AN ABSTRACT OF THE DISSERTATION OF

Ann M. Willyard for the degree of Doctor of Philosophy in Botany and Plant Pathology, presented on June 7, 2007.

Title: New Perspectives on Evolutionary Relationships within *Pinus* (Pinaceae) and Within Subsection *Ponderosae* (Subgenus *Pinus*).

Abstract approved:

Aaron I. Liston Richard C. Cronn

A fossil-calibrated phylogenetic framework based on exemplars from each of the four taxonomic sections within *Pinus* was created using multiple nuclear and chloroplast loci. Calibration at the well-defined subgeneric split within *Pinus* with either fossil leaves and cones (ca. 45 million years ago) or fossil wood (ca. 85 million years ago) provides a reasonable starting point. Despite a wide difference in rates between loci, this work infers a moderate tempo of mutation rate in *Pinus,* and concludes that the within-locus rate variation and the leaves vs. wood question are both minor factors in comparison to the enormous effects of incorrect fossil/node association.

Next, I used full taxonomic sampling with multiple individuals per species and two nuclear loci to create independent phylogenies for 18 species in subsection *Ponderosae* (Section *Trifoliae*, subgenus *Pinus*). I estimate that some species have enormous effective population sizes ($> 10^5$ individuals). This factor and an inferred origin 15 million years ago contribute to the lack of species-level coalescence and the incongruence between *Ponderosae* gene trees. *Pinus jeffreyi* is allied with *Sabinianae*, as proposed from terpene biochemistry. I observed one instance where reticulate evolution is likely. *Pinus coulteri* accessions resolve with *Sabinianae* at one locus, but in a monophyletic clade sister to other *Ponderosae* at the other locus, suggesting that this species originated as a diploid hybrid. Similar incongruence between loci for accessions of the *P. ponderosa*/*P. washoensis* complex could also be the result of introgression.

Using this phylogenetic framework, I focused on three sympatric *Ponderosae* (*P. washoensis*, *P. ponderosa*, and *P. jeffreyi*) to evaluate the genetic distinctiveness and specific status of *P. washoensis*. Population-level sampling with faster-evolving nuclear

microsatellite loci reveals a clear divergence between *P. jeffreyi* and the *P. ponderosa/P. washoensis* complex. Using allele frequencies, I observed a weak cluster of traditional *P. washoensis* with high-elevation putative *P. washoensis* populations from Oregon, but other methods show no differentiation between *P. ponderosa* and *P. washoensis.* Nuclear admixture and chloroplast haplotype analysis suggest a low level of introgression between *P. jeffreyi* and *P. ponderosa/P. washoensis*, providing an important source of migrants into the genetic milieu of these species.

New Perspectives on Evolutionary Relationships within *Pinus* (Pinaceae) and within Subsection *Ponderosae* (Subgenus *Pinus*)

by
Ann M. Willyard

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 7, 2007

Commencement June 2008

Doctor of Philosophy dissertation of Ann M. Willyard presented on June 7, 2007

APPROVED:

Co-Major Professor, representing Botany and Plant Pathology

Co-Major Professor, representing Botany and Plant Pathology

Chair of the Department of Botany and Plant Pathology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Ann M. Willyard, Author

ACKNOWLEDGEMENTS

I am forever indebted to a long list of people who contributed to this research and to my growth as a scientist. They are gratefully acknowledged at the end of each manuscript. A special thanks to my enthusiastic and exceedingly supportive advisers Richard Cronn and Aaron Liston, and to my committee, Steven Strauss, Joseph Spatafora, and Paul Doescher. This thesis is lovingly devoted to my husband Gene.

CONTRIBUTION OF AUTHORS

Drs. Aaron Liston and Richard Cronn were intensively involved in the design of all three experiments, in the analysis of the results, and in the editing of each of the three manuscripts. Dr. John Syring was involved in collecting sequencing data and editing the first manuscript. Dr. David Germandt was involved in collecting sequencing data and editing the first manuscript, and sharing unpublished results and advice that greatly assisted the progress of the second manuscript.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

New Perspectives on Evolutionary Relationships within *Pinus* (Pinaceae)
and Within Subsection *Ponderosae* (Subgenus *Pinus*)

GENERAL INTRODUCTION

Dissertation Overview. This dissertation examines the evolutionary relationships among *Pinus* species from several angles. A classification of the genus (largely following (Gernandt et al. 2005)) is summarized in Table 1.1. In the broadest view, chapter two presents a time-calibrated phylogenetic framework using an exemplar species from each of the four taxonomic sections within the genus (Willyard et al. 2007). In this chapter, entitled "*Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for Pinus*", I integrate a careful review of fossil literature with multiple nuclear and chloroplast locus sequences to estimate mutation rates and divergence dates for major lineages within the genus. The third chapter, entitled "*Phylogenetic inference for reticulate evolution in subsection* Ponderosae *(Pinaceae;* Pinus; *subgenus* Pinus, *Section* Trifoliae*)*", is a phylogenetic analysis focusing on the most species-rich subsection within subgenus *Pinus,* namely subsect. *Ponderosae*. Using full taxonomic sampling with multiple individuals per species and two independent nuclear loci, I examine evidence for phylogenetic signal in these closely related species, and try to discern reticulate evolution in my phylogenetic trees. I build on the inferred ages for lineages and the mutation rates presented in the second chapter to compare patterns of species-level monophyly in *Ponderosae* with those reported in subgenus *Strobus* (Syring et al. 2007). In the fourth chapter, I further narrow the focus to evaluate the genetic uniqueness and specific status of the enigmatic species *P. washoensis*. This chapter, entitled "*The affinities of* Pinus washoensis *(Pinaceae) to* P. ponderosa *and* P. jeffreyi *based on nuclear microsatellites*", couples population-level sampling with faster-evolving nuclear microsatellite loci to look for species-level resolution between *P. washoensis, P. ponderosa* and *P. jeffreyi* that could not be revealed by phylogenetic analyses of nucleotide polymorphisms in low-copy nuclear genes. A variety of statistical methods that use genetic distances among individuals as well as allele frequencies among populations were used to look for evidence of ancient or current introgression. This fourth chapter uses the phylogenetic framework that is constructed in the third chapter to place these three species in the context of the closely related taxa in *Ponderosae*.

The research for this dissertation builds upon previous work by another member of our research group that established the utility of low-copy nuclear loci for *Pinus* phylogeny (Syring et al. 2005). Application of low-copy nuclear sampling of the taxa in subg. *Pinus* was guided by those results. Within subg. *Strobus*, a species-level phylogeny revealed widespread nonmonophyly which was attributed to shared ancestral polymorphism (Syring et al. 2007). That research suggested that the enormous effective population sizes estimated for many pine species will delay coalescence for an 'average' locus beyond the age inferred for their taxonomic subsection, let alone the projected divergence between species pairs (Syring et al. 2007; Willyard et al. 2007). Given that cladogenesis within subsect. *Ponderosae* is inferred to be even more recent than for many of the groups that exhibited nonmonophyly in subg. *Strobus* (Syring et al. 2007), nonmonophyly was expected in the *Ponderosae* as well. An overall sharing of ancestral polymorphism was, indeed, observed. Unlike results from subg. *Strobus*, I also found evidence that reticulate evolution may have played a substantial role in the evolution of *Ponderosae* species. In chapter three of this dissertation, I will argue that ancient reticulation is the simplest explanation for the phylogenetic patterns that I observe for *P. coulteri* and perhaps for the *P. ponderosa/P. washoensis* species complex. In the fourth chapter, I will report evidence for a clear divergence, yet examples of recent admixture, between *P. jeffreyi* and the *P. ponderosa/P. washoensis* complex. I will suggest that these occasional migrants into the gene pool are likely to contribute to the enormous level of within-population variability that is observed in most *Pinus* species.

The Genus *Pinus*. *Pinus* is the most species-rich (perhaps 116 extant species; Table 1.1) and widespread genus in Pinaceae. It is recognized as a monophyletic lineage within the Pinaceae based on leaves clustered in fascicles with a sheath of bud scales at the base (leading to the interpretation of the fascicles as short shoots) as well as the distinctive woody portions of the ovulate cone left exposed in the first and second year (umbo and apophysis) (Price, Liston, and Strauss 1998). A clear differentiation of the genus is also supported by molecular evidence (Wang, Tank, and Sang 2000). Following Gernandt et al. (2005), intrageneric classification (Table 1.1) resolves two subgenera, subg. *Pinus* and subg. *Strobus* (Gernandt et al. 2005; Syring et al. 2005), supported morphologically by the number of fibrovascular bundles per needle (Gernandt et al. 2005). Molecular evidence shows a large

genetic distance between the subgenera (Gernandt et al. 2005; Syring et al. 2005; Willyard et al. 2007) but divergence between many species is limited (Gernandt et al. 2005).

In spite of the ecological and economic importance of pines, our knowledge of the genus is – in the words of Mirov (1967) - still "rather uneven". This assessment seems as true today as it was when Mirov wrote those words 40 years ago, despite a steady stream of research into *Pinus* biochemistry, physiology, genetics, and phylogeny that continues unabated. Recent *Pinus* research topics have covered a wide range of topics, from a karyotype map (Islam-Faridi, Nelson, and Kubisiak 2007) to association mapping of complex traits (Neale and Savolainen 2004) and candidate genes (Gonzalez-Martinez et al. 2006). The coevolution of pines and the many species with which they interact is of special interest. For example, investigations have been made into interactions between *Pinus* species, *Dendroctonus* beetles, and their fungal associates (Six and Paine 1999); genetic interactions with *Cronartium* pathogens (Vogler 2000; Yu et al. 2002), and a recently-described coevolutionary arms race that appears to be driving speciation of the South Hills crossbills that feed on *P. contorta* var. *latifolia* (Smith and Benkman 2007).

Despite these focused efforts, species-level relationships in *Pinus* remain mostly unresolved, frustrating attempts to draw correlations between pine speciation and that of dependent organisms and hampering the study of adaptive traits. In order to test trait correlations, supertrees have been built for *Pinus* species by combining a number of different published phylogenetic trees (e.g. Schwilk and Ackerly 2001; Grotkopp et al. 2004). In addition to suffering from the inherent supertree problem that all input trees and their features are treated equally (Felsenstein 2004), the *Pinus* supertrees are still not fully resolved. This lack of species-level resolution may seem surprising when one considers their long history of taxonomic attention spanning more than a century (Price, Liston, and Strauss 1998). Since that review, various levels of species sampling within *Pinus* have been used for numerous phylogenetic studies, for example: Liston et al. 1999; Wang et al. 1999; Wu, Krutovskii, and Strauss 1999; Dvorak et al. 2000; Wang, Szmidt, and Nguyen 2000; Wolff et al. 2000; Boratynska and Bobowicz 2001; Gernandt, Liston, and Pinero 2001; Zitzetmuller and 5 2001; Hizume, Matsusaki, and Garajova 2002; Nkongolo, Michael, and Gratton 2002; Faldt et al. 2004; Gerson and Kelsey 2004; Grotkopp et al. 2004; Gernandt et al. 2005; Syring et al. 2005; Eckert and Hall 2006; Mendoza and Rodriguez-Banderas 2006; Marcysiak and Boratynski 2007; Syring et al. 2007. These studies have settled relationships

for the subgenera and sections, although questions remain as to whether traditionally recognized subsections are monophyletic (Table 1.1; Gernandt et al. 2005). For some lower level taxa, e.g. species and species 'complexes', phylogenetic relationships remain largely unanswered questions.

Some aspects of *Pinus* biology make these species desirable and tractable research targets. For example, all species are diploid, yet haploid megagametophytes can be used for genetic analysis. *Pinus* fossils are abundant, unlike many herbaceous plant genera. The economic importance of pines has inspired a wealth of information regarding phenology, environmental interactions, and population dynamics. Compared with other non-model species, rich molecular tools have been developed, including biochemical profiles and expressed sequence libraries that have been used as the basis for genetic linkage mapping.

Nonetheless, there are a number of factors affecting pines that make them challenging subjects in which to study evolutionary relationships. The second chapter of this dissertation describes the complexities involved with attributing *Pinus* fossils to evolutionary lineages. The plastic nature of *Pinus* morphological character traits (Gernandt et al. 2005) contributes immensely to the confusion regarding both extinct and extant taxa. The long generation time has severely limited the characters that have been measured for pine species in common gardens. Of necessity, these studies often rely on *Pinus* seedling traits; a few studies have measured growth over several decades, but experiments that compare sexually mature trees are extremely rare. For example, a hybridization experiment that extends past the creation of viable seed into a second generation will greatly exceed a researcher's career span, e.g. Critchfield (1986). Molecular evidence indicates that states for most characters are homoplasious even between the subgenera, despite their wide divergence (Gernandt et al. 2005).

Subsection *Ponderosae*. Subgenus *Pinus* (the diploxylon or "hard" pines) contain lower levels of genetic variation than subg. *Strobus* (Geada Lopez, Kamiya, and Harada 2002; Willyard et al. 2007) but are clearly reconciled into two sections in phylogenies based on chloroplast sequences (Geada Lopez, Kamiya, and Harada 2002; Gernandt et al. 2005). No morphological character state has been found that consistently distinguishes sect. *Pinus* from sect. *Trifoliae* (Gernandt et al. 2005). Section *Pinus* encompasses about 26 species that are mainly Eurasian and Mediterranean but include *P. resinosa* and *P. tropicalis* in Eastern

North America and western Cuba, respectively. Section *Trifoliae* (Gernandt et al. 2005) corresponds to the "New World Diploxylon Pine" group of Price et al. (1998), containing the remaining 50 North American species from subg. *Pinus*. Within sect. *Trifoliae*, subsect. *Ponderosae* (including the three species traditionally placed in subsect. *Sabinianae*) forms a monophyletic group based on chloroplast sequences (Gernandt et al. 2005; Eckert and Hall 2006). About 21 species from Mexico, Central America, and the western United States are included in subsect. *Ponderosae*. These species display relatively low levels of cpDNA divergence from each other, resulting in many terminal polytomies.

The *Ponderosae* are an intriguing group for several reasons. The taxa include one of the widest ranging and most economically important pine species in western North America (*P. ponderosa*), as well as one of the narrowest endemics in North America (*P. washoensis*). Several groupings have been proposed within the *Ponderosae*, e.g. the 'Montezumae' and 'Pseudostrobus' groups (Martinez 1948) and subsect. *Sabinianae* (Little and Critchfield 1969), but cpDNA phylogenies (Gernandt et al. 2005; Eckert and Hall 2006) provide no resolution with which to test the distinctiveness of these or alternative groups. Investigators have studied several species complexes (e.g. Niebling and Conkle 1990; Matos 1995; Rehfeldt 1999a; Matos and Schaal 2000), but to my knowledge this is the first research that includes full taxonomic sampling of the *Ponderosae* with multiple accessions per species, and the first to use independent loci from the nuclear genome. Since the phylogeny of this species rich group is surrounded by considerable uncertainty, it is the focus of the third chapter of this dissertation.

Pinus washoensis. The distinctiveness of *P. washoensis* and its evolutionary relationship to *P. ponderosa* and *P. jeffreyi* remain open and interesting questions. Many morphological features are within the range of variability of *P. ponderosa*, and especially fit within the character states reported for the North Plateau race of *var. ponderosa* and/or for var. *scopulorum*. Thus, even though *P. washoensis* resembles *P. jeffreyi* in its high-altitude habitat, higher cone scale phyllotaxy, and low seed wing to seed body ratio, it is likely to be more closely related (or even conspecific with) *P. ponderosa* (Niebling and Conkle 1990; Lauria 1996; Brayshaw 1997; Lauria 1997). Some have suggested that *P. washoensis* merely represents outlying populations of the North Plateau race (Niebling and Conkle 1990; Lauria 1997; Rehfeldt 1999b). However, as presented in the fourth chapter of this dissertation, not

all of the evidence that has been presented is clear-cut, and some studies have resolved a sample of *P. washoensis* more closely related to other *Ponderosae* than *P. ponderosa* (Prager, Fowler, and Wilson 1976; Eckert and Hall 2006). Since the genetic uniqueness and specific status of *P. washoensis* are still unclear, it is the focus of the fourth chapter of this dissertation.

Table 1.1. Intrageneric classification of *Pinus*

subgenus	section	subsection	approximate number of species
Strobus	*Parrya*	*Balfourianae*	3
		Cembroides	11
		Nelsoniae	1
	Quinquefoliae	*Gerardianae*	3
		Krempfianae	1
		Strobus	21
	total		40
Pinus	*Pinus*	*Pinaster*	7
		Pinus	19
	Trifoliae	*Australes*	25
		Contortae	4
		Ponderosae	21
	total		76
		total	116

LITERATURE CITED

Boratynska, K., and M. Bobowicz. 2001. *Pinus uncinata* Ramond taxonomy based on needle characters. Plant Syst Evol **227**:183-194.

Brayshaw, T. C. 1997. Washoe and ponderosa pines on Promontory Hill near Merritt, B.C., Canada. Ann Naturhist Mus Wien **99B**:673-680.

Critchfield, W. 1986. Hybridization and classification of the white pines (*Pinus* section *Strobus*). Taxon **35**:647-656.

Dvorak, W., A. Jordan, G. Hodge, and J. Romero. 2000. Assessing evolutionary relationships of pines in the *Oocarpae* and *Australes* subsections using RAPD markers. New Forests **20**:163-192.

Eckert, A. J., and B. D. Hall. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenetics and Evol **40**:166-182.

Faldt, J., K. Sjodin, M. Persson, I. Valterova, and A.-K. Borg-Karlson. 2004. Correlations between selected monoterpene hydrocarbons in the xylem of six *Pinus* (Pinaceae) species. Chemoecology **11**:97-106.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, MA.

Geada Lopez, G., K. Kamiya, and K. Harada. 2002. Phylogenetic relationships of Diploxylon pines (subgenus *Pinus*) based on plastid sequence data. Int J Plant Sci **163**:737-747.

Gernandt, D., G. Geada Lopez, S. Garcia, and A. Liston. 2005. Phylogeny and classification of *Pinus*. Taxon **54**:29-42.

Gernandt, D., A. Liston, and D. Pinero. 2001. Variation in the nrDNA *ITS* of *Pinus* Subsection *Cembroides*: implications for molecular systematics studies of pine species complexes. Mol Phylogenetics and Evol **21**:450-468.

Gerson, E. A., and R. G. Kelsey. 2004. Piperidine alkaloids in North American *Pinus* taxa: implications for chemosystematics. Biochemical Systematics and Ecology **32**:63-74.

Gonzalez-Martinez, S., E. Ersoz, G. R. Brown, N. C. Wheeler, and D. B. Neale. 2006. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics **172**:1915-1926.

Grotkopp, E., M. Rejmanek, M. J. Sanderson, and T. L. Rost. 2004. Evolution of genome sze in pines (*Pinus*) and its life-history correlates: supertree analyses. Evolution **58**.

Hizume, M., Shibata, F., Y. Matsusaki, and Z. Garajova. 2002. Chromosome identification and comparative karyotypic analyses of four *Pinus* species. Theor Appl Genet **105**:491-497.

Islam-Faridi, M. N., C. D. Nelson, and T. L. Kubisiak. 2007. Reference karyotype and cytomolecular map for loblolly pine (*Pinus taeda* L.). Genome **50**:241-251.

Lauria, F. 1997. The taxonomic status of *Pinus washoensis* H. Mason & Stockw. (Pinaceae). Ann Naturhist Mus Wien **99B**:655-671.

Lauria, F. 1996. The identity of *Pinus ponderosa* Douglas ex. C. Lawson (Pinaceae). Linzer biologische Beitrag **28/2**:999-1052.

Liston, A., W. Robinson, D. Pinero, and E. Alvarez-Buylla. 1999. Phylogentics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenetics and Evol **11**:95-109.

Little, E., and W. Critchfield. 1969. Subdivisions of the Genus *Pinus* (Pines). U.S.D.A. Forest Service **Misc Pub 1144**.

Marcysiak, K., and A. Boratynski. 2007. Contribution to the taxonomy of *Pinus uncinata* (Pinaceae) based on cone characters. Plant Syst Evol **264**:57-73.

Martinez, M. 1948. Los Pinos Mexicanos, Segunda Edicion, Mexico.

Matos, J., and B. Schaal. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. Evolution **54**:1218-1233.

Matos, J. A. 1995. *Pinus hartwegii* and *P. rudis*: A critical assessment. Syst Biol **20**:6-21.

Mendoza, C. F. V., and A. Rodriguez-Banderas. 2006. Evolutionary analysis of *Pinus leiophylla*: a study using an Intron II sequence fragment of mitochondrial *nad1*. Can J Bot **84**:172-177.

Mirov, N. T. 1967. The genus *Pinus*. The Ronald Press Company, New York, NY.

Neale, D. B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. Trends in Plant Science **9**:325-330.

Niebling, C., and M. Conkle. 1990. Diversity of Washoe pine and comparisons with allozymes of ponderosa pine races. Can J For Res **20**:298-308.

Nkongolo, K., P. Michael, and W. Gratton. 2002. Identification and characterization of RAPD markers inferring genetic relationships among pine species. Genome **45**:51-58.

Prager, E., D. Fowler, and A. Wilson. 1976. Rates of evolution in conifers (Pinaceae). Evolution **30**:637-649.

Price, R., A. Liston, and S. Strauss. 1998. Phylogeny and Systematics of *Pinus*. Pp. 49-68. in Richardson DM, Editor, Ecology and Biogeography of *Pinus*.

Rehfeldt, G. 1999a. Systematics and genetic structure of *Ponderosae* taxa (Pinaceae) inhabiting the mountain islands of the Southwest. Am J Bot **86**:741-752.

Rehfeldt, G. 1999b. Systematics and genetic structure of Washoe pine: Applications in conservation genetics. Silvae Genet **48**:167-173.

Schwilk, D. W., and D. D. Ackerly. 2001. Flammability and serotiny as strategies: correlated evolution in pine. Oikos **94**:326-346.

Six, D. L., and T. D. Paine. 1999. Allozyme diversity and gene flow in *Ophiostoma clavigerum* (Ophiostomatales: Ophiostomataceae), the mycangial fungus of the Jeffrey pine beetle, *Dendroctonus jeffreyi* (Coleoptera: Scolytidae). Can J For Res **29**:324-331.

Smith, J. W., and C. W. Benkman. 2007. A Coevolutionary Arms Race Causes Ecological Speciation in Crossbills. The American Naturalist **169**:455-465.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among pine (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot **92**:2086-2100.

Syring, J. V., K. Farrell, R. Businsky, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst. Biol. **56**:163-181.

Vogler, D. 2000. Coevolution of *Cronartium* with its host. Hort Technology **10**:518.

Wang, X.-Q., D. Tank, and T. Sang. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol **17**:773-781.

Wang, X.-R., A. Szmidt, and H. Nguyen. 2000. The phylogenetic position of the endemic flat-needle pine *Pinus krempfii* Lec. (Pinaceae) from Vietnam, based on PCR-RFLP analysis of chloroplast DNA. Plant Syst Evol **220**:21-36.

Wang, X.-R., Y. Tsumura, H. Yoshimaru, K. Nagasaka, and A. Szmidt. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer and *trnV* intron sequences. Am J Bot **86**:1742-1753.

Willyard, A., J. V. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol **24**:90-101.

Wolff, R. L., F. Pedrono, E. Pasquier, and A. M. Marpeau. 2000. General characteristics of *Pinus* spp. seed fatty acid compositions, and importance of delta5-olefinic acids in the taxonomy and phylogeny of the genus. Lipids **35**:1-22.

Wu, J., K. Krutovskii, and S. Strauss. 1999. Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. Genome **41**:893-908.

Yu, X., A. Ekramoddoullah, D. Taylor, and N. Piggott. 2002. Cloning and characterization of a cDNA of *cro rl* from *Cronartium ribicola*. Fungal Genetics and Biology **35**:53-66.

Zitzetmuller, K., and 5. 2001. Fatty acid composition of Pinaceae as taxonomic markers. American Oil Chemists' Society **36**:439-452.

Fossil Calibration of Molecular Divergence Infers a Moderate Mutation Rate and Recent Radiations for *Pinus*

Ann Willyard, John Syring, David S. Gernandt**,** Aaron Liston, and Richard Cronn

Molecular Biology and Evolution

Oxford University Press

2001 Evans Road

Cary, North Carolina 27513 USA

24(1):90-1001. 2007

ABSTRACT

Silent mutation rate estimates for *Pinus* vary 50-fold, ranging from angiosperm-like to among the slowest reported for plants. These differences either reflect extraordinary genomic processes or inconsistent fossil calibration, and they have important consequences for population and biogeographical inferences. Here we estimate mutation rates from four *Pinus* species that represent the major lineages using eleven nuclear and four chloroplast loci. Calibration was tested at the divergence of *Pinus* subgenera with the oldest leaf fossil from subg. *Strobus* (Eocene; 45 MYA) or a recently published subg. *Strobus* wood fossil (Cretaceous; 85 MYA). These calibrations place the origin of *Pinus* 190–102 MYA, and give absolute silent rate estimates of 0.70–1.31 × 10^{-9} and 0.22–0.42 × 10^{-9}·site^{-1}·year^{-1} for the nuclear and chloroplast genomes, respectively. These rates are approximately four- to 20-fold slower than angiosperms, but unlike many previous estimates, they are more consistent with the high per-generation deleterious mutation rates observed in pines. Chronograms from nuclear and chloroplast genomes show that the divergence of subgenera accounts for about half of the time since *Pinus* diverged from *Picea*, with subsequent radiations occurring more recently. By extending the sampling to encompass the phylogenetic diversity of *Pinus*, we predict that most extant subsections diverged during the Miocene. Moreover, subsections *Australes, Ponderosae*, and *Contortae*, containing over 50 extant species, radiated within a 5 MY time span starting as recently as 18 MYA. An Eocene divergence of pine subgenera (using leaf fossils) does not conflict with fossil-based estimates of the *Pinus*-*Picea* split, but a Cretaceous divergence using wood fossils accommodates Oligocene fossils that may represent modern subsections. Since homoplasy and polarity of character states have not been tested for fossil pine assignments, the choice of fossil and calibration node represents a significant source of uncertainty. Based on several lines of evidence (including agreement with ages inferred using calibrations outside of *Pinus*), we conclude that the 85 MYA calibration at the divergence of pine subgenera provides a reasonable lower bound and that further refinements in age and mutation rate estimates will require a synthetic examination of pine fossil history.

INTRODUCTION

Morphological and molecular analyses of the pine genus (*Pinus*; Pinaceae) reveal conflicting estimates of age and evolutionary rates that are difficult to reconcile. *Pinus* contains two monophyletic subgenera, *Pinus* (diploxylon or "hard pines") and *Strobus* (haploxylon or "soft pines"), diagnosable by two vs. one fibrovascular bundle per leaf (Gernandt et al. 2005). Origin of the genus is thought to date to the early Cretaceous (Millar 1998), while estimates for the divergence of the subgenera range from the late Cretaceous (Millar 1998) to the mid Eocene (Miller 1976). Limited morphological differentiation in the ca. 110 species has been attributed to an exceptionally slow rate of change, but morphological homoplasy (Gernandt et al. 2005) and retention of ancestral molecular polymorphism (Syring et al. 2006) are common. Several studies have reported slow molecular divergence rates for pines (Krupkin, Liston, and Strauss 1996; Dvornyk et al. 2002; Geada López, Kamiya, and Harada 2002; Brown et al. 2004; Sokol and Williams 2005; Ma, Szmidt, and Wang 2006), but these rates appear inconsistent with estimates of per-generation deleterious mutation rates which are known to be at least 10-fold higher in pines than in self-compatible annual flowering plants (Kärkkäinen, Koski, and Savolainen 1996; Klekowski 1998). In contrast, mutation rate estimates comparable to angiosperms have been reported for pines based on antigenic distances (Prager, Fowler, and Wilson 1976), and for a retrotransposon (Kossack and Kinlaw 1999).

Unlike many plant groups, *Pinus* is prominent in paleofloras (Miller 1976; Millar 1998; Price, Liston, and Strauss 1998), has a long history of taxonomic inquiry (Price, Liston, and Strauss 1998; Gernandt et al. 2005), and extensive genomic resources are available (Brown et al. 2001; Temesgen et al. 2001; Chagné et al. 2003; Komulainen et al. 2003; Krutovsky et al. 2004). These enhance its usefulness for studying the interrelationship between morphological and molecular evolution. However, the paleontological data cannot be utilized to full advantage until critical tests of fossil-phylogenetic associations are conducted (e.g. Magallón and Sanderson 2005). The earliest fossil attributable to *Pinus, P. belgica* (Alvin 1960), is a Cretaceous ovulate cone apparently originating from the Wealden Formation in Belgium 145–125 million years ago (MYA). Attempts to incorporate this and other *Pinus* fossils into molecular phylogenetic analyses have taken strikingly different approaches. For calibration, *P. belgica* has been placed at the divergence between *Pinus* and

other modern Pinaceae genera (Wang, Tank, and Sang 2000; García-Gil, Mikkonen, and Savolainen 2003) or at the divergence of subgenera (Sokol and Williams 2005; Ma, Szmidt, and Wang 2006). Because Alvin (1960) further ascribed *P. belgica* to subsection *Pinus*, its age has been commonly applied to the divergence between representatives from the sections of subg. *Pinus* (Krupkin, Liston, and Strauss 1996; Dvornyk et al. 2002; Geada López, Kamiya, and Harada 2002; Brown et al. 2004; Eckert and Hall 2006). Alternative calibrations used 195 MYA based on a presumed Jurassic origin of the genus that lacks explicit fossil evidence (Kutil and Williams 2001), or 45 MYA as the divergence of subgenera based on the earliest fossils representing both subgenera (Kossack and Kinlaw 1999). As a consequence of inconsistent calibration, synonymous mutation rate estimates vary 50-fold, ranging from angiosperm-like (2.8×10^{-9} substitutions·site^{-1}·year^{-1} for *gypsy*-like retrotransposons; Kossack and Kinlaw 1999) to far slower than angiosperms (0.05×10^{-9} for *Adh*; Dvornyk et al. 2002). Similarly, divergence rates calculated for pine chloroplast DNA (cpDNA; 0.06×10^{-9}; Krupkin, Liston, and Strauss 1996) are among the slowest reported for any plant.

Here we evaluate hypotheses concerning the age of *Pinus* and explore the impact of different calibrations on estimated absolute mutation rates across eleven nuclear and four chloroplast loci. A phylogenetic framework was created with an exemplar from each of the four monophyletic sections (a *quartet* of species). Using a critical evaluation of the fossil record, we test two calibrations at the crown, inferred as the divergence between subg. *Strobus* and subg. *Pinus,* and represented by competing hypotheses for the oldest fossils from subg. *Strobus* (putative lower age of 85 MYA and an upper bound of high certainty at 45 MYA). We also use our multi-locus data set to recalculate ages and rates with the calibration scenarios used in recent studies (Krupkin, Liston, and Strauss 1996; Dvornyk et al. 2002; Geada López, Kamiya, and Harada 2002; Brown et al. 2004; Eckert and Hall 2006; Ma, Szmidt, and Wang 2006) to test whether highly heterogeneous rates are an intrinsic feature of the loci and taxa sampled in those studies, or if the variation can be attributed to calibration.

MATERIALS AND METHODS

Plant Materials. Each subgenus is comprised of two monophyletic groups, sections *Trifoliae* and *Pinus* from subg. *Pinus*, and sections *Parrya* and *Quinquefoliae* from subg.

Strobus. For this study, one species was chosen from each section: *Pinus taeda* L. (Mississippi, USA) for section *Trifoliae; P. thunbergii* Parl. (Cheonnam, South Korea) or *P. merkusii* Jungh. & deVriese *s.l.* (Thailand) for section *Pinus*; *P. monticola* D. Don (Oregon, USA) for section *Quinquefoliae;* and *P. nelsonii* Shaw (Nuevo León, Mexico) for section *Parrya*. Haploid genomic DNA from seed megagametophyte tissue was isolated for amplifying nuclear DNA (nDNA) (FastDNA®, Qbiogene, Irvine, CA, USA), while leaf DNA was the source for cpDNA (Gernandt et al. 2005).

Loci evaluated. Nuclear loci included in this study map to seven of 12 linkage groups (LG) in *Pinus* (see Table 2.S1 in Supplementary Material online). Represented are: LG2, *aquaporin*; LG3, chlorophyll binding protein type II precursor (*LHC-CAB*) and late embryogenesis abundant-like protein (*LEA-like*); LG5, arabinogalactan-like protein (*AGP6*) and *ferritin*; LG6, phenylalanine ammonia-lyase (*PAL1*); LG7, 4-coumarate:CoA ligase (*4CL*); LG9, cinnamyl alcohol dehydrogenase (*CAD*) and open stomata (*OST1*); LG10, chloroplast-localized Cu-Zn superoxide dismutase (SOD_{chl}); and the unmapped early-response to dehydration (*ERD3*). Loci from LG3, LG5, and LG9 are sufficiently distant that they are effectively unlinked (Krutovsky et al. 2004). Chloroplast loci included in this study are *matK*, *rbcL*, the *rpL20/rpsS18* intergenic spacer, and the *trnV* intron. See Table 2.S1 in Supplementary Material online for primer and amplicon descriptions.

Amplification, Sequencing, and Analysis. PCR products were directly sequenced using BigDye® v. 3.1 (Applied Biosystems, USA) and visualized on an Applied Biosystems 3730 Genetic Analyzer. DNA alignments were made using ClustalW (Thompson, Higgins, and Gibson 1994) as implemented in BioEdit v. 7.0.1 (Hall 1999) or an iterative process of BLASTN analysis (http://www.ncbi.nih.gov), followed by hand alignment, to align cDNA sequences from GenBank (http://www.ncbi.nlm.nih.gov/Genbank) to genomic sequences. Alignments were made at three levels: *(i)* species within a subgenus, i.e., *P. taeda* to *P. thunbergii*, and *P. monticola* to *P. nelsonii*; *(ii*) the quartet of *Pinus* species, excluding unalignable regions (Syring et al. 2005); and *(iii*) the quartet of *Pinus* species with an outgroup sequence. For ten loci (*4CL*, *AGP6*, *aquaporin*, *CAD*, *LHC-CAB*, *PAL1*, *matK*, *rbcL*, *rpl20/S18*, and *trnV*; (see Table 2.S1 in Supplementary Material online for details), outgroup sequences from *Picea* were obtained using PCR (as described for ingroup

sequences) or by searching GenBank for putative genomic or expressed sequence orthologs. Outgroup sequences were not isolated for *ERD3* and *SODchl* because of orthological concerns regarding some ingroup amplicons (see Results). Amplification of the *LEA-like* locus failed in *Picea*, so a GenBank sequence from *Pseudotsuga* was used as the outgroup. The closest GenBank matches available for *ferritin* and *OST1* (*Arabidopsis thaliana*) were used to provide a root for these two loci. New nucleotide sequences were submitted to GenBank (see Table 2.S2 in Supplementary Material online for details); alignments are available as Supplementary Material online and Figure 2.S1in this dissertation.

Pairwise substitution rates for silent (dS; synonymous plus noncoding) and nonsynonymous (dN) sites were calculated using DnaSP (Rozas et al. 2003) using the approximation method of Nei and Gojobori (1986) with a Jukes-Cantor correction for multiple substitutions. DnaSP (Rozas et al. 2003) was also used to calculate GC content. Models of sequence evolution (number of substitution categories, base frequencies, shape parameter, and proportion of invariant sites) were tested independently for all sites and for the silent sites (approximated by including noncoding and third codon positions) of each locus using ModelTest 3.7 (Posada and Crandall 1998). Models selected by the hierarchical likelihood ratio tests in ModelTest 3.7 were used to obtain maximum likelihood (ML) estimates of phylogeny with PAUP* ver. 4.0b10 (Swofford 2002). To test for rate equality (i.e. clock-like evolutionary history) among lineages, we used the likelihood ratio test (Muse and Weir 1992) to compare clock-enforced and clock-relaxed likelihood scores that were obtained using the selected evolutionary models described above (one-tailed probability, $\alpha = 0.05$).

A partition homogeneity test (Cunningham 1997; 100 replicates using a heuristic search) was used to test for significant conflict between loci (P=0.01) for all sites as well as silent sites from three different data sets (see Table 2.S3 in Supplementary Material online for details). The three data sets are: *(i)* nine nuclear loci with different outgroup species (*Picea*, *Pseudotsuga*, or *Arabidopsis*; *4CL, AGP6, aquaporin, CAD, ferritin, LEA-like, LHC-CAB, OST1, PAL1*; 8346 bp; 6098 silent bp); *(ii)* five nuclear loci with *Picea* as the outgroup (*4CL, AGP6, aquaporin, CAD, LHC-CAB,* 3973 bp; 2348 silent bp); and *(iii)* four chloroplast loci with *Picea* as the outgroup (*matK, rbcL, rpL20/S18, trnV*; 3903 bp; 1914 silent bp). Since data sets did not exhibit significant conflict, they were concatenated, and models of sequence evolution were selected for concatenated data sets as described above.

To investigate the impact of increased taxon density on rate and divergence estimates, we examined data from Syring et al. (2005) that include four loci evaluated in this study (*4CL, AGP6, LEA-like, LHC-CAB*; 5338 bp; 3766 silent bp) for 12 pine species serving as exemplars for subsections with *Picea* as the outgroup. Evolutionary model selection, likelihood ratio tests and partition homogeneity tests were performed on each locus and on the concatenated silent sites as described above.

Fossil calibration. Since all pine fossil reports that were considered as calibration sources lack radiometric dating, ages were adjusted to the midpoint of the currently assigned range for their geological Epoch or Stage (Gradstein and Ogg 2004). *Pinus* pollen, pollen cones, and ovulate cone casts lack diagnostic characters for subgeneric identification (Miller 1976; Phipps, Osborn, and Stockey 1995), so only anatomical reports from ovulate cones, leaves, and wood were considered for calibration sources. The earliest *Pinus* fossil, *P. belgica* (145–125 MYA; Alvin 1960), shows affinities to subg. *Pinus* (Miller 1976) and has been considered part of subsection *Pinus* (Millar 1998). However, the collection locality for *P. belgica* (Wealden Formation, Belgium) was inferred from the lignitic state and adhering particles (Alvin 1960), leaving its geographic and stratigraphic origins uncertain. There is a ca. 35 MY gap until the next *Pinus* fossils (100–75 MYA ovulate cones from North America, Japan, and Europe), and these fossils also have affinity to subg. *Pinus* (Fliche 1896; Alvin 1960; Robison 1977; Blackwell 1984; Miller and Malinky 1986; Stockey and Nishida 1986; Saiki 1996). Since these early fossils show symplesiomorphic features and lack synapomorphies to support infrageneric nodes, they cannot be used as reliable calibration sources (Magallón and Sanderson 2001).

The node corresponding to the divergence of the subgenera is well supported with both morphological and molecular synapomorphies (Gernandt et al. 2005). Since all of the oldest *Pinus* fossils have affinity to subg. *Pinus*, the first fossil representing subg. *Strobus* supports the minimum age of divergence of the subgenera. The appearance of subg. *Strobus* dates to the late Cretaceous based on permineralized wood anatomy (Santonian, 83.5–85.8 MYA, midpoint ca. 85 MYA, *Pinuxylon* sp.; Meijer 2000), or to the mid Eocene based on either leaf anatomy (37.2–48.6 MYA, midpoint ca. 45 MYA, *Pinus similkameenensis*; Miller 1973), or ovulate cones (ca. 43 MYA; Axelrod 1986). It is important to note that Cretaceous fossils have been attributed to subg. *Strobus* (Jeffrey 1908; Stopes and Kershaw 1910; Penny

1947), but these fossils are not useful for calibration since they have been reassigned to subg. *Pinus* or to other genera (see Discussion). Therefore, we calibrated the divergence of the subgenera with a putative lower age of 85 MYA based on wood anatomy and an upper age of 45 MYA based on leaves and ovulate cones.

Rates and ages. Absolute rates of silent site changes in the concatenated data sets were estimated using two approaches. First, we computed silent site divergence (dS) for comparisons between subgeneric representatives using nine nuclear and four chloroplast loci. Lower and upper rate estimates (i.e. two calibration ages) were calculated using the formula $\mu = dS / 2T$, where μ is the silent divergence$\cdot site^{-1}\cdot year^{-1}$, dS is the mean of silent substitutions$\cdot site^{-1}$ (weighted by number of sites in each locus), and T is time in years. Second, we calculated ML branch lengths (using the selected model of sequence substitution described above; see Table 2.S3 in Supplementary Material online for details) from clock-enforced phylograms using silent sites from pine quartet alignments of nDNA and cpDNA. Silent-site data sets were used for age inference to maintain consistency with silent rate estimation and to minimize the potential for selection to distort the evolutionary history or rate at a locus. Ages were also inferred using *all* nucleotide sites to provide corroboration of silent site estimates. To account for departures from clock-like behavior in the 12-species data set, we used a non-clock-enforced tree based on silent sites to assign dates and to estimate a range of local rates using Penalized Likelihood (PL), with a smoothing factor estimated using cross-validation (r8s ver. 1.70; Sanderson 2002).

RESULTS

Sequence variation across loci, genomes, and lineages. Alignments of eleven nuclear and four chloroplast loci from four *Pinus* species included 11,481 bp and 10,643 bp from subg. *Pinus* and *Strobus*, respectively (Table 2.1). Mean lengths of nuclear sequences were shorter than chloroplast sequences (690 bp vs. 972 bp in subg. *Pinus*), but the larger sample of nuclear loci yielded a substantially larger nDNA data set (e.g. 7594 bp nDNA vs. 3887 bp cpDNA in subg. *Pinus*). In total, 65% of nDNA and 77% of cpDNA was exonic, while the noncoding portion included introns (*4CL*, *aquaporin*, *CAD*, *ERD*3, *ferritin*, *LEA-like*, *OST1*, SOD_{chl}, *trnV*), 5′ (*OST1*) and 3′ (*ferritin*) untranslated regions, and an intergenic spacer (*rpl20/S18*). Noncoding regions provided 51% of nDNA silent sites (1959 bp vs.

1882 third-codon positions) and 40% of cpDNA silent sites (876 bp vs. 1289 third-codon positions). Nine of the mapped nuclear loci lacked heterozygosity when amplified from haploid megagametophyte tissue, providing evidence for target specificity and orthology. Two loci – SOD_{chl} and *ERD3* – showed evidence of paralogy and were excluded from rate analyses. These loci were unexpectedly heterogeneous in subg. *Pinus*, indicative of gene duplication or non-specific priming. The SOD_{chl} sequence from *P. monticola* also lacked four expected introns, raising the possibility that it is a reverse transcribed pseudogene. GC% showed considerable variation across loci. For nDNA, the average GC content was 47.3% (range = 32.9–66.7%), with coding slightly higher (51.3%; range = 40.5–67.0%) and noncoding considerably lower (35.8%; range = 23.6–44.5%). For cpDNA, the average GC content was 37.6% (range = 31.4–45.4%), and the GC content in coding (37.8%; range = 32.5–45.4%) and noncoding (36.2%; range = 34.4–39.2%) regions were very similar.

Sequence divergence in nDNA and cpDNA showed two important trends. First, while silent substitutions per site (dS) averaged ca. three fold higher in nDNA than cpDNA, mean nonsynonymous substitutions per site (dN) are almost identical among eight nuclear (0.020) and four chloroplast (0.021) loci (Table 2.1). Second, divergence at nDNA was significantly greater between representatives of subg. *Strobus* (mean dS = 0.063) than representatives of subg. *Pinus* (mean dS = 0.042; $F = 15.19$, $P = 0.001$). This trend was apparent in cpDNA, but the difference was not significant (dS *Strobus* = 0.015, dS *Pinus* = 0.010; $F = 0.982$, $P = 0.359$).

Fossil calibrated mutation rates and divergence dates. Single-locus ML phylogenies (results not shown) using *all* characters in the quartet alignments showed the expected species relationships in all cases except *OST1* (section *Quinquefoliae* is sister to section *Parrya* and subg. *Pinus*) and *ferritin* (ingroup nodes were unresolved). Rooting of these two loci with *Arabidopsis* is likely responsible for the topological differences, and we excluded these loci from our chronograms. Rate equivalence (as inferred by the likelihood ratio test) was statistically supported for all individual loci except *LHC-CAB*, *ferritin,* and *PAL1* using *all* characters and for all individual loci except *PAL1* using *silent* sites (see Table 2.S3 in Supplementary Material online for details). A concatenated alignment of the five clock-like loci that share *Picea* as outgroup (*4CL, AGP6, aquaporin, CAD, LHC-CAB*) exhibited rate equality using either *silent* sites or *all* sites (see Table 2.S3 in Supplementary

Material online for details). Partition homogeneity tests among these loci (data not shown) indicated that these data sets do not reflect conflicting topologies, supporting concatenation. Rate equivalence was supported for *all* sites as well as *silent* sites for each cpDNA locus and for a concatenated cpDNA alignment (see Table 2.S3 in Supplementary Material online for details). Rates were inferred from silent sites for each genome; ages were calculated using all-site and silent-site data sets.

Based on a calibration with 85 or 45 MYA, absolute silent mutation rates (μ) for nine nuclear loci average 0.70 or 1.31×10^{-9} silent substitutions·site^{-1}·year^{-1} for nDNA, and 0.22 or 0.42×10^{-9} for cpDNA, respectively (Table 2.2). To address the impact of unequal base frequencies and among-site rate heterogeneity on rate estimates, the absolute silent mutation rate (μ) was also calculated from ML branch lengths for each calibration test (Table 2.3). ML-based rates are marginally higher than dS-based rates for nDNA ($\mu = 0.75$ or 1.41×10^{-9}), while cpDNA rates are almost identical ($\mu = 0.21$ or 0.40×10^{-9}).

Chronograms based on silent sites from nDNA and cpDNA reveal identical topologies, but disparities in the branch lengths (especially at deeper nodes) yield different estimated divergence dates for *Pinus* lineages (Figure 2.1, Table 2.3). Most notably, estimates from the 5-locus nDNA data set predict older divergence events between *Pinus* and *Picea* (190–102 MYA) than does cpDNA (164–136 MYA). Both genomes show sections of subg. *Strobus* diverging before those of subg. *Pinus*, although nDNA estimates are more ancient (48–25 for subg. *Strobus* vs. 30–16 MYA for subg. *Pinus*) than those from cpDNA (37–19 MYA for subg. *Strobus* vs. 25–13 MYA for subg. *Pinus*). Divergence dates inferred from the same data sets using *all* nucleotides were nearly identical at sectional divergences, but somewhat younger for the *Pinus*/*Picea* node (Table 2.3). Divergence dates were also estimated using silent sites in the clock-enforced data set that included four additional nuclear loci (*ferritin*, *LEA-like*, *OST1*, *PAL1*; 6098 bp). The use of different outgroups for three of these loci precluded us from estimating the divergence of *Pinus* from *Picea* (node D), but estimated sectional divergences within subg. *Pinus* (31–16 MYA; node A) and subg. *Strobus* (48–25 MYA; node B) were not different from estimates based on five clock-like loci (Table 2.3).

Cross-validation yielded an optimum smoothing factor of 63 for PL in the 12-taxon data set. In general, results from PL indicate more recent divergence events than predicted from the analysis of only four exemplar species (Figure 2.2; Table 2.3). For example, using

an 85 MYA calibration, subg. *Strobus* sections are predicted to have diverged 37 MYA, or 10 MY more recently than predicted from four-taxon comparisons (Figure 2.1). Similarly, subg. *Pinus* sections are predicted to have diverged 28 MYA, which is 2 MY more recent than indicated by four-taxon estimates (Figure 2.1). An important finding highlighted by this additional taxon sampling is that modern pine subsections radiated in a narrow time span in the relatively recent past. For example, using the 85 MYA calibration, the three subsections of *Trifoliae* (Figure 2.2b, nodes E-G), with ca. 51 extant species, radiated within a 5 MY time span starting 18 MYA.

DISCUSSION

Mutation rate estimates are used widely for hypothesis testing in molecular, population, and evolutionary genetic studies (Muse 2000), and they are increasingly used to estimate genetic parameters of conifers (Dvornyk et al. 2002; García-Gil, Mikkonen, and Savolainen 2003; Brown et al. 2004; Ma, Szmidt, and Wang 2006). Numerous estimates of mutation rates have been made for angiosperms (e.g. Wolfe, Li, and Sharp 1987; Wolfe, Sharp, and Li 1989; Gaut et al. 1996; Koch, Haubold, and Mitchell-Olds 2000; Clark, Tavaré, and Doebley 2005), but comparable conifer estimates are limited to studies of single-locus variation across multiple species (Geada López, Kamiya, and Harada 2002), multi-locus variation in a single species (García-Gil, Mikkonen, and Savolainen 2003; Brown et al. 2004) or multi-locus variation in a few closely related species (Ma, Szmidt, and Wang 2006). The perspective added by our synthesis of multiple nuclear and chloroplast loci based on exemplar taxa and two fossil calibration points, indicates that divergence times separating pine lineages have been frequently overestimated, with a concomitant underestimation of absolute mutation rates (Krupkin, Liston, and Strauss 1996; Dvornyk et al. 2002; Geada López, Kamiya, and Harada 2002; Brown et al. 2004; Sokol and Williams 2005; Ma, Szmidt, and Wang 2006). This new perspective supports a relatively recent radiation of extant pine sections in the early Miocene, a hypothesis that has been proposed (e.g. Miller 1973; Strauss and Doerksen 1990) (but often ignored) to explain the modest morphological and genetic divergence between pine species.

Sources of error. Confidence intervals for molecular clock estimations must consider the simultaneous uncertainties of rate and time that are confounded within

divergences, as well as the age and placement of calibration points. Since fossils document minimum divergence times, the elapsed time from taxon origin to the earliest fossil cannot be known. Additional uncertainty is introduced by variation in deposition, discovery, and the quality of our understanding of the age of the source formation as well as the taxonomic affinity of the fossil. There are promising approaches for estimating confidence intervals for divergence times (e.g. Kumar et al. 2005; Yang and Rannala 2006), but these will require multiple robust fossil calibrations for *Pinus*. In addition, molecular divergence rates may vary across the genome and across lineages. We addressed this complex problem using several approaches to provide a perspective on the relative impact of various factors on the error ranges of our rate estimates.

First, both of the fossils selected for calibration lack radiometric dating. Since midpoints of the ranges for geological Epochs or Stages were used to create the chronograms (Figures 2.1, 2.2; Table 2.3), we recalculated divergence dates using the upper and lower bounds of these periods. The stratigraphic age ranges for wood from the Santonian *Pinuxylon* sp. (Meijer 2000) and leaves from the Middle Eocene *Pinus similkameenensis* (Phipps, Osborn, and Stockey 1995) are currently considered to be 83.5–85.8 MYA, and 37.2–48.6 MYA, respectively (Gradstein and Ogg 2004). The inferred age ranges for node D (cf. Table 2.3), using these stratigraphic ranges to calibrate at node C, are 187–192 MYA and 83–109 MYA for calibrations based on wood and leaves, respectively. Inferred age ranges for node A are 29.6-30.4 and 13–17 MYA for calibrations based on wood and leaves, respectively, while node B ages are 47–48 and 21–27 MYA, respectively. Clearly, the use of midpoint ages introduces uncertainty in our estimates, particularly for Eocene fossils since that Epoch is much longer than the Santonian (11.4 vs. 2.3 MY). Nonetheless, this uncertainty is small relative to the differences in alternative calibration points suggested by fossil wood (85 MYA) vs. fossil leaves (45 MYA) for the origin of *Pinus* subgenera.

Second, the use of the mean substitution rates from multiple loci reduces stochastic deviation resulting from rate variation among genomic regions, and limiting the sample to silent sites reduces the potential for selection to influence rates. To assess the amplitude of this variation among our nine nuclear loci, we also calculated absolute substitution rates using the mean dS plus and minus one S.D. This results in $\mu = 0.43–0.96 \times 10^{-9}$ for a calibration based on wood and $0.81–1.81 \times 10^{-9}$ for leaves (cf. Table 2.2). We also projected the *combined* effects of locus-specific rate variation (mean dS plus one S.D. and the upper

bound for the stratigraphy; dS minus one S.D. and the lower bound) for each calibration. This results in *inclusive* ranges of $\mu = 0.42–0.98 \times 10^{-9}$ for a calibration based on wood and $0.75–2.2 \times 10^{-9}$ for leaves (cf. Table 2.2). Among the nine loci that were used for rate calculation, distortion may have been introduced by our use of different outgroups for *ferritin*, *LEA-like*, and *OST1* (see Table 2.S2 in Supplementary Material online for details). Further, *PAL1* silent sites showed a significant departure from rate equivalence even with *Picea* as the outgroup. However, ages and rates inferred from the five- and nine-locus data sets (Table 2.3) are largely consistent. Further, data accumulated for 50 *Pinus* loci (Cronn, unpublished data) show a median silent rate that is very similar to the silent rate described in these results. This implies that, despite locus-specific rate variation, our five-locus data set is useful as a first approximation of μ in *Pinus*.

Finally, we derived confidence intervals for age estimates using the method of Haubold and Wiehe (2001), which uses nonoverlapping pairs of phylogenetic distances to infer the unknown mutation rate for the other pair. Using this method at node A (Figure 2.1b), with subg. *Strobus* as the reference and ages calculated from the 85 MYA crown calibration, yields 95% confidence intervals of 14–67 MYA for the divergence of subg. *Pinus*. Applying this technique to node B (with subg. *Pinus* as the reference) yields intervals of 22–106 MYA for subg. *Strobus*. The magnitude of these intervals far exceeds error ranges introduced by stratigraphic ranges, intra-locus rate variation, or even fossil choice between permineralized wood (85 MYA) or leaves (45 MYA). These large uncertainties clearly highlight the challenge inherent in the simultaneous uncertainties of rate and time over evolutionary time scales.

While the magnitude of these sources of error highlight the importance of considering molecular rate variation and stratigraphic uncertainty, past calibration scenarios applied to *Pinus* show that incorrect fossil assignment can be a far more dramatic source of error. For example, the common practice of using *P. belgica* to calibrate the divergence of sections within subg. *Pinus* (node A, Figure 2.3e) pushes the divergence of pine subgenera (node C) to 339 MYA, and the divergence of *Pinus* – *Picea* (node D) to 758 MYA in the Precambrian. The attendant silent mutation rate in the nuclear genome is exceptionally slow ($\mu = 0.19 \times 10^{-9}$). This unrealistic (but commonly cited) calibration is clearly responsible for many of the exceptionally low μ values reported for pine nDNA and cpDNA (Krupkin, Liston, and Strauss 1996; Dvornyk et al. 2002; Geada López, Kamiya, and Harada 2002; Brown et al.

2004; Ma, Szmidt, and Wang 2006). This calibration can also cause substantial distortion in biogeographical interpretations. For example, Eckert and Hall (2006) recently hypothesized dispersal and vicariance events using *P. belgica* to calibrate the divergence of sections within subg. *Pinus* (node A, Figure 2.3f) in a cpDNA data set. In doing so, they fail to consider the unrealistic estimate that this calibration indicates for the divergence of *Pinus* and *Picea* (i.e. 720 MYA) based on our clock-like cpDNA branch lengths (Figure 2.3f). In the same manner, associating *P. belgica* with the divergence of pine subgenera also conflicts with the fossil record. Such a calibration (e.g. 136 MYA, Sokol and Williams 2005; 130 MYA, Ma, Szmidt, and Wang 2006), indicates a *Pinus*/*Picea* split of ca. 300 MYA (Figure 2.3d), and produces absolute silent mutation rates of $\mu = 0.47–0.49 \times 10^{-9}$. These projected ages for *Pinus* (Carboniferous or early Permian, respectively) predate *P. belgica* by more than 150 MY. In summary, the errors resulting from incorrect fossil/node association far exceed the cumulative error estimated with confidence intervals that account for the confounding effect of simultaneous rate and time variation. In contrast, using *P. belgica* to represent the divergence of *Pinus* and *Picea* in our clock-like data set (node D, Figure 2.3c; Wang, Tank, and Sang 2000; García-Gil, Mikkonen, and Savolainen 2003) places the divergence of *Pinus* subgenera at 63 MYA. This estimate lies between the earliest putative subg. *Strobus* wood and leaf fossils.

Implications of a moderate tempo for pine mutation rates. Several authors have argued that a high mutation rate is necessary to account for the extremely high level of inbreeding depression found in conifer species capable of partial self-fertilization (Lande, Schemske, and Schultz 1994; summarized in Scofield and Schultz 2006). Since plants do not segregate a germ line, somatic mutations may accumulate in meristems and be incorporated into gametes. Hence, longevity and large stature contribute to a large number of mitoses, both of which may serve to elevate the per-generation mutation rate (Scofield and Schultz 2006). Evidence for a high per-generation mutation rate in *Pinus* compared to other plant groups is provided by observations of the frequency of chlorophyll-deficient mutants. Assuming an equal number of loci capable of mutating to chlorophyll deficiency and equivalent lengths across loci, the per-generation rate of deleterious mutations for *Pinus sylvestris* ($U = 1–3 \times 10^{-2}$; Kärkkäinen, Koski, and Savolainen 1996) is comparable to the mutation rate for *Rhizophora mangle*, another long-lived woody perennial ($U = 1.5 \times 10^{-2}$;

Klekowski & Godfrey 1989). These mutation rates are ca. 100-fold higher than the average for 10 annual flowering plant species ($U = 1–3 \times 10^{-4}$; Klekowski 1992).

Mutation rates in *Pinus* can be evaluated by comparing absolute rates derived from fossil calibrations in other plant groups. Our 85 MYA calibration provides an estimate for nDNA ($\mu = 0.70 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}; Table 2.2) that is 14-fold higher than previous *Pinus* estimates (Dvornyk et al. 2002), but ca. four-fold slower than the rate for palms ($\mu = 2.61 \times 10^{-9}$; Gaut et al. 1996) and seven- to 40-fold slower than the range reported for herbaceous angiosperms ($\mu = 5–33 \times 10^{-9}$; Gaut et al. 1996; Koch, Haubold, and Mitchell-Olds 2000; Clark, Tavaré, and Doebley 2005). Similarly, our estimated rates for chloroplast silent sites ($\mu = 0.22 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}; Table 2.2) are ca. six-fold slower than those reported for angiosperm cpDNA ($\mu = 1.1–1.6 \times 10^{-9}$; Wolfe, Li, and Sharp 1987). Expressed on a per-year basis, these comparisons suggest that substitution rates in *Pinus* are still far slower than most angiosperms. In this context, it's important to note that molecular substitution rates can also show a "generation-time effect" (e.g. Gaut et al. 1992; Gaut et al. 1996; Kay, Whittall and Hodges 2006). Based on our estimate of the average rate ($\mu = 0.70 \times 10^{-9}$ synonymous substitutions·site^{-1}·year^{-1}) and correcting for longevity (assuming a 25-year generation time; Brown et al., 2004), the per-generation nuclear substitution rate for *Pinus* averages 1.75×10^{-8} substitutions·site^{-1}·generation^{-1}. This rate is nearly equivalent to the 1.5×10^{-8} substitutions·site^{-1}·year^{-1} rate inferred for short-lived Brassicaceae (Koch, Haubold, and Mitchell-Olds 2000). The near-equivalence between these values closes the gap between gymnosperm and angiosperm rates relative to prior rate estimates (e.g., Brown et al., 2004; Dvornyk et al., 2002), but they still fall short of the 100-fold difference that might be expected based on chlorophyll-deficiency mutations.

Clearly, these calculations hinge on an appropriate association between a calibration time and a phylogenetic node; the choice of a node with which to associate a fossil is the largest determinant of the rate. The availability of *Pinus* fossils with synapomorphies supporting the divergence of *Pinus* subgenera (a well-supported node in a molecular phylogeny) is more robust than options available for many plant groups, and we suggest that this will be an important factor for future, more refined, rate comparisons. This study illustrates the variability of divergence rates among loci since dS for nine nuclear loci ranges from 0.063 to 0.205 (Table 2.1). Even though fossil dates were used as fixed calibration

points in our calculations, each fossil represents a *minimum* age for a lineage. It follows, then, that the rates calculated from mean dS ($\mu = 0.70–1.31 \times 10^{-9}$ substitutions·site^{-1}·year^{-1}; Table 2.2) represent *maximum* absolute rates, although the same caveat applies to angiosperm rates used for comparison.

Our revised estimates of mutation rates in pines have important implications for studies of population and molecular genetic parameters. For example, an effective population size (N_e) of 5.6×10^5 was calculated for *P. taeda* using 19 loci and a substitution rate ($\mu = 1.17 \times 10^{-10}$; estimated by calibrating divergence between members of section *Trifoliae* with the age of *P. belgica*; Brown et al. 2004). If we apply our estimate of $\mu = 0.70 \times 10^{-9}$ (calibrated at 85 MYA between the subgenera; Table 2.2), we find that the predicted N_e for *P. taeda* is far lower at 9.4×10^4. This value of N_e contrasts sharply with census population estimates for *P. taeda* (which may exceed 10^{10}; Brown et al. 2004), and it suggests that this species has experienced dramatic population growth following a relatively recent genetic bottleneck in the history of this species.

***Pinus* fossil status.** Despite the diversity and abundance of *Pinus* fossils (see Millar 1998), several obstacles need to be overcome before integrating additional fossils into a multi-point calibration. To date, pine fossil reports have been based on unattached organs (ovulate or pollen cones, leaves, or wood fragments). Two of these organs have yet to be found on a contiguous fossil, although species have been named based on the hypothesized common origin of separate organs. Retention of ancestral character states and homoplasy among extant *Pinus* species are sufficiently frequent that characters from a single organ are inadequate for discriminating among extant pine subsections or sections (Gernandt et al. 2005). Since *Pinus* subgenera are diagnosable by leaf vasculature (Gernandt et al. 2005), and by a combination of wood features (Van der Burgh 1973), this key divergence event provides the most reliable calibration point for *Pinus*.

Historically, fossil descriptions used a typological approach for assigning affiliations with extant taxa. This process is complicated by historical revisions of *Pinus* sectional affiliations and it lacks the necessary phylogenetic framework for making assignments. As previously noted, the number of fibrovascular bundles per leaf is the only non-homoplasious character known to date that diagnoses *Pinus* subgenera (Gernandt et al. 2005). However, the presence of a single fibrovascular bundle at the base of a leaf fossil may not be

diagnostic for the subgenera since bundles branch distally into the two bundles characteristic of subg. *Pinus* (Stockey and Nishida 1986). Based on this information, *Pinus* sp. leaf fossils from Staten Island, New York (Jeffrey 1908), are now thought to represent subg. *Pinus* rather than subg. *Strobus*. Similarly, Stockey and Nishida (1986) suggest that Cretaceous *P. yezoensis* leaf fossils from Japan (Stopes and Kershaw 1910), also once believed to represent subg. *Strobus*, are more representative of *Cedrus*.

While needles are informative within *Pinus*, ovulate cones are considered the most dependable evidence for assigning fossils to the genus. Miller (1976) outlined four characteristics that together define *Pinus*: inflated scale apex, bract and scale traces united at origin, all resin canals abaxial to vascular tissue in scale base, and scale strands curved on adaxial side. Critically, members of the extinct cone genus *Pityostrobus* can share at least two of these features, complicating the identification of Cretaceous pine ovulate cones. Indeed, using Miller's criteria, some early 'pine' fossils have been reassigned to *Pityostrobus*, *Picea*, or *Cedrus*. For example, Miller and Malinky (1986) referred cone scales from the Magothy Formation of Delaware, originally described as representing subg. *Strobus* (Penny 1947) to *Pityostrobus*. Since many fossil cones are found in marine deposits, abrasion during transport can remove important cone scale features (Smith and Stockey 2002), misleading attempts at classification (Wolfe and Schorn 1989).

Evidence for an Eocene divergence of subgenera. Some authors have suggested a relative recent origin for subg. *Strobus* in the Eocene (Miller 1973; Phipps, Osborn, and Stockey 1995; Kossack and Kinlaw 1999), as reflected by our 45 MYA crown calibration. This hypothesis is based on numerous ovulate cone fossils assigned to subg. *Strobus* that appear in the fossil record starting ca. 40 MYA (Millar 1998), closely following the oldest known leaves attributable to subg. *Strobus* (45 MYA). Unattached leaves (with subg. *Strobus* features), deposited near ovulate cones (with subg. *Pinus* features), in the Princeton Chert in British Columbia were originally described as two species: *Pinus similkameenensis* and *P. arnoldii* (Miller 1973). More recently, Phipps, Osborn, and Stockey (1995) reinterpreted these fossils as a single species (*P. similkameenensis*), and proposed that the mosaic of characters in this hypothetical species may represent a lineage that predates the divergence of subg. *Strobus.* These authors did note that alternative explanations (e.g.,

organs derive from a lineage with no extant descendent, or from sympatric pine species) cannot be ruled out.

A 45 MYA calibration also shows surprising agreement with a recent cpDNA-based estimate of seed plant divergences using entirely different calibration sources (Magallón and Sanderson 2005). In that analysis, the divergence of Gnetophytes and Pinaceae was constrained with a fossil date of 216 MYA; estimated divergence dates for *Pinus* subgenera (ca. 50 MYA) and synonymous cpDNA substitution rates ($\mu = 0.26 \times 10^{-9}$) nearly match our results. Since fossils represent minimum ages, calibration at the divergence of *Pinus* subgenera with 45 MYA leaf fossils appears to provide a reasonable upper bound for this event.

Evidence for a Cretaceous divergence of subgenera. The description of a *Pinuxylon* sp. from the Aachen (Meijer 2000) may push the age of subg. *Strobus* more than 40 MY before the oldest leaf and cone fossils, but this is not a simple determination. The cone genus *Pityostrobus* was abundant and diverse during the Cretaceous, but associated wood remains undescribed and its exact relationship to *Pinus* is unresolved (Smith and Stockey 2002). Further, the two characters used to affiliate the Aachen *Pinuxylon* wood with subg. *Strobus* (crossfield pitting and ray tracheid dentations) are not unequivocal synapomorphies (data not shown), even in extant pine species, indicating the possibility of retention of ancestral polymorphism or parallel evolution. Both of these tracheid features are quantitative traits for which intra-species variation has been noted (Shaw 1914; Hudson 1960; Van der Burgh 1973) and for which the ancestral state is unclear (Hart 1987). Perhaps most challenging is that diagnosis of subg. *Strobus* fossil wood is made by determining that a specimen has *weak to absent* dentations. In this context, the description of the Aachen *Pinuxylon* sp. as possessing "ray tracheids faintly dentate but owing to rather poor state of preservation only locally visible" (Meijer 2000), may best be considered tantalizing evidence for the existence of subg. *Strobus* in the Cretaceous, but evidence requiring corroboration. The evidence for subg. *Strobus*-like wood in the Cretaceous is bolstered by fossil wood of about the same age that shows similarities to modern-day members of subg. *Pinus* (Blackwell 1984), hence providing wood fossils flanking the crown node. While the possibility that the subg. *Pinus*-like fossils may simply be the ancestral state in Pinaceae

cannot be excluded, they do lend credence to the hypothesis that the Aachen *Pinuxylon* represents *Pinus* rather than another member of the Pinaceae.

The 45 MYA calibration (even when viewed as a minimum age) suggests a more recent *Pinus–Picea* split (102–87 MYA; Figure 2.1) than is indicated by the existence of *P. belgica* at ca. 135 MYA (Miller 1976). However, there are unresolved issues concerning the age of the genus. First, the ca. 35 MY gap between *P. belgica* and the next *Pinus* fossil, combined with uncertainties about the source of *P. belgica*, deserves note. Second, the only other evidence for an *early* Cretaceous origin for *Pinus* is pollen from Alaska (Langenheim, Smiley, and Gray 1960). The age of the Kuk River fossil formation that is the source for these reported 145–100 MYA *Pinus* microfossils is now estimated to be the mid to late Albian (104–97 MYA; Koteja and Poinar 2001). In addition, Erwin and Schorn (2006) recommend caution when relying on pollen without associated megafossils due to the unknown status of pollen associated with seed cone genera such as *Pityostrobus*. If one were to discount the inference of an *early* Cretaceous age for *P. belgica* and the previous estimate for the age of the Alaskan pollen, the oldest evidence for *Pinus* would be ovulate cones from the Albian-Cenomanian (ca. 103 MYA; Fliche 1896). This does not appear to fit with the 190–164 MYA *Pinus–Picea* split implied by a Cretaceous divergence of the subgenera (Figure 2.1). On the other hand, the earlier age estimate for our node D (Figure 2.1) that results from the 85 MYA calibration appears consistent with an older divergence of *Pinus* and *Picea*, since there is support for an early independent divergence of these two genera from the *Pityostrobus* grade (Smith and Stockey 2002). The 85 MYA calibration also agrees with a study of Pinaceae genera which calibrated the *Pinus-Picea* split at 140 MYA, yielding an estimated age of ca. 84 MYA for the divergence of *Pinus* subgenera (Wang, Tank, and Sang 2000).

Perhaps the strongest evidence that subg. *Strobus* diverged in the Cretaceous are the numerous Oligocene (ca. 26 MYA) fossils that potentially represent recent lineages. Oligocene ovulate cone fossils have been affiliated with six different *Pinus* subsections: *Pinaster* and *Pinus* in section *Pinus* (Mai 1986; Erwin and Schorn 2006); *Ponderosae* in section *Trifoliae* (Wolfe and Schorn 1989); *Balfourianae* and *Cembroides* in section *Parrya* (Wolfe and Schorn 1989); and *Strobus* in section *Quinquefoliae* (Mai 1986). Existence of these subsections ca. 26 MYA is in better agreement with our predicted age of these groups using an 85 MYA calibration (Figure 2.2b, nodes E-I). This assessment carries two

important caveats. First, most subsectional assignments have been made by typological matching that considered extant taxa from only one continent (but see Erwin and Schorn 2006). Second, a cladistic analysis of the characters used to affiliate fossils to subsections has yet to be conducted, so their value at this taxonomic level remains speculative. Based on the framework provided by our clock-like data, many of these Oligocene fossils must reflect symplesiomorphic character states for the simple reason that they infer too ancient a divergence for *Pinus* lineages. Even our 85 MYA chronogram does not project subsections diverging early enough to support the affiliation of late Cretaceous leaf fossils (ca. 85–78 MYA) with three different subsections: *Pinaster* (Stockey and Nishida 1986), *Pinus* (Robison 1977), and *Ponderosae* (Stockey and Nishida 1986). Integrative fossil validation, such as the method used to discard "outlier" fossils in a time-calibrated phylogeny of turtles (Near, Meylan, and Shaffer 2005) can be used to take advantage of the numerous pine fossils, but only as putative synapomorphies supporting fossil associations are identified.

Concluding remarks. Our interpretations of mutation rates and divergence ages are relatively insensitive to the choice of genome (nDNA vs. cpDNA), the selection of silent sites vs. all sites, the number of taxa, or clock constraints (four taxa constrained vs. 12 taxa unconstrained). Instead, the choice of fossil and node for calibration have a pronounced impact. This suggests that refinements of estimated mutation rates will benefit more from a re-evaluation of the *Pinus* fossil record based on a cladistic analysis of morphological characters in extant taxa rather than increased genomic sampling. Because *Pinus* has a rich fossil record, a critical evaluation of the morphological synapomorphies that support additional phylogenetic nodes should be made a priority. The dramatic distortions of ages and rates (Figure 2.3) resulting from incorrect placement of fossils should serve as a cautionary note to studies of fossil-poor families. As in angiosperms (Sanderson and Doyle 2001), a relatively recent divergence of the crown group in *Pinus* may have been obscured by the antiquity of the stem lineage. Others have suggested that a recent origin for most extant pine taxa, especially in subg. *Pinus*, is the simplest explanation for low divergence rates (e.g. Strauss and Doerksen 1990; Govindaraju, Lewis, and Cullis 1992). Regardless of which absolute calibration is chosen, the relative divergence times inferred from multiple nuclear and chloroplast loci provide an important perspective for studying pine species

relationships. For example, *relative* divergences between sections in subg. *Pinus* are only one-sixth of the total divergence within the genus.

We conclude that a 45 MYA subgeneric divergence may be too young, but yields an upper limit for *Pinus* evolutionary rates. Because a subgeneric calibration at 85 MYA based on permineralized wood yields realistic age projections for both older and younger nodes, it provides a useful lower rate limit. Together, these rate estimates (Table 2.2) reveal a moderate tempo for pine divergence, and provide a framework that can be used to compare future conifer gene- and taxon-specific rates.

SUPPLEMENTARY MATERIAL

Supplementary Figure 2.S1 (the alignment file) is available at *Molecular Biology Evolution* online (http://www.mbe.oxfordjournals.org/)

ACKNOWLEDGMENTS

We thank J. Berdeen, D. Johnson, W. Kwan-Soo, and M. McGregor for providing plant collections; G. Brown, S. Gonzalez-Martinez, K. Krutovsky, D. Neale, and C. Plomion for sharing linkage maps and primer sequences; D. Erwin, S. Manchester, G. Poinar, and J. van der Burgh for advice on fossils; S. Muse, D. Remington, R. Small, J. Wendel, and two anonymous reviewers for helpful comments; K. Farrell, C. Streng, and O. Zerón Flores for laboratory assistance. Funding for this study was provided by National Science Foundation grant DEB 0317103 to AL and RC, Secretaria de Educación Pública grant to DSG, and the USDA Forest Service Pacific Northwest Research Station.

Table 2.1. Length (L), noncoding length (NC); silent (dS; synonymous plus noncoding), and nonsynonymous (dN) substitutions per site across eleven nuclear and four chloroplast genes within subg. *Pinus*, subg. *Strobus*, and across subgenera.

	Section *Trifoliae – Pinus*				**Section *Quinquefoliae – Parrya***				**Subg. *Pinus – Strobus***			
Locus	**L**	**NC**	**dS**	**dN**	**L**	**NC**	**dS**	**dN**	**L**	**NC**	**dS**	**dN**
nDNA												
4CL	996	321	0.040	0.002	740	110	0.083	0.009	700	75	0.189	0.013
AGP6	444	0	0.047	0.003	465	0	0.086	0.016	444	0	0.147	0.015
aquaporin	653	338	0.039	0.000	624	311	0.058	0.004	571	258	0.100	0.003
CAD	360	246	0.050	0.000	612	498	0.075	0.011	272	158	0.124	0.017
ERD3 [a]	656	202	0.162	0.046	687	234	0.070	0.015	634	198	0.104	0.024
ferritin	767	367	0.042	0.018	725	517	0.051	0.033	690	482	0.129	0.039
LEA-like	1115	1037	0.040	0.018	967	889	0.051	0.000	386	308	0.087	0.146 [a]
LHC-CAB	701	0	0.072	0.006	701	0	0.091	0.017	701	0	0.205	0.029
OST1	641	476	0.036	0.000	642	477	0.064	0.000	640	475	0.063	0.000
PAL1	409	0	0.021	0.010	409	0	0.065	0.016	409	0	0.142	0.029
SOD_{chl} [a]	852	658	0.060	0.043	194	6	0.153	0.153	200	5	0.127	0.129
mean [b]	**690**	**331**	**0.042**	**0.005**	**615**	**277**	**0.063**	**0.013**	**513**	**178**	**0.118**	**0.020**
(S.D.)			(0.013)	(0.007)			(0.015)	(0.010)			(0.045)	(0.013)
cpDNA												
matK	1539	107	0.008	0.013	1545	113	0.020	0.013	1545	113	0.039	0.036
rbcL	1262	0	0.023	0.004	1262	0	0.023	0.004	1262	0	0.064	0.005
rpL20/S18	540	254	0.003	0.016	521	231	0.011	0.008	511	221	0.051	0.012
trnV	546	543	0.007	0.000 [a]	549	543	0.009	0.000 [a]	548	542	0.015	0.000 [a]
mean [b]	**972**	**226**	**0.010**	**0.009**	**969**	**222**	**0.015**	**0.009**	**967**	**219**	**0.038**	**0.021**
(S.D.)			(0.008)	**(**0.005)			(0.006)	(0.004)			(0.019)	(0.014)
Total	**11481**	**4549**			**10643**	**3929**			**9513**	**2835**		

[a] Values are excluded from mean due to possible paralogy (SOD_{chl}, *ERD*3), short partition length (*trnV*), or unusually high dN (> 9 S.D. from mean; *LEA-like*); [b] dS and dN means are weighted by number of sites

Table 2.2. Estimated absolute mutation rates (μ; substitutions·site^{-1}·year^{-1}) based on silent (dS) and nonsynonymous (dN) substitutions in comparisons between subg. *Pinus* and subg. *Strobus*. Rates are averages of nine nuclear loci or four chloroplast loci, using divergence dates of 85 or 45 MYA. Values in parentheses are one standard deviation.

			μ (× 10^{-9})	
Genome	**rate**	**d**	**T = 85 MYA**	**T = 45 MYA**
Nuclear	**dS**	0.118 (0.045)	0.70 (0.27)	1.31 (0.50)
	dN	0.020 (0.013)	0.12 (0.04)	0.22 (0.08)
Chloroplast	**dS**	0.038 (0.018)	0.22 (0.11)	0.42 (0.20)
	dN	0.021 (0.016)	0.12 (0.06)	0.23 (0.11)

Table 2.3. Estimated divergence dates for *Pinus* nodes shown in Figure 2.1, based on silent sites (or all sites) from nuclear or chloroplast DNA. Calibration is at node C at 85 or 45 MYA. Ages and silent mutation rates (μ) are estimated using ML branch lengths for data showing rate equality, or using PL for the non-clock-like 12-taxon data set.

				Ages (MYA) for nodes				μ
Data set [a]	**Outgroup**	**Method**	**sites**	**A**	**B**	**C**	**D**	**(× 10^{-9})**
Nuclear								
5 loci–4 taxa	*Picea*	ML	silent	30	48	85	190	0.75
			all	29	47		172	---
			silent	16	25	45	102	1.41
			all	15	25		91	---
9 loci–4 taxa	Multiple [b]	ML	silent	31	48	85	n/a	0.66
				16	25	45	n/a	1.24
4 loci–12 taxa	*Picea*	PL	silent	28	37	85	n/a	0.74
				15	20	45	n/a	1.39
Chloroplast								
4 loci–4 taxa	*Picea*	ML	silent	25	37	85	164	0.21
			all	27	33		136	---
			silent	13	19	45	87	0.40
			all	14	18		72	---

[a] Data set composition described in text

Table 2.S1. Loci included in this study, including genetic linkage map information, amplification primers and conditions, and relationship to other GenBank accessions.

Locus	LG[a]	Mapped species	Primer Sequences (5' – 3')	Primer source	PCR[b]	Amplicon description
4CL	LG7	*P. sylvestris* (Komulainen et al. 2003); *P. taeda* (Krutovsky et al. 2004); *Pseudotsuga menziesii* (Krutovsky et al. 2004)	CAGAGGTGGAAYTGATTTC CTGCTTCTGTCATGCCGTA GCSAATCCTTTCTACAAGC GCCAATCCTTTTTACAAGC CTGTTAGATGCAAATCAGGTAG	(Syring et al. 2005)	2.0 mM 52 °C	Exon 1 and intron 1 (pos. 461 to 1531) of *P. taeda* complete genomic (PTU39405)
AGP6	LG5	*P. taeda* (Chagné et al. 2003); *P. pinaster* (Chagné et al. 2003)	TCAGGGTCAACAATGGCGTTC GGGCTTTTCAGTGCGGACG	(Syring et al. 2005)	2.5 mM 56 °C	All exonic; pos. 48 to 509 of *P. taeda* mRNA (AF101785)
Aqua porin	LG2	*P. taeda* (Temesgen et al. 2001); *P. pinaster* (Chagné et al. 2003)	TGTCACTGCCCAGAGCTATTC ATCACAGCCGCTCCAAAAC	PtIFG464; (Temesgen et al. 2001)	2.0 mM 52 °C	Pos. 111 to ca. 661 bp downstream of *P. taeda* EST (PtIFG464; H75150); pos. 111 to 701 in *Picea mariana* complete mRNA (AF051202). Intron has substantial length variation in *Pinus* and *Picea* and is absent in *Pinus banksiana* (Perry and Bousquet 1998a; Perry and Bousquet 1998b)

CAD	LG9	*P. taeda* (Krutovsky et al. 2004); *P. pinaster* (Chagné et al. 2003); *Pseudotsuga menziesii* (Krutovsky et al. 2004)	GTCCCCTTACACTTACAATC ATCCCCACCACTTCAT	http://www.pierroton.inra.fr/genetics/pinus/primers.html	2.0 mM 50 °C	Intron 1, exon 1, and intron 2 of genomic *P. radiata*: pos. 104—474 of 3540 bp; AF60491)
ERD3	Not mapped	n/a	CCCATGCGGTCTTGTATAGT ATCCTGGGAAAATGAAACG	(Plomion et al. 1999; González-Martinez, personal communication)	2.0 mM 55 °C	Extends from 17 bp upstream to pos. 873 of *P. taeda* genomic (AY874639)
ferritin	LG5	*P. pinaster* (Chagné et al. 2003)	ATGGAGCTGACTTTGTCTTT GAGGGCTAATTGAGAAACAG	http://www.pierroton.inra.fr/genetics/pinus/primers.html	2.5 mM 55 °C	4 exons, 3 introns, and 163 bp (subg. *Pinus*) to 139 bp (subg. *Strobus*) of 3' UTR; pos 183—589 of *P. taeda* cDNA AF028072). Translation has high a.a. similarity to *A. thaliana* (At5g01600), although only *P. nelsonii* shows the same stop codon location; the other three species include an additional 27 bp that encode the amino acids ADAGALAAA
LEA-like	LG3	*P. taeda* (Krutovsky et al. 2004)	TGTTAGCATGCAATCAATCAC CTGACAGAGTGGGCAGCTTCAT GAGTGCAAATAGCTACTC TCAGG CAACAATAGGAATGTGCATGGTAAGG	PtIFG8612 (Syring et al. 2005)	2.0 mM 55 °C	Exon 1 (pos. 51 to 128 of *P. taeda* EST AA739606), and variable-length intron
LHC-CAB	LG3	*P. taeda* (Temesgen et al. 2001)	CTACTGCTGCAACAATGGCAA TAGGCCCAGGCGTTGTTGTT	PtIFG1934 (Syring et al. 2005)	1.5 mM 52 °C	All exonic; pos 1149 to 1849 of *P. contorta* full genomic (X67714)

OST1	LG9	*P. taeda* (Temesgen et al. 2001)	CACAATTGCCAGATGGGTC CTTCTCTAGCAACGATCCGG	PtIFG624 (Temesgen et al. 2001; Mustilli et al. 2002)	2.5 mM 55 °C	275 bp of 5' UTR; exon 1 and intron 1 of *P. taeda* cDNA (PtIFG624); BLASTN = 82% similar to *A. thaliana* (AJ316009; Mustilli et al. 2002)
PAL1	LG6	*P. sylvestris* (Komulainen et al. 2003)	TAGCCAAGAAAACCCTGAG ACTGATAGCGTCGTAAACCA	(Plomion et al. 1999)	1.5 mM 52 °C	Pos. 636 to 2044 of *P. taeda* genomic (PTU39792).
SOD chl	LG10	*P. pinaster* (Chagné et al. 2003); *P. sylvestris* (Komulainen et al. 2003)	TCCGTTTGACAGGATTGACT CCCCAGGTCATCCTCTAACT	(Plomion et al. 1999)	1.5 mM 50 °C	Extends 70 bps upstream and 105 bps downstream of *P. taeda* genomic (AY866724).
matK	cpDNA	n/a	(Gernandt et al. 2005)	(Gernandt et al. 2005)	(Gernandt et al. 2005)	Pos. 1608 to 3146 of *P. thunbergii* complete cp (NC001631)
rbcL	cpDNA	n/a	(Gernandt et al. 2005)	(Gernandt et al. 2005)	(Gernandt et al. 2005)	Pos. 43145 to 44406 of *P. thunbergii* complete cp (NC001631)
rpL20/ S18	cpDNA	n/a	CTTCGTCGTTTGTGGATTAC AGTCGATTTATTAGTGAGCA	(Wang et al. 1999)	(Gernandt et al. 2005)	*rpL20* exon, intergenic spacer, *rpS18* exon
trnV	cpDNA	n/a	GTAGAGCACCTCGTTTACAC CTCGAACCGTAGACCTTCTC	(Wang et al. 1999)	(Gernandt et al. 2005)	$trnV^{GUA}$, 3 to 6 bp of exon; remainder is intron

[a] published linkage groups (LG) correspond to *P. taeda* map (Brown et al. 2001)
[b] $MgCl_2$ concentration and annealing temperature. Cycling parameters for all loci followed Syring et al. (2005)

Table 2.S2. GenBank accession numbers for *Pinus* and outgroup taxa.

Locus	***P. taeda***	***P. thunbergii***	***P. monticola***	***P. nelsonii***	**Outgroup, Accession**
4CL	U39405	DQ370109	AY634357	AY634362	*Picea sitchensis*, AY634363
AGP6	AF101785	*P. merkusii, s.l.* AY634320	AY634325	AY634330	*Picea sitchensis*, AY634331
aquaporin	DQ370110	DQ370111	DQ370112	DQ370113	*Picea abies*, Z93764[a]
CAD	DQ370114	DQ370115	DQ370116	DQ370117	*Picea abies*, AJ001926
ERD3	DQ370118	DQ370119	DQ370120	DQ370121	n/a
ferritin	DQ370122	DQ370123	DQ370124	DQ370125	*Arabidopsis thaliana*, AF229850
LEA-like	AY634332	DQ370126	AY634342	AY634347	*Pseudotsuga menziesii*, AJ012483[a]
LHC CAB	H75103	*P. merkusii, s.l.* AY617085	AY617090	AY617095	*Picea sitchensis*, AY617096
OST1	DQ370127	DQ370128	DQ370129	DQ370130	*Arabidopsis thaliana*, AJ316009
PAL1	DQ370131	DQ370132	DQ370133	DQ370134	*Picea abies*, AY952468[a]
SODchl	DQ370135	DQ370136	DQ370137	DQ370138	n/a
matK	AY724750	NC00163 (pos 1608—3146)	AY497259	AY115793	*Picea sitchensis*, AY035203
rbcL	AY724758	NC00163 (pos 43145—44406)	AY497223	AY11577	*Picea sitchensis*, X63660
rpL20/S18	DQ359718	NC00163 (pos 31404—31945)	DQ359719	DQ359720	*Picea sitchensis*, DQ375446
trnV	DQ359721	NC00163 (pos 47465—48012)	DQ359722	DQ359726	*Picea sitchensis*, DQ375447

[a]expressed sequences; all other sequences are genomic

Table 2.S3. Tree statistics and likelihood ratio tests for each locus and each concatenated data set. Loci showing significant deviation ($\alpha \leq 0.05$) from rate constancy based on likelihood ratio tests (LRT) are indicated by gray shading, with p-values shown in bold.

						Ln likelihood score		LRT				
Genome	Locus	*Pinus* taxa	Outgroup	Sites	bp	clock-relaxed	clock-enforced	df	p	I	shape	Sequence model
nDNA	*4CL*	4	*Picea*	all	1091	2320	2321	2	0.225	0	n/a	HKY
nDNA	*4CL*	4	*Picea*	silent	641	1484	1487	2	0.068	0	n/a	HKY85
nDNA	*4CL*	12	*Picea*	all	1315	3067	3079	10	**0.006**	0	0.645	HKY85+G
nDNA	*4CL*	12	*Picea*	silent	865	2128	2142	10	**0.003**	0	2.815	HKY85
nDNA	*AGP6*	4	*Picea*	all	610	1175	1177	2	0.137	0	n/a	HKY85
nDNA	*AGP6*	4	*Picea*	silent	203	400	403	2	0.164	0	n/a	HKY85
nDNA	*AGP6*	12	*Picea*	all	609	1440	1446	10	0.217	0	0.452	GTR+G
nDNA	*AGP6*	12	*Picea*	silent	203	379	381	10	0.979	0	n/a	F81
nDNA	*aquaporin*	4	*Picea*	all	776	1641	1641	2	0.868	0	n/a	HKY85+G
nDNA	*aquaporin*	4	*Picea*	silent	551	1172	1172	2	0.854	0	n/a	GTR
nDNA	*CAD*	4	*Picea*	all	796	1559	1560	2	0.284	0	n/a	HKY85
nDNA	*CAD*	4	*Picea*	silent	720	1415	1416	2	0.355	0	n/a	HKY85
nDNA	*ferritin*	4	*Arabidopsis*	all	847	2001	2004	2	**0.047**	0	0.617	HKY85+G
nDNA	*ferritin*	4	*Arabidopsis*	silent	689	1529	1532	2	0.127	0	0.334	HKY85+G
nDNA	*LEA-like*	4	*Pseudotsuga*	all	2447	3972	3975	2	0.121	0	n/a	HKY85
nDNA	*LEA-like*	4	*Pseudotsuga*	silent	2375	3779	3781	2	0.050	0	n/a	HKY85
nDNA	*LEA-like*	12	*Pseudotsuga*	all	2644	5333	5340	10	0.243	0	0.473	HKY85+G
nDNA	*LEA-like*	12	*Pseudotsuga*	silent	2441	4801	4812	10	**0.016**	0	0.379	HKY85+G
nDNA	*LHC-CAB*	4	*Picea*	all	701	1555	1631	2	**0.000**	0	0.012	HKY85+G
nDNA	*LHC-CAB*	4	*Picea*	silent	234	663	665	2	0.169	0	0.493	HKY85+G
nDNA	*LHC-CAB*	12	*Picea*	all	770	2092	2110	10	**0.000**	0.6	0.673	HKY85+G+I
nDNA	*LHC-CAB*	12	*Picea*	silent	257	980	992	10	**0.009**	0	0.434	HKY85+G
nDNA	*OST1*	4	*Arabidopsis*	all	671	1355	1358	2	0.101	0	n/a	HKY85
nDNA	*OST1*	4	*Arabidopsis*	silent	551	1144	1145	2	0.286	0	0.511	K2P+G
nDNA	*PAL1*	4	*Picea*	all	409	772	776	2	**0.029**	-	n/a	K2P
nDNA	*PAL1*	4	*Picea*	silent	136	295	299	2	**0.024**	0	n/a	K2P

nDNA: 5 loci	*4CL*, *AGP6*, *aquaporin*, *CAD*, *LHC-CAB*	4	*Picea*	all	3973	8449	8452	2	0.101	0	0.505	HKY85+G
				silent	2348	5292	5293	2	0.247	0	1.600	GTR+G+I
nDNA: 9 loci	*4CL*, *AGP6*, *aquaporin*, *CAD*, *LHC-CAB*, *ferritin*, *LEA-like*, *OST1*, *PAL1*	4	*Picea*, *Pseudotsuga*, *Arabidopsis*	all	8346	16694	16698	2	**0.009**	0	0.526	HKY85+G
				silent	6098	12116	12118	2	0.096	0	1.279	HKY+G
nDNA: 4 loci	*4CL*, *AGP6*, *LEA-like*, *LHC-CAB*	12	*Picea*	all	5338	12166	12180	10	**0.001**	0	0.290	HKY85+G
				silent	3766	8497	8508	10	**0.021**	0	0.550	HKY85+G
cpDNA	*matK*	4	*Picea*	all	1545	2918	2918	2	0.586	0	0.526	HKY85+G
cpDNA	*matK*	4	*Picea*	silent	591	1135	1137	2	0.176	0	0.526	HKY85+G
cpDNA	*rbcL*	4	*Picea*	all	1262	2117	2117	2	0.880	0	0.526	HKY85+G
cpDNA	*rbcL*	4	*Picea*	silent	421	810	811	2	0.441	0	0.526	HKY85+G
cpDNA	*rpL20/S18*	4	*Picea*	all	546	1061	1062	2	0.680	0	0.526	HKY85+G
cpDNA	*rpL20/S18*	4	*Picea*	silent	356	736	737	2	0.220	0	0.526	HKY85+G
cpDNA	*trnV*	4	*Picea*	all	550	920	922	2	0.243	0	0.526	HKY85+G
cpDNA	*trnV*	4	*Picea*	silent	546	915	916	2	0.243	0	0.526	HKY85+G
cpDNA: 4 loci	*matK*, *rbcL*, *rpl20/S18*, *trnV*	4	*Picea*	all	3903	7042	7043	2	0.592	0	0.526	HKY85+G
				silent	1914	3619	3619	2	0.587	0	0.526	HKY85+G

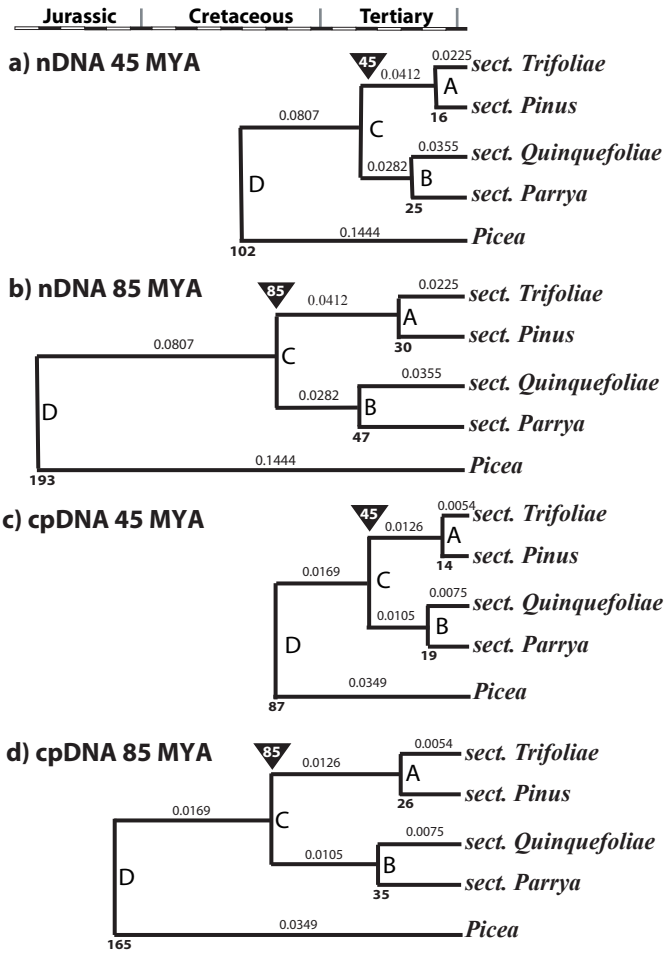

Figure 2.1. ML chronograms of major lineages of *Pinus* using silent sites from: five nuclear loci (2a, 2b; 2348 bp) or four chloroplast loci (2c, 2d; 1914 bp). Calibration is at node C with either 45 based on leaves (2a, 2c) or 85 based on wood (2b, 2d) MYA. Branch lengths are shown above each branch, estimated ages are shown below each node. See Supplementary Material online for tree statistics.

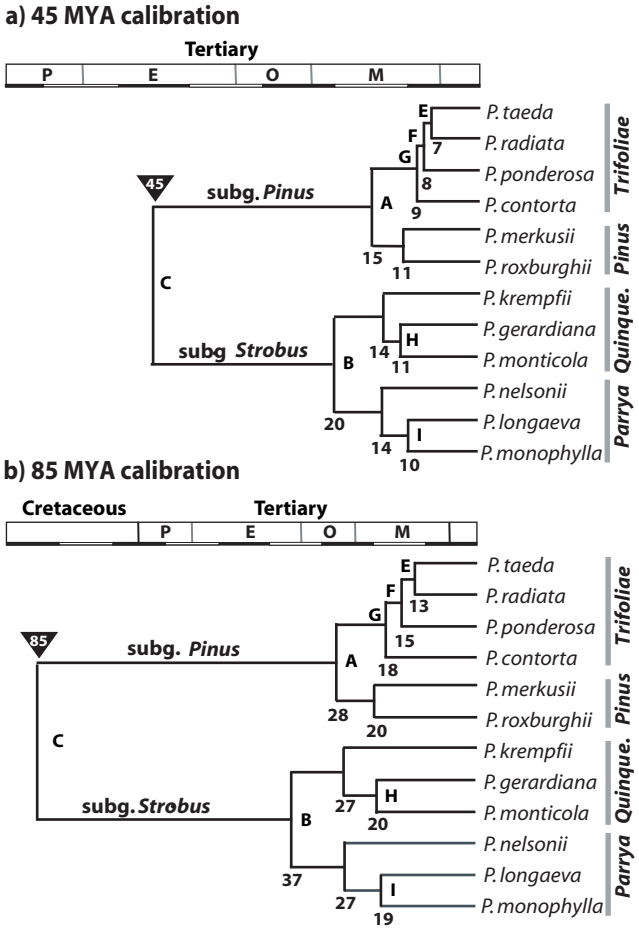

Figure 2.2. PL chronograms of 12 taxa using silent sites from four nuclear loci (3766 bp), based on calibration at node C with a) 45 or b) 85 MYA. Estimated ages are shown below each node. See Supplementary Material online for tree statistics. Nodes E-I represent subsections referred to in the text: E-Australes; F-Ponderosae; G-Contortae; H-Strobus; I-Balfourianae/Cembroides. All nodes have greater than 70% bootstrap support except f or nodes E and F, which collapse in the strict consensus tree.

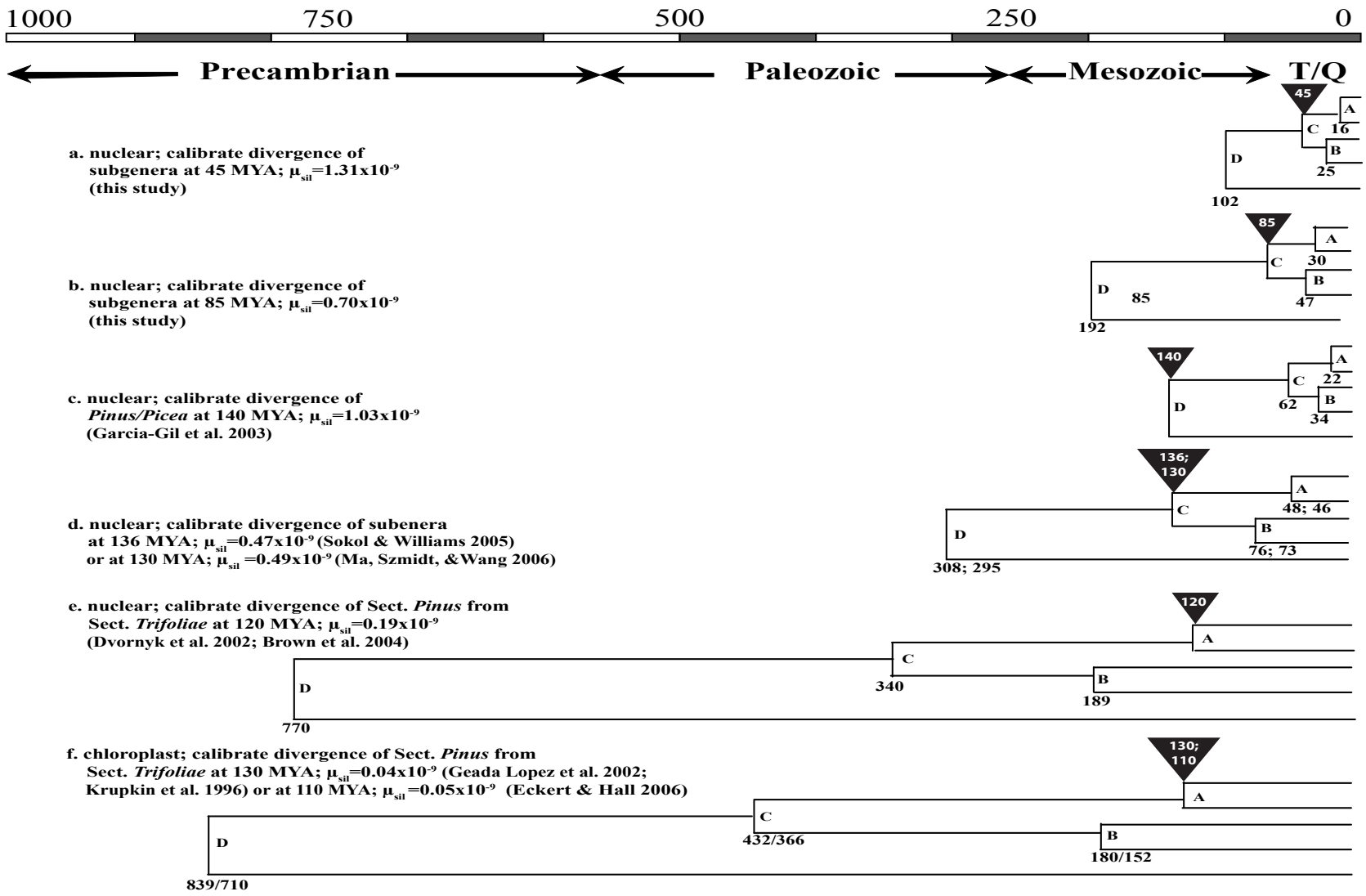

Figure 2.3. Application of some recently published fossil calibration scenarios (see Discussion), using clock-enforced branch lengths (shown in Fig. 1) from our nDNA or cpDNA data sets. ML-projected ages are shown below each node; ML-calculated silent rates are given for each scenario. Node letters correspond to Fig. 2.1.

LITERATURE CITED

Alvin K. 1960. Further conifers of the Pinaceae from the Wealden Formation of Belgium. Institut Royal des Sciences Naturelles de Belgique, Mémoires 146:1-39.

Axelrod D. 1986. Cenozoic history of some western American pines. Ann Mo Bot Gard 73:565-641.

Blackwell W. 1984. Fossil ponderosa-like pine wood from the Upper Cretaceous of northeast Mississippi. Ann Bot London 53:133-136.

Brown G, Gill G, Kuntz R, Langley C, Neale D. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101:15255-15260.

Brown G, Kadel E, Bassoni D, Kiehne K, Temesgen B, van Buijtenen J, Sewell M, Marshall K, Neale D. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159:799-809.

Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C. 2003. Comparative genome and QTL mapping between maritime and loblolly pines. Mol Breeding 12:185-195.

Clark R, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Mol Biol Evol 22:2304-2312.

Cunningham, C. 1997. Can three incongruence tests predict when data should be combined? Mol Biol Evol 14:733-740.

Dvornyk V, Sirviö A, Mikkonen M, Savolainen O. 2002. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19:178-188.

Eckert A, Hall B. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenet Evol 40:166-182.

Erwin D, Schorn H. 2006. *Pinus baileyi* (Section *Pinus*, Pinaceae) from the Paleogene of Idaho, USA. Am J Bot 93:197-205.

Fliche P. 1896. Etude sur la flore fossile de l'Argonne (Albien-Cénomanien), Nancy, France.

García-Gil M, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12:1195-1206.

Gaut B, Morton B, McCaig B, Clegg M. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA. 93:10274-10279.

Gaut B, Muse S, Clark W, Clegg M. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J Mol Evol 35:292-303.

Geada López G, Kamiya K, Harada K. 2002. Phylogenetic relationships of Diploxylon pines (subgenus *Pinus*) based on plastid sequence data. Int J Plant Sci 163:737-747.

Gernandt D, Geada López G, Ortiz García S, Liston A. 2005. Phylogeny and classification of *Pinus*. Taxon 54:29-42.

Govindaraju D, Lewis P, Cullis C. 1992. Phylogenetic analysis of pines using ribosomal DNA restriction fragment polymorphisms. Plant Syst Evol 179:141-153.

Gradstein F, Ogg J. 2004. Geologic time scale 2004: why, how and where next! Lethaia 37:175-181.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Hart J. 1987. A cladistic analysis of conifers: preliminary results. J Arnold Arboretum 68:269-307.

Haubold B, Wiehe T. 2001. Statistics of divergence times. Mol Biol Evol 18:1157-1160.

Hudson R. 1960. The anatomy of the genus *Pinus* in relation to its classification. J I Wood Sci 6:26-46.

Jeffrey E. 1908. On the structure of the leaf in Cretaceous pines. Ann Bot London 22:207-220.

Kärkkäinen K, Koski K, Savolainen O. 1996. Geographical variation in the inbreeding depression of Scots pine. Evolution 50:111-119.

Kay K, Whittall J, Hodges S. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. Evol Biol 6:36-44.

Klekowski E. 1992. Review: mutation rates in diploid annuals-are they immutable? Int J Plant Sci 153:462-465.

Klekowski E. 1998. Mutation rates in mangroves and other plants. Genetica 102:325-331.

Klekowski E, Godfrey P. 1989. Ageing and mutation in plants. Nature 340:389-391.

Koch M, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17:1483-1498.

Komulainen P, Brown G, Mikkonen M, Karhu A, García-Gil M, O'Malley D, Lee B, Neale D, Savolainen O. 2003. Comparing EST based genetic maps between *Pinus sylvestris* and *P. taeda*. Theor Appl Genet 107:667-678.

Kossack D, Kinlaw C. 1999. *IFG*, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39:417-426.

Koteja J, Poinar G. 2001. A new family, genus, and species of scale insect (Hemiptera: Coccinea: Kukaspididae, new family) from Cretaceous Alaskan amber. Proc Entomol Soc Wash 103:356-363.

Krupkin A, Liston A, Strauss S. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot 83:489-498.

Krutovsky K, Troggio M, Brown G, Jermstad K, Neale D. 2004. Comparative mapping in the Pinaceae. Genetics 168:447-461.

Kumar S, Filipski A, Swarna V, Walker A, Hedges S. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA. 102:18842-18847.

Kutil B, Williams C. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J Hered 92:327-332.

Lande R, Schemske D, Schultz S. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evolution 48:965-978.

Langenheim R, Smiley C, Gray J. 1960. Cretaceous amber from the Arctic coastal plain of Alaska. Bulletin of the Geologic Society of America 71:1345-1356.

Ma X-F, Szmidt AE, Wang X-R. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol 23:807-816.

Magallón S, Sanderson M. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762-1780.

Magallón S, Sanderson M. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution 59:1653-1670.

Mai D. 1986. Über typen und originale tertiärer arten von *Pinus* L. (Pinaceae) in mitteleuropÜäischen sammlungen - ein beitrag zur geschichte der gattung in Europa. Feddes Repertorium 97:571-605.

Meijer J. 2000. Fossil woods from the Late Cretaceous Aachen Formation. Rev Palaeobot Palyno 112:297-336.

Millar C. 1998. Early evolution of pines. In: Richardson D, ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge, U.K. p. 69-91.

Miller C. 1973. Silicified cones and vegetative remains of *Pinus* from the Eocene of British Columbia. Contributions from the Museum of Paleontology, The University of Michigan 24:101-118.

Miller C. 1976. Early evolution in the Pinaceae. Rev Palaeobot Palyno 21:101-117.

Miller C, Malinky J. 1986. Seed cones of *Pinus* from the Late Cretaceous of New Jersey, U.S.A. Rev Palaeobot Palyno 46:257-272.

Muse S. 2000. Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol 42:25-43.

Muse S, Weir B. 1992. Testing for equality of evolutionary rates. Genetics 132:269-276.

Mustilli A-C, Merlot S, Vavasseur A, Fenzi F, Giraudat J. 2002. *Arabidopsis OST1* protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell 14:3089-3099.

Near T, Meylan P, Shaffer H, 2. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat 165:137-146.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.

Penny J. 1947. Studies on the conifers of the Magothy flora. Am J Bot 34:281-296.

Perry DJ, Bousquet J. 1998a. Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. Genetics 149:1089-1098.

Perry DJ, Bousquet J. 1998b. Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor Appl Genet 97:735-743.

Phipps C, Osborn J, Stockey R. 1995. *Pinus* pollen cones from the middle Eocene Princeton Chert (Allenby Formation) of British Columbia. Int J Plant Sci 156:117-124.

Plomion C, Hurme P, Frigerio J-M, Ridolfi M, Pot D, Pionneau C, Avila C, Gallardo F, David H, Neutelings G, et al. 1999. Developing SSCP markers in two *Pinus* species. Mol Breeding 5:21-31.

Posada D, Crandall K. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Prager E, Fowler D, Wilson A. 1976. Rates of evolution in conifers (Pinaceae). Evolution 30:637-649.

Price R, Liston A, Strauss S. 1998. Phylogeny and systematics of *Pinus*. In: Richardson D, ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge, U.K. p. 49-68.

Robison C. 1977. *Pinus triphylla* and *Pinus quinquefolia* from the Upper Cretaceous of Massachusetts. Am J Bot 64:726-732.

Rozas J, Sánchez-DelBarrio J, Messeguer X, Rozas R. 2003. DnaSP, DNA sequence polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Saiki K. 1996. *Pinus mutoi* (Pinaceae), a new species of permineralized seed cone from the Upper Cretaceous of Hokkaido, Japan. Am J Bot 83:1630-1636.

Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101-109.

Sanderson M, Doyle J. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and *18s* rDNA data. Am J Bot 88:1499-1516.

Scofield D, Schultz S. 2006. Mitosis, stature and evolution of plant mating systems: low-Φ and high-Φ plants. Proc Royal Soc Lond B Bot 273:275-282.

Shaw G. 1914. The genus *Pinus*. Riverside Press, Cambridge, MA.

Smith S, Stockey R. 2002. Permineralized pine cones from the Cretaceous of Vancouver Island, British Columbia. Int J Plant Sci 163:185-196.

Sokol K, Williams C. 2005. Evolution of a triplet repeat in a conifer. Genome 48:417-426.

Stockey R, Nishida M. 1986. *Pinus harborensis* sp. nov. and affinities of permineralized leaves from the Upper Cretaceous. Can J Botany 64:1856-1866.

Stopes M, Kershaw E. 1910. The anatomy of Cretaceous pine leaves. Ann Bot London 24:395-402.

Strauss S, Doerksen A. 1990. Restriction fragment analysis of pine phylogeny. Evolution 44:1081-1096.

Swofford D. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods). Sinauer Assoc., Sunderland, MA.

Syring J, Farrell K, Businský R, Cronn R, Liston A. 2006. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst Biol 56:163-181.

Syring J, Willyard A, Cronn R, Liston A. 2005. Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot 92:2086-2100.

Temesgen B, Brown G, Harry D, Kinlaw C, Sewell M, Neale D. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). Theor Appl Genet 102:664-675.

Thompson J, Higgins D, Gibson T. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.

Van der Burgh J. 1973. Hölzer der niederrheinischen braunkohlenformation, 2. hölzer der braunkohlengruben "Maria Theresia" zu herzogenrath, "zukunft west" zu eschweiler und "victor" (zülpich mitte) zu zülpich. nebst einer systematisch-anatomischen bearbeitung der gattung *Pinus* L. Rev Palaeobot Palyno 15:73-275.

Wang X-Q, Tank D, Sang T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol 17:773-781.

Wang X-R, Tsumura Y, Yoshimaru H, Nagasaka K, Szmidt AE. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. Am J Bot 86:1742-1753.

Wolfe J, Schorn H. 1989. Taxonomic revision of the Spermatopsida of the Oligocene Creede Flora, southern Colorado. U.S. Geological Survey Bulletin 1923:1-40.

Wolfe K, Li W-H, Sharp P. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054-9058.

Wolfe K, Sharp P, Li W-H. 1989. Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29:208-211.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 23:212-226.

Phylogenetic Inference for Reticulate Evolution in Subsection *Ponderosae*

(Pinaceae; *Pinus*; Subgenus *Pinus;* Section *Trifoliae*)

Ann Willyard, Aaron Liston, and Richard Cronn

Prepared for submission

INTRODUCTION

Despite persistent attempts to classify *Pinus* species based on morphological, biochemical, and cytological characters, as well as a growing number of molecular phylogenetic studies, taxonomic relations among many species remain unsolved. Our understanding of the causes, if not always the solutions, has been gradually improving. First, *Pinus* is an ancient genus, diverging at least 100 million years ago in the Cretaceous (Willyard et al. 2007). Nevertheless, recent genetic and fossil evidence suggests that most of the ca. 100 species of pine arose rather recently. For example, the 20 or so subsect. *Ponderosae* (subg. *Pinus*, Section *Trifoliae*) species, are likely to have diverged within the last 8 to 15 million years (Willyard et al. 2007). Thus, we should expect pines to share inference problems with other recently-diverged, "young" species.

Second, we expect species-level evolutionary relationships to be a serious conundrum in a genus where morphological character states are plastic as well as homoplasious. These factors are well documented in *Pinus* (Mirov 1929; Gernandt et al. 2005). Examples of plasticity from *P. ponderosa* include substantial differences in leaves growing on the shady side of trees, on older trees, and on trees exposed to strong winds (Helmers 1943). Serotonous cones, identified by Shaw (1914) as a character state that "should convey taxonomic significance of some kind" have been found to be a result of convergent evolution in several lineages (Gernandt et al. 2005).

Third, molecular phylogenies have revealed that lineage sorting between pine species is often incomplete. When multiple pine individuals are sampled per species, there are numerous cases of conspecific samples which lack monophyly, and this has been attributed to incomplete lineage sorting that leads to the observed retention of ancestral polymorphism (Syring et al. 2007). Because loci coalesce at different rates and are affected by stochastic processes, different relationships may be inferred from different loci (Syring et al. 2005). Thus, conflicting tree topologies are not unexpected when the same taxa are sampled using different types of evidence. Within the nuclear ribosomal internal transcribed spacer (nrITS), intra-individual polymorphism and presumed recombination yield conflicting phylogenetic estimates from chimeric regions of the spacer (Gernandt, Liston, and Pinero 2001). Chloroplast phylogenies reveal some conflicts with those based on nrITS, as well as a general lack of species-level resolution. For example, a strict consensus tree built over

combined *matK* and *rbcL* sequences resulted in one 11-species polytomy and one three-species polytomy for 14 species of *Ponderosae* (Gernandt et al. 2005). Many of the factors that contribute to a lack of genetic coalescence within species are present in woody trees in general (and specifically in *Pinus*): mating systems that are predominantly outcrossing, high within-species (and within-population and within-individual) heterozygosity, long generation time, and large effective population sizes where alleles are rarely purged. The retention of ancestral polymorphism may be especially troubling in *Pinus* because the speciation has been rapid in comparison to the effective population size (Willyard et al. 2007). Abundant within-population variation appears to be the norm; generally more than 90% of variation is contained among vs. between populations, whether the evidence is from quantitative traits, allozymes, or molecular markers (reviewed by Conkle and Critchfield 1988; Ledig 1998; also see Sorensen, Mandel, and Aagaard 2001), although the limited number of pine species with fragmented populations can exhibit much more differentiation among populations (Ledig 1998).

Fourth, hybridization between diverged lineages is an important factor in the evolution of many plants (Anderson and Stebbins 1954; Stebbins 1959; Ellstrand and Schierenbeck 2000). Although evidence is minimal for reticulate evolution in *Pinus*, many have speculated about how reticulation may have influenced the observed complexities, e.g. the "intricate and reticulate evolutionary history" of *Pinus* (Styles 1993). Whether this process is viewed as intra- or inter-specific introgression (or slow-motion speciation) is largely subjective, depending on what level of coalescence along a continuum is deemed distinctive. However, current introgression is an important consideration in species relationships, as it may be difficult to distinguish introgression from ancient reticulation when we observe genetic evidence for admixture. Furthermore, introgression might provide migrant alleles that increase the diversity within populations and slow the process of allelic coalescence.

Secondary contact between divergent *Pinus* taxa has only been well documented in a few cases where the contact is extremely recent, e.g. between varieties of *P. ponderosa*; (Johansen and Latta 2003), between *P. ponderosa* var. *scopulorum* and *P. arizonica* (Epperson et al. 2001; Epperson, Chung, and Telewski 2003) and between *P. palustris* and *P. taeda* after disturbance from intensive logging (Namkoong 1966). A reticulate origin has been hypothesized in diploid hybrid speciation events that formed *P. densata* Masters (subg.

Pinus, Sect. *Pinus*) during the uplift of the Tibetan Plateau (Ma, Szmidt, and Wang 2006 and references therein), although the evidence presented in the *P. densata* example may not exclude the retention of ancestral polymorphism shared with extant species in addition to the putative parents as an alternative explanation.

Elsewhere, natural pine hybrids have been documented between sympatric species pairs, e.g. *P. coulteri* and *P. jeffreyi* (morphology, Zobel 1951; Libby 1958); *P. hartwegii* and *P. montezumae* (morphology and chloroplast, Matos 1995; Matos and Schaal 2000); *P. brutia* and *P. halepensis* (cpSSRs; Bucci et al. 1998); and *P. mugo* and *P. sylvestris* (morphology; Christensen and Dar 1999). Putative natural hybrids have been reported (but not confirmed genetically) between many other sympatric species. With the exception of the pinyon pines *P. edulis* and *P. monophylla* (Lanner and Phillips 1992), introgression is generally localized geographically.

Artificial hybridizations demonstrate that incomplete mating barriers between allopatric *Pinus* species are also common. Interspecific crosses between many pine species within taxonomic subsections are fertile (e.g. Critchfield 1986). If these currently allopatric species come into secondary contact over evolutionary timescales, reticulate ancestry in *Pinus* might actually be the norm rather than the exception. In support of this possibility, there is substantial evidence that pine species may, in fact, be quite mobile (reviewed in Petit et al. 2004). For example, *P. ponderosa* is currently considered an alien invader from recent introductions to Argentina, Australia, Chile, and New Zealand, and there are similar examples across the genus (Richardson and Higgins 1998). This suggests that populations of pines may be capable of expanding rather quickly in response to changing climate or habitat opportunities. It seems, then, that the opportunity for reticulate evolution within *Pinus* is provided by incomplete mating barriers and the ability of a mobile species to come into secondary contact. Further evidence regarding reticulate events will help understand the impact that this hypothetical process may have had on *Pinus* lineages.

Thus, we have a fair idea of the main causes for a lack of species-level resolution in the *Ponderosae*, a species-rich subsection of *Pinus*: young species, high variability within populations and species, slow genetic coalescence leading to retention of ancestral polymorphism, and the potential for reticulate ancestry. This study combines multiple samples of almost every *Ponderosae* species for two independent nuclear loci to unravel part of this complicated evolutionary story.

***Ponderosae* Taxonomy.** Species delineations, as well as relationships, in *Ponderosae sensu lato* (*s.l.*; including subsect. *Sabinianae*; Table 3.1) remain unsettled. About seven taxa are predominantly distributed in the western United States, with ranges extending into southwestern Canada and into northwestern Mexico. The remaining thirteen *Ponderosae* taxa occur predominantly in the species-rich mountain ranges of Mexico and Central America, an area that supports about 45% of the world's pine species (Styles 1993).

The type species, *P. ponderosa,* grows from western Canada into Mexico (Critchfield and Little 1966) or almost into Mexico (Rehfeldt et al. 1996) and from sea level to altitudes approaching 3100 m (Mirov 1967) when considered in the broadest sense. However, some uncertainty between only partially-delineated geographic varieties and "races" has endured in this wide-ranging species (or species-complex) since David Douglas' collection in 1836 introduced the species to European botanists (Lauria 1996). I treat two taxa that are closely affiliated with *P. ponderosa* (*P. arizonica* and *P. washoensis*) as distinct species. Rehfeldt (1999a) provided evidence for elevating *P. arizonica* from its varietal status under *P. ponderosa*. Similarly, the narrowly endemic *P. washoensis* has been suggested to fit within *P. ponderosa* by some (Lauria 1997), but is recognized as a distinct species by the Flora of North America (Kral 1993). *Pinus jeffreyi* is also considered to be a close affiliate of *P. ponderosa*, but evidence from successful natural and artificial hybridizations with *P. coulteri* (Critchfield 1966) as well as biochemical evidence (Mirov 1961) hint that its relationships are more complicated as well. In Chapter 4 of this dissertation, I test the genetic similarity of *P. jeffreyi*, *P. ponderosa*, and *P. washoensis* using population-level sampling. The remaining three species from the U.S. are the California big-coned pines (*P. sabiniana*, *P. coulteri*, and *P. torreyana*). These three species were once treated as subsection *Macrocarpae* Shaw, synonymous with *Sabinianae* (Loud.)A E Murray (Little and Critchfield 1969). More recently, this subsection was subsumed within *Ponderosae* because maintaining it would have created a paraphyletic *Ponderosae* based on phylogenetic resolution of chloroplast restriction sites (Krupkin, Liston, and Strauss 1996) and chloroplast sequences (Gernandt et al. 2005).

For the Mexican and Central American *Ponderosae*, many species delineations remain controversial, with floristic treatments recognizing as many as 15 taxa at the species level (Perry 1991), as few as 10 species (Martinez 1948), or just seven (Farjon and Styles

1997). Martínez (1948) described Ponderosa, Montezumae, and Pseudostrobus groups within *Ponderosae* that are still useful as working hypotheses. The species Martinez included in these groups are listed in Table 3.1 along with groups suggested by other major floristic treatments. Mirov (1967) adopted Martinez' group names, calling them the "three great pine complexes of America". Little and Critchfield (1969) adopted *Sabinianae* for the California big-coned pines, but did not suggest any other divisions below the level of subsect. *Ponderosae*. Based largely on wood characters, Van der Burgh (1973) combined Martinez' Montezumae and Pseudostrobus groups into subsect. *Pseudostrobi*, maintained *Ponderosae* and *Sabinianae*, and added a monotypic *Torreyanae* as well (Table 3.1). Perry (1991) added Michoacana, Oaxacana, and Rudis to Martinez' three groups, and placed *P. durangensis* in the Ponderosae group (Table 3.1). Farjon and Styles (1997) applied Van der Burgh's combination of the Montezumae and Pseudostrobus groups into a subsect. *Pseudostrobi*, but placed *P. cooperi* under *Ponderosae* and *P. durangensis* under subsect. *Oocarpae.* Price, Liston, and Strauss (1998) accepted a single subsection, but subdivided it into four informal groups. They agreed with the inclusion of *P. cooperi* and *P. durangensis* within the Ponderosae group, and maintained the Sabinianae group for the California big-coned pines (Table 3.1). As a working hypothesis, I have adopted the taxonomic groupings suggested by Price, Liston, and Strauss, placing the recently described *P. yecorensis* with the Pseudostrobus group.

Experiment Overview. Low-copy nuclear, rather than ribosomal, chloroplast or mitochondrial loci were chosen for this species-level phylogeny for several reasons. First, this is the first species-level molecular phylogeny for the *Ponderosae* using the nuclear genome, and will provide an important comparison with broader taxonomic studies based on the chloroplast (Krupkin, Liston, and Strauss 1996; Gernandt et al. 2005) and nuclear ribosomal DNA (nrITS; Liston et al. 1999) and with a *Ponderosae* phylogeny being constructed based on chloroplast sequences (Gernandt, pers. comm.). Second, intra-individual variation in nrITS (Liston et al. 1999; Gernandt, Liston, and Pinero 2001) limits the usefulness of this genomic region in *Pinus*. Third, mitochondrial genomes exhibit substantial divergence in *Pinus*, but are affected by the "rearrangement of multiple, convergent subgenomic domains" (Wu, Krutovskii, and Strauss 1998), complicating the use of mitochondrial loci for phylogenetic comparisons in *Pinus*. Fourth, the use of two unlinked

low-copy nuclear loci allows us to create two independent phylogenies. This is not possible using the completely linked chloroplast genome. Because of factors already described that affect the coalescence of young species, I consider independent estimates to be an important factor for *Pinus* phylogenetic inference. Fifth, coalescent theory predicts that given the same effective population sizes, faster-evolving loci will coalesce sooner than slower-evolving loci (Rosenberg 2003). If rates were comparable, the two-fold smaller effective population size of the haploid chloroplast genome compared to the nuclear genome would lead to faster predicted coalescence for chloroplast loci (Birky, Maruyama, and Fuerst 1983). However, *Pinus* chloroplast substitution rates are estimated to be ca. three-fold slower than *Pinus* nuclear loci (on the order of 0.04 vs. 0.12 substitutions per site, respectively; Willyard et al. 2007), suggesting that without the complication of recombination, nuclear loci may coalesce faster than chloroplast loci in *Pinus*. These advantages come at an unknown cost when recombination is considered. Finally, interspecific introgression of chloroplast organelles (without observed morphological or ecological differences) between sympatric *Pinus* species has been documented (e.g. Matos and Schaal 2000, Hong, Krupkin, and Strauss 1993, Liston et al. in press). This makes chloroplast powerful markers for detecting hybridization, especially when used in conjunction with nuclear markers, but complicates inferences of species trees. After a hybridization event, organelle genealogy may not track species genealogy, and unlike nuclear introgression, this cannot be tested by sampling other chloroplast regions since they are not independent. For this reason, I looked for chloroplast haplotypes across these samples to infer more recent, pollen-mediated gene flow by comparing nuclear and organellar genotypes in the same accession. Because of the low level of divergence that has been reported between *Pinus* species within a subsection (Syring et al. 2005), I selected loci with the highest divergence within subg. *Pinus* for which homology can be inferred from alignments (Syring et al. 2005, Willyard et al. 2007).

I included essentially every species in *Ponderosae* because comprehensive sampling is crucial for inferring relationships at the species level, and I sequenced multiple individuals per species in order to assess reciprocal monophyly. For the purpose of this phylogenetic framework, I sampled taxa that have been treated as varieties only where there is evidence for the distinction of the taxon as a unique species. However, I discuss assignment of my samples to variety where this is informative. The two to six individuals from each species were selected to represent the geographic range wherever possible.

MATERIALS AND METHODS

Plant Materials. My molecular work was based on DNA extracted from the megagametophyes of single seeds, thus allowing the amplification of low-copy nuclear loci without cloning. I sampled 18 *Ponderosae* species (Table 3.1). The geographic sources for collections are given in Table 3.2. I use the state (or locality if needed) to distinguish between multiple samples of each species. Following the treatment of Price et al. (1998), I have fully sampled the subsection, except that *P. nubicola*, sometimes synonymized with *P. pseudostrobus* (Farjon and Styles 1997), is missing from my data set. In addition, I included two species that were not treated separately by Price et al.: *P. arizonica* (listed as *P. ponderosa* var. *arizonica*) and *P. yecorensis* (Debreczy and Racz 1995). I included only one individual of *P. torreyana*. All of my *P. cooperi* individuals were collected in the same population, as were my two *P. donnell-smithii* individuals. I included two *P. douglasiana* individuals from the same Guerrero population in *WD-40*. My two *P. engelmannii* accessions were collected from geographically adjacent locations, so do not adequately sample the range of this species. Based on a published phylogeny (Syring et al. 2005), I selected *P. contorta* Douglas ex Loudon (section *Trifoliae*, subsect. *Contortae*) as the outgroup and used a sequence from the same individual for both loci.

Loci Sequenced. I chose two independent (i.e. unlinked) nuclear loci that yield suitable-sized amplicons and contain large amounts of presumably neutral noncoding sequence (intron and/or 3' untranslated region) using primers anchored in or near exons. My first PCR amplifications used published primers based on expressed sequences: IFG8612 (GenBank AA739606) was linkage-mapped to *P. taeda* linkage group 3 (Krutovsky et al. 2004) and IFG8898 (GenBank AA739796) was mapped to *P. taeda* linkage group 4 (Temesgen et al. 2001).

The IFG8612 amplicon has the highest similarity (using a nucleotide-nucleotide search in GenBank; http://130.14.29.110/BLAST/) to a genomic sequence for a late embryogenesis abundant-like (*LEA-like*) locus in another member of Pinaceae, *Pseudotsuga menziesii* (PME012483). I used the published translation of this *Pseudotsuga* exon and its intron sequence to infer that my amplicon has 53 bps of exon with the remainder of the sequence intronic. Based on early sequencing results, I redesigned primers for optimized

amplification in species of subg. *Pinus*, using a reverse primer that is anchored at the intron/exon boundary: 8612F1: TGT TAG CAT GCA ATC AAT CAC; 8612R5: TTG TTC CAG ACG CTA TTT CT.

My translation of the IFG8898 amplicon has highest similarity (protein-protein BLAST search in GenBank) to an *Arabidopsis thaliana* plasma membrane intrinsic protein (*WD-40*; NP175413). Based on an alignment of my amplicon with *P. taeda* cDNA (AA739796), and my translation of this cDNA sequence, I infer that my amplicon consists of an intron, 103 bps of exon, with the remainder consisting of 3' untranslated region.

Chloroplast Haplotypes. Sixteen individuals (including some diploid samples described in Chapter 4 of this dissertation) were sequenced for the chloroplast *trnG* (UCC) intron by PCR amplification using previously published primers (3'trnG and 5'trnG2G) (Shaw, Lickey, and Beck 2005). Sequences represented five *P. jeffreyi*, seven *P. ponderosa*, and four *P. washoensis* individuals. I aligned these 16 sequences with positions 8857 through 9610 in the *P. thunbergii* whole-chloroplast sequence (GenBank D17510). Two major haplotypes were identified within my 16 samples (see results) that differ by three substitutions. I used this information to assay amplicons of *trnG* for other *Pinus* samples by taking advantage of a *Dra*I restriction site that is inferred to cleave once, thus being diagnostic for, the "ponderosa-style" haplotype (Liston 1992). This recognition site is present in eight of the 16 sequences, but absent from *P. koraiensis* or *P. thunbergii* sequences.

Amplification and Sequencing. Haploid genomic DNA was isolated from the megagametophyte tissue of single seeds using the (FastDNA Kit®, Qbiogene, Irvine, CA, USA). PCR was performed on this haploid DNA as described in Willyard et al. (2007), products were directly sequenced using BigDye® v. 3.1 (Applied Biosystems, Foster City, CA, USA), and visualized on an Applied Biosystems 3730 Genetic Analyzer. The DNA yield from my single-seed isolation was inadequate for direct amplification in two samples (*P. douglasiana,* Atenquique, Mexico and *P. durangensis*; Table 3.2). These two samples were pre-amplified with a whole-genome multiple displacement amplification (Hosono et al. 2003) using phi29 DNA polymerase and pyrophosphatase (New England Biolabs, Ipswitch, MA, USA) and random hexamer primers (Operon, Huntsville, AL, USA). The whole-

genome product was then used as the template for PCR amplifications of these two samples. Five samples included in this study (footnoted in Table 3.2) required gel isolation because the PCR amplified two fragment sizes. For these samples, DNA was isolated from the excised band with the Ultra Clean™ DNA purification kit (MoBio Laboratories, Carlsbad, CA) prior to sequencing.

Analysis. Nucleotide sequences from the forward and reverse reads for each locus in each individual were assembled using CodonCode software (vers. 1.4.6 CodonCode Corporation, Dedham, MA) and edited by hand to create a consensus sequence. DNA alignments were made by eye to minimize the number of inferred indels. Nucleotide sequences and alignments are presented in Appendix 1. Each locus was analyzed independently using a Bayesian framework. I selected a nucleotide substitution model for each locus using the Akaike Information Criterion (AIC) implemented in MrModeltest (vers. 2.0; Nylander et al. 2004). For *LEA-like*, I chose the GTR + G model (six variable substitution types and gamma-distributed among-site rate variation). For *WD-40*, I selected HKY + G (transition/transversion model and gamma-distributed among-site rate variation).

Gaps were treated as missing data in the nucleotide partition and coded as present/absent using the simple gap coding method (Simmons and Ochoterena 2000) using web-based software available at: http://maen.huh.harvard.edu:8080/services/gap_recoder. Gap characters were analyzed along with the nucleotide substitutions using a binary model that specified equal rates. To account for the inability to observe gap presence/absence characters that are constant in all taxa, I implemented the "variable coding" option in the Bayesian analysis. I established three partitions for each locus: nonsynonymous (1^{st} and 2^{nd} codon positions of inferred exons); synonymous (3^{rd} codon positions plus noncoding sequences); and gap presence/absence codes. Rates were allowed to vary by partition. In order to assess the usefulness of these partitions, I calculated AIC scores using a formula that adds a penalty for extra parameters: $2lnL + 2K$, where lnL is the natural log of the total harmonic mean of the estimated marginal likelihoods, and K is the number of parameters (Felsenstein 2004). I compared AIC scores, topology, and the number of supported nodes for analyses using two partitions (nucleotides and gap-coding characters) with analyses using three partitions (nonsynonymous, synonymous, and gap codes). For each locus, I selected the number of partitions that resulted in the lowest AIC score.

I performed two runs using MrBayes software (vers. 3.1; Huelsenbeck and Ronquist 2003) for each locus. Each run used four simultaneous chains and 10 million generations of Metropolis–coupled Monte Carlo simulations, sampling every 1000 generations in order to save 10,000 trees per run. One majority-rule consensus tree was built for each locus by combining the trees generated by the two runs, discarding the first 1000 trees from each run as a burn-in. Branch lengths were estimated by averaging across all retained trees. Nodes with less than 0.95 posterior probability were considered weakly supported and were collapsed.

Evidence of recombination was evaluated in each locus using all six methods available in RDP2 software (Martin, Williamson, and Posada 2005), interpreting the inference of a recombination event by more than one method as strong evidence. Gaps were excluded prior to testing for recombination. I took advantage of multiple samples per species to calculate the coancestry coefficient Θw (Watterson 1975) for each species at each locus using DnaSP software (vers. 4.10.9; Rozas et al. 2003). The mean Θw for my two loci was used to estimate effective population sizes (N_e) for each species using the formula $N_e = \Theta w / (4 \mu G)$, assuming a generation time of $G = 25$ and $\mu = 0.70 \text{ X } 10^{-9}$ substitutions per site per year for *Pinus* estimated by Willyard et al. (2007) across nine nuclear loci based on a fossil calibration of the divergence of *Pinus* subgenera 85 million years ago. My two-locus estimate of effective population size was then used under coalescent theory to estimate the number of years for each species until allelic monophyly is more likely than paraphyly using the formula: $1.665 \times 2N_eG$ (Rosenberg 2003).

RESULTS

Alignments. I aligned sequences from 46 individuals representing 18 species for the *LEA-like* locus with an aligned length of 1630 characters. The simple gap coding method inferred 71 gaps, 38 of which were unique to a single sample and 33 were shared by two or more species. I aligned sequences from 52 individuals that represent 17 species for the *WD-40* locus. These included all of the samples aligned for *LEA-like*, with an additional *P. cooperi*, two additional *P. douglasiana*, one additional *P. hartwegii*, and two additional *P. ponderosa* individuals. My *WD-40* alignment does not include a sample from *P.*

durangensis. Aligned length for *WD-40* was 1182 characters. The simple gap coding method inferred 18 gaps for my *WD-40* alignment, of which four were unique to a single sample.

Trees. For each locus and parameter set, the Bayesian analyses converged on the same gene tree in every run. For two runs, the average standard deviation of split frequencies were 0.00384 and 0.003971 for *LEA-like* and *WD-40* loci, respectively. I compared the topology and summary statistics for each locus with results run under the same conditions using data sets that did not include the gap presence/absence codes. Topologies were identical for both loci with and without gap coding. The likelihoods are not comparable because the data sets are of different sizes. More importantly, the inferred relationships differed only in the posterior probabilities; some nodes were moved above or below my threshold of 0.95 for consideration as well-supported nodes. For *LEA-like*, this resulted in 19 supported nodes compared to 14 nodes inferred when gaps are coded. However, there were more deep nodes supported with gap coding; the additional nodes inferred without gaps were confined to the tips of the tree. For *WD-40*, there were 18 nodes in the data set that included gap coding, compared to 14 nodes without gap coding. I interpret these results as support that gap-coding adds information, and present only the trees inferred with the addition of a gap-coded partition. The total harmonic mean of the estimated marginal likelihoods was -4605.57 for *LEA-like* and -3116.71 for *WD-40*. For *LEA-like* with 53 bps of inferred exon, two partitions (nucleotides and gap-codes) was preferred, but for *WD-40* with 103 bps of inferred exon, three partitions (synonymous, nonsynonymous, and gap-coding) performed better (Table 3.3). Topologies were identical and number of nodes and branch lengths were nearly identical for inferences based on both partitioning schemes for each locus. Therefore, I elected to present the two-partition results for *LEA-like* and the three-partition results for *WD-40*.

Recombination. For each locus, a recombination event was inferred by only one of the six methods implemented in RDP2 software. In *LEA-like*, the "GENECONV" method infers a recombination point near the end of the alignment, while in *WD-40*, the "SiScan" method infers an event near an indel. No recombination events are inferred by the remaining five methods, including the more conservative Maximum X^2 test.

Chloroplast Haplotypes. Across 785 bps of aligned nucleotides, my sequences of the cpDNA *trnG* intron in 16 *Pinus* individuals revealed two predominant haplotypes with fixed differences at three sites. The "ponderosa-style" is distinguished by T, A, and G at positions 146, 203, and 573, respectively. The other major haplotype in my sequences is distinguished by G, G, and A at these positions. I observed one unique haplotype in my 16 samples. A *P. ponderosa* sample from near Crater Lake, Oregon, has G, A, and A at these positions and two additional autapomorphic substitutions at positions 188 and 680. The *trnG* intron sequences from GenBank create two additional haplotypes. Compared to my samples, the *P. thunbergii* sequence differs at 16 sites and requires three indels for alignment. I identified a *Dra*I recognition site (TTTAAA) at position 146 that is present in eight of my 16 samples: five *P. ponderosa* and three *P. washoensis*. The other samples (five sequences of *P. jeffreyi* and two sequences of *P. ponderosa*) have TTGAAA at these positions, while *P. thunbergii* has TTTTCA. Thus, I termed the haplotype with this *Dra*I recognition site a "ponderosa-style" haplotype and proceeded under the hypothesis that it may be unique in that lineage. I classified samples as having or lacking the "ponderosa-style" haplotype based on either my observed nucleotide sequence or the results of the *Dra*I restriction assay. Those results are summarized in Table 3.4 and mapped onto Figure 3-1 and Figure 3-2 using solid or open circles next to taxon names. I have identified the "ponderosa-style" haplotype in three species: *P. washoensis* (28 of 29 accessions tested), *P. ponderosa* (9 of 11 accessions tested), and *P. jeffreyi* (2 of 12 accessions tested). Using the *Dra*I restriction site, I did not observe the "ponderosa-style" haplotype in any other *Ponderosae* species, or in species outside of this subsection.

Intraspecific Diversity. Based on the mean intraspecific θ_W, the *LEA-like* locus has a higher diversity (θ_W = 0.01167) than the *WD-40* locus (θ_W = 0.00815) in *Ponderosae* species (Table 3.5). I did not find a correlation among species' θ_W estimates between the two loci (R=0.450, p=0.12, t=1.671). One θ_W estimate was more than four standard deviations above the mean: *P. maximinoi* for *LEA-like*. Excluding this estimate, the mean divergence for *LEA-like* (θ_W = 0.00869) is comparable to *WD-40*, but I did not find a correlation among species' θ_W estimates (R=0.251, p=0.43, t=0.82). The difference in N_e estimates for the 14 species where I sampled more than one individual per species (Table

3.5) was more than 8-fold (range 30.6 to 265.9 x 10^3). Five species had estimated N_e under 100 x 10^3 (*P. cooperi*, *P. coulteri*, *P. devoniana*, *P. jeffreyi*, and *P. pseudostrobus*).

DISCUSSION

Tree Overview. The phylograms shown in Figure 3-1 and Figure 3-2 have been collapsed to retain only nodes with 0.95 or higher posterior probabilities. Showing only well-supported nodes on each gene tree allows a straightforward comparison between these two independent inferences. Despite my use of highly divergent nuclear intronic regions, both trees leave many relationships unresolved. This suggests that the low interspecific differentiation observed in species-level phylogenies using chloroplast loci (e.g. *matK* and *rbcL*; Gernandt et al. 2005) and rDNA (Gernandt, Liston, and Pinero 2001) is likely to be mirrored across much of the nuclear genome.

LEA-like resolves two clades: one containing all members of the Sabinianae group plus all samples of *P. jeffreyi*; the other with the remainder of *Ponderosae s.s.* In contrast, *WD-40* resolves three clades. One clade contains all Sabinianae individuals plus all *P. jeffreyi,* but without all *P. coulteri* samples. With *WD-40*, there is also support for a clade with all of the *P. arizonica* samples plus *P. ponderosa* var. *scopulorum* from Utah and the *P. montezumae* sample from Hidalgo. The third clade is made up of the remaining *Ponderosae s.s*. plus a derived, monophyletic grouping of all three samples of *P. coulteri*.

Because the gene-trees inferred by *LEA-like* and *WD-40* are different, I do not present a combined phylogeny (Kubatko and Degnan 2007). In contrast, another study of *Pinus* phylogeny found congruent trees when comparing four low-copy nuclear loci (Syring et al. 2005). That experiment utilized one of the same loci (*LEA-like*), but tested 12 exemplars representing the subsections of *Pinus.* Thus, the taxonomic focus conducted by Syring et al. (2005) was on more divergent taxa than the full (i.e. species-level) taxonomic sampling within *Ponderosae* that was conducted in this study. Applying the formula $2N_eG$ (Tajima 1983) for the number of years for an 'average' locus in *Ponderosae* species to coalesce, I estimate a range from 1.5 (*P. devoniana*) to 9.7 (*P. ponderosa*) million years. Thus, for some species, the average coalescent time for these two loci may approach the inferred age for divergence of the *Ponderosae* clade (9 to 18 million years ago; Willyard et al. 2007).

I examined my alignments to see if inconsistencies could explain the conflict between loci. For *WD-40*, automated alignments inferred by Clustal W (Thompson, Higgins, and Gibson 1994) and MUSCLE (Edgar 2004) resulted in nearly identical alignments. The *LEA-like* locus, however, contains long stretches of imperfect repeats. Each of the automated alignment methods listed above 'scrambled' the sections of the alignment containing long indels which consist of imperfect repeats. After comparing automated and manual alignments of this locus, I agree with the opinion expressed by Kjer et al. (2007) in reference to the use of rRNA structural information to guide a manual alignment: "Manual alignment permits flexible and appropriate mental gap costs…", and I also agree that there is no guarantee that automated alignments will be more accurate or more repeatable under complicated conditions. In addition, I found that alignments of the *LEA-like* "indel-imperfect-repeat regions" between more diverged sequences (i.e. between species in *Ponderosae* and species from other subsections) were problematic as the number of imperfect repeats increased. I attribute this to two mechanisms that are both likely to play a role. First, some of the similar regions may have arisen from non-homologous insertion/deletion events. Second, some likely reflect the true sharing of ancestral polymorphism in repeat number. Another possibility is that some of the amplicons lack homology. Complex gene families have been described in *Pinus* (Kinlaw and Neale 1997); paralogs have been identified in *Pinus* low-copy nuclear loci (e.g. the *4CL* locus; Wang, Tank, and Sang 2000), and there is a possibility that nearly-identical paralogs occur in *Pinus* similar to those described for *Zea* (Emrich et al. 2007). Because I have no method to distinguish the relative contribution of these possibilities, I suggest that this highly divergent intron reaches its maximum usefulness at the subsectional level in subg. *Pinus* for phylogenetic inferences.

Introgression Inferred from Chloroplast Haplotypes. My observation of a "ponderosa-type" chloroplast haplotype in some individuals of *P. ponderosa*, *P. washoensis*, and *P. jeffreyi*, but not in other *Pinus* species (Table 3.4) provides evidence for introgression patterns among these three species. Although *P. ponderosa* and *P. washoensis* contain predominantly the "ponderosa-style" haplotype, I found three individuals that lack this haplotype. These samples are a *P. washoensis* from Babbitt Peak (included in this study), a *P. ponderosa* from the Warner Mountains (included in Chapter 4 of this dissertation), and

the *P. ponderosa* from Crater Lake whose sequence of the *trnG* intron reveals a unique haplotype. The two *P. jeffrey*i samples with the "ponderosa-type" haplotype are both from the Warner Mountains (included in Chapter 4 of this dissertation). The *LEA-like* gene tree provides a clear differentiation between *P. jeffreyi* and the clade containing all *P. ponderosa* and *P. washoensis* samples, suggesting that interspecific hybridization may have resulted in the sharing of a chloroplast, but not a nuclear, genealogy for these five samples. On the *WD-40* gene tree, the two samples of *P. ponderosa* and three samples of *P. washoensis* that join the Sabinianae group comprise four examples of the "ponderosa-style" haplotypes and only one example that also contains the "other" haplotype. Thus, the *P. washoensis* from Babbitt Peak is my only example that shares similarity with *P. jeffreyi* in both nuclear (*WD-40*) and chloroplast genomes, but notably not in *LEA-like*. Taken together, these results support a low, but not negligible, level of shared nuclear and organellar similarity between *P. jeffreyi* and both *P. ponderosa* and *P. washoensis*. Recent introgression is the simplest explanation for the occurrence of presumably identical chloroplast haplotypes (based on my restriction enzyme assay) in samples that resolve with their conspecifics in one or both nuclear loci. Nonetheless, older reticulate evolution is implied in the four samples that resolve with the Sabinianae group yet have the "ponderosa-style" haplotype because they apparently retain similarity to their conspecifics in the *LEA-like* locus as well as their chloroplast genome. I will use this information to further address the relationship between *P. ponderosa* and *P. washoensis* in Chapter 4 of this dissertation.

Shared Ancestral Polymorphism vs. Reticulate Ancestry. Overall, these gene trees exhibit ubiquitous retention of ancestral polymorphism, as the alleles of most species are not monophyletic. The inference that this incomplete lineage sorting is an accurate measure of their evolutionary history is strengthened by the similar results from two independent nuclear loci. Further, the substitution rates for these two loci are among the highest of nine low-copy nuclear loci tested in a recent study (Willyard et al. 2007), suggesting that this lack of coalescence may be representative of the majority of low-copy loci in the *Ponderosae* nuclear genome. There are two examples with potential evidence for reticulate ancestry: *P. coulteri* and the five *P. ponderosa*/*P. washoensis* samples that resolve with the Sabinianae group in *WD-40*. In both cases, the results are not easily explained by

incomplete lineage sorting alone, although it likely plays a role. Each species is discussed in turn below.

The well-supported resolution of a monophyletic *P. coulteri* clade within *Ponderosae s.s.* using the *WD-40* locus is surprising because this gene tree infers a strongly-supported clade that encompasses all of the remaining Sabinianae. In addition, *LEA-like* infers the traditional grouping of *P. coulteri* within the Sabinianae group. It is important to note that natural, as well as artificial, crosses between *P. coulteri* and *P. jeffreyi* have been reported (Critchfield 1966). Ledig (2000) suggested that *P. coulteri* has undergone a series of bottlenecks in population size and that introgression from sympatric *P. jeffreyi* may be the source of unique allozyme alleles observed in some *P. coulteri* populations. An alternative explanation is that *P. coulteri* is a diploid hybrid offspring of an ancient hybridization between *P. jeffreyi* and *P. ponderosa*. This would account for shared biochemical attributes, especially heptane, between *P. jeffreyi* and *P. coulteri*. A second alternative is that one or both of the ancestral lineages of *P. coulteri* is/are extinct. Either way, the dramatic differences in *P. coulteri* cone morphology are hypothesized to result from strong selection for a combination of serotiny and protection from seed predators (Borchert 1985). This scenario fits the framework proposed by Anderson and Stebbins (Anderson and Stebbins 1954) for hybridization as an evolutionary stimulus and for hybridization to set the stage for the invasion of new habitats (Ellstrand and Schierenbeck 2000). Further tests that address the origin of *P. coulteri* as a hybrid species will be bolstered by this phylogenetic framework. Given the results I present here, candidates for putative parents include a long list of extant taxa as well as the possibility that one or both parental lineages are extinct.

Neither of my nuclear loci provides support for a monophyletic *P. washoensis*. *WD-40* resolves four of the samples inside the Sabinianae group. The other four samples are sister to *P. ponderosa s.s*. samples or to Mexican and Central American species. The *LEA-like* locus places all of the *P. washoensis*, as well as all of the *P. ponderosa s.s.* samples in a mostly-unresolved polytomy with *P. arizonica*, *P. engelmannii*, *P. yecorensis*, and sister to clades containing the remaining Mexican taxa. This suggests reticulate evolution resulting in a portion of the *P. ponderosa s.l.* genome (i.e. that portion sampled by *WD-40*) that shares ancestry with Sabinianae. Again, *P. jeffreyi* is a likely link, because natural as well as artificial crosses between *P. ponderosa* and *P. jeffreyi* and between *P. washoensis* and *P. jeffreyi* have been reported (Haller 1962). My finding that all but two *P. ponderosa* and all

but one *P. washoensis* contain the "ponderosa-style" haplotype suggests that introgression is ancient, rather than current. If supported by further work, this hypothesis might explain why some exemplar-based studies have placed an allele from *P. washoensis* sister to *P. jeffreyi* (Patten and Brunsfeld 2002) or sister to *P. sabiniana* (Prager, Fowler, and Wilson 1976). It might also provide an explanation for the morphological discontinuity that lead to the original description of *P. washoensis* (Mason and Stockwell 1945) and lead Haller (1965) to suggest that high-elevation populations of *P. ponderosa* in the Blue Mountains of Oregon and in British Columbia appear more "Washoe-like" than those at lower altitudes. Perhaps the populations presently restricted to high elevations represent a lineage that has retained more of the 'footprint' of reticulation because of traits that are more useful in these harsh climates. With the exception of *P. jeffreyi*, all species of Sabinianae currently grow in hotter and dryer climates than *P. ponderosa*, while *P. jeffreyi* grows in colder climates or on ultramafic soil (Haller 1962). I believe that this question deserves a closer inspection. In the next chapter of this dissertation, I present results from a population-level study that attempts to shed more light on the relationship between *P. washoensis*, *P. ponderosa*, and *P. jeffreyi* and to determine if current gene flow shapes these species.

The Phylogenetic Affinity of *P. jeffreyi.* *Pinus jeffreyi* has traditionally been included in *Ponderosae s.s.* because of its morphological similarity with *P. ponderosa*. If ovulate cones are unavailable, morphological differences are subtle, but include non-resinous buds and a softer, "more-lemony" smell in fresh-cut wood (Kral 1993). In areas where *P. jeffreyi* is sympatric with *P. ponderosa*, individuals can be found that appear intermediate between these two species for some character states (Haller 1962). In addition, putative natural hybrids between these two species have been documented at a low, but consistent level (Haller 1962), and artificial crosses of *P. jeffreyi* with other *Ponderosae* (e.g. *P. washoensis*, *P. engelmannii*, and *P. montezumae*) are fertile, as are back crosses to *P. jeffreyi* (Critchfield 1966). In the fourth chapter of this dissertation, I show that *P. jeffreyi* is clearly differentiated from *P. ponderosa s.l.*, but also present evidence for genetic admixture in some individuals. This supports previous conclusions that there is a low level of natural first-generation introgression between *P. jeffreyi* and *P. ponderosa*. However, *P. jeffreyi* has a unique softwood oleoresin profile that is shared with *P. sabiniana*: turpentine predominantly comprised of heptane rather than the terpenes produced by most other

Ponderosae (Mirov 1929). *Pinus sabiniana* is traditionally grouped with two other species of California big-coned pines (*P. coulteri* and *P. torreyana*) in *Sabinianae* (Little and Critchfield 1969). Oleoresin chemistry lends support for the distinctiveness of the Sabinianae group, as *P. coulteri* and *P. torreyana* also produce heptane and other paraffin hydrocarbons in at least trace amounts. These compounds have only been found in a few other species, including *P. oaxacana* Mirov (21% heptane, 1% undecane), which I treat as a variety of *P. pseudostrobus* (Mirov 1929). The seed fatty acid profile of *P. jeffreyi* is also more similar to *P. coulteri* than to any Ponderosae species tested (Wolff et al. 2000). Further, *P. jeffreyi* forms occasional putative natural hybrids with *P. coulteri*, and artificial crosses between these species are fertile (Critchfield 1966). Mirov used this biochemical and crossability evidence, as well as cones that are larger than most *Ponderosae,* as evidence to include *P. jeffreyi* and *P. oaxacana* in his subsection *Macrocarpae* (Mirov 1929). Nonetheless, the alkaloids reported to uniquely occur in the *Sabinianae* (Tallent and Horning 1955) have now been identified in most *Ponderosae* species (Gerson and Kelsey 2004). Because recognizing *Sabinianae* would have rendered *Ponderosae s.s.* paraphyletic on previous chloroplast-based phylogenies, *Sabinianae* was subsumed into *Ponderosae s.l.* (Gernandt et al. 2005, Price, Liston, and Strauss 1998). Other evidence now exists for the eccentricity of *P. jeffreyi*. The pine bark beetle genus *Dendroctonus* has species that are specialized for their pine hosts: one species is monophagous on *P. jeffreyi* (*Dendroctonus jeffreyi*), while *D. ponderosae* is highly polyphagous, utilizing 12 different *Pinus* host species (Paine and Hanlon 1994 and references therein). More recently, Krupkin (1996) found weak resolution for *P. ponderosa* as sister to *P. jeffreyi* and for *P. jeffreyi* as sister to a *P. coulteri*/*P. torreyana* clade based on chloroplast restriction sites. Gernandt et al. (2005) found no resolution within a clade that encompassed all of *Ponderosae s.l.,* except that *P. hartwegii*, *P. montezumae*, and *P. pseudostrobus* were weakly resolved as sister to this clade.

Here, I find *P. jeffreyi* strongly nested within the Sabinianae group in both *LEA-like* and *WD-40*. These results are further corroborated by an expanded phylogeny based on a large sample of cpDNA sequence data (Gernandt, pers. comm.). In that nearly-complete study, *P. jeffreyi* is placed with the Sabinianae group in some phylogenetic trees, although this placement lacks statistical support in the strict consensus.

Montezumae, Ponderosae, and Pseudostrobus Groups. Neither gene tree infers support for these traditional subdivisions (color-coded on Figure 3-1 and Figure 3-2), even if the Montezumae and Pseudostrobus groups are combined (Farjon and Styles 1997). This supports a similar finding from chloroplast nucleotide sequences. Using sequences of *matK* and *rbcL*, Gernandt et al. resolved *P. hartwegii* and *P. montezumae* (Montezumae group) and *P. pseudostrobus* (Pseudostrobus group) as sister to a polytomy containing taxa from all three groups (Gernandt et al. 2005).

Despite the polymorphism evident within each species, nearly every clade contains samples from more than one of these groups. This suggests that these groupings are artificial or are sufficiently young that ancestral polymorphism (perhaps aided by occasional reticulation) has yet to be sorted via isolation. Similar results have been reported based on isozyme analysis (Karalamangala and Nickrent 1989).

Species-level Monophyly and Paraphyly. Coalescent theory predicts that, in the absence of geographic structure, faster-evolving loci will coalesce more rapidly than slower-evolving loci in the same lineage (Rosenberg 2003). However, the *LEA-like* locus, which is slightly more divergent within the *Ponderosae* than *WD-40* based on intraspecific θ_W, only yields 14 well-resolved nodes in this data set compared to the 18 nodes resolved using *WD-40*. A number of mechanisms could contribute to this, including stochastic effects, balancing selection on this locus or linked genes, undetected recombination, a paralogous locus in one or more lineages, and geographic structure. I also note that I have seven more samples in the *WD-40* data set. The amplification of a paralogous locus seems unlikely for most samples, as *P. maximinoi* θ_W in *LEA-like* is the only extreme outlier (Table 3.5).

The lack of resolution at the tips of the trees prevents most tests of monophyly or paraphyly, despite the availability of multiple samples per species. Short branches like those observed in this data set (indicating low species-level divergences) can hide paraphyly by increasing the sharing of ancestral polymorphisms. In fact, only four of my 18 species have individuals that resolve in separate, well-supported clades on both *LEA-like* and *WD-40* trees (Figure 3-1 and Figure 3-2). This is in strong contrast to Syring et al. (2007), who report that 19 of 33 species in subg. *Strobus* (Lemmon)A E Murray were nonmonophyletic for the *LEA-like* locus.

Two of the species are well resolved to more than one clade, *P. ponderosa* and *P. washoensis*, appear the most polyphyletic on these gene trees. In *LEA-like*, *P. ponderosa* from the Blue Mountains (OR) resolves within a clade containing two samples of *P. washoensis* from the Warner Mountains (CA) while other *P. ponderosa* and *P. washoensis* samples either pair with conspecifics or are unresolved. In contrast, the *P. ponderosa* sample from the Blue Mountains resolves as sister to two species from Mexico, *P. douglasiana* and *P. yecorensis* in *WD-40* (Figure 3-2). In *WD-40*, two *P. ponderosa* and three *P. washoensis* samples join a clade comprised of *P. jeffreyi*, *P. sabiniana*, and *P. torreyana*. However, these same accessions are part of one unresolved clade in *LEA-like*. In a third example for this pair, the *P. ponderosa* var. *scopulorum* sample resolves as sister to *P. arizonica* plus *P. montezumae* in *WD-40*, but sister to the *P. ponderosa* var. *ponderosa* sample from Southern California in *LEA-like*. While these examples that lack reciprocal monophyly lend support to the suggestion that the narrowly endemic *P. washoensis* may be synonymous with the wide-ranging *P. ponderosa* (Niebling and Conkle 1990, Rehfeldt 1999b, Lauria 1997, Brayshaw 1997), I do not consider this proof for conspecificity. This is because shared ancestral polymorphism maintained within each species could account for these patterns, particularly since no sample of alleles shows taxon-level monophyly in this analysis. In other words, minor genetic differences ("speciation genes") that are unlinked to the loci that I sampled may actually differentiate these taxa. It is important to note that while I have few cases of polyphyly, essentially every species lacks reciprocal monophyly in one or both of my gene trees. Further, the contrast between my two gene trees suggests that lack of coalescence is the most important contributing factor. However, the group of five *P. ponderosa*/*P. washoensis* individuals that resolve inside the Sabinianae group on the *WD-40* tree (but not the *LEA-like* tree) are a noteworthy result, and are discussed below as potentially supporting reticulate ancestry with members of the Sabinianae.

In the third species with alleles in two well-supported clades, the behavior of my *P. maximinoi* samples in both loci appears suspicious, but with different symptoms. The extreme values for *P. maximinoi LEA-like* intraspecific θ_W are more than four standard deviations away from the mean, and the unusual placement of the Oaxaca sample and its long branch on that tree suggest aberrant behavior for the amplification of this locus in this collection. This does not appear to be a misidentified accession because the Oaxaca individual resolves as expected in *WD-40*, and it is the Honduras accession of *P. maximinoi*

that resolves in a rather unexpected clade in the *WD-40* tree. Repeated amplification and sequencing of these samples confirm these results. While previously-undescribed variability within this species may be indicated by the *WD-40* tree, the amplification of a nearly identical paralog (Emrich et al. 2007) is the most likely explanation for the highly divergent allele in *LEA-like*.

The fourth example is *P. montezumae*. This species exhibits high intraspecific divergence in both loci (Table 3.5). In both gene trees, the accession from Hidalgo resolves with unexpected clades. In my *LEA-like* tree, it joins *P. durangensis* (Ponderosae group) and *P. maximinoi* (Pseudostrobus group), while it is sister to *P. arizonica* (Ponderosae group) in *WD-40*. This result was confirmed by verifying the identity of the collection and repeating the amplifications and re-sequencing. *Pinus montezumae* is always described as a highly variable species. Two samples of this species resolved into separate groups using a distance matrix from isozyme variation (Karalamangala and Nickrent 1989). Putative natural hybrids between *P. montezumae* and *P. devoniana*, *P. hartwegii*, and *P. pseudostrobus* have been described (Martinez 1948; Mirov 1967; Perry 1991; Farjon and Styles 1997). Natural hybridization has been documented between *P. montezumae* and *P. hartwegii* using chloroplast restriction sites (Matos and Schaal 2000), and Matos (1995) concluded that the taxon which Engelmann named *P. rudis* represents hybrids between *P. montezumae* and *P. hartwegii*. Artificial crosses between *P. montezumae* and *P. ponderosa* and between *P. montezumae* and *P. engelmannii* have been successful (Conkle and Critchfield 1988), and Matos suggested that the Montezumae complex links the Ponderosa complex with the Pseudostrobus complex. However, I have no direct evidence for introgression between *P. montezumae* and any of the species with which my sample from Hidalgo groups in these two gene trees. Further, the ranges of those species (*P. durangensis, P. maximinoi*, and *P. arizonica*) do not extend into Hidalgo. Therefore, current or recent introgression is unlikely. This suggests that the genomic composition of *P. montezumae* is diverse, and that the evolutionary relationships inferred by my two loci may be complicated by undeciphered diversity within this species, disguised here by shared ancestry with the other major groups within *Ponderosae*.

Intraspecific Genetic Diversity. Ours are the first nuclear estimates of θ_W for many of these species, and they are useful for species comparisons despite being based on a small

number of samples. Importantly, I estimated an eight-fold range in the estimated effective population sizes within the species of *Ponderosae*. Syring et al. (2007) reported comparable estimates of intraspecific nucleotide diversity (π) for 33 species in *Pinus* subg. *Strobus* that ranged more than 20-fold (0.001 to 0.021). After removing the potentially unreliable estimate for *P. maximinoi* (for reasons discussed above), I report a range in my estimates based on the average of two loci of the number of years until allelic monophyly is more likely than paraphyly (Table 3.5) from 2.6 million years (*P. devoniana*) to 16.2 million years (*P. ponderosa*). Across subg. *Strobus*, previous estimates for years until monophyly is more likely than paraphyly range from 1.7 (*P. flexilis*) to 24 (*P. lambertiana*) million years (Syring et al. 2007). Both studies used the *LEA-like* locus, allowing a direct comparison. For subg. *Strobus*, N_e estimates for 12 species ranged from 17 $x10^3$ to 120 $x10^3$ (Syring et al. 2007). Excluding estimates for *P. maximinoi*, I estimate *Ponderosae* species N_e based on *LEA-like* to range from 35 $x10^3$ (*P. devoniana*) to 215 $x10^3$ (*P. ponderosa*) and 218 $x10^3$ (*P. arizonica*). Interestingly, N_e for *P. arizonica* from the *WD-40* locus is among the smallest of my estimates (179 $x10^3$), a reminder that measurements of diversity for species from only two loci should be interpreted with caution.

Only a limited number of *Pinus* species have previously been reported to have low intraspecific genetic diversity. Prominent examples include *P. resinosa* (DeVerno and Mosseler 1997; Echt et al. 1998; Boys, Cherry, and Dayanandan 2005) and *P. torreyana* (Ledig and Conkle 1983; Haller 1986). In fact, I chose to include only one example of *P. torreyana* in this experiment based on those reports of low diversity within and between its populations.

The three samples of *P. devoniana* from states in the central portion of its range (Hidalgo, Guerrero, and Michoacán; Table 3.2) yield the lowest N_e (30.6 x 10^3) of any of my species, and a mean Θw for two nuclear loci of only 0.00215. Thus, *P. devoniana* appears to have a very restricted nuclear, as well as chloroplast, genetic diversity, despite its wide geographic range. Syring et al. (2007) attributed similar patterns of low genetic diversity but large geographic ranges in *P. koraiensis and P. strobus* (subsection *Strobus*) to recent range expansions from a narrow genetic base. However, the low values for Θ (0.00152 and 0.00010) previously reported for geographically isolated populations of *P. devoniana* based on chloroplast restriction sites were attributed to lack of gene flow between eastern and

western regions for this species that grows in "small, scattered populations" (Matos 1998), so geographic isolation appears an important factor for *P. devoniana.*

I am unaware of any previously published estimates of genetic diversity within *P. cooperi*, but note that my number of intraspecific samples is small in this study. In addition, unlike most other species included in this paper, I used several seeds from a bulk collection at one locality for *P. cooperi*. It is possible that the limited divergence I observed for this species indicates a family relationship (perhaps even as full siblings) between these seeds, rather than species-wide homogeneity.

Ledig (2000) attributed a species-wide low level of heterozygosity (and a clinal reduction in heterozygosity from south to north) at allozyme loci in *P. coulteri* to a "cascading series of founder effects" as this species expanded its range northwards from putative Pleistocene refugia. My samples from the central portion of the range of this species (Table 3.2) confirm that *P. coulteri* has one of the lowest estimated N_e of any of the *Ponderosae* (Table 3.5). I have already discussed possible causes in the context of phylogenetic resolution and my confirmation of limited diversity for *P. coulteri* (Ledig 2000) is in accord with a recent origin.

Despite its broad distributional range and morphological variability (Haller 1962), *P. jeffreyi* exhibits a low estimated N_e similar to *P. coulteri* (Table 3.5). Furnier and Adams (1986) reported that *P. jeffreyi* populations in the Klamath region (which are geographically marginal and presumably adapted to ultramafic soils) contained lower levels of heterozygosity than other populations. However, they reported a mean expected heterozygosity ($H_e = 0.255$) from allozyme loci for Sierra Nevada and southern California populations that is typical for *Pinus* species. I estimate a low effective population size ($N_e = 62.9 \times 10^3$) from my sample of six *P. jeffreyi* individuals from locations that ranged from the Warner Mountains to the San Gabriel Mountains.

The final species for which I observed a rather low inferred N_e is *P. pseudostrobus*. This species is quite variable and has been considered to be 'difficult' taxonomically, although Stead (1983) presented a suite of morphological characters that distinguish it from *P. maximinoi* and *P. douglasiana*. *Pinus pseudostrobus* has a large geographical distribution with at least three varieties (Farjon and Styles 1997), and I am not aware of any previous reports of limited genetic variability.

At the other end of the spectrum is the exceptionally wide-ranging *P. ponderosa* with estimates of $N_e = 194.1 \times 10^3$. Given its incredibly large latitudinal and altitudinal distribution, my six samples are unlikely to have sampled the breadth of diversity for this species. Because *P. ponderosa* encompasses at least two well-documented varieties, this large estimate for N_e seems reasonable. For comparison, an estimate of $N_e = 120 \times 10^3$ has been made for *P. lambertiana*, another *Pinus* species with a wide geographic range (Syring et al. 2007).

Conclusions. The *Ponderosae* encompass a group of species with very different levels of diversity. Estimates for the number of years until monophyly is more likely than paraphyly for many species (Table 3.5) approach or exceed the estimated age of the subsection (8 to 15 million years ago; Willyard et al. 2007). Due to incomplete lineage sorting, I expect that the genomes of these species are still "mosaics of conflicting genealogies" (Pollard et al. 2006). This comparison predicts that the lack of phylogenetic resolution observed using two nuclear loci is reflective of their evolutionary relationships. On the other hand, there are a few species with rather limited diversity, including *P. devoniana*, *P. coulteri*, and *P. jeffreyi*.

I found support for Mirov's (1929) proposed grouping of *P. jeffreyi* with the California big-coned pines (Sabinianae group). I also provide tantalizing evidence for a hypothesized reticulation between it and *Ponderosae s.s.* The strongest evidence for hybridization between these groups is *P. coulteri*. The small N_e estimated for both loci in *P. coulteri* suggests that it is very unlikely that the *Ponderosae s.s.*-like alleles I observe in *WD-40* for *P. coulteri* were retained from a polymorphic ancestor. This origin of *P. coulteri* could be described as a hybrid speciation event between a taxon in *Ponderosae s.s.* and one related to the Sabinianae group, with the possibility that one or both ancestors is extinct. This hypothesis merits further investigation.

A legacy of reticulate evolution could also explain the resolution of some *P. washoensis* and some *P. ponderosa* individuals with the Sabinianae group in *WD-40*. However, the large N_e for these species means that a retention of ancestral polymorphism is also a viable alternative explanation. In addition, a low level of ongoing introgression with *P. jeffreyi* from the Sabinianae group may contribute to this pattern.

My results suggest that *P. washoensis* may not be distinct from *P. ponderosa*. In order to make a reasonable delineation between species and to distinguish ancient reticulation from current introgression (especially considering the high level of intraspecific variation I have observed within *Pinus* species), I propose that population-level sampling will be needed. The clustering of individuals as well as the comparison of allele frequencies between populations has already proven useful in a number of experiments for *Pinus* using both morphological and molecular characters (e.g. Stead 1983, Haller 1962, Rehfeldt 1999a, Matos and Schaal 2000). In the next chapter of this dissertation, I present the results of such an experiment for *P. washoensis* and its sympatric species.

Table 3.1. Comparison of taxonomic subsection (italics) or group (plain text) assignments in major floristic treatments that have suggested divisions within *Ponderosae*. Parentheses in cells are alternate names used in that treatment.

Pinus species	Martínez 1948	Van der Burgh 1973	Perry 1991	Farjon & Styles 1997	Price, Liston, & Strauss 1998
devoniana Lind.	Montezumae (*P. michoacana*)	n/a	Michoacana (*P. michoacana*)	*Pseudostrobi*	Montezumae
donnell-smithii Masters	n/a	n/a	Rudis	n/a	Montezumae
hartwegii Lind.	Montezumae	*Pseudostrobi*	Rudis	*Ponderosae*	Montezumae
montezumae Lamb.	Montezumae	*Pseudostrobi*	Montezumae	*Pseudostrobi*	Montezumae
arizonica Engelm. ex Rothr.	Ponderosa	*Ponderosae*	Ponderosae	*Ponderosae*	Ponderosae (*P. ponderosa* var. *arizonica*)
cooperi Blanco	Montezumae (*P. lutea*)	*Pseudostrobi*	Montezumae	*Ponderosae* (*P. arizonica* var. *cooperi*)	Ponderosae
durangensis Martínez	Montezumae	n/a	Ponderosae	*Oocarpae*	Ponderosae
engelmannii Carrière	Ponderosa	*Ponderosae*	Ponderosae	*Ponderosae*	Ponderosae
jeffreyi Balf.	Ponderosa	*Ponderosae*	Ponderosae	*Ponderosae*	Ponderosae
ponderosa Dougl. ex C. Lawson	Ponderosa	*Ponderosae*	n/a	*Ponderosae*	Ponderosae
washoensis Mason & Stockw.	n/a	n/a	n/a	n/a	Ponderosae

yecorensis Debreczy & Rácz	n/a	n/a	n/a	n/a	n/a
douglasiana Martínez	Pseudostrobus	n/a	Montezumae	*Pseudostrobi*	Pseudostrobus
maximinoi H.E. Moore	Pseudostrobus (*P. tenuifolia*)	n/a	Pseudostrobus	*Pseudostrobi*	Pseudostrobus
nubicola Perry	n/a	n/a	Oaxacana	n/a	Pseudostrobus
pseudostrobus Lind.	Pseudostrobus	*Pseudostrobi*	Pseudostrobus	*Pseudostrobi*	Pseudostrobus
coulteri D. Don	Coulteri	*Sabinianae*	n/a	n/a	Sabinianae
sabiniana Dougl. ex C. Lawson	n/a	*Sabinianae*	n/a	n/a	Sabinianae
torreyana Parry ex Carrière	n/a	*Torreyanae*	n/a	n/a	Sabinianae

Table 3.2. Geographic sources for seed samples of 18 *Ponderosae* species and one outgroup species.

Species	Location	State / Province, Country	Latitude	Longitude	Voucher
P. arizonica	Silver City	New Mexico, USA	33.13°N	-108.00°W	OSC
P. arizonica	Saguaro National Park	Arizona, USA	32.20°N	-110.53°W	OSC
P. arizonica[1,2]	Paila	Coahuila, Mexico	25.50°N	-102.50°W	OSC
P. cooperi[3]	El Salto	Durango, Mexico	23.80°N	-105.40°W	OSC
P. coulteri	Santa Barbara	California, USA	34.92°N	-116.92°W	RSA
P. coulteri	San Gabriel Mts	California, USA	34.35°N	-117.98°W	RSA
P. coulteri	Anza-Borrego	California, USA	33.50°N	-116.50°W	OSC
P. devoniana	Huasca	Hidalgo, Mexico	20.20°N	-98.60°W	MEXU
P. devoniana	Jujucato	Michoacan, Mexico	19.42°N	-101.82°W	OSC
P. devoniana	Filo de Caballos	Guerrero, Mexico	17.65°N	-99.84°W	OSC
P. donnell-smithii[4]	Quetzaltenango	Quet., Guatemala	14.08°N	-91.52°W	OSC
P. douglasiana[1]	Atenquique	Jalisco, Mexico	19.53°N	-103.52°W	OSC
P. douglasiana[4]	Yerba Santa	Guerrero, Mexico	16.97°N	-98.58°W	OSC
P. douglasiana	Concordia	Sinaloa, Mexico	23.48°N	-105.85°W	ARIZ
P. durangensis	San Juanita	Chihuahua, Mexico	28.00°N	-107.58°W	E
P. engelmannii	Cave Creek	Arizona, USA	31.72°N	-110.78°W	OSC
P. engelmannii	Florida Canyon	Arizona, USA	31.73°N	-110.83°W	OSC
P. hartwegii	Temascaltepec	State of Mexico, Mexico	19.12°N	-99.76°W	OSC
P. hartwegii	Yerba Santa	Guerrero, Mexico	17.52°N	-99.96°W	OSC
P. jeffreyi	San Gabriel Mts.	California, USA	34.03°N	-117.92°W	OSC
P. jeffreyi	Bishop	California, USA	37.37°N	-118.39°W	OSC
P. jeffreyi	Susanville	California, USA	40.30°N	-120.87°W	OSC
P. jeffreyi	Warner Mts.	California, USA	41.17°N	-120.28°W	OSC
P. jeffreyi	Mt. Rose	Nevada, USA	39.33°N	-119.88°W	OSC
P. jeffreyi	Reno	Nevada, USA	39.24°N	-119.84°W	OSC
P. maximinoi	San Jeronomo	Oaxaca, Mexico	17.82°N	-97.83°W	OSC
P. maximinoi	Minas de Oro	Comayugua, Honduras	13.53°N	-86.56°W	OSC
P. maximinoi	Alta Verapaz	Guatemala	15.47°N	-90.37°W	OSC
P. montezumae	Epazoyucan	Hidalgo, Mexico	20.11°N	-98.61°W	MEXU

P. montezumae	Malacatancito	Huehue., Guatemala	15.22°N	-91.52°W	OSC
P. ponderosa var. *ponderosa*[2]	Curlew	Washington, USA	48.88°N	-118.77°W	OSC
P. ponderosa var. *ponderosa*	Southern California (Big Bear Lake)	California, USA	34.15°N	-116.85°W	OSC
P. ponderosa var. *scopulorum*	Price Canyon	Utah, USA	39.77°N	-110.92°W	OSC
P. ponderosa var. *ponderosa*	Warner Mts.	California, USA	41.03°N	-120.32°W	OSC
P. ponderosa var. *ponderosa*	Abert Rim	Oregon, USA	42.38°N	-120.23°W	OSC
P. ponderosa var. *ponderosa*	Blue Mts.	Oregon, USA	44.07°N	-118.78°W	OSC
P. pseudostrobus[1]	Patio de Bolas	Chiantla, Guatemala	15.38°N	-91.43°W	OSC
P. pseudostrobus	Filo de Caballos	Guerrero, Mexico	17.65°N	-99.84°W	OSC
P. sabiniana	Weaverville	California, USA	40.73°N	-122.94°W	OSC
P. sabiniana	Clearlake	California, USA	39.18°N	-122.70°W	OSC
P. sabiniana	Redding	California, USA	40.55°N	-122.46°W	OSC
P. torreyana	Santa Rosa Is.	California, USA	33.95°N	-120.10°W	OSC
P. washoensis	Babbitt Peak	California, USA	39.43°N	-120.08°W	OSC
P. washoensis	Mt. Rose	Nevada, USA	39.33°N	-119.52°W	OSC
P. washoensis	Warner Mts. 1	California, USA	41.16°N	-120.12°W	OSC
P. washoensis	Warner Mts. 2	California, USA	41.18°N	-120.12°W	OSC
P. washoensis	Warner Mts. 3	California, USA	41.17°N	-120.25°W	OSC
P. washoensis	Warner Mts. 4	California, USA	41.17°N	-120.25°W	OSC
P. yecorensis	Yecora	Sonora, Mexico	28.38°N	-108.87°W	ARIZ
P. contorta (outgroup)	Tagish Lake	Yukon, Canada	60.14°N	-134.24°W	OSC

[1] Amplicon was gel-isolated for *LEA-like* locus
[2] Amplicon was gel-isolated for *WD-40* locus
[3] Three individuals from the same population were sampled
[4] Two individuals from the same population were sampled

Table 3.3. Comparison of Akaike Information Criteria (AIC) between two schemes for each locus. Two-partition allows rates to vary between nucleotide and gap-coding partitions; three-partition varies by synonymous, nonsynonymous, and gap codes. For each locus, the partitioning scheme with the lowest AIC is favored. ln L, natural log of the total harmonic mean of the estimated marginal likelihoods; K, number of parameters; delta, difference between AIC scores.

Locus	Model	ln L	K	AIC	delta
LEA-like	GTR+G		9		
	GTR+G+2 partitions	-4582.22	11	9186.44	
	GTR+G+3 partitions	-4605.57	12	9235.14	48.70
WD-40	HKY+G		5		
	HKY+G+3 partitions	-3116.71	8	6249.42	
	HKY+G+2 partitions	-3132.33	7	6278.66	29.24

Table 3.4. Summary of chloroplast haplotypes by species. Includes sampling of individuals from this study, diploid samples described in Chapter 4 of this dissertation, a survey of *Pinus* species outside of subsect. *Ponderosae,* and two sequences from GenBank. Species are arranged in descending order for number of samples with ponderosa-style haplotype. S, inferred from nucleotide sequence; R, inferred from restriction digest.

	ponderosa-style			other		
Species	S	R	**Total**	S	R	**Total**
P. washoensis	3	25	**28**	1	0	**1**
P. ponderosa	5	4	**9**	2	0	**2**
P. jeffreyi	0	2	**2**	5	5	**10**
Other *Ponderosae*						
P. arizonica	0	0	**0**	0	3	**3**
P. cooperi	0	0	**0**	0	2	**2**
P. coulteri	0	0	**0**	0	3	**3**
P. devoniana	0	0	**0**	0	3	**3**
P. donnell-smithii	0	0	**0**	0	1	**1**
P. douglasiana	0	0	**0**	0	5	**5**
P. engelmannii	0	0	**0**	0	2	**2**
P. hartwegii	0	0	**0**	0	2	**2**
P. maximinoi	0	0	**0**	0	3	**3**
P. montezumae	0	0	**0**	0	2	**2**
P. pseudostrobus	0	0	**0**	0	2	**2**
P. sabiniana	0	0	**0**	0	3	**3**
P. torreyana	0	0	**0**	0	1	**1**
P. yecorensis	0	0	**0**	0	1	**1**
Other *Pinus*						
P. attenuata	0	0	**0**	0	1	**1**
P. contorta	0	0	**0**	0	1	**1**
P. koraiensis	0	0	**0**	1	0	**1**
P. radiata	0	0	**0**	0	1	**1**
P. thunbergii	0	0	**0**	1	0	**1**

Table 3.5. Effective population sizes (N_e) inferred from the mean intraspecies coancestry coefficient (Θw) for two loci. Species are ranked by increasing N_e. N, number of samples; L, number of aligned nucleotides excluding gaps; Θw, the Watterson estimator.

	LEA-like			*WD-40*			N_e ($x10^3$)	Years for monophyly to be more likely than paraphyly ($x10^6$)
	N	L	Θw	N	L	Θw		
P. devoniana	3	551	0.00242	3	1067	0.00187	30.6	2.6
P. cooperi	2	1093	0.00457	3	1068	0.00062	37.1	3.1
P. coulteri	3	911	0.00659	3	948	0.00070	52.1	4.3
P. jeffreyi	5	887	0.00433	6	882	0.00447	62.9	5.2
P. pseudostrobus	n/a			2	1068	0.00562	80.3	6.7
P. sabiniana	3	892	0.00897	3	1022	0.00522	101.4	8.4
P. hartwegii	2	1436	0.00279	2	958	0.01253	109.4	9.1
P. donnell-smithii	2	1502	0.00266	2	977	0.01331	114.1	9.5
P. arizonica	3	918	0.01525	3	1066	0.00125	117.9	9.8
P. washoensis	6	730	0.01020	6	539	0.00650	119.3	9.9
P. engelmannii	2	1077	0.01393	2	1068	0.00281	119.6	9.9
P. ponderosa	5	860	0.01507	6	723	0.01211	194.1	16.2
P. maximinoi	2	864	0.04745*	3	1056	0.01578	225.4	18.8
P. montezumae	2	1087	0.01748	2	1064	0.01974	265.9	22.1
mean	3.1	985	0.01167	3.3	965	0.00815	108.0	9.0
std. dev.	1.4	260	0.01200	1.5	157	0.00485	61.8	5.4

* Excluded from calculation of N_e because Θw is >4 S.D. from mean

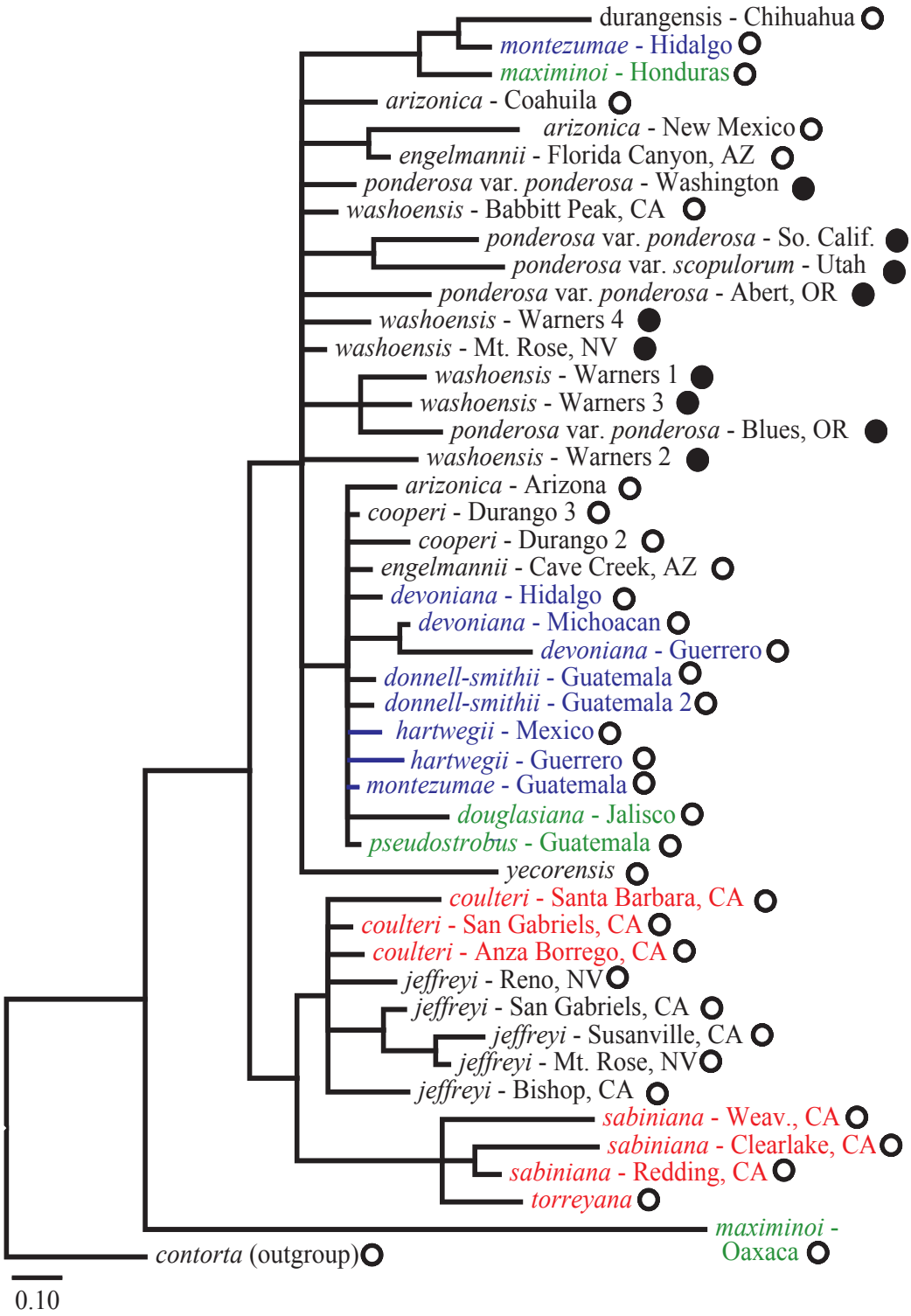

Fig. 3.1. Majority-rule consensus tree for *LEA-like* showing only nodes with 0.95 posterior probability or higher. Blue, Montezumae; green, Pseudostrobus; black, Ponderosae group. Branches are proportional to length; scale bar is substitutions per site. Geographic localities are given next to epithets to distinguish multiple collections (Table 3.2). Filled circles have "ponderosa-style" chloroplast haplotype; open circles lack this haplotype.

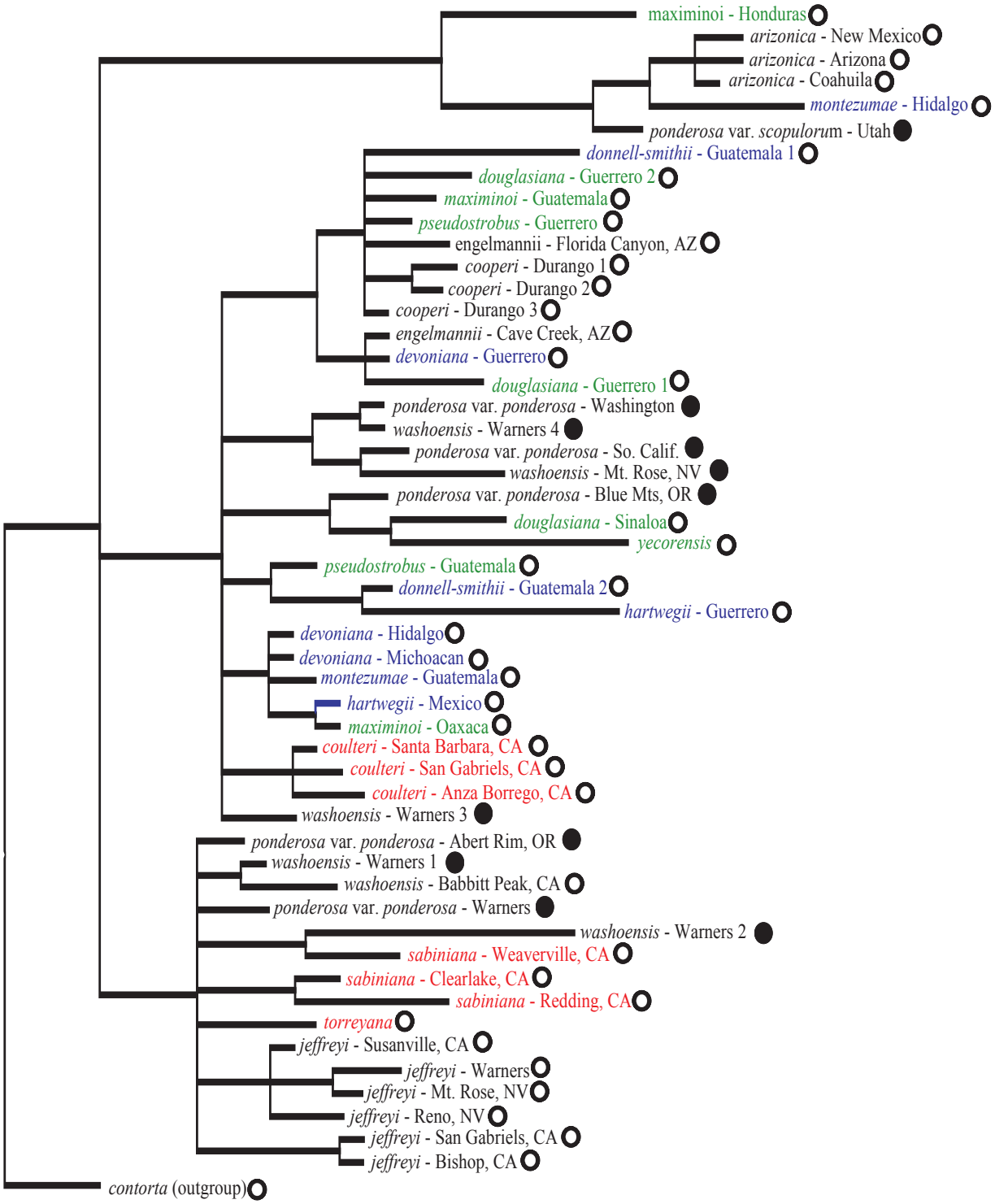

Fig. 3.2. Majority-rule consensus tree for *WD-40* showing only nodes with 0.95 posterior probability or higher. Blue, Montezumae; green, Pseudostrobus; black, Ponderosae group. Branches are proportional to length; scale bar is substitutions per site. Geographic localities are given next to epithets to distinguish multiple collections (Table 3.2). Filled circles have "ponderosa-style" chloroplast haplotype; open circles lack this haplotype.

ACKNOWLEDGMENTS

I appreciate outstanding laboratory results produced by Mariah Parker-Defeniks, Anthony Shireman, Carla Streng, and Kathleen Farrell. Jerry Berdeen (USDA Forest Service, Dorena Genetic Resource Center), Leonel Eduardo Barrios Escobras (Instituto Nacional de Bosques, Guatemala), George Ferguson (University of Arizona), David Gernandt (U. N.A.M., Ciudad Universitaria, Mexico), James Henrickson (University of Texas), Jesus Vargas Hernandez (Instituto de Recursos Naturales, Mexico), Jon P. Kazmierski (USDA Forest Service), Steve Laws (USDA Forest Service), E. Durant McArthur (USDA Forest Service, Rocky Mt. Research Station), Gerald Rehfeldt (USDA Forest Service, Rocky Mt. Research Station), Logan Sander, Frank Sorensen (USDA Forest Service, Pacific Northwest Research Station), Dale Simpson (Natural Resources Canada), John Syring (Montana State University), Philip Thomas (Royal Botanic Garden, Edinburgh), Michael Wall (Rancho Santa Ana Botanic Garden, Claremont, CA), and Walter Wisura (Rancho Santa Ana Botanic Garden) generously contributed plant material. I am indebted to David Gernandt for sharing unpublished information as well as helpful comments. I am grateful to, John Syring, Susan Huber, Conrad Schoch, Gi-Ho Sung, Mathew Horning, Mark Dasanko, and Christopher Sullivan for advice and guidance with laboratory work and data analysis. Funding was provided by National Science Foundation grants DEB 0317108 and ATOL 0629508 to Aaron Liston and Richard Cronn and the USDA Forest Service Pacific Northwest Research Station.

LITERATURE CITED

Anderson, E., and G. L. Stebbins. 1954. Hybridization as an evolutionary stimulus. Evolution **8**:378-388.

Birky, C. W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103**:513-527.

Borchert, M. 1985. Serotiny and cone-habit variation in populations of *Pinus coulteri* (Pinaceae) in the southern Coast Ranges of California. Madroño **32**:29-48.

Boys, J., M. Cherry, and S. Dayanandan. 2005. Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). Am J Bot **92**:833-841.

Brayshaw, T. C. 1997. Washoe and ponderosa pines on Promontory Hill near Merritt, B.C., Canada. Ann Naturhist Mus Wien **99B**:673-680.

Bucci, G., M. Anzidei, A. Madaghiele, and G. G. Vendramin. 1998. Detection of haplotypic variation and natural hybridization in halepensis-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol Ecol **7**:1633-1643.

Christensen, K. I., and G. H. Dar. 1999. A morphometric study of hybridization between *Pinus mugo* and *P. sylvestris* (Pinaceae). Acta Horti **615**.

Conkle, M. T., and W. B. Critchfield. 1988. Genetic variation and hybridization of ponderosa pine. In "Ponderosa pine: the species and its management", D.M. Baumgartner & James E. Lotan, eds.

Critchfield, W. 1986. Hybridization and classification of the white pines (*Pinus* section *Strobus*). Taxon **35**:647-656.

Critchfield, W., and E. Little. 1966. Geographic Distribution of the Pines of the World. USDA Forest Service Misc. Publication 991:91 pp.

Critchfield, W. B. 1966. Crossability and relationships of the California big-cone pines. USDA Forest Service Research Paper **NC-6**:36-44.

Debreczy, and Racz. 1995. New species and varieties of conifers from Mexico. Phytologia **78**:233-234.

DeVerno, L., and A. Mosseler. 1997. Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis. Can J For Res **27**:1316-1320.

Echt, C. S., L. L. DeVerno, M. Anzidei, and G. G. Vendramin. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol Ecol **7**:307-316.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Ellstrand, N., and K. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci USA **97**:7043-7050.

Emrich, S. J., L. Li, T.-J. Wen, M. Yandeau-Nelson, Y. Fu, L. Guo, H.-H. Chou, S. Aluru, D. A. Ashlock, and P. S. Schnable. 2007. Nearly identical paralogs: implications for maize (*Zea mays* L.) genome evolution. Genetics **175**:429-429.

Epperson, B., F. Telewski, A. Plovanich-Jones, and J. Grimes. 2001. Clinal differentiation and putative hybridididization in a contact zone of *Pinus ponderosa* and *P. arizonica* (Pinaceae). Am J Bot **88**:1052-1057.

Epperson, B. K., M. G. Chung, and F. W. Telewski. 2003. Spatial pattern of allozyme variation in a contact zone of *Pinus ponderosa* and *P. arizonica* (Pinaceae). Am J Bot **90**:25-31.

Farjon, A., and B. Styles. 1997. Flora Neotropica, Monograph 75, *Pinus* (Pinaceae). The New York Botanical Garden, New York, NY, USA.

Furnier, G., and W. Adams. 1986. Geographic patterns of allozyme variation in Jeffrey pine. Am J Bot **73**:1009-1015.

Gernandt, D., G. Geada Lopez, S. Garcia, and A. Liston. 2005. Phylogeny and classification of *Pinus*. Taxon **54**:29-42.

Gernandt, D., A. Liston, and D. Pinero. 2001. Variation in the nrDNA *ITS* of *Pinus* Subsection *Cembroides*: implications for molecular systematics studies of pine species complexes. Mol Phylogenetics and Evol **21**:450-468.

Gerson, E. A., and R. G. Kelsey. 2004. Piperidine alkaloids in North American *Pinus* taxa: implications for chemosystematics. Biochemical Systematics and Ecology **32**:63-74.

Haller, J. R. 1962. Variation and hybridization in ponderosa and Jeffrey pines. Univ Calif Publ Bot **34**:123-165.

Haller, J. R. 1986. Taxonomy and relationships of the mainland and island populations of *Pinus torreyana* (Pinaceae). Syst Bot **1**:39-50.

Haller, J. R. 1965. *Pinus washoensis* in Oregon: taxonomic and evolutionary implications. Am J Bot **52**:646 (abstract).

Helmers, A. E. 1943. The ecological anatomy of ponderosa pine needles. The American Midland Naturalist **29**:55-71.

Hong, Y.-P., A. Krupkin, and S. Strauss. 1993. Chloroplast DNA transgresses species boundaries and evolves at variable rates in the California closed-cone pines (*Pinus radiata, P. muricata*, and *P. attenuata*). Mol Phylogenetics and Evol **2**:322-329.

Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm, and R. S. Lasken. 2003. Unbiased whole-genome amplification directly from clinical samples. Genome Research **13**:954-964.

Huelsenbeck, J. P., and J. P. Ronquist. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572-1574.

Johansen, A. D., and R. G. Latta. 2003. Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. Mol Ecol **12**:293-298.

Karalamangala, R., and D. Nickrent. 1989. An electrophoretic study of representatives of subgenus Diploxylon of *Pinus*. Can J Bot **67**:1750-1759.

Kinlaw, C., and D. Neale. 1997. Complex gene families in pine genomes. Trends in Plant Science **2**:356-359.

Kjer, K. M., J. J. Gillespie, and K. A. Ober. 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst Biol **56**:133-146.

Kral, R. 1993. *Pinus in* N. Morin, ed. Flora of North America. Oxford University Press, Oxford, U.K.

Krupkin, A., A. Liston, and S. Strauss. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot **83**:489-498.

Krutovsky, K., M. Troggio, G. Brown, K. Jermstad, and D. Neale. 2004. Comparative mapping in the Pinaceae. Genetics **168**:447-461.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol **1**:17-24.

Lanner, R. M., and A. M. Phillips. 1992. Natural hybridization and introgression of pinyon pines in northwestern Arizona. Int J Plant Sci **153**:250-257.

Lauria, F. 1996. The identity of *Pinus ponderosa* Douglas ex. C. Lawson (Pinaceae). Linzer biologische Beitrag **28/2**:999-1052.

Lauria, F. 1997. The taxonomic status of *Pinus washoensis* H. Mason & Stockw. (Pinaceae). Ann Naturhist Mus Wien **99B**:655-671.

Ledig, F., and M. Conkle. 1983. Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex Carr.). Evolution **37**:79-85.

Ledig, F. T. 2000. Founder effects and the genetic structure of Coulter pine. J Heredity **91**:307-315.

Ledig, F. T. 1998. Genetic variation in *Pinus*. Pp. 251-280. *in Richardson DM, Editor, Ecology and Biogeography of* Pinus.

Libby, W. 1958. The backcross hybrid Jeffrey x (Jeffrey x Coulter) pine. J Forest **56**:840-842.

Liston, A. 1992. Variation in the chloroplast genes *rpoC1* and *rpoC2* of the genus *Astragalus* (Fabaceae): Evidence from restriction site mapping of a PCR-amplified fragment. . Am J Bot **79**:953-961.

Liston, A., W. Robinson, D. Pinero, and E. Alvarez-Buylla. 1999. Phylogentics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenetics and Evol **11**:95-109.

Liston, A., M. Parker-Defeniks, J. V. Syring, A. Willyard, and R. Cronn. Interspecific phylogenetic analysis enhances intraspecific phylogeographic inference: A case study in *Pinus lambertiana*. Mol. Ecol. **in press**.

Little, E., and W. Critchfield. 1969. Subdivisions of the Genus *Pinus* (Pines). U.S.D.A. Forest Service **Misc Pub 1144**.

Ma, X.-F., A. E. Szmidt, and X.-R. Wang. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol **23**:807-816.

Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21**:260-262.

Martinez, M. 1948. Los Pinos Mexicanos, Segunda Edicion, Mexico.

Mason, H. L., and W. P. Stockwell. 1945. A new pine from Mount Rose, Nevada. Madroño **8**:61-63.

Matos, J., and B. Schaal. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. Evolution **54**:1218-1233.

Matos, J. A. 1998. A coalescent approach to chloroplast genome relationships within and between populations of *Pinus devoniana* in Mexico. Aliso **17**:145-156.

Matos, J. A. 1995. *Pinus hartwegii* and *P. rudis*: A critical assessment. Syst Biol **20**:6-21.

Mirov, N. T. 1961. Composition of gum turpentines of pines. USDA Forest Service Technical Bulletin **1239**.

Mirov, N. T. 1967. The genus *Pinus*. The Ronald Press Company, New York, NY.

Mirov, N. T. 1929. Chemical anlysis of the oleoresins as a means of distinguishing Jeffrey pine and western yellow pine. J Forest **27**:176-187.

Namkoong, G. 1966. Statistical analysis of introgression. Biometrics **22**:488-502.

Niebling, C., and M. Conkle. 1990. Diversity of Washoe pine and comparisons with allozymes of ponderosa pine races. Can J For Res **20**:298-308.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst Biol **53**:47-67.

Paine, T. D., and C. C. Hanlon. 1994. Influence of oleoresin constituents from *Pinus ponderosa* and *Pinus jeffreyi* on growth of mycangial fungi from *Dendroctonus ponderosae* and *Dendroctonus jeffreyi*. Journal of Chemical Ecology **20**:2551-2563.

Patten, A., and S. Brunsfeld. 2002. Evidence of a novel lineage within the *Ponderosae*. Madroño **49**:189-192.

Perry, J. P. 1991. The pines of Mexico and Central America. Timber Press, Portland, OR, USA.

Petit, R. J., R. Bialozyt, P. Garnier-Gere, and A. Hampe. 2004. Ecology and genetics of tree invasons: from recent introductions to Quaternary migrations. Forest Ecology and Management **197**:117-137.

Pollard, D. A., N. I. Venky, A. M. Moses, and M. B. Eisen. 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. PLOS Genetics **2**:1634-1646.

Prager, E., D. Fowler, and A. Wilson. 1976. Rates of evolution in conifers (Pinaceae). Evolution **30**:637-649.

Price, R., A. Liston, and S. Strauss. 1998. Phylogeny and Systematics of *Pinus*. Pp. 49-68. in Richardson DM, Editor, Ecology and Biogeography of *Pinus*.

Rehfeldt, G. 1999a. Systematics and genetic structure of *Ponderosae* taxa (Pinaceae) inhabiting the mountain islands of the Southwest. Am J Bot **86**:741-752.

Rehfeldt, G. 1999b. Systematics and genetic structure of Washoe pine: Applications in conservation genetics. Silvae Genet **48**:167-173.

Rehfeldt, G., B. C. Wilson, S. P. Wells, and R. M. Jeffers. 1996. Phytogeograpic, taxonomic, and genetic implications of phenotypic variation in the Ponderosae of the Southwest. The Southwestern Naturalist **41**:409-418.

Richardson, D. M., and S. I. Higgins. 1998. Pines as invaders in the southern hemisphere. Pp. 450-473. in Richardson DM, Editor, Ecology and Biogeography of *Pinus*.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution **57**:1465-1477.

Rozas, J., J. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Shaw, G. R. 1914. The Genus *Pinus*. Publ. of the Arnold Arboretum **5**.

Shaw, J., E. B. Lickey, and J. T. Beck. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot **92**:142-166.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst Biol **49**:369-381.

Sorensen, F. C., N. L. Mandel, and J. E. Aagaard. 2001. Role of selection versus historical isolation in racial differentiation of ponderosa pine in southern Oregon: an investigation of alternative hypotheses. Can J For Res **31**:1127-1139.

Stead, J. W. 1983. Studies in Central American Pines V: a numerical study of variation in the Pseudostrobus group. Silvae Genet **32**:101-115.

Stebbins, G. L. 1959. The role of hybridization in evolution. Proc of the American Philosophical Society **103**:231-251.

Styles, B. 1993. Genus Pinus: A Mexican Purview. in: Ramamoorthy, editor; Biological diversity of Mexico.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among pine (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot **92**:2086-2100.

Syring, J. V., K. Farrell, R. Businsky, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst Biol **56**:163-181.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437-460.

Tallent, W. H., and E. C. Horning. 1955. *Pinus* alkaloids: The alkaloids of *P. sabiniana* Dougl. and related species. Am Chem Soc Jour **78**:4467-4469.

Temesgen, B., G. R. Brown, D. E. Harry, C. S. Kinlaw, M. M. Sewell, and D. B. Neale. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). Theor Appl Genet **102**:664-675.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research **22**:4673-4680.

Van der Burgh, J. 1973. Holzer der niederrheinischien braunkohlenformation, 2. holzer der braunkohlengruben "Maria Theresia" zu herzogenrath, "zukunft west" zu eschweiler und "victor" (Zulpick mitte) zu zulpich. Nebst einer systematisch-anatomischen bearbeitung der gattung *Pinus*.

Wang, X.-Q., D. Tank, and T. Sang. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol **17**:773-781.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology **7**:256-276.

Willyard, A., J. V. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol **24**:90-101.

Wolff, R. L., F. Pedrono, E. Pasquier, and A. M. Marpeau. 2000. General characteristics of *Pinus* spp. seed fatty acid compositions, and importance of delta5-olefinic acids in the taxonomy and phylogeny of the genus. Lipids **35**:1-22.

Wu, J., K. Krutovskii, and S. Strauss. 1998. Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics **150**:1605-1614.

Zobel, B. 1951. The natural hybrid between Coulter and Jeffrey pines. Evolution **5**:405-413.

The Affinities of *Pinus washoensis* (Pinaceae) to *P. ponderosa* and *P. jeffreyi* Based on Nuclear Microsatellite Loci

Ann Willyard, Aaron Liston, and Richard Cronn

Prepared for submission

INTRODUCTION

The enigmatic pine species *Pinus washoensis* Mason & Stockw. (subg. *Pinus*, section *Trifoliae*) is sometimes considered conspecific with *P. ponderosa* and is sympatric with both *P. ponderosa* Dougl. ex C. Lawson and *P. jeffreyi* Balf. In Chapter 3 of this dissertation, I present results from an exemplar-based study with two to six samples per taxon of all species in subsect. *Ponderosae* using nucleotide sequences of two nuclear loci. This phylogenetic approach provides an important framework for lineages, but reveals a substantial retention of ancestral polymorphism that prevents inference of close relationships. In one gene-tree (*LEA-like*), all samples of *P. ponderosa* and *P. washoensis* resolve apart from all samples of *P. jeffreyi*. However, in the other gene tree (*WD-40*), some *P. ponderosa* and some *P. washoensis* samples resolve in a clade with *P. jeffreyi*, suggesting that a history of reticulate evolution may have shaped these lineages. I review relevant research regarding these three species, then describe this experiment, which uses population-level sampling to explore the relationships between *P. washoensis*, *P. ponderosa*, and *P. jeffreyi*. I will integrate the conclusions from both studies to further my understanding of the origins of *P. washoensis*, testing whether the integration of two very different approaches helps disentangle intraspecific variation and shared ancestral polymorphism from recent introgression, and whether a hypothesized 'footprint' of ancient reticulate evolution is discernable using these methods.

Pinus ponderosa. This impressively wide-ranging species grows from western Canada to Mexico (Critchfield and Little 1966) and from sea level to altitudes approaching 3100 m (Mirov 1967) when considered in the broadest taxonomic sense. Weidman (1939) described morphological and physiological characteristics that differentiated what he termed "ecotypes" of *P. ponderosa* from diverse geographic areas: *P. ponderosa* var. *scopulorum* Engelmann in the Rocky Mountain region; and *P. ponderosa* var. *ponderosa* in the west. Another variety, *P. ponderosa* var. *arizonica* (Engelmann) Shaw that grows in the southwestern United States and Mexico is treated by many taxonomists as a separate species, *P. arizonica* Engelmann (Rehfeldt 1999b). The distinction between var. *scopulorum* and var. *ponderosa* has been supported by growth response in common gardens (Wells 1964b; Wells 1964a), allozymes (Niebling and Conkle 1990) and by test crosses (Conkle and Critchfield

1988). However, biochemistry presents conflicting evidence (see below) that var. *scopulorum* shares a monoterpene profile with one of the races of var. *ponderosa*.

Within var. *ponderosa*, evidence has been presented for three races. Conkle and Critchfield (1988) recommend recognition of a southern California race based on Smith's monoterpene results (Smith 1971). Little more is known about the genetic structure of those localized populations, as the two large geographic races with heritable physiological differences have received the bulk of research attention (Figure 4.1). Compared to the North Plateau race, the Pacific race exhibits rapid one-year height growth, but suffers severe first year winter injury when grown in cold-climate common gardens (Squillace and Silen 1962; Niebling and Conkle 1990). Morphological differences are subtle and describe the frequency of character states within populations rather than fixed differences, hence the taxonomic recognition as races rather than varieties. In comparison to the Pacific race, trees from the North Plateau race tend to have shorter leaves, smaller cones, smaller seeds, a thicker layer of hypodermal cells in leaves (Weidman 1939), darker green foliage, exserted second-year bud scales (Wells 1964b), and dark purple (rather than green) immature ovulate cones (Smith 1981). The racial boundary roughly coincides with the crest of the cascades (Sorensen, Mandel, and Aagaard 2001 and references therein), with the Pacific race generally west of the Cascades and the North Plateau race east of the crest (Figure 4.1). Even where the geographic topology is not abrupt, steep clines have been found between the races in physiological features (Sorensen, Mandel, and Aagaard 2001). Adaptive differences between the races are evident, as the Pacific race grows in summer-dry habitats. Nonetheless, not all evidence is in agreement. Surprisingly, Smith (1967) found that the North Plateau race shares a xylem monoterpene profile with var. *scopulorum.* This variety is morphologically (Squillace and Silen 1962; Conkle and Critchfield 1988) and ecologically (Norris, Jackson, and Betancourt 2006) distinct from var. *ponderosa.* Except for a narrow zone of recent contact, var. *scopulorum* has presumably been geographically separated from var. *ponderosa* for at least 40,000 years by the Great Basin (Critchfield 1984). A fascinating complication to the geographical race story is the evidence that "altitudinal races" exist for both *P. ponderosa* and *P. jeffreyi* (Mirov 1967). When grown in common gardens, high-altitude sources grow slower, and there are heritable differences in both height and diameter (Callaham and Metcalf 1959; Callaham and Liddicoet 1961). The altitudinal differences have been recognized and integrated into seed transfer guidelines required for reforestation

with the goal of replanting locally-adapted plants. However, the hypothesized "racial" differences by altitude have not been directly addressed by taxonomists. Nevertheless, some have proposed that the differences between the North Plateau race and the Pacific race are substantial and well established enough to warrant recognition as distinct species. For the Pacific race, Lauria (1996) suggested that the correct name would be *P. benthamiana* Hartweg.

Pinus jeffreyi. Jeffrey pine overlaps a large portion of the var. *ponderosa* range in California and southern Oregon (Figure 4.1) and can superficially appear quite similar to *P. ponderosa* (Haller 1962). *Pinus jeffreyi* has the clear distinction of being one of only two pine species that produce substantial amounts of the oleoresin heptane; *P. sabiniana* Douglas ex D. Don being the other (Mirov 1929; Mirov 1961). Outside these two species, production of aliphatic hydrocarbons (heptane and undecane) has only been reported in *P. torreyana* Parry ex Carrière, *P. coulteri* D. Don, and *P. oaxacana* Mirov. Despite isolated reports of trace amounts of heptane in leaves and fruits of other conifers (Kolesnikova, Latysh, and Chernodubov 1977), heptane production in xylem is limited to these pine species. Further, production of more than trace amounts plants is considered extremely rare, and only one other plant species in the world (*Pittosporum resiniferum*) has been found that produces a substantial amount of heptane (Savage, Hamilton, and Croteau 1996). Morphological characters distinguishing *P. jeffreyi* from var. *ponderosa* include larger ovulate cones, more numerous scales per cone with long, but generally deflexed prickles, non-resinous buds, and details of needle anatomy (Haller 1962). *Pinus jeffreyi* tends to grow in colder or harsher habitats: high altitude sites, cold-air drainages, or on ultramafic soils (Haller 1962). These habitat differences result in a familiar pattern of low- to mid-altitude var. *ponderosa* stands being gradually replaced by *P. jeffreyi* on higher slopes (ca. 1800 m on the western slope of the Sierra Nevada and ca. 2000 m on the eastern slope). Krupkin et al. (1996) using chloroplast restriction sites, placed *P. jeffreyi* as sister to the California big-cone pines (subsect. *Sabinianae*: *P. sabiniana*, *P. coulteri*, and *P. torreyana*). In Chapter 3 of this dissertation, I show that *P. jeffreyi* resolves in a clade with *Sabinianae* in phylogenies based on two different nuclear loci. Artificial hybrids between *P. jeffreyi* and *P. washoensis* and, as well as between *P. jeffreyi* and *P. ponderosa* are fertile, but natural hybridization (documented by comparing a suite of morphological characters for population-level

sampling) is thought to be limited by the narrow zone of contact due to different habitat preferences and by timing of pollen release (Haller 1962). Nonetheless, *P. jeffreyi* and *P. washoensis* share some ecological (high-altitude sites) and some morphological features (more scales per ovulate cone), are sympatric, and the level of introgression has not been tested at the molecular level. For these reasons, I include *P. jeffreyi* in my sampling and test the hypothesis that it is an isolated species.

Pinus washoensis. As traditionally treated, *P. washoensis sensu stricto* (*s.s.*) occurs in only three high-altitude localities near the California-Nevada border, making it "one of the most narrowly distributed pines of North America" (Niebling and Conkle 1990). At the type locality on Mt. Rose, Nevada, and on nearby Babbitt Peak, California, *P. washoensis* grows on sites that are several hundred meters higher in altitude than nearby var. *ponderosa* (Critchfield and Allenbaugh 1965). In fact, low-altitude var. *ponderosa* stands in this area are replaced by *P. jeffreyi* at increasing altitudes. It is only at the highest altitudes (2100-2500 m) that the ponderosa-like pine named *P. washoensis* is found growing intermixed with *P. jeffreyi*. The third widely recognized *P. washoensis* locality is in the Warner Mountains in the far northeastern corner of California. In this area, there seems to be more of a morphological continuum between the low-altitude var. *ponderosa* and the high-altitude *P. washoensis,* as well as higher morphological variability within *P. washoensis* than on Mt. Rose or Babbitt Peak (Haller 1961; Critchfield 1984). The story in the Warner Mountains is complicated because the low-altitude *P. ponderosa* are apparently in a transition zone between the Pacific race and the North Plateau race (Smith 1981; Critchfield 1984; Niebling and Conkle 1990). Other localities have also been reported where the 'ponderosa pine' resembles *P. washoensis*. Haller (1965) described several high-altitude populations from the Blue Mountains in Oregon and in British Columbia where some individuals fall within the range of morphological variation of *P. washoensis*. A similar report describes the affinity of high-altitude populations in Crater Lake National Park with *P. washoensis* (Mastrogiuseppe and Mastrogiuseppe 1995).

Highly fragmented populations are characteristic of some other high-elevation pine species. In the bird-dispersed *P. albicaulis* Engelmann and *P. flexilis* James, Clark's nutcrackers can disperse seed as far as 22 km, easily transporting seeds of these two species between mountain-top islands (Vander Wall and Balda 1977). In contrast, seeds of *P. jeffreyi*

and *P. ponderosa* are more locally dispersed by wind and scatter-hoarding by rodents and jays (Johnson, Vander Wall, and Borchert 2003). Because *P. washoensis* seed size is similar to that of *P. ponderosa*, seed dispersal is presumably local for this species as well. Thus, recent long-distance dispersal is not a likely explanation for shared ancestry among these disjunct putative *P. washoensis* populations.

Pinus washoensis trees tend to be compact in overall growth form, with short, wide leaves, and smaller ovulate cones that are dark purple when immature, with more scales per cone and more phyllotactic spirals (5, 8 and 13) than typical for *P. ponderosa* (Mason and Stockwell 1945). Despite their presumed narrow endemism, *P. washoensis* populations display allele frequency and heterozygosity patterns comparable to widely distributed conifer species (Niebling and Conkle 1990). In addition, 'physiological plasticity' has been reported in *P. ponderosa* from comparisons between growth of trees in common gardens (Zhang and Cregg 2005). Two of the traits (tree height and leaf length) that differed in *P. ponderosa* trees grown at a harsher site are important in the recognition of *P. washoensis*.

The distinctiveness of *P. washoensis* and its evolutionary relationship to *P. ponderosa* and *P. jeffreyi* remain persistent and interesting questions. Many morphological features are within the range of variability of *P. ponderosa*, and especially fit within the character states reported for the North Plateau race and/or for var. *scopulorum*. Thus, even though *P. washoensis* resembles *P. jeffreyi* in its high-altitude habitat, higher cone scale phyllotaxy, and low seed wing to seed body ratio, it is likely to be more closely related (or even conspecific with) *P. ponderosa* (Niebling and Conkle 1990; Lauria 1996; Brayshaw 1997; Lauria 1997). One early hypothesis suggested that *P. washoensis* arose as a hybrid between *P. ponderosa* and *P. jeffreyi* (Axelrod 1986). This theory has not been supported by allozyme or monoterpene analysis, but the possibility of some introgression (ancient or current) from *P. jeffreyi* into *P. washoensis* has not been ruled out.

Similar monoterpene profiles between *P. washoensis* and the disjunct var. *scopulorum* (Smith 1967) have lead to a second theory that *P. washoensis* arose through contact between the Pacific race (or the North Plateau race) and var. *scopulorum* (Haller 1965). Further, crosses of *P. washoensis* with var. *scopulorum* yield a higher seed set than intraspecific *P. washoensis* crosses (Critchfield 1984). There are several counter arguments to this theory. First, the inheritance of monoterpene profiles is poorly understood and evidence that the shared profile implies identity-by-descent rather than differences in

expression in response to environmental cues is lacking. Second, this intravarietal hybridization theory would imply that the contact between these varieties predated the geographical separation enforced by the Great Basin. Third, the mounting evidence that incomplete mating barriers between distantly related pine species are commonplace diminishes the argument that crossability with var. *scopulorum* implies direct ancestry.

The third theory is that the taxon described as *P. washoensis* merely represents outlying populations of the North Plateau race. In fact, expanding the North Plateau description to include *P. washoensis* accentuates the differences between it and the Pacific race, and perhaps argues that the Pacific race is different enough to be treated as a separate species (Lauria 1997). Arguments for including *P. washoensis* within the North Plateau race include: 1) nearly adjacent distributional ranges (Figure 4.1); 2) genetic similarity based on allozyme data (Niebling and Conkle 1990); 3) phylogenetic clustering inferred from one nuclear gene fragment (287 bps of a nitrate reductase; Patten and Brunsfeld 2002), 4) overlapping values for quantitative traits observed in common-garden tests (Wells 1964a; Rehfeldt 1999a); 5) crossability patterns (Critchfield 1984); and 6) morphological similarities: small ovulate cones, purple-colored immature ovulate cones, and short leaves (Lauria 1997); leaf length, number of resin canals, cone length, number of cone scales, ratio of seed wing length to seed body length (Critchfield 1984; Brayshaw 1997). In further support of this notion, *P. washoensis* habitat may be more similar to typical North Plateau race sites than it appears at first glance. As Critchfield (1984) noted, all three traditional populations are "within sight of the Great Basin sagebrush steppe". Indeed, the morphological continuum observed in the Warner Mountains and at other high-altitude sites in Oregon, Washington, and British Columbia lead Haller (1965) to doubt the status of *P. washoensis* as a separate species.

Not all of the evidence is clear cut, however. An antigenic distance between *P. ponderosa* and *P. washoensis* was reported to be greater than the distance between *P. ponderosa* and either *P. jeffreyi* or *P. sabiniana* (Prager, Fowler, and Wilson 1976). It is worth noting that the ponderosa pine used for this comparison was provided by the Institute of Forest Genetics, Placerville, California, and was likely representative of the Pacific race. As shown in Chapter 3 of this dissertation, further complexity is revealed when other taxa in *Ponderosae* are considered. Species in the western United States appear to be closely related to those in Mexico and Central America as well as to the taxa in subsect. *Sabinianae*. A

study based on four chloroplast loci recently placed *P. washoensis* and the Mexican *P. douglasiana* as unresolved sisters to *P. jeffreyi*, *P. arizonica,* and the Mexican *P. devoniana* and *P. engelmannii* (Eckert and Hall 2006).

Rehfeldt (1999a) used the absence of differentiation in growth characters in common-garden grown seedlings to conclude that *P. washoensis* from Mt. Rose, Babbitt Peak, and the Warner Mountains share a relationship with the North Plateau race. In this study, seeds from Promontory Hill, a putative *P. washoensis* locality in British Columbia (Haller 1965) were used to represent the North Plateau race. However, the altitude for the British Columbia seed collections is unclear. Therefore, an alternative possibility is that the ponderosa pine collections from the British Columbia site are not representative of the North Plateau race, i.e. their shared morphological features ally them instead with *P. washoensis*. Thus, the lack of differentiation reported by Rehfeldt (1999a) might argue for *inclusion* of the British Columbia population within *P. washoensis* rather than for subsuming *P. washoensis* into *P. ponderosa*. In addition, Brayshaw (1997) drew a similar conclusion using seedlings grown from seeds collected at Promontory Hill. Because the altitudes for seed collections in that study are unclear as well, it is uncertain whether it was actually the high-altitude trees that have been attributed to *P. washoensis* that were sampled in his common garden test.

Shared Ancestral Polymorphism vs. Reticulate Ancestry. Conkle and Critchfield (1988) noted that *Ponderosae* are generally recent colonizers, having moved into their present-day ranges within ca. the last 6,000 to 8,000 years. They suggested that "sharp transition zones" may represent recent secondary contact between races, varieties, or species. Thus, while subsuming *P. washoensis* within the North Plateau race may appear to be a simple answer, it may not be the whole story. The phylogenetic trees presented in Chapter 3 of this dissertation do not resolve relationships between *P. ponderosa* and *P. washoensis*, and the different topology inferred from each locus hints that reticulate evolution may have played a role. Unless gene trees are congruent with species trees, a very large number of loci may be needed to infer a species phylogeny (Rokas et al. 2003). In contrast, the lack of coalescence suggests that these nascent species may exhibit a better fit to models applied in population genetics, i.e. panmictic breeding may have been recent enough (as measured by genetic composition, not time) that it remains a useful assumption. If I make the assumption

of panmictic breeding, I can apply a number of different techniques for inferring clustering among individuals as well as measuring genetic distances using allele frequencies. This approach has been demonstrated to be effective using morphological (e.g. Stead 1983) and molecular (e.g. Matos and Schaal 2000) characters to infer relationships among related *Pinus* species.

Experiment Overview. For population-level sampling, one must sample enough alleles per population to make reasonable inferences about their frequencies. Just as important are the number of populations sampled, and an adequate coverage of putative relatives and sources of migrants. In Chapter 3 of this dissertation, I presented a phylogenetic analysis using exemplar samples that infers mutual monophyly between *P. jeffreyi* and a clade that contains all samples of *P. ponderosa* and *P. washoensis* for the *LEA-like* locus. Using the *WD-40* locus, the two clades containing *P. jeffreyi* samples resolve within a comb that also contains two samples of *P. ponderosa* and three samples of *P. washoensis*. Neither locus resolved samples of *P. ponderosa* or *P. washoensis* as mutually monophyletic. In fact, many *P. ponderosa* and *P. washoensis* individuals exhibit quite different relationships on these two locus trees. I used unpublished morphological research (generously shared by John R. Haller) to identify sites where putative *P. washoensis* could be collected in Oregon for comparison with traditional *P. washoensis* from Nevada and California. As a working hypothesis, I called these populations putative *P. washoensis*. My sampling design included 11 populations from three species (described in more detail in Methods): a) two traditional *P. washoensis*; b) two putative *P. washoensis*; c) two var. *ponderosa* populations that are sympatric with *P. washoensis*); d) two var. *ponderosa* populations that are allopatric with *P. washoensis* (one Pacific race and one North Plateau race); and e) three sympatric *P. jeffreyi* populations. I did not collect *P. jeffreyi* near the Pacific race population, but the forest contains that species as well. However, because *P. jeffreyi* does not grow in the Blue Mountains or Idaho (Figure 4.1), my var. *ponderosa* collections from those locations provide populations that are allopatric with *P. jeffreyi* for comparison.

Because divergence times are likely to be much shorter between populations than between species, faster-evolving loci are needed to observe differences between them. This follows directly from coalescent theory, which predicts that under constant population size,

the time to coalescence for a lineage is inversely proportional to the mutation rate (Tajima 1983). The mutation rates for low-copy nuclear loci and chloroplast loci in *Pinus* are on the order of 10^{-9}, and 10^{-10} substitutions per site per year, respectively (Willyard et al. 2007). In contrast, mutation rates for nuclear microsatellite loci (nSSRs) in plants are many orders of magnitude higher, perhaps on the order of 10^{-3} substitutions per site per year (Udupa and Baum 2001).

Exonic nuclear microsatellites (nSSRs) have been developed for a number of pine species, and limited trans-species amplification has been demonstrated, even across distantly related pine taxa (e.g. Shepherd et al. 2002). In addition to biparental inheritance, nSSRs offer an opportunity to make numerous, independent genetic estimates. Further, sampling the diploid nuclear genome provides twice as many alleles per reaction as a haploid organellar genome and allows direct estimation of heterozygosity. It has been suggested that nSSRs may not be neutral (Turpeinen et al. 2001), but this risk is inherent in any molecular or morphological marker. A further problem with SSRs in either the nuclear or chloroplast genome is that their relatively fast evolutionary rate in combination with unknown constraints on repeat number may lead to size homoplasy that obscures relationships when used in genetically distant comparisons. Despite these limitations, nSSRs have proven to be useful genetic markers between closely related species (e.g. Matsuoka et al. 2002 and references therein). A substantial effort is required for *de novo* development of nSSRs, but several loci developed for *P. taeda* (another species in subg. *Pinus*) have been shown to amplify in *P. ponderosa* (Liewlaksaneeyanawin et al. 2004). Microsatellites developed for other subg. *Pinus* species (Smith and Devey 1994; Karhu et al. 2000; Shepherd et al. 2002) are also available. The low-cost, high-throughput options available for fragment analysis provide an opportunity for the analysis of multiple loci, thus with the potential to compensate for potential limitations.

Chloroplast microsatellites (cpSSRs) also meet the criterion of faster evolution, and have been used to infer population structure in pines, e.g. (Marshall, Newton, and Ritland 2002). In all plants, the non-recombining nature of the chloroplast genome creates a single organellar genealogy, which can be an advantage in inferring relationships. In *Pinus*, mating barriers are frequently incomplete between species within a section (Critchfield 1966; Critchfield 1984; Critchfield 1986; Conkle and Critchfield 1988), leading to occasional natural interspecific hybridization. These hybrids may be less fit and/or less fertile, and

backcrosses to one parent will reduce the nuclear introgression from the second parental species in a lineage. However, an intact chloroplast lineage from a second species can persist in a population. This tendency may be exaggerated in pines due to paternal inheritance of chloroplasts (Neale and Sederoff 1989), which increases the possibility of long-distance chloroplast movement via pollen. Inter-specific chloroplast introgression from what may have been a limited number hybridization events has been documented in pines (Hong, Krupkin, and Strauss 1993; Matos and Schaal 2000; Liston et al. in press), suggesting that chloroplast data can be used to infer hybridization when used in conjunction with nuclear data in closely related, sympatric *Pinus* species. I elected to combine a chloroplast restriction enzyme recognition site that appears to differentiate the ponderosa lineage from other *Pinus* species with my nSSR data to gain information about potential introgression patterns.

Hypotheses. I selected two hypotheses to test (Table 4.1). The first hypothesis (H_1), states that *P. jeffreyi* is distinct from var. *ponderosa s.l.* (i.e. var. *ponderosa* + *P. washoensis*). Support for H_1 will allow us to focus my second hypothesis on relationships between var. *ponderosa* and *P. washoensis*. The second hypothesis tests whether *P. washoensis* fits one of three alternatives: H_{2A} states that *P. washoensis* represents relictual populations of a formerly widespread, unique taxon that have become isolated on these mountainous 'islands'. Specifically, the first alternative tests whether the high-altitude populations in Oregon are more similar to the traditional *P. washoensis* populations than these four populations are to var. *ponderosa s.s.* An alternative (H_{2B}) also states that *P. washoensis* is unique, but restricted to the traditional sites. Under H_{2B}, the putative *P. washoensis* populations from Oregon are grouped with var. *ponderosa*. The third alternative (H_{2C}) states that *P. washoensis* populations represent high-altitude stands of the North Plateau race. This alternative implies that the observed subtle morphological differences are due to either selection (Callaham's (1959) "altitudinal races"), or are merely exaggerated by variable expression of the highly plastic ponderosa pine genome under these harsh conditions and short growing seasons. Thus, populations of both the North Plateau race of var. *ponderosa* and *P. washoensis* would be considered "var. *ponderosa s.l.*". Under alternative H_{2C}, I also ask if there is evidence that reticulate evolution (including introgression with *P. jeffreyi*) has contributed to the variation observed in extant populations.

MATERIALS AND METHODS

Plant Sampling. Leaf samples of *P. washoensis*, var. *ponderosa*, and *P. jeffreyi* were collected from ca. 20 individual trees at each of seven geographic locations in California, Idaho, Nevada, and Oregon, yielding a total of 11 sampled populations (Figure 4.1; Table 4.2). At three of these sites (Blue Mountains, Likely, and Mt. Rose), population collections were made for more than one putative species.

My assignment of collections to *P. washoensis* was made on geographical rather than morphological criteria. Traditional *P. washoensis* populations were collected at Mt. Rose, Nevada (WR; the type locality) and the Warner Mountains (WL, near Likely, California). For WR, individuals were sampled at a well-documented population adjacent to the type locality (Mason and Stockwell 1945). The WL collections were made at another well-documented site at Patterson Meadows (Smith 1971). In addition, putative *P. washoensis* collections were made at two high-altitude sites in Oregon. Locations for the Oregon *P. washoensis* collections were based on unpublished information provided by John R. Haller (pers. comm.). At the Abert Rim and Blue Mountains sites, I collected the highest altitude individuals as putative *P. washoensis*. At the Blue Mountain site, I also collected the lowest-altitude individuals as var. *ponderosa*.

The study also included collections from two var. *ponderosa* locations that are allopatric with *P. washoensis*: Grangeville, Idaho (North Plateau race) and Shasta County, California (Pacific race). None of my collection sites has a record of previous reforestation, but this cannot be known with certainty. I minimized within-family collections by choosing trees at least 30 m apart. Where possible, I collected foliage from trees where immature cones containing mature seeds were also available. Samples were refrigerated for up to one week, and then frozen until processed. A herbarium voucher from each of the 11 populations has been deposited (OSC).

Nuclear Microsatellite Loci (nSSRs). DNA was isolated using the CTAB procedure described by Chen and Ronald (1999), modified by lysing leaves in 2 ml tubes with garnet sand and a ceramic bead in a FastPrep® instrument (Qbiogene, Morgan Irvine, California). The proprietary polysaccharide-binding resin from the Nucleon PhytoPure System (Tepnel Lifecodes Corporation, Samford, Connecticut) was added to DNA samples

after isolation. Further purification of recalcitrant DNA samples was performed using the Qiaquick Gel Extraction Kit (Qiagen), designed for extraction of DNA fragments from agarose gel. I found that many DNA samples extracted from *Pinus* leaves which initially failed to PCR-amplify my nSSR loci were successful after processing with this kit.

I tested primer pairs for nSSRs developed from the *P. taeda* (subg. *Pinus*, sect. *Trifoliae*, subsect. *Australes*) genome for cross-species transfer. Loci that have been shown to amplify polymorphic alleles in one or more species of subsect. *Ponderosae* were given highest priority (Liewlaksaneeyanawin et al. 2004), followed by loci that were polymorphic in other species in sect. *Trifoliae* (Kutil and Williams 2001). Seventeen nSSR loci (Table 4.3) were screened for PCR amplification in my three target species. Eight loci that yielded single amplicons of the expected size (as visualized by agarose gel electrophoresis) in all three species were chosen to examine population structure and species relationships (Table 4.4).

Forward primers for these eight loci were labeled with a 5' fluorophore (6-FAM or HEX, Integrated DNA Technology, Coralville, Iowa; or NED, Applied Biosystems, Foster City, California; Table 4.4). Polymerase chain reactions (PCR) were carried out in 10 µl volumes with ca. 100 ng genomic DNA, 1.5 mM $MgCl_2$, 0.5 µM of each primer, 0.2 mM dNTPs, and 10% bovine serum albumin in a buffer of 10 mM Tris-H_{CL} [pH 9.0] and 50 mM K_{cl}. The cycling program used an initial denaturing step of 4 minutes at 94°C; 32 cycles at 94°C for 30 seconds, annealing at either 53°C or 60°C (Table 4.4) for 30 seconds and extension at 65°C for 1 minute; then a final extension step of 65°C for 3 minutes.

PCR amplifications were diluted 1:20 and amplifications from the same individual were pooled into two different multiplexed mixtures for fragment size analysis without overlap of color channels (Table 4.4). The first multiplexed mixture contained: *PtTX3098* (FAM, 173 to 186 bps) + *PtTX2123* (HEX, 185 to 203 bps) + *LOP11* (NED, 244 bps) + *PtTX3030* (FAM, 312 to 352 bps). The second multiplexed mixture contained *LOP5* (FAM, 160 to 206 bps) + *PtTX3118* (HEX, 205 to 229 bps) + *PtTX2128* (NED, 237 to 246 bps) + *PtTX3025* (FAM, 264 to 272 bps). Further economy was realized by pooling some batches with up to three smaller-sized amplicons from another experiment. Fragment lengths were estimated by the Center for Genome Research and Biocomputing (Oregon State University, Corvallis, Oregon) on an ABI 3100 capillary DNA sequencer (Applied Biosystems, Foster City, California) with a GS500 ROX (red, 35 to 500 bp) internal size standard. Reverse

primers for three loci where "stuttering" was observed in the first batch (*PtTX2128, PtTX3025*, and *PtTX3030*) were modified by adding a PIG-tail sequence (GTTTCTT) to the 5' end. This modification results in nearly 100% adenylation of the 3' end of the amplicon, yielding consistent-sized fragments for genotyping (Brownstein, Carpten, and Smith 1996). Fragment sizes reported here include the extra 7 bps for the PIG-tail sequence and one bp for the adenosine for those loci. PCR was repeated for several putative homozygotes for each species and locus. The amplified fragments were sequenced as described in Chapter 3 of this dissertation and used to infer repeat motifs, number of repeats, and the length of flanking sequences.

Chloroplast haplotypes. As described in Chapter 3 of this dissertation, I have identified a "ponderosa-style" chloroplast haplotype in three species: *P. washoensis*, *P. ponderosa*, and *P. jeffreyi*. Using the *Dra*I restriction site, I did not observe the "ponderosa-style" haplotype in any other *Ponderosae* species, or in species outside of this subsection. Thirty-one individuals from this study were assayed for this haplotype: all individuals from the WR population, six *P. jeffreyi*, five *P. ponderosa*, and an additional two *P. washoensis* samples.

Data Analysis. GENOTYPER software (Applied Biosystems) was used to set the internal size standards and generate electropherograms according to manufacturer recommendations. Electropherograms were reviewed manually to verify that samples generated one or two peaks as expected with single-locus amplification in a diploid species. Where necessary, machine-generated sizes were "binned" for di- and tri-nucleotide motifs by rounding up to 1 bp to the nearest repeat size. Only six polymorphic loci that produced single (presumed homozygotic) or double (presumed heterozygotic) fragment sizes in every individual were used to estimate population parameters.

I assumed that missing data points were due to null homozygotes rather than technical failures after two PCR reactions and at least one additional PCR optimization using less stringent PCR conditions failed. However, amplification failure in 18 individuals (~8% of the original 216 samples) of five or more of seven attempted loci was attributed to poor DNA quality and those samples were removed from the data set. For analyses that are sensitive to missing data, calculations were further limited to the 186 individuals for which I

obtained allele sizes for every locus (i.e. 0% missing data). This strategy avoids noise caused by interpolation of missing data, but has the potential to introduce bias by excluding individuals with null alleles. To address this issue, the frequency of null alleles was estimated using the EM algorithm (Dempster, Laird, and Rubin 1977) in FREENA software (Chapuis and Estoup 2007) using 10,000 bootstrap replicates for the 198 individuals where three or more loci successfully amplified. The average frequency of estimated null alleles was calculated by locus and by population. The FREENA software also calculates a 'corrected' estimate of the coancestry correlation (F_{st}; Weir and Cockerham 1984) between population pairs using an "excluding null alleles" method. First, the null and visible allele frequencies are estimated as in Dempster, Laird, and Rubin (1977), and then F_{st} is estimated from the corrected data set by restricting the estimation to visible allele sizes. The corrected pairwise F_{st} matrix inferred using this method was compared to an AMOVA-based estimate using a Mantel test (9,999 permutations) with GENALEX software (vers. 6; Peakall and Smouse 2006). In addition, I used the population pairwise chord distance measure (D_c), corrected by FREENA for the estimated null allele frequency (D_{cc}), and compared it to the uncorrected D_c distance described below.

Expected heterozygosity (H_e) and allele counts were calculated in GENALEX. Exact P-values for Hardy-Weinberg equilibrium (HWE) were estimated by locus (across 11 populations) and by population (across six loci) with the Markov chain method (10,000 dememorizations, 100 batches, 5,000 iterations per batch) using GENEPOP software (vers. 3.4; Raymond and Rousset 1995).

Genetic linkage information on the nSSRs selected for this study is limited. Zhou et al. (2003) observed weak linkage between *PtTX2128* and *PtTX3018* (20 cM distance), and found that *PtTX2123, PtTX3030,*and *PtTX3118* are not linked to each other or to *PtTX2128*/*PtTX3018*. Their linkage map does not include four of the loci selected for this study (*LOP5, LOP11, PtTX3025*, and *PtTX3098*). Loci *PtTX2123* and *PtTX3018* were also mapped to different linkage groups in *P. caribaea* var. *hondurensis* (Shepherd et al. 2003). In lieu of genetic linkage mapping, genotypic linkage disequilibrium was tested for six loci in the 198 individuals sampled in this study. A Fisher exact test was performed on contingency tables for pairs of loci in each sample using a Markov chain (10,000 dememorizations, 100 batches, 5,000 iterations per batch) with GENEPOP.

Pairwise population coancestry coefficient was estimated with the F_{st} statistic for six loci and 186 individuals (0% missing data) using a hierarchical analysis of molecular variance (AMOVA; Excoffier, Smouse, and Quattro 1992) in GENALEX. The codominant allelic option was chosen to implement an infinite alleles model (IAM), and 9,999 random permutations were used to estimate the probability (after Bonferroni correction for multiple substitutions (Rice 1989)) that there is no genetic differentiation between the population pair.

Pairwise population distances were estimated four ways to provide a comparison between three distinct evolutionary models for nSSR divergence. The first method, Nei's standard genetic distance (D_a; Nei 1978), assumes an IAM. The second method, $\delta\mu^2$ (Goldstein et al. 1995) assumes a stepwise mutation model (SMM) by using the inferred number of repeats in fragments of a given length (Table 4.4). The third method uses a geometric chord distance (D_c; Cavalli-Sforza and Edwards 1967) that is not based on a nucleotide substitution model. Fourth, I recalculated the D_c distance using a correction for the estimated frequency of null alleles as described above (D_{cc}; Chapuis and Estoup 2007). All distance estimates were calculated using MSA software (vers. 2.32; Dieringer and Schlotterer 2003), and each was used to create a neighbor-joining dendrogram with MEGA software (vers. 3.1; Kumar et al. 2004). Because the relationships inferred by these four metrics differed, I used a series of pairwise tests to compare the distance matrices. Mantel tests (Mantel 1967) were conducted in GENALEX using 9,999 permutations for all 11 populations, and also for the data set restricted to just the eight populations representing var. *ponderosa s.l.*

Geographic distances between each of the seven collection sites (Table 4.2) were estimated with a great-circle approximation using a web-based calculator available at http://home.hiwaay.net/~taylorc/toolbox/geography/geodist.cgi. Mantel tests (Mantel 1967; Smouse, Long, and Sokal 1986; Smouse and Long 1992) were performed in GENALEX using 9,999 permutations to evaluate the relationship between log 10-transformed genetic distance matrices and the geographic distance matrix. Genetic differences estimated by each of the four methods (D_a, $\delta\mu^2$, D_c, and D_{cc}), were tested for four data sets (all 11 populations, eight populations excluding *P. jeffreyi*, seven populations also excluding PS, and six populations also excluding PG). My reason for repeating the test without the PS site is that it is the only population representing the Pacific race, and may thus be a genetic outlier. By

also excluding PG, I test just the geographically proximal PB, PL, WA, WB, WL, and WR populations.

Principal Components Analysis (PCA) was conducted using NTSYSpc (vers. 2.11T; Exeter Software, Setauket, NY) to search for patterns between the populations. Similarity was computed over allele frequencies to create a covariance matrix and an Eigen analysis was used to extract the first three axes of this matrix to project population relationships. Calculations were made for two data sets using six loci: 1) all 11 populations (186 samples; 0% missing data); and 2) eight populations (just var. *ponderosa s.l.* populations; 131 samples; 0% missing data).

Principal COordinates analysis (PCO) was performed with six loci using a multivariate matrix between individuals of: 1) all 11 populations (186 samples; 0% missing data) and 2) between individuals of 8 populations (just var. *ponderosa s.l.* populations; 131 samples; 0% missing data). Calculations were performed in GENALEX. Individuals were plotted along the first three major axes of variation to visualize clustering.

Population structure was also investigated with model-based clustering simulations in STRUCTURE software (vers. 2; Pritchard, Stephens, and Donnelly 2000). This Bayesian method estimates genetic admixture within individuals for multiple loci based on a range of prior assumptions of the number of homogeneous groups (K). This yields two inferences. First, a likelihood estimate ($\Pr(X|K)$) for each prior value of K can provide evidence for the most probable value of K (i.e. the number of groups in the sample). Second, inferred admixture between individuals can be used to assign individuals to groups and to suggest the level of shared ancestry. I applied two ancestry models to my data: admixture (assumes unlinked loci), and linkage (attempts to account for the admixture linkage disequilibrium that results from the correlation between weakly linked loci (Falush, Stephens, and Pritchard 2003)). Because of linkage disequilibrium inferred between *PtTX3030* and *PtTX3098* (see Results), I compared results of the admixture and linkage models to assess the reliability of an assumption of independence for these two weakly linked loci in this data set. I tested 24 parameter combinations to evaluate the robustness and consistency of the inferences. Specifically, I varied 1) ancestry model as admixture or linkage; 2) allele frequencies as independent or correlated; 3) whether clustering was assisted with the sample's population origin; 4) whether or not only samples from the "true" species (i.e. excluding putative *P. washoensis* populations WA and WB) were used to update the allele frequencies; and 5) by

using data sets that included 186 samples from all 11 populations or a data set restricted to the 131 samples from 8 populations (just var. *ponderosa s.l.*). For each parameter combination, two simulations (to observe convergence) for each value of *K* from 1 to 5 were conducted using a burn-in period of 50,000 discarded iterations, followed by one million Markov Chain Monte Carlo iterations. I plotted *ln* Pr(X|K) from the run with the lowest variance for each parameter combination and each value of *K* to look for a consistent "plateau" to aid in choosing the most probable value for K. One of the models (admixture ancestry, correlated allele frequencies, and clustering assisted with the sample's population origin) was chosen to display inferred admixture. Results of ten independent simulations using this model were permuted using CLUMPP (vers 1.0; Jakobsson and Rosenbeg 2006). Bar charts were created from the mean permuted admixture for each individual. I also computed the posterior probability (Pr(*K*|X) by assuming a uniform prior (K=1 to 5) using Bayes rule (Pritchard, Stephens, and Donnelly 2000).

RESULTS

One of the eight loci selected for population-level sampling (*LOP11*), was monomorphic in 48 individuals, and was excluded from analysis. The remaining seven loci (Table 4.4) were polymorphic, yielding 69 alleles and from 5 to 22 alleles per locus in my population sample. These seven nSSRs included four tri-nucleotide repeats, two di-nucleotide repeats, and one mono-nucleotide repeat, with inferred number of repeats ranging from 4 to 28 (Table 4.4). At least four of the seven loci were successfully amplified in 198 individuals, yielding a mean sample size of 18.0 individual plants per population (range 7 to 22; Table 4.5). Although locus *PtTX3118* amplified 184 apparently well-behaved samples, 14 individuals (evenly spaced across the three target species) revealed three to four comparable-strength fragment sizes. The amplification of more than one locus with a primer pair violates presumptions of homozygosity for single bands and heterozygosity for double bands that are required for an analysis of diploid individuals based on fragment sizes. For this reason, *PtTX3118* was excluded and population inferences were built over the remaining six loci. For these six useful nSSR loci, mean H_e was 0.573. The range across populations was 0.456 to 0.684 (Table 4.5), while H_e across loci ranged from 0.375 to 0.815 (Table 4.4).

When all individuals with a missing locus score are removed from the data set, 186 samples remain, yielding 59 alleles with a mean of 5.1 alleles per population per locus (Table 4.5). Sample sizes across populations were fairly evenly distributed, but the number of individuals with six loci was low for two populations: WA with N=7 and PS with N=10. However, only WA was more than one S.E. below the mean for the total number of alleles (N_a=25).

Private alleles for populations were infrequent. For the data set with no missing data, eight alleles were private to five populations (three in JT, two in WR, and one each in JL, JR, and WL). There were 20 alleles which differentiated *P. jeffreyi* from var. *ponderosa s.l.* Eight of these alleles were private to *P. jeffreyi*, and twelve were shared by var. *ponderosa* and *P. washoensis.* By locus, *LOP5* provided nine of these distinguishing alleles, *PtTX3030* yielded seven, and *PtTX2123*, *PtTX2128*, *PtTX3025*, and *PtTX3098* provided one each. Therefore, private alleles support *P. jeffreyi* as an independent species (H_1). Of the 59 alleles in the entire data set, 43 alleles were observed in *P. jeffreyi* and 51 alleles were observed in var. *ponderosa s.l.*

Frequencies of null alleles in this data set are not trivial. Using the techniques described in Methods, the estimated mean frequency of null alleles in the 198-sample data set is 0.095 (Table 4.6). The range across loci is substantial (from zero for *PtTX3098* to ca. 0.23 for *PtTX3025* and *PtTX3030*). The range across populations is more evenly distributed (0.074 to 0.135), but still varies two-fold. Comparisons between the corrected vs. raw F_{st} and genetic distance are presented below.

For 11 populations, random union of gametes cannot be rejected for *LOP5* (P=0.81) and *PtTX2123* (P=0.46), but HWE is rejected for four loci: *PtTX2128*, *PtTX3025, PtTX3030*, and *PtTX3098* (P<0.00). In addition, random mating was rejected for each of the 11 populations using allelic counts across 6 loci. Overall, an excess of homozygotes were scored (537 observed vs. 460 expected).

Tests of genotypic linkage disequilibrium between six pairs of loci were not significant for 14 pairwise comparisons, including the pair where weak linkage was previously reported (*PtTX2128* and *PtTX3018*). However, the null hypothesis that genotypes at one locus are independent from another was rejected between *PtTX3030* and *PtTX3098* (P=0.00934). This inferred linkage may violate assumptions of independence expected by

several model-based analyses. Therefore, I included several additional tests (see below) to assess the impact of linkage disequilibrium between two of my six loci.

Population Relationships Based on Allele Frequencies. F_{st} estimates across six loci between population pairs as inferred from an AMOVA revealed significant differentiation between each population of *P. jeffreyi* compared to each var. *ponderosa s.l.* population (Table 4.7). Differences within var. *ponderosa s.l.* populations are more complicated. Significant differences were found for five of the 28 population pairs using F_{st}. Two pairs with significant differences are between a traditional *P. washoensis* population and a North Plateau race population (PB-WR and PG-WL). Two pairs with significant differences are between putative *P. washoensis* populations and either the North Plateau race or the Pacific race (PG-WB and PS-WA). One significant difference is between populations within the North Plateau race (PB-PG). A Mantel test to compare the F_{st} matrix estimated using AMOVA (above the diagonal in Table 4.7) with the F_{st} matrix corrected for the inferred frequency of null alleles (below the diagonal in Table 4.7) shows that the two methods are positively correlated (R^2=0.8407; y=0.786 + 0.0062). However, corrected F_{st} estimates are ca. 79% lower than raw estimates.

When populations are pooled to test Hypothesis 1, a significant difference is found (F_{st} = 0.219; Table 4.8) for *P. jeffreyi* vs. var. *ponderosa s.l.* Thus, this test supports *P. jeffreyi* as an independent species. Because H_1 was supported by this test, I used pairwise F_{st} to evaluate alternatives for Hypothesis 2 (Table 4.1). Despite the low F_{st} between var. *ponderosa s.s.* and *P. washoensis s.l.* (F_{st} = 0.010; Table 4.8, alternative H_{2A}), this result is statistically significant, and alternative H_{2B} is not. Thus, there is some support, albeit weak, for grouping the putative *P. washoensis* populations from Oregon (WA and WB) with traditional *P. washoensis*. However, this result does not provide overwhelming evidence to reject the third alternative (H_{2C}), namely that there is no significant structure among var. *ponderosa s.l.* populations.

The first three PCA axes account for a cumulative 83.3% of an 11-population comparison and 75.5% of an 8-population comparison analyzing only var. *ponderosa s.l.* (Figure 4.2). A strong differentiation between the three *P. jeffreyi* populations and the eight var. *ponderosa s.l.* populations is evident using this method (Figure 4.2, A), lending further support for H_1. When I "zoom" into var. *ponderosa s.l.* by reanalyzing just those eight

populations (Figure 4.2, B), it is possible to draw lines intersecting axis 1 that divide populations into the Pacific race (a circle), the North Plateau race (squares), and *P. washoensis* (triangles). The "clustering" is weak, and this axis represents only ca. 34% of the variability. Nonetheless, this result lends some support for H_{2A} over H_{2B}. As with the F_{st} results, PCA does not provide evidence for rejecting H_{2C}, namely that there is no significant structure among var. *ponderosa s.l.*

Population relationships inferred from genetic distances. Results of four different metrics for pairwise genetic distances are presented in Figure 4.3. All four methods place the three *P. jeffreyi* populations in a separate group, lending additional support for hypothesis H_1. Within var. *ponderosa s.l.*, a comparison of the trees reveals that each method of genetic distance infers different relationships. This suggests that the inter-population differences are either too weak to be revealed at this level, or that there is no structure among them (perhaps supporting alternative H_{2C}). A cluster that unites PB and WL is present in two of the four NJ trees. In two NJ trees, PG branches near the remaining var. *ponderosa s.l.* populations, while this position is represented by PB for the other two trees. There is no support for groups representing either alternative for Hypothesis 2 (Table 4.1). Further, the chord distance corrected as described in Methods for the inferred frequency of null alleles (Figure 4.3, D) shows completely different relationships within var. *ponderosa s.l.* than the uncorrected chord distance (Figure 4.3, B). Interspecific genetic distances between *P. jeffreyi* and var. *ponderosa s.l.* appear to be adequately measured with this technique. Thus, there is no indication that geographically distant populations (PS and PG), or genetically less related (Pacific race population PS), are distorting this estimation by acting as outliers.

Because relationships visualized with NJ trees are so different within var. *ponderosa s.l.*, I investigated whether the distance matrices are significantly different. Table 4.9 shows the coefficient of determination (R^2) for pairwise distance matrix Mantel tests. When all 11 populations were included (below the diagonal), tests show that each pair of distance matrices are positively correlated. Further, the chord distance corrected for the inferred frequency of null alleles (D_{cc}) is positively correlated with the uncorrected chord distance (D_c). However, when the distance matrices are restricted to the eight populations representing var. *ponderosa s.l.* (above the diagonal), only one correlation is observed (between D_a and D_c), and I find that D_{cc} is not correlated with D_c.

Next, Mantel tests between each genetic matrix and the geographic distance matrix indicate that there is no significant difference between them ($\alpha = 0.05$) when all 11 populations are compared. For matrices that exclude *P. jeffreyi*, only the D_{cc} method shows a significant departure from the geographic matrix: $R^2 = 0.2747$, $P = 0.007$ with eight populations; $R^2 = 0.2883$, $P = 0.002$ with seven populations excluding PS; and $R^2 = 0.2188$, $P = 0.045$ with six populations also excluding PG (see Methods for rationale).

Population relationships inferred from individuals. The first three PCO axes account for 74.9% of the variance in the 186-individual comparison (Figure 4.4). This method shows a cluster of all *P. jeffreyi* samples at one end of the first axis, except for two field-identified putative hybrid individuals that are intermingled with var. *ponderosa s.l.* samples. In the 131-individual data set with just var. *ponderosa s.l.*, the first three axes account for 65.9% of the distance (Figure 4.5). No clustering by population or by either alternative for H_2 (Table 4.1) was observed.

Inferences of K, the number of homogeneous groups, with STRUCTURE, are complex. Figure 4.6 plots the results of *ln* Pr(X|K) for the 24 models I tested for five values of K. For the 11-population data set, the graph of *ln* Pr(X|K) "begins to plateau" at K=2 in all 12 parameter combinations that were tested (Figure 4.6; triangles). With *P. jeffreyi* accounting for one of the 'populations', this leaves an inference of K=1 for all individuals within *var. ponderosa s.l.* The graph for the eight-population data set could be interpreted as "beginning to plateau" at either K=1, or K=2 (Figure 4.6; squares), indicating that var. *ponderosa s.l.* is comprised of either one or two homogeneous groups.

The estimated posterior probabilities (Pr(K|X; see Methods) indicate a higher probability for K=4 or K=5 for most models in the 11-population data set and K=3 or K=4 in the eight-population data set. These estimates are clearly unreasonable because the variances increase rapidly (Evanno, Regnaut, and Goudet 2005), and they are not supported by the individual admixture inferences described below.

Figure 4.7 presents the mean permuted admixture for each individual sample from ten simulations using one set of model parameters with the 11-population data set. For K=2 to K=5, differentiation between the three populations of *P. jeffreyi* and all eight populations of var. *ponderosa s.l.* individuals is clear. Two of the four *P. jeffreyi* individuals that were noted as putative hybrids based on some intermediate morphological characteristics show a

substantial proportion of admixture, and these are the same two individuals that were intermingled with var. *ponderosa s.l.* individuals in PCO.

There are also two var. *ponderosa s.l.* individuals with high proportions of inferred *P. jeffreyi* ancestry. Notably, one of these var. *ponderosa s.l.* individuals was collected at PB (Blue Mountains) and the other at PG (Grangeville, Idaho). Both of these sites are well outside the current geographic range of *P. jeffreyi*. These results corroborate support for hypothesis H_1; putative hybridization is detected, but appears to be limited to less than 1% of the individuals sampled in this study. Further, the admixture results do not provide evidence of introgression from *P. jeffreyi* into var. *ponderosa,* from *P. jeffreyi* into *P. washoensis,* or vice versa. Except for three clearly admixed *P. jeffreyi* individuals, inferred admixture from var. *ponderosa s.l.* is extremely low. As noted,, var. *ponderosa* individuals that are sympatric (PL, PS) and allopatric (PB, PG) with *P. jeffreyi* have similar low levels of inferred nuclear admixture with *P. jeffreyi*. Further, *P. washoensis* and var. *ponderosa* individuals have similar low levels of inferred admixture with *P. jeffreyi*. In fact, *P. ponderosa* shows more inferred admixture than *P. washoensis*.

Within var. *ponderosa s.l.*, the bar charts (Figure 4.7) suggest that each *individual*, rather than each *population*, is admixed for the number of homogeneous groups (K) being tested. In other words, no grouping of populations is inferred by these simulations. I do not find evidence from admixture to support either alternative for *P. washoensis* as a distinct species (H_{2A} or H_{2B}). Thus, this test supports alternative H_{2C}, that *P. washoensis* is encompassed within the diversity of the North Plateau race. The slight plateau of *ln* Pr(X|K) at K=2 when *P. jeffreyi* is excluded (Figure 4.6, squares) and the admixture WITHIN each individual rather than by population (e.g. at K=3 when *P. jeffreyi* is included; Figure 4.7) could be interpreted as lending some support for the possibility that ancient reticulate evolution left a legacy of genetic diversity that is still maintained within these populations .

I compared models with and without the assumption of linkage because of previously reported weak linkage between two loci and the inferred linkage I found between two different loci. There was not a substantial difference between the ancestry and linkage models. In fact, models with independent allele frequencies performed slightly better than those with correlated allele frequencies in the var. *ponderosa s.l.* data set, suggesting that an assumption of independence for these six loci in this data set may be warranted.

Chloroplast haplotypes. Results for 31 individuals are shown along the top of Figure 4.7. The "ponderosa-style" haplotype was identified in all 20 *P. washoensis* individuals that were tested, including the entire WR population. In *P. ponderosa*, this haplotype was evident in four samples, including the individual with the highest inferred nuclear admixture in the PG population. One *P. ponderosa* sample from the WL population lacked the "ponderosa-style" haplotype. Four *P. jeffreyi* individuals lacked the "ponderosa-style" haplotype, as expected from the results presented in Chapter 3 of this dissertation. However, three *P. jeffreyi* individuals contained the "ponderosa-style" haplotype, including the two accessions with a high proportion of inferred nuclear admixture.

DISCUSSION

This study analyzed 11 populations using a mean of 16.9 individuals and 30.5 alleles per population with no missing data for six nSSRs (Table 4.5). Thus, my sampling sizes were comparable to many recent studies reporting population-level inferences. The number of individuals sampled at the Abert Rim (population code WA) was small due to the limited number and scattered distribution of trees at this site. I decided to retain this population because analyses at the individual level (e.g. PCO and admixture) are informative irrespective of population sample size.

These data and analyses were adequate to address Hypothesis 1. *Pinus jeffreyi* is clearly distinct from var. *ponderosa s.l.*, supporting hypothesis H_1. Admixture presumed to be the result of recent introgression between *P. jeffreyi* and var. *ponderosa s.l.* was inferred with nSSR data in two *P. jeffreyi* individuals and not apparent in the remaining samples of this species (Figure 4.7). Chloroplast haplotypes also support the distinctiveness of *P. jeffreyi* from var. *ponderosa s.l.* While none of the six *P. jeffreyi* individuals tested in Chapter 3 of this dissertation contained the "ponderosa-style" haplotype, three *P. jeffreyi* individuals in this study were identified with that haplotype. One individual was randomly selected, and it lacks inferred nuclear admixture based on nSSRs. This suggests that a chloroplast lineage may have persisted without the retention of nuclear introgression in this individual (Figure 4.7). The two *P. jeffreyi* individuals with substantial nuclear admixture also have the "ponderosa-style" haplotype. These two individuals were field-identified as "putative hybrids", but exhibit greater molecular similarity to var. *ponderosa s.l.* than to *P.*

jeffreyi in both genomes. This indicates that the overlapping ranges of field-identifiable morphological characteristics can occasionally mask the identity of these species, and these two samples would clearly be assigned to var. *ponderosa* based on molecular evidence.

Inferred admixture from *P. jeffreyi* into var. *ponderosa s.l.* was also very limited. Surprisingly, molecular evidence indicates that admixture is slightly higher in allopatric populations (PB, PG, WA, and WB) than in sympatric populations (PL, PS, WL, and WR; Figure 4.7). A substantial amount of admixture from *P. jeffreyi* into *P. ponderosa* was inferred for several individuals in allopatric populations (PB and PG). For these individuals, retention of ancestral polymorphism or stochastic processes are more likely explanations than introgression. Only one individual (from population PL) of the 25 var. *ponderosa s.l.* that were assayed in this study did not contain the "ponderosa-style" chloroplast haplotype. Notably, this individual has a predominantly ponderosa-like nuclear assignment (Figure 4.7). Similar to the observation of "ponderosa-style" haplotypes in *P. jeffreyi*, this suggests that a *P. jeffreyi* chloroplast lineage may have persisted in this var. *ponderosa s.l.* individual without evidence of nuclear introgression. One *P. washoensis* individual from Babbitt Peak, CA was identified in Chapter 3 of this dissertation without the "ponderosa-style" haplotype.

These data, along with the nuclear sequences and chloroplast haplotypes analyzed in Chapter 3 of this dissertation for all *Ponderosae*, do not support recent introgression as the likely explanation for some *P. washoensis* individuals resolving closer to *P. jeffreyi* than to *P. ponderosa* in some studies (Prager, Fowler, and Wilson 1976; Patten and Brunsfeld 2002). However, evidence for a low level of introgression suggests that sporadic but continuing interspecific hybridization between *P. jeffreyi* and var. *ponderosa s.l.* provides an important source of migrants that increase the level of genetic diversity and extend the coalescence time for nuclear loci. It also appears that chloroplast lineages from the other species can persist in both taxa without concomitant nuclear introgression detectable with these six nSSR loci.

This study does not provide a clear answer for Hypothesis 2. I found extremely weak differentiation in nSSRs between populations of var. *ponderosa s.l.*, despite sampling eight populations separated by as much as 750 km, and with the few noted exceptions, they all share the "ponderosa-style" chloroplast haplotype. None of my results provided evidence to support the traditional *P. washoensis* populations as distinct from *var. ponderosa s.s.* (alternative H_{2B}). Some support for a distinct *P. washoensis* was suggested by F_{st} and PCA

when the putative *P. washoensis* populations from Oregon are included (alternative H_{2A}). In contrast, results from four different genetic distance methods, from PCO, and from admixture all lend support for *P. washoensis* being encompassed within the North Plateau race (alternative H_{2C}). I note the philosophical difficulty of demonstrating that a lack of structure is biological rather than due to weak signal in methodology. For example, the lack of correspondence for population relationships inferred with different genetic distance metrics could be interpreted as either the absence of a biological structuring, or weak signal in this data set. However, structuring (albeit weak) was inferred using methods based on allele frequencies (F_{st} and PCA), but absent in methods based on individuals (PCO, admixture). This could be interpreted as an indication that weak biological structuring is present that is below the power of individual-based methods to detect.

Additional populations representing the Pacific race and several populations from var. *scopulorum* may reveal more structure within var. *ponderosa s.l.* The sole population representing the Pacific race (PS), appeared quite different from the remainder of var. *ponderosa s.l.* in two tests (pairwise F_{st} and PCA), and it may be acting as an outlier in genetic distance calculations. I note that *P. jeffreyi* populations are much further outliers, and the similarity of analyses with and without these populations suggests that they do not distort the results. This study was not designed to address suggestions that *P. washoensis* may share ancestry with var. *scopulorum* (Haller 1965; Smith 1967). Now that I have confirmed the morphological observation that introgression with *P. jeffreyi* is minor, albeit not negligible, it would be reasonable to focus future work on sampling more populations of each of the ponderosa races and varieties. More importantly, deeper sampling of the genome may help determine whether the lack of resolution within var. *ponderosa s.l.* is really biological (i.e. supporting alternative H_{2C}), or due to the low information content and/or background noise in this data set.

Potential Error Sources. I found that PCR amplification of *P. taeda*-designed nSSRs in *Ponderosae* was quite successful, with eight of 17 loci amplifying single-banded fragments without a great deal of optimization. Of those eight loci, one apparently had undergone a duplication (*PtTX3118*), and one was monomorphic (*LOP11*). The remaining six loci were polymorphic, yielding length variation that created a total of 69 alleles in 198 samples (Table 4.5). Is this cross-species transfer of nSSRs appropriate, and should nSSR

loci be used to deduce inter- as well as intra-species relationships? First, the differentiation of all three *P. jeffreyi* populations from all eight of my var. *ponderosa s.l.* populations is robust to method of inference. I observed a clear, unique clustering of *P. jeffreyi* in every test, whether based on allele frequency or individual clustering. Thus, I can conclude that nSSRs appear to be useful for inter-species comparisons at this close level of relatedness in *Pinus*. The second question of resolution between races within var. *ponderosa s.l.* is obviously more challenging. Because of the clear distinction between *P. jeffreyi* and var. *ponderosa s.l.* provided by these nSSRs, it is unlikely that their evolution is too fast (i.e. that observed allele sizes are homoplasious rather than identical-by-descent) for intra-specific comparisons. Rather, the extremely high level of *Pinus* within-population compared to between-population (and between-race) variation is likely to make these inferences more sensitive to any noise that may distort the genealogy of each locus. Several potential causes of distortion are discussed below, along with the potential for this multi-locus assessment to overcome them.

First, every population measured in this study violates assumptions of random mating using my six-locus data set. The source of disequilibrium is not even across loci, but is driven by four of the six loci (see Results). This lack of HWE leaves us trapped in a circular argument; it is either caused by, or reduces the power to detect, null alleles and/or linkage disequilibrium. It is possible that sampling additional individuals would reduce the imbalance, although underlying biological causes could be present as well. Obviously, sampling more loci and discarding the unbalanced ones would improve my confidence that this lack of resolution is biological rather than an artifact of my measurements. The important question at hand is how much doubt this violation of HWE casts on inferences in support of my alternative hypotheses.

Second, most (but not all) of my analyses explicitly make the assumption that inferences are based on multiple independent loci. The pine genome is large, is divided into 12 chromosomes (Saylor 1961), and Zhou et al. (2003) found no clustering of nSSRs in the pine genome. Despite fairly extensive genetic linkage mapping in *Pinus*, including syntenic comparisons between species (Brown et al. 2001) and the prior application of several of the loci chosen for this project to other pine species, not all of these six nSSRs have been linkage-mapped. Further, linkage disequilibrium in conifers tends to decay rapidly, e.g. less than 1500 bps, or within the size of an average gene (reviewed in Neale and Savolainen

2004). Testing for linkage disequilibrium in my data set inferred linkage between two of the six loci. Unlike genetic linkage mapping, inferences of linkage disequilibrium are not robust enough to infer the linkage distance, and inferences may be limited where populations violate assumptions of HWE (Excoffier and Slatkin 1998). To assess whether this linkage might distort my measurements, I used two methods in STRUCTURE: a) repeating the simulation excluding one of the two linked loci; b) comparing simulation results from the "admixture" model (which assumes independence) with the "linkage" model where linkage between these two (but not the other four) loci is taken into account. I did not observe a difference in either test. The five-locus simulations were nearly identical to the six-locus results presented here, and each of the linkage model simulations (Figure 4.7) were similar to the corresponding model that assumes unlinked loci. Although I did not detect any distortion based on this violation of the assumption of independence, I cannot rule out the possibility that it contributes to the background 'noisiness' in this data set.

A third, and potentially major, source of noise in projects relying on fragment-length polymorphism is the presence of undetected null alleles. Many SSR studies mention the number of failed amplifications (presumably due to changes within the primer binding site), but few emphasize the underlying assumption that a single fragment size in a heterozygotic genome represents a homozygote (rather than one heterozygotic allele and one null allele). Chapuis and Estoup (2007) suggest that null SSR alleles are likely to be found in populations with large effective size, and in populations that have diverged from the source for primer design. Because this study shares both of these features, I gave careful consideration to the pattern of null alleles and their potential for distortion. My observed excess of homozygotes, combined with the method suggested by Chapuis and Estoup (2007), indicate that null alleles may be a substantial source of distortion in this data set. Inferred frequencies for null alleles exceed 20% at two loci (Table 4.6), and the inferred null allele frequency ranges from 6% (for the PB and WA populations) to over 13% for the PG population. These inferences for null alleles may be overstated due to departures from HWE. In addition to undetected null alleles, departures from HWE can be caused by non-random mating, selection, finite population size, and an excess of homozygotes due to the Wahlund effect. As these biological causes should affect all loci, the wide range of null frequencies that were observed among loci (Table 4.6) suggests that null alleles are present in some of my loci. Nonetheless, in a review of 233 published studies that reported frequencies of SSR null alleles, Dakin and

Avise found typical null frequencies to be < 0.2, but about 20% of studies reported frequencies of 0.2 to 0.4 (Dakin and Avise 2004). Thus, my two loci with the highest inferred frequencies of null alleles (ca. 22% and ca. 23%) appear to be close to a 'typical' range for SSRs.

I implemented several tests to determine the impact that null alleles may have on my conclusions. Table 4.7 presents pairwise F_{st}'s that allow a comparison between uncorrected and 'corrected-for-null-allele' estimates. A Mantel test comparing the two F_{st} matrices shows that they are significantly correlated (R^2=0.841, y=0.786 + 0.0062). I also present neighbor-joining diagrams that allow a comparison between uncorrected (D_c) and 'corrected-for-null-allele' (D_{cc}) chord distances (Figure 4.3). Except for the consistent separation of all three *P. jeffreyi* populations from all eight var. *ponderosa s.l.* populations, every distance metric (including the one correcting for null alleles) yields a different pattern. Nevertheless, each pairwise Mantel test yields a positive correlation (Table 4.9) when the distance matrices with all 11 populations are compared. The Mantel test between the D_c and the D_{cc} distance matrices yields R^2=0.8013, y=0.8803x + 0.2532. This may support the notion that the bias introduced by undetected nulls may be somewhat uniform across populations, presumably by chance. However, when the distance matrices are limited to the eight var. *ponderosa s.l.* populations, the only positive correlation is between the D_a and the D_c matrices. The eight-population D_c and D_{cc} matrices do not show a correlation, but neither do the comparisons between other distance metrics. Despite terming D_{cc} the 'corrected' distance, I actually have no way to corroborate its inferences, which are ultimately based on departure from HWE. Lacking a way to identify the 'correct' pattern, uncertainty remains about how much noise my inference of a substantial frequency of null alleles, as well as the two-fold difference in the frequency of inferred nulls between populations, introduces into these analyses. As a final test, I repeated the PCO analysis for data sets excluding each locus in turn, then excluding both loci with the highest frequency of inferred null alleles (PtTX3025 and PtTX3030). Resolution within var. *ponderosa s.l.* was not altered in these tests, again suggesting that the lack of signal is inherent to this system rather than an artifact of the analysis.

Fourth, the evolutionary history that underlies the length variation in SSRs is not fully understood (reviewed in Ellegren 2004). Some have suggested that the SMM model better accounts for the most likely source of differences in repeat number, namely slippage

of the polymerase during replication. This model also assumes that differences in repeat number account for all observed differences in fragment length. However, the SMM model is known to be confounded by different repeat motifs expanding in sister lineages. This lends supports to the notion that these loci reside in fast-evolving regions of the genome, suggesting that substitutions and indels of all kinds may be frequent in SSRs. Further, duplication of microsatellite loci has been reported in conifers (Karhu et al. 2000) and references therein). In locus *PtTX3018*, I observed an apparent case of gene duplication, with 14 out of 198 samples amplifying multiple loci. Size homoplasy in my observed fragment lengths could thus result from unrelated events as well as from a convergence to the same number of repeats that can arise in different lineages due to constraints on number of repeats (Karhu et al. 2000). One way to address this known complicating factor is to use a large number of loci. The number of loci required will likely depend on the study system. In human forensics, for example, 30 or more loci are commonly used in order to achieve very high levels of statistical inference (Weir 1996). A cursory review of the recent application of nSSRs to various questions regarding conifer population structure show a wide range in the number of loci used, although some combined nSSRs with other types of data. Some research was based on as few as three (Mariette et al. 2001) to five (Boys, Cherry, and Dayanandan 2005) loci, while others have applied 15 (Al-Rabab'ah and Williams 2002) to 18 (Williams, Elsik, and Barnes 2000) polymorphic loci. Because of weak linkage between two of the six loci, this study may have only provided five independent estimates. Further, additional loci would provide enough data to support meaningful tests that exclude one or more loci in order to test the consistency of their inferences.

Kalinowski (2002) found that for most genetic distance measures, the number of total alleles effected the precision of estimates as much as the number of loci. In the present study, my number of alleles with no missing data (59), exceeded the highest number simulated by Kalinowski (32). After subtracting the eight alleles that were private to *P. jeffreyi*, the data set still contains 51 alleles in var. *ponderosa s.l.* Whether an even larger sample of alleles from var. *ponderosa s.l.* will overcome complications from introgression and probable reticulation remains to be seen. It seems likely that increasing the number of sampled loci and alleles as well as increasing the number of sampled populations will be needed in future work to disentangle these relationships or provide firm evidence that they are indeed not differentiated.

Fifth (and last), my nSSR results (this chapter) and nuclear sequence data (Chapter 3) indicate that the retention of ancestral polymorphism may confound even fast-evolving nSSRs in comparisons between populations and races. The same factors (outcrossing, high heterozygosity, and large effective population sizes) are in effect. If the evolutionary rate of the nSSRs is slow relative to the age of the population divergence and N_e, then one would expect incomplete lineage sorting at this level as well. Ideally, I would like to measure the few minor changes that are responsible for the major effects on morphological and physiological differences that have been observed between ponderosa races. Because I have no way to do this at present outside of model species, the only way to address this problem in var. *ponderosa s.l.* is by sampling more populations and more alleles to gradually increase the statistical power.

Conclusions. These results clearly support the separation between *P. jeffreyi* and var. *ponderosa s.l.*, and confirm that ongoing introgression is occurring at a low frequency, as reported by Haller (1965). In fact, the observed admixture patterns (a few mixed individuals and the others all pure) suggests that first-generation hybrids may not survive to reproductive maturity, since later-generation introgressants are uncommon. I found unanimous support from a number of different methods for the distinctiveness of *P. jeffreyi* from var. *ponderosa s.l.*: private alleles, PCO, PCA, genetic distances based on four different metrics, pairwise F_{st}, and admixture simulations, as well as chloroplast haplotypes. From the 131 *P. jeffreyi* and var. *ponderosa s.l.* individuals sampled within sympatric ranges, evidence for recent introgression was found in only four individuals: inferred nuclear admixture from a ponderosa-like parent into two putative hybrid individuals in the LJ population, a "ponderosa-style" chloroplast haplotype lineage persisting in a different *P. jeffreyi* individual, and a *P. ponderosa* individual that lacked the "ponderosa-style" chloroplast haplotype (Figure 4.7). While I did not specifically target putative hybrids in my field sampling, I noted only a few morphologically intermediate individuals, all of them at the LJ site. Further, my admixture simulations from areas of sympatry (PL, PS, WL, and WR) show a similar level of inferred admixture as the allopatric populations (PB, WA, WB, and PG). In Chapter 3 of this dissertation, I identified one *P. washoensis* individual from Babbitt Peak, California that lacked the "ponderosa-style" chloroplast haplotype. Thus, neither ancient nor current introgression from *P. jeffreyi* can fully explain the phylogenetic placement of some

var. *ponderosa s.l.* alleles for some loci with a *P. jeffreyi/P. sabiniana* clade (Chapter 3 of this dissertation; Patten and Brunsfeld 2002; Eckert and Hall 2006), although occasional migrants may substantially increase diversity and delay nuclear coalescence. This leaves retention of shared ancestral polymorphism as the likely mechanism for phylogenetic affinity between *P. jeffreyi* and var. *ponderosa s.l.*

In support of alternative hypothesis H_{2A}, I observed a weak clustering by grouping the putative Washoe populations from Oregon (WA and WB) with the Mt. Rose and Warner-Mountain *P. washoensis s.s.* (WL and WR) using PCA (Figure 4.2, B). Similarly, alternative H_{2A} receives weak support from F_{st} comparisons with these same groupings (Table 4.8). Thus, I have a limited amount of support for Haller's unpublished morphological evidence for shared ancestry between the high-altitude populations of var. *ponderosa s.l.*, separate from either the North Plateau or the Pacific race. This finding could be explained by those populations continuing as relicts of a formerly widespread taxon. Alternatively, these populations might retain alleles gained by ancient reticulate ancestry, perhaps between incompletely-diverged varieties (e.g. var. *ponderosa* and var. *scopulorum*). None of the genetic metrics employed yielded a cluster supporting any of my alternative hypotheses (Figure 4.3). In fact, because the only resolution is between *P. jeffreyi* and var. *ponderosa s.l.*, genetic distances serve only to highlight the low level of population divergence within var. *ponderosa s.l.* This would not be in conflict with a reticulate ancestry, as it is difficult to explain this long-distance lack of population differentiation except for the general lack of population structure in woody plants (Hamrick and Godt 1996). It is tempting to speculate that this is what populations that harbor genetic diversity for long periods of time would look like if they were the result of secondary contact that allowed a remixing of diverged genetic pools. Perhaps over time, the very narrow zone in western Montana where var. *scopulorum* and var. *ponderosa* have recently met that Johansen and Latta call a "suture zone" (Johansen and Latta 2003) will eventually result in a mixture of two diverged genetic milieus with little purging of alleles.

It is possible that the cryptic structure that was observed is an artifact of the high level of intra-population variability and/or one or more of the potentially distorting noise factors previously discussed. In support of this, simulations that inferred admixture in two allopatric individuals (one from the PB and one from the PG population), are likely due to one of the background noise and/or stochastic factors. However, a simple retention of

polymorphism would be unlikely to exhibit this pattern over six loci. Because introgression with *P. jeffreyi* has been excluded (and var. *ponderosa* s.l. populations well outside the *P. jeffreyi* range exhibit this pattern as well), the putative ancestors would likely have been races and/or varieties of *P. ponderosa.* A logical expansion of this work would be to include more populations of the Pacific race and several populations of var. *scopulorum* in this data set. However, the substantial portion of inferred null alleles is troubling, and is an issue that should be considered in cross-species transfers of nSSRs. The extent to which null alleles in *Pinus* nSSRs can be overcome by scoring more loci remains to be tested. Because my results indicate that relationships as distant as those between var. *ponderosa s.l.* and *P. jeffreyi* are evident using nSSRs, this technique may also be useful to unravel some species delimitations among *Ponderosae* native to Mexico and Central America.

Table 4.1. Two hypotheses regarding the assignment of populations to species. Population codes are described in Table 4.2. For the purposes of the AMOVAs presented in Table 4.8, population PS (Shasta, California, representing the Pacific race) was omitted.

Description		Population groups
H_1	Hypothesis 1: *P. jeffreyi* is distinct from var. *ponderosa s.l.* (i.e. var. *ponderosa* + *P. washoensis*)	J: JL+JT+JR vs. Psl: PB+PG+PL+PS+WA+WB+WL+WR
Hypothesis 2: three alternatives for the relationship between var. *ponderosa* and *P. washoensis*		
H_{2A}	*P. washoensis* is a distinct, relictual, species, encompassing high-elevation populations in Oregon as well as traditional locations (i.e. comparing var. *ponderosa s.s.* with *P. washoensis s.l.*)	Pss: PB+PG+PL+PS vs. Wsl: WA+WB+WL+WR
H_{2B}	*P. washoensis* is a distinct, relictual, species, but confined to traditional locations	P: PB+PG+ PL+PS+WA+WB vs. W: WL+WR
H_{2C}	*P. washoensis* is encompassed within the diversity of the North Plateau race of var. *ponderosa*	supported by a lack of significant groupings among: PB+PG+PL+WA+WB+WL+WR

Table 4.2. Geographic locations for population-level sampling.

Map Code	Population	State	County	Latitude	Longitude	Elevation	Species collected	Population Code
A	Abert Rim	Oregon	Lake	42.39°N	120.23°W	2290 m	*P. washoensis**	WA
B	Blue Mountains	Oregon	Grant	44.07°N	118.79°W	1550 m 1920 m	*P. ponderosa* *P. washoensis**	PB WB
G	Grangeville	Idaho	Idaho	45.75°N	116.37°W	910 m	*P. ponderosa*	PG
L	Likely	California	Modoc	41.23°N	120.41°W	1880 m 1400 m 2090 m	*P. jeffreyi*; *P. ponderosa* *P. washoensis*	JL PL WL
R	Mt. Rose	Nevada	Washoe	39.33°N	119.87°W	2260 m 2260 m	*P. jeffreyi*; *P. washoensis*	JR WR
S	Shasta	California	Shasta	40.67°N	122.70°W	570 m	*P. ponderosa***	PS
T	Thomas Creek	Nevada	Washoe	39.39°N	119.84°W	1820 m	*P. jeffreyi*	JT

* putative *P. washoensis*; **Pacific race – sympatric with *P. jeffreyi*

Table 4.3. Seventeen *P. taeda* nSSR loci that were tested for amplification in three target species: *P. jeffreyi*, *P. ponderosa*, and *P. washoensis*.

Locus	Source
LOP1	Liewlaksaneeyanawin et al. 2004
LOP3	Liewlaksaneeyanawin et al. 2004
LOP5	Liewlaksaneeyanawin et al. 2004
LOP8	Liewlaksaneeyanawin et al. 2004
LOP9	Liewlaksaneeyanawin et al. 2004
LOP11	Liewlaksaneeyanawin et al. 2004
LOP12	Liewlaksaneeyanawin et al. 2004
PtTX2123	Elsik et al. 2000
PtTX2128	Elsik et al. 2000
PtTX2146	Elsik et al. 2000
PtTX3011	Elsik et al. 2000
PtTX3020	Elsik et al. 2000
PtTX3025	Elsik et al. 2000
PtTX3030	Elsik et al. 2000
PtTX3034	Elsik et al. 2000
PtTX3098	Kutil and Williams 2001
PtTX3118	Kutil and Williams 2001

Table 4.4. Eight nSSR loci used to sample population variation in 198 individuals of three target species: *P. jeffreyi*, *P. ponderosa*, and *P. washoensis*. Repeat motif, flanking bps, and number of repeat motifs were inferred from a comparison of sequences from three to five homozygotes with their fragment lengths. H_e, heterozygosity, was averaged over 11 populations; S.E., standard error; n/a, not available.

Locus	5' primer fluorophore	PCR annealing temperature	(Repeat motif) number of repeats	Size range (bps)	Bps flanking repeats	No. of alleles	H_e (S.E.)
LOP5	FAM	53	(TA)5 to 28	160 to 206	150	22	0.815 (0.022)
LOP11	NED	60	n/a	244	n/a	1	n/a
PtTX2123	HEX	60	(AGC)5 to 11	185 to 203	170	6	0.525 (0.021)
PtTX2128	NED	53	(GAC)5 to 8*	237 to 246	222	4	0.375 (0.059)
PtTX3025	FAM	53	(A)6 to 14	264 to 272	258	7	0.451 (0.056)
PtTX3030	FAM	60	(TA)5 to 25*	312 to 352	302	15	0.727 (0.041)
PtTX3098	FAM	60	(GTT)4 to 8	173 to 186	161	5	0.551 (0.026)
PtTX3118	HEX	60	(CAT)n/a	205 to 229	n/a	9	n/a
total						69	

*Repeat motifs for *PtTX2128* and *PtTX3030* were inferred from published sequences due to read failure in the repeat regions of homozygote sequences

Table 4.5. Diversity within 11 populations for six nSSR loci (*LOP5, PtTX2123, PtTX2128, PtTX3025, PtTX3030*, and *PtTX3098*). Comparison between: all 198 samples (1.59% missing data) and 186 samples (restricted to samples with no missing data for these six loci). Population codes are described in Table 4.2; N, number of individuals; N_a, number of alleles; MNA, mean Na per locus; H_e, expected heterozygosity (Nei 1987).

		198 samples				186 samples			
Species	Population Code	N	N_a	MNA	H_e	N	N_a	MNA	H_e
P. jeffreyi	JL	20	36	6.0	0.493	19	36	6.0	0.503
	JR	17	29	4.8	0.464	16	28	4.7	0.456
	JT	22	31	5.2	0.471	20	31	5.2	0.475
P. ponderosa	PB	20	32	5.3	0.572	20	32	5.3	0.572
	PG	21	31	5.2	0.612	17	29	4.8	0.601
	PL	17	31	5.2	0.624	17	31	5.2	0.624
	PS	14	28	4.7	0.573	10	31	5.2	0.563
P. washoensis	WA*	7	25	4.2	0.633	7	25	4.2	0.633
	WB*	19	30	5.0	0.609	19	30	5.0	0.609
	WL	22	30	5.0	0.578	22	30	5.0	0.578
	WR	19	37	6.2	0.684	19	37	6.2	0.684
	mean per population (S.E.)	18.0 (1.3)	30.9 (3.4)	5.2 (0.2)	0.551 (0.03)	16.9 (1.4)	30.5 (3.7)	5.1 (0.2)	0.573 (0.02)
	total	198	69			186	59		

* putative *P. washoensis*

Table 4.6. Inferred frequency of null alleles by locus and by population.
S.E., standard error

Locus	Frequency (S.E.)
LOP5	0.008 (0.005)
PtTX2123	0.009 (0.007)
PtTX2128	0.104 (0.028)
PtTX3025	0.224 (0.017)
PtTX3030	0.232 (0.018)
PtTX3098	0.000 (0.000)
Population	Frequency (S.E.)
JL	0.117 (0.053)
JR	0.088 (0.029)
JT	0.096 (0.062)
PB	0.057 (0.035)
PG	0.135 (0.061)
PL	0.074 (0.047)
PS	0.099 (0.043)
WA	0.057 (0.038)
WB	0.107 (0.051)
WL	0.113 (0.052)
WR	0.115 (0.055)
overall mean (S.E.)	0.095 (0.014)

Table 4.7. F_{st} estimates across six loci between each pair of populations. Above the diagonal: AMOVA-based estimates of F_{st} (footnoted with probability based on an AMOVA with 9,999 permutations). Below the diagonal: F_{st} estimates corrected for inferred null alleles (see Methods).

	JL	JR	JT	PB	PG	PL	PS	WA	WB	WL	WR
JL	---	0.000	0.000	0.235**	0.200**	0.193**	0.207**	0.211**	0.252**	0.248**	0.158**
JR	0.004	---	0.004	0.288**	0.242**	0.248**	0.273**	0.265**	0.303**	0.304**	0.207**
JT	0.009	0.012	---	0.246**	0.238**	0.225**	0.238**	0.231**	0.273**	0.260**	0.195**
PB	0.146	0.183	0.206	---	0.076**	0.022	0.076	0.016	0.019	0.000	0.047**
PG	0.162	0.209	0.216	0.014	---	0.021	0.078	0.039	0.072**	0.083**	0.026
PL	0.156	0.217	0.198	0.067	0.022	---	0.029	0.007	0.012	0.029	0.007
PS	0.190	0.235	0.225	0.058	0.021	0.049	---	0.071**	0.082	0.077	0.058
WA	0.183	0.221	0.224	0.024	0.006	0.052	0.011	---	0.000	0.013	0.010
WB	0.203	0.251	0.242	0.055	0.030	0.057	-0.003	0.009	---	0.008	0.024
WL	0.127	0.174	0.184	0.014	0.009	0.051	0.040	0.014	0.033	---	0.043
WR	0.207	0.248	0.253	0.047	0.009	0.058	0.017	-0.004	0.006	0.017	---

** $P<0.01$ (after Bonferroni correction for multiple substitutions; $\alpha=0.000182$) that there is no genetic differentiation between the population pair

Table 4.8. F_{st} estimates between pairs of population groups to compare the alternative hypotheses for population groupings described in Table 4.1. F_{st} values are below the diagonal; probabilities based on AMOVAs with 9,999 permutations are shown above the diagonal.

H_1

J	Psl	
----	0.000	J
0.219**	----	Psl

H_{2A}

Pss	Wsl	
----	0.007	Pss
0.010**	----	Wsl

H_{2B}

P	W	
----	.165	P
0.003^{NS}	----	W

**P<0.01 that there is no genetic differentiation between the population pair; NS, not significant.

Table 4.9. Coefficient of determination (R^2) from pairwise Mantel tests comparing the genetic distances matrices resulting from four different metrics to infer population distances. Below diagonal: comparing all 11 populations; above diagonal: comparing the eight populations for var. *ponderosa s.l.*

	D_a	$\delta\mu^2$	D_c	D_{cc}
D_a	---	0.2070	0.8325	0.0228
$\delta\mu^2$	0.8414	---	0.1213	0.0000
D_c	0.9884	0.8019	---	0.0213
D_{cc}	0.7885	0.6209	0.8013	---

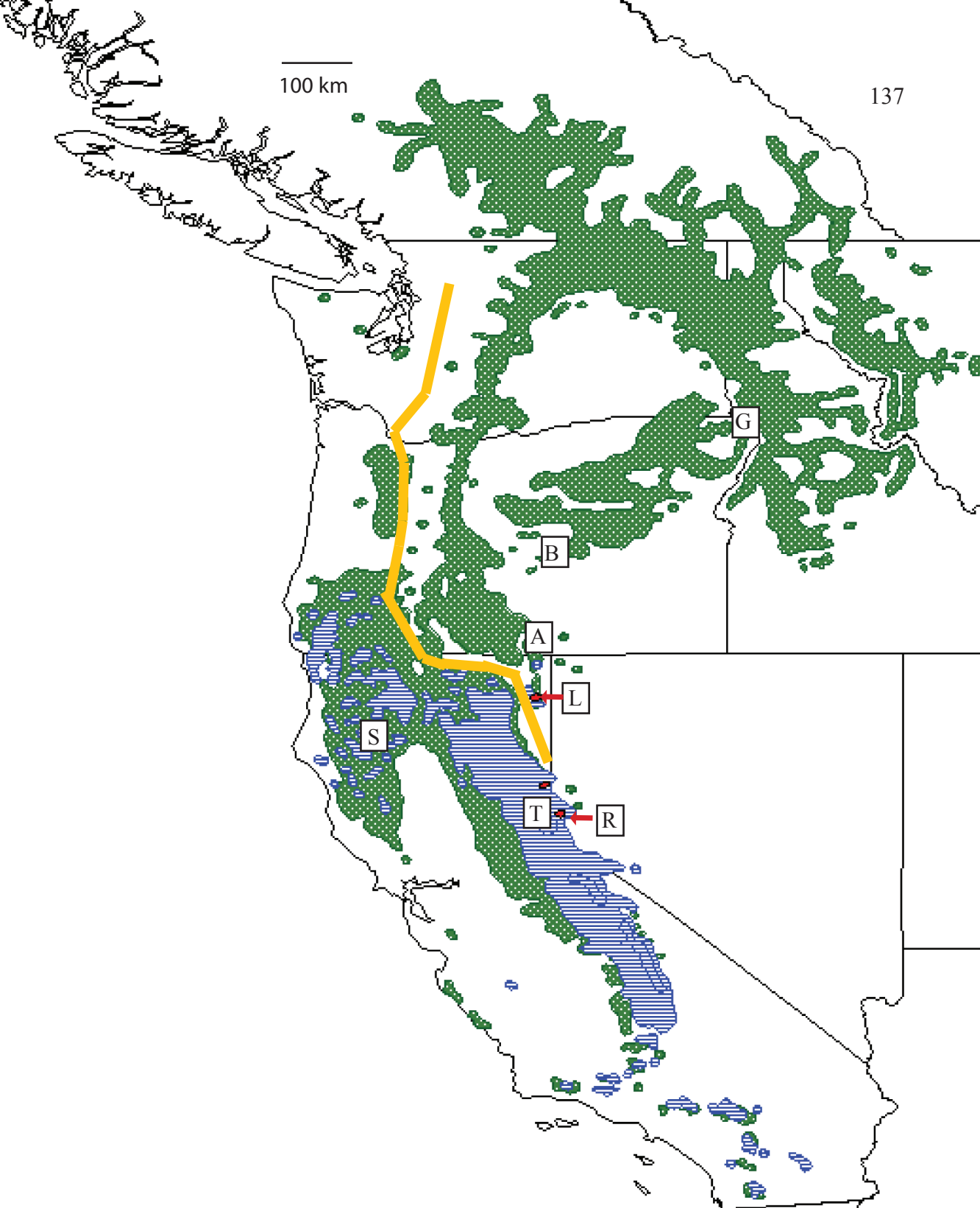

Figure 4.1. Geographic locations sampled (Table 4.2). Distribution ranges: *Pinus ponderosa* var. *ponderosa* (stippling; thick line shows approximate boundary between Pacific and North Plateau races.); *P. jeffreyi* (horizontal stripes); and traditional *P. washoensis* (red areas at locations L and R, and at Babbitt Peak, CA).

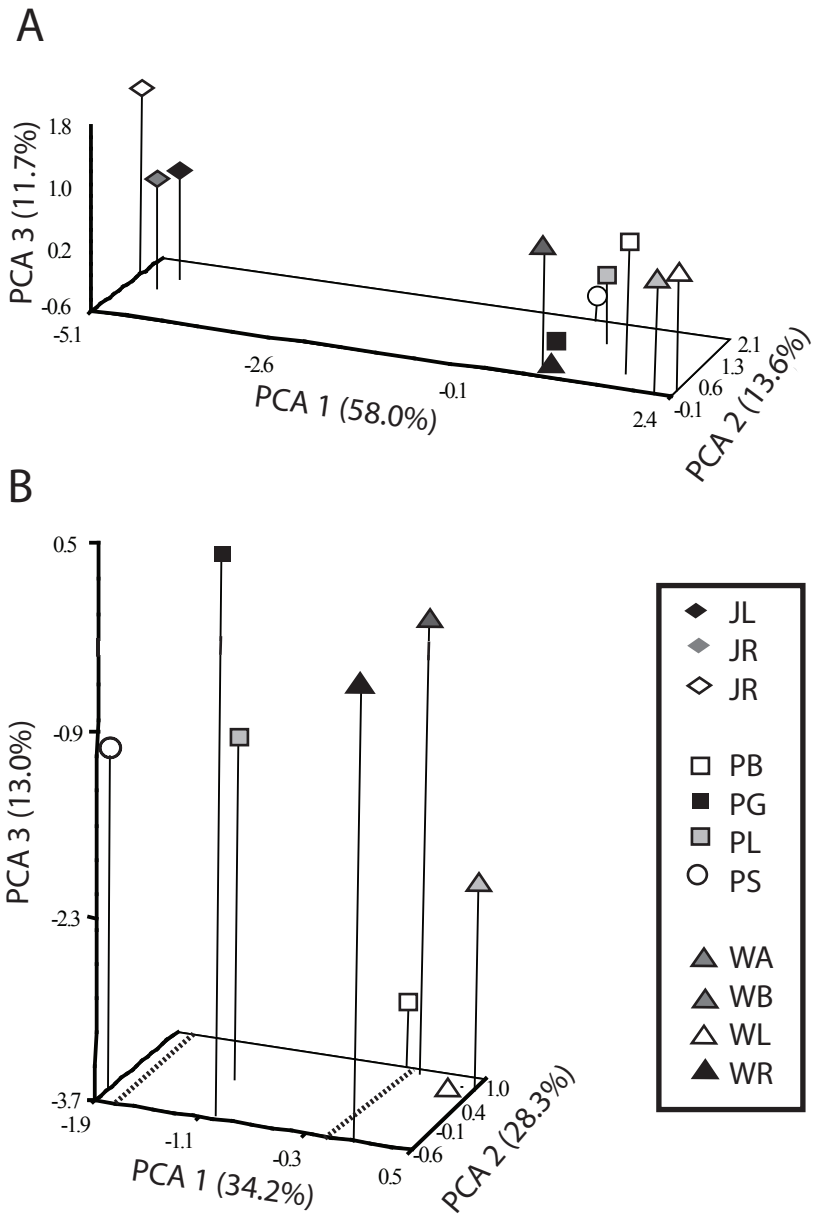

Figure 4.2. Principal component analyses based on a covariance matrix of the allele frequencies of six nSSR loci in A) 11 populations or B) eight populations (excluding *P. jeffreyi*). The first three axes account for a cumulative 83.8% in the 11-population comparison (A) and 75.5% in the eight-population comparison (B). Diamonds , *P. jeffreyi*; squares, var. *ponderosa* North Plateau race; circle, var. *ponderosa* Pacific race; triangles, traditional or putative *P. washoensis*.

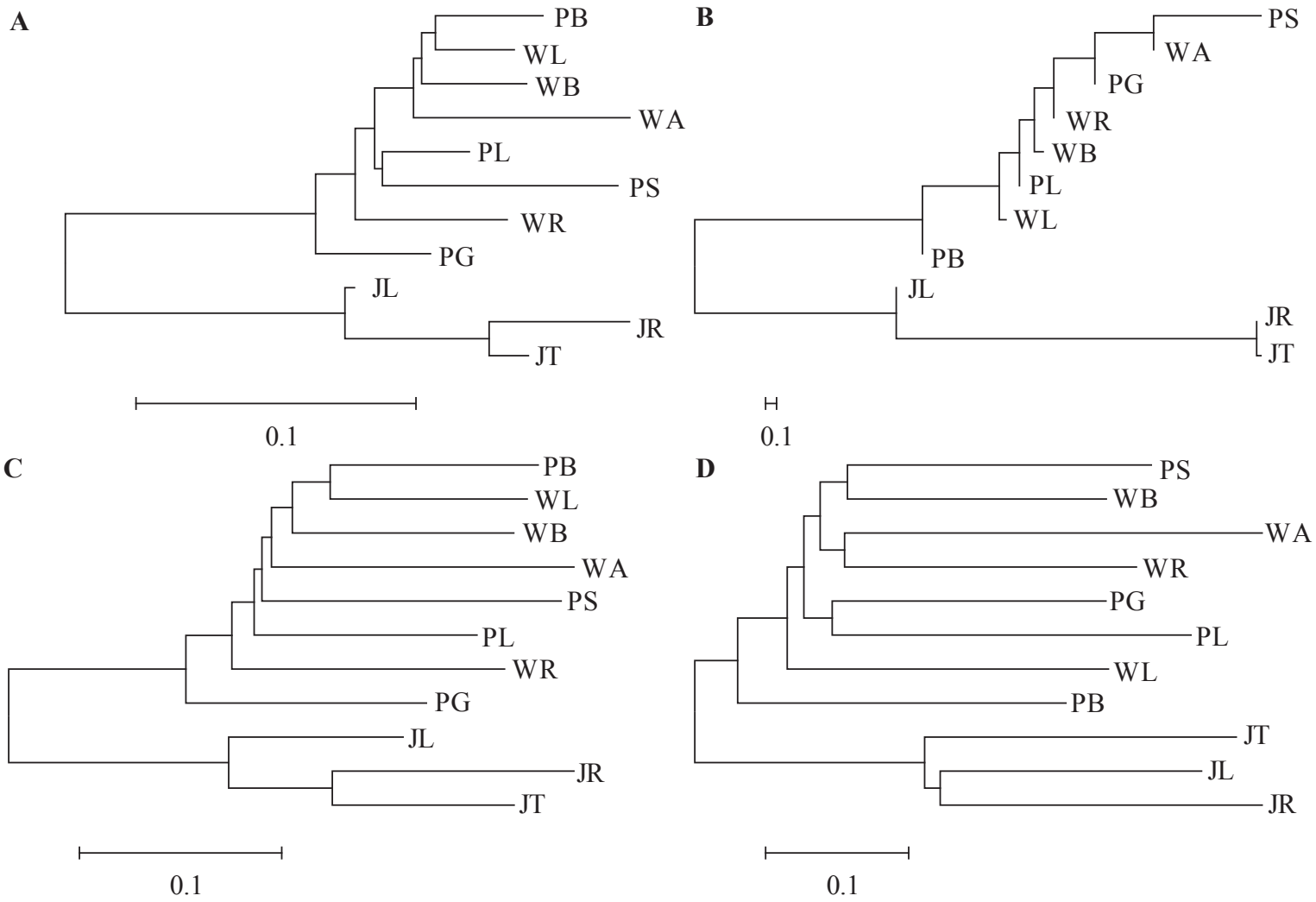

Figure 4.3. Visualization of four different metrics for population pairwise genetic distances in six nSSR loci with the neighbor-joining algorithm. A) Da (Nei); B) δμ2 (Goldstein); C) Dc (Chord); D) Dcc, chord distance corrected for inferred frequency of null allleles.

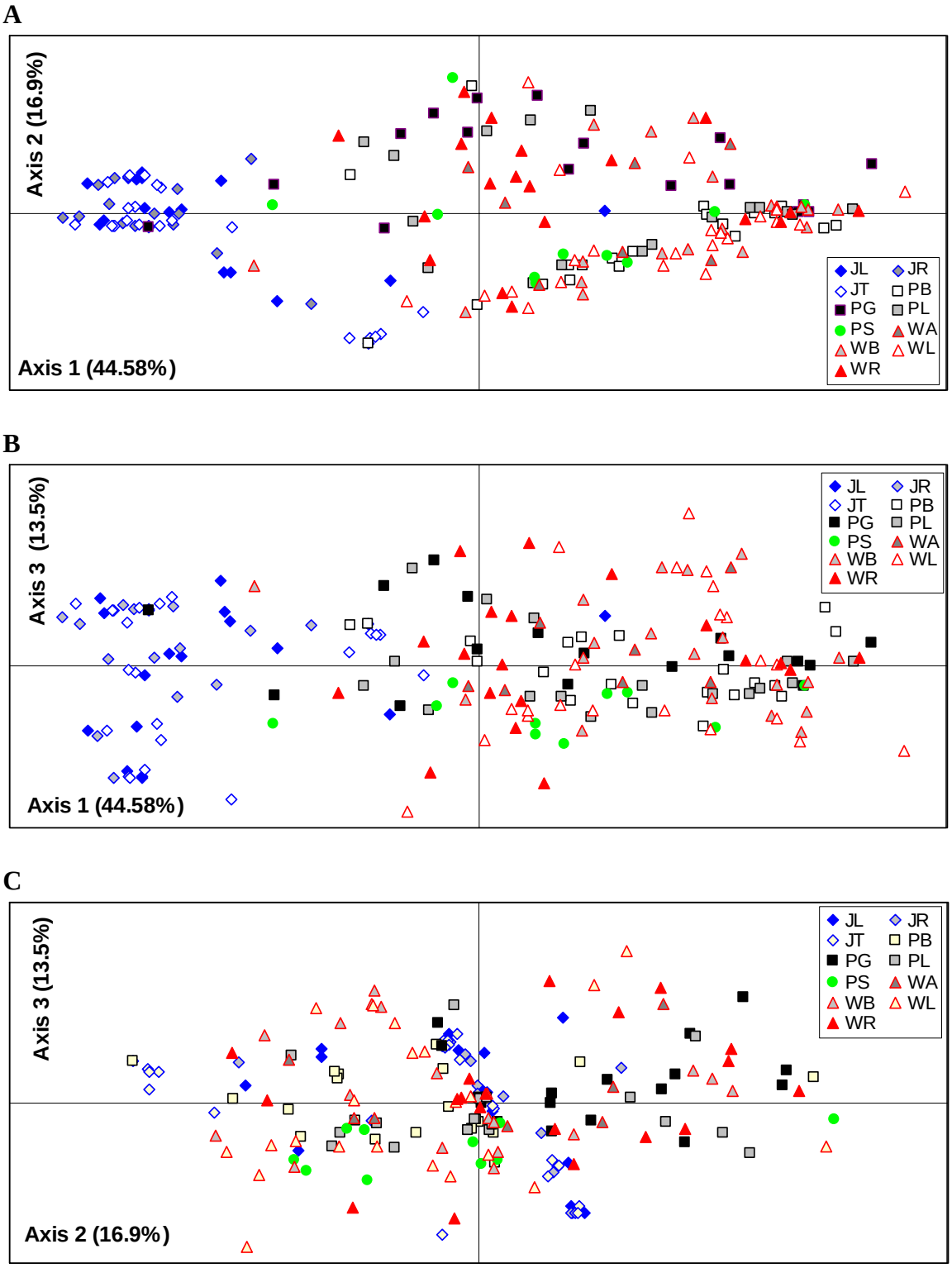

Figure 4.4. Principal coordinates analysis using a distance matrix to reveal similarities between 186 individuals sampled form11 populations for six nSSR loci. The first three axes account for a cumulative 74.91% of the difference. A) axis 1 vs. axis 2; B) axis 2 vs. axis 3; C) axis 2 vs. axis 3. Diamonds , *P. jeffreyi*; squares, var. *ponderosa* North Plateau race; circle, var. *ponderosa* Pacific race; triangles, traditional or putative *P. washoensis*.

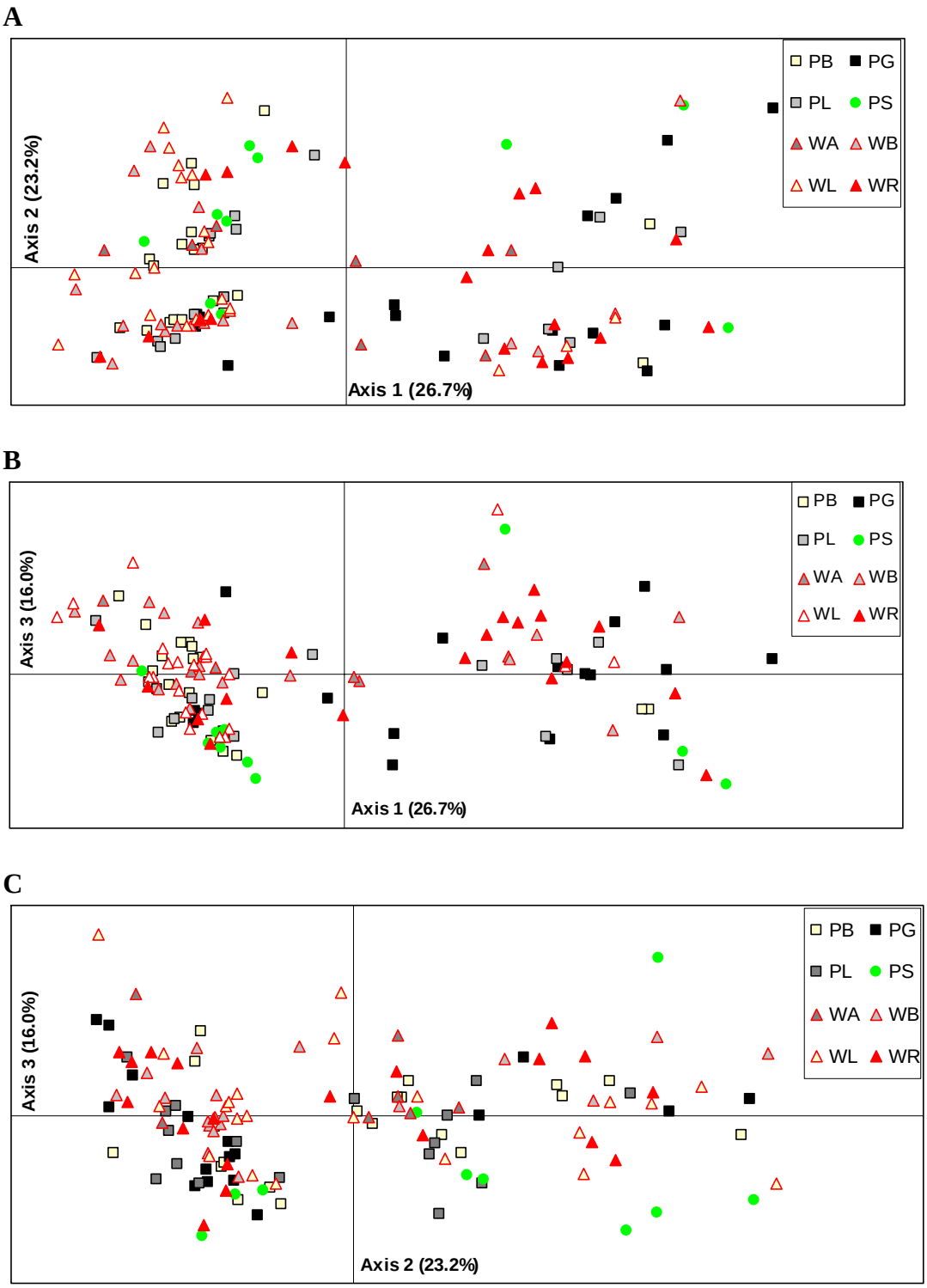

Figure 4.5. Principal coordinates analysis using a distance matrix to reveal similarities between 131 individuals sampled from eight populations of *P. ponderosa* and *P. washoensis* for six nSSR loci. The first three axes account for a cumulative 65.9% of the difference. A) axis 1 vs. axis 2; B) axis 2 vs. axis 3; C) axis 2 vs. axis 3. Diamonds , *P. jeffreyi*; squares, var. *ponderosa* North Plateau race; circle, var. *ponderosa* Pacific race; triangles, traditional or putative *P. washoensis*.

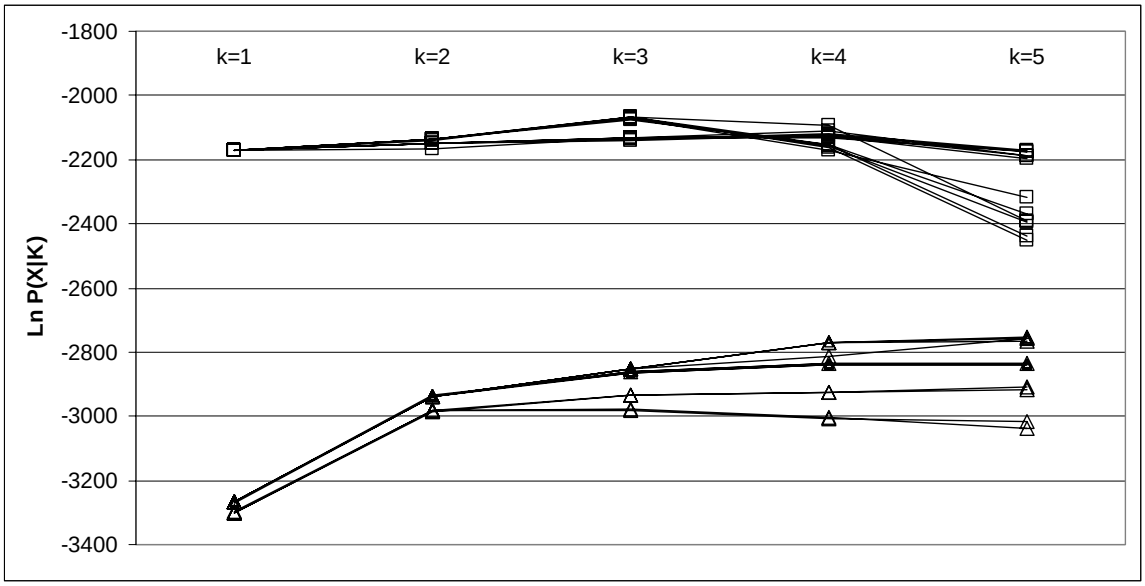

Figure 4.6. Estimated lnP(X|K) using STRUCTURE sofware, where K is the inferred number of populations. For each of 24 combinations of data sets and model parameters, K from 1 to 5 was estimated from one million iterations after discarding a 50,000-iteration burn-in. Triangles: data set includes *P. jeffrey*i, squares: data set was restricted to eight populations (excluding *P. jeffreyi*).

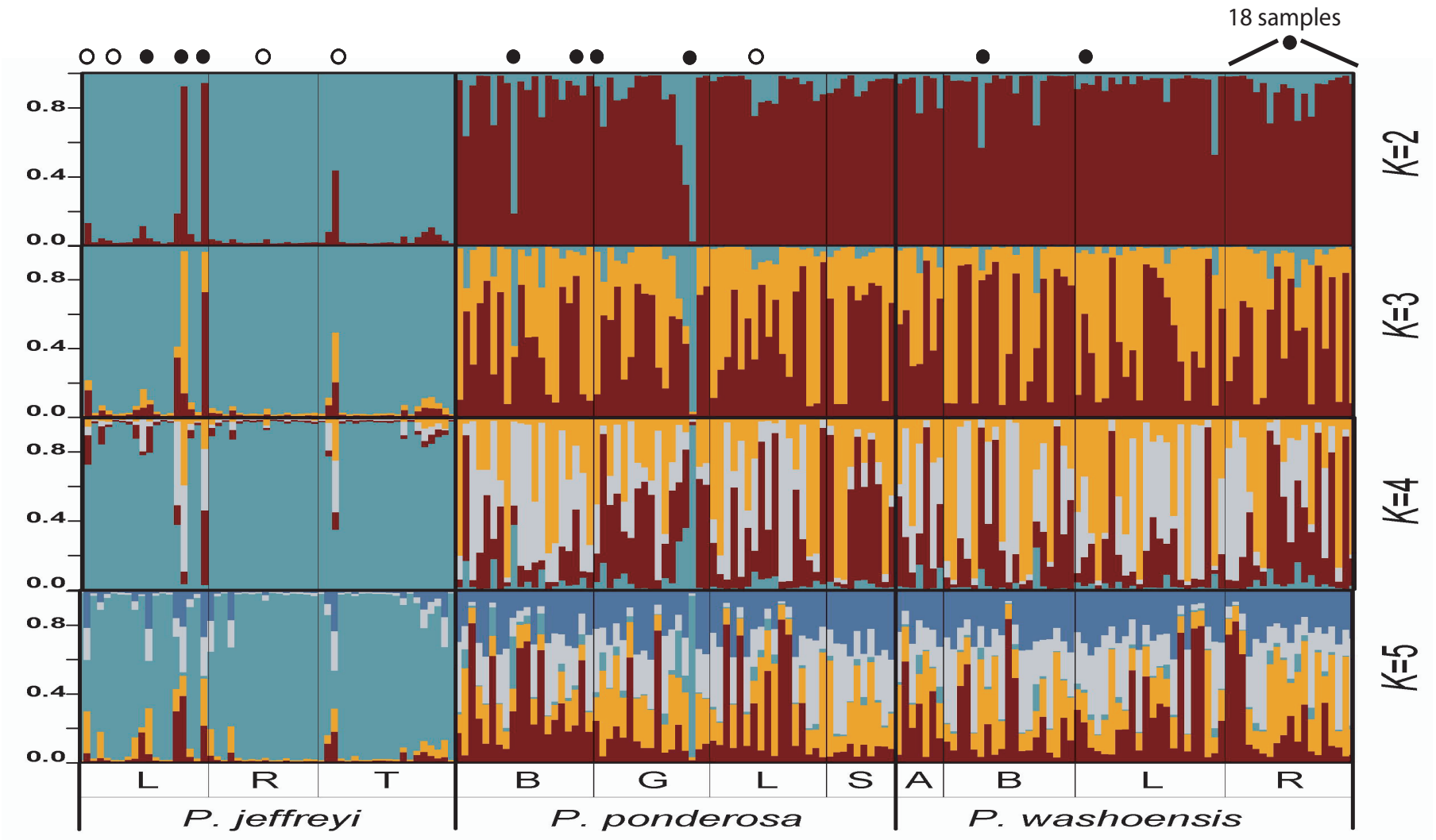

Figure 4.7. Inferred admixture for 186 individuals from 11 populations showing the mean permuted value of 10 independent simulations in STRUCTURE. The four graphs represent values of K (the inferred number of groups) from K=2 to K=5. Letters above species names correspond to the map codes described in Figure 1. Circles above bar chart are chloroplast haplotypes: filled circle, "ponderosa-style", open circle, non-ponderosa style.

ACKNOWLEDGMENTS

I am especially grateful to John R. Haller for sharing unpublished information on morphometrics and collection localities, for helpful discussions, and for his enthusiasm for these species. I am indebted to David Gernandt for sharing unpublished information as well as helpful comments. Assistance with the selection of collection sites was provided by Amanda Brinnand, Elizabeth Bergstrom, Shelly Schaff, Cheryl Beyer, Sydney Smith, and Rick Vetter (USDA Forest Service). Doyle E. Willyard and Paul Macias cheerfully helped with field collections. Thanks are extended to Mariah Parker-Defeniks for assistance in the laboratory, and to Matthew Horning, Alex Krupkin, Brian Knaus, Paul Severns, Nancy Mandel, and Leslie Gottlieb for helpful comments. Funding for this study was provided by National Science Foundation grants DEB 0317108 and ATOL 0629508 to Aaron Liston and Richard Cronn, by the USDA Forest Service Pacific Northwest Research Station, and by the Leslie and Vera Gottlieb Research Fund in Plant Evolutionary Biology.

LITERATURE CITED

Al-Rabab'ah, M. A., and C. G. Williams. 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. Forest Ecology and Management **163**:263-271.

Axelrod, D. 1986. Cenozoic history of some western American pines. Ann Missouri Bot Garden **73**:565.

Boys, J., M. Cherry, and S. Dayanandan. 2005. Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). Am J Bot **92**:833-841.

Brayshaw, T. C. 1997. Washoe and ponderosa pines on Promontory Hill near Merritt, B.C., Canada. Ann Naturhist Mus Wien **99B**:673-680.

Brown, G., E. Kadel, D. Bassoni, K. Kiehne, B. Temesgen, J. van Buijtenen, M. Sewell, K. Marshall, and D. Neale. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics **159**:799-809.

Brownstein, M. J., J. D. Carpten, and J. R. Smith. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques **20**:1004-1010.

Callaham, R. Z., and A. R. Liddicoet. 1961. Altitudinal variation at 20 years in ponderosa and Jeffrey pines. J Forest **59**:814-820.

Callaham, R. Z., and W. Metcalf. 1959. Altitudinal races of *Pinus ponderosa* confirmed. J Forest **57**:50-502.

Cavalli-Sforza, L. L., and W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution **21**:550-570.

Chapuis, M.-P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol **24**:621-631.

Chen, D.-H., and P. C. Ronald. 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR applications. Plant Molecular Biology Reporter **17**:53-57.

Conkle, M. T., and W. B. Critchfield. 1988. Genetic variation and hybridization of ponderosa pine. In "Ponderosa pine: the species and its management", D.M. Baumgartner & James E. Lotan, eds.

Critchfield, W. 1986. Hybridization and classification of the white pines (*Pinus* section *Strobus*). Taxon **35**:647-656.

Critchfield, W., and E. Little. 1966. Geographic Distribution of the Pines of the World. USDA Forest Service Misc. Publication 991:91 pp.

Critchfield, W. B. 1966. Crossability and relationships of the California big-cone pines. USDA Forest Service Research Paper **NC-6**:36-44.

Critchfield, W. B. 1984. Crossability and relationships of Washoe pine. Madroño **31**:144-170.

Critchfield, W. B., and G. L. Allenbaugh. 1965. Washoe pine on the Bald Mountain Range, California. Madroño **18**:63-64.

Dakin, E. E., and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity **93**:504-509.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B **39**:1-38.

Dieringer, D., and C. Schlotterer. 2003. Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes **3**:167-169.

Eckert, A. J., and B. D. Hall. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenetics and Evol **40**:166-182.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics **5**:435-445.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**:2611-2620.

Excoffier, L., and M. Slatkin. 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. Am J Hum Genet **62**:171-180.

Excoffier, L., P. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA and restriction data. Genetics **131**:479-491.

Falush, D., M. Stephens, and J. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.

Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA **92**:6723-6727.

Haller, J. R. 1961. Some recent observations on ponderosa, Jeffrey, and Washoe pines in Northeastern California. Madroño **16**:126-132.

Haller, J. R. 1965. *Pinus washoensis* in Oregon: taxonomic and evolutionary implications. Am J Bot **52**:646 (abstract).

Haller, J. R. 1962. Variation and hybridization in ponderosa and Jeffrey pines. Univ Calif Publ Bot **34**:123-165.

Hamrick, J. L., and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. Phil Trans R Soc Lond **351**:1291-1298.

Hong, Y.-P., A. Krupkin, and S. Strauss. 1993. Chloroplast DNA transgresses species boundaries and evolves at variable rates in the California closed-cone pines (*Pinus radiata, P. muricata*, and *P. attenuata*). Mol Phylogenetics and Evol **2**:322-329.

Jakobsson, M., and N. A. Rosenbeg. 2006. CLUMPP: CLUster Matching and Permutation Program.

Johansen, A. D., and R. G. Latta. 2003. Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. Mol Ecol **12**:293-298.

Johnson, M., S. B. Vander Wall, and M. Borchert. 2003. A comparative analysis of seed and cone characteristics and seed-dispersal strategies of three pines in the subsection Sabinianae. Plant Ecology **168**:69-84.

Kalinowski, S. 2002. How many alleles per locus should be used to estimate genetic distances? Heredity **88**:62-65.

Karhu, A., J.-H. Dieterich, O. Savolainen. 2000. Rapid expansion of microsatellite sequences in pines. Mol Biol Evol **17**:259-265.

Kolesnikova, R. D., V. G. Latysh, and A. I. Chernodubov. 1977. A chromatographic investigation of the n-heptane in the essential oil of representatives of the family Pinaceae. Chemistry of natural compounds **12**:548-550.

Krupkin, A., A. Liston, and S. Strauss. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot **83**:489-498.

Kumar, S., T. K., and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics. Briefings in Bioinformatics **5**:150-163.

Kutil, B., and C. Williams. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J Heredity **92**:327-332.

Lauria, F. 1997. The taxonomic status of *Pinus washoensis* H. Mason & Stockw. (Pinaceae). Ann Naturhist Mus Wien **99B**:655-671.

Lauria, F. 1996. The identity of *Pinus ponderosa* Douglas ex. C. Lawson (Pinaceae). Linzer biologische Beitrag **28/2**:999-1052.

Liewlaksaneeyanawin, C., C. Ritland, Y. El-Kassaby, and K. Ritland. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet **109**:361-369.

Liston, A., M. Parker-Defeniks, J. V. Syring, A. Willyard, and R. Cronn. Interspecific phylogenetic analysis enhances intraspecific phylogeographic inference: A case study in *Pinus lambertiana*. Mol. Ecol. **in press**.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research **27**:209-220.

Mariette, S., D. Chagne, S. Decroocq, G. Vendramin, C. Lalanne, D. Madur, and C. Plomion. 2001. Microsatellite markers for *Pinus Pinaster* Ait. Ann For Sci **58**:203-206.

Marshall, H. D., C. Newton, and K. Ritland. 2002. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). Theor Appl Genet **104**:367-378.

Mason, H. L., and W. P. Stockwell. 1945. A new pine from Mount Rose, Nevada. Madroño **8**:61-63.

Mastrogiuseppe, R., and J. Mastrogiuseppe. 1995. Mimicry among the pines? Nature Notes **XXVI**.

Matos, J., and B. Schaal. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. Evolution **54**:1218-1233.

Matsuoka, Y., Y. Vigouroux, M. Goodman, J. Sanchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA **99**:6080-6084.

Mirov, N. T. 1961. Composition of gum turpentines of pines. USDA Forest Service Technical Bulletin **1239**.

Mirov, N. T. 1967. The genus *Pinus*. The Ronald Press Company, New York, NY.

Mirov, N. T. 1929. Chemical anlysis of the oleoresins as a means of distinguishing Jeffrey pine and western yellow pine. J Forest **27**:176-187.

Neale, D., and R. Sederoff. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet **77**:212-216.

Neale, D. B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. Trends in Plant Science **9**:325-330.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**:583-590.

Niebling, C., and M. Conkle. 1990. Diversity of Washoe pine and comparisons with allozymes of ponderosa pine races. Can J For Res **20**:298-308.

Norris, J. R., S. T. Jackson, and J. L. Betancourt. 2006. Classification tree and minimum-volume ellipsoid analyses of the distribution of ponderosa pine in the western USA. Journal of Biogeography **33**:342-.

Patten, A., and S. Brunsfeld. 2002. Evidence of a novel lineage within the *Ponderosae*. Madroño **49**:189-192.

Peakall, R., and P. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Prager, E., D. Fowler, and A. Wilson. 1976. Rates of evolution in conifers (Pinaceae). Evolution **30**:637-649.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity **86**:248-249.

Rehfeldt, G. 1999a. Systematics and genetic structure of Washoe pine: Applications in conservation genetics. Silvae Genet **48**:167-173.

Rehfeldt, G. 1999b. Systematics and genetic structure of *Ponderosae* taxa (Pinaceae) inhabiting the mountain islands of the Southwest. Am J Bot **86**:741-752.

Rice, W. 1989. Analyzing tables of statistical tests. Evolution **43**:223-225.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature **425**:798-804.

Savage, T. J., B. S. Hamilton, and R. Croteau. 1996. Biochemistry of short-chain alkanes. Plant Physiol **110**:179-186.

Saylor, L. C. 1961. A karyotypic analysis of selected species of Pinus. Silvae Genet **10**:77-84.

Shepherd, M., M. Cross, M. Dieters, and R. Henry. 2003. Genetic maps for *Pinus elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. Theor Appl Genet **106**:1409-1419.

Shepherd, M., M. Cross, T. Maguire, M. Dietes, C. Williams, and R. Henry. 2002. Transpecific microsatellites for hard pines. Theor Appl Genet **104**:819-827.

Smith, D., and M. Devey. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome **37**:977-983.

Smith, R. H. 1981. Variation in immature cone color of ponderosa pine (Pinaceae) in Northern California and Southern Oregon. Madroño **28**:272-275.

Smith, R. H. 1971. Xylem monoterpenes of *Pinus ponderosa*, *P. washoensis*, and *P. jeffreyi* in the Warner Mountains of California. Madroño **21**:26-32.

Smith, R. H. 1967. Variations in the monoterpene composition of the wood resin of Jeffrey, Washoe, Coulter and Lodgepole Pines. Forest Science **13**:1967.

Smouse, P., and J. Long. 1992. Matrix correlation analysis in anthropology and genetics. Yearbook Phys Anthropol **35**:187-213.

Smouse, P., J. Long, and R. Sokal. 1986. Multiple regression extensions of the Mantel test of matrix correspondence. Systematic Zoology **35**:627-632.

Sorensen, F. C., N. L. Mandel, and J. E. Aagaard. 2001. Role of selection versus historical isolation in racial differentiation of ponderosa pine in southern Oregon: an investigation of alternative hypotheses. Can J For Res **31**:1127-1139.

Squillace, A. E., and R. R. Silen. 1962. Racial variation in ponderosa pine. Forest Science Monograph **2**:27pp.

Stead, J. W. 1983. Studies in Central American Pines V: a numerical study of variation in the Pseudostrobus group. Silvae Genet **32**:101-115.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437-460.

Turpeinen, T., T. Tenhola, O. Manninen, E. Nevo, and E. Nissila. 2001. Microsatellite diversity associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol Ecol **10**:1577-1591.

Udupa, S. M., and M. Baum. 2001. High mutation rate and mutational bias at (TAA)n microsatellite loci in chickpea (*Cicer arietinum* L.). Mol Genet Genomics **265**:1097-1103.

Vander Wall, S. B., and R. P. Balda. 1977. Coadaptatons of the Clark's nutcracker and the Pinon pine for efficient seed harvest and dispersal. Ecological Monographs **47**:89-111.

Weidman, R. H. 1939. Evidences of racial influence in a 25-year test of ponderosa pine. Journal of Agricultural Research **59**:855-887.

Weir, B. S. 1996. Genetic data analysis II: methods for discrete population genetic data. Sinaur Assoc., Sunderland, Mass.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

Wells, O. O. 1964a. Geographic variation in ponderosa pine II. Correlations between progeny performance and characteristics of the native habitat. Silvae Genet **13**:126-132.

Wells, O. O. 1964b. Geographic varation in ponderosa pine I. The ecotypes and their distribution. Silvae Genet **13**:89-103.

Williams, C. G., C. G. Elsik, and R. D. Barnes. 2000. Microsatellite analysis of *Pinus taeda* L. in Zimbabwe. Heredity **84**:261-268.

Willyard, A., J. V. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol **24**:90-101.

Zhang, J., and B. M. Cregg. 2005. Growth and physiological responses to varied environments among populations of *Pinus ponderosa*. Forest Ecology and Management **219**:1-12.

Zhou, Y., D. Gwaze, M. Reyes-Valdes, T. Bui, and C. Williams. 2003. No clustering for linkage map based on low-copy and undermethylated microsatellites. Genome **46**:809-816.

GENERAL DISCUSSION

The overall theme for these three manuscripts is a recommendation of caution with the interpretation of results regarding *Pinus* evolution. The cliché regarding losing sight of the forest through all of the trees seems to be exceptionally prophetic for the study of *Pinus* systematics. Hopefully, the perspectives I have provided will contribute to a more solid foundation from which to view the valuable contributions that have been made to *Pinus* biology over the centuries.

Mutation rate estimates, whether based on a clock-like data set or a likelihood-based estimate, require a framework for calibration. I took advantage of the *Pinus* fossil affinities suggested by generations of paleobotanists to select a fossil/node association that offers a clear starting point for such a calibration. The second chapter of this dissertation illustrates the dramatic distortion in ages (and inferred rates) that occurs if a fossil is placed at a tip, rather than a stem, node in a molecular phylogeny. These inconsistent fossil calibrations are the likely cause of widely differing estimates that have been made regarding *Pinus* mutation rates. I suggest that calibration at the well-defined subgeneric split within *Pinus* is only the beginning of this process. Important questions remain about whether the age of this node is best supported by fossil leaves and cones (ca. 45 million years ago) or fossil wood (ca. 85 million years ago). I also report a wide difference in rates between loci, and suggest that single-locus estimates be compared to multi-locus measurements. Regardless, I suggest that my range of locus-specific rate estimates paired with either fossil support a moderate tempo of mutation rate in *Pinus,* and conclude that the within-locus rate variation and leaves vs. wood question are both minor factors in comparison to the enormous effects of incorrect fossil/node association. I am cautious in my inferences, as I think that placing the rate within a reasonable order of magnitude is all that my evidence supports. Further accuracy will await paleobotanical evaluation of both fossil and extant taxa so that more focused calibrations can be made.

A direct consequence of a moderate, rather than very slow, mutation tempo in *Pinus* is the inference that many *Pinus* clades radiated rather recently. I note that despite evidence that the genus is ancient, a large portion of the relative divergence is near the base of the phylogenetic tree. For example, in the second chapter of this dissertation, I note that in terms of relative rates, the divergence between sections in subg. *Pinus* contains only one-sixth of the total divergence within the genus. This indicates that sections, subsections, and especially many species, are quite young. While not the first to suggest this pattern, my framework provides multi-locus nuclear and chloroplast data to support it. I avoid biogeographical speculation that I feel is premature for my evidence and the accuracy with which I can infer ages. Rather, I build upon these projected recent divergences to aid my interpretation of species-level coalescence in the third chapter of this dissertation.

Within the *Ponderosae*, I find rather different phylogenetic trees inferred from two independent nuclear loci in chapter three of this dissertation. Despite the widespread lack of species-level resolution, I have some examples of multiple within-species accessions that resolve into more than one clade. The simplest explanation for two of these examples (the *P. ponderosa/P. washoensis* complex and *P. coulteri*) is that reticulate evolution has occurred in the history of these lineages.

For *P. coulteri*, I observed a derived, monophyletic clade containing all three samples in one locus, yet retrieved the expected clustering of these same three samples with the other California big-cone pines using the second locus and for the chloroplast haplotype. One of the possible explanations for this pattern would be a diploid hybrid speciation event, with *P. coulteri* arising from introgression between a member of the Sabinianae group and the Ponderosae group. Alternative explanations include introgression after differentiation, as well as the retention of ancestral polymorphism at these two loci. In keeping with my theme of applying caution in the interpretation of *Pinus* results, I suggest that the origin of this species merits further work.

At the population level, chapter four of this dissertation found a clear divergence between three populations of *P. jeffreyi* compared to eight populations of

the *P. ponderosa/P. washoensis* complex using six nSSRs. I found no differentiation between *P. ponderosa* and *P. washoensis* individuals or populations using most analytical methods. However, there is very weak evidence for some clustering of the traditional *P. washoensis* populations and two high-elevation putative *P. washoensis* populations from Oregon apart from the *P. ponderosa s.s.* populations. This result is only evident using F_{st} estimates between pooled populations and using PCA based on allele frequencies. Results from admixture simulations using six nSSRs suggest a low, but not negligible, level of introgression between *P. jeffreyi* and members of the *P. ponderosa/P. washoensis* complex. Surprisingly, this pattern was found in samples collected in allopatric populations as well, suggesting that it may be retained from ancient polymorphism, perhaps affected by ancient reticulation.

I observed a chloroplast haplotype that is inferred to be unique to the *P. ponderosa/P. washoensis* complex and to have introgressed into a few *P. jeffreyi* accessions. One *P. ponderosa* individual from each study lacked this "ponderosa-style" haplotype. The low level of ongoing bi-directional introgression that I have observed would provide an important source of migrants into the genetic milieu of both species. These new alleles would be expected to add to the already enormous effective population size, further slowing the process of coalescence.

Using the phylogenetic framework presented in the third chapter of this dissertation and the population-level sampling presented in the fourth chapter, I offer the explanation that the *P. ponderosa/P. washoensis* complex retains an incredibly large amount of variability. Genealogy may vary by locus due to selection, stochastic events, and/or occasional migrants from introgression with *P. jeffreyi*. These factors are likely to have produced the differing patterns observed in phylogenetic trees built using independent loci. We think that it would be an over-interpretation to ascribe species status to the Oregon populations sampled here as 'putative *P. washoensis*'. Likewise, it would be premature to suggest subsuming traditional *P. washoensis* within *P. ponderosa s.l.* based on results from presumably neutral loci. Demonstration of a lack of distinctiveness can be challenging, as this question bears the burden of proof that adequate and appropriate populations, accessions, and alleles

have been sampled. Further, the presence, or at least the frequency, of non-neutral alleles may differ in these distinctive, high-elevation populations. Thus, *P. washoensis* remains an enigmatic taxon.

My inference that reticulation has played a role in the evolutionary history of pine species supports the suggestions by Anderson, Stebbins, Ellstrand, and Schierenbeck (Anderson and Stebbins 1954; Stebbins 1959; Ellstrand and Schierenbeck 2000) that hybridization can make an important contribution to speciation. In the case of *Pinus*, the hybridization events are suggested to occur between divergent lineages with incomplete mating barriers and result in diploid progeny. In the case of the vast populations of var. *ponderosa s.l.*, these hybridization events may contribute to the maintenance of their diversity and to their slow coalescence. In contrast, hybridization events may have created the opportunity for a strikingly different lineage to develop into *P. coulteri,* which exhibits a limited level of genetic diversity and a very unique morphology as well as a distinctive habitat.

LITERATURE CITED

Anderson, E., and G. L. Stebbins. 1954. Hybridization as an evolutionary stimulus. Evolution **8**:378-388.

Ellstrand, N., and K. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proc. Natl. Acad. Sci. U.S.A. **97**:7043-7050.

Stebbins, G. L. 1959. The role of hybridization in evolution. Proc. of the American Philosophical Society **103**:231-251.

BIBLIOGRAPHY

Al-Rabab'ah, M. A., and C. G. Williams. 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. Forest Ecology and Management **163**:263-271.

Alvin K. 1960. Further conifers of the Pinaceae from the Wealden Formation of Belgium. Institut Royal des Sciences Naturelles de Belgique, Mémoires 146:1-39.

Anderson, E., and G. L. Stebbins. 1954. Hybridization as an evolutionary stimulus. Evolution **8**:378-388.

Axelrod D. 1986. Cenozoic history of some western American pines. Ann Mo Bot Gard 73:565-641.

Birky, C. W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103**:513-527.

Blackwell W. 1984. Fossil ponderosa-like pine wood from the Upper Cretaceous of northeast Mississippi. Ann Bot London 53:133-136.

Boratynska, K., and M. Bobowicz. 2001. *Pinus uncinata* Ramond taxonomy based on needle characters. Plant Syst Evol **227**:183-194.

Borchert, M. 1985. Serotiny and cone-habit variation in populations of *Pinus coulteri* (Pinaceae) in the southern Coast Ranges of California. Madroño **32**:29-48.

Boys, J., M. Cherry, and S. Dayanandan. 2005. Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). Am J Bot **92**:833-841.

Brayshaw, T. C. 1997. Washoe and ponderosa pines on Promontory Hill near Merritt, B.C., Canada. Ann Naturhist Mus Wien **99B**:673-680.

Brown G, Gill G, Kuntz R, Langley C, Neale D. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101:15255-15260.

Brown G, Kadel E, Bassoni D, Kiehne K, Temesgen B, van Buijtenen J, Sewell M, Marshall K, Neale D. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159:799-809.

Brownstein, M. J., J. D. Carpten, and J. R. Smith. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques **20**:1004-1010.

Bucci, G., M. Anzidei, A. Madaghiele, and G. G. Vendramin. 1998. Detection of haplotypic variation and natural hybridization in halepensis-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol Ecol **7**:1633-1643.

Callaham, R. Z., and A. R. Liddicoet. 1961. Altitudinal variation at 20 years in ponderosa and Jeffrey pines. J Forest **59**:814-820.

Callaham, R. Z., and W. Metcalf. 1959. Altitudinal races of *Pinus ponderosa* confirmed. J Forest **57**:50-502.

Cavalli-Sforza, L. L., and W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution **21**:550-570.

Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C. 2003. Comparative genome and QTL mapping between maritime and loblolly pines. Mol Breeding 12:185-195.

Chapuis, M.-P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol **24**:621-631.

Chen, D.-H., and P. C. Ronald. 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR applications. Plant Molecular Biology Reporter **17**:53-57.

Christensen, K. I., and G. H. Dar. 1999. A morphometric study of hybridization between *Pinus mugo* and *P. sylvestris* (Pinaceae). Acta Horti **615**:211-221.

Clark R, Tavaré S, Doebley J. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Mol Biol Evol 22:2304-2312.

Conkle, M. T., and W. B. Critchfield. 1988. Genetic variation and hybridization of ponderosa pine. In "Ponderosa pine: the species and its management", D.M. Baumgartner & James E. Lotan, eds.

Critchfield, W. 1986. Hybridization and classification of the white pines (*Pinus* section *Strobus*). Taxon **35**:647-656.

Critchfield, W. B. 1966. Crossability and relationships of the California big-cone pines. USDA Forest Service Research Paper **NC-6**:36-44.

Critchfield, W. B. 1984. Crossability and relationships of Washoe pine. Madroño **31**:144-170.

Critchfield, W. B., and G. L. Allenbaugh. 1965. Washoe pine on the Bald Mountain Range, California. Madroño **18**:63-64.

Critchfield, W., and E. Little. 1966. Geographic Distribution of the Pines of the World. USDA Forest Service Misc. Publication 991:91 pp.

Cunningham, C. 1997. Can three incongruence tests predict when data should be combined? Mol Biol Evol 14:733-740.

Dakin, E. E., and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity **93**:504-509.

Debreczy, and Racz. 1995. New species and varieties of conifers from Mexico. Phytologia **78**:233-234.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B **39**:1-38.

DeVerno, L., and A. Mosseler. 1997. Genetic variation in red pine (*Pinus resinosa*) revealed by RAPD and RAPD-RFLP analysis. Can J For Res **27**:1316-1320.

Dieringer, D., and C. Schlotterer. 2003. Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes **3**:167-169.

Dvorak, W., A. Jordan, G. Hodge, and J. Romero. 2000. Assessing evolutionary relationships of pines in the *Oocarpae* and *Australes* subsections using RAPD markers. New Forests **20**:163-192.

Dvornyk V, Sirviö A, Mikkonen M, Savolainen O. 2002. Low nucleotide diversity at the *pal1* locus in the widely distributed *Pinus sylvestris*. Mol Biol Evol 19:178-188.

Echt, C. S., L. L. DeVerno, M. Anzidei, and G. G. Vendramin. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol Ecol **7**:307-316.

Eckert, A. J., and B. D. Hall. 2006. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenetics and Evol **40**:166-182.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics **5**:435-445.

Ellstrand, N., and K. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci USA **97**:7043-7050.

Emrich, S. J., L. Li, T.-J. Wen, M. Yandeau-Nelson, Y. Fu, L. Guo, H.-H. Chou, S. Aluru, D. A. Ashlock, and P. S. Schnable. 2007. Nearly identical paralogs: implications for maize (*Zea mays* L.) genome evolution. Genetics **175**:429-429.

Epperson, B. K., M. G. Chung, and F. W. Telewski. 2003. Spatial pattern of allozyme variation in a contact zone of *Pinus ponderosa* and *P. arizonica* (Pinaceae). Am J Bot **90**:25-31.

Epperson, B., F. Telewski, A. Plovanich-Jones, and J. Grimes. 2001. Clinal differentiation and putative hybridididization in a contact zone of *Pinus ponderosa* and *P. arizonica* (Pinaceae). Am J Bot **88**:1052-1057.

Erwin D, Schorn H. 2006. *Pinus baileyi* (Section *Pinus*, Pinaceae) from the Paleogene of Idaho, USA. Am J Bot 93:197-205.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**:2611-2620.

Excoffier, L., and M. Slatkin. 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. Am J Hum Genet **62**:171-180.

Excoffier, L., P. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA and restriction data. Genetics **131**:479-491.

Faldt, J., K. Sjodin, M. Persson, I. Valterova, and A.-K. Borg-Karlson. 2004. Correlations between selected monoterpene hydrocarbons in the xylem of six *Pinus* (Pinaceae) species. Chemoecology **11**:97-106.

Falush, D., M. Stephens, and J. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.

Farjon, A., and B. Styles. 1997. Flora Neotropica, Monograph 75, *Pinus* (Pinaceae). The New York Botanical Garden, New York, NY, USA.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, MA.

Fliche P. 1896. Etude sur la flore fossile de l'Argonne (Albien-Cénomanien), Nancy, France.

Furnier, G., and W. Adams. 1986. Geographic patterns of allozyme variation in Jeffrey pine. Am J Bot **73**:1009-1015.

García-Gil M, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12:1195-1206.

Gaut B, Morton B, McCaig B, Clegg M. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA. 93:10274-10279.

Gaut B, Muse S, Clark W, Clegg M. 1992. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. J Mol Evol 35:292-303.

Geada López G, Kamiya K, Harada K. 2002. Phylogenetic relationships of Diploxylon pines (subgenus *Pinus*) based on plastid sequence data. Int J Plant Sci 163:737-747.

Gernandt D, Geada López G, Ortiz García S, Liston A. 2005. Phylogeny and classification of *Pinus*. Taxon 54:29-42.

Gernandt, D., A. Liston, and D. Pinero. 2001. Variation in the nrDNA *ITS* of *Pinus* Subsection *Cembroides*: implications for molecular systematics studies of pine species complexes. Mol Phylogenetics and Evol **21**:450-468.

Gernandt, D., G. Gaeda Lopez, S. Garcia, and A. Liston. 2005. Phylogeny and classification of *Pinus*. Taxon **54**:29-42.

Gerson, E. A., and R. G. Kelsey. 2004. Piperidine alkaloids in North American *Pinus* taxa: implications for chemosystematics. Biochemical Systematics and Ecology **32**:63-74.

Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA **92**:6723-6727.

Gonzalez-Martinez, S., E. Ersoz, G. R. Brown, N. C. Wheeler, and D. B. Neale. 2006. DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. Genetics **172**:1915-1926.

Govindaraju D, Lewis P, Cullis C. 1992. Phylogenetic analysis of pines using ribosomal DNA restriction fragment polymorphisms. Plant Syst Evol 179:141-153.

Gradstein F, Ogg J. 2004. Geologic time scale 2004: why, how and where next! Lethaia 37:175-181.

Grotkopp, E., M. Rejmanek, M. J. Sanderson, and T. L. Rost. 2004. Evolution of genome sze in pines (*Pinus*) and its life-history correlates: supertree analyses. Evolution **58**.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Haller, J. R. 1961. Some recent observations on ponderosa, Jeffrey, and Washoe pines in Northeastern California. Madroño **16**:126-132.

Haller, J. R. 1962. Variation and hybridization in ponderosa and Jeffrey pines. Univ Calif Publ Bot **34**:123-165.

Haller, J. R. 1965. *Pinus washoensis* in Oregon: taxonomic and evolutionary implications. Am J Bot **52**:646 (abstract).

Haller, J. R. 1986. Taxonomy and relationships of the mainland and island populations of *Pinus torreyana* (Pinaceae). Syst Bot **1**:39-50.

Hamrick, J. L., and M. J. W. Godt. 1996. Effects of life history traits on genetic diversity in plant species. Phil Trans R Soc Lond **351**:1291-1298.

Hart J. 1987. A cladistic analysis of conifers: preliminary results. J Arnold Arboretum 68:269-307.

Haubold B, Wiehe T. 2001. Statistics of divergence times. Mol Biol Evol 18:1157-1160.

Helmers, A. E. 1943. The ecological anatomy of ponderosa pine needles. The American Midland Naturalist **29**:55-71.

Hizume, M., Shibata, F., Y. Matsusaki, and Z. Garajova. 2002. Chromosome identification and comparative karyotypic analyses of four *Pinus* species. Theor Appl Genet **105**:491-497.

Hong, Y.-P., A. Krupkin, and S. Strauss. 1993. Chloroplast DNA transgresses species boundaries and evolves at variable rates in the California closed-cone pines (*Pinus radiata, P. muricata*, and *P. attenuata*). Mol Phylogenetics and Evol **2**:322-329.

Hosono, S., A. F. Faruqi, F. B. Dean, Y. Du, Z. Sun, X. Wu, J. Du, S. F. Kingsmore, M. Egholm, and R. S. Lasken. 2003. Unbiased whole-genome amplification directly from clinical samples. Genome Research **13**:954-964.

Hudson R. 1960. The anatomy of the genus *Pinus* in relation to its classification. J I Wood Sci 6:26-46.

Huelsenbeck, J. P., and J. P. Ronquist. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572-1574.

Islam-Faridi, M. N., C. D. Nelson, and T. L. Kubisiak. 2007. Reference karyotype and cytomolecular map for loblolly pine (*Pinus taeda* L.). Genome **50**:241-251.

Jakobsson, M., and N. A. Rosenbeg. 2006. CLUMPP: CLUster Matching and Permutation Program.

Jeffrey E. 1908. On the structure of the leaf in Cretaceous pines. Ann Bot London 22:207-220.

Johansen, A. D., and R. G. Latta. 2003. Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. Mol Ecol **12**:293-298.

Johnson, M., S. B. Vander Wall, and M. Borchert. 2003. A comparative analysis of seed and cone characteristics and seed-dispersal strategies of three pines in the subsection Sabinianae. Plant Ecology **168**:69-84.

Kalinowski, S. 2002. How many alleles per locus should be used to estimate genetic distances? Heredity **88**:62-65.

Karalamangala, R., and D. Nickrent. 1989. An electrophoretic study of representatives of subgenus Diploxylon of *Pinus*. Can J Bot **67**:1750-1759.

Karhu, A., J.-H. Dieterich, O. Savolainen. 2000. Rapid expansion of microsatellite sequences in pines. Mol Biol Evol **17**:259-265.

Kärkkäinen K, Koski K, Savolainen O. 1996. Geographical variation in the inbreeding depression of Scots pine. Evolution 50:111-119.

Kay K, Whittall J, Hodges S. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. Evol Biol 6:36-44.

Kinlaw, C., and D. Neale. 1997. Complex gene families in pine genomes. Trends in Plant Science **2**:356-359.

Kjer, K. M., J. J. Gillespie, and K. A. Ober. 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst Biol **56**:133-146.

Klekowski E, Godfrey P. 1989. Ageing and mutation in plants. Nature 340:389-391.

Klekowski E. 1992. Review: mutation rates in diploid annuals-are they immutable? Int J Plant Sci 153:462-465.

Klekowski E. 1998. Mutation rates in mangroves and other plants. Genetica 102:325-331.

Koch M, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17:1483-1498.

Kolesnikova, R. D., V. G. Latysh, and A. I. Chernodubov. 1977. A chromatographic investigation of the n-heptane in the essential oil of representatives of the family Pinaceae. Chemistry of natural compounds **12**:548-550.

Komulainen P, Brown G, Mikkonen M, Karhu A, García-Gil M, O'Malley D, Lee B, Neale D, Savolainen O. 2003. Comparing EST based genetic maps between *Pinus sylvestris* and *P. taeda*. Theor Appl Genet 107:667-678.

Kossack D, Kinlaw C. 1999. *IFG*, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39:417-426.

Koteja J, Poinar G. 2001. A new family, genus, and species of scale insect (Hemiptera: Coccinea: Kukaspididae, new family) from Cretaceous Alaskan amber. Proc Entomol Soc Wash 103:356-363.

Kral, R. 1993. *Pinus in* N. Morin, ed. Flora of North America. Oxford University Press, Oxford, U.K.

Krupkin, A., A. Liston, and S. Strauss. 1996. Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot **83**:489-498.

Krutovsky K, Troggio M, Brown G, Jermstad K, Neale D. 2004. Comparative mapping in the Pinaceae. Genetics 168:447-461.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol **1**:17-24.

Kumar S, Filipski A, Swarna V, Walker A, Hedges S. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA. 102:18842-18847.

Kumar, S., T. K., and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics. Briefings in Bioinformatics **5**:150-163.

Kutil B, Williams C. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J Hered 92:327-332.

Lande R, Schemske D, Schultz S. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evolution 48:965-978.

Langenheim R, Smiley C, Gray J. 1960. Cretaceous amber from the Arctic coastal plain of Alaska. Bulletin of the Geologic Society of America 71:1345-1356.

Lanner, R. M., and A. M. Phillips. 1992. Natural hybridization and introgression of pinyon pines in northwestern Arizona. Int J Plant Sci **153**:250-257.

Lauria, F. 1996. The identity of *Pinus ponderosa* Douglas ex. C. Lawson (Pinaceae). Linzer biologische Beitrag **28/2**:999-1052.

Lauria, F. 1997. The taxonomic status of *Pinus washoensis* H. Mason & Stockw. (Pinaceae). Ann Naturhist Mus Wien **99B**:655-671.

Ledig, F. T. 1998. Genetic variation in *Pinus*. Pp. 251-280. *in Richardson DM, Editor, Ecology and Biogeography of* Pinus.

Ledig, F. T. 2000. Founder effects and the genetic structure of Coulter pine. J Heredity **91**:307-315.

Ledig, F., and M. Conkle. 1983. Gene diversity and genetic structure in a narrow endemic, Torrey pine (*Pinus torreyana* Parry ex Carr.). Evolution **37**:79-85.

Libby, W. 1958. The backcross hybrid Jeffrey x (Jeffrey x Coulter) pine. J Forest **56**:840-842.

Liewlaksaneeyanawin, C., C. Ritland, Y. El-Kassaby, and K. Ritland. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet **109**:361-369.

Liston, A. 1992. Variation in the chloroplast genes *rpoC1* and *rpoC2* of the genus *Astragalus* (Fabaceae): Evidence from restriction site mapping of a PCR-amplified fragment. . Am J Bot **79**:953-961.

Liston, A., M. Parker-Defeniks, J. V. Syring, A. Willyard, and R. Cronn. Interspecific phylogenetic analysis enhances intraspecific phylogeographic inference: A case study in *Pinus lambertiana*. Mol. Ecol. **in press**.

Liston, A., W. Robinson, D. Pinero, and E. Alvarez-Buylla. 1999. Phylogentics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Mol Phylogenetics and Evol **11**:95-109.

Little, E., and W. Critchfield. 1969. Subdivisions of the Genus *Pinus* (Pines). U.S.D.A. Forest Service **Misc Pub 1144**.

Ma, X.-F., A. E. Szmidt, and X.-R. Wang. 2006. Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. Mol Biol Evol **23**:807-816.

Magallón S, Sanderson M. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762-1780.

Magallón S, Sanderson M. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution 59:1653-1670.

Mai D. 1986. Über typen und originale tertiärer arten von *Pinus* L. (Pinaceae) in mitteleuropÜäischen sammlungen - ein beitrag zur geschichte der gattung in Europa. Feddes Repertorium 97:571-605.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research **27**:209-220.

Marcysiak, K., and A. Boratynski. 2007. Contribution to the taxonomy of *Pinus uncinata* (Pinaceae) based on cone characters. Plant Syst Evol **264**:57-73.

Mariette, S., D. Chagne, S. Decroocq, G. Vendramin, C. Lalanne, D. Madur, and C. Plomion. 2001. Microsatellite markers for *Pinus Pinaster* Ait. Ann For Sci **58**:203-206.

Marshall, H. D., C. Newton, and K. Ritland. 2002. Chloroplast phylogeography and evolution of highly polymorphic microsatellites in lodgepole pine (*Pinus contorta*). Theor Appl Genet **104**:367-378.

Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21**:260-262.

Martinez, M. 1948. Los Pinos Mexicanos, Segunda Edicion, Mexico.

Mason, H. L., and W. P. Stockwell. 1945. A new pine from Mount Rose, Nevada. Madroño **8**:61-63.

Mastrogiuseppe, R., and J. Mastrogiuseppe. 1995. Mimicry among the pines? Nature Notes **XXVI**.

Matos, J. A. 1995. *Pinus hartwegii* and *P. rudis*: A critical assessment. Syst Biol **20**:6-21.

Matos, J. A. 1998. A coalescent approach to chloroplast genome relationships within and between populations of *Pinus devoniana* in Mexico. Aliso **17**:145-156.

Matos, J., and B. Schaal. 2000. Chloroplast evolution in the *Pinus montezumae* complex: A coalescent approach to hybridization. Evolution **54**:1218-1233.

Matsuoka, Y., Y. Vigouroux, M. Goodman, J. Sanchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA **99**:6080-6084.

Meijer J. 2000. Fossil woods from the Late Cretaceous Aachen Formation. Rev Palaeobot Palyno 112:297-336.

Mendoza, C. F. V., and A. Rodriguez-Banderas. 2006. Evolutionary analysis of *Pinus leiophylla*: a study using an Intron II sequence fragment of mitochondrial *nad1*. Can J Bot **84**:172-177.

Millar C. 1998. Early evolution of pines. In: Richardson D, ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge, U.K. p. 69-91.

Miller C, Malinky J. 1986. Seed cones of *Pinus* from the Late Cretaceous of New Jersey, U.S.A. Rev Palaeobot Palyno 46:257-272.

Miller C. 1973. Silicified cones and vegetative remains of *Pinus* from the Eocene of British Columbia. Contributions from the Museum of Paleontology, The University of Michigan 24:101-118.

Miller C. 1976. Early evolution in the Pinaceae. Rev Palaeobot Palyno 21:101-117.

Mirov, N. T. 1929. Chemical anlysis of the oleoresins as a means of distinguishing Jeffrey pine and western yellow pine. J Forest **27**:176-187.

Mirov, N. T. 1961. Composition of gum turpentines of pines. USDA Forest Service Technical Bulletin **1239**.

Mirov, N. T. 1967. The genus *Pinus*. The Ronald Press Company, New York, NY.

Muse S, Weir B. 1992. Testing for equality of evolutionary rates. Genetics 132:269-276.

Muse S. 2000. Examining rates and patterns of nucleotide substitution in plants. Plant Mol Biol 42:25-43.

Mustilli A-C, Merlot S, Vavasseur A, Fenzi F, Giraudat J. 2002. *Arabidopsis OST1* protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell 14:3089-3099.

Namkoong, G. 1966. Statistical analysis of introgression. Biometrics **22**:488-502.

Neale, D. B., and O. Savolainen. 2004. Association genetics of complex traits in conifers. Trends in Plant Science **9**:325-330.

Neale, D., and R. Sederoff. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet **77**:212-216.

Near T, Meylan P, Shaffer H, 2. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat 165:137-146.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**:583-590.

Niebling, C., and M. Conkle. 1990. Diversity of Washoe pine and comparisons with allozymes of ponderosa pine races. Can J For Res **20**:298-308.

Nkongolo, K., P. Michael, and W. Gratton. 2002. Identification and characterization of RAPD markers inferring genetic relationships among pine species. Genome **45**:51-58.

Norris, J. R., S. T. Jackson, and J. L. Betancourt. 2006. Classification tree and minimum-volume ellipsoid analyses of the distribution of ponderosa pine in the western USA. Journal of Biogeography **33**:342-.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst Biol **53**:47-67.

Paine, T. D., and C. C. Hanlon. 1994. Influence of oleoresin constituents from *Pinus ponderosa* and *Pinus jeffreyi* on growth of mycangial fungi from *Dendroctonus ponderosae* and *Dendroctonus jeffreyi*. Journal of Chemical Ecology **20**:2551-2563.

Patten, A., and S. Brunsfeld. 2002. Evidence of a novel lineage within the *Ponderosae*. Madroño **49**:189-192.

Peakall, R., and P. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Penny J. 1947. Studies on the conifers of the Magothy flora. Am J Bot 34:281-296.

Perry DJ, Bousquet J. 1998a. Sequence-tagged-site (STS) markers of arbitrary genes: development, characterization and analysis of linkage in black spruce. Genetics 149:1089-1098.

Perry DJ, Bousquet J. 1998b. Sequence-tagged-site (STS) markers of arbitrary genes: the utility of black spruce-derived STS primers in other conifers. Theor Appl Genet 97:735-743.

Perry, J. P. 1991. The pines of Mexico and Central America. Timber Press, Portland, OR, USA.

Petit, R. J., R. Bialozyt, P. Garnier-Gere, and A. Hampe. 2004. Ecology and genetics of tree invasons: from recent introductions to Quaternary migrations. Forest Ecology and Management **197**:117-137.

Phipps C, Osborn J, Stockey R. 1995. *Pinus* pollen cones from the middle Eocene Princeton Chert (Allenby Formation) of British Columbia. Int J Plant Sci 156:117-124.

Plomion C, Hurme P, Frigerio J-M, Ridolfi M, Pot D, Pionneau C, Avila C, Gallardo F, David H, Neutelings G, et al. 1999. Developing SSCP markers in two *Pinus* species. Mol Breeding 5:21-31.

Pollard, D. A., N. I. Venky, A. M. Moses, and M. B. Eisen. 2006. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. PLOS Genetics **2**:1634-1646.

Posada D, Crandall K. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Prager E, Fowler D, Wilson A. 1976. Rates of evolution in conifers (Pinaceae). Evolution 30:637-649.

Prager, E., D. Fowler, and A. Wilson. 1976. Rates of evolution in conifers (Pinaceae). Evolution **30**:637-649.

Price R, Liston A, Strauss S. 1998. Phylogeny and systematics of *Pinus*. In: Richardson D, ed. Ecology and Biogeography of *Pinus*. Cambridge University Press, Cambridge, U.K. p. 49-68.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity **86**:248-249.

Rehfeldt, G. 1999. Systematics and genetic structure of *Ponderosae* taxa (Pinaceae) inhabiting the mountain islands of the Southwest. Am J Bot **86**:741-752.

Rehfeldt, G. 1999. Systematics and genetic structure of Washoe pine: Applications in conservation genetics. Silvae Genet **48**:167-173.

Rehfeldt, G., B. C. Wilson, S. P. Wells, and R. M. Jeffers. 1996. Phytogeograpic, taxonomic, and genetic implications of phenotypic variation in the Ponderosae of the Southwest. The Southwestern Naturalist **41**:409-418.

Rice, W. 1989. Analyzing tables of statistical tests. Evolution **43**:223-225.

Richardson, D. M., and S. I. Higgins. 1998. Pines as invaders in the southern hemisphere. Pp. 450-473. in Richardson DM, Editor, Ecology and Biogeography of *Pinus*.

Robison C. 1977. *Pinus triphylla* and *Pinus quinquefolia* from the Upper Cretaceous of Massachusetts. Am J Bot 64:726-732.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature **425**:798-804.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution **57**:1465-1477.

Rozas, J., J. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Saiki K. 1996. *Pinus mutoi* (Pinaceae), a new species of permineralized seed cone from the Upper Cretaceous of Hokkaido, Japan. Am J Bot 83:1630-1636.

Sanderson M, Doyle J. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and *18s* rDNA data. Am J Bot 88:1499-1516.

Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101-109.

Savage, T. J., B. s. Hamilton, and R. Croteau. 1996. Biochemistry of short-chain alkanes. Plant Physiol **110**:179-186.

Saylor, L. C. 1961. A karyotypic analysis of selected species of Pinus. Silvae Genet **10**:77-84.

Schwilk, D. W., and D. D. Ackerly. 2001. Flammability and serotiny as strategies: correlated evolution in pine. Oikos **94**:326-346.

Scofield D, Schultz S. 2006. Mitosis, stature and evolution of plant mating systems: low-Φ and high-Φ plants. Proc Royal Soc Lond B Bot 273:275-282.

Shaw G. 1914. The genus *Pinus*. Riverside Press, Cambridge, MA.

Shaw, J., E. B. Lickey, and J. T. Beck. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot **92**:142-166.

Shepherd, M., M. Cross, M. Dieters, and R. Henry. 2003. Genetic maps for *Pinus elliottii* var. *elliottii* and *P. caribaea* var. *hondurensis* using AFLP and microsatellite markers. Theor Appl Genet **106**:1409-1419.

Shepherd, M., M. Cross, T. Maguire, M. Dieters, C. Williams, and R. Henry. 2002. Transpecific microsatellites for hard pines. Theor Appl Genet **104**:819-827.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst Biol **49**:369-381.

Six, D. L., and T. D. Paine. 1999. Allozyme diversity and gene flow in *Ophiostoma clavigerum* (Ophiostomatales: Ophiostomataceae), the mycangial fungus of the Jeffrey pine beetle, *Dendroctonus jeffreyi* (Coleoptera: Scolytidae). Can J For Res **29**:324-331.

Smith S, Stockey R. 2002. Permineralized pine cones from the Cretaceous of Vancouver Island, British Columbia. Int J Plant Sci 163:185-196.

Smith, D., and M. Devey. 1994. Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome **37**:977-983.

Smith, J. W., and C. W. Benkman. 2007. A Coevolutionary Arms Race Causes Ecological Speciation in Crossbills. The American Naturalist **169**:455-465.

Smith, R. H. 1967. Variations in the monoterpene composition of the wood resin of Jeffrey, Washoe, Coulter and Lodgepole Pines. Forest Science **13**:1967.

Smith, R. H. 1971. Xylem monoterpenes of *Pinus ponderosa*, *P. washoensis*, and *P. jeffreyi* in the Warner Mountains of California. Madroño **21**:26-32.

Smith, R. H. 1981. Variation in immature cone color of ponderosa pine (Pinaceae) in Northern California and Southern Oregon. Madroño **28**:272-275.

Smouse, P., and J. Long. 1992. Matrix correlation analysis in anthropology and genetics. Yearbook Phys Anthropol **35**:187-213.

Smouse, P., J. Long, and R. Sokal. 1986. Multiple regression extensions of the Mantel test of matrix correspondence. Systematic Zoology **35**:627-632.

Sokol K, Williams C. 2005. Evolution of a triplet repeat in a conifer. Genome 48:417-426.

Sorensen, F. C., N. L. Mandel, and J. E. Aagaard. 2001. Role of selection versus historical isolation in racial differentiation of ponderosa pine in southern Oregon: an investigation of alternative hypotheses. Can J For Res **31**:1127-1139.

Squillace, A. E., and R. R. Silen. 1962. Racial variation in ponderosa pine. Forest Science Monograph **2**:27pp.

Stead, J. W. 1983. Studies in Central American Pines V: a numerical study of variation in the Pseudostrobus group. Silvae Genet **32**:101-115.

Stebbins, G. L. 1959. The role of hybridization in evolution. Proc of the American Philosophical Society **103**:231-251.

Stockey R, Nishida M. 1986. *Pinus harborensis* sp. nov. and affinities of permineralized leaves from the Upper Cretaceous. Can J Botany 64:1856-1866.

Stopes M, Kershaw E. 1910. The anatomy of Cretaceous pine leaves. Ann Bot London 24:395-402.

Strauss S, Doerksen A. 1990. Restriction fragment analysis of pine phylogeny. Evolution 44:1081-1096.

Styles, B. 1993. Genus Pinus: A Mexican Purview. in: Ramamoorthy, editor; Biological diversity of Mexico.

Swofford D. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods). Sinauer Assoc., Sunderland, MA.

Syring, J. V., K. Farrell, R. Businsky, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Syst. Biol. **56**:163-181.

Syring, J., A. Willyard, R. Cronn, and A. Liston. 2005. Evolutionary relationships among pine (Pinaceae) subsections inferred from multiple low-copy nuclear loci. Am J Bot **92**:2086-2100.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics **105**:437-460.

Tallent, W. H., and E. C. Horning. 1955. *Pinus* alkaloids: The alkaloids of *P. sabiniana* Dougl. and related species. Am Chem Soc Jour **78**:4467-4469.

Temesgen, B., G. R. Brown, D. E. Harry, C. S. Kinlaw, M. M. Sewell, and D. B. Neale. 2001. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). Theor Appl Genet **102**:664-675.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research **22**:4673-4680.

Turpeinen, T., T. Tenhola, O. Manninen, E. Nevo, and E. Nissila. 2001. Microsatellite diversity associated with ecological factors in *Hordeum spontaneum* populations in Israel. Mol Ecol **10**:1577-1591.

Udupa, S. M., and M. Baum. 2001. High mutation rate and mutational bias at (TAA)n microsatellite loci in chickpea (*Cicer arietinum* L.). Mol Genet Genomics **265**:1097-1103.

Van der Burgh J. 1973. Hölzer der niederrheinischen braunkohlenformation, 2. hölzer der braunkohlengruben "Maria Theresia" zu herzogenrath, "zukunft west" zu eschweiler und "victor" (zülpich mitte) zu zülpich. nebst einer systematisch-anatomischen bearbeitung der gattung *Pinus* L. Rev Palaeobot Palyno 15:73-275.

Vander Wall, S. B., and R. P. Balda. 1977. Coadaptatons of the Clark's nutcracker and the Pinon pine for efficient seed harvest and dispersal. Ecological Monographs **47**:89-111.

Vogler, D. 2000. Coevolution of *Cronartium* with its host. Hort Technology **10**:518.

Wang, X.-Q., D. Tank, and T. Sang. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol **17**:773-781.

Wang, X.-R., A. Szmidt, and H. Nguyen. 2000. The phylogenetic position of the endemic flat-needle pine *Pinus krempfii* Lec. (Pinaceae) from Vietnam, based on PCR-RFLP analysis of chloroplast DNA. Plant Syst Evol **220**:21-36.

Wang, X.-R., Y. Tsumura, H. Yoshimaru, K. Nagasaka, and A. Szmidt. 1999. Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer and *trnV* intron sequences. Am J Bot **86**:1742-1753.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology **7**:256-276.

Weidman, R. H. 1939. Evidences of racial influence in a 25-year test of ponderosa pine. Journal of Agricultural Research **59**:855-887.

Weir, B. S. 1996. Genetic data analysis II: methods for discrete population genetic data. Sinaur Assoc., Sunderland, Mass.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

Wells, O. O. 1964a. Geographic variation in ponderosa pine II. Correlations between progeny performance and characteristics of the native habitat. Silvae Genet **13**:126-132.

Wells, O. O. 1964b. Geographic varation in ponderosa pine I. The ecotypes and their distribution. Silvae Genet **13**:89-103.

Williams, C. G., C. G. Elsik, and R. D. Barnes. 2000. Microsatellite analysis of *Pinus taeda* L. in Zimbabwe. Heredity **84**:261-268.

Willyard, A., J. V. Syring, D. S. Gernandt, A. Liston, and R. Cronn. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol **24**:90-101.

Wolfe J, Schorn H. 1989. Taxonomic revision of the Spermatopsida of the Oligocene Creede Flora, southern Colorado. U.S. Geological Survey Bulletin 1923:1-40.

Wolfe K, Li W-H, Sharp P. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054-9058.

Wolfe K, Sharp P, Li W-H. 1989. Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29:208-211.

Wolff, R. L., F. Pedrono, E. Pasquier, and A. M. Marpeau. 2000. General characteristics of *Pinus* spp. seed fatty acid compositions, and importance of delta5-olefinic acids in the taxonomy and phylogeny of the genus. Lipids **35**:1-22.

Wu, J., K. Krutovskii, and S. Strauss. 1998. Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics **150**:1605-1614.

Wu, J., K. Krutovskii, and S. Strauss. 1999. Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. Genome **41**:893-908.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 23:212-226.

Yu, X., A. Ekramoddoullah, D. Taylor, and N. Piggott. 2002. Cloning and characterization of a cDNA of *cro rl* from *Cronartium ribicola*. Fungal Genetics and Biology **35**:53-66.

Zhang, J., and B. M. Cregg. 2005. Growth and physiological responses to varied environments among populations of *Pinus ponderosa*. Forest Ecology and Management **219**:1-12.

Zhou, Y., D. Gwaze, M. Reyes-Valdes, T. Bui, and C. Williams. 2003. No clustering for linkage map based on low-copy and undermethylated microsatellites. Genome **46**:809-816.

Zitzetmuller, K., and 5. 2001. Fatty acid composition of Pinaceae as taxonomic markers. American Oil Chemists' Society **36**:439-452.

Zobel, B. 1951. The natural hybrid between Coulter and Jeffrey pines. Evolution **5**:405-413.