experiments with plants or with plant DNA (note: if any plant materials were used, they would have involved plants distantly related to conifers). Macrofossils were separated from the sediment over a sink and sorted under a microscope on one of the bench tops in Spaulding G07. Attempts to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

extract and PCR-amplify DNA from the macrofossils were also carried out in this room. Macrofossil DNA extractions and PCR preparations were carried out in separate hooded enclosures. These enclosures were illuminated with ultra-violet (UV) light for 10-15 minutes prior to use.

All work with fossil pollen, including separating it from core sediment and attempts to PCR-amplify pollen DNA, was carried out in the Microbiology teaching lab room G30 on the ground floor of Rudman Hall. This work was conducted in laminar-flow hoods equipped with UV lights with the exception of sorting individual pollen grains from sediment suspensions. Sorting was carried out under a microscope on one of the bench tops in Rudman G30. The only equipment from the main research lab that was used in for these experiments was a thermal cycler used for pollen DNA amplifications in Rudman G30. Every surface of the machine was decontaminated with RNase Away and 8% bleach; the heating block was soaked with each solution and thoroughly rinsed with distilled water.

Separating Macrofossils and Pollen from Core Sections

Macrofossils of core BR92D and the 1999 cores were separated from the sediment by rinsing one 1-cm section at a time with distilled water through sterile 500 μm and 125 μm brass sieves. Brass sieves were rinsed well with deionized water followed by distilled water between each use, and autoclaved each day (between every 5-10 uses). *Picea* macrofossils were sorted from other macrofossils under a dissecting microscope, and then transferred to sterile 1.5-ml microcentrifuge tubes. Larger macrofossils, such as whole needles or almost whole needles, seeds, seed wings, twigs or buds, were transferred into individual 1.5-ml tubes. Smaller macrofossils, such as small

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

needle fragments or sterigmata, were pooled together into single tubes. Each tube contained about 1-15 mg of tissue (wet weight); the contents of each tube served as one sample from which DNA was extracted. Large seeds, such as *Rubus* (raspberry) or *Abies* (fir), were retained for future AMS radiocarbon dating from each core section and stored in individual tubes refrigerated (4°C) or frozen (−20°C).

Pollen was separated from the sediment of core 99-1 by passing 1-2 cm^3 pieces of a section through 1) a 500 μm brass sieve, 2) a 120 μm nylon screen and 3) a 7 μm nylon screen using copious amounts of distilled water. Sometimes the 7 μm nylon screen was placed in a Nalgene filter sterilizer unit (with the 0.2 μm filter removed) and attached to a vacuum in order to speed the process. The apparatus was sterilized between separating each core section, the sieve by autoclaving, and the nylon screens by soaking in 8% bleach and exposing them to UV light. The pollen suspension was collected from the 7 μm and 120 μm nylon screens with a pipet and transferred to separate microcentrifuge tubes. Before any sediment was separated, distilled water was passed through the entire apparatus and collected in order to serve as a blank negative control. Pollen grains were identified as *Picea* under a 40X magnification dissecting microscope according to morphological descriptions provided by Drs. Margaret Kneller and Dorothy Peteet. Individual grains were removed from underneath the microscope with a P-2 Gilson Pipetman and transferred to a microcentrifuge tube.

DNA Extraction from Macrofossils

DNA was extracted from macrofossil tissue using two types of extractions. Extractions using a traditional CTAB extraction and isopropanol precipitation (Doyle and Doyle, 1987) were performed as described for small scale preparations in Germano and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Klein (1999). The macrofossil samples used for these extractions were from cores BR89W1 and BR92D. A phenol extraction step was incorporated into some of the extractions just before the chloroform extraction step. A Proteinase K step was also added to the end of some extractions. The DNA pellet was resuspended in a Proteinase K buffer (0.01 M Tris, pH 8, 5 mM EDTA, 0.5% SDS), Proteinase K (50 µg/ml; Promega) was added, and the solution was incubated at 37°C for one hour. The solution was re-extracted with phenol and chloroform, brought to 1.4 M NaCl and re-precipitated.

A silica extraction method for fossil DNA extraction was modified from: Poinar *et al.* (1998), Pääbo (1989), Pääbo (1988), Höss and Pääbo (1993), Boom *et al.* (1990), and Hendrik Poinar (pers. comm). Macrofossil tissue (1-15 mg) was ground to a fine powder in a 1.5-ml microcentrifuge tube resting in liquid nitrogen using a conical mini-pestle. The powdered tissue was suspended in 200 µl Extraction Buffer (10 mM Tris [pH 8.0], 2 mM EDTA, 10 mM NaCl) and Proteinase K was added to a final concentration of 200 µg/ml. The suspensions were incubated at 37°C with constant agitation for 48 hours. *N*-phenacylthiazolium bromide (PTB) was added to a final concentration of 10 mM. PTB was synthesized according to Vasan *et al.* (1996), stored dry under a vacuum as crystals and dissolved in a solution of 10 mM $NaPO_4$ buffer (pH 7.4) immediately before use. The solution was extracted twice with an equal volume of saturated phenol (pH ~7) and once with an equal volume of chloroform:isoamyl alcohol (24:1). Solutions were mixed for 5-10 minutes and centrifuged at 10-13,000 × g for 3-5 minutes. The supernatant was removed to a Microcon-30 sample reservoir (Amicon Inc., Danvers, MA) and concentrated by centrifugation to 20-30 µl. L6 Buffer (10M GuSCN, 0.1M Tris [pH 6.4], 3.6mM EDTA, 21 mg/ml Triton X-100) and Silica Suspension (Boom *et al.*, 1990) were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

added to the concentrate (100 µl and 4 µl, respectively, per 20 µl concentrate), and the solutions were rotated for several hours to overnight at room temperature. The silica was pelleted by centrifuging 10-13,000 × g for 8 seconds and the supernatant was discarded. The pellet was washed once in 150 µl L2 Buffer (10M GuSCN, 0.1M Tris [pH 6.4]) and twice in 150 µl ice-cold Bio 101 New Wash (Bio 101). After discarding the supernatant, the pellet was allowed to dry at 56°C for 2-3 minutes. DNA was eluted from the pellet with 20 µl TE buffer (56°C) for 1-2 minutes at 56°C. Note: the L6 and L2 Buffers contained a small amount of Silica Suspension in order to bind any contaminating DNAs, therefore these buffers were centrifuged before each use. Reagent blanks, where no macrofossil tissue was added, were carried out with each batch of extractions.

DNA was "extracted" from pollen grains by crushing grains suspended in distilled water with a pipet tip or glass rod. Either the crushed grains in suspension or the supernatant after pelleting the grains were added to PCR amplification reactions.

Southern Blot and Hybridization Analysis

In an attempt to detect ancient *Picea* DNA, macrofossil extracts (CTAB extractions, core BR92D) were subjected to electrophoresis through an agarose gel, and transferred to a nylon membrane by Southern blot according to Sambrook *et al.* (1989). Whole genomic DNAs from extant *Picea, Escherichia coli* (bacteria), *Trichoderma* (a common soil fungus) and *Phaeocryptopus* (a conifer endophytic fungus) were also loaded into different lanes of the gel as well as λ *Hind* III and ΦX *Hae* III DNA size markers. Radioactive probes were synthesized from *P. rubens* ITS1 PCR amplicon (Germano and Klein, 1999) using Prime-It® Random Primer Labeling Kit (Stratagene, La Jolla, CA). Hybridization and wash buffers were prepared according to Sambrook *et al.*

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(1989). Pre-hybridization was carried out in 5 ml pre-hybridization solution (50% deionized formamide, 4X SSPE buffer, 0.5% BLOTTO, 1% SDS, 100μg/ml sheared, denatured, quenched herring sperm DNA) at 42°C for 2 hours. Hybridization was carried out in 5 ml reactions (50% deionized formamide, 4X SSPE buffer, 9.3% dextran sulfate, 0.5% BLOTTO, 1% SDS, 100μg/ml sheared, denatured, quenched herring sperm DNA, probe) at 42°C for 16 hours with constant rolling. The membrane was washed twice with 2X SSC buffer, 0.1% SDS for 15 minutes, once at room temperature and once at 65°C. KODAK XOMAT Blue X-Ray film was exposed for 10 days.

PCR Amplification

Attempts to amplify macrofossil DNA were carried out using the primers R/BvsW SSCP-1 and R/BvsW SSCP-1A (Germano and Klein, 1999). Each 25-μl reaction contained 1-4 μl DNA extract, 1X Taq Buffer (Promega), 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.4 μM each primer, and 1-12.5 units Taq Polymerase (Promega). Some reactions contained an amount of Taq Antibody (Gibco BRL) equal to the amount of Taq. Bovine Serum Albumin (Promega) was also added to some reactions to a final concentration of 1 mg/ml. Thermal cycling conditions were as follows: initial denaturation 94°C for 3 minutes, 35-50 cycles of 94°C for 30 seconds, 47-63°C for 30-60 seconds, 72°C for 0-30 seconds, final extension step 72°C for 10 minutes.

Primers were designed to amplify fossil ITS DNA as follows: ITS1-5.8S-1 and ITS1-5.8S-1A (GTC TTG TGG GGT GGG AGG GTT GTT G and AGA GCC GAG ATA TCC GTT GCC GAG AG) and 5.8S-ITS2-1 and 5.8S-ITS2-1A (CAG AAT CCC GTG AAT CAT CGA GTT TTT G and CTT GCA GGG AGC GCG TGT ATG TAG G). Each 25-μl PCR reaction contained 10-100 crushed pollen grains or 14 μl crushed pollen

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

supernatent, 1X Taq Buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.4 μM each primer, 6.25 units Taq Polymerase and 6.25 units Taq Antibody. Bovine Serum Albumin was also added to some reactions to a final concentration of 1 mg/ml. Thermal cycling conditions were as follows: initial denaturation 94°C for 4 minutes, 39 cycles of 94°C for 20 seconds, 55°C for 1 minute, 72°C for 1 minute, final extension step 72°C for 10 minutes.

Results and Discussion

Core Samples: Effects of Handling and Time since Collection from the Ground

Core BR89W1. Macrofossils from core BR89W1 were unsuitable for fossil DNA analysis for several reasons. Firstly, they were subjected to a number of chemical treatments including immersion in a basic solution (KOH) to completely remove the sediment from the macrofossils. These treatments are customary for analysis of pollen and macrofossil core records because they reveal details of the fossils' anatomy allowing easier identification. However, the chemical treatments likely compromised the integrity of any DNA remaining in the BR89W1 fossils. Furthermore, after separating them from the sediment, the macrofossils were stored in tap water and refrigerated for at least five years. DNA is most stable in a buffered, bacteriostatic solution that is stored at –80°C. Any DNA remaining in the macrofossils after the chemical treatments would have been hydrolyzed, or degraded by bacteria or fungus that could have easily grown under these storage conditions. In fact, high molecular weight DNAs were detected by agarose gel in the CTAB extracts from these macrofossils (data not shown). Bacterial and fungal DNAs are common contaminants of ancient DNA extracts (Höss *et al.*, 1996b; Austin *et al.*, 1997), and are thought to be recent in origin and the source of high molecular weight

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DNAs. The presence of high molecular weight DNAs in the BR89W1 extracts suggested that they were indeed contaminated with contemporary microorganisms.

Core BR92D. The macrofossils from core BR92D also yielded high-molecular weight DNA (CTAB extraction; see Figure 2). Following a Southern blot, the *Picea* ITS1 radiolabeled probe hybridized to the contemporary *Picea* DNA on the blot, and it did not cross-react with any of the microbial or size marker DNAs. The high molecular weight DNA seen in the macrofossil lane on the agarose gel did not hybridize to the *Picea* probe, supporting the hypothesis that it is microbial in origin (Figure 3). The growth of bacteria and/or fungi is not surprising considering the length of time that core BR92D had been stored refrigerated (approximately seven years).

Interestingly, the macrofossil lane of the Southern blot revealed a very faint smear corresponding to DNAs ranging from ~250 to 650 bp (Figure 3). This result suggests that authentic ancient *Picea* DNA was present in the macrofossil extract. The size range inferred for the macrofossil DNAs may be skewed due to DNA-DNA cross-links between ancient DNA molecules (Pääbo, 1990). Even if the macrofossil DNAs are half as long as 250-650 bp, they are within the range that can typically be PCR-amplified (Pääbo, 1989).

The 1999 Cores. The macrofossils from the 1999 cores did not yield a high molecular weight DNA band visible on an agarose gel (data not shown). This suggests that the contemporary microorganisms hypothesized to be the source of the high molecular weight DNA in BR89W1 and BR92D macrofossils developed throughout the cores/fossils *after* the cores were extracted from the ground and were not inherent in the sediments below Browns Pond. The proper handling and storage of the 1999 cores prevented such growth from occurring. This also confirms the expectation that few

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

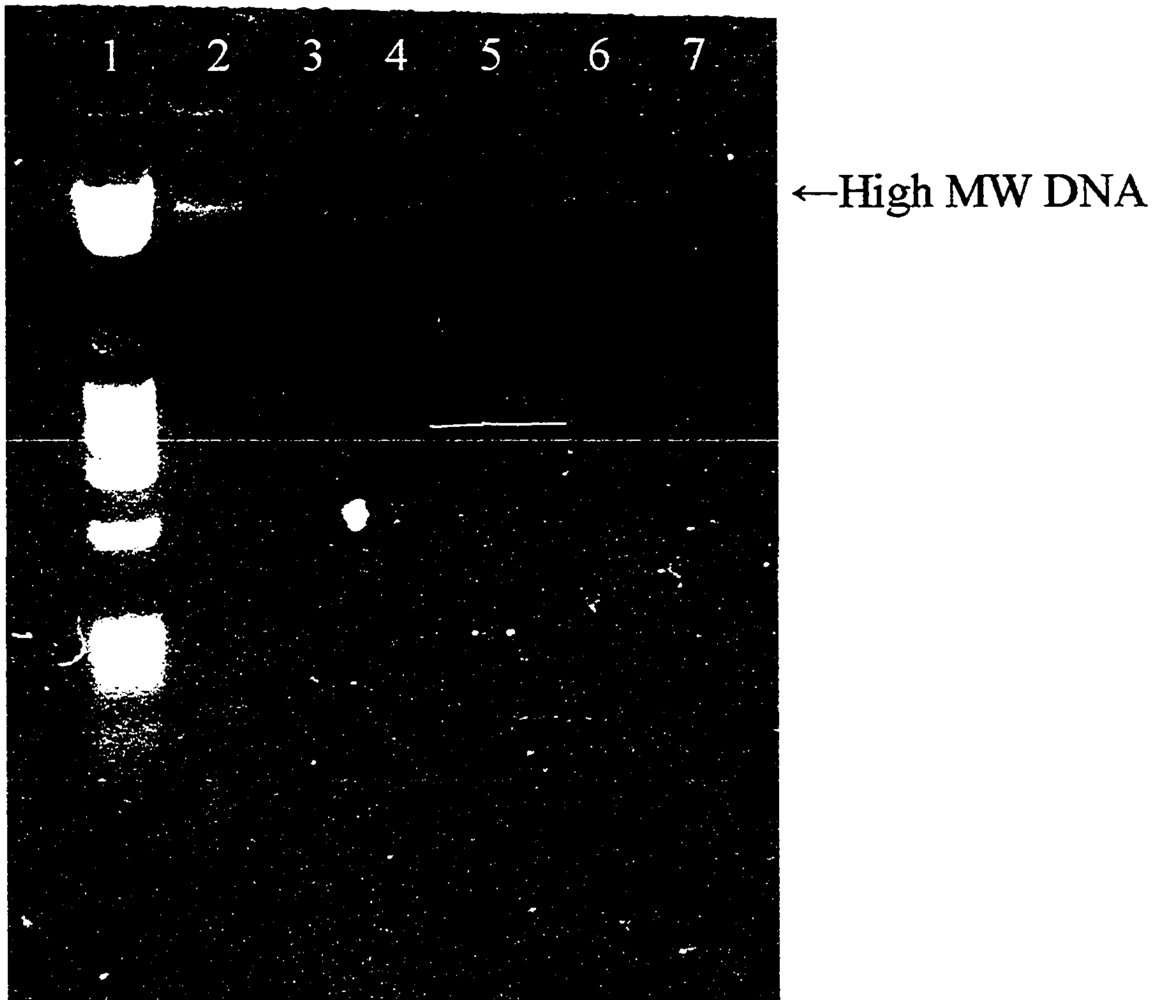


Figure 2. Agarose gel showing DNA extracts (CTAB method) from BR92D *Picea* macrofossils. Lane 1: Lambda *Hind*III and PhiX *Hae*III DNA size markers, Lanes 2-7: macrofossil extracts (note: high molecular weight DNA visible in some extracts).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 3. Hybridization of Southern blot with *Picea*-specific probe. Lanes 1 and 2: *Picea rubens* contemporary DNA, Lane 3: DNA size markers, Lane 4: *Trichoderma*, Lane 5: *E. coli*, Lane 6: *Picea* macrofossil, Lane 7: *Phaeocryptopus*. Bracket indicates presence of putative ancient DNA.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

microorganisms grow in the anoxic environment of the sediment layers below lakes and ponds, a reason why the fossils below Browns Pond are so well preserved structurally.

Extraction of one of the macrofossils from core 99-1 (silica method), a small twig with nine pulvini and a bud, yielded a very faint DNA smear visible on the agarose gel used for a second Southern blot (data not shown). This smear of DNA, from about 100-1000 bp, contains larger DNAs than those detected in the BR92D macrofossil extract by hybridization. This suggests that the fossil DNA in core BR92D degraded significantly *after* the core was extracted from the ground, possibly due to bacterial/fungal growth or oxidation. These results demonstrate the importance of properly handling fossils once they are removed from their original place of deposit, in order to preserve the integrity of their DNA. The putative fossil DNA visualized on the agarose gel, however, was not detected by the *Picea* radiolabeled probe (data not shown). This could be because 1) the probe had a 2-fold lower specific activity than the probe used for the BR92D blot, 2) there was a lot of residual background left on the blot after the washes, or 3) the film was exposed to the blot for less than half the amount of time as the BR92D blot.

The Southern blot and hybridization experiment to detect authentically ancient *Picea* DNA from the macrofossils of the 1999 cores warrants repetition, with several changes to the protocol. A large-scale DNA extraction from numerous *Picea* macrofossils should be carried out in lieu of extracting DNA from individual fossils, and the entire contents of the extract should be loaded on the gel. Furthermore, the DNAs in the gel should not be depurinated with hydrochloric acid (HCl) prior to blotting on the nylon membrane since ancient DNA is already extensively damaged and depurinated (Lindahl, 1993), and further depurination may completely abolish the ability of the target

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

DNA to anneal to a probe. Also, the probe should have a specific activity of at least 4 X 10^9 dpm/μg. Finally, the radioactivity should be detected over time with a phosphor-imager. This would eliminate the guessing involved in determining how much time the X-ray film should be exposed, and there is no risk of over-exposing X-ray film.

Problems Encountered with DNA Extraction of Spruce Macrofossils

CTAB Extraction. The CTAB extraction method was used to extract DNA from BR89W1 and BR92D macrofossils. Following the isopropanol precipitation and ethanol washes, the precipitant pellet was brown in color. When subjected to electrophoresis through an agarose gel, this brown substance traveled towards the anode at a slightly greater speed than the bromophenol blue. Interestingly, when visualized under UV light, it appeared as though all of the ethidium bromide had been stripped away from the lane containing brown extracts, or as if the brown molecules prevented the ethidium bromide from traveling towards the cathode.

Ancient DNA extracts are almost invariably brown in color, commonly due to the presence of Maillard products (Pääbo, 1989; Pääbo, 1990; Poinar *et al.*, 1998). The Maillard reaction involves the condensation of reducing sugar carbonyl groups with primary amines; advanced stages of the reaction can cross-link DNA and proteins over long periods of time (Poinar *et al.*, 1998). Humic acid, a form of Maillard products (Richard Blakemore, pers. comm.), is a common component of organic soil and it is a common contaminant of DNA extracts from soil microorganisms (Torsvik, 1995). Humic acid is brown in color, and it has an electrophoretic mobility that is slightly faster than bromophenol blue through agarose gel (Richard Blakemore, pers. comm.). It is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

likely that a component of the brown contaminants of the macrofossil CTAB extracts was a form of humic acid that co-precipitated with the DNA.

Silica Extraction. The silica based extraction method purified much of the brown contaminants away from the DNA extract. After each wash of the silica pellet, the brown color became lighter; following the elution with TE buffer, any remaining brown substance stayed on the silica and did not elute into the final product. This purification may have been aided by the addition of PTB during the extraction. PTB has been shown to break down Maillard products (Vasan *et al.*, 1996). It is hypothesized to help chemically release DNA molecules that may be covalently bound in Maillard products ancient tissue extracts (Poinar *et al.*, 1998). This purification is also important because humic acid is known to inhibit PCR (Richard Blakemore, pers. comm.).

In addition to yielding a better purified product, silica based extractions are generally preferred over precipitation methods for fossil DNA because DNAs that are small in size (such as ancient DNAs) do not always precipitate on their own. Many times a "seed" such as tRNA is added to assist the small DNA fragments in precipitating. The CTAB method may have worked for the BR89W1 and BR92D macrofossils because the contemporary high molecular weight DNAs that were present may have acted like a "seed" for the precipitation step. For these reasons, it was concluded that the silica extraction method was better suited for extracting DNA from Browns Pond macrofossils.

PCR Amplification

All attempts to PCR-amplify authentically ancient DNA from macrofossils and from fossil pollen failed. The primary impediments included inhibition of PCR, non-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

specific amplification, contamination, lack of template DNA (regarding the pollen) and primer design.

Inhibition. Inhibition is a common problem when using to PCR to amplify ancient DNAs because many different substances that co-purify with the fossil DNA in the extraction can interfere with the PCR reaction. There are two ways to determine if inhibition is the reason behind failed PCR. Firstly, one can conduct an experiment in which quality template (known to be amplifiable) is added to a reaction containing fossil extract. If amplification does not occur, yet a positive control reaction does (demonstrating the functionality of all the PCR reagents), then it may be concluded that some component of the fossil extract inhibited the amplification reaction. Second, the absence of primer-dimer amplification products in PCRs containing fossil extracts is also an indicator of inhibition, especially if such products are present in negative control reactions. If inhibition is determined to be the problem, the next step is to determine the nature of the inhibition; i.e. is the inhibitor affecting the polymerase, or is it compromising some other component of the reaction?

In many of the attempts to amplify DNA from the macrofossil CTAB extracts, inhibition was apparent by both indicators discussed above. Since positive controls were positive, and negative controls exhibited primer-dimer amplicon, it was concluded that the inhibitor was derived from the fossil extracts. Working under the hypothesis that the inhibitor was a form of DNase (digesting the primers, hence no primer-dimer), the CTAB extraction method was conducted with two additional steps: a phenol extraction step, and a Proteinase K digestion step. Neither of these prevented PCR inhibition, therefore it was concluded that the inhibitor was not DNase or any other protein-based molecule.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Switching to the silica extraction method greatly reduced the amount of PCR inhibition. This can probably be attributed to the increased purity of the DNA extracts with using silica and to the fact that the method employs both Proteinase K and phenol. Since Maillard products, including humic acid, inhibit the Taq DNA polymerase (Poinar *et al.*, 1998), two additional strategies were employed in assembling the amplification reactions. First, 6-12 fold more polymerase (than is used when amplifying fresh DNA) was added to the reactions to combat the effects of the inhibitor. This strategy worked to overcome inhibition for Pääbo (1990). Second, bovine serum albumin (BSA), was added to stabilize the polymerase as has been employed in multiple ancient DNA studies (Pääbo *et al.*, 1988; Pääbo, 1990; Taylor, 1996). BSA is often used to stabilize other molecular biology enzymes such as DNA ligase and restriction endonucleases. Theoretically, if the inhibitor was chemically reacting with and hence disabling the polymerase, the BSA would act as a competitor with which the inhibitor could react. These strategies, both singly and together, in addition to switching to the silica extraction method, significantly overcame many of the inhibition problems.

PCR Amplification Attempts from Fossil Pollen. Attempts were also made to amplify DNA from Browns Pond pollen. The sporopollenin wall of pollen grains is durable, chemically inert, and even resistant to treatments with strong acid; therefore pollen grains can remain intact for thousands of years. Theoretically, the pollen wall can act to protect and preserve the pollen DNA over long periods of time. PCR amplification has been demonstrated from DNA template from single pollen grains (Petersen *et al.*, 1999). Moreover, putatively ancient DNA was detected with DAPI staining and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

amplified by PCR from 150,000 year old *Abies* pollen from peat deposits (Suyama *et al.*, 1996).

Attempts to PCR-amplify DNA from Browns Pond pollen yielded no positive results. Under 40X magnification, the *Picea* pollen grains appeared compressed, and many were broken open. The poor condition of the pollen was not due to the vacuum apparatus used during sifting of the sediment. Separating the intact pollen grains from a suspension of the sediment was tedious and time consuming. The low number of well-preserved pollen grains and the amount of time it took to isolate them from the sediment limited the chances of successful amplification of pollen DNA.

Inhibition of PCR was also a problem when attempting to amplify DNA from fossil pollen. Although no experiments were carried out where quality template was added to reactions containing pollen, the accumulation of primer-dimer amplicon was affected by the pollen extract. An experiment was carried out in which 1) smashed pollen grains were added directly to the PCR or 2) just the supernatant (distilled water atop the pollen grains after smashing them and pelleting them by centrifugation) was added. The reactions containing pollen grains (approximately 50-100 grains per reaction) produced no primer-dimer amplicons, whereas the reactions containing supernatant did yield primer-dimer amplicon (Figure 4). This led me to conclude that either the substance that makes up the outer layer of pollen grains (sporopollenin), or something associated with the pollen from the sediment, inhibits PCR.

Non-Specific Amplification. Non-specific amplification can be problematic in conducting PCR of ancient DNAs. Ancient DNAs are extensively damaged. Not only are the phosphodiester bonds cleaved by hydrolysis, effectively shortening the DNA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

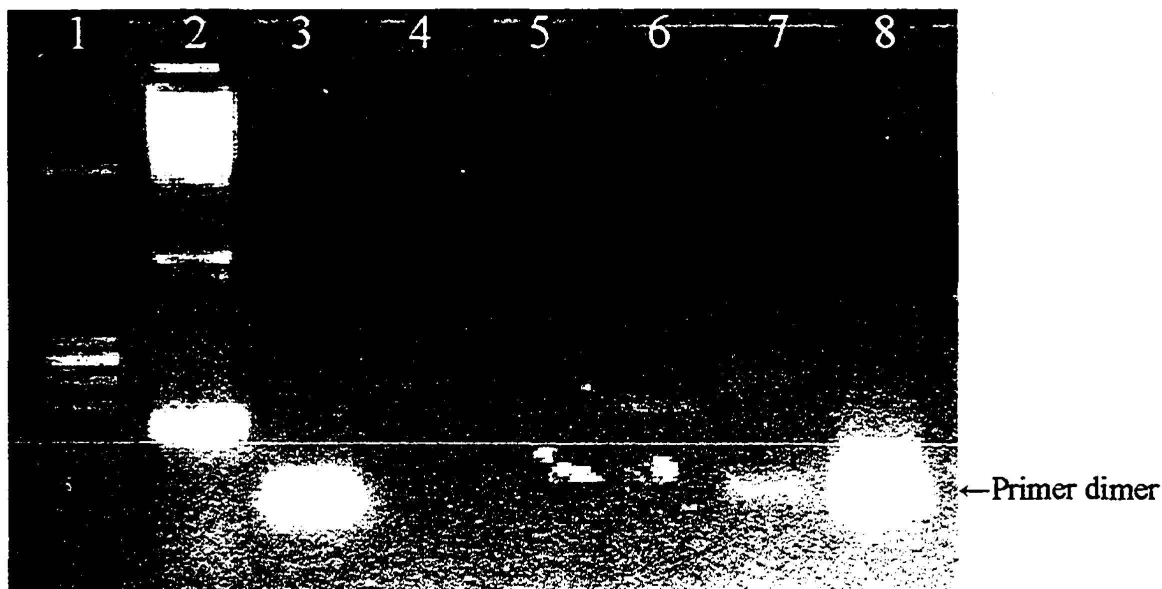


Figure 4. Agarose gel showing PCR products using primers ITS1-5.8S-1 and ITS1-5.8S-1A. Lane 1: PhiX HaeIII DNA size marker, Lane 2 *Picea* positive control, Lane 3: supernatent collected after crushing pollen, Lanes 4 and 5: crushed pollen grains, Lanes 6, 7 and 8: negative control of water left open on bench top.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

fragments, but the nucleobases are compromised as well through depurination and hydrolytic deamination (Lindahl, 1993). This damage of the nucleobases directly affects the effectiveness with which the PCR primers anneal with the template DNA. Therefore, annealing temperatures are usually set much lower than usual so that primers can anneal to the damaged template. Even if primers are designed to be specific to the species of fossil organism, such low annealing temperatures would enable them to anneal to contemporary DNA from other contaminating organisms.

Many PCR reactions resulted in the production of multiple amplicons between 100 and 500 bp in length. This problem was exacerbated by carrying out more than 40 cycles of PCR. Contamination is an obvious source of non-specific amplification. However experiments with and without BSA showed that most of these random amplification products were due to the addition of BSA (Figure 5). According to their tech-support, Promega does not carry out quality control measures for the presence of DNA in their BSA products, therefore residual bovine DNA was likely present in the BSA preparations. This hypothesis could be tested directly by sequencing some of the non-specific PCR products and searching for homologous sequences by a BLAST search of GenBank. Since the addition of excess Taq polymerase overcame much of the inhibition caused by the fossil extracts, the addition of BSA is probably not necessary.

Contamination. The risk of contamination of ancient DNA with contemporary DNA is problematic because ancient DNA molecules are present in very low numbers, and PCR is so sensitive, it can yield amplified products from just a few starting template molecules (Pääbo, 1990; Austin *et al.*, 1997). Contaminating modern DNA usually out-competes ancient DNA during PCR because ancient DNA is significantly damaged, thus

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

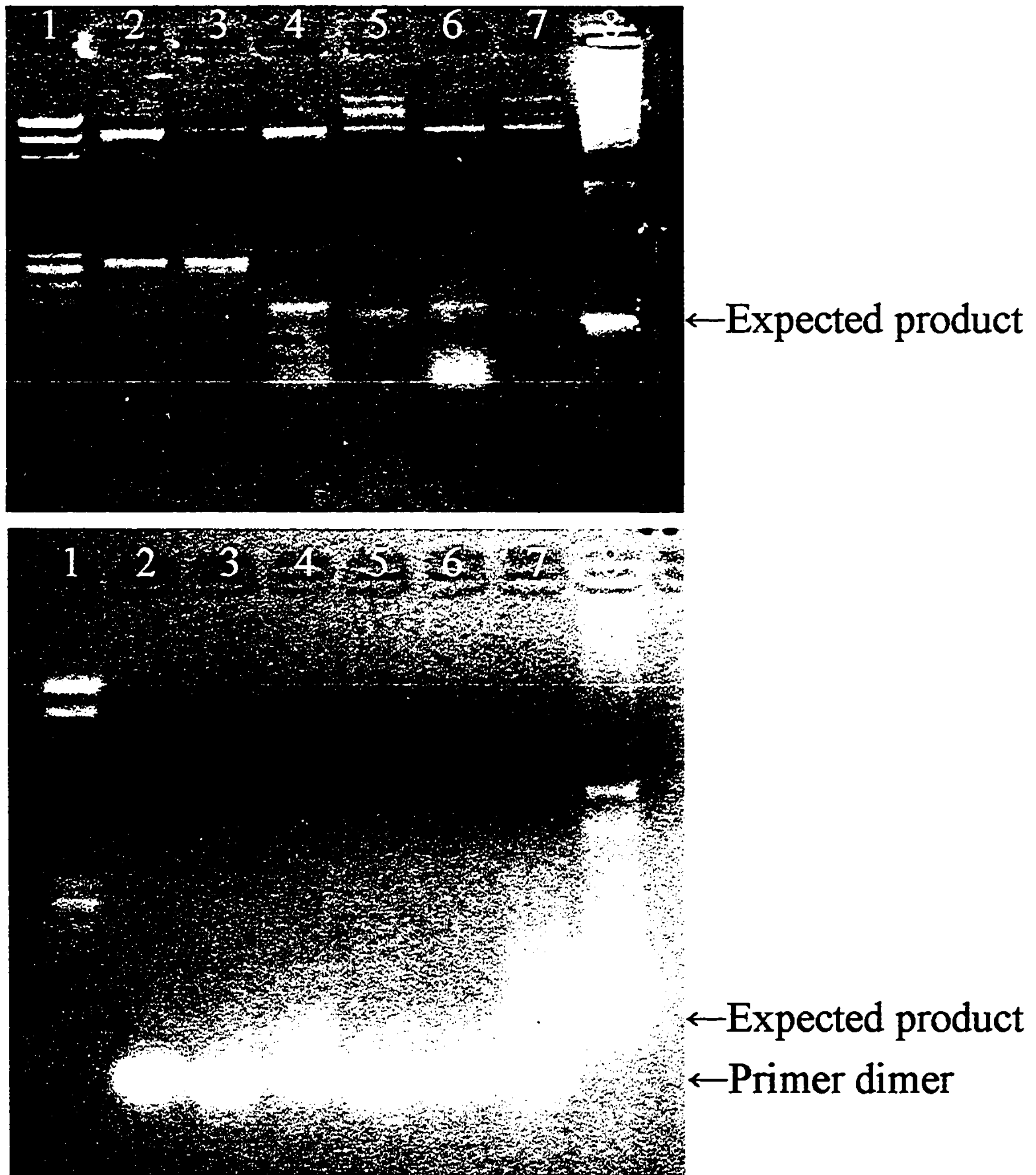


Figure 5. Agarose gel showing PCR products using primers ITS1-5.8S-1 and ITS1-5.8S-1A. Top panel: reactions contained 1 mg/ml BSA, bottom panel, reactions contained no BSA. Lane 1: PhiX HaeIII DNA size marker, Lanes 2-7: negative controls (no DNA template added), Lane 8: *Picea* positive control (DNA template from contemporary *Picea* sample).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

modern DNA makes a better template (Höss *et al.*, 1996b). Multiple PCR reactions containing macrofossil extracts yielded amplicons of the expected size. Some of these amplicons, of both macrofossil extracts and of negative control reactions, were directly sequenced. The sequences were compared to homologous sequences of a variety of conifers including *Picea sitchensis* (Sitka spruce), *Picea engelmannii* (Engelman spruce), *Picea pungens* (Colorado blue spruce), *Picea breweriana* (Brewers spruce), *Picea chihuahuana* (Chihuahua spruce), *Picea mexicana* (Mexican spruce), *Abies balsamea* (Balsam fir), *Larix decidua* (European larch), *Pinus contorta* (shore pine), *Pinus sylvestris* (Scotch pine), *Pinus thunbergii* (Japanese black pine), *Pinus banksiana* (jack pine) and *Pinus strobus* (eastern white pine). The PCR sequences matched those of eastern white pine. It was concluded that these amplifications were of contemporary contaminating white pine DNA because the PCR products were also found in the negative controls. It is likely that the source of the contamination was white pine pollen, which was particularly abundant during 1999 when many of these experiments were performed.

Multiple negative controls were carried out throughout all fossil manipulation procedures. These included 1) a blank run of distilled water through the sifting apparatuses (the brass sieves and nylon membranes), 2) a reagent blank carried out along side of each batch of macrofossil extractions to test for contamination of extraction reagents and 3) a template blank along side of each batch of PCR preparations to test for contamination of PCR reagents. Many times, the PCRs using these negative controls as templates were negative for amplification, yet the fossil extract-containing reactions were positive for the expected-size product (later identified as white pine DNA). Therefore,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the contamination must have occurred at some point in time other than during the sifting, extraction or PCR preparations. It is important to note here that the macrofossil extractions and preparation of PCR reagents were carried out under hooded enclosures. In the case of pollen sifting, the pollen was separated from the sediment under a laminar flow hood. This led to the suspicion that the contamination occurred during the sorting and picking of macrofossils and pollen because during this step, 1) the samples are open to the environment while under the microscope, 2) the sorting under the microscope takes a relatively long time, and 3) the sorting and picking are carried out on the bench top because it was not possible to use a microscope in the laminar flow hood.

Issues with contamination were more prevalent when working in Spaulding G07. This laboratory was chosen because it was in a separate building as Dr. Klein's laboratory. The rationale was to get as far away from the ubiquitous *Picea* genomic and amplified DNAs present in the Klein lab. Spaulding was probably a poor choice for a "clean" lab however since it does not have a filtered air handling system. Many of the rooms in Spaulding, including G07, have windows in them. During the hot summer months these windows are opened to let in fresh air. Large amounts of wind-borne pollen, including white pine pollen, enter the building this way.

Working in the Microbiology lab G30 solved many of the contamination problems because the lab is effectively closed to the outside environment, and because of the availability of laminar flow hoods. There were still some issues with contamination however when working in the Microbiology lab. An experiment was conducted in which a sterile Petri dish containing distilled water, like the ones used for sorting and picking pollen grains, was left on the bench top next to the microscope for the length of time it

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

took to sift and pick through one batch of pollen grains. This water was then used as template in a negative control PCR reaction. Amplicons of the expected size range were produced in these negative controls indicating that the water in the open Petri dish had become contaminated by simply standing open on the bench top, even after the bench was disinfected with bleach (Figure 4). The source of the contamination was likely air-borne pollen.

Primer Design. When attempting to amplify the macrofossil DNA, the primers used (R/BvsW SSCP-1 and R/BvsW SSCP-1A) were chosen based on 1) the location of their target in the multi-copy chloroplast DNA and 2) the short length of their product (119 bp). When attempting to amplify the pollen DNA, different primers (ITS1-5.8S-1 and ITS1-5.8S-1A / 5.8S-ITS2-1 and 5.8S-ITS2-1A; see Materials and Methods for primer sequences) were used. Their target was in also in a multi-copy region: the ITS region of the rDNA repeat, which is tandemly repeated on multiple chromosomes in thousands to millions of copies in the *Picea* cell nucleus (Bobola *et al.*, 1992b). The amplicon (product) lengths were also short: 147 and 108 bp, respectively. These primers, however, were designed with additional conditions in mind. They were significantly longer (25-28 nucleotides) and therefore had much higher melting temperatures (63-65°C). The rationale was the longer the primer, the higher the melting temperature, the "stickier" the primer, the better the chances of it annealing to damaged DNA templates. A high GC content was also preferred; G and C nucleotides form a complimentary base pair by three hydrogen bonds, whereas A and T only anneal by two hydrogen bonds. Additionally, the GA content was considered because purines are damaged in DNA over time much more rapidly than pyrimidines (Lindahl, 1993). The newly designed primers

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

had %GC contents between 43 and 60% and %GA contents between 46 and 60%. Since these primers are "stickier" they have a greater capability of annealing non-specifically to contaminating contemporary DNAs. Therefore, they were designed to anneal to highly variable regions of the ITS region. Sequence alignment of this region in spruce with species of pine and hemlock indicate that there are multiple mismatches between the primer and the non-spruce conifer DNAs (data not shown). At PCR annealing temperatures near the melting temperatures of these primers, they should not amplify contaminating conifer DNAs. However, when positive controls were performed using contemporary *Picea* DNA with an annealing temperature of 55°C, longer than expected amplicons were produced in addition to the fragment of expected size. This was likely caused by non-specific annealing to other regions of the *Picea* genome because of the low annealing temperature. These primers most likely had a greater chance of successfully amplifying ancient DNA, however, their "stickiness" increases the risk of amplifying contemporary contaminating DNAs.

Future Directions

Hindsight is always 20/20. If this project were to be attempted again, I would carry it out the following way. All the work would be done in the Microbiology teaching laboratories. These rooms are sufficiently far enough away from the main research lab and its amplified *Picea* DNA. All equipment, supplies and reagents would be borrowed or brand new. Experiments would have designated rooms, i.e. one room for sediment separation and extraction, one room for amplification preparation, one room for thermal cycling and one room for detecting PCR products. Each room would have its own designated equipment and supplies that would not travel from room to room. All

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

experiments would be carried out in laminar flow hoods equipped with UV light. A special hooded enclosure would be necessary for microscope work. *Picea* macrofossils would be pooled together and large-scale extractions (from 100-1000 mg tissue) would be carried out using the silica extraction method. All of the negative controls discussed above would be used. If DNA cannot be detected with agarose gel electrophoresis, then Southern blot and hybridization analyses would be done to show that authentic ancient DNA is present. PCR would be carried out with the ITS1-5.8S-1 / ITS1-5.8S-1A and 5.8S-ITS2-1 / 5.8S-ITS2-1A primers. Ten times the normal amount of Taq polymerase would be used in each reaction, and a higher quality polymerase with proof-reading function would be used (e.g. AmpliTaq Gold, Applied Biosystems Inc.). BSA would not be used in the reactions unless a source could be found where quality controls for the presence of DNA are performed. In addition, experiments would be done to check for the presence of contaminating DNA in the BSA. PCR products would be cloned in lieu being directly sequenced since multiple species of spruce could have been combined when pooled together for the extraction process. Any positive results would be repeated by a collaborator in another laboratory in a remote location.

Specific Applications of *Picea* Species Identifications Using DNA Technology. Upon successful amplification of macrofossil DNA, the following hypotheses would be addressed. Estimates of climate change during the LGM have been made for the unglaciated region of eastern North America based on pollen records (Watts, 1980b; Whitehead, 1981; Prentice *et al.*, 1991; Webb, III *et al.*, 1993). Interestingly, these estimates are significantly colder and drier than LGM climates simulated using GCMs (Hansen *et al.*, 1984; Rind and Peteet, 1985; Broccoli and Manabe, 1987; Kutzbach,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1993; Kutzbach and Ruddiman, 1993; Webb, III *et al.*, 1998; Kutzbach *et al.*, 1998). *Picea glauca* is considered the most cold-hardy of the *Picea* (Fowells, 1965), and it is most tolerant of dry conditions (Nienstaedt and Zasada, 1990). Therefore, we hypothesize[2] that *P. glauca* was the dominant *Picea* species at Browns Pond, Virginia during the LGM. If *P. glauca* was dominant at this time, then this strongly supports previous vegetation-based inferences that LGM conditions were much colder and drier than present, and also colder and drier than conditions simulated by GCMs. If *P. rubens* or *P. mariana* were dominant, then a less severe LGM climate would be implied.

Between 12,730 and 12,260 ^{14}C yr BP, a local warming trend is proposed to have taken place based on pollen and macrofossil assemblages from Browns Pond, Virginia (Kneller and Peteet, 1999). This warming trend is correlative with warming interpreted in southern New England and Western Europe, hence it may be a regional event (Kneller and Peteet, 1999). Of the eastern *Picea* species, *P. rubens* currently has the most southern range (Morgenstern and Farrar, 1964; Little, Jr., 1971), and it tolerates the highest mean July temperatures (Fowells, 1965; Blum, 1990). We hypothesize[3] that *P. rubens* became the dominant or only *Picea* in western Virginia between 12,730 and 12,260 ^{14}C yr BP. If *P. rubens* was dominant at this time, then the assemblage can be considered analogous to modern higher elevation tree assemblages in the Appalachians, providing further support of the warming trend. If *P. glauca* or *P. mariana* are identified, then the assemblage present between 12,730 and 12,260 ^{14}C yr BP was more like the modern forest assemblages of the Adirondack or White Mountains.

---

[2] This hypothesis was formed with input from Drs. Margaret Kneller and Dorothy Peteet, Lamont Doherty Earth Observatory of Columbia University.

[3] This hypothesis was formed with input from Drs. Margaret Kneller and Dorothy Peteet.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Previously, it was believed that trees were displaced latitudinally as whole forest units during glacial/interglacial cycles in response to climate change. In contrast, it is now widely accepted that different taxa respond to climate change individually and migrate at different rates (Davis, 1983b; Bartlein *et al.*, 1986; Delcourt and Delcourt, 1987a; Prentice *et al.*, 1991). Therefore, it is not unreasonable to expect that different species also respond to environmental changes individually, especially when their individual environmental preferences and reproductive characteristics (Table 1) are considered. Following the LGM as the Laurentide ice sheet retreated, pollen records show that the leading northern margin of *Picea*'s range advanced northward thousands of years before the southern margin of its range retreated from the south (Delcourt and Delcourt, 1987a). I hypothesize that this significant expansion of the range of the genus (*Picea*) was caused by differential migration rates of the individual species (*P. glauca*, *P. mariana* and *P. rubens*). Furthermore, I hypothesize that *P. glauca* and *P. mariana* were the dominant *Picea* species at the northern margin of *Picea*'s range and that *P. rubens* was dominant at the southern margin of the genus' range throughout the 10,000 years which followed the LGM. If mainly *P. glauca* and *P. mariana* are identified in older Browns Pond sediments (18,000-17,000 ^{14}C yr BP) and *P. rubens* is identified in younger sediments (11,000-10,000 ^{14}C yr BP), then these hypotheses would be supported.

A primary goal of this project was to determine which species of *Picea* was dominant at Browns Pond, Virginia during the LGM. Using the modern analog method, Watts (1980b) and Whitehead (1981) estimated the climate of the southeastern United States during the LGM to be 17.8-25°C colder than modern mean January temperatures and 7.3-10°C colder than modern mean July temperatures. Pollen response surfaces were

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

used to estimate 8-10°C cooler temperatures and 20-40% lower precipitation during the LGM (Prentice *et al.*, 1991; Webb, III *et al.*, 1993). Interestingly, all these estimates are for a colder and drier LGM climate than that seen in GCM simulations of LGM climate (Hansen *et al.*, 1984; Rind and Peteet, 1985; Broccoli and Manabe, 1987; Kutzbach, 1993; Kutzbach and Ruddiman, 1993; Webb, III *et al.*, 1998; Kutzbach *et al.*, 1998). Identifying predominantly *P. glauca* macrofossils from Browns Pond sediments dated 18,000 to 14,000 ^{14}C yr BP would support the vegetation-based climate estimates and suggest that some GCM climate models cannot accurately simulate LGM climate of eastern North America. The species-identification of *P. glauca* would also allow a more precise estimate of the magnitude of the temperature depression and aridity during the LGM. Identification of predominantly *P. mariana* or *P. rubens* would show that 1) there may have been a moisture gradient from the edge of the ice sheet southward, or 2) there may have been more than one species of *Picea* present during this time. The latter case could be tested by further species-identifications of LGM macrofossils from other southeastern United States sites.

Another goal of this project is to determine which species of *Picea* was dominant at Browns Pond, Virginia between 12,730 and 12,260 ^{14}C yr BP. A local and probably regional warming trend between 12,730 and 12,260 ^{14}C yr BP was inferred from an increase in deciduous hardwood pollen, an increase in *Tsuga* macrofossils and a decrease in *Alnus* macrofossils from Browns Pond (Kneller and Peteet, 1999). Presently, *Tsuga*, hardwood deciduous trees and *P. rubens* grow together at higher elevations in the central Appalachians. If *P. rubens* were identified at Browns Pond during this time period, then this would support the inference of warming, and it would suggest that the pollen and

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

macrofossil assemblage at the time is analogous to modern forests at higher elevations in the central Appalachians. If *P. mariana* or *P. glauca* are present at this time, then an assemblage analogous to lower elevations of the Adirondack or White Mountains existed. The Adirondack and White Mountains have much cooler climates than the central Appalachians. Based on which analogue is supported, a more precise local climate inference can be made. Species identifications of *Picea* macrofossils of this time period from other eastern United States sites could be used to test the hypothesis that this warming trend is regional in extent.

The final goal of this project is to determine which species of *Picea* were present at Browns Pond, Virginia at ~1,000 year intervals spanning from the LGM to the Early-Holocene. Delcourt and Delcourt (1987a) estimated that between 12,000 and 10,000 years ago, *Picea* advanced across New England at a rate of 368 m/year. The highest mean rate of advance for the leading margin of *Picea*'s range was 242 m/year between 10,000 and 8,000 yr BP as it invaded central and eastern Canada. Interestingly, the migration of *Picea*'s southern range margin was markedly different than its northern limit. In fact, the northern margin of *Picea*'s range began to spread northward 4,000 years before its southern margin began to retreat from the south (Delcourt and Delcourt, 1987a). This expansion of the genus' range could be explained by differential migration rates of the different *Picea* species.

Acknowledgements

This project would not have been possible without collaborators Dr. Margaret Kneller and Dr. Dorothy Peteet of the Lamont Doherty Earth Observatory, Columbia University. Dr. Kneller is credited with the original idea of using ancient DNA to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

identify botanical species of fossils recovered from sediment cores for the purpose of generating more precise vegetation histories that could then be applied to paleoclimate studies. Some core samples were provided by Drs. Kneller and Peteet, while others were obtained directly with the help of Dr. Kneller, Mr. Thaddeus Presby and Mr. Aaron Wallace. Technical advice and guidance was provided by Dr. Richard Blakemore, Department of Microbiology, UNH, Dr. Hendrik Poinar, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Dr. Paul Mayewski, Bryand Global Sciences Center, University of Maine and Dr. Queta Boese, Department of Biochemistry and Molecular Biology, UNH. The synthesis of PTB was carried out with the help of Dr. Andrew Laudano, Department of Biochemistry and Molecular Biology, UNH. Laboratory space and equipment for conducting work with ancient DNA was provided by Mrs. Gail Stewart and the Department of Biochemistry and Molecular Biology, UNH and by Mr. Robert Mooney and the Department of Microbiology, UNH. Bacterial and fungal DNA samples for Southern blot experiments were provided by Dr. John Collins and Dr. Clyde Denis, Department of Biochemistry and Molecular Biology, UNH, Ms. Therese Thompson and Dr. Robert Blanchard, Department of Plant Biology, UNH, and Dr. Francisco Camacho, Department of Botany and Plant Pathology, Oregon State University. Funding for the coring field trip was provided by the New Hampshire Space Grant Consortium. This research was supported by a grant from the New Hampshire Agricultural Experiment Station (NHAES) MacIntire-Stennis Project MS27.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF REFERENCES

Abbott,R.J., Smith,L.C., Milne,R.I., Crawford,R.M., Wolff,K., Balfour,J. (2000). Molecular analysis of plant migration and refugia in the Arctic. Science *289*, 1343-1346.

Albert,V.A., Backlund,A., Bremer,K., Chase,M.W., Manhart,J.R., Mishler,B.D., Nixon,K.C. (1994). Functional constraints and *rbcL* evidence for land plant phylogeny. Annals of the Missouri Botanical Garden *81*, 534-567.

Aldén,B. (1987). Taxonomy and geography of the genus *Picea*. In: International Dendrology Society Yearbook 85-96.

Alley,R.B., Meese,D.A., Shuman,C.A., Gow,A.J., Taylor,K.C., Grootes,P.M., White,J.W.C., Ram,M., Waddington,E.D., Mayewski,P.A., Zielinski,G.A. (1993). Abrupt increase in Greenland snow accumulation at the end of the Younger Dryas event. Nature *362*, 527-529.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., Lipman,D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.

Austin,J.J., Smith,A.B., Thomas,R.H. (1997). Paleontology in a molecular world: the search for authentic ancient DNA. Tree *12*, 303-306.

Averof,M., Rokas,A., Wolfe,K.H., Sharp,P.M. (2000). Evidence for a high frequency of simultaneous double-nucleotide substitutions. Science *287*, 1283-1286.

Baldwin,B.G., Sanderson,M.J., Porter,J.M., Wojciechowski,M.F., Campbell,C.S., Donoghue,M.J. (1995). The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden *82*, 247-277.

Bartlein,P.J., Prentice,I.C., Webb,T., III (1986). Climatic response surfaces from pollen data for some eastern North American taxa. Journal of Biogeography *13*, 35-57.

Beiswenger,J.M. (1991). Late Quaternary vegetational history of Grays Lake, Idaho. Ecological Monographs *61*, 165-182.

Birks,H.J.B., Peglar,S.M. (1980). Identification of *Picea* pollen of Late Quaternary age in eastern North America: a numerical approach. Canadian Journal of Botany *58*, 2043-2058.

Björck,S., Kromer,B., Johnsen,S.J., Bennike,O., Hammarlund,D., Lemdahl,G., Possnert,G., Rasmussen,T.L., Wohlfarth,B., Hammer,C.U., Spurk,M. (1996). Synchronized terrestrial-atmospheric deglacial records around the North Atlantic. Science *274*, 1155-1160.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Blum,B.M. (1990). *Picea rubens* Sarg. Red spruce. In: Silvics of North America, ed. R.M.Burns, B.H.Honkala. Washington, D.C.: Forest Service, United States Department of Agriculture, 250-259.

Bobola,M.S., Eckert,R.T., Klein,A.S. (1992a). Restriction fragment variation in the nuclear ribosomal DNA repeat unit within and between *Picea rubens* and *Picea mariana*. Canadian Journal of Forest Research *22*, 255-263.

Bobola,M.S., Eckert,R.T., Klein,A.S., Stapelfeldt,K., Hillenberg,K.A., Gendreau,S.B. (1996a). Hybridization between *Picea rubens* and *Picea mariana*: differences observed between montane and coastal island populations. Canadian Journal of Forest Research *26*, 444-452.

Bobola,M.S., Eckert,R.T., Klein,A.S., Stapelfeldt,K., Smith,D.E., Guenette,D. (1996b). Using nuclear and organelle DNA markers to discriminate among *Picea rubens*, *Picea mariana,* and their hybrids. Canadian Journal of Forest Research *26*, 433-443.

Bobola,M.S., Smith,D.E., Klein,A.S. (1992b). Five major nuclear ribosomal repeats represent a large and variable fraction of the genomic DNA of *Picea rubens* and *P. mariana*. Molecular Biology and Evolution *9*, 125-137.

Boom,R., Sol,C.J.A., Salismans,M.M.M., Jansen,C.L., Wertheim-van Dillen,P.M.E., van der Noordaa,J. (1990). Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology *28*, 495-503.

Broccoli,A.J., Manabe,S. (1987). The effects of the Laurentide ice sheet on North American climate during the last glacial maximum. Géographie physique et Quaternaire *XLI*, 291-299.

Chang,J., Hanover,J.W. (1991). Geographic variation in the monoterpene composition of black spruce. Canadian Journal of Forest Research *21*, 1796-1800.

Chase,M.W., Soltis,D.E., Olmstead,R.G., Morgan,D., Les,D.H., Mishler,B.D., Duvall,M.R., Price,R.A., Hills,H.G., Qiu,Y.L., Kron,K.A., Retting,J.H., Conti,E., Palmer,J.D., Manhart,J.R., Systma,K.J., Michaels,H.J., Kress,W.J., Karol,K.G., Clark,W.D., Hedren,M., Gaut,B.S., Jansen,R.K., Kim,K.J., Wimpee,C.F., Smith,J.F., Furnier,G.R., Strauss,S.H., Xiang,G.Y., Plunkett,G.M., Soltis,P.S., Swensen,W.M., Williams,S.E., Gadek,P.A., Quinn,C.J., Equiarte,L.E., Golenberg,E., Learn,G.H.J., Graham,S.W., Barrett,S.C.H., Dayanandan,S., Albert,V.A. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Annals of the Missouri Botanical Garden *80*, 528-580.

Clark,J.S., Fastie,C., Hurtt,G., Jackson,S.T., Johnson,C., King,G.A., Lewis,M., Lynch,J., Pacala,S., Prentice,C., Schupp,E.W., Webb,T., III, Wyckoff,P. (1998). Reid's paradox of rapid plant migration. BioScience *48*, 13-24.

Collignon,A.M., Favre,J.M. (2000). Contribution to the postglacial history at the western margin of *Picea abies*' natural area using RAPD markers. Annals of Botany *85*, 713-722.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Comes,H.P., Kadereit,J.W. (1998). The effect of Quaternary climatic changes on plant distribution and evolution. Trends in plant science *3*, 432-438.

Critchfield,W.B. (1984). Impact of the Pleistocene on the genetic structure of North American conifers. In: Proceedings of the eighth North American Forest Biology Workshop, ed. R.M.Lanner. International Union of Forest Research Organizations, 70-118.

D'Aquila,R.T., Bechtel,L.J., Videler,J.A., Eron,J.J., Gorczyca,P., Kaplan,J.C. (1991). Maximizing sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Research *19*, 3749.

Dansgaard,W., Johnsen,S.J., Clausen,H.B., Dahl-Jenson,D., Gundestrup,N., Hammer,C.U., Oeschger,H. (1984). North Atlantic Climatic Oscillations Revealed by Deep Greenland Ice Cores. In: Climate processes and climate sensitivity, ed. J.E.Hansen, T.Takahashi. Washington D.C.: American Geophysical Union, 288-298.

David,A.J., Keathley,D.E. (1996). Inheritance of mitochondrial DNA in interspecific crosses of *Picea glauca* and *Picea omorika*. Canadian Journal of Forest Research *26*, 428-432.

Davis,M.B. (1958). Three pollen diagrams from Central Massachusetts. American Journal of Science *256*, 540-570.

Davis,M.B. (1983a). Holocene vegetation history of the eastern United States. In: Late-Quaternary environments of the United States, ed. H.E.J.Wright. Minneapolis: University of Minnesota Press, 166-181.

Davis,M.B. (1983b). Quaternary history of deciduous forests of eastern North America and Europe. Annals of the Missouri Botanical Garden *70*, 550-563.

Davis,M.B., Shaw,R.G. (2001). Range shifts and adaptive responses to Quaternary climate change. Science *292*, 673-679.

Delcourt,H.R., Delcourt,P.A. (1985). Quaternary palynology and vegetational history of the southeastern United States. In: Pollen records of Late-Quaternary North American sediments, ed. V.M.Bryant, Jr., R.G.Holloway. American Association of Stratigraphic Palynologists Foundation, 1-37.

Delcourt,P.A., Delcourt,H.R. (1987a). Late-Quaternary dynamics of temperate forests: applications of paleoecology to issues of global environmental change. Quaternary Science Reviews *6*, 129-146.

Delcourt,P.A., Delcourt,H.R. (1987b). Long-term forest dynamics of the temperate zone. New York: Springer-Verlag.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Demesure,B., Sodzi,N., Petit,R.J. (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology *4*, 129-131.

Don,R.H., Cox,P.T., Wainwright,B.J., Baker,K., Mattick,J.S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research *19*, 4008.

Dong,J., Wagner,D.B. (1994). Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. Genetics *136*, 1187-1194.

Doyle,J.J., Doyle,J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin *19*, 11-15.

Duman,M.G. (1957). Resin cavity pattern in the needles of black, red and white spruce. Bulletin of the Torrey Botanical Club *84*, 388.

Dumolin-Lapegue,S., Pemonge,M.H., Petit,R.J. (1997). An enlarged set of consensus primers for the study of organelle DNA in plants. Molecular Ecology *6*, 393-397.

Eckert,R.T. (1989). Genetic variation in red spruce and its relation to forest decline in the northeastern United States. In: Air Pollution and Forest Decline, ed. J.B.Bucher, I.Bucher-Wallin. Birmensdorf, Switzerland: International Union of Forest Research Organizations P2.05, 319-324.

Eckert, R. T. Genetic variation in red spruce pollution response. Final report. 1990. Radnor, PA, Spruce-Fir Research Cooperative, Northeast Forest Experiment Station.

Ennos,R.A. (1994). Estimating the relative rates of pollen and seed migration among plant populations. Heredity *72*, 250-259.

Fowells,H.A. (1965). Silvics of forest trees of the United States. Washington D.C.: U.S. Department of Agriculture.

Fowler,D.P. (1966). A new spruce hybrid *Picea schrenkiana* X *P. glauca*. In: Proceedings of the 2nd Genetics Workshop of Soc. Am. For U.S. Forest Service North Central Experiment Station, 44-47.

Fowler,D.P., Park,Y.S., Gordon,A.G. (1988). Genetic variation of red spruce in the Maritimes. Canadian Journal of Forest Research *18*, 703-709.

Furnier,G.R., Stine,M. (1995). Interpopulation differentiation of nuclear and chloroplast loci in white spruce. Canadian Journal of Forest Research *25*, 736-742.

Furnier,G.R., Stine,M., Mohn,C.A., Clyde,M.A. (1991). Geographic patterns of variation in allozymes and height growth in white spruce. Canadian Journal of Forest Research *21*, 707-712.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Germano, J., Cox, M., Wright, W. A., Arsenault, M. P., Klein, A. S., and Campbell, C. S. A chloroplast DNA phylogeny of *Picea* (Pinaceae). Botany 2002 Proceedings (Abstract). 2002.

Germano,J., Klein,A.S. (1999). Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca, P. mariana* and *P. rubens*. Theoretical and Applied Genetics *99*, 37-49.

Gielly,L., Taberlet,P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. Molecular Biology and Evolution *11*, 769-777.

Givens,C.R., Givens,F.M. (1987). Age and significance of fossil white spruce (*Picea glauca*), Tunica Hills, Louisiana-Mississippi. Quaternary Research *27*, 283-296.

Goldman,N., Anderson,J.P., Rodrigo,A.G. (2000). Likelihood-based tests of topologies in phylogenetics. Syst Biol. *49*, 652-670.

Gordon,A.G. (1976). The taxonomy and genetics of *Picea rubens* and its relationship to *Picea mariana*. Canadian Journal of Botany *54*, 781-813.

Gordon, A. G. Crossability in the genus *Picea* with special emphasis on the Mexican species. Joint meeting of Western Forest Genetics and IUFRO Working Parties S2.02-05,12 2, 114-116. 1992.

Grivet,D., Jeandroz,S., Favre,J.M. (1999). *Nad1* b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. Theoretical and Applied Genetics *99*, 346-349.

Gugerli,F., Senn,J., Anzidei,M., Madaghiele,A., Büchler,U., Sperisen,C., Vendramin,G.G. (2001a). Chloroplast microsatellites and mitochondrial nad1 intron 2 sequences indicate congruent phylogenetic relationships among Swiss stone pine (*Pinus cembra*), Siberian stone pine (*Pinus sibirica*), and Siberian dwarf pine (*Pinus pumila*). Molecular Ecology *10*, 1489-1497.

Gugerli,F., Sperisen,C., Buchler,U., Brunner,I., Brodbeck,S., Palmer,J.D., Qiu,Y.L. (2001b). The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. Molecular Phylogenetics and Evolution *21*, 167-175.

Hamburg,S.P., Cogbill,C.V. (1988). Historical decline of red spruce populations and climatic warming. Nature *331*, 428-431.

Hamrick,J.L., Godt,M.J.W. (1990). Allozyme Diversity in Plant Species. In: Plant population genetics, breeding, and genetic resources, ed. A.H.D.Brown. Sunderland, Mass: Sinauer Associates, 43-63.

Hansen,B.C.S., Engstrom,D.R. (1985). A comparison of numerical and qualitative methods of separating pollen of black and white spruce. Canadian Journal of Botany *63*, 2159-2163.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Hansen,J.E., Lacis,A., Rind,D., Russell,G., Stone,P., Fung,I., Ruedy,R., Lerner,J. (1984). Climate Sensitivity: Analysis of Feedback Mechanisms. In: Climate processes and climate sensitivity, ed. J.E.Hansen, T.Takahashi. Washington, D.C.: American Geophysical Union, 130-163.

Hawley,G.J., DeHayes,D.H. (1994). Genetic diversity and population structure of red spruce (*Picea rubens*). Canadian Journal of Botany *72*, 1778-1786.

Hewitt,G.M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society *58*, 247-276.

Hewitt,G.M. (2000). The genetic legacy of the Quaternary ice ages. Nature *405*, 907-913.

Hills,L.V., Ogilvie,R.T. (1970). *Picea banksii* n. sp. Beaufort formation (tertiary), northwestern Banks Island, Artic Canada. Canadian Journal of Botany *48*, 457-464.

Hoffman, D and Kleinschmidt, J. An utilization program for spruce provenances and species hybrids. S2.03.22-S2.02.22, 216-236. 1979. Bucharest. IUFRO Norway spruce meeting.

Holloway,R.G., Bryant,V.M., Jr. (1984). *Picea glauca* pollen from late glacial deposits in central Texas. Palynology *8*, 21-32.

Höss,M., Dilling,A., Currant,A., Pääbo,S. (1996a). Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. Proceedings of the National Academy of Sciences USA *93*, 181-185.

Höss,M., Jaruga,P., Zastawny,T.H., Dizdaroglu,M., Pääbo,S. (1996b). DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Research *24*, 1304-1307.

Höss,M., Pääbo,S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Research *21*, 3913-3914.

Höss,M., Pääbo,S., Vereshchagin,N.K. (1994). Mammoth DNA sequences. Nature *370*, 333.

Hughen,K.A., Overpeck,J.T., Peterson,L.C., Trumbore,S. (1996). Rapid climate changes in the tropical Atlantic region during the last deglaciation. Nature *380*, 51-54.

Hultén,E. (1968). Flora of Alaska and neighboring territories; a manual of the vascular plants. Stanford, CA: Stanford University Press.

Huntley,B., Birks,H.J.B. (1983). An atlas of past and present pollen maps for Europe, 0-13,000 years ago. Cambridge: Cambridge University Press.

Isoda,K., Shiraishi,S., Kisanuki,H. (2000). Classifying *Abies* species (Pinaceae) based on the sequence variation of a tandemly repeated array found in the Chloroplast DNA *trnL* and *trnF* intergenic spacer. Silvae Genetica *49*, 161-165.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Jackson,S.T., Givens,C.R. (1994). Late Wisconsinan vegetation and environment of the Tunica Hills region, Louisiana/Mississippi. Quaternary Research *41*, 316-325.

Jackson,S.T., Overpeck,J.T., Webb,T., III, Keattch,S.E., Anderson,K.H. (1997). Mapped plant-macrofossil and pollen records of Late Quaternary vegetation change in eastern North America. Quaternary Science Reviews *16*, 1-70.

Jackson,S.T., Weng,C. (1999). Late quaternary extinction of a tree species in eastern North America. Proc.Natl.Acad.Sci.U.S.A *96*, 13847-13852.

Jacobson,G.L., Webb,T., III, Grimm,E.C. (1987). Patterns and rates of vegetation change during the deglaciation of eastern North America. In: North America and adjacent oceans during the last deglaciation, ed. H.E.J.Wright, W.F.Ruddiman. Boulder, Colorado: Geological Society of America, 277-288.

Jarne,P., Lagoda,P.J.L. (1996). Microsatellites, from molecules to populations and back. Tree *11*, 424-429.

Johnsen,S.J., Clausen,H.B., Dansgaard,W., Fuhrer,K., Gundestrup,N., Hammer,C.U., Iversen,P., Jouzel,J., Stauffer,B., Steffensen,J.P. (1992). Irregular glacial interstadials recorded in a new Greenland ice core. Nature *359*, 311-313.

Johnson,L.A., Soltis,D.E. (1994). *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Systematic Botany *19*, 143-156.

Khalil,A.K. (1987). Genetic variation in red spruce (*Picea rubens* Sarg.). Silvae Genetica *36*, 164-171.

Kneller,M., Peteet,D.M. (1993). Late-Quaternary climate in the Ridge and Valley of Virginia, U.S.A.: changes in vegetation and depositional environment. Quaternary Science Reviews *12*, 613-628.

Kneller,M., Peteet,D.M. (1999). Late-glacial to early Holocene climate changes from a central Appalachians pollen and macrofossil record. Quaternary Research *51*, 133-147.

Koç Karpuz,N., Jansen,E. (1992). A high-resolution diatom record of the last deglaciation from the SE Norwegian Sea: documentation of rapid climatic changes. Paleoceanography *7*, 499-520.

Krings,M., Stone,A., Schmitz,R.W., Krainitzki,H., Stoneking,M., Pääbo,S. (1997). Neandertal DNA sequences and the origin of modern humans. Cell *90*, 19-30.

Krutovskii,K.V., Bergmann,F. (1995). Introgressive hybridization and phylogenetic relationships between Norway, *Picea abies* (L.) Karst., and Siberian, *P. obovata* Ledeb., spruce species studied by isozyme loci. Heredity *74*, 464-480.

Kutzbach,J.E. (1993). Simulated climatic changes: results of the COHMAP climate-model experiments. In: Global climates since the Last Glacial Maximum, ed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

H.E.J.Wright, J.E.Kutzbach, T.Webb, III, W.F.Ruddiman, F.A.Street-Perrott, P.J.Bartlein. Minneapolis: University of Minnesota Press, 24-93.

Kutzbach,J.E., Gallimore,R., Harrison,S.P., Behling,P., Selin,R., Laarif,F. (1998). Climate and biome simulation for the past 21,000 years. Quaternary Science Reviews *17*, 473-506.

Kutzbach,J.E., Ruddiman,W.F. (1993). Model description, external forcing, and surface boundary conditions. In: Global climates since the Last Glacial Maximum, ed. H.E.J.Wright, J.E.Kutzbach, T.Webb, III, W.F.Ruddiman, F.A.Street-Perrott, P.J.Bartlein. Minneapolis: University of Minnesota Press, 12-23.

Laroche,J., Li,P., Maggia,L., Bousquet,J. (2000). Molecular evolution of angiosperm mitochondrial introns and exons. Proceedings of the National Academy of Sciences USA *94*, *5722-5727.*

Larson,A. (1994). The Comparison of Morphological and Molecular Data in Phylogenetic Systematics. In: Molecular ecology and evolution: Approaches and applications, ed. B.Schierwater, B.Streit, G.B.Wagner, R.DeSalle. Basel: Birkhäuser Verlag, 371-390.

Latta,R.G., Linhart,Y.B., Fleck,D., Elliot,M. (1998). Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. Evolution *52*, 61-67.

Latta,R.G., Mitton,J.B. (1997). A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). Genetics *146*, 1153-1163.

Lehman,S.J., Keigwin,L.D. (1992). Sudden changes in North Atlantic circulation during the last deglaciation. Nature *356*, 757-762.

LePage,B.A. (2001). New species of *Picea* A. Dietrich (Pinaceae) from the middle Eocene of Axel Heiberg Island, Arctic Canada. Botanical Journal of the Linnean Society *135*, 137-167.

Levesque,A.J., Mayle,F.E., Walker,I.R., Cwynar,L.C. (1993). A previously unrecognized late-glacial cold event in eastern North America. Nature *361*, 623-626.

Lindahl,T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

Lindahl,T. (1997). Facts and artifacts of ancient DNA. Cell *90*, 1-3.

Little,E.L., Jr. (1971). Conifers and important hardwoods. Washington, D.C.: U.S. Department of Agriculture, Forest Service.

Little,E.L., Jr., Pauley,S.S. (1958). A natural hybrid between black and white spruce in Minnesota. American Midland Naturalist *60*, 202-211.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Lowe,A.J., Gillies,A.C., Wilson,J., Dawson,I.K. (2000). Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Molecular Biology and Evolution *9*, 831-841.

MacDonald,G.M., Edwards,T.W.D., Moser,K.A., Pienitz,R., Smol,J.P. (1993). Rapid response of treeline vegetation and lakes to past climate warming. Nature *361*, 243-246.

Manley,S.A.M. (1972). The occurrence of hybrid swarms of red and black spruces in central New Brunswick. Canadian Journal of Forest Research *2*, 381-391.

Mayewski,P.A., Meeker,L.D., Whitlow,S., Twickler,M.S., Morrison,M.C., Alley,R.B., Bloomfield,P., Taylor,K. (1993). The atmosphere during the Younger Dryas. Science *261*, 195-197.

Mayr,E. (1992). A local flora and the biological species concept. American Journal of Botany *79*, 222-238.

McLeod,T.K., MacDonald,G.M. (1997). Postglacial range expansion and population growth of *Picea mariana*, *Picea glauca* and *Pinus banksiana* in the western interior of Canada. Journal of Biogeography *24*, 865-881.

Miller,C.N., Jr. (1977). Mesozoic conifers. The Botanical Review *43*, 217-281.

Miller,C.N., Jr. (1989). A new species of *Picea* based on silicified seed cones from the Oligocene of Washington. American Journal of Botany *76*, 747-754.

Mitton,J.B., Kreiser,B.R., Latta,R.G. (2000a). Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. Molecular Ecology *9*, 91-97.

Mitton,J.B., Kreiser,B.R., Rehfeldt,G.E. (2000b). Primers designed to amplify a mitochondrial *nad1* intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*. Theoretical and Applied Genetics *101*, 1269-1272.

Morgenstern,E.K. (1978). Range-wide genetic variation of black spruce. Canadian Journal of Forest Research *8*, 463-473.

Morgenstern,E.K., Farrar,J.L. (1964). In: Introgressive hybridization in red spruce and black spruce. Toronto, Ontario: Faculty of Forestry, University of Toronto, 1-46.

Mott,R.J., Grant,D.R., Stea,R., Occhietti,S. (1986). Late-glacial climatic oscillation in Atlantic Canada equivalent to the Allerod/Younger Dryas event. Nature *323*, 247-250.

Mullis,K.B., Faloona,F.A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. In: Recombinant DNA Part F, ed. R.Wu. San Diego: Academic Press Inc., 335-350.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Newton,A.C., Allnutt,T.R., Gillies,A.C.M., Lowe,A.J., Ennos,R.A. (1999). Molecular phylogeography, intraspecific variation and the conservation of tree species. Trends in Ecology and Evolution *14*, 140-145.

Nienstaedt,H., Zasada,J.C. (1990). *Picea glauca* (Moench) Voss: White Spruce. In: Silvics of North America, ed. R.M.Burns, B.H.Honkala. Washington, DC: Forest Service, United States Department of Agriculture, 204-226.

Nishizawa,S., Kubo,T., Mikami,T. (2000). Variable number of tandem repeat loci in the mitochondrial genomes of beets. Current Genetics *37*, 34-38.

Okayama,H., Curiel,D.T., Brantly,M.L., Holmes,M.D., Crystal,R.G. (1989). Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification. The Journal of Laboratory and Clinical Medicine *114*, 105-113.

Overpeck,J.T., Webb,R.S., Webb,T., III (1992). Mapping eastern North American vegetation change of the past 18 ka: no-analogs and the future. Geology *20*, 1071-1074.

Palmer,J.D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. *6*, 115-120.

Palmer,J.D. (1992). Mitochondrial DNA in plant systematics: Applications and limitations. In: Molecular Systematics of Plants, ed. P.S.Soltis, D.E.Soltis, J.L.Doyle. New York: Chapman and Hall, 36-49.

Pääbo,S. (1989). Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification. Proceedings of the National Academy of Sciences USA *86*, 1939-1943.

Pääbo,S. (1990). Amplifying Ancient DNA. In: PCR protocols: a guide to methods and applications, ed. M.A.Innis, D.H.Gelfand, J.J.Sninsky, T.J.White. San Diego: Academic Pres Inc., 159-166.

Pääbo,S., Gifford,J.A., Wilson,A.C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. Nucleic Acids Research *16*, 9775-9787.

Perez de la Rosa,J., Harris,S.A., Farjon,A. (1995). Noncoding chloroplast DNA variation in Mexican pines. Theoretical and Applied Genetics *91*, 1101-1106.

Perron,M., Bousquet,J. (1997). Natural hybridization between black spruce and red spruce. Molecular Ecology *6*, 725-734.

Perron,M., Gordon,A.G., Bousquet,J. (1995). Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. Theoretical and Applied Genetics *91*, 142-149.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Perron,M., Perry,D.J., Andalo,C., Bousquet,J. (2000). Evidence from sequence-tagged-site markers of a recent progenitor-derivative species pair in conifers. Proceedings of the National Academy of Sciences USA *97*, 11331-11336.

Perry,D.J., Bousquet,J. (1998). Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theoretical and Applied Genetics *97*, 735-743.

Peteet,D.M. (1991). Postglacial migration history of lodgepole pine near Yakutat, Alaska. Canadian Journal of Botany *69*, 786-796.

Peteet,D.M., Alley,R.B., Bond,G., Chappellaz,J., Clapperton,C., Del Genio,A., Keigwin,L.D. (1993). Global Younger Dryas? EOS, Transactions, American Geophysical Union *74*, 587-589.

Peteet,D.M., Daniels,R., Heusser,L.E., Vogel,J.S., Southon,J.R., Nelson,D.E. (1994). Wisconsinan Late-glacial environmental change in southern New England: a regional synthesis. Journal of Quaternary Science *9*, 151-154.

Peteet,D.M., Vogel,J.S., Nelson,D.E., Southon,J.R., Nickmann,R.J., Heusser,L.E. (1990). Younger Dryas climatic reversal in northeastern USA? AMS ages for an old problem. Quaternary Research *33*, 219-230.

Petersen,G., Johansen,B., Seberg,O. (1999). PCR and sequencing from a single pollen grain. Plant Molecular Biology *31*, 189-191.

Petit,R.J., Kremer,A., Wagner,D.B. (1993). Geographic structure of chloroplast DNA polymorphisms in European oaks. Theoretical and Applied Genetics *87*, 122-128.

Pla,M., Mathieu,C., De Paepe,R., Chetrit,P., Vedel,F. (1995). Deletion of the last two exons of the mitochondrial *nad7* gene results in lack of the NAD7 polypeptide in a *Nicotiana sylvestris* CMS mutant. Molecular and General Genetics *248*, 79-88.

Poinar,H.N., Hofreiter,M., Spaulding,W.G., Martin,P.S., Stankiewicz,B.A., Bland,H., Evershed,R.P., Possnert,G., Pääbo,S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. Science *281*, 402-406.

Pons,O., Petit,R.J. (1995). Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. Theoretical and Applied Genetics *90*, 462-470.

Prentice,I.C., Bartlein,P.J., Webb,T., III (1991). Vegetation and climate change in eastern North America since the last glacial maximum. Ecology *72*, 2038-2056.

Rehfeldt,G.E., Wykoff,W.R., Ying,C.C. (2001). Physiologic plasticity, evolution, and impacts of a changing climate on *Pinus contorta*. Climatic Change *50*, 355-376.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Riemenschneider,D., Mohn,C.A. (1975). Chromatographic analysis of an open-pollinated Rosendahl spruce progeny. Canadian Journal of Forest Research *5*, 414-418.

Rind,D., Peteet,D.M. (1985). Terrestrial conditions at the Last Glacial Maximum and CLIMAP sea-surface temperature estimates: are they consistent? Quaternary Research *24*, 1-22.

Ritchie,J.C. (1980). Towards a Late-Quaternary palaeoecology of the ice-free corridor. Canadian Journal of Anthropology *1*, 15-28.

Ritchie,J.C., MacDonald,G.M. (1986). The patterns of post-glacial spread of white spruce. Journal of Biogeography *13*, 527-540.

Rowe,J.S. (1972). Forest Regions of Canada. Ottawa: Canadian Forestry Service Department of the Environment.

Sambrook,J., Fritch,E.F., Maniatis,T. (1989). Molecular Cloning. A Laboratory Manual. Cold Harbor Spring Laboratory Press.

Schmidt-Vogt,H. (1977). Die Fichte, Ein Handbuch in Zwei Banden Band I Taxonomie, Verbreitun, Morphologie, Okologie, Waldgesellschaften. Hamburg: Verlag Paul Parey.

Schneider, S., Roessli, D., and Excoffier, L. Arlequin ver. 2.000: A software for population genetics data analysis. University of Geneva, Switzerland, Genetics and Biometry Laboratory. 2000.

Scotti,I., Vendramin,G.G., Matteotti,L.S., Scarponi,C., Sari-Gorla,M., Binelli,G. (2000). Postglacial recolonization routes for *Picea abies* K. in Italy as suggested by the analysis of sequence-characterized amplified region (SCAR) markers. Molecular Ecology *9*, 699-708.

Sheffield,V.C., Beck,J.S., Kwitek,A.E., Sandstrom,D.W., Stone,E.M. (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics *16*, 325-332.

Sigurgeirsson, A. Insights into the evolution of *Picea* inferred from chloroplast DNA (Dissertation). 1-25. 1992. Swedish University of Agricultural Sciences.

Sigurgeirsson,A., Szmidt,A.E. (1993). Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. Nordic Journal of Botany *13*, 233-246.

Sigurgeirsson,A., Szmidt,A.E., Karpinska,B. (1991). Alaskan *Picea sitchensis* populations infiltrated with *Picea glauca* genes: a study using DNA markers. In: Biochemical markers in the population genetics of forest trees, ed. S.Fineschi, M.E.Malvolti, F.Cannata, H.H.Hattemer. The Hague, The Netherlands: SPB Academic Publishing, 197-207.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Soltis,D.E., Gitzendanner,M.A., Strenge,D.D., Soltis,P.S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution *206*, 353-373.

Sommer,S.S., Cassady,J.D., Sobell,J.L., Bottema,C.D. (1989). A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. Mayo Clinic Proceedings *64*, 1361-1372.

Sperisen,C., Büchler,U., Gugerli,F., Mátyás,G., Geburek,T., Vendramin,G.G. (2001). Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Molecular Ecology *10*, 257-263.

Sperisen,C., Büchler,U., Mátyás,G. (1998). Genetic variation of mitochondrial DNA reveals subdivision of Norway spruce (*Picea abies* (L.) Karst.). In: Molecular tools for screening biodiversity, ed. A.Karp, P.G.Isaac, D.S.Ingram. London: Chapman & Hall, 413-417.

Stewart,J.R., Lister,A.M. (2001). Cryptic northern refugia and the origins of the modern biota. Trends in Ecology and Evolution *16*, 608-613.

Stine,M., Keathley,D.E. (1990). Paternal inheritance of plastids in Engelmann spruce X blue spruce hybrids. Journal of Heredity *81*, 443-446.

Stine,M., Sears,B.B., Keathley,D.E. (1989). Inheritance of plastids in interspecific hybrids of blue spruce and white spruce. Theoretical and Applied Genetics *78*, 768-774.

Sutton,B.C.S., Flanagan,D.J., Gawley,J.R., Newton,C.H., Lester,D.T., El-Kassaby,Y.A. (1991). Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. Theoretical and Applied Genetics *82*, 242-248.

Sutton,B.C.S., Pritchard,S.C., Gawley,J.R., Newton,C.H., Kiss,G.K. (1994). Analysis of Sitka spruce--interior spruce introgression in British Columbia using cytoplasmic and nuclear DNA probes. Canadian Journal of Forest Research *24*, 278-285.

Suyama,Y., Kawamuro,K., Kinoshita,I., Yoshimura,K., Tsumura,Y., Takahara,H. (1996). DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. Genes and Genetic Systems *71*, 145-149.

Swofford, D. L. PAUP* Phylogenetic Analysis Using Parsimony (* and Other Methods). [4.06b]. Sunderland, MA, Sinauer Associates. 1998.

Taberlet,P., Fumagalli,L., Wust-Saucy,A.G., Cossons,J.F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology *7*, 453-464.

Taberlet,P., Gielly,L., Pautou,G., Bouvet,J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology *17*, 1105-1109.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Taylor,P.G. (1996). Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Molecular Biology and Evolution *13*, 283-285.

Taylor,R.J., Patterson,T.F. (1980). Biosystematics of Mexican spruce species and populations. Taxon *29*, 421-440.

Taylor,R.J., Patterson,T.F., Harrod,R.J. (1994). Systematics of Mexican spruce--revisited. Systematic Botany *19*, 47-59.

Torsvik,V. (1995). Cell extraction method. In: Molecular microbial ecology manual, ed. A.D.L.Akkermans, J.D.van Elsas, F.J.de Bruijn. Dordrecht: Kluwer Academic Publishers, 1-15.

Tremblay,N.O., Schoen,D.J. (1999). Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. Molecular Ecology *8*, 1187-1198.

Tsay,R.C., Taylor,I.E.P. (1978). Isoenzyme complexes as indicators of genetic diversity in white spruce, *Picea glauca*, in southern Ontario and the Yukon Territory. Formic, glutamic, and lactic dehydrogenases and cationic peroxidases. Canadian Journal of Botany *56*, 80-90.

Tsudzuki,J., Nakashima,K., Tsudzuki,T., Hiratsuka,J., Shibata,M., Wakasugi,T., Sugiura,M. (1992). Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trn*Q, *trn*K, *psb*A, *trn*I and *trn*H and the absence of *rps*16. Molecular and General Genetics *232*, 206-214.

Vann,D.R., Johnson,A.H., Casper,B.B. (1994). Effect of elevated temperatures on carbon dioxide exchange in *Picea rubens*. Tree Physiology *14*, 1339-1349.

Vasan,S., Zhang,X., Kapurniotu,A., Bernhagen,J., Teichberg,S., Basgen,J., Wagle,D., Shih,D., Terlecky,I., Bucala,R., Cerami,A., Egan,J., Ulrich,P. (1996). An agent cleaving glucose-derived protein crosslinks *in vitro* and *in vivo*. Nature *382*, 275-278.

Vendramin,G.G., Anzidei,M., Madaghiele,A., Sperisen,C., Bucci,G. (2000). Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). Genome *43*, 68-78.

Vidakovic,M. (1991). Conifers: morphology and variation. Zagreb: Graficki zavod Hrvatske.

Viereck,L.A., Johnston,W.F. (1990). *Picea mariana* (Mill.) B.S.P. Black spruce. In: Silvics of North America, ed. R.M.Burns, B.H.Honkala. Washington, D.C.: Forest Service, United States Department of Agriculture, 227-237.

Viereck,L.A., Little,E.L., Jr. (1972). Alaska trees and shrubs. Washington, D.C.: Forest Service, U.S. Dept. of Agriculture.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Wang,X.Q., Tank,D.C., Sang,T. (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol.Biol.Evol. *17*, 773-781.

Watano,Y., Imazu,M., Shimizu,T. (1996). Spatial distribution of CpDNA and MtDNA haplotypes in a hybrid zone between *Pinus pumila* and *P. parviflora* var. *pentaphylla* (Pinaceae). Journal of Plant Research *109*, 403-408.

Watts,W.A. (1979). Late Quaternary vegetation of Central Appalachia and the New Jersey coastal plain. Ecological Monographs *49*, 427-469.

Watts,W.A. (1980a). Late-Quaternary vegetation history at White Pond on the inner coastal plain of South Carolina. Quaternary Research *13*, 187-199.

Watts,W.A. (1980b). The Late Quaternary vegetation history of the southeastern United States. Annual Review of Ecology and Systematics *11*, 387-409.

Watts,W.A. (1983). Vegetational history of the eastern United States 25,000 to 10,000 Years Ago. In: Late-Quaternary environments of the United States, ed. S.C.Porter. Minneapolis: University of Minnesota Press, 294-310.

Watts,W.A., Hansen,B.C.S., Grimm,E.C. (1992). Camel Lake: a 40,000-yr record of vegetational and forest history from northwest Florida. Ecology *73*, 1056-1066.

Webb,T., III (1988). Eastern North America. In: Vegetation history, ed. B.Huntley, T.Webb, III. Dordrecht: Kluwer Academic Publishers, 385-414.

Webb,T., III, Anderson,K.H., Bartlein,P.J., Webb,R.S. (1998). Late Quaternary climate change in eastern North America: a comparison of pollen-derived estimates with climate model results. Quaternary Science Reviews *17*, 587-606.

Webb,T., III, Bartlein,P.J., Harrison,S.P., Anderson,K.H. (1993). Vegetation, Lake Levels, and Climate in Eastern North America for the Past 18,000 Years. In: Global climates since the Last Glacial Maximum, ed. H.E.J.Wright, J.E.Kutzbach, T.Webb, III, W.F.Ruddiman, F.A.Street-Perrott, P.J.Bartlein. Minneapolis: University of Minnesota Press, 415-467.

Weir,B.S., Cockerham,C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358-1370.

Whitehead,D.R. (1973). Late-Wisconsin vegetational changes in unglaciated eastern North America. Quaternary Research *3*, 621-631.

Whitehead,D.R. (1981). Late-Pleistocene vegetational changes in northeastern North Carolina. Ecological Monographs *51*, 451-471.

Whittle,C.A., Johnston,M.O. (2002). Male-driven evolution of mitochondrial and chloroplastidial DNA sequences in plants. Molecular Biology and Evolution *19*, 938-949.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Wilkinson,R.C., Hanover,J.W., Wright,J.W., Flake,R.H. (1971). Genetic variation in the monoterpene composition of white spruce. Forest Science *17*, 83-90.

Williams,N. (1995). The trials and tribulations of cracking the prehistoric code. Science *269*, 923-924.

Wright,J.W. (1955). Species crossability in spruce in relation to distribution and taxonomy. For Sci *1*, 319-349.

Wright, W. A., Cox, M., and Campbell, C. S. Internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): structure and sequence divergence. Botany 2001 Abstracts 104. 2001.

Wu,D.Y., Ugozzoli,L., Pal,B.K., Wallace,R.B. (1989). Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proceedings of the National Academy of Sciences USA *86*, 2757-2760.

Yang,H. (1997). Ancient DNA from Pleistocene fossils: preservation, recovery, and utility of ancient genetic information for quaternary research. Quaternary Science Reviews *16*, 1145-1161.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX A

List of abbreviations and useful definitions.

AMOVA	analysis of molecular variance
AMS	Accelerator Mass Spectrometry
A_p	average number of haplotypes per population
A_s	number of haplotypes per species
ASPCR	allele-specific PCR
BMA	BioWhittaker Molecular Applications
bp	base pairs
BSA	bovine serum albumin
CCM1	Community Climate Model
cp	chloroplast
cpDNA	chloroplast DNA
F_{ST}, F_{SC}, F_{CT}	fixation indices
GCM	General Circulation Model
G_{ST}	degree of differentiation
h_k	diversity within each population
Holocene epoch	Geologic time period ~11,000 years ago to the present
h_S	average within-population diversity
H_T	total genetic diversity
I	number of indels per site
indel	insertion/deletion
ITS	internal transcribed spacer
kb	kilobase pairs
KH test	Kishino-Hasegawa test
K_o	number of substitutions per site
K_o/I	ratio of substitutions to indels
LGM	Last Glacial Maximum; ~21,000 calendar years ago
ME	minimum evolution
MP	maximum parsimony
mt	mitochondrial
mtDNA	mitochondrial DNA
nad1 B/C	The intron between exons B and C (i.e. exons 2 and 3) of the mitochondrial *nad1* gene
NCAR	National Center for Atmospheric Research
PAUP	Phylogenetic Analysis Using Parsimony
PCR	polymerase chain reaction
Pleistocene epoch	Geologic time period ~2 million to 10,000 years ago
PTB	*N*-phenacylthiazolium bromide
Quaternary period	Geologic time period of the last two million years
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
SCAR	sequence-characterized amplified region
SNP	single nucleotide polymorphism

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SSCP	single strand conformation polymorphism
SSR	simple sequence repeat
STS	sequence tagged site
UV	ultra-violet
VNTR	variable number tandem repeat
Wisconsin period	The most recent ice age; ~100,000-10,000 years ago

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX B

TrnK intron insertion/deletion data expressed as binary code and positioned at the end of the sequence alignment in the nexus file for phylogenetic analysis in Chapter III.

[	ABCDEFGHI J KL]
Picea abies	1 0 1 1 0 0 1 1 0 1 1 1
Picea asperata	1 0 1 1 0 0 1 1 0 1 1 1
Picea breweriana	1 0 1 1 1 0 1 1 1 1 0 1
Picea chihuahuana	1 0 1 1 0 0 1 1 1 1 1 1
Picea engelmannii	1 0 1 1 0 0 1 1 1 1 1 1
Picea glauca	1 0 1 1 0 0 1 1 1 1 1 1
Picea jezoensis	1 0 1 1 0 0 1 1 0 1 1 1
Picea mariana	1 0 1 1 0 0 1 1 1 1 1 1
Picea mexicana	1 0 1 1 0 0 1 1 1 1 1 1
Picea omorika	1 0 1 1 0 0 1 1 1 1 1 1
Picea pungens	1 0 1 1 0 0 1 1 1 1 1 1
Picea rubens	1 0 1 1 0 0 1 1 1 1 1 1
Picea schrenkiana	1 0 1 1 0 0 1 1 1 1 1 1
Picea sitchensis	0 0 1 1 1 0 1 1 1 1 1 1
Picea smithiana	1 0 1 1 0 0 1 1 1 1 1 1
Picea wilsonii	1 0 1 1 0 0 1 1 0 1 1 1
Pinus armandii	? ? ? 0 1 1 0 0 1 1 1 1
Pinus banksiana	2 0 1 0 1 1 0 0 1 1 ? ?
Pinus thunbergii	2 1 0 0 1 1 0 0 1 0 1 0

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX C

Nadl intron 2 polymorphic sequence data expressed as binary code for phylogenetic analysis in Chapter III. The polymorphisms SNPs A-E and Indels 1-9 were identified in Chapter II, therefore their names were retained here. Positions correspond to location of polymorphisms within the sequence alignment. Mutation types include transitions (transit), transversions (transvers), insertion/deletions (indels; in/del), tandem duplications (dup) or mononucleotide tandem repeats (e.g. C repeat).

Polymorphism	SNP A	Indel 11	SNP F	Indel 12	SNP G	Indel 13	SNP H	SNP I	Indel 8
Position	19	466	467	467-472	522	549-553	656	657	724-734
Mutation type	transvers	indel	transit	dup	transvers	dup	transvers	transvers	C repeat
Code definition	0=A 1=C	0=A 1=del	0=T 1=C	0=in 1=del	0=T 1=G	0=del 1=in	0=T 1=G	0=C 1=A	0=9C 1=10C 2=11C
P. abies	0	0	0	0	0	0	0	0	0
P. asperata	0	0	0	0	0	0	0	0	0
P. breweriana	0	0	0	0	0	1	0	0	1
P. chihuahuana	0	0	0	0	0	0	0	0	0
P. engelmannii	0	0	0	0	0	0	0	0	2
P. glauca	0	0	0	0	0	0	0	0	2
P. jezoensis	0	0	0	0	0	0	0	0	0
P. mariana	1	0	0	0	0	0	0	0	0
P. mexicana	0	0	0	0	0	0	0	0	0
P. omorika	0	0	0	0	0	0	0	0	0
P. pungens	0	0	0	0	1	0	0	0	0
P. rubens	1	0	0	0	0	0	0	0	0
P. schrenkiana	1	0	0	0	0	0	1	1	0
P. sitchensis	0	0	0	0	0	0	0	0	1
P. smithiana	1	1	1	1	0	0	0	0	0
P. wilsonii	0	0	0	0	0	0	0	0	0

Polymorphism	SNP B	SNP K	Indel 2	SNP L	Indel 14	Indel 15	Indel 17	Indel 16	SNP BBB
Position	753	777	782-786	858	869-1131	902-904	932-1033	932-964	969
Mutation type	transvers	transit	5bp indel	transvers	160bp indel	C repeat	102bp indel	33bp indel	transit
Code definition	0=G	0=C	0=in	0=A	0=del	0=2C	0=del	0=del	0=T
	1=T	1=T	1=del	1=C	1=in	1=3C	1=in	1=in	1=C
						in Indel 14	in Indel 14	in Indel 17	in Indel 17
P. abies	0	0	0	0	1	1	1	1	0
P. asperata	0	0	0	1	1	0	0	?	?
P. breweriana	0	0	0	0	0	?	?	?	?
P. chihuahuana	0	0	0	0	0	?	?	?	?
P. engelmannii	0	0	0	0	0	?	?	?	?
P. glauca	0	0	0	0	0	?	?	?	?
P. jezoensis	0	0	0	1	1	0	0	?	?
P. mariana	0	0	0	0	0	?	?	?	?
P. mexicana	0	0	0	0	0	?	?	?	?
P. omorika	0	0	0	0	1	1	1	0	1
P. pungens	0	0	0	0	0	?	?	?	?
P. rubens	1	0	1	0	0	?	?	?	?
P. schrenkiana	0	1	0	0	0	?	?	?	?
P. sitchensis	0	0	0	0	0	?	?	?	?
P. smithiana	0	0	0	0	0	?	?	?	?
P. wilsonii	0	0	0	1	1	0	0	?	?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Polymorphism	Indel 18	SNP M	Indel 19	Indel 20	SNP N	Indel 21	Indel 22	SNP O	SNP P
Position	974-976	1013	1132-1267	1275-1291	1318	1333-2613	1495-1525	1538	1570
Mutation type	G repeat	transvers	34bp dup	indel	transvers	1281bp indel	31bp dup	transvers	transvers
Code definition	0=2G 1=3G in Indel 17	0=T 1=G in Indel 17	0=del 1=in 4X	0=del 1=in	0=C 1=A	0=del 1=in	0=del 1=in in Indel 21	0=G 1=T in Indel 21	0=A 1=C in Indel 21
P. abies	1	0	0	0	0	0	?	?	?
P. asperata	?	?	0	0	0	0	?	?	?
P. breweriana	?	?	0	0	0	1	0	0	1
P. chihuahuana	?	?	0	0	0	1	0	1	0
P. engelmannii	?	?	0	0	0	1	0	0	0
P. glauca	?	?	0	0	0	1	0	0	0
P. jezoensis	?	?	0	0	0	0	?	?	?
P. mariana	?	?	0	0	0	1	0	0	0
P. mexicana	?	?	0	0	0	1	0	1	0
P. omorika	0	1	1	0	0	0	?	?	?
P. pungens	?	?	0	0	0	1	0	0	0
P. rubens	?	?	0	0	0	1	0	0	0
P. schrenkiana	?	?	0	1	1	1	0	0	0
P. sitchensis	?	?	0	0	0	1	1	0	0
P. smithiana	?	?	0	0	0	1	0	0	1
P. wilsonii	?	?	0	0	0	0	?	?	?

Polymorphism	SNP Q	SNP R	SNP S	Indel 9	SNP T	SNP U	SNP V	SNP C	Indel 23
Position	1613	1614	1615	1632-1634	1637	1638	1738	1786	1803-2057
Mutation type	transvers	transvers	transvers	A repeat	transvers	transvers	transvers	transvers	255bp indel
Code definition	0=A 1=T in Indel 21	0=G 1=T in Indel 21	0=C 1=A in Indel 21	0=2A 1=3A in Indel 21	0=T 1=G in Indel 21	0=C 1=A in Indel 21	0=A 1=C in Indel 21	0=G 1=T in Indel 21	0=del 1=in in Indel 21
P. abies	?	?	?	?	?	?	?	?	?
P. asperata	?	?	?	?	?	?	?	?	?
P. breweriana	0	0	0	1	1	1	1	0	1
P. chihuahuana	0	0	0	0	0	0	0	1	1
P. engelmannii	0	0	0	0	0	0	0	0	1
P. glauca	0	0	0	0	0	0	0	0	1
P. jezoensis	?	?	?	?	?	?	?	?	?
P. mariana	0	0	0	1	0	0	0	1	1
P. mexicana	0	0	0	0	0	0	0	1	1
P. omorika	?	?	?	?	?	?	?	?	?
P. pungens	0	0	0	0	0	0	0	1	1
P. rubens	0	0	0	1	0	0	0	1	1
P. schrenkiana	0	0	0	1	0	0	0	1	1
P. sitchensis	0	0	0	0	0	0	0	0	1
P. smithiana	1	1	1	1	0	0	0	1	0
P. wilsonii	?	?	?	?	?	?	?	?	?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Polymorphism	Indel 24	Indel 7	Indel 3	Indel 5	Indel 4	Indel 25	SNP W	SNP X	Indel 26
Position	1865	1865-1872	1900-1905	2126-2130	2201-2220	2299-2307	2346	2347	2348-2460
Mutation type	indel	6bp dup	6bp dup	5bp indel	20bp dup	9bp dup	transvers	transvers	113bp indel
Code definition	0=del	0=del	0=del	0=del	0=del	0=del	0=T	0=C	0=del
	1=A	1=in	1=in	1=in	1=in	1=in	1=G	1=A	1=in
	in Indel 23	in Indel 23	in Indel 23	in Indel 21	in Indel 21	in Indel 21	in Indel 21	in Indel 21	in Indel 21
P. abies	?	?	?	?	?	?	?	?	?
P. asperata	?	?	?	?	?	?	?	?	?
P. breweriana	1	0	0	1	0	0	1	1	0
P. chihuahuana	1	0	0	1	0	0	0	0	1
P. engelmannii	1	0	0	0	1	0	0	0	1
P. glauca	1	0	0	0	1	0	0	0	1
P. jezoensis	?	?	?	?	?	?	?	?	?
P. mariana	1	1	0	1	0	0	0	0	1
P. mexicana	1	0	0	1	0	0	0	0	1
P. omorika	?	?	?	?	?	?	?	?	?
P. pungens	1	0	0	1	0	0	0	0	1
P. rubens	1	0	1	1	0	0	0	0	1
P. schrenkiana	0	1	0	1	0	0	0	0	1
P. sitchensis	1	0	0	0	0	0	0	0	1
P. smithiana	?	?	?	1	0	1	0	0	1
P. wilsonii	?	?	?	?	?	?	?	?	?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Polymorphism	SNP Y	Indel 27	SNP D	SNP E	SNP Z	Indel 6	SNP AA	SNP BB	SNP CC
Position	2361	2407-2412	2447	2492	2506	2508-2513	2532	2544	2546
Mutation type	transit	6bp indel	transit	transvers	transvers	6bp dup	transvers	transvers	transvers
Code definition	0=A 1=G in Indel 26	0=del 1=in in Indel 26	0=T 1=C in Indel 26	0=C 1=A in Indel 21	0=A 1=C in Indel 21	0=del 1=in in Indel 21	0=C 1=A in Indel 21	0=T 1=A in Indel 21	0=A 1=C in Indel 21
P. abies	?	?	?	?	?	?	?	?	?
P. asperata	?	?	?	?	?	?	?	?	?
P. breweriana	?	?	?	0	0	0	0	0	0
P. chihuahuana	0	1	0	0	0	0	0	0	0
P. engelmannii	0	1	0	1	0	0	0	0	0
P. glauca	0	1	0	1	0	0	0	0	0
P. jezoensis	?	?	?	?	?	?	?	?	?
P. mariana	0	1	1	0	0	1	0	0	0
P. mexicana	0	1	0	0	0	0	0	0	0
P. omorika	?	?	?	?	?	?	?	?	?
P. pungens	0	1	0	0	0	0	0	0	0
P. rubens	0	1	1	0	0	1	0	0	0
P. schrenkiana	1	0	0	0	0	0	0	0	0
P. sitchensis	0	1	0	1	0	0	0	0	0
P. smithiana	1	0	0	0	1	0	1	1	1
P. wilsonii	?	?	?	?	?	?	?	?	?

Polymorphism	Indel 28	SNP DD	SNP EE	Indel 29	SNP FF	Indel 30	Indel 32	SNP LL	Indel 33
Position	2549	2567	2599	2628-2632	2641	2691-2694	2842-2846	2875	2910-2938
Mutation type	indel	transvers	transvers	5bp indel	transvers	4bp indel	5bp indel	transvers	29bp indel
Code definition	0=del 1=T in Indel 21	0=G 1=T in Indel 21	0=A 1=C in Indel 21	0=del 1=in	0=C 1=A	0=del 1=in	0=del 1=in	0=C 1=A	0=del 1=in
P. abies	?	?	?	1	0	1	1	1	0
P. asperata	?	?	?	1	0	1	1	1	0
P. breweriana	1	1	0	1	0	1	0	0	0
P. chihuahuana	1	0	0	1	0	1	1	0	0
P. engelmannii	1	0	0	1	0	1	1	0	0
P. glauca	1	0	0	1	0	1	1	0	0
P. jezoensis	?	?	?	1	0	1	1	1	0
P. mariana	1	0	0	1	0	1	1	0	0
P. mexicana	1	0	0	1	0	1	1	0	0
P. omorika	?	?	?	1	0	1	1	1	0
P. pungens	1	0	0	1	0	1	1	0	0
P. rubens	1	0	0	1	0	1	1	0	0
P. schrenkiana	1	0	0	1	1	0	1	0	0
P. sitchensis	1	0	0	1	0	1	1	0	0
P. smithiana	0	0	1	0	0	0	1	0	1
P. wilsonii	?	?	?	1	0	1	1	1	0

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Polymorphism	SNP MM	Indel 34	SNP NN	SNP OO	SNP PP	SNP QQ	Indel 35	Indel 36	SNP YY
Position	2943	2951-2956	2965	2569	2570	3013	3088-3093	3176-3186	3472
Mutation type	transvers	6bp dup	transvers	transvers	transvers	transvers	6bp dup	11bp dup	transvers
Code definition	0=G 1=T	0=del 1=in	0=C 1=A	0=T 1=G	0=C 1=A	0=G 1=T	0=del 1=in	0=del 1=in	0=A 1=C
P. abies	1	0	0	0	0	0	0	0	0
P. asperata	0	0	1	1	1	0	0	0	0
P. breweriana	1	0	0	0	0	0	0	0	1
P. chihuahuana	0	0	0	0	0	0	0	0	0
P. engelmannii	0	0	0	0	0	0	0	0	0
P. glauca	0	0	0	0	0	0	0	0	0
P. jezoensis	0	0	1	1	1	0	0	0	0
P. mariana	0	0	0	0	0	0	0	0	0
P. mexicana	0	0	0	0	0	0	0	0	0
P. omorika	0	0	0	0	0	0	1	0	0
P. pungens	0	0	0	0	0	0	0	0	0
P. rubens	0	0	0	0	0	0	0	0	0
P. schrenkiana	1	1	0	0	0	1	0	1	1
P. sitchensis	1	0	0	0	0	0	0	0	0
P. smithiana	0	1	0	0	0	0	0	0	1
P. wilsonii	0	0	1	1	1	0	0	0	0

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX D

Pairwise distances between taxa as generated in PAUP.

Chloroplast *trnK* intron

1 *Picea abies*
2 *Picea asperata*
3 *Picea breweriana*
4 *Picea chihuahuana*
5 *Picea engelmannii*
6 *Picea glauca*
7 *Picea jezoensis*
8 *Picea mariana*
9 *Picea mexicana*
10 *Picea omorika*
11 *Picea pungens*
12 *Picea rubens*
13 *Picea schrenkiana*
14 *Picea sitchensis*
15 *Picea smithiana*
16 *Picea wilsonii*
17 *Pinus armandii*
18 *Pinus banksiana*
19 *Pinus thunbergii*

Pairwise distances between taxa: chloroplast *trnK* intron

Below diagonal: Total character differences
Above diagonal: Mean character differences (adjusted for missing data)

	1	2	3	4	5	6	7	8
1	-	0.00000	0.00844	0.00361	0.00281	0.00241	0.00040	0.00401
2	0	-	0.00844	0.00361	0.00281	0.00241	0.00040	0.00401
3	21	21	-	0.00641	0.00802	0.00762	0.00885	0.00681
4	9	9	16	-	0.00320	0.00280	0.00401	0.00280
5	7	7	20	8	-	0.00040	0.00321	0.00440
6	6	6	19	7	1	-	0.00281	0.00400
7	1	1	22	10	8	7	-	0.00441
8	10	10	17	7	11	10	11	-
9	6	6	19	7	1	0	7	10
10	9	9	16	6	10	9	10	1
11	8	8	15	3	7	6	9	6
12	9	9	16	6	10	9	10	1
13	11	11	18	6	10	9	12	9
14	19	19	10	14	18	17	20	15
15	11	11	18	6	10	9	12	7
16	1	1	22	10	8	7	0	11
17	104	104	110	104	107	106	104	106
18	127	127	125	127	131	130	127	125
19	133	133	133	133	136	135	133	133

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pairwise distances between taxa (continued): chloroplast *trnK* intron

Below diagonal: Total character differences
Above diagonal: Mean character differences (adjusted for missing data)

	9	10	11	12	13	14	15	16
1	0.00241	0.00361	0.00321	0.00361	0.00441	0.00762	0.00441	0.00040
2	0.00241	0.00361	0.00321	0.00361	0.00441	0.00762	0.00441	0.00040
3	0.00762	0.00641	0.00601	0.00641	0.00721	0.00400	0.00721	0.00885
4	0.00280	0.00240	0.00120	0.00240	0.00240	0.00560	0.00240	0.00401
5	0.00040	0.00400	0.00280	0.00400	0.00400	0.00720	0.00400	0.00321
6	0.00000	0.00360	0.00240	0.00360	0.00360	0.00680	0.00360	0.00281
7	0.00281	0.00401	0.00361	0.00401	0.00481	0.00803	0.00481	0.00000
8	0.00400	0.00040	0.00240	0.00040	0.00360	0.00600	0.00280	0.00441
9	-	0.00360	0.00240	0.00360	0.00360	0.00680	0.00360	0.00281
10	9	-	0.00200	0.00000	0.00320	0.00560	0.00240	0.00401
11	6	5	-	0.00200	0.00200	0.00520	0.00200	0.00361
12	9	0	5	-	0.00320	0.00560	0.00240	0.00401
13	9	8	5	8	-	0.00560	0.00160	0.00481
14	17	14	13	14	14	-	0.00560	0.00803
15	9	6	5	6	4	14	-	0.00481
16	7	10	9	10	12	20	12	-
17	106	105	104	105	105	107	105	104
18	130	124	126	124	129	123	129	127
19	135	132	130	132	135	130	135	133

Pairwise distances between taxa (continued): chloroplast *trnK* intron

Below diagonal: Total character differences
Above diagonal: Mean character differences (adjusted for missing data)

	17	18	19
1	0.05103	0.05235	0.05374
2	0.05103	0.05235	0.05374
3	0.05376	0.05123	0.05356
4	0.05083	0.05218	0.05356
5	0.05230	0.05382	0.05477
6	0.05181	0.05341	0.05437
7	0.05103	0.05235	0.05374
8	0.05181	0.05136	0.05356
9	0.05181	0.05341	0.05437
10	0.05132	0.05094	0.05316
11	0.05083	0.05177	0.05236
12	0.05132	0.05094	0.05316
13	0.05132	0.05300	0.05437
14	0.05214	0.05043	0.05225
15	0.05132	0.05300	0.05437
16	0.05103	0.05235	0.05374
17	-	0.03808	0.04057
18	76	-	0.01397
19	83	34	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Mitochondrial *nad1* intron 2

1 *Picea abies*
2 *Picea asperata*
3 *Picea breweriana*
4 *Picea chihuahuana*
5 *Picea engelmannii*
6 *Picea glauca*
7 *Picea jezoensis*
8 *Picea mariana*
9 *Picea mexicana*
10 *Picea omorika*
11 *Picea pungens*
12 *Picea rubens*
13 *Picea schrenkiana*
14 *Picea sitchensis*
15 *Picea smithiana*
16 *Picea wilsonii*

Pairwise distances between taxa: mitochondrial *nad1* intron 2

Below diagonal: Total character differences
Above diagonal: Mean character differences (adjusted for missing data)

	1	2	3	4	5	6	7	8
1	-	0.20000	0.21212	0.12121	0.15152	0.15152	0.20000	0.15152
2	7	-	0.36364	0.21212	0.24242	0.24242	0.00000	0.24242
3	7	12	-	0.25397	0.26984	0.26984	0.36364	0.26984
4	4	7	16	-	0.09091	0.09091	0.21212	0.09091
5	5	8	17	6	-	0.00000	0.24242	0.15152
6	5	8	17	6	0	-	0.24242	0.15152
7	7	0	12	7	8	8	-	0.24242
8	5	8	17	6	10	10	8	-
9	4	7	16	0	6	6	7	6
10	7	8	10	5	6	6	8	6
11	5	8	16	2	6	6	8	6
12	7	10	19	8	12	12	10	4
13	15	20	25	19	23	23	20	17
14	4	9	15	7	4	4	9	11
15	13	16	31	25	29	29	16	24
16	7	0	12	7	8	8	0	8

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pairwise distances between taxa (continued): mitochondrial *nad1* intron 2

Below diagonal: Total character differences
Above diagonal: Mean character differences (adjusted for missing data)

	9	10	11	12	13	14	15	16
1	0.12121	0.17949	0.15152	0.21212	0.45455	0.12121	0.39394	0.20000
2	0.21212	0.22857	0.24242	0.30303	0.60606	0.27273	0.48485	0.00000
3	0.25397	0.30303	0.25397	0.30159	0.39683	0.23810	0.51667	0.36364
4	0.00000	0.15152	0.03030	0.12121	0.28788	0.10606	0.39683	0.21212
5	0.09091	0.18182	0.09091	0.18182	0.34848	0.06061	0.46032	0.24242
6	0.09091	0.18182	0.09091	0.18182	0.34848	0.06061	0.46032	0.24242
7	0.21212	0.22857	0.24242	0.30303	0.60606	0.27273	0.48485	0.00000
8	0.09091	0.18182	0.09091	0.06061	0.25758	0.16667	0.38095	0.24242
9	-	0.15152	0.03030	0.12121	0.28788	0.10606	0.39683	0.21212
10	5	-	0.18182	0.24242	0.54545	0.21212	0.42424	0.22857
11	2	6	-	0.12121	0.28788	0.10606	0.39683	0.24242
12	8	8	8	-	0.31818	0.19697	0.41270	0.30303
13	19	18	19	21	-	0.33333	0.41270	0.60606
14	7	7	7	13	22	-	0.47619	0.27273
15	25	14	25	26	26	30	-	0.48485
16	7	8	8	10	20	9	16	-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.