AN ABSTRACT OF THE THESIS OF

Junyuan Wu for the degree of Master of Science in Genetics presented on August 26, 1998. Title: Nuclear and Mitochondrial DNA Polymorphism and Phylogeny in the California Closed-Cone Pines

Redacted for Privacy

Abstract approved: _____

Steven H. Strauss

We studied genetic polymorphism and phylogeny using nuclear random amplified polymorphic DNA markers (RAPDs) and mitochondrial DNA (mtDNA) restriction fragment length polymorphisms (RFLPs) in the three California Closed-Cone Pines: *Pinus attenuata* Lemm., *P. muricata* D. Don, and *P. radiata* D. Don. A total of 343 to 384 trees derived from 13 populations were analyzed using 13 mitochondrial gene probes and two restriction enzymes, and more than 90 RAPD loci generated by 22 primers. Southern hybridization was used to test homology among comigrating RAPD markers. Segregation analysis and Southern hybridization were carried out to distinguish between RAPD fragments of nuclear and organellar origin. Estimates of genetic diversity and population differentiation, and phylogenetic analyses based on RAPD and RFLP markers, were compared with those based on allozymes from a similar study.

Twenty-eight distinct mtDNA haplotypes were detected among the three species. All three species showed limited variability within populations, but strong differentiation among populations. Based on haplotype frequencies, genetic diversity within populations (H_S) averaged 0.22, and population differentiation (G_{ST} and θ) exceeded 0.78. Analysis of molecular variance (AMOVA) also revealed that more than 90% of the variation resided among populations. Species and populations could be readily distinguished by unique haplotypes, often using the combination of only a few probes.

Twenty-eight of 30 (93%) comigrating RAPD fragments tested were homologous by Southern hybridization. Hybridization with enriched mtDNA, and chloroplast DNA (cpDNA) clones, identified one fragment as being of mtDNA origin and two as being of cpDNA origin, among 142 RAPD fragments surveyed. RAPD markers revealed

moderately higher intrapopulation gene diversity and significantly higher total genetic diversity and population differentiation than did allozyme markers for each species. Simulation analysis to study effects of dominance on RAPD diversity suggested that dominance substantially depressed values of diversity within populations and inflated values of differentiation among populations. By comparison to our empirical analyses, we inferred that the underlying diversity of RAPD markers is substantially greater than that of allozymes.

Results of phylogenetic analysis of RAPD markers were largely consistent with those from allozyme analysis, though they had many minor differences. Joint phylogenetic analysis of both the RAPD and allozyme markers strongly supported a common ancestor for *P. radiata* and *P. attenuata*, and south to north migration histories for all three species. Dendrograms based on mtDNA analysis, however, strongly disagreed with those based on allozymes, RAPDs, chloroplast DNA and morphological traits, suggesting convergent genome evolution.

Nuclear and Mitochondrial DNA Polymorphism and Phylogeny in the California Closed-Cone Pines

by

Junyuan Wu

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the degree of

Master of Science

Presented August 26, 1998
Commencement June 1999

Master of Science thesis of Junyuan Wu presented on August 26, 1998

APPROVED:

Redacted for Privacy

Major professor, representing Genetics

Redacted for Privacy

Chair of Genetics Program

Redacted for Privacy

Dean of Graduate School

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

Redacted for Privacy

Junyuan Wu, Author

CONTRIBUTION OF AUTHORS

The assays were performed in the laboratory of Dr. Steven Strauss who also wrote the project proposal, provided financial support, and assisted in the sample collection, experimental design, interpretation of data, and writing of each manuscript. Dr. Konstantian Krutovskii provided primer data to make mitochondrial probes, and was also involved in the data analysis and writing of each manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

PREFACE

The thesis is written in a manuscript format to facilitate submission of its main chapters to scholarly journals for publication. It is organized into four chapters followed by a bibliography and appendices. Chapter 1 provides general background and review of literature related to the thesis. Chapter 2, entitled "Abundant mitochondrial genome diversity, population differentiation, and convergent evolution in pines", and chapter 3, entitled "Nuclear DNA diversity, population differentiation and phylogenetic relationships in the California Closed-Cone Pines based on RAPD and allozyme markers" have been submitted in modified form for publication in *Genetics* (accepted) and *Genome*, respectively. The introduction and discussion sections of both manuscripts therefore include information that is used and/or discussed in the general review sections of the thesis. Chapter 4 provides a summary of key results and major conclusions common to the entire thesis and is followed by a bibliography containing all references cited within the thesis. The appendices contain supplemental, useful information pertinent to this thesis.

NUCLEAR AND MITOCHONDRIAL DNA POLYMORPHISM AND PHYLOGENY IN THE CALIFORNIA CLOSED-CONE PINES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

THE CALIFORNIA CLOSED-CONE PINES

Geographic distribution

The California Closed-Cone Pines (CCCP) are composed of three species: one montane, interior species, *Pinus attenuata* Lemmon (knobcone pine), and two maritime, insular species, *P. muricata* D. Don (bishop pine) and *P. radiata* D. Don (Monterey pine) (Figure 1.1). *P. attenuata* occurs at scattered locations on the dry, interior mountains of southern Oregon and California, and at a single location in northern Baja California. *Pinus. radiata* and *P. muricata* are distributed disjunctly along the coasts of California and Baja California, and on four offshore islands: Guadalupe and Cedros Islands for Monterey pine in Mexico, and Santa Cruz and Santa Rosa Islands for bishop pine in California. The three species are sympatric in only a few locations; *P. radiata* and *P. muricata* coexist on the Monterey Peninsula, while *P. attenuata* is associated with *P. radiata* in a few stands near Pt. Año Nuevo (Critchfield and Little 1966; Millar et al. 1988).

Studies on morphological traits, biochemical compounds, and crossability

Monterey pine includes three mainland populations (Año Nuevo, Monterey and Cambria) and two island populations (Guadalupe and Cedros Islands). Although there is variability among the mainland populations in cone and seed size, growth rate, needle

traits, terpene composition, seed proteins, and disease and cold resistance (Axelord 1980; Bannister et al. 1962; Forde 1964; Guinon et al. 1982; Hood and Libby 1980; Murphy 1981), most of the differences are relatively small; the three populations form a closely related unit. The evolutionary relationships among the mainland populations are uncertain because patterns of resemblance vary among different traits. The two island populations, however, have diverged from one another and from the mainland populations in many morphological and biochemical characteristics (reviewed in Millar 1986). The two-needled pines on Guadalupe Island and pines on Cedros Island were once named *P. radiata* var. *binata* (Mason 1932) and *P. radiata* var. *cedrosensis* (Axelord 1980), respectively, because of their distinct differences from other populations. However, all five populations of Monterey pine are interfertile (Critchfield 1967).

Bishop pine contains considerable variation both within and among populations, and patterns of variation are both clinal and discontinuous (reviewed in Millar 1986). Based on cone morphology, Axelord (1983) described *P. muricata* var. *borealis*, first proposed by Duffield (1951), to describe the populations from Monterey northward. Populations north of Sonoma also differ distinctly from those south of Monterey in many traits including growth, stem form, cone abundance, cone sterotiny and bark characteristics, while the Sonoma, Marin and Monterey populations have intermediate characteristics (Doran 1974; Duffield 1951; Everard and Fourt 1974; Fielding 1961; Shelbourne et al. 1982). *Pinus. muricata* var. *borealis* was further separated into two distinct groups based on needle color and anatomy: blue bishop pine for Trindad and Mendocino populations; and green bishop pine for Sonoma, Marin and Monterey populations. The southern mainland populations called *P. muricata* var. *muricata* form a distinct monoterpene type and are cross compatible (Critchfield 1967; Millar et al. 1988) despite being heterogeneous in many traits. Two island populations, Santa Cruz and Santa Rosa, form another distinguishable monoterpene type (Mirov et al. 1966) and they also differ in cone morphology (Mason 1930) and phenology from the mainland bishop pine populations. They are, however, more closely related to southern bishop pine populations than to the island or mainland populations of Monterey pine (Crowley 1974; Shelbourne et al. 1982).

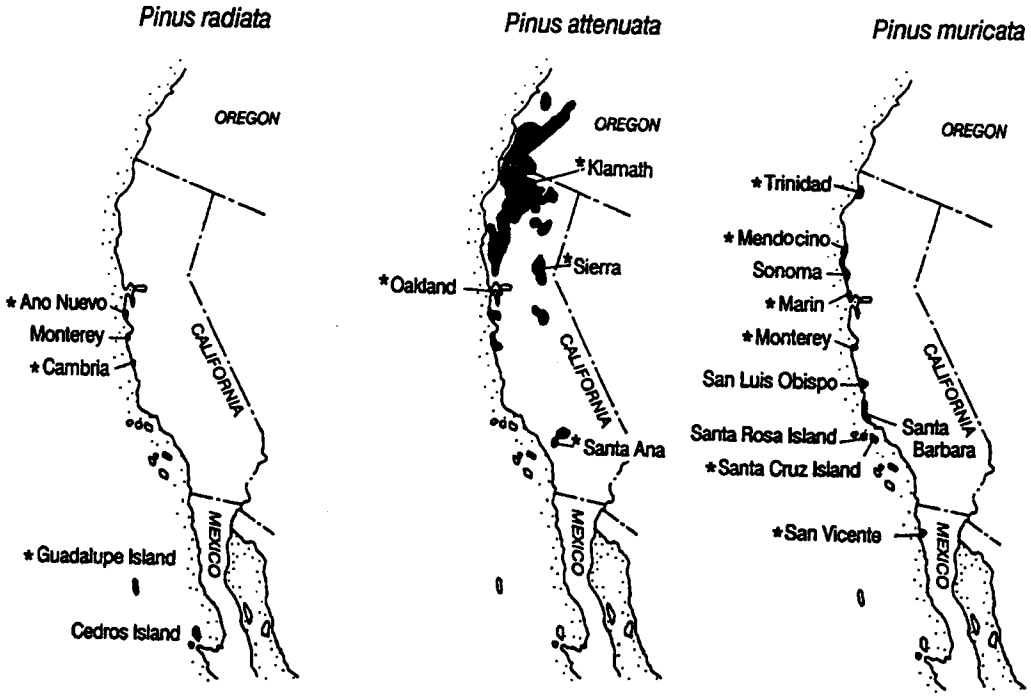


Figure 1.1 Distribution of *Pinus radiata, P. attenuata, and P. muricata* (Hong et al. 1993a) and the origins of sampled populations (*). All populations were studied for chloroplast DNA polymorphism by Hong et al. (1993a).

The island populations were once named as *P. remorata* (Mason 1930) and *P. muricata* var. *remorata* (Duffield 1951) due to the high frequency of trees with thin-scaled, symmetric cones (Axelord 1983; Linhart 1978; Mason 1930).

Fewer investigations on geographic variation in knobcone pine have been made than in the two maritime species. Less variation exists within and among populations of knobcone pine in comparison to the maritime species (Mirov et al. 1966). Several authors have described two patterns of variation of morphology and growth traits in knobcone pine: clinal variation among the northern populations and discontinuous variation with respect to the northern vs. southern California populations. Studies of crossability indicate that knobcone pine populations are interfertile (Critchfield 1967).

Crossing experiments have revealed that Monterey pine can hybridize freely with knobcone pine (Critchfield 1967; Millar et al. 1988). Southern populations of bishop pine are interfertile and can also cross with knobcone pine and Monterey pine. However, the Mendocino population of bishop pine does not cross with the southern mainland nor the island populations, and is highly genetically differentiated from the other populations and species (Critchfield 1967; Millar et al. 1988).

Allozyme studies

There have been several studies of allozyme variation in the CCCP (Millar 1983; Millar et al. 1988; Moran et al. 1988; Plessas and Strauss 1986). Compared with other conifer species, low to moderate levels of allozyme variation exist in Monterey pine (H_S = 0.098 averaged for all five populations, Moran et al. 1988; H_S = 0.127 averaged for three mainland populations only, Plessas and Strauss 1986). Substantial genetic differentiation occurs among populations (G_{ST} = 0.130, Millar et al. 1988; G_{ST} = 0.162, Moran et al. 1988), primarily due to the differences between the island and mainland populations. The three mainland populations form a closely related group, whereas the two island populations are distinct from each other and from the mainland populations. However, the relationships between the mainland populations are uncertain because results from different studies are not consistent, and their affinities even differed when two different methods of cluster analyses were used for the same data set. The genetic

distances between *P. radiata* populations and *P. oocarpa*, a widespread Latin American species and putative progenitor of the CCCP, increases from south to north (Millar et al. 1988). The Cedros Island population is most divergent from the others and most resembles *P. oocarpa*.

Knobcone pine also possesses moderate diversities within populations and substantial differentiation among populations ($H_S = 0.131$ and $G_{ST} = 0.120$, Millar et al. 1988). A cline of allelic variation is apparent among populations of knobcone pine from Sierra Nevada through the Klamath region and into the north and south coast areas. The Sierran population is the closest to the progenitor, while the south coast population is the most derived and distinct from others. A close relationship is also found between geographic distances and genetic distances (Millar et al. 1988).

Bishop pine shows the largest genetic divergence among populations of the three CCCP species ($G_{ST} = 0.22$), though its intrapopulation genetic diversity is low ($H_S = 0.118$) (Millar et al. 1988). The large G_{ST} is mainly attributed to its extensive and disjunct geographic distribution, and lack of interfertility between northern and southern populations. Clinal divergence with respect to *P. oocarpa* is consistent from south to north, with the northern mainland populations being most divergent.

In sum, allozyme studies have revealed that considerable genetic diversity exists within populations of the CCCP, though less than that of more widespread conifers (Hamrick and Godt 1990). Population genetic differentiation, however, is often higher than that of other conifer species due to their disjunct population distribution. Phylogenetic analyses based on allozyme data indicate that populations are clearly clustered into three monophyletic species groups, approximately equally differentiated from each other.

Chloroplast DNA studies

Hong et al. (1993a; 1993b) comprehensively studied the nature and distribution of chloroplast DNA (cpDNA) genetic diversity and phylogenetic relationships through analysis of restriction fragment length polymorphisms (RFLPs) in the CCCP. Based on the analyses of restriction site mutations, a majority of genetic diversity is attributed to

the differences among species [G_{ST} = 0.84 (±0.13)]. Almost no genetic variation is detected within or among populations of knobcone pine and Monterey pine. Although little variation is observed within populations of bishop pine, population differentiation is substantial [G_{ST} = 0.88 (±0.08)], which is explained by three distinct regional groups (northern, intermediate and southern groups). Thus cpDNA RFLPs and allozymes revealed discordant patterns of genetic variation. However, population genetic diversity and differentiation based on RFLP length variants differed significantly from those based on site mutations. Bishop pine showed no diversity and Monterey pine had high intrapopulation gene diversity and low population differentiation, whereas knobcone pine had high differentiation but low diversity.

Hong et al. (1993b) included five taxa in their phylogenetic analyses: knobcone pine, Monterey pine, and the northern, intermediate and southern races of bishop pine. Wagner and Dollo parsimony analyses clearly distinguishes knobcone pine from Monterey and bishop pines, and the northern and intermediate races of bishop from the southern race. Interestingly, the southern race of bishop pine was found to be much closer to Monterey pine rather than to its conspecific intermediate and northern races. This is inconsistent with allozyme studies, but is supported by crossing studies (Critchfield 1967; Millar and Critchfield 1988) and some morphological and biochemical traits.

MITOCHONDRIAL DNA DIVERSITY

Genome size variation

The mitochondrial genome, one of the two main extrachromosomal genomes in plant cells, consists of a single closed, circular and double-stranded DNA molecule (mtDNA). The mtDNA genome of animals is relatively small (16-20 kb) (Avise et al. 1987). In land plants, however, mtDNA is much larger and highly variable (200 to 2500 kb). It is also several times larger and much more variable in size than chloroplast DNA (cpDNA, 120-217 kb), most of which are between 140-160 kb (Palmer 1990).

The variation in mtDNA genome size is not believed to be caused by large differences in the number of different genes, but rather by variation in the amount and

distribution of repeated DNA (Birky 1988; Palmer 1990). Changes in the complexity of intergenic spacer sequences play a major role in mtDNA size evolution (Gary 1989; Palmer 1985). Some of these spacer DNAs are derived from the chloroplast and the nucleus (Gary 1989; Lonsdale 1989). For example, maize mtDNA contains sequences homologous to at least six regions of cpDNA, while spinach has at least 13 (Birky 1988). However, sequences of cpDNA origin comprise less than 10% of the mtDNA genome and thus can not account for the large differences in genome size. Palmer (1990) suggested that nuclear DNA (nucDNA) was the source for the most of plant mtDNA and could account for the variable mtDNA size. In contrast, the size of animal mtDNA and cpDNA is very uniform. Two thirds of cpDNA size variation result from expansion or contraction of the large rRNA-encoding inverted repeat (IR) of 10-76 kb (Birky 1988). This IR, however, was lost or much reduced in conifer cpDNA (Strauss et al. 1988) and some legumes (Downie & Palmer 1992).

Gene content and function

Animal mtDNA lacks short dispersed repeats and thus is tightly packed and gene content is very conservative. It encodes a small and large rRNA, 22 tRNAs and 13 proteins. All of these proteins code for subunits of enzyme complexes that carry out electron transport and ATP synthesis in the main energy-generating pathway, oxidative phosphorylation of cells (Wallace, 1994). In contrast, plant mtDNA contains large and small dispersed repeats, and often large untranscribed and/or untranslated regions. Many foreign sequences from cpDNA or nucDNA are found in plant mtDNA. Only a small proportion of the genome is conserved and is termed 'core DNA'. The conserved sequence encodes approximately 5% of the proteins found in the mitochondrion; the rest are nuclear encoded and imported from the cytoplasm. Most (12-15) of the mtDNA-encoded proteins are used for subunits of respiratory chain enzyme complexes. Plant mtDNA also encodes rRNAs (26s, 18s, 5s), tRNAs and other proteins of unknown functions (Palmer 1992). Ninety-four possible protein-encoding genes have been found in the complete sequence of liverwort mtDNA (Oda et al. 1992).

Structural evolution

Unlike plant cpDNA and animal mtDNA that have compact gene arrangements and very conservative gene orders, plant mtDNA evolves rapidly in gene structure and almost every plant species has a unique gene order. Several factors may account for this phenomenon (reviewed in Palmer 1990, 1992b; Sederoff 1987). First, plant mtDNA contains large direct repeat elements that enable a high frequency of intragenomic and intergenomic recombination. This causes a highly variable genome conformation and composition. Almost all plant mtDNAs contain sets of direct repeats that mediate production of subgenomic circles via intramolecular and intermolecular recombination. Recombination between plasmids and other homologous sequences provides another source of variability. For example, recombination between maize mtDNA and free plasmids linearizes the majority of the normally circular genomes (Schardl et al. 1984). Second, plant mtDNA is characterized by an abundance of short dispersed repeats that can serve as sites of inversional recombination. This will accordingly promote a high rate of rearrangement in mtDNA and changes in gene order. Finally, large intergenic spacer sequences are present in mtDNA so that the high rate of inversions and structural rearrangements can be tolerated without causing gene mutations. In addition, transposition of sequences within and between molecules, and transfer of sequences from foreign genomes such as cpDNA and nucDNA sequences present in mtDNA genomes, are also responsible for the extensive variation in mtDNA genome structure (reviewed in Birky 1988).

Opportunities for rearrangement of cpDNA genomes are limited due to a paucity of dispersed repeats and short intergenic spacers. Furthermore, the large IR element of cpDNA appears to inhibit certain types of inversions (Palmer 1990; Strauss et al. 1988). Thus cpDNA has very slow structural evolution and strong constraints, mechanistic and/or selective, are proposed to maintain its compact and mainly genic genome.

Sequence evolution

In contrast to the rapid changes in its genome organization, plant mtDNA evolves very slowly in its primary sequence. Although the non-synonymous substitution rates are

similar in plant mtDNA and cpDNA, the rates of synonymous substitution are three-fold lower in plant mtDNA than in cpDNA (Wolfe et al. 1987). In addition, the substitution rates of nucDNAs of plants are at least two times higher than that in cpDNA and five times higher than that in mtDNA. However, the change of primary sequences is the fastest in animal mtDNA and its rate of synonymous substitution is 5-10 times higher than that of nucDNA. Given that most of the plant mtDNA genome is noncoding DNA, it is still a puzzle why sequence evolution is so slow (Palmer 1990).

Inheritance

Organelle genes are often inherited from one parent. Mitochondrial DNA is inherited primarily or exclusively from the maternal parent in animals (Avise 1991). Many studies have also revealed maternal inheritance in plant mtDNA (Neale and Sederoff 1989; Rajora et al. 1992; Wagner et al. 1991). However, some exceptions have been found in Cupressaceae and Taxodiaceae, where mtDNA is inherited paternally in *Sequoia sempervirens* (Neale et al. 1989) and *Calocedrus decurrens* (Neale et al. 1991).

In most angiosperms, cpDNA is inherited maternally (Schaal et al. 1991). However, cpDNA has shown predominant paternal inheritance in conifers studied to date, which includes members of the families Cupressaceae and Pinaceae (e.g., Boscherini et al. 1994; Neale and Sederoff 1988; Sutton et al. 1991b; Wagner et al. 1992). These findings are generally consistent with ultrastructural observations of organelle transmission (Owens and Morris 1990, 1991; Wagner et al. 1992).

MtDNA markers for population genetic studies

Restriction fragment length polymorphisms (RFLPs) are based on DNA fragment size variation due to the presence or absence of restriction sites, changes in site orientation, or insertions and deletions between sites. RFLPs have been particularly useful in providing genetic information for organellar genomes in both plants and animals due to their high copy number per cell and relatively simple polymorphisms compared to nucDNA (reviewed in Avise 1994; Dong and Wagner 1993, 1994; Mitton 1994; Strauss et al. 1993). In animals, mtDNA RFLPs have been used in a large number of studies

engaged in questions of mating systems, kinship, population structure, geographic variation, gene flow and speciation (reviewed by Avise 1994). Animal mtDNA is highly variable because it accumulates mutations 5-10 times faster than nuclear loci. For example, Avise et al. (1984) found 32 mtDNA haplotypes in a sample of 33 fishes called menhaden, *Brevoortia tyrannus* and *B. patrons*. The degree of population differentiation is also expected to be substantially higher in animal mtDNA than that in nucDNA because the effective population size of mtDNA loci is considered to be approximately one quarter that of nuclear loci due to the haploid genome and maternal inheritance of animal mtDNA. G_{ST} values for animal mtDNA have been observed to be 3.3 to 80-fold higher than those for allozymes (Crease et al. 1990; Davis 1986; DeSalle et al. 1987).

There have been few studies on genetic diversity and phylogeny using RFLP markers from mtDNA in plants (Deu et al. 1995; Dong and Wagner 1993, 1994; Strauss et al. 1993). In general, mtDNA have shown high levels of population differentiation in conifers. Strauss et al. (1993) detected nine mtDNA haplotypes by RFLPs with the *coxI* gene as a probe in a survey of 268 trees from 19 populations of three CCCP species. Average intrapopulation haplotype diversity was low ($H_S = 0.07$), which was less than for allozymes ($H_S = 0.15$, Millar et al. 1988) and for cpDNA length mutations ($H_S = 0.17$, Hong et al. 1993a), and similar to cpDNA site mutations ($H_S = 0.06$, Hong et al. 1993a). Population differentiation (G_{ST}) for mtDNA, which varies from 0.75 (*Pinus radiata*) to 0.96 (*P. muricata*), is significantly higher than for allozymes which varies from 0.12 (*P. attenuata*) to 0.22 (*P. muricata*). However, mtDNA G_{ST} values are comparable to those based on cpDNA for *P. muricata* (0.88, Hong et al. 1993a).

Two mtDNA *coxI*- and *coxII*-associated RFLPs were used to investigate the patterns of variation among 741 individuals from 16 allopatric populations throughout the ranges of jackpine (*P. banksiana* Lamb), and lodgepole pine (*P. contorta* Dougl.) (Dong and Wagner 1993). *CoxI*-associated restriction fragments revealed a diagnostic difference between two species but no variation within populations or species. In contrast, *coxII*-associated restriction fragments demonstrated a high level of variation within *P.contorta* species ($H_S = 0.68$), though only a small amount of variation within *P. banksiana* excluding the Saskatchewan population which is a putative hybrid between the two

species ($H_S = 0.03$). A majority of the variation within lodgepole pine resided among populations within species ($F_{ST} = 0.66$) and among subspecies ($F_{ST} = 0.31$).

Hong et al. (1995) analyzed genetic diversity and population differentiation of three genomes among 18 populations of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) in three geographic regions of British Columbia, Canada. CpDNA and mtDNA were studied by RFLP markers, while the nucDNA was analyzed by random amplified DNA markers (RAPDs). MtDNA revealed the highest level of population differentiation ($G_{ST} = 0.45$) and geographic regional differentiation ($G_{ST} = 0.29$), whereas cpDNA showed the highest total genetic diversity ($H_T = 0.77$) of the three genomes. However, only four trees were collected and analyzed for each population in their study.

Aagaard et al. (1995) identified a large number of mtDNA-derived fragments in RAPD profiles of Douglas-fir. Genetic differentiation between the coastal (var. *menziesii*) and interior (var. *glauca*) varieties based on frequencies of haplotypes formed from the combinations of those mtDNA RAPD fragments ($G_{ST} = 0.62$) were considerably higher than that obtained from allozyme frequencies ($G_{ST} = 0.26$).

Several mechanisms have been proposed to account for strong population subdivision in plant mtDNA genomes compared with nucDNA. First, maternally inherited mtDNA can only be dispersed through seeds, while nucDNA and paternally inherited organelle DNA can migrate via both seeds and pollen. Thus dispersal will generally be less and population subdivision will be higher for organelle genes with maternal inheritance than for nuclear genes and organelle genes with paternal inheritance (Birky 1988). Indeed, like animal mtDNAs and most plant mtDNAs, maternally inherited cpDNAs have also exhibited strong subdivision among populations (Mason-Gamer et al. 1995; Petit et al. 1993; Soltis et al. 1992). Second, the organelle genomes are haploid and uniparentally inherited, and thus their effective population size is approximately one fourth of that for nuclear genes in dioecious species and one-half in monoecious species such as pines. Low effective population size will increase the probability and rate of fixation of new variants, which will consequently accelerate population subdivision (Li and Graur 1991). Third, Birky et al. (1989) showed that population subdivision at equilibrium is inversely related to mutation rate. The rates of mutation are significantly

lower for organelle genes than for nuclear genes, potentially leading to greater differentiation than for nucDNAs. Finally, because organelle genomes do not undergo sexual recombination, they can be subject to an extreme form of hitchhiking called periodic selection. This results in a striking reduction of genetic diversity of a population following natural selection for a new mutation, and can amplify population subdivision when different variant genomes are selected in different populations (Maruyama and Birky 1991).

MtDNA markers for phylogenetic studies

Animal mtDNA RFLPs have been successfully employed for many phylogenetic analyses (reviewed by Avise 1987, 1994). However, RFLP markers from plant mtDNA have been only rarely used for such studies. Strauss et al. (1993) found that three CCCP species were all polyphyletic based on the phenogram produced from mtDNA RFLP analysis, a result that was strongly incompatible with morphology and allozymes. Hong et al. (1994) was unable to distinguish three races of Douglas-fir using mtDNA polymorphisms and all three regions were polyphyletic. However, *P. contorta* and *P. banksiana* did form their own species clusters on the basis of mtDNA data (Dong and Wagner 1993), though populations often did not cluster within subspecies.

Because RFLP polymorphisms of plant mtDNA are mostly length mutations and structural rearrangements, it is difficult to be confident of the evolutionary homology among different haplotypes (or fragments) due to the complex and overlapping nature of structural changes (Palmer 1992a). In addition, the recombinogenic sequences of plant mtDNA will often lead to homoplasic and convergent evolution. Thus it is not surprising that mtDNA restriction fragments often appear to cause misleading phylogenetic results in plants. MtDNA site mutations might be as useful as animal mtDNA or cpDNA for reconstruction of phylogenies; however, they are very infrequent and thus have not, to our knowledge, been employed.

RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD) POLYMORPHISM

Development and advantages of RAPD markers

Two related PCR-based molecular markers were developed almost simultaneously in 1990 for the detection of genetic polymorphisms, arbitrary-primer PCR (AP-PCR, Welsh and McClelland 1990) and random amplified polymorphic DNA (RAPD, Williams et al. 1990). Both techniques utilize a single oligonucleotide primer of arbitrary sequence to amplify DNA sequences between complementary inverted repeats within an amplifiable distance. The RAPD procedure uses shorter primers (usually 10 bp) and lower annealing temperatures (34-37^{0}C), while the AP-PCR procedure uses longer primers (20 bp or larger) and higher annealing temperatures (40-60^{0}C).

RAPD has gained widespread applications since its development, which could be attributed to its several apparent advantages: (1) no DNA sequence information is needed for primer design; (2) a far smaller amount of genomic DNA is required than for RFLPs; (3) no radioactivity is employed; (4) data can be produced quickly with much less complex and labor-intensive procedures than RFLP; and (5) a potentially unlimited number of RAPD markers can be generated. In addition, RAPDs may provide a more arbitrary and representative sample of the genome compared to allozymes. Thus RAPDs may provide a more accurate or complete measure of genetic variability in the genome.

Drawbacks, limitations and countermeasures

Despite their advantages over allozymes and/or RFLPs, RAPDs have some drawbacks and limitations. Foremost is the problem of dominant allelic expression. Heterozygotes are not usually distinguishable from dominant homozygotes and the accuracy of estimates of genetic diversity is thus reduced relative to estimates obtained from codominant systems. The problem of dominance may be somewhat overcome for population studies by using large sample sizes and many loci (discussed below). However, codominant markers can be identified among RAPD profiles, although they are relatively rare. With the help of Southern hybridization, Rieseberg et al. (1993) found about 7% of RAPD loci were codominant in their genome mapping study of *Helianthus*.

Roehrdanz et al. (1993) identified codominant markers in RAPD profiles by digesting a consistently amplified fragment with restriction enzymes having four-bp recognition sites. Davis et al. (1995) obtained 10 codominant RAPD markers by detecting the common occurrence of heteroduplex bands in individuals heterozygous for a codominant locus. However, these methods are either labor-intensive or difficult to utilize when the RAPD profiles are complex. Lack of homology between RAPD fragments with identical molecular weight is another potential problem, especially when comparing different species. The uses of RAPDs depend on the assumption that comigrating fragments are the indicators of homologous loci. However, fragments comigrating in two species may not represent homologous loci, especially on the basis of relatively low resolution agarose gels. The use of more accurate resolution methods (e.g., polyacrylamide gels) and silver staining could reduce the proportion of non-homologous comigrating fragments. However, most current RAPD studies have used agarose gel electrophoresis and ethidium bromide staining. Based on Southern hybridization, Williams et al. (1993) found nine of 10 comigrating fragments were homologous among several *Glycine* species. Thormann et al. (1994) revealed that three of 15 fragments, scored as identical, were not homlogous among six *Brassica* species, while all individuals showed the expected hybridization patterns at the intraspecific level. Van de Zande and Bijlsma (1995) found that the reliability with which fragment size could be equaled with homology was inversely related to genetic distance. More recently, a relatively comprehensive study has been made on the homology among RAPD fragments in three *Helianthus* species (Rieseberg 1996). As a result, 200 of 220 comigrating fragments were confirmed to be homologous using Southern hybridization and/or restriction digestion of gel-isolated fragments. However, 13% of these homologus loci were determined to represent paralogous loci rather than orthologous loci. Paralogous loci are non-identical, but derive from common genetic loci, and may thus bias estimates of phylogenetic relationships; only 174 (79.1%) of the identified loci appeared to be useful for comparative genetic analyses (Doyle 1992). It appears therefore important to test homology of RAPD markers for reliable genetic inferences, and is essential when RAPD are used for comparisons of species and genera.

The numbers, sizes and intensities of amplified fragments are very sensitive to experimental factors, causing low repeatability of RAPD reactions. For example, variation in DNA purity, the concentrations of template DNA, primer, dNTP, magnesium, DNA polymerase, and cycling parameters, and especially temperatures and ramp times, can all affect the RAPD banding patterns (Ellsworth et al. 1993; Roedhranz et al. 1993; Williams et al. 1993). In addition, different sources of Taq polymerase and thermocyclers can even cause variable results (Sheila Vollmer, personal communication). However, these problems can usually be mitigated by strict standardization of all reagents and protocols, and scoring only strong and clear fragments (e.g., Dowling et al. 1996; Haig et al. 1994).

Finally, non-Mendelian inheritance has been commonly observed for RAPDs. The proportion of Mendelian loci expressed as a fraction of the total number of amplification products can be as low as 32.7% (Bucci and Menozzi 1993), a possible result of amplification of many DNA fragments of organelle origin. Lorenz et al. (1994) found that many fragments in RAPD patterns of *Beta vulgaris L.* originated from mtDNA. In *Brassica,* results of Southern hybridization indicated that among the 142 scored fragments, seven were of mtitochondrial origin, and one was of chloroplast origin (Thormann et al. 1994). The most extreme case occurred in Douglas-fir, where 45% of scored RAPD markers showed maternal inheritance and were considered as mtDNA-derived fragments (Aagaard et al. 1995). The competition for amplification among different priming sites could be another factor contributing to non-Mendelian inheritance (Heun and Helentjaris 1993). Smith et al. (1994) revealed that some markers are amplified in one genetic background but not in another because of priming competition from unlinked sites. Segregation analysis may be essential for conclusively identifying nuclear genes with Mendelian behavior, but it is time-consuming. Southern hybridization using enriched organelle DNAs as probes can also be used to exclude RAPD fragments of organellar origin (Aagaard 1997).

Population genetic studies using RAPD markers

Despite their limitations, RAPD markers have been used in hundreds of studies in plants (for reviews see Weising et al. 1995; Whitkus et al. 1994; Williams et al.1993, Wolfe and Liston 1997). For example, they have been applied for identification of species, cultivars, and clones (Hsiao and Rieseberg 1995; Perron et al. 1995; Schnell et al. 1995); interspecific hybridization and introgression (Dean and Arnold 1996; Smith et al. 1996); genome mapping (Bradshaw et al. 1995, Rieseberg 1993; Tulsieram et al. 1992); conservation genetics (Haig et al. 1994; Virk et al. 1995); estimation of genetic diversity and population differentiation (Isabel et al. 1995; Szmidt et al. 1996); and estimation of phenetic and cladistic relationships (reviewed by Wolfe and Liston 1997).

RAPD markers have been increasingly applied for population genetic analysis, where comparisons are often made to allozyme and/or RFLP markers. RAPDs have generally shown greater genetic diversity and/or higher population differentiation than allozymes. Since RAPD markers are inherited as dominant markers and thus heterozygous individuals will be indistinguishable from dominant homozygous; estimates of allele frequencies for diploids can be made only from the frequencies of null homozygotes under the assumption of Hardy-Weinberg equilibrium. Several authors have instead used the Shannon's information index $H_0 = - p_i log_2 p_i$ (p_i: the RAPD product frequency) to calculate population genetic diversities, which are then partitioned to among- and within-population components (Chalmers et al. 1992; Dawson et al. 1995; Yeh et al. 1995). Subpopulation differntiation was found to be three-fold higher for RAPDs than for allozymes in a single population of *Gliricidia sepium* (Dawson et al. 1995). Chalmers et al. (1992) found that Shannon's indices were high both within and among populations (1.19 and 1.79, respectively) and G_{ST} was substantial (0.60). Baruffi et al. (1995) utilized a simple dissimilarity index $D = N_{AB}/(N_A+N_B)$ (N_A and N_B are the number of fragments in individuals A and B, respectively, and N_{AB} is the number of fragments that differ between the two individuals) to approximately represent heterozygosity according to the model of Gilbert et al. (1990) and Lynch (1990). As a result, RAPD markers revealed higher genetic variation than did allozymes, and improved discrimination within and

between populations and strains of the medfly (*Ceratitis capitata*), a polyphagous and multivoltine dipteran species.

Analysis of molecular variance (AMOVA), which was originally developed for analysis of mtDNA RFLP haplotypes by Excoffier et al. (1992), has recently been incorporated into data analysis of RAPD markers (Haig et al. 1994; Huff et al. 1993; Nesbitt et al. 1995; Peakall et al. 1995; Yeh et al. 1995). To analyze data with AMOVA, a phenotypic data file must first be provided in which each fragment is scored as 1 (the presence of a fragment) or 0 (the absence of a fragment) for all selected RAPD markers. Then a suitable distance matrix is chosen to calculate the differences between any two individuals. Finally a classical analysis of variance is employed to the resultant distance matrix for partitioning of RAPD variation into molecular variance components of different hierarchical levels. A nonparametric permutational procedure is used simultaneously to compute the significance of the variance components. In buffalograss (*Buckloë dactyloides*), AMOVA analysis of RAPD data revealed that population differentiation was substantial and significant in two geographic regions (28% and 20%, respectively) (Huff et al. 1993). In *Eucalyptus globulus*, among population variation accounted for between 6% and 26% of the total variation in different geographic regions (Nesbitt et al. 1995).

The several methods mentioned above are different from the approaches of genetic analyses for allozymes based on allele frequencies. As Dawson et al. (1995) discussed, it is difficult to make direct comparisions between allozymes and RAPDs due to the different approaches used for analysis. Fortunately, a few studies have been reported that compared allozymes and RAPDs directly using the same method of analysis. Peakall et al. (1995) transformed allozyme data into a distance matrix like RAPDs and then analyzed both markers using AMOVA. RAPD markers were found to provide higher population differentiation than allozymes (58.4% vs. 45.2%). In a few other studies, the estimations of population genetic parameters from different markers were all based on allele frequencies. Liu and Furnier (1993) compared allozyme, RFLP and RAPD markers for detecting genetic variation within and between trembling aspen (*Populus tremuloides*) and bigtooth aspen (*P. grandidentata*). RAPDs revealed greater genetic diversity than did

allozymes and RFLPs in both aspen species, with the difference being particularly substantial in bigtooth aspen. RAPDs also provided more polymorphic loci and species-specific loci than did allozymes. Isabel et al. (1995) estimated allele frequencies directly for both allozymes and RAPDs by using haploid megagametophytes. Averaged expected heterozygosities were very similar between two markers (0.30 for allozyme; 0.32 for RAPDs). However, population differentiation was more than three times higher for RAPDs than for allozymes, though F_{ST} was very low in both cases (0.027 for RAPDs; 0.008 for allozymes). Haploid tissues were also utilized for the comparison of RAPDs and allozymes in *Pinus sylvestris* (Szmidt et al. 1996). The "direct" allele frequencies were calculated for the two markers based on segregation patterns in haploid macrogametophytes. Diploid genotypes were then inferred from haplotypes and used to estimate "indirect" allele frequencies for RAPDs on the basis of number of null homozygotes. The objective was to determine how dominance of RAPDs might affect comparisons of genetic parameters between the two makers where diploid tissues were used. The results of the analysis indicated that genetic diversities within populations were higher for direct RAPDs than for allozymes, and lower for indirect RAPDs than for allozymes. Population differentiation, however, was very similar for allozymes and direct RAPDs, but it was three-fold greater for indirect RAPDs than for direct RAPDs. They concluded that indirect estimation of allele frequencies might bias the estimation of population genetic parameters. However, both their population sample sizes and locus number were small (20 maternal trees per population, 22 direct RAPD loci, and 12 indirect loci).

Lynch and Milligan (1994) suggested excluding RAPD markers with a frequency of null homozygotes of less than 3/N (N: sample size of a population) in any population from quantitative analysis because their low frequency can result in errors in diversity estimates. However, this would result in exclusion of many monomorphic and highly differentiated loci (e.g., population specific markers), causing estimates of intrapopulation diversity to be biased upward, and interpopulation diversity to be biased downward. For example, according to the method of Lynch and Milligan (1994), Szmidt et al. (1996)

excluded 10 loci from their total of 22 loci and found that the expected heterozygosity within populations was substantially higher than that estimated from their total dataset.

Although RAPDs have shown higher genetic variation than allozymes in many studies, some studies have reported little difference between these two markers. Yeh et al. (1995) found that RAPD variation in *Populus tremuloides* Michx. largely resided within populations rather than among populations (97.4% vs. 2.6%). A similar low level of among population variation had also been observed for allozyme loci (Jelinski and Cheliak 1992). Mosseler et al. (1992) detected low level of diversity and differentiation for RAPDs as had been observed for allozymes in red pine (*Pinus resinosa* Ait.). Haig et al. (1994) obtained similar levels of population differentiation for allozyme and RAPD markers in red-cockaded woodpeckers (*Picoides borealis*).

Phylogenetic studies using RAPD markers

RAPD markers have been utilized both in phenetic and cladistic studies of phylogenies (reviewed in Wolfe and Liston 1997). Phenetic methods such as UPGMA and neighbor-joining (NJ) are the preferred tools for analyzing RAPD phylogenetic data. These methods, also known as distance-matrix methods, construct phenetic trees based on similarities or distances between different operational taxonomic units (OTUs) without considering the evolutionary pathways of specific characters. In the majority of RAPD studies, phenotypic binary matrices are transformed to similarity (or distance) measures by the use of the coefficients of Jaccard (e.g., Buren et al. 1994), Dice (e.g., Marillia et al. 1996), or Nei and Li (1985) (e.g., Apostol et al. 1993). Several studies have also used simple matching coefficients, which include negative matches, as similarity measures (Apostol et al. 1993; Baruffi et al. 1995; Lifante and Aguinagalde 1996). However, the shared absence of a RAPD fragment might arise from many causes such as point mutations at primer annealing sites, inversions flanking the annealing sites, secondary structure constraints, competition with other loci for amplification, and insertions increasing distances to a greater length than can be amplified with routine PCR techniques. Because of this heterogeneity in the null phenotypic class, a simple matching

coefficient is not recommended for RAPD phylogenetic analysis, especially when comparing different species (Apostol et al. 1993).

There has been general concordance between the results derived from RAPDs and from other techniques using phenetic analyses (reviewed by Wolfe and Liston 1997). For example, Lifante and Aguinagalde (1996) obtained similar dendrograms for RAPDs and allozymes with 15 populations clustered in the same three *Asphodelus* species. The results obtained from 372 RAPD markers for 13 *Rubus* species were generally in agreement with previous classifications (Graham et al. 1995). Marillia and Scoles (1996) investigated the phylogenetic relationships among *Hordem* species using both Jaccard and Dice coefficients of similarity. The phenograms generated were identical and consistent with those based on morphological characters. However, Thormann and Osborn (1992) found phenogram differences between RAPD and RFLP markers. It was speculated that the absence of homology between shared RAPD fragments could explain this discrepancy.

There are several disadvantages, however, of using distance matrices for phylogenetic analyses: (1) information is lost when character data are transformed to pairwise distances; (2) it may be difficult to combine different types of data into the same analysis because different methods may be used to calculate the distances for various data, thus it might be difficult to get a generalized distance from various data types for phylogenetic references; and (3) the most highly informative characters cannot be directly identified (Swofford et al. 1996).

Cladistic analysis is the other main approach to phylogenetic inferences. It is used to study evolutionary pathways by analyzing the gains and losses of individual characters. Among the cladistic approaches that infer phylogenies directly from character data, methods based on the principle of maximum parsimony have been the mostly widely used (Swofford et al. 1996). However, they are rarely employed for RAPD analysis because they require that RAPD fragments are homologous and independent (Swofford and Olsen 1990), assumptions that are often violated (e.g., Rieseberg 1996; Smith et al. 1994; Van de Zande and Bijlsma 1995). Even in a study at the intraspecific level, Van Heusden and Bachmann (1992a, b, c) estimated that up to 10% of the RAPD markers might violate the

assumptions of independence and size homology. In addition to the problems of homology and independence, Backeljau et al. (1995) have explained three other conceptual factors which could limit the reliability of RAPDs in parsimony analyses: (1) a suitable model for RAPD character state is not available among currently utilized parsimony methods; (2) the diallelic interpretation of RAPD loci is equivalent to the independent allele model for allozymes, which is invalid because alleles are not independent. The alternative is to use the locus as a character with the different allelic combinations as character states. However, this approach is not applicable to RAPDs because of dominance; and (3) RAPD primers with high GC contents tend to amplify GC rich regions which often contain rapidly mutating sites (e.g., "hotspots", CpG islands). This biased priming in rapidly mutating regions may result in more frequent homoplasy and multiallelism.

CHAPTER 2

ABUNDANT MITOCHONDRIAL GENOME DIVERSITY, POPULATION DIFFERENTIATION, AND EVOLUTION IN PINES

ABSTRACT

Because of its abundant repetitive DNA and frequent genome rearrangements, the plant mitochondrial genome has rarely been the subject of microevolutionary and phylogenetic studies. We have examined mitochondrial DNA polymorphisms via the analysis of restriction fragment length polymorphisms in three closely related species of pines from western North America: knobcone pine (*Pinus attenuata* Lemm.), Monterey pine (*P. radiata* D. Don) and bishop pine (*P. muricata* D. Don). The species, and their many disjunct populations and varieties, represent a wide range of the early stages of phyletic differentiation. A total of 343 trees derived from 13 populations were analyzed using 13 homologous mitochondrial gene probes amplified from three species by polymerase chain reaction. Twenty-eight distinct mtDNA haplotypes were detected and no common haplotypes were found among the species. All three species showed limited variability within populations, but strong differentiation among populations. Based on haplotype frequencies, genetic diversity within populations (H_S) averaged 0.22, and population differentiation (G_{ST} and θ) exceeded 0.78. Analysis of molecular variance (AMOVA) also revealed that more than 90% of the variation resided among populations. For the purposes of genetic conservation and breeding programs, species and populations could be readily distinguished by unique haplotypes, often using the combination of only a few probes. Neighbor-Joining phenograms, however, strongly disagreed with those based on allozymes, chloroplast DNA and morphological traits. Thus, despite its diagnostic haplotypes, the genome appears to evolve via rearrangement of multiple, convergent subgenomic domains.

INTRODUCTION

Plant organelle genomes have been increasingly applied to study population genetic structure and phylogenetic relationships in plants (Dumolin *et al.* 1997; see Hipkins *et al.* 1994; Olmstead and Palmer 1994 for reviews). The use of molecular markers derived from different genomes provides a more complete description of population structure (*e.g.,* Dong and Wagner 1994; Hong *et al.* 1993a; McCauley *et al.* 1996), and thus aids in identification of species, races, and populations in breeding (*e.g.,* Grabau *et al.* 1992) and conservation (*e.g.,* Furman *et al.* 1996) programs. For species such as woody perennials, nuclear genes (*e.g.,* allozymes) provide little power for discrimination among populations because the large majority of their diversity resides within populations (Brown and Schoen 1992; Hamrick and Godt 1990). In contrast, cytoplasmic organelle genomes (chloroplasts and mitochondria) are often strongly differentiated among populations. This difference may be due to a low rate of sequence mutation, small effective population size, and limited gene flow for maternally inherited organelles (Birky 1988; Dong and Wagner 1994). Plant mitochondrial DNA (mtDNA) has a very low rate of gene sequence evolution (Wolfe *et al.* 1987), suggesting a much lower rate of point mutation in plant mtDNA than in chloroplast DNA (cpDNA) and animal mtDNA (Palmer 1992a; Sederoff 1987). However, it is extremely variable in size and gene arrangement (Palmer 1992a; Pring and Lonsdale 1985), and shows maternal inheritance in pines (Neale and Sederoff 1989) and most other plant species. Taken together, these distinctive features of plant mtDNA make it a potentially powerful tool for the analysis of population differentiation. The few studies of mtDNA polymorphism in plants have revealed high variability within and/or among populations (Belhassen *et al.* 1993; Dong and Wagner 1993; Luo *et al.* 1995), including a previous study of the California Closed-Cone Pines (CCCP) using a single mtDNA gene probe (Strauss *et al.* 1993).

The aim of this study was to intensively assess the level and distribution of mtDNA genetic diversity in the CCCP via sampling of a number of regions of the genome. The CCCP contains three closely related species, and includes several disjunct populations and distinctive taxonomic varieties, thus providing samples of several early stages of speciation. It is comprised of one interior species, *Pinus attenuata* (knobcone pine), and

two maritime species, *P. muricata* (bishop pine) and *P. radiata* (Monterey pine). Knobcone pine grows on interior sites of southern Oregon and California as disjunct populations. The two other species are distributed discontinuously along the California coast and on four islands (Figure 2.1) (Critchfield and Little 1966). Many characteristics of these species have been studied in previous population genetic analyses, including morphology, secondary compound chemistry, allozyme and chloroplast DNA (cpDNA) (Hong *et al.* 1993a, b; reviewed in Millar 1986; Millar *et al.* 1988;), providing reference points for comparison. The specific objectives of this study were to: (1) document patterns of mtDNA diversity within and among populations; (2) compare these patterns with those of other genetic markers; (3) evaluate the phylogenetic value of mtDNA genome polymorphisms; and (4) assess the capability of mtDNA to quantitatively differentiate species and populations, and thus assist in germplasm identification for conservation and breeding programs.

MATERIALS AND METHODS

Plant materials

Trees were sampled from natural populations or from gene conservation and genetic test plantations as described in Hong *et al.* (1993a). Two different collections contributed to this study. The Año Nuevo, Cambria and Guadalupe populations of Monterey pine; the Sierra Nevada and Santa Ana populations of knobcone pine; and the Santa Cruz population of bishop pine were primarily collected by Hong *et al.* (1993a). The other populations were collected specifically for this study. For knobcone pine, the Klamath population was sampled over a 6.0-kilometer (km) transect adjacent to the Lakehead Exit on U.S. Interstate 5, California (latitude 40°55', longitude 122°30'), and the Oakland population was sampled over a 2.6-km transect along Flicker Ridge adjacent to the town of Moraga in the hills east of Oakland, California (latitude 37°50', longitude 122°30'). For bishop pine, the San Vicente population was sampled in several small scattered populations along a road north of San Vicente that goes out to the town of Erendira, Mexico (latitude 31°15', longitude 116°30'); the Monterey population was sampled over

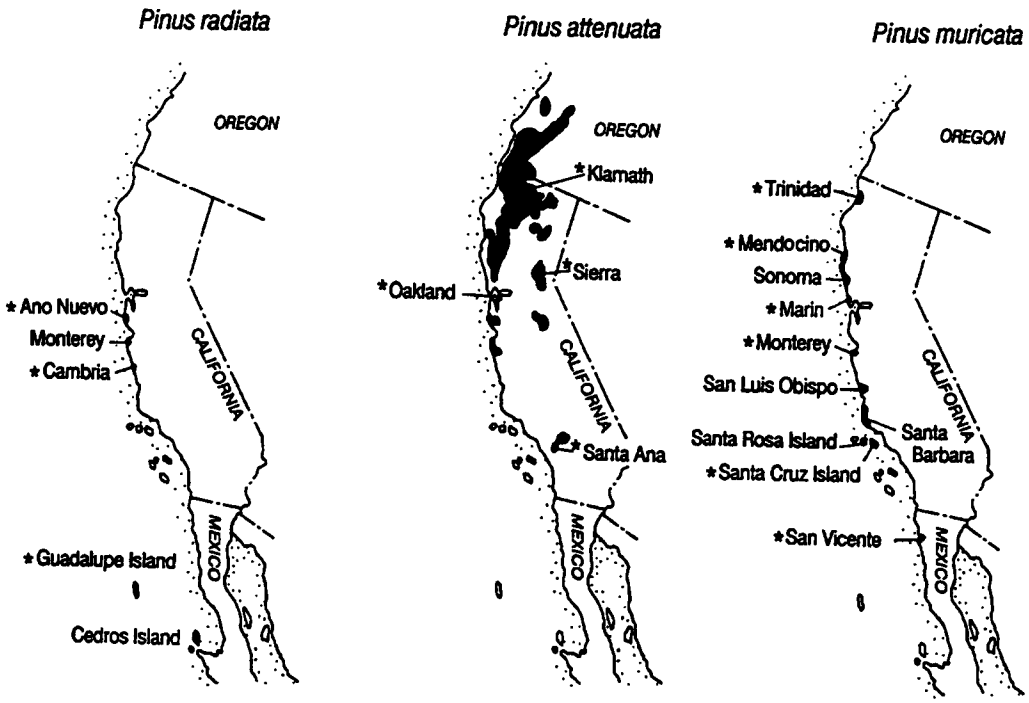


Figure 2.1 Distribution of *Pinus radiata, P. attenuata, and P. muricata* (Hong et al. 1993a) and the origins of sampled populations (*). All populations were studied for chloroplast DNA polymorphism by Hong et al. (1993a).

one linear mile in the woods of Samuel F. B. Morse Botanical Reserve located south of Monterey, California (latitude 36°40', longitude 121°50'); the Marin population was sampled over a 1.7-km transect about 5.1-km southwest from the town of Inverness, California (latitude 38°08', longitude 122°45'); the Mendocino population was sampled for 2.6-km on both sides of U.S. Highway 101, 7.5-km south of the town of Point Arena, California (latitude 39°20', longitude 123°50'); and the Trinidad population was sampled over 1.7 linear kilometers up Fox Farm Road adjacent to the town of Trinidad, California (latitude 41°05', longitude 124°10').

Probe preparation and universal mtDNA primers

A total of 13 different probes were used in the restriction fragment length polymorphism (RFLP) analysis. Ten probes were specific for different single mtDNA genes: *atp1, atp6, cob, cox1, cox2* (exon 1), *cox3, nad1* (exon 1), *nad3, nad4* (exons 1 and 2 including intron 1), and *rps14*. Two probes were specific to different parts of *nad5*: one probe (*nad5a*) hybridized to the exons 1 and 2 including intron 1, and the other (*nad5d*) to exons 4 and 5 including intron 4. One probe hybridized to the intergenic region between *nad3* and *rps12*.

Probes were amplified using universal mtDNA specific primers (Table 2.1) via the polymerase chain reaction (PCR). To design universal mtDNA specific primers, we retrieved and aligned as many genes of fungal, algal and higher plant mtDNA sequences as were available from international DNA sequence data bases, including GenBank, EMBL, DDBJ, and others. Among plant species, we used monocots, dicots, and gymnosperms when available. GeneRunner (Hastings Software, Inc., 1994, v. 3.04) was used for multiple alignment, oligonucleotide analysis, and primer design. Our primary criteria for choosing primer sites and sequences were: (1) high conservation of amino acid sequences across all available organisms; (2) exact or nearly exact matches of DNA sequences across seed plants; (3) sites of likely *C* to *T* editing were avoided when possible; (4) the primer's *3'*-end had nearly perfect matches for the last 7-8 nucleotides, and in no cases had mismatches for the last 4-5 nucleotides; (5) amino acids with highly degenerate codons were avoided, and those encoding unique and low degeneracy codons

were preferred; (6) internal repeats, hairpins, internal loops, and dimers were avoided; (7) primers had a relatively high T_m, usually not less than 55°C, with a high G-C ratio; and (8) they had no significant homologies to cpDNA sequences based on database searches. For cpDNA homology searches, we used all published cpDNA genome sequences, including liverwort (*Marchantia polymorpha*), maize (*Zea mays*), rice (*Oryza sativa*), tobacco (*Nicotiana tabacum*), and black pine (*Pinus thunbergii*), employing the Organelle Genome Database (GOBASE) extensively for searches: http://megasun.bch.umontreal.ca/gobase/content.html.

Primers were synthesized in Central Service Laboratory of OSU Gene Research and Biotechnology Center using ABI 380B or 394 DNA synthesizers (Perkin-Elmer Applied Biosystems Division, Foster City, CA, USA). To test synthesized primers we used DNA samples from a large variety of plant species, and from enriched cpDNA and mtDNA samples provided by V. Hipkins and J. Aagaard (Aagaard et al. 1998), respectively. PCR products amplified from CCCP DNA samples were recovered from a 2% agarose gel under long-wave UV light, purified using a QIAquick gel extraction kit (QIAGEN Inc., Chatsworth, California) or a GENECLEAN kit (BIO101 Inc., La Jolla, California), and radioactively labeled with [^{32}P] by primer extension using a random hexamer labeling kit (Boehringer Mannheim GmbH, Germany).

RFLP procedures

Total genomic DNA was extracted from needles using a CTAB-based DNA extraction protocol (Wagner *et al.* 1987), followed by three phenol/chloroform purifications and a final ethanol precipitation. Modifications of this protocol, and procedures for restriction enzyme digestion, agarose electrophoresis, Southern blotting, hybridization, and washing and stripping of blots were as described in Strauss and Doerksen (1990) and Hong (1991). However, we added three final high stringency washes (0.1x SSC and 1% SDS solutions at 65°).

Table 2.1 Nucleotide sequence, name, melting temperature (T_m), GC content, and expected size of amplified PCR products for universal primers used to amplify mitochondrial genes

Gene [a] and probe	End	Name	Sequence	*Size, bp*	*G+C, %*	T_m, C°	PCR product, bp
atp1	*5'*	atpain51	TTTGCCAGCGGTGT(G=I[b])AAAGG [c]	20	55	67	1039
	3'	atpain32	CTTCGCGATATTGTGCCAATTC	22	46	70	
atp6	*5'*	atp6in51	GGAGG(A=I)GGAAA(C=I)TCAGT(A=I)CCAA	22	48	60	604
	3'	atp6in31	TAGCATCATTCAAGTAAATACA	22	27	56	
cob	*5'*	cob-in52	AGTTATTGGTGGGGGTTCGG	20	55	68	350
	3'	cob-in33	CCCCAAAAGCTCATCTGACCCC	22	59	74	
cox1 [d]	*5'*	cox1in51	GGTGCCATTGC(T=I)GGAGTGATGG [c]	22	59	70	1485
	3'	cox1in32	TGGAAGTTCTTCAAAAGTATG [c]	21	33	57	
	5'	cox1in53	GGCT(G=I)TTCTCCAC(T=I)AACCACAA	22	50	66	1507
	3'	cox1in33	GGAGGGCTTTGTACCA(A=I)CCATTC	23	52	69	

Table 2.1 (continued)

cox2	*5'*	cox2in51	GATGC(A=I)GC(G=I)GAACC(A=I)TGGCA [c]	20	55	58	340
(exon 1)	*3'*	cox2in32	TCCGATACCATTGATGTCC	19	47	53	
cox3	*5'*	cox3in51	GTAGATCCAAGTCCATGGCCT [c,e]	21	52	65	692
	3'	cox3in31	GCAGCTGCTTCAAAGCC [c]	17	59	61	
nad1	*5'*	nad1in52	CTAGCTGAACGTAAAGTAATGGC	23	44	64	306
(exon 1)	*3'*	nad1in32	CCAACC(T=I)GCTATAAT(A=I)ATTCC	21	38	54	
nad3	*5'*	nad3in51	AATTGTCGGCCTACGAATGTG [c]	21	48	67	215
	3'	nad3in31	TTCATAGAGAAATCCAATCGT	21	33	58	
nad3-	*5'*	nad3in51	AATTGTCGGCCTACGAATGTG [c]	21	48	67	~370
rps12	*3'*	rps12o51	GCTCG(A=I)GTACGGTC(C=I)GTGCG	20	65	62	
nad4a	*5'*	nad4ai52	ATACGATTGATTGGTCTGTG (exon 1)	20	40	57	~1500
(intron 1)	*3'*	nad4ai32	TGAACTGGTACCATAGGCACTTT (exon 2)	23	46	64	-2000

Table 2.1 (continued)

nad5a	*5'*	nad5in51	GAAATGTTTGATGCTTCTTGGG (exon 1)	22	41	66	~1000
(intron 1)	*3'*	nad5in31	ACCAACATTGGCATAAAAAAAGT (exon 2)	23	30	64	
nad5d	*5'*	nad5in54	ATAAGTCAACTTCAAAGTGGA (exon 4)	21	33	56	~1000
(intron 4)	*3'*	nad5in34	CATTGCAAAGGCATAATGAT [f] (exon 5)	20	35	61	
rps14	*5'*	rps14i51	ATACGAGATCACAAACGTAGA [c]	21	38	56	282
	3'	rps14i31	CCAAGACGATTT(C=I)TTTATGCC	21	38	61	

[a] *atp1* (or *atpA*) = F_1-adenosine triphosphatase (ATPase) subunit 1 (alpha) gene; *atp6* (or *atpF*) = F_0-ATPase subunit 6 gene; *cob*: apocytochrome b gene; *cox1* (or *coxI*) = cytochrome c oxidase subunit 1 gene; *cox2* (or *coxII*) = cytochrome c oxidase subunit 2 gene; *cox3* (or *coxIII*) = cytochrome c oxidase subunit 3 gene; *nad1* (or *nadA*, or *ndhA*, or *ndh1*, or *nd1*) = NADH-ubiquinone oxidoreductase subunit 1 gene; *nad3* (or *nadC*, or *ndhC*, or *ndh3*, or *nd3*) = NADH-ubiquinone oxidoreductase subunit 3 gene; *nad4* (or *nadD*, or *ndhD*, or *ndh4*, or *nd4*) = NADH-ubiquinone oxidoreductase subunit 4 gene; *nad5* (or *nadF*, or *ndhF*, or *ndh5*, or *nd5*) = NADH-ubiquinone oxidoreductase subunit 5 gene; *rps12* = ribosomal protein subunit 12 gene; *rps14* = ribosomal protein subunit 14 gene.

[b] Inosine was used in the synthesis instead of the corresponding nucleotide because of nucleotide variation observed among plant sequences.

[c] unstable can form hairpins.

[d] Any of the four possible combinations - cox1in51/cox1in32, cox1in51/cox1in33, cox1in53/cox1in32, or cox1in53/cox1in33 - can be used to produce practically identical *cox1*-specific probes, but cox1in51/cox1in32 combination was used to obtain data.

[e] Overlaps with primers designed by Hiesel *et al.* 1994.

[f] "Hot" PCR start is recommended because of a possibly stable hairpin.

Preliminary detection of polymorphisms

For a preliminary survey, two trees were randomly chosen from each of the 13 populations under study. Thirteen mtDNA probes and two restriction enzymes (*Bam*HI and *Xba*I) previously identified as showing high polymorphism (Strauss *et al.* 1993) were employed for the detection of mtDNA polymorphisms. Only those probe-enzyme combinations that detected polymorphisms either within or between species in this preliminary survey, and did not give information redundant with other probes (see below), were retained for the full analysis of all 343 sampled trees (*atp6*, *cox1*, *cox2*, *nad3*, *nad4*, *nad5a*, *nad5d*, and *rps14*).

Data analysis

Haplotype analysis: Haplotypes were determined based on having unique restriction fragment patterns over the various combinations of restriction enzymes and probes. Haplotype frequency (where haplotypes are treated as alleles at a single genetic locus) in each population was used to estimate genetic diversity and population differentiation. Genetic diversity and Nei's (1986) G_{ST} adjusted for sample size and population number were calculated using the GeneStat-PC 3.3 program (Lewis 1994). Weir and Cockerham's (1984) theta (θ) value for population subdivision, and standard deviation derived by jackknifing over populations, were calculated from individual haplotypes using the Genetic Data Analysis (GDA) program (Lewis and Zaykin 1996).

Probe-enzyme-based multi-locus analysis: To better understand the nature of diversity in different parts of the mtDNA genome, we analyzed the data where each probe-enzyme combination was considered as a genetic locus and each restriction fragment profile variant was considered as an allele. Allele frequencies at each locus in each population were then employed to estimate the genetic diversity parameters and Nei's (1986) G_{ST} using the GeneStat-PC 3.3 program. To give a more accurate estimate of gene diversity than the inflated value that would be obtained if only polymorphic combinations were used (see discussion), probe-enzyme combinations monomorphic in the preliminary sample were assumed to be monomorphic in all trees.

Analysis of Molecular Variance (AMOVA): AMOVA was used to partition molecular variance into different hierarchical levels (Excoffier *et al.* 1992). Each tree was scored by a vector of 1's (presence of a band) and 0's (absence of a band) representing the components of their multibanded RFLP phenotype. The proportion of shared fragments was calculated for each possible pairwise comparison according to the formula proposed by Nei and Li (1985): S (similarity) = $2N_{AB}/(N_A+N_B)$, where N_{AB} is the number of bands shared by individuals A and B; N_A and N_B are the number of bands in individuals A and B, respectively. The distance index was: D = 1-S. All similarity and distance indices were obtained with the RAPDPLOT program (Black 1996). An analysis of molecular variance was then performed on the resultant matrix for partitioning the total RFLP variation into the within- and among-group variance components, and producing Φ-statistics analogous to F-statistics. The significance of the variance component was computed using a non-parametric permutation test.

Phylogenetic analysis

To incorporate the phenotypic similarities of different haplotypes into genetic distances, pairwise Manhattan Distances (Wright 1978) between populations were computed on the basis of the fragment phenotype vector described above. A distance matrix, or a set of matrices via bootstrapping, was generated using the RAPDDIST program (Black 1996). The matrices were then subjected to the NJTREE and CONSENSE programs in the PHYLIP package (Felsenstein 1995) to produce a Neighbor-Joining consensus tree indicating the phylogenetic relationships between populations and species.

RESULTS

Diversity

Of the 13 probes tested, three probes (*atp1*, *cob*, and *cox3*) failed to reveal any polymorphism regardless of the restriction enzymes used. Four probes (*nad3*, *nad3-rps12*, *nad4*, and *nad5a*) detected polymorphism with only *Xba*I, whereas another three probes (*atp6*, *nad1* and *rps14*) exhibited polymorphism with only *Bam*HI. The other three probes

(*cox1, cox2*, and *nad5d*) showed polymorphism with both enzymes (*e.g.*, Figure 2.2). However, two pairs of probes (*nad3* and *nad3-rps12*, and *nad1* and *rps14*) detected identical mtDNA-RFLP patterns for the screened individuals. Thus only one gene from each pair (*e.g.*, *nad3* and *rps14*) was used in the full analysis. In sum, 22 probe-enzyme combinations were used to generate our data.

Restriction fragment and haplotype polymorphisms

The 22 probe-enzyme combinations produced a total of 76 scored fragments (Appendix A). The number of fragments produced per combination varied from one to five for each haplotype, suggesting that several genes have multiple copies in many of the populations. The multiple fragments usually had identical relative hybridization intensities among individuals, which would result from simple duplications and deletions. A total of 28 haplotypes were identified based on the RFLP patterns of all probe-enzyme combinations. The number of fragments for each haplotype ranged from 25 to 38 out of a total of 76 fragments (Appendix A). Only 13 (22%) of the fragments were present among all haplotypes. Sixteen (27%) of the fragments were unique to a single haplotype, and 60 fragments (73%) were shared by two or more haplotypes. There were 6 haplotypes for Monterey pine, 11 for knobcone pine, and 11 for bishop pine. There was no haplotype common among any of the populations from the different species (Table 2.2).

Genetic diversity within populations

Eight population samples had two or more haplotypes and five samples had only a single haplotype. The frequency of common haplotypes in the polymorphic population samples varied from 52% to 96% (Table 2.2). The Oakland population of knobcone pine and the Santa Cruz population of bishop pine each contained five different haplotypes, while four haplotypes each were detected for the Cambria population of Monterey pine and the Sierra Nevada population of bishop pine.

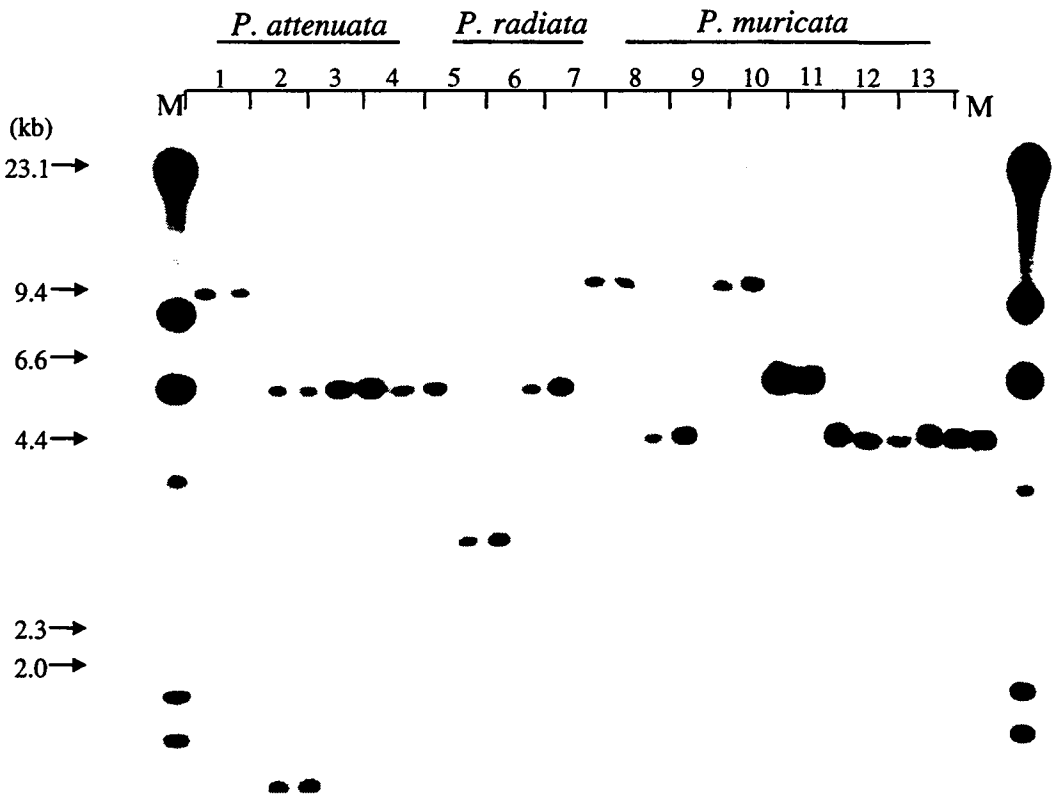


Figure 2.2 An example of autoradiograms showing interspecific and interpopulation mtDNA diversity. DNA was digested with *Bam*HI and probed with the *cox2* gene. M: DNA size markers (Lamda/*Hin*dIII digest DNA). Numbers 1-4: *P. attenuata* (1: Santa Ana; 2: Sierra Nevada; 3: Oakland; 4: Klamath); Numbers 5-7: *P. radiata* (5: Año Nuevo, 6: Cambria; 7: Guadalupe). Numbers 8-13: *P. muricata* (8: Mendocino; 9: San Vicente; 10: Santa Cruz; 11: Marin; 12: Monterey; 13: Trinidad).

Table 2.2 Haplotype frequencies in sampled populations and number of trees assayed (N)

Species	Population	Haplotype frequencies	N
P. radiata	Año Nuevo	a: 1.0	26
	Cambria	a: 0.04, b: 0.64, c: 0.08, d: 0.24	25
	Guadalupe	e: 0.96, f: 0.04	25
P. attenuata	Klamath	g: 1.0	25
	Sierra Nevada	h: 0.05, i: 0.85, j: 0.05, k: 0.05	22
	Oakland	l: 0.54, m: 0.04, n: 0.04, o: 0.34, p: 0.04	26
	Santa Ana	q: 1.0	23
P. muricata	Trinidad	r: 1.0	25
	Mendocino	s: 0.27, t: 0.73	29
	Marin	t: 0.48, u: 0.52	25
	Monterey	r: 0.92, v: 0.08	25
	Santa Cruz	w: 0.04, x: 0.04, y: 0.04, z: 0.04 aa: 0.84	24
	San Vicente	bb: 1.0	43

On average, gene diversity within populations based on haplotype frequencies was 0.22 (Table 2.3), ranging from 0.21 to 0.23 between species. As expected, diversity was substantially lower (0.03) (Table 2.4) when each probe-enzyme combination was considered as a genetic locus and each fragment profile variant was designated as an allele; averaged over populations, the number of effective alleles per locus only slightly higher than one. The percentage of polymorphic loci ranged from a relatively high value of 22.7% in *Pinus attenuata* to a low of 11.4% in *P. muricata*.

Levels of diversity differed greatly for some populations, depending on whether multi-locus or haplotype analysis was used. For example, the Cambria and Mendocino populations each had two main haplotypes occurring in roughly equal frequencies, but the haplotypes differed due to fragment changes at four loci in the Cambria population and at only one locus in the Mendocino population (Appendix A and Table 2.2). As a result, the haplotype diversity of the Mendocino population was similar to that of the Cambria population (0.41 vs. 0.54), but its multi-locus diversity was much lower (0.02 vs. 0.08)

(data not shown). AMOVA analysis showed low molecular variance within populations (0.017 on average, Table 2.5), which was comparable with our estimate of multi-locus diversity ($H_S = 0.03$).

Table 2.3 Estimates of population subdivision based on haplotype frequencies

Level of analysis	H_S [a]	D_{ST} [b]	G_{ST} [c]	θ [d]
Pooled populations	0.22 (0.06)	0.77	0.78	0.78 (0.07)
Species	0.78 (0.03)	0.22	0.22	0.21 (0.02)
Populations in species				
P. radiata	0.21 (0.17)	0.78	0.79	0.79 (0.14)
P. attenuata	0.21 (0.14)	0.79	0.79	0.78 (0.13)
P. muricata	0.23 (0.09)	0.69	0.75	0.77 (0.06)
Regional [e]	0.60 (0.05)	0.30	0.34	0.35 (0.01)

[a] H_S = Haplotype diversity within populations; standard errors in parentheses.
[b] D_{ST} = Haplotype diversity among populations (H_T - H_S).
[c] G_{ST} = Nei's (1986) G_{ST} unbiased for sample size and population number.
[d] θ = Subdivision estimate of Weir and Cockerham (1984) with jackknife-derived standard deviation over populations.
[e] Regional analyses investigate the diversity among three groups of populations within *P. muricata*: Trinidad and Mendocino (Northern); Marin and Monterey (Intermediate); and Santa Cruz and San Vicente (South).

Table 2.4 Population genetic statistics based on 22 individual probe-enzyme combinations

Level of analysis	A_{EP} [a]	P_{99} [b] (%)	H_S [c]	D_{ST} [d]	G_{ST} [e]
Pooled populations	1.04 (0.02)	16.78 (6.05)	0.03 (0.01)	0.31	0.91
Species	1.64 (0.05)	50.00 (0.00)	0.25 (0.06)	0.11	0.31
Populations in species:					
P. radiata	1.04 (0.04)	19.70 (15.38)	0.03 (0.01)	0.37	0.93
P. attenuata	1.06 (0.05)	22.73 (13.25)	0.04 (0.01)	0.25	0.87
P. muricata	1.04 (0.02)	11.36 (7.85)	0.02 (0.01)	0.23	0.91
Regional[f]	1.27 (0.18)	22.73 (13.89)	0.09 (0.03)	0.17	0.65

[a] A_{EP} = effective number of alleles per locus. Each restriction fragment profile variant was counted as an allele and each probe-enzyme combination as a genetic locus; standard errors in parentheses.
[b] P_{99} = percentage of loci polymorphic, where the frequency of the most common allele was less than 0.99.
[c] H_S = unbiased average gene diversity within populations.
[d] D_{ST} = unbiased gene diversity among populations.
[e] G_{ST} = Nei's (1986) G_{ST} unbiased for sample size and population number.
[f] Regional groups defined in the legend of Table 2.3.

Population differentiation

Haplotype frequencies differed substantially among populations (Table 2.2). With the exception of one individual of the Cambria population that showed the same haplotype as the Año Nuevo population, every population of Monterey pine and knobcone pine had a distinctive haplotype. The southern populations of bishop pine, San Vicente and Santa Cruz each had unique haplotypes. No haplotype was shared among species.

Based on haplotype frequencies, Nei's (1986) G_{ST} and Weir and Cockerham's (1984) θ value were very similar (Table 2.3). Differentiation among populations within species was 0.79 for Monterey pine, 0.78-0.79 for knobcone pine, and 0.75-0.77 for bishop pine. Differentiation among species (0.21-0.22) and differentiation among northern, intermediate, and southern regions of bishop pine (0.34-0.35) were substantially lower than that among populations in the total species complex (0.78) and in bishop pine as a

whole (0.75-0.77). Thus, the strong population differentiation observed does not accumulate linearly at higher phyletic levels.

Table 2.5 Nested Analysis of Molecular Variance (AMOVA) for mitochondrial DNA-RFLP phenotypes

Source of variation	df [a]	SSD [b]	MSD [c]	Variance component	% of total [d]	*P*-value [e]
Among species	2	14.15	7.08	0.014	5.80	0.22
Among populations (within species)	10	55.94	5.59	0.213	87.26	0.01
Within populations	328	5.56	0.02	0.017	6.94	0.01
Among *P. muricata* regions	2	17.39	8.70	0.085	36.34	0.33
Among populations (within regions)	3	11.21	3.74	0.137	58.88	0.01
Within populations	163	1.82	0.01	0.011	4.78	0.01

[a] df = degrees of freedom.
[b] SSD = sums of squared deviations.
[c] MSD = Mean squared deviations.
[d] Percent variance of total.
[e] *P*-value = probability of obtaining a larger variance component by chance alone.

G_{ST} was considerably higher when the probe-enzyme multi-locus analysis was used. Population differentiation varied from 0.87 to 0.93 for three species. AMOVA analysis also demonstrated that the total mtDNA-RFLP polymorphism was mainly attributed to the variance among populations within species (87.3%) (Table 2.5). Variance among species and within populations each accounted for less than 7% of total variance. The Φ

value (F_{ST} analog), like G_{ST} and θ, can be interpreted as the fraction of among group variance compared to the total amount of variance in the reference group. The Φ values for populations were all above 0.90 within the three species (Table 2.6). Although the Φ values among species and among regions of bishop pine appeared to be high (0.22 and 0.66, respectively), they were not statistically significant ($P > 0.20$).

Table 2.6 Φ-statistics from Analysis of Molecular Variance

Level of analysis	V(A) [a]	V(B) [b]	Φ_{ST} [c]	P-value [d]
Pooled species	0.223	0.017	0.93	0.01
Species	0.065	0.182	0.26	0.22
Populations in species:				
P. radiata	0.230	0.023	0.91	0.01
P. attenuata	0.216	0.023	0.90	0.01
P. muricata	0.205	0.011	0.95	0.01
Regional	0.154	0.078	0.66	0.33

[a] V(A) = variance among a hierarchical level (*e.g.*, among populations).
[b] V(B) = variance within a hierarchical level (*e.g.*, within populations).
[c] $\Phi_{ST} = V(A)/[V(A)+V(B)]$.
[d] P-value = probability of obtaining a larger V(A) and Φ_{ST} by chance under permutation test.

Phylogenetic analysis

The Neighbor-Joining phylogenetic tree indicated that mitochondrial genomes representing the species and populations were often polyphyletic (Figure 2.3). The phenogram topology had three main clusters. One cluster (at bottom) contained four populations from the two species and had low bootstrapping support. The four northern populations of bishop pine were grouped into a second paraphyletic cluster with strong bootstrapping support (98%). The third cluster (at the top) included populations from all three species, yet had very high bootstrapping support (100%).

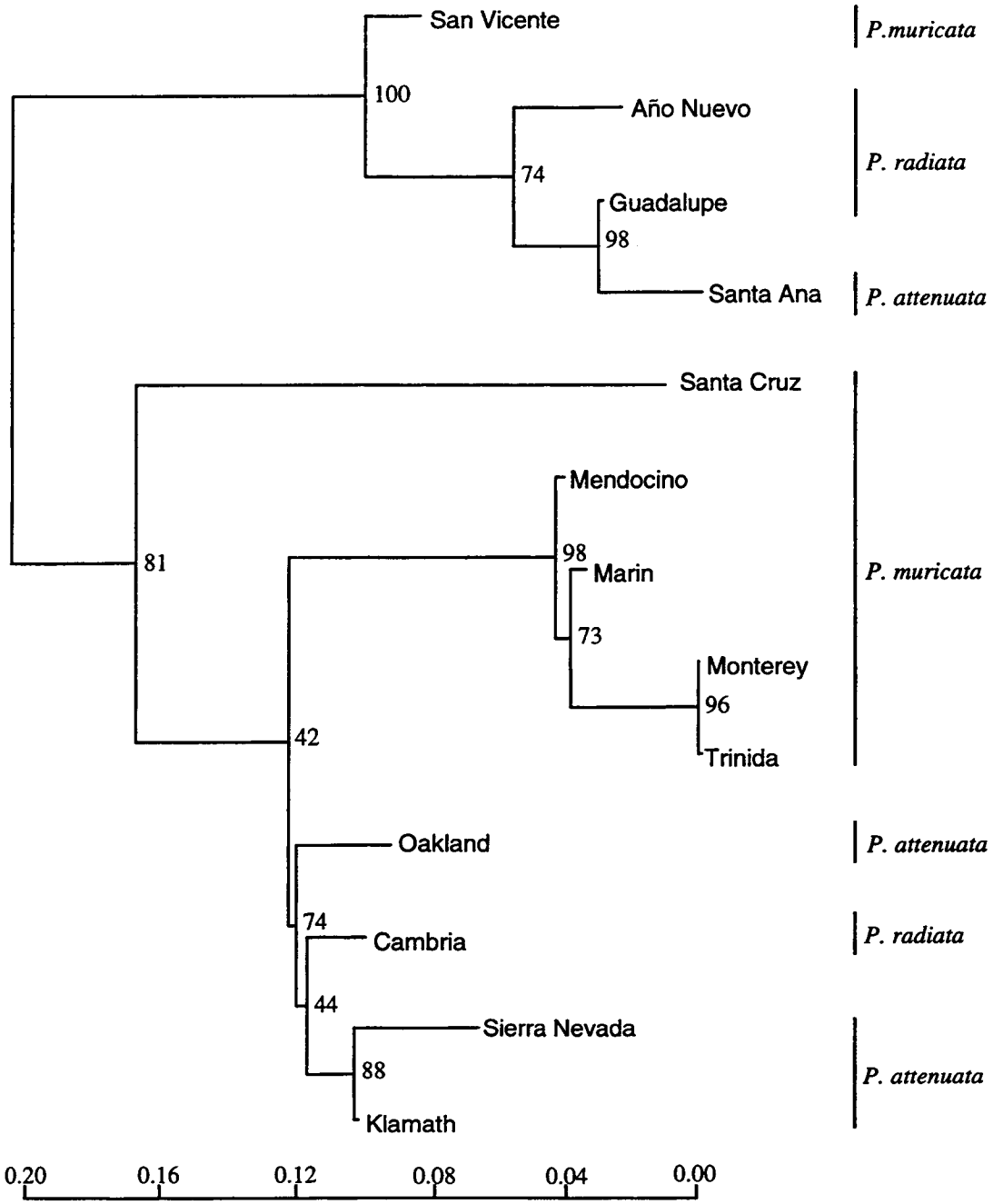


Figure 2.3 Neighbor Joining phenogram derived from Manhattan distances on the basis of mtDNA phenotypes. The tree was rooted at midpoint between the pair of taxa with the greatest patristic distance. Number to the right of relevant nodes are percentage of 200 bootstrap replicates.

DISCUSSION

Diversity

We used 22 probe-enzyme combinations (11 independent probes and two enzymes) to detect mtDNA-RFLP polymorphisms in 343 individuals from 13 populations. This is the most intensive genome sample employed in plant population genetic studies of plant mtDNA to date, providing a window on polymorphism and micro-evolution of the entire genome. By contrast, Belhassen *et al.* (1993) used one heterologous probe in their study of 52 individuals from three populations of *Thymus vulgaris*. Dong and Wagner (1993) surveyed 741 individuals, by hybridization with two probes, from 16 allopatric populations of *P. banksiana* and *P. contorta*. Strauss *et al.* (1993) examined RFLP polymorphisms associated with one gene sequence in 268 trees derived from 19 CCCP populations. Hong *et al.* (1995) applied three gene probes when analyzing 72 trees from 18 populations of Douglas-fir.

Based on our preliminary survey, probes *nad3 and nad3-rps12*, *nad1* and *rps14* provided identical mtDNA polymorphisms for both enzymes, confirming that *nad3* and *rps12*, and *nad1* and *rps14* genes are very closely located in the pine mitochondrial genome, similar to that of angiosperms (Perrotta *et al.* 1996). Thus, only 22 out of 26 probe-enzyme combinations were used in the preliminary survey. Eleven of these 22 combinations detected intraspecific and/or interspecific polymorphism and were employed for the full survey of all collected samples. Two of remaining 11 monomorphic combinations were used in an extended study (100 trees), but no additional polymorphism was detected; they were therefore not studied further. Because of the strong population differentiation for mtDNA, it is unlikely that significant additional polymorphisms existed that were missed by the preliminary surveys.

The level of mtDNA haplotype diversity is often high in plants. Total gene diversity ranges from 0.68 for lodgepole pine (*Pinus contorta* Dougl.) (Dong and Wagner 1993) to 0.78 for Douglas-fir (*Pseudotsuga menziesii* (Mirb) Franco) (Hong *et al.* 1995), similar to our results. Strauss *et al.* (1993) found earlier that intrapopulation haplotype diversities over 19 CCCP populations averaged only 0.07 for mtDNA, less than half of that for allozymes and approximately one-third of the present estimate of 0.22. However, only a

single gene probe was used, and thus only a small portion of the mtDNA genome was surveyed in that study. In contrast, we used 11 probes from 11 different genes, allowing us to resolve 28 haplotypes in 13 CCCP populations, while only nine haplotypes were detected in 19 CCCP populations studied earlier (Strauss *et al.* 1993). Haplotype diversity within populations of lodgepole pine (0.21) is very similar to that for the CCCP, though it is very low for jack pine (*Pinus banksiana* Lamb.) (0.03) when a species-hybrid population was excluded (Dong and Wagner 1993). In Douglas-fir, the mtDNA genetic diversity within populations is 0.33, higher than values in most other species (Hong *et al.* 1995).

Although mtDNA has the lowest sequence mutation rate among the three plant genomes, its presumed high rate of structural rearrangement is likely to be the cause of its high level of diversity. All the polymorphisms that we detected appeared to result from structural rearrangements, particularly large duplications and deletions, rather than point mutations. This is in agreement both with results from previous studies in CCCP (Strauss *et al.* 1993), jack and lodgepole pines (Dong and Wagner 1993), and Douglas-fir (Hong *et al.* 1995), and with most observations of plant mtDNA polymorphisms (Palmer 1990, 1992a). MtDNA intrapopulation haplotype diversity for the CCCP in this study (0.22) was higher than that obtained from allozymes (0.15, Millar *et al.* 1988), cpDNA site mutations (0.06, Hong *et al.* 1993a) and cpDNA length mutations (0.17, Hong *et al.* 1993a). MtDNA diversity within populations (0.03) calculated from allele frequencies was substantially lower than for mtDNA haplotype diversity (0.22). This is not surprising as diversity calculated from allele frequencies is averaged over all loci, whereas haplotype diversity is an integrated measure of variation that treats the entire mtDNA genome as a single locus.

Differentiation

More than three-quarters of mtDNA diversity was distributed among populations in all three species, in contrast to the low population differentiation for nuclear genes in typical long-lived woody species (G_{ST} = 0.10, Hamrick and Godt 1990), and allozyme studies for the CCCP (G_{ST} = 0.12 to 0.22, Millar *et al.* 1988). The few other surveys of mtDNA

polymorphisms have also found high levels of population differentiation. For lodgepole pine, F_{ST} was 0.31 among subspecies and was up to 0.82 among populations within subspecies (Dong and Wagner 1993), whereas F_{ST} at a single allozyme locus was rarely larger than 0.06 in this species (Wheeler and Guries 1982). Of the two organelle genomes and nuclear genome in Douglas-fir, mtDNA also showed the highest degree of genetic differentiation (G_{ST}) among populations (0.45) and geographic regions (0.29), while differentiation among populations was 0.20 for cpDNA and 0.14 for nuclear RAPD markers (Hong *et al.* 1995). Mitochondrial *cox1*-associated G_{ST} values were as high as 0.88 in the previous study of mtDNA variation in the CCCP (Strauss *et al.* 1993).

Maternally inherited cytoplasmic polymorphisms in plants are expected to exhibit greater population differentiation at equilibrium than nuclear polymorphisms. This is due to the influence of maternal inheritance on both gene flow and effective population size, and a consequence of the lower effective population size of haploid vs. diploid genomes (Birky 1988; Petit *et al.* 1993a). As a result, maternally inherited cpDNA, like mtDNA, also can show strong population subdivision. Petit *et al.* (1993b) reported that more than 85% of cpDNA diversity resided among populations within *Quercus* species, while the value for allozymes was less than 5% (Kremer and Petit 1993). Similar results have been obtained in *Eucalyptus nitens* (Byrne and Moran 1994), where the majority of cpDNA variation was distributed among populations as a result of population isolation and genetic drift (N_{ST} and $G_{ST} = 0.78$).

However, cpDNA in conifers shows predominant paternal inheritance (Neale and Sederoff 1989; Wagner *et al.* 1992), which allows cpDNA to migrate through both seeds and pollen. Consistent with drift-migration equilibrium predictions for paternally inherited markers, population differentiation for cpDNA in Douglas-fir is less than half of that for maternally inherited mtDNA (Hong *et al.* 1995). In bishop pine, however, mtDNA subdivision is similar to that for cpDNA restriction site mutations, where strong differences among populations ($G_{ST} > 87\%$) were observed (Hong *et al.* 1993a). However, most bishop pine populations are geographically and reproductively isolated (Critchfield and Little 1966; Millar and Critchfield 1988), and gene flow among populations is likely to be infrequent. As a result, the proportion of allozyme variation

attributable to population differentiation in bishop pine (22%) is also much greater than is typical for other conifers.

It thus appears that the strong population subdivision we observed for mtDNA in the CCCP can be explained by maternal inheritance of mtDNA combined with rapid structural evolution, and genetic drift due to population isolation. In contrast, the strong population differentiation for cpDNA site mutations in bishop pine results mainly from very limited gene flow among isolated geographic regions. However, the low effective population size of a haploid organelle genomes, and the possibility of periodic selection (Birky 1988; Maruyama and Birky 1991), could also contribute to the high subdivision of organelle DNA compared with nuclear gene markers.

Phylogenetic relationships

Despite high haplotype differentiation, genetic distances between populations were often low. Although nearly every population had unique haplotypes, most of the fragments were shared by other populations, including those of the other species. For example, the Guadalupe population of Monterey pine and the San Vicente population of bishop pine shared no haplotypes, yet had 24 fragments in common out of 31 total fragments; their genetic distance was only 0.06 (data not shown).

The phylogenetic trees based on our mtDNA analyses roughly agree with those reported in the mtDNA study of Strauss *et al.* (1993), but both disagree strongly with those based on morphology, allozymes and RAPDs. Allozymes have strongly confirmed monophylly of the three species (Millar *et al.* 1988), and the close relationships of three mainland populations of Monterey pine. Allozymes, terpenes (Mirov *et al.* 1966) and cpDNA (Hong *et al.* 1993b) have recognized the strong divergence of the northern versus southern populations of bishop pine. In contrast, our phylogenetic trees would suggest that the species are all polyphyletic. Similar results have been found in other studies of conifers. Dong and Wagner (1993) found that populations of lodgepole pine did not generally cluster by subspecies, discordant with traditional taxonomy. Hong *et al.* (1995) observed that Douglas-fir populations in each of three geographic regions of British Columbia failed to cluster based on geographic affinity.

The complex nature of mtDNA evolution is probably the cause of its poor performance as a phylogenetic marker. The assumptions that the presence or absence of a mtDNA fragment is due to the same mutational event and that the phenotypes reflect the underlying mutational events in mtDNA are likely violated. Thus, it should be stated clearly that the result found with mtDNA differs from earlier phylogenies because of the violation of the assumption that the absence/presence of a fragment reflects the same mutational event. RFLP polymorphisms of plant mtDNA are mostly length mutations and complex rearrangements, rather than site mutations. It is therefore difficult to infer the evolutionary homology among different haplotypes (or fragments) due to the complex and overlapping nature of structural changes (Palmer 1992a), causing apparent homoplasic and convergent evolution. For example, the Guadalupe population of Monterey pine, the Santa Ana population of knobcone pine, and the San Vicente population of bishop pine shared most of their restriction fragments (see e, q, and bb haplotypes, respectively, in Appendix A), although they are widely separated geographically. We hoped that because of our large sample of the genome we might be able to "average over" individual homoplasious rearrangements reported earlier (Strauss *et al.* 1993); however, this clearly was not the case.

The high frequency of convergent evolution is likely to be associated with the repetitive nature of mtDNA. MtDNA structural rearrangements are associated with recombination among major repeat elements (Palmer 1992a); if there are a finite number of sections of the genome that recombine in predictable ways across these hotspots, similar genome structures could evolve repeatedly (Strauss *et al.* 1993).

Although mtDNA rearrangements do not appear to be of value for phylogenetic interpretations in pines, they may be of use in other taxa for grouping closely related genomes (Palmer 1992a). For example, most of 345 rubber (*Hevea brasiliensis*) accessions can be grouped according to their geographical distributions and hydrographical origin (Luo *et al.* 1995). Deu *et al.* (1995) also successfully employed mtDNA to cluster several races in wild and cultivated sorghum (*Sorghum bicolor* ssp. *arundinaceum* and *S. bicolor* ssp. *bicolor*). The value of mtDNA for phylogenetic

ACKNOWLEDGEMENT

We wish to thank Tony Cario for sampling the San Vicente population of bishop pine; Jan Aagaard and Nathan Strauss for their help in field collections; Bill Libby for his advice and help in accessing the study populations; Steve DiFazio for help with editing the manuscript; and the National Science Foundation (NSF Conservation and Restoration Biology, DEB-9300083) for grant support.

CHAPTER 3

NUCLEAR DNA DIVERSITY, POPULATION DIFFERENTIATION AND PHYLOGENETIC RELATIONSHIPS IN THE CALIFORNIA CLOSED-CONE PINES BASED ON RAPD AND ALLOZYME MARKERS

ABSTRACT

We studied nuclear gene diversity and population differentiation using 91-98 random amplified polymorphic DNA (RAPD) loci in the California Closed-Cone Pines: knobcone (*Pinus attenuata* Lemm.), bishop (*P. muricata* D. Don) and Monterey (*P. radiata* D. Don) pines. A total of 384 trees from 13 populations were analyzed for RAPDs and 242 trees from 12 populations were analyzed at 32-36 allozyme loci using a published dataset. Twenty-eight of 30 (93%) comigrating RAPD fragments tested were found to be homologous by Southern hybridization in all three species. Using an enriched mitochondrial DNA (mtDNA) preparation and a chloroplast DNA (cpDNA) library as probes, two fragments of cpDNA origin and one of mtDNA origin present among RAPD profiles were excluded from analysis. RAPD markers revealed moderately higher intrapopulation gene diversity and significantly higher total genetic diversity and population differentiation than did allozyme markers for each species. We performed a simulation using our allozyme data to study whether these discrepancies could be explained by the dominant and biallelic nature of RAPD markers. Transforming allozyme data to simulate dominance gave substantially lower values of diversity within populations compared to those based on the original codominant, multiallelic allozyme dataset; the higher RAPD diversities observed suggest that RAPD loci contain much higher levels of inherent genetic polymorphisms than do allozyme loci. Results of joint phylogenetic analysis of both the RAPD and allozyme markers strongly supported a common ancestor for *P. radiata* and *P. attenuata*, and south to north migration histories for all three species.

INTRODUCTION

Assessment of genetic variation in plant populations has many potential uses for evolutionists, breeders and conservation biologists. The amount and distribution of genetic diversity reflect the interaction of various evolutionary processes such as gene flow, mutation, genetic drift, and natural selection (e.g., Wright 1978). Knowledge of genetic variation within a species can help plant breeders to collect and utilize different genetic resources and predict potential genetic gain in breeding programs. Implementation of conservation policies also requires a careful analysis of genetic variation in target species.

Over the last a few decades, allozymes have been the markers of choice for quantifying population structure at the molecular level (e.g., Hamrick and Godt 1990). Allozymes usually exhibit simple Mendelian inheritance and codominant expression, facilitating genetic interpretations. However, allozymes have some well-known limitations, such as highly biased genomic sampling (only detecting protein-encoded genes and a fraction of all mutational events) and small number of available loci which may not adequately represent genome-wide diversity.

Random amplified polymorphic DNA (RAPD) has been widely used to study population genetic structure and phylogenetic relationships (see Aagaard 1997 and Aagaard *et al.* 1998a, b for references). It provides a more representative sample of the genome than allozymes and a very large number of loci can be readily studied. Several reports have compared genetic diversity and population differentiation using allozyme and RAPD markers (e.g., Aagaard *et al.* 1998a; Baruffi *et al.* 1995; Isabel *et al.* 1995; Lannér-Herrera *et al.* 1996a; Latta and Mitton 1997; le Corre *et al.* 1997; Peakall *et al.* 1995; Vicario *et al.* 1995), and have usually found RAPDs to show somewhat greater polymorphism and/or differentiation than allozymes (e.g., Liu and Furnier 1993; Szmidt *et al.* 1996). RAPD markers have also been utilized extensively in the reconstruction of phenetic and cladistic relationships (reviewed in Wolfe and Liston 1997). There has been general concordance among the results derived from RAPDs and other techniques (e.g., Graham *et al.* 1995; Lifante and Aguinagalde 1996; Marillia and Scoles 1996).

However, RAPDs also have some significant limitations. First, the dominant allelic expression of RAPDs precludes direct estimates of allele frequencies from diploid material, and can thus bias the calculation of genetic diversity and population differentiation (Lynch and Milligan 1994; Isabel *et al.* 1995; Szmidt *et al.* 1996). Secondly, lack of homology among comigrating RAPD fragments has been found in some species, confounding phenotypic interpretation (e.g., Rieseberg 1996; Thormann *et al.* 1994). Finally, RAPDs of organellar origin have been reported in several studies (e.g., Aagaard *et al.* 1995, 1998b; Lorenz *et al.* 1994), which will skew the comparison of genetic parameters between allozymes and RAPDs.

The goals of this study were to estimate nuclear DNA diversity using RAPD markers and to compare it to results from allozyme and organelle DNAs. The specific objectives were to: (1) test the homology of comigrating RAPD fragments among populations and species; (2) identify the frequency of organellar DNA markers among RAPD profiles; (3) verify bi-parental, Mendelian inheritance for a sample of putative nuclear markers; (4) compare the genetic diversity and differentiation of RAPD and allozyme markers; (5) simulate the effects that dominance and biallelism would have on allozyme markers to facilitate comparison to RAPD markers; and (6) study phylogenetic relationships among populations and species using the two markers separately and together.

The California Closed-Cone Pines (CCCP) are composed of three species: one interior species, *Pinus attenuata* Lemm. (knobcone pine), and two maritime species, *P. muricata* D. Don (bishop pine) and *P. radiata* D. Don (Monterey pine). *P. attenuata* grows on interior sites of southern Oregon and California as disjunct populations. The two other species are distributed discontinuously along the California coast and on four islands (Figure 3.1). Many characteristics of these species have been intensively studied for population genetic variation, including morphology, secondary compound chemistry, allozymes, chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA) (Hong *et al.* 1993a, b; reviewed in Millar 1986; Millar *et al.* 1988; Strauss *et al.* 1993; Wu *et al.* 1998), thus providing ideal taxa for comparisons among genetic markers. For example, a recent phylogenetic study using chloroplast DNA markers showed unexpected phylogenetic affinities and variable rates of evolution (Hong *et al.* 1993b). This result will

be re-examined in this study using the large set of nuclear DNA markers provided by RAPDs.

MATERIALS AND METHODS

Plant materials

Trees were sampled from natural populations or from gene conservation and genetic test plantations as described in Hong *et al.* (1993a). Two different collections contributed to this study (Figure 3.1). The Año Nuevo, Cambria and Guadalupe populations of *P. radiata*; the Sierra Nevada and Santa Ana populations of *P. attenuata*; and the Santa Cruz population of *P. muricata* were primarily collected by Hong *et al.* (1993a). The other populations were collected specifically for this study. For *P. attenuata*, the Klamath population was sampled over a 6.0-kilometer (km) transect adjacent to the Lakehead Exit on U.S. Interstate 5, California (latitude 40°55', longitude 122°30'), and the Oakland population was sampled over a 2.6-km transect along Flicker Ridge adjacent to the town of Moraga in the hills east of Oakland, California (latitude 37°50', longitude 122°30'). For *P. muricata*, the San Vicente population was sampled in several small scattered populations along a road north of San Vicente that goes out to the town of Erendira, Mexico (latitude 31°15', longitude 116°30'); the Monterey population was sampled over one linear mile in the woods of Samuel F. B. Morse Botanical Reserve located south of Monterey, California (latitude 36°40', longitude 121°50'); the Marin population was sampled over a 1.7-km transect about 5.1-km southwest from the town of Inverness, California (latitude 38°08', longitude 122°45'); the Mendocino population was sampled for 2.6-km on both sides of U.S. Highway 101, 7.5-km south of the town of Point Arena, California (latitude 39°20', longitude 123°50'); and the Trinidad population was sampled over 1.7 linear kilometers up Fox Farm Road adjacent to the town of Trinidad, California (latitude 41°05', longitude 124°10').

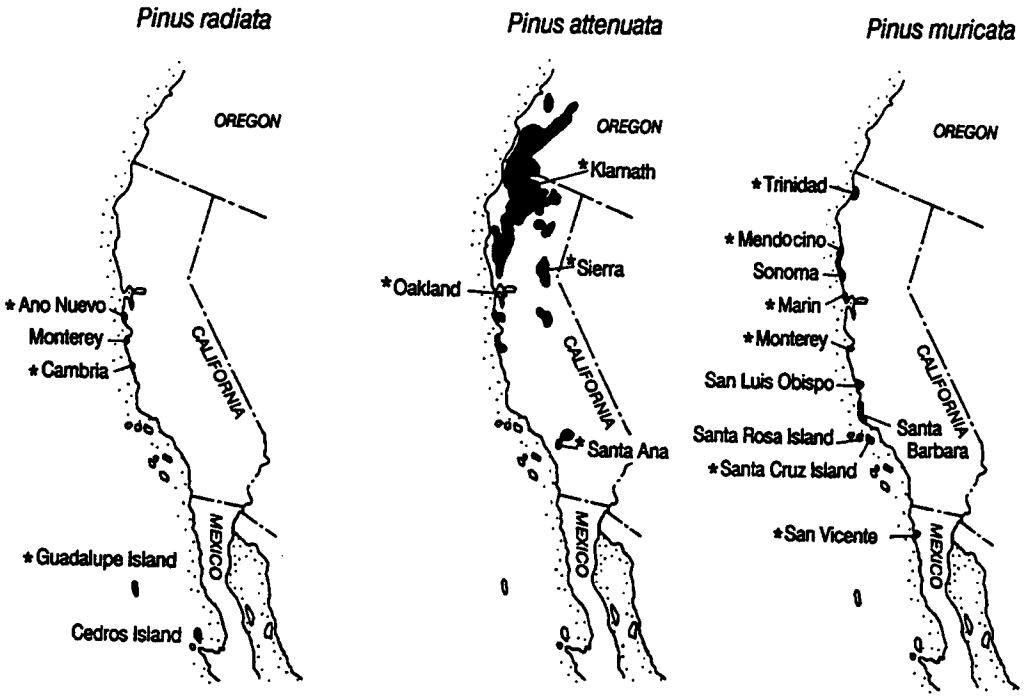


Figure 3.1 Distribution of *Pinus radiata, P. attenuata, and P. muricata* (Hong et al. 1993a) and the origins of sampled populations (*). All populations were studied for chloroplast DNA polymorphism by Hong et al. (1993a).

Allozyme data

We reanalyzed the allozyme data collected by Millar *et al.* (1988) as well as some additional data (R. Westfall, pers. comm.) for populations that were geographically matched to the populations used for the RAPD study (Figure 3.1). The population identities and sample sizes were also shown in Table 3.1 and 3.2 for both markers. Allozyme data were not available for the San Vicente population.

DNA extraction and RAPD procedures

Total genomic DNA was extracted from megagametophytes and needles using a modified CTAB protocol described previously (Aagaard *et al.* 1995), followed by three phenol/chloroform purifications and a final ethanol precipitation. RAPD reactions were carried out in a volume of 25 µl with final concentrations of 10 mM Tris-HCl (pH 8.3); 1.8 mM $MgCl_2$; 100 µM of each dNTP, 0.2 µM of primer; 1.3 mg/ml BSA; 2 ng template DNA and one unit of *Taq* Polymerase (Stratagene, USA). Amplification was performed in a PTC-100 thermocycler (MJR Research, Inc., USA) using 3 minutes of initial denaturation at 93°C, followed by 44 cycles of 1 minute of denaturation at 93°C, 1 minute of annealing at 37°C, 2 minutes of extension at 72°C, and a final extension step of 10 minutes at 72°C. All amplification products were electrophoresed through 2.0% agarose gels (TBE buffer pH 8.0) for 6 hours. Gels were stained with 1 µg/ml of ethidium bromide for 30 minutes, then destained in distilled water for 2 hours or more, and photographed under UV light.

A total of 500 primers from University of British Columbia (UBC, Vancouver, Canada) and Operon Technologies, Inc. (USA) were screened using one representative of each of three populations per species. Twenty-two primers that gave distinct, repeatable fragment patterns were selected for final analysis – 17 UBC primers: #106 (CGTCTGCCCG), #137 (GGTCTCTCCC), #154 (TCCATGCCGT), #184 (CAAACGGCAC), #195 (GATCTCAGCG), #203 (CACGGCGAGT), #219 (GTGACCTCAG), #254 (CGCCCCCATT), #268 (AGGCCGCTTA), #299 (TGTCAGCGGT), #337 (TCCCGAACCG), #352 (CACAACGGGT), #429 (AAACCTGGAC), #485 (AGAATAGGGC), #503 (ATCGTCCAAC), #536

(GCCCCTCGTC), #587 (GCTACTAACC), and 5 Operon primers: #OPA11 (CAATCGCCGT), #OPC04 (CCGCATCTAC), #OPE17 (CTACTGCCGT), #OPG09 (CTGACGTCAC), and #OPY17 (GACGTGGTGA). Due to interpretation problems primers UBC #154, UBC #203, and OPA11 were scored only in *P. attenuata* and *P. radiata*, and OPY17 was scored only in *P. radiata*. In order to avoid biasing the estimation of polymorphism, the selection of primers for full analysis was dependent only on the clearness and reproducibility of RAPD fragments, not on the level of polymorphism.

Homology of RAPD markers

We used Southern analysis to test homology among populations and species for 30 different RAPD markers. Some of these markers had an uneven distribution of fragment staining intensities within or between species. The number of trees used for each test ranged from 30 to 150. RAPD fragments used as probes were excised individually from agarose gels, purified using the QIAquick gel extraction kit (QIAGEN, Inc., USA) or GENECLEAN kit (BIO101, Inc., USA), and then radioactively labelled with [^{32}P] by random hexamer labelling using Boehringer Mannheim's (Germany) random primed DNA labelling kit. RAPD fragments were first transferred to Zetabind nylon filters (CUNO, Inc., USA) and then hybridized to isolated probes. Hybridization protocols followed Hong (1991), but were followed by final high stringency washes (3 times of 0.1x SSC, 1% SDS, 65°C).

Identification of organellar origin of RAPD markers

As described previously (Aagaard 1997), Southern hybridization of RAPD blots with organelle DNA was used to identify organellar DNA products among RAPD profiles. MtDNA was extracted from embryogenic suspension cultures of *P. radiata* (J. Aitken-Christie, unpublished) following the protocol of Aagaard (1997). The degree of enrichment of mtDNA was quantified using Southern hybridization: 0.2 μg mtDNA extracted from embryogenic suspension cultures and 1 μg total genomic DNA extracted from needles were hybridized with a ^{32}P-labelled mitochondrion-specific *coxIII* probe

that had been amplified from *P. radiata* using a pair of universal primers (K. Krutovskii, unpublished data; Wu *et al.* 1998). Phosphorimage analysis (PhosphorImager model PSI-PC, Molecular Dynamics, USA) was used to quantify the ratio of hybridization signals of mtDNA and genomic DNA to estimate relative mtDNA enrichment.

For the cpDNA hybridizations, we probed with mixtures of cloned cpDNA fragments from *P. contorta* (Lindholm and Gustafsson 1991). Each hybridization used half of 24 clones that together cover 92% of the chloroplast genome. Electrophoretically-separated RAPD products were blotted and sequentially hybridized with the mtDNA and cpDNA probes to determine the origin of RAPD markers. High (3 times of 0.1x SSC, 1% SDS, 65°C) and low (3 times of 2x SSC, 0.5% SDS, 65°C) stringency washes were used for the mtDNA and cpDNA probes, respectively. RAPD fragments which showed clear hybridization were excluded from nuclear DNA analyses.

Segregation analysis

Nuclear markers were also identified and confirmed by segregation analysis. Since normally mtDNA is maternally and cpDNA is paternally inherited in pines (Hipkins *et al.* 1995) and should not segregate in megagametophytes, it was assumed that markers segregating among megagametophytes from the same trees were nuclear markers. We used seeds from two hybrid trees between *P. attenuata* and *P. radiata*, three hybrid trees between different populations of *P. radiata*, and 10 trees collected in eight natural populations for segregation analysis. A small number of megagametophytes (4-20) were used from each tree because our goal was simply to identify segregating markers, rather than test for quantitative conformance to Mendelian ratios. However, in order to conservatively select segregating markers, homogeneity of segregation ratios among trees, and the segregation ratio deviation for pooled samples taken from different trees (including hybrids and different species), were also tested using a *G*-test (Sokal and Rohlf 1981).

Genetic diversity analysis

RAPD markers show mainly dominant expression, usually with only a few percent showing codominance (e.g., Heun and Helentjaris 1993; Krutovskii *et al.* 1998a; Lu *et al.* 1995). Thus, allele frequencies for RAPD loci were calculated from the phenotypic frequencies of null homozygotes assuming dominance and Hardy-Weinberg equilibrium. Wind-pollinated species, including the CCCP and other conifers, generally show only modest and transient departures from Hardy-Weinberg proportions (e.g., Isabel *et al.* 1995; Moran *et al.* 1988; Plessas and Strauss 1986).

Because of dominance, corrected allele frequencies (q) were used for estimation of gene diversity parameters (Lynch and Milligan 1994): $q = x^{1/2}[1\text{-}Var(x)/8x^2]^{-1}$, where $Var(x) = x(1\text{-}x)/N$, and x is the proportion of the N sampled individuals that do not have amplified RAPD fragments (null-allele homozygous phenotype). Mean (A) and effective (A_e) number of alleles per locus, number of private alleles (U) for each population, percentage of polymorphic loci (95% criterion, P_{95}), Nei's (1978) unbiased expected heterozygosity (H_e), and population differentiation (G_{ST}) unbiased for sample size and population number (Nei 1986) were calculated using GENESTAT-PC 3.3 (Lewis 1994). To facilitate comparison of our data with that in a review of estimates of plant diversity based on allozymes, we also calculated G_{ST} by averaging over polymorphic loci (Hamrick and Godt 1989). In addition, Wright's (1931) F_{ST}, Weir and Cockerham's (1984) θ and Lynch and Milligan's (1994) F_{ST} were also calculated using Black's RAPDFST program (1996) for comparison with Nei's G_{ST}. Hierarchical F-statistics describing subdivision between populations within species and between species were analyzed using the BIOSYS-1 version 1.7 (Swofford and Selander 1989) that we recompiled for large sample sizes.

RAPD phenotypic (presence/absence) data were used to produce F-statistic analogs called Φ-statistics (e.g., Φ_{ST} = variance among populations/total variance within species) using the analysis of molecular variance procedure (WINAMOVA, Excoffier *et al.* 1992; Excoffier 1993). For each tree we created a binary vector of 1's (presence of a band) and 0's (absence of a band) for RAPD loci. The proportion of shared fragments was calculated for each possible pairwise comparison according to the formula proposed by Nei and Li (1985): S (similarity) = $2N_{AB}/(N_A+N_B)$ (where N_{AB} is the number of bands

of 1,000 individuals each having genotypes that maintained the original allele frequencies within each population. A total of 200 subpopulations of n individuals were then sampled with replacement from each of the 12 populations. For the codominant data set, genetic parameters (H_S, H_T and G_{ST}) were calculated from allele frequencies inferred directly from each set of 200 subpopulations. For the dominance analysis, one of the alleles at each locus was randomly selected as a dominant allele and others pooled as a null recessive allele. The genotype frequency of null homozygotes was then used to calculate null and dominant allele frequencies assuming Hardy-Weinberg equilibrium, and diversity and differentiation parameters were estimated based on these frequencies. Means and variances of H_S, H_T and G_{ST} were then calculated from the 200 sampled subpopulations.

Phylogenetic analysis

Nei's (1972) genetic distances between both populations and species were calculated from allele frequencies for both RAPD and allozyme markers using BIOSYS-1 program (Swofford and Selander 1989) and GENEDIST program in the PHYLIP package (Felsenstein 1995). In addition, the pairwise Manhattan distances between populations (Prevosti distance in Wright 1978) were computed on the basis of dissimilarities in RAPD phenotypes (presence/absence of an amplified fragment) using RAPDDIST program (Black 1996). A distance matrix or a set of matrices generated by bootstrapping over loci were produced using the RAPDDIST or RAPDBOOT programs (Black 1996). The distance matrices based on both allele frequencies and phenotypes were then used to produce phylogenetic trees using Neighbor-Joining Tree (NJTREE) and UPGMA clustering programs in PHYLIP (Felsenstein 1995). The SEQBOOT program was also used to generate bootstrap datasets, and the majority-rule and strict consensus trees were generated from bootstrap trees using the program CONSENSE in the PHYLIP package. The phylogenetic trees were viewed and drawn using TREEVIEW program (Page 1996). We also combined both allozyme and RAPD loci for phylogenetic analysis.

RESULTS

A total of 94 to 100 fragments were scored in each of *P. attenuata*, *P. muricata* and *P. radiata*. More than 70% of the scored fragments showed polymorphism across populations within species. The percentage of polymorphic loci for each species ranged from 70% to 85%, with the highest percentage for *P. muricata*. Seventy-six fragments were scored as homologous bands among all three species.

Homology of RAPD markers

Southern blots confirmed homology for 28 out of 30 (93%) randomly chosen RAPD markers tested by hybridizing comigrating DNA fragments to blots of DNAs from different populations and species. For two of the 30 markers, some comigrating fragments did not give a hybridization signal. In addition to these 30 comigrating fragments, several other markers were also tested for hybridization because of their inconsistent band phenotypes. In two cases, the fragments comigrated but had very different intensities among individuals within populations: the homology tests showed that the fragments were non-homologous, and controlled by at least two genetic loci. In two other cases, hybridization revealed that fragments with slight mobility differences were homologous and presumably controlled by two codominant alleles of the same locus. However, in two other cases, fragments of slightly different sizes were found to be completely non-homologous.

Identification of organelle DNAs among RAPD profiles

Quantification of mtDNA enrichment indicated that mtDNA extracted from embryogenic cultures was more than 50 times enriched for mtDNA compared to total genomic DNA (data not shown). However, of all of the RAPD fragments scored, only one fragment amplified by primer UBC #268 hybridized strongly with the enriched mtDNA. Two fragments amplified by primer UBC #337 were identified to be of putative chloroplast origin. These three fragments were excluded from further RAPD data analysis. Thus, 91 RAPD markers for *P. radiata* and 98 markers for each of *P. attenuata* and *P. muricata* pine were retained as putative nuclear loci.

Segregation analysis

To establish a data set of nuclear loci showing evidence of Mendelian segregation, we analyzed inheritance in several trees. A total of 48, 54, and 58 loci were segregating among megagametophytes in *P. attenuata*, *P. radiata* and *P. muricata*, respectively, including several hybrid trees. Altogether, 92 loci and 180 tree-locus combinations were tested in three species. Of these loci, seven showed statistically significant ($P \leq 0.05$) segregation distortion in pooled samples and were excluded from our "segregation data set". Of 52 loci that had segregation data for more than one tree and thus provided an opportunity to test segregation ratio homogeneity among trees, three loci demonstrated significant segregation ratio heterogeneity. Those loci were also excluded from the data set. Seven loci that were analyzed in only four megagametophytes from a single tree and that showed only one "absence" phenotype were also discarded because the null phenotype could have resulted from failed amplification. Thus a total of 17 loci were excluded for the three species combined, and 42, 44 and 50 loci retained for estimation of genetic parameters for *P. attenuata*, *P. radiata* and *P. muricata*, respectively. The overall (pooled for all trees and loci) ratio for the number of presence vs. absence bands was 629:653, thus showing no tendency for an excess of null phenotypes due to poor amplification ($\chi^2_{1:1} = 0.45$, $P > 0.05$). Deviations from 1:1 segregation ratio were apparently stochastic; the ratio favored the dominant allele at 36 loci and favored the null-allele at 38 loci, while the remaining 18 loci had an exact 1:1 ratio.

RAPD genetic diversity and differentiation

Different locus sets: Estimates of genetic diversity within populations derived from the set of normally segregating loci were slightly higher than those based on all RAPD loci (Table 3.1). However, differentiation among populations estimated from only segregating loci was nearly the same as that obtained from all RAPD loci. When many monomorphic and weakly polymorphic loci were excluded from the data set of all loci following the recommendation of Lynch and Milligan (1994), genetic diversities were higher and G_{ST} values decreased significantly. Using their criterion, only one-third to approximately one-half of all loci could be retained for the different species (Table 3.1).

Diversity: Based on the data set with all loci (Table 3.2), the average number of alleles per locus (A) ranged from 1.3-1.7, the effective number of alleles per locus (A_E) from 1.2 to 1.3, the percent of polymorphic loci (P_{95}) from 26.5% to 67.4%, and the expected heterozygosity (H_E) from 0.09 to 0.20. Among the three species, *P. muricata* had the lowest average number of alleles per locus (1.4), percent of polymorphic loci (38.6%) and expected heterozygosity (0.13). Total genetic diversities within species were very similar among the three species (H_T = 0.21-0.22, Table 3.2).

Differentiation: Allele frequencies were highly variable at many loci among populations within each species (data not shown). For example, there were 14 private alleles for *P. attenuata* (unique to different populations); 11 of those 14 alleles belonged to the Oakland population. In *P. radiata*, 15 private alleles were detected; however, only one marker was strictly population-specific in our sample (monomorphic in one population, absent elsewhere, UBC primer #219, 0.52 kb). Nine of those 15 alleles were present in the Guadalupe Island population. Similar results were observed with *P. muricata*: 15 private alleles were detected, and most of them belonged to the southern California mainland population (five alleles, San Vicente) and an island population (six alleles, Santa Cruz). One marker (UBC primer #268, 1.82kb) was specific to the San Vicente population sample.

Populations within each species generally showed high differentiation from one another. In contrast to intrapopulation diversity, *P. muricata* revealed the highest population differentiation due to its large regional differences (G_{ST} = 0.45 compared to 0.36 for *P. attenuata* and 0.26 for *P. radiata*, Table 3.1). Several different methods of analysis gave congruent trends of differentiation, although their absolute values sometimes differed substantially. Nei's (1986) G_{ST} were nearly the same as Lynch and Milligan's (1994) F_{ST} and Weir and Cockerham's (1984) θ (Table 3.3). Hamrick and Godt's (1989) G_{ST}, and Wright's (1931) F_{ST} were also very similar, but both were lower than Nei's (1986) G_{ST}. The F-statistic analog obtained from WINAMOVA analysis, Φ_{ST}, was the highest among all population differentiation parameters. When the species complex was considered as a unit, as expected the differentiation among all populations

Table 3.1 Estimates of population genetic diversity and differentiation in three California Closed-Cone Pines based on three sets of data: all putative RAPD loci, loci selected based on segregation analysis, and loci selected using the polymorphism criteria of Lynch and Milligan (1994)

Average[1]	N_L[2]	A[3]	A_E[4]	P_{95}[5]	H_E[6]	G_{ST}[7]
			All loci			
P. attenuata	98	1.48 ± 0.07	1.24 ± 0.04	48.21 ± 7.16	0.15 ± 0.02	0.36
P. radiata	91	1.50 ± 0.02	1.30 ± 0.02	49.82 ± 2.04	0.17 ± 0.01	0.26
P. muricata	98	1.41 ± 0.04	1.22 ± 0.02	40.82 ± 4.17	0.13 ± 0.01	0.45
		Loci selected based on segregation analysis				
P. attenuata	42	1.59 ± 0.07	1.32 ± 0.04	59.38 ± 6.38	0.19 ± 0.02	0.36
P. radiata	44	1.62 ± 0.02	1.38 ± 0.02	61.73 ± 2.47	0.22 ± 0.01	0.22
P. muricata	50	1.51 ± 0.04	1.30 ± 0.02	51.15 ± 4.22	0.18 ± 0.02	0.41
		Loci selected using polymorphism criteria[8]				
P. attenuata	27	1.64 ± 0.09	1.22 ± 0.04	64.81 ± 8.88	0.17 ± 0.03	0.17
P. radiata	31	1.78 ± 0.05	1.38 ± 0.06	78.49 ± 4.69	0.23 ± 0.03	0.17
P. muricata	43	1.57 ± 0.05	1.27 ± 0.02	56.59 ± 5.19	0.16 ± 0.01	0.27

[1] Average = parameters of genetic diversity averaged over all populations within each species for each data set.
[2] N_L = number of loci.
[3] A = number of alleles per locus.
[4] A_E = number of effective alleles per locus.
[5] P_{95} = percent of polymorphic loci (0.95 standard).
[6] H_E = Expected heterozygosity within populations.
[7] G_{ST} = Nei's G_{ST} (1986) unbiased for small sample size and population number.
[8] Loci that meet the criterion recommended by Lynch and Milligan (1994) that number of null homozygotes is not less than three at each locus selected for analysis within each population.

Tble 3. 2 Genetic diversity within populations and species of three California Closed-Cone Pines based on RAPD and allozyme markers

Species: Population	N[1]	A[2]	A_E[3]	P_{95}[4]	H_E[5]	N	A	A_E	P_{95}	H_E
			RAPD					Allozyme		
P. attenuata			H_T[6] = 0.21 ± 0.02 (98[7])					H_T = 0.13 ± 0.02 (36)		
Klamath	27.9	1.47 ± 0.05	1.26 ± 0.04	46.94 ± 5.07	0.15 ± 0.02	50.0	1.83 ± 0.14	1.22 ± 0.05	58.33 ± 8.33	0.15 ± 0.03
Sierra Nevada	22.0	1.46 ± 0.05	1.26 ± 0.04	45.92 ± 5.06	0.16 ± 0.02	30.0	1.92 ± 0.16	1.21 ± 0.06	63.89 ± 8.12	0.14 ± 0.03
Oakland	29.0	1.67 ± 0.05	1.32 ± 0.03	67.35 ± 4.76	0.20 ± 0.02	10.0	1.25 ± 0.07	1.11 ± 0.04	25.00 ± 7.32	0.07 ± 0.03
Santa Ana	23.0	1.33 ± 0.05	1.14 ± 0.03	32.65 ± 4.76	0.09 ± 0.02	5.0	1.14 ± 0.06	1.10 ± 0.05	13.89 ± 5.85	0.06 ± 0.03
Average	25.5	1.48 ± 0.07	1.24 ± 0.04	48.21 ± 7.16	0.15 ± 0.02	23.8	1.55 ± 0.20	1.16 ± 0.03	40.28 ± 12.3	0.11 ± 0.02
P. radiata			H_T = 0.22 ± 0.02 (91)					H_T = 0.15 ± 0.02 (32)		
Año Nuevo	27.0	1.47 ± 0.05	1.26 ± 0.04	47.25 ± 5.26	0.15 ± 0.02	15.0	1.75 ± 0.18	1.22 ± 0.07	46.88 ± 8.96	0.13 ± 0.04
Cambria	28.0	1.48 ± 0.05	1.30 ± 0.04	48.35 ± 5.27	0.18 ± 0.02	15.0	1.81 ± 0.17	1.24 ± 0.07	50.00 ± 8.98	0.14 ± 0.03
Guadalupe	27.0	1.54 ± 0.05	1.33 ± 0.04	53.85 ± 5.25	0.19 ± 0.02	15.0	1.72 ± 0.15	1.23 ± 0.07	46.88 ± 8.96	0.14 ± 0.03
Average	27.3	1.50 ± 0.02	1.30 ± 0.02	49.82 ± 2.04	0.17 ± 0.01	15.0	1.76 ± 0.03	1.23 ± 0.00	47.92 ± 1.04	0.14 ± 0.04
P. muricata			H_T = 0.22 ± 0.02 (98)					H_T = 0.15 ± 0.02 (32)		
Trinidad	24.9	1.27 ± 0.04	1.16 ± 0.03	26.53 ± 4.48	0.09 ± 0.02	10.0	1.38 ± 0.12	1.13 ± 0.06	28.13 ± 8.08	0.07 ± 0.03
Mendocino	49.8	1.40 ± 0.05	1.20 ± 0.03	39.80 ± 4.97	0.12 ± 0.02	15.0	1.59 ± 0.10	1.18 ± 0.05	56.25 ± 8.91	0.12 ± 0.03

Table 3.2 (continued)

Marin	25.0	1.35 ± 0.05	1.21 ± 0.04	34.69 ± 4.83	0.13 ± 0.02	36.0	1.69 ± 0.16	1.18 ± 0.06	46.88 ± 8.96	0.11 ± 0.03
Monterey	25.0	1.39 ± 0.05	1.21 ± 0.03	38.78 ± 4.95	0.13 ± 0.02	12.0	1.28 ± 0.08	1.17 ± 0.06	28.13 ± 8.08	0.11 ± 0.03
Santa Cruz	25.0	1.53 ± 0.05	1.27 ± 0.04	53.06 ± 5.07	0.16 ± 0.02	29.0	1.53 ± 0.12	1.27 ± 0.09	43.75 ± 8.91	0.14 ± 0.04
San Vicente	49.7	1.52 ± 0.05	1.29 ± 0.04	52.04 ± 5.07	0.17 ± 0.02	-	-	-	-	-
Average[8]	33.2	1.41 ± 0.04	1.22 ± 0.02	40.82 ± 4.17	0.13 ± 0.01	-	-	-	-	-
Average[9]	29.9	1.39 ± 0.04	1.21 ± 0.02	38.57 ± 4.31	0.13 ± 0.01	20.4	1.49 ± 0.07	1.19 ± 0.02	40.63 ± 5.50	0.11 ± 0.01
Pooled species[10]	28.0	1.44	1.24	44.13	0.15	20.2	1.58	1.19	42.45	0.11

[1] N = average number of individuals analyzed per locus.
[2] A = average number of alleles per locus.
[3] A_E = average number of effective alleles per locus.
[4] P_{95} = percent of polymorphic loci (95% criteria).
[5] H_E = expected heterozygosity within populations or within species.
[6] H_T = total genetic diversity within species.
[7] Number of loci studied in the species.
[8] Average over all six *P. muricata* populations.
[9] Average over five *P. muricata* populations shared by RAPD and allozyme studies (excluding San Vicente population).
[10] Genetic diversity parameters averaged over all populations in the three species.

Table 3.3 Parameters of genetic differentiation among populations and species of three California Closed-Cone Pines based on RAPD, allozyme, mitochondrial (mtDNA) and chloroplast (cpDNA) RFLP markers

Parameter	*P. attenuata*	*P. radiata*	*P. muricata*[1]	*P. muricata*[2]	Species	Populations in all species
			RAPDs			
G_{ST}[3]	0.36	0.26	0.45	0.41	0.42	0.60
G_{ST}[4]	0.24 ± 0.03	0.18 ±0.02	0.34 ± 0.03	0.29 ± 0.03		0.51 ± 0.03
F_{ST}[5]	0.23 ± 0.03	0.18 ± 0.02	0.35 ± 0.03	0.29 ± 0.03		
θ[6]	0.37 ± 0.01	0.26 ± 0.01	0.45 ± 0.01	0.41 ± 0.01		
F_{ST}[7]	0.37 ± 0.03	0.26 ± 0.03	0.45 ± 0.03	0.42 ± 0.03		
Φ_{ST}[8]	0.41	0.32	0.52	0.44		
F_{ST}[9]					0.31	0.58
			Allozymes			
G_{ST}[3]	0.22	0.12		0.32	0.31	0.44
G_{ST}[4]	0.13 ± 0.03	0.08 ± 0.02		0.17 ± 0.03		0.28 ± 0.04
F_{ST}[9]					0.23	0.43
			mtDNA[10]			
G_{ST}[3]	0.79	0.79	0.75		0.22	0.78
θ[11]	0.78 ± 0.13	0.79 ± 0.14)	0.77 ± 0.06		0.21	0.78
Φ_{ST}[8]	0.90	0.91	0.95		0.26	0.91
			cpDNA[12]			
G_{ST}[3]	-[13]	-	0.88		0.84	0.93
θ[11]	-	-	0.82		0.82	0.93
N_{ST}[14]	-	-	0.98		0.88	0.99

[1] Based on all *P. muricata* populations.
[2] Based on five *P. muricata* populations (without the San Vicente population) that were shared by RAPD and allozyme studies.
[3] G_{ST} (Nei 1986) unbiased for both sample size and population number and calculated using GENESTAT-PC (Lewis 1994).
[4] G_{ST} (Nei 1986) averaged over polymorphic loci (Hamrick and Godt 1989) using GENESTAT-PC (Lewis 1994).
[5] F_{ST} (Wright 1931) averaged over polymorphic loci using RAPDFST (Black 1996).
[6] θ (Weir and Cockerham 1984) estimated using RAPDFST (Black 1996).
[7] F_{ST} (Lynch and Milligan 1994) calculated using RAPDFST (Black 1996).
[8] Φ_{ST} (Excoffier *et al.* 1992) calculated using WINAMOVA (Excoffier 1993).

[9] F_{ST} (Wright 1978) calculated when all species and populations were analyzed hierarchically using BIOSYS-1 program (Swofford and Selander 1989).
[10] For complete mtDNA data see in Wu *et al.* (1998).
[11] θ (Weir and Cockerham 1984) calculated using Genetic Data Analysis program, GDA (Lewis and Zaykin 1996).
[12] CpDNA restriction site data were obtained from Hong *et al.* (1993).
[13] "-" = No polymorphism to allow estimates.
[14] N_{ST} calculated using HAPLO program (Lynch and Crease 1990).

was substantially increased (G_{ST} = 0.60) (Table 3.3), and differentiation among species incomplex was also high (G_{ST} = 0.42). When all species and populations were analyzed hierarchically using BIOSYS-1 (Swofford and Selander 1989), F_{ST} among species was decreased to 0.31 (F coefficients were 0.39 for populations within species and 0.58 for populations relative to the total complex).

Allozyme genetic diversity and differentiation

Diversity: Allozyme diversity within populations (H_E) averaged for each species ranged from 0.11 to 0.14 (Table 3.2). Similar to RAPDs, the estimates of genetic diversity (H_E = 0.06 to 0.15) were highly variable between different populations. However, unlike RAPDs, the population sample sizes were also highly variable. Averaged over all populations, the estimates of A, A_E, P_{95}, and H_E were 1.58, 1.19, 42.45% and 0.11, respectively. Similar to RAPDs, allozyme diversity was highest in the populations of *P. radiata*. Total diversities (H_T = 0.13 to 0.15) within species were very similar among the three species.

Differentiation: Allele frequencies varied greatly among species and among populations within species (data not shown). For example, allele $Lap2^{100}$ frequencies were high in *P. attenuata*, low in *P. radiata* and moderate in *P. muricata*. $Got1^{114}$ was present only in *P. muricata*, while Got^{114} and Got^{157} were observed only in *P. radiata*. All three species had geographic trends in allelic variation. Nei's (1986) G_{ST} was 0.22 for *P. attenuata*, 0.12 for *P. radiata*, and 0.32 for *P. muricata* (Table 3.3).

Differentiation among all populations in the three species complex was 0.44 and the differentiation among species was 0.31. When the complex was analyzed hierarchically using the BIOSYS-1 program (Swofford and Selander 1989), F_{ST} among species was reduced to 0.23, while F_{ST} coefficients were 0.25 for populations within species and 0.43 for populations relative to the total complex (Table 3.3).

Simulation analysis: Estimates of population gene diversity and differentiation were obtained from allozyme allele frequencies inferred from both the codominant multiallelic data set and the transformed dominant biallelic data set in simulated samples of different population sizes (Figure 3.2). Diversities within populations and species were substantially lower for the simulated dominant data set than for the codominant data set over a large range of sample sizes. At our average sample size of approximately 25, the dominant markers were 36 to 49 percent lower in their mean diversity values (H_S) within populations for each species. The simulated bias in population differentiation values (G_{ST}) for each species depended strongly on sample size, decreasing as sample size increased. At our sample size of 25, a 20 to 30 percent increase of G_{ST} values were predicted (Figure 3.2). Comparison of genetic diversity and differentiation between RAPDs and allozymes

Both diversity and differentiation for RAPDs were higher than for allozymes (Tables 3.2, 3.3). Averaged within-population diversity over all populations for RAPDs was nearly 30% higher than those for allozymes (0.15 vs. 0.11; Table 3.2). Total diversity within species averaged 0.21 over three species for RAPDs, more than one-third greater than for allozymes (0.13). Population differentiation within each species (G_{ST} = 0.26 to 0.41) for RAPDs was substantially larger than observed for allozymes

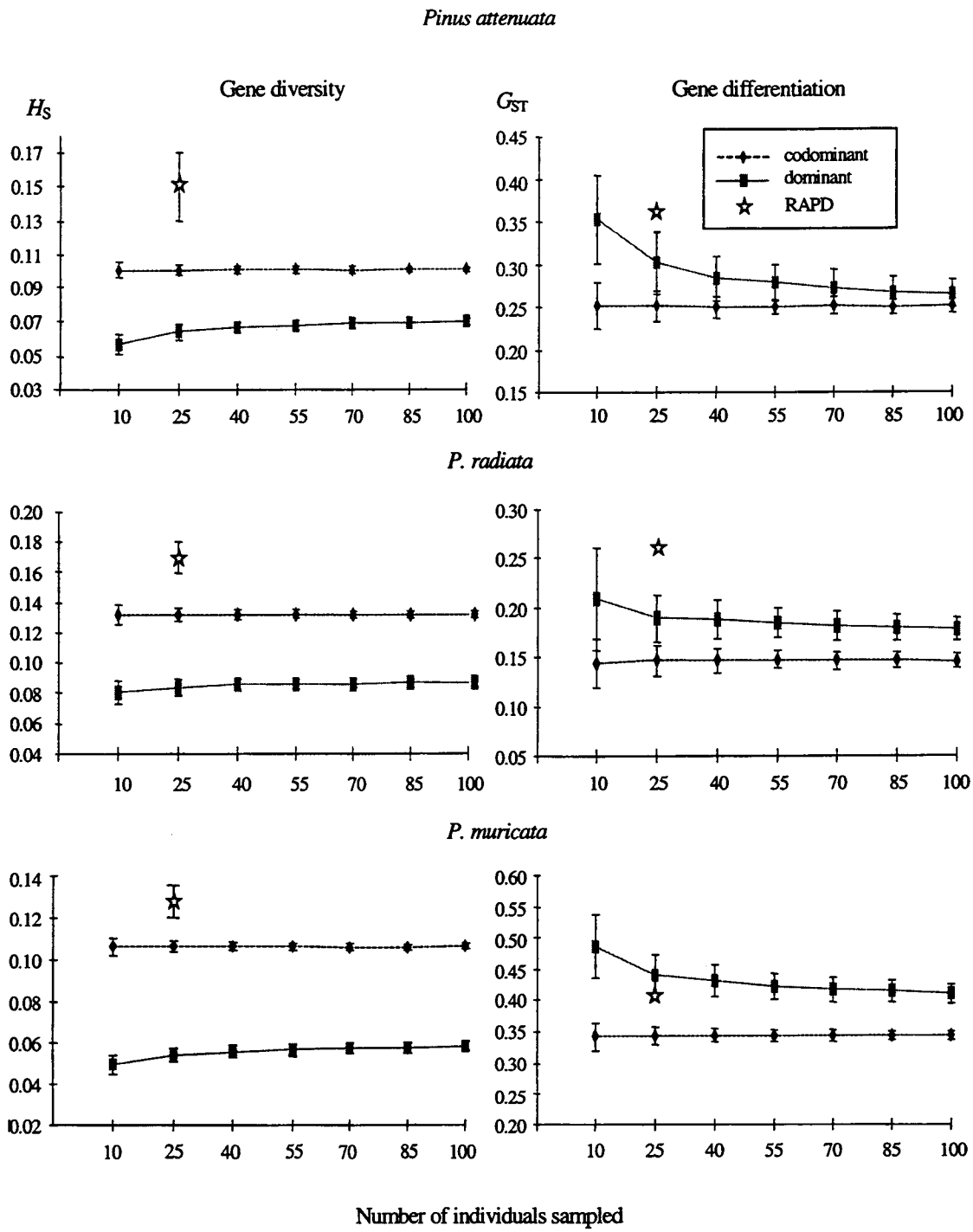


Figure 3.2 Genetic diversity (H_S) and differentiation (G_{ST}; Nei 1986) values averaged over populations of each California Closed-Cone Pine species for codominant multiallelic allozyme and dominant biallelic markers simulated in the samples of different sizes. Standard deviations (error bars) were calculated from the variance among 200 resamples simulated for each species. Observed RAPD values are also shown as a star.

(G_{ST} = 0.12 to 0.32; Table 3.3). Differentiation among species also differed between RAPDs and allozymes (0.42 vs. 0.31, respectively).

When estimates of within population diversity (H_S) for RAPD markers were compared to those based on the dominant data set that we simulated based on allozymes, the diversities were more than twice as high as those of transformed allozymes at $N = 25$ (Table 3.2, Figure 3.2). However, observed RAPD population differentiation (G_{ST}) was similar to that based on the transformed data set. RAPD estimates were either moderately higher than those for the transformed allozymes (G_{ST} = 0.36 vs. 0.30 for *P. attenuata*; 0.26 vs. 0.19 for *P. radiata*), or slightly lower (0.41 vs. 0.44 in *P. muricata*).

Phylogenetic relationships

UPGMA and Neighbor-Joining (NJ) cluster analyses were carried out to estimate phylogenetic relationships among the 12 (allozymes) or 13 (RAPDs) populations using genetic distance matrices. For RAPDs, phenogram topologies were very similar regardless of whether distance matrices were derived from allele frequencies or direct phenotypic data (presence/absence), except that the *P. attenuata* subgroup and the group including both *P. attenuata* and *P. radiata* species from NJ were better supported by bootstrap values using the latter data set (bootstrap values of 93% and 94%, respectively).

In both UPGMA and NJ analyses of RAPDs, 13 populations were clustered into three main groups, clearly corresponding to the three species. *P. attenuata* and *P. radiata* were more similar to one another (D = 0.20; Table 3.4) than were *P. attenuata* to *P. muricata* (D = 0.27) or *P. radiata* to *P. muricata* (D = 0.32), and they were considered a monophyletic group at high confidence in UPGMA (100%; data not shown) and modest confidence in NJ (73%) trees.

Table 3.4 Nei's (1978) unbiased genetic distance within and among species based on allele frequencies of allozyme and RAPD markers. The values in the parentheses are the ranges of genetic distance between populations within species

Species	*P. attenuata*	*P. radiata*	*P. muricata*
	RAPDs (76 loci)		
P. attenuata	0.070 (0.031-0.095)		
P. radiata	0.204 (0.178-0.233)	0.050 (0.029-0.067)	
P. muricata	0.271 (0.166-0.337)	0.319 (0.231-0.381)	0.111 (0.037-0.209)
	Allozymes (32 loci)		
P. attenuata	0.035 (0.009-0.054)		
P. radiata	0.144 (0.089-0.222)	0.021 (0.013-0.030)	
P. muricata	0.143 (0.071-0.249)	0.122 (0.077-0.179)	0.062 (0.034-0.112)

Among populations, the mainland populations (Año Nuevo and Cambria) of *P. radiata* were closer to each other than to the island population (Guadalupe) based on RAPD data. The Santa Ana population of *P. attenuata* was most distinct from the other, northern populations. The three regional races of *P. muricata* formed proximal phylogenetic groups (South = San Vicente and Santa Cruz; Central = Marin and Monterey; North = Mendocino and Trinidad). In all cases the most southern population(s) of each species were most basal (Santa Ana, Guadalupe, and San Vicente). In the genetically distinct populations of *P. muricata*, the populations were arrayed in a clear south (basal) to north (derived) phylogenetic sequence. UPGMA and NJ phenograms based on allozymes had many minor differences from those based on RAPDs, though all were supported by low to modest bootstrap values in the NJ tree (Figures 3.3, 3.4).

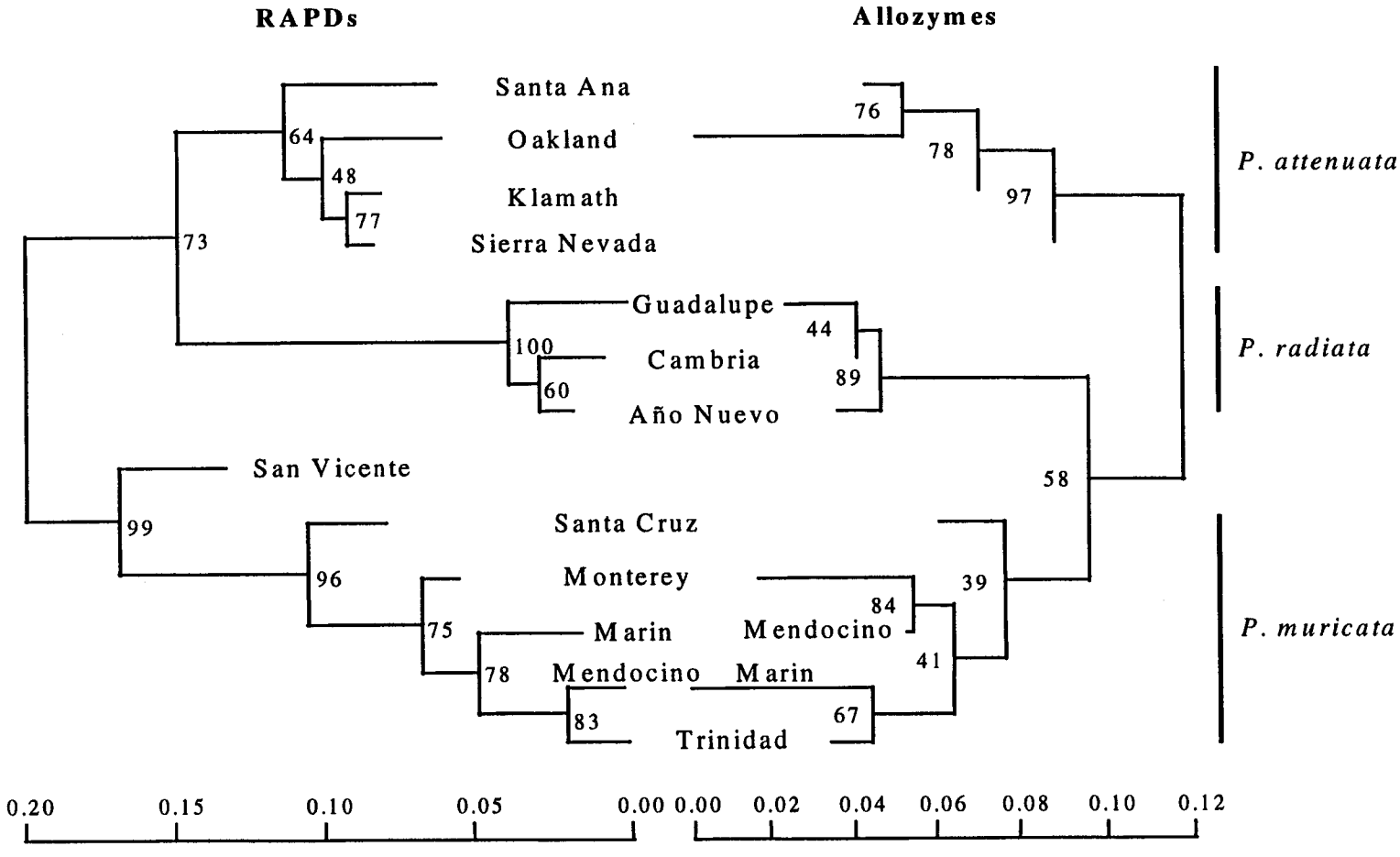


Figure. 3.3 Neighbor-Joining trees based on Nei's (1972) genetic distance among populations of three California Closed-Cone Pine species estimated from allele frequencies of 76 RAPD and 32 allozyme loci. Numbers near relevant nodes are percentage of 200 bootstrap replicates.

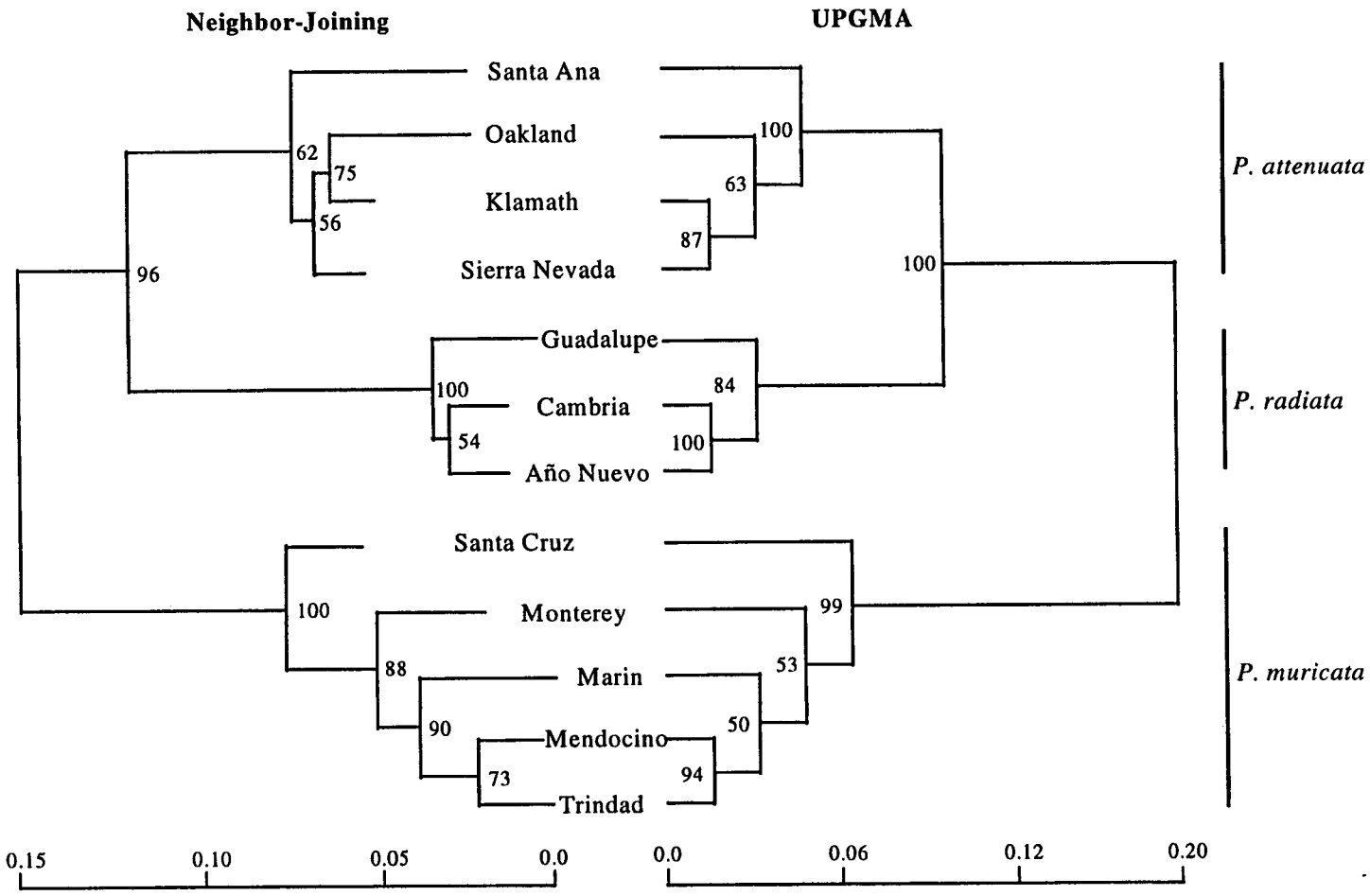


Figure 3.4 Neighbor-Joining and UPGMA trees based on Nei's (1972) genetic distance among populations of three California Closed-Cone Pine species estimated from allele frequencies of both 76 RAPD and 32 allozyme loci. Numbers near relevant nodes are percentage of 200 bootstrap replicates.

When the RAPD loci were combined with the allozyme loci for phylogenetic analysis, both UPGMA and NJ phenograms (Figure 3.4) were similar to those based only on RAPD loci (Figure 3.3). The main differences were that most clusters were supported at higher bootstrap levels. Most notably, *P. attenuatta* and *P. radiata* formed a monophyletic group at high confidence (96-100%) in NJ and UPGMA trees, respectively.

DISCUSSION

Homology of comigrating RAPD fragments

A large majority of comigrating RAPD markers among CCCP were homologous, so similarity of fragment size appears to be reasonably reliable predictor of homology, at least among closely related populations or species. These results are similar to those recently published by Rieseberg (1996) who found that 200 out of 220 comigrating fragments (91%) were homologous among three species of sunflower (*Helianthus*) based on Southern hybridization and/or restriction enzyme analysis. Williams *et al.* (1993) also found nine out of 10 comigrating fragments were homologous among several *Glycine* species from Southern hybridization. Lannér-Herrera *et al.* (1996b) observed an even higher degree of homology in 10 wild *Brassica* species, as only three of 250 RAPD comigrating markers tested did not cross-hybridize as expected. A lower level of homology, however, was reported by Thormann *et al.* (1994), who revealed that three of 15 comigrating fragments were not homologous among six *Brassica* species, although all fragments showed the expected hybridization within species. Thus, it appears that most RAPD markers can be used for analysis of closely related species. However, homology tests are very helpful for excluding some doubtful RAPD fragments such as those with large differences in intensity or slight differences in mobility. Without these tests, it is advisable to exclude any markers whose phenotypes are not identical.

Genomic origin of RAPD fragments

Based on Southern hybridization with organellar DNAs as probes, we found only one mtDNA and two cpDNA fragments among more than 130 tested in these pines. This is

consistent with the results obtained by Lu *et al.* (1995) and Heun and Helentjaris (1993), who did not detect any RAPD fragments of cytoplasmic origin in *Pinus sylvestris* L. and *Zea mays* by segregation analysis of F_1 progeny, respectively. However, Lorenz *et al.* (1994) found that a number of fragments in "genomic" RAPD patterns of *Beta vulgaris L.* originated from mtDNA and to a lesser extent from cpDNA. In *Brassica,* results of Southern hybridization indicated that among the 142 scored fragments, seven (4.9%) were of mitochondrial origin, and one (0.7%) was of chloroplast origin (Thormann *et al.* 1994). The most extreme case occurred in Douglas-fir, where 45% of scored RAPD markers showed maternal inheritance and were considered to be mtDNA-derived fragments (Aagaard *et al.* 1995). However, based on Southern analysis, Aagaard *et al.* (1998b) recently re-estimated the mtDNA value to be 16%, with 3% of chloroplast origin.

Variation in the frequency of cytoplasmic fragments in RAPD profiles among species could be explained by variable contributions of organellar genomes to total cellular DNA among species (Heun and Helentjaris 1993; Williams *et al.* 1990). This can result from varying cellular states of development at the time of DNA extraction and variation in relative genome sizes. In addition, there will be varying numbers of inverted repeat sequences (necessary for single primer amplifications) in cytoplasmic genomes relative to nuclear genomes (Lorenz *et al.* 1994). This would be expected to range widely, depending greatly on relative genome sizes and complexity. Finally, values based on segregation are expected to provide only a minimum estimate because of the modest levels of polymorphism expected for plant organelle genomes compared to nuclear genomes.

Segregation analysis

To eliminate organellar and unreliable fragments, we studied segregation among haploid megagametophytes. More than 40 loci segregated among megagametophytes of several hybrid and/or natural cones for each species. Lu *et al.* (1995) studied eight parents and 80 F_1 progeny of *Pinus sylvestris*, and found that the segregation of all polymorphic fragments was consistent with the biparental, diploid mode of inheritance expected for a dominant trait. Similar results have also been reported in other organisms (e.g., Heun and

Helentjaris 1993; Kazan *et al.* 1993; Krutovskii *et al.* 1998a). However, the proportion of Mendelian loci expressed as a fraction of the total number of amplification products can be as low as 33% (*Picea abies*, Bucci and Menozzi 1993). Because of the rarity of suitable trees or crosses, selection of loci based on segregation analysis is difficult for weakly polymorphic or monomorphic loci. Therefore, segregating loci will rarely represent a random set of RAPD loci, and can bias the estimation of genetic diversity.

RAPDs vs. allozymes

The genetic data obtained at both RAPD and allozyme loci allowed us to make a direct comparison of genetic diversity within populations, and differentiation among populations, for these two types of nuclear loci. RAPD markers generally detected more intrapopulation diversity than did allozymes in terms of effective number of alleles per locus, proportion of polymorphic loci, and expected heterozygosity, but not for number of alleles per locus. Allozymes are expected to have a larger number of alleles per locus because many allozyme loci are multiallelic, although in most cases only one or two alleles are common and others are rare. Population and species differentiation were also significantly higher for RAPDs than for allozymes.

However, dominance of RAPDs is expected to bias estimates of null allele frequency when the null homozygotes are rare within populations (Lynch and Milligan 1994). For example, only one null homozygote per 25 individuals is expected even when null allele frequency is 0.20. Such rare homozygotes can go easily undetected in small samples, biasing frequency estimates downward and potentially affecting estimates of genetic diversity and differentiation. When there is an excess of heterozygotes and Hardy-Weinberg equilibrium is violated, the bias will even be greater. Szmidt *et al.* (1996) inferred allele frequencies of RAPDs directly for maternal trees from segregation patterns in haploid megagametophytes, and indirectly estimated allele frequencies based on frequency of null homozygotes among diploid phenotypes converted from haploid material in Scots pine, *Pinus sylvestris*. The comparison of genetic parameters based on these two sets of allele frequencies indicated that genetic diversity within populations was substantially higher for direct RAPDs than for indirect RAPDs (H_S = 0.37 vs. 0.22),

whereas differentiation among populations was three-fold higher for indirect RAPDs than for direct RAPDs (G_{ST} = 0.06 vs. 0.02). Not surprisingly, dominance of RAPDs can affect calculation of genetic parameters, although the small sample sizes and low number of loci in this study may have increased the extent of bias. A lower expected heterozygosity, and inflated among-population differentiation, were also observed with predicted diploid phenotypes for RAPDs in black spruce, *Picea mariana* (Isabel *et al.* 1995).

The simulation analysis allowed us to estimate the effect of RAPD dominance and biallelism and thus make a more accurate comparisons of genetic parameters between RAPDs and allozymes. Genetic diversities within populations based on allozyme allele frequencies that were transformed into biallelic dominant data were significantly lower than for nontransformed multiallelic codominant data, while population differentiation was biased upwardly in each species (Figure 3.2). Nonetheless, RAPD markers detected more than twice the intrapopulation diversity than the transformed dominant allozymes. Population differentiation was similar (*P. muricata*) or only moderately larger (*P. radiata* and *P. attenuata*) for RAPDs compared to transformed allozymes. This suggests that RAPDs are predictably underestimating diversity, yet should be more sensitive markers for detecting variation in the level of genetic diversity than are allozymes. However, comparisons of population differentiation between RAPDs and allozymes can be seriously biased in a complex manner that depends on allele frequencies and sample sizes.

The higher population diversity and differentiation of RAPDs in this study are consistent with the results reported in other organisms. Liu and Furnier (1993) found that RAPD diversity (H_S) is much higher than allozymes and RFLPs within populations of bigtooth aspen (*Populus grandidentata*) (H_S = 0.31 for RAPDs, 0.08 for allozymes and 0.13 for RFLPs), despite the dominance of RAPDs. Diversity within populations of *Pinus sylvestris* is also substantially higher for haploid-based RAPDs than for allozymes (H_S = 0.36 vs. 0.26). Peakall *et al.* (1995) transformed allozyme data into binary presence-absence (1,0) vectors to enable comparison to RAPDs and then computed distance matrices for both markers and analyzed them using AMOVA (Excoffier *et al.* 1992). RAPD markers provided higher population differentiation than allozymes (58.4% vs. 45.2%).

The greater population diversity and divergence for RAPDs than for allozymes may result from several causes (summarized in Aagaard *et al.* 1995). First, only functional genes encoding soluble enzymes are studied for allozymes, whereas RAPDs anneal randomly to inverted repetitive DNA, and are expected to reflect primarily random variation of noncoding repetitive DNA. It is well known that the repetitive component of genomes can diverge and change in sequence rapidly (Cabot *et al.* 1993; Charlesworth *et al.* 1994). Second, allozymes often fail to detect many kinds of genetic variation because only nucleotide differences that lead to electrophoretically detectable changes in amino acid composition of the gene product are observed. RAPDs, however, should be sensitive to different types of mutations, including point mutations at primer annealing sites, inversions flanking the annealing sites, secondary structure constraints, and insertions that cause a greater length than can be amplified with routine PCR techniques. Third, it has been hypothesized (e.g.,, Black *et al.* 1992) that highly repetitive regions of genomes such as telomeres and minisatellites evolve more rapidly than allozyme loci, and are preferably represented in RAPD fragments. Finally, allozymes may be subject to stronger natural selection since they represent functional gene products. Directional selection may reinforce the action of genetic drift in reducing the variability at protein level (Baruffi *et al.* 1995; Begun and Aquadro 1993). It has also been suggested that allozymes are under the influence of balancing selection (Altukhov 1990, 1991; Karl and Avise 1992), which helps to maintain intrapopulation polymorphism, but could reduce population differentiation.

Population differentiation based only on loci selected by segregation analysis was similar to those based on all RAPD loci, though the genetic diversities were inflated. Thus, segregation analysis could be used to select nuclear RAPD markers to estimate differentiation when enriched organellar DNA is unavailable for excluding RAPD markers of cytoplasmic origin.

For diploid data, Lynch and Milligan (1994) recommended that unbiased estimates of population genetic parameters can be obtained with RAPDs if the analysis is limited to loci with dominant alleles that are not prevalent (recessive null-allele homozygote occurs not less than in three individuals). This restriction can exclude many population-specific

loci and highly differentiated loci. Loci that are polymorphic in some populations or species, but monomorphic in others, will also be excluded. It is therefore not surprising that our estimates of population differentiation based on this limited set of loci were much lower compared to those based on all loci (Table 3.1). Szmidt *et al.* (1996) also found that population differentiation was substantially underestimated based on loci selected by the criteria of Lynch and Milligan (1994) compared to estimates from all RAPD loci.

The degree of population differentiation for nuclear markers (RAPDs and allozymes) was also compared with that for cytoplasmic DNA RFLP markers (mitochondrial and chloroplast DNAs) in the same groups (Table 3.3). The levels of population subdivision for mtDNA length mutations and cpDNA site mutations greatly exceeded those observed with RAPD and allozyme markers. This is consistent with the theoretical predictions for organelle genomes, which have small effective population size, low sequence mutation rate, and limited gene flow for maternally inherited mtDNA and/or paternally inherited cpDNA.

Phylogenetic analysis

Cluster analysis of RAPD markers resulted in unambiguous distinction of the three species, *P. attenuata, P. radiata* and *P. muricata*. These results are in agreement with the allozyme study of Millar *et al.* (1988), where three species were distinguished clearly and each species was about equally differentiated from the other two. In their results there was no obvious phylogenetic affinity between pairs of species. However, our NJ analysis of the joint RAPD-allozyme data provided strong support for a common ancestor for *P. attenuata* and *P. radiata*. CpDNA analysis, on the other hand, divided the CCCP into three major groups: 1) a very strongly supported clade with *P. attenuata*, 2) a weakly supported clade with both *P. radiata* and the southern race of *P. muricata*, and 3) a very strongly supported clade with the northern and intermediate races of *P. muricata* (Hong *et al.* 1993b). Rates of cpDNA evolution were also highly variable. It appears that the chloroplast and nuclear genomes have different phylogenetic origins in the CCCP, a result also reported in several other plant taxa (Rieseberg and Soltis 1991).

Geographic clines of RAPD were apparent within the three species. RAPD relationships of *P. attenuata* populations were similar to allozyme patterns (Millar *et al.* 1988) and to latitudinal clines observed in growth and morphological characteristics (Newcomb 1962). Clinal patterns of variation in the NJ tree appear to follow a gene flow path from Santa Ana northward along Sierra Nevada through the Klamath mountains, and then southward along the coast mountain ranges. The large genetic distance of the Oakland population from the other three northern populations is likely to have forced it to its basal positions among those populations in the UPGMA tree (Figure 3.4).

In *P. radiata*, RAPD analysis indicated that the two mainland populations are more closely related to each other than to the Guadalupe Island population. This is also consistent with allozyme variation (Millar *et al.* 1988; Moran *et al.* 1988; Plessas and Strauss 1986), and many morphological and biochemical traits such as needle number, stem form, seed proteins, and terpenes (reviewed in Millar 1986). Based on cone morphology (Axelrod 1980) and allozymes, the island populations were inferred to be most ancestor-like, consistent with our RAPD results.

Separate taxonomic designations have been proposed for the northern and southern groups of *P. muricata* based on studies of a number of traits, including morphology (Fielding 1961; Shelbourne *et al.* 1982), monoterpene composition (Forde and Blight 1964; Mirov *et al.* 1966), and crossability (Critchfield 1967; Millar and Critchfield 1988). They are also supported by frequencies at some specific allozyme loci (Millar 1983; Millar *et al.* 1988) and cpDNA restriction site variation (Hong *et al.* 1993b). However, both groups still clustered tightly within *P. muricata* based on allozymes, and genetic distances between groups were not much higher than among distant populations in *P. attenuata* and *P. radiata*. Phenetic analysis of RAPD data in the present study confirms that populations of *P. muricata* are monophyletic though they are highly differentiated. Both RAPDs and allozymes indicate clinal patterns of divergence from north to south, with the southern populations appearing most ancestral (similar to outgroups) and the northern mainland populations being most divergent (Millar *et al.* 1988). The several phylogenetic discrepancies between RAPDs and allozymes in our analysis are very likely

to be the result of small sample size in several populations in the allozyme data set (Millar *et al.* 1988).

The broad relationships of populations and species based on RAPDs agrees well with those based on allozymes (Millar *et al.* 1988) and many morphological and biochemical traits. This suggests that RAPDs are appropriate molecular markers for phylogenetic analysis at this level. RAPD markers have been widely used in the reconstruction of phylogenetic relationships for many organisms (reviewed by Wolfe and Liston 1997), and there has been general concordance among the results derived from RAPDs and other techniques. These and our own results confirm that despite the complications of dominance and phenotype interpretation, for which RAPDs are infamous, the large number of loci resolved can make them valuable phylogenetic markers within and between closely related species.

ACKNOWLEDGMENTS

We wish to thank Tony Cario for sampling the San Vicente population of *P. muricata*, Nathan Strauss for his help in field collections; Jan Aagaard for his help in field collections and extraction of mitochondrial DNA; Bill Libby for his advice and help in accessing the study populations; Jenny Aitken-Christie of Carter Holt Harvey; New Zealand for providing and transporting embryogenic cultures of *P. radiata*; Tom Conkle and David Johnson for collecting hybrid cones; Sheila Vollmer for her technical assistance; Bob Westfall for help with our reanalysis of allozyme data; Steve DiFazio for help editing the manuscript; and the National Science Foundation (NSF Conservation and Restoration Biology, DEB-9300083) for grant support.

CHAPTER 4

CONCLUSIONS

SUMMARY OF KEY RESULTS

The key results of this thesis are:

1. The majority of mtDNA diversity resided within species rather than among species ($G_{ST} = 0.21$).
2. Based on haplotype frequency of mtDNA-RFLPs, relatively high haplotype diversity was observed within populations ($H_S = 0.22$), which was significantly larger than that obtained from cpDNA-RFLPs ($H_S = 0.06$).
3. MtDNA revealed strong population subdivision within species (average $G_{ST} = 0.78$), which was in contrast with nuclear DNA markers such as allozymes (average $G_{ST} = 0.22$), but comparable with cpDNA markers of bishop pine ($G_{ST} = 0.88$).
4. Probe-enzyme based mtDNA showed substantially less genetic diversity within populations ($H_S = 0.03$), but even higher differentiation among populations ($G_{ST} = 0.91$).
5. Species and most of the populations can be readily distinguished by their unique haplotypes.
6. Pine mtDNAs were not suitable for reconstruction of phylogenetic relationships due to a high frequency of homoplasy and convergent evolution.
7. More than 90% of tested comigrating RAPD markers were highly homologous among populations and species.
8. Few fragments (2 for cpDNA and 1 for mtDNA) were identified to be of organellar DNA origin among 142 total RAPD fragments surveyed.
9. RAPD markers detected higher levels of genetic diversity within populations and differentiation among populations and species than did allozymes.
10. Simulation analyses indicated that the RAPD diversity within populations is likely to be greatly underestimated (36 to 49%) due to its dominance and biallelism. However,

its level of differentiation among populations is likely to be moderately overestimated (20 to 30%).

11. RAPDs are reliable molecular markers for the inference of species and population relationships.

MAJOR CONCLUSIONS

The major conclusions of this thesis are:

1. MtDNA markers are powerful for detecting population differences, but are inappropriate for phylogenetic analysis due to a high degree of homoplasy and convergent evolution.
2. MtDNA markers are very effective for distinguishing populations and species for plant conservation and breeding programs.
3. RAPD products are primarily of nuclear origin in the three California Closed-Cone Pines.
4. RAPD markers detected higher levels of genetic diversity within populations, and differentiation among populations and species than allozymes, but dominance and biallelism of RAPDs can lead to underestimates of within population and overestimates of between population differentiation. Thus, RAPD loci may possess more than twice as much twice population diversity than do allozymes loci.
5. RAPDs are effective markers for phylogenetic analysis.

RECOMMENDATIONS FOR FUTURE RESEARCH

MtDNA-RFLP markers

Strong population differentiation was revealed by mtDNA-RFLP markers, suggesting they would be highly effective genetic markers for tree improvement studies to identify seeds of unknown origin, and genetic conservation programs to select populations and provenances. Based on this study, most of the populations could be distinguished by their unique haplotypes or combination of just a few RFLP fragments. Future work could

include more restriction enzymes and/or probes to make it possible to distinguish every population definitively. The value of mtDNA-RFLP markers to differentiate populations should also be studied in other species.

RAPD markers

In this study, only three organelle DNA markers were identified among RAPD profiles by Southern hybridization. However, in a study of Douglas-fir (Aagaard 1997), mtDNA markers are fairly frequent among RAPD fragments, severely affecting the estimation of genetic parameters. The methods of identifying organelle origin of RAPD markers could be readily applied to other studies if enriched organellar DNAs are available. A survey of a variety of plant species would help us to better understand this phenomenon and its mechanism.

Future analysis of RAPD patterns of parent species, hybrids, and hybrid progenies could generally distinguish nuclear and cytoplasmic origin of RAPD markers. This will not only confirm the results of Southern hybridization with organelle DNA probes, but also determine the efficiency and validity of Southern analysis.

Megagametophyte segregation analysis would allow direct comparison of genetic parameters based on RAPDs and allozymes (Isabel et al. 1995; Szmidt et al. 1996). Segregation patterns among megagametophytes would provide direct estimates of allele frequencies for both RAPDs and allozymes. Thus, the conclusions based on simulation analyses in the current study could be empirically tested.

Microsatellites or simple sequence repeats (SSRs) are powerful new genetic markers that are being more and more extensively used in genome mapping and population genetic studies (e.g., Chase et al. 1996; Innan et al. 1997). They provide a substantial level of polymorphism, which is characterized as a large number of alleles per locus. Microsatellites are also codominant markers and can be easily scored and assayed using PCR-based methods. Thus it would be very useful to include microsatellite makers in our further studies to compare their results with RAPDs and allozymes.

BIBLIOGRAPHY

Aagaard, J. E., S. S. Vollmer, F. C. Sorensen, and S. H. Strauss. 1995. Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. Mol. Ecol. 4: 441-447.

Aagaard, J. E. 1997. Genetic diversity and differentiation in Douglas-fir from RAPD markers of nuclear and mitochondrial origin. M. S. Thesis, Oregon State University, Corvallis, Oregon.

Aagaard J. E., K. V. Krutovskii, and S. H. Strauss, 1998a. RAPDs and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. Heredity 80, in press.

Aagaard J. E., K. V. Krutovskii, and S. H. Strauss, 1998b. RAPD markers of mitochondrial origin exhibit lower population diversity and higher differentiation than RAPDs of nuclear origin in Douglas-fir. Mol. Ecol. 7, in press.

Altukhov, Yu. P., 1990. Population genetics: diversity and stability. Harwood Academic Publishers, London.

Altukhov, Yu. P., 1991. The role of balancing selection and overdominance in maintaining allozyme polymorphisms. Genetica 85: 79-90.

Apostol, B. L., W. C. Black IV, B. R. Miller, P. Reiter, and B. J. Beaty. 1993. Estimation of family members at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor. Appl. Genet. 86: 991-1000.

Avise, J. C., B. W. Bowen and T. Lamb. 1984. DNA fingerprints from hypervariable mitochondrial genotypes. Mol. Biol. Evol. 4: 258-269.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Niegel, C. A. Reeb and N. C. Saunders. 1987. Intraspecific Phylogeography: the mitochondrial bridge between population genetics and systematics. Annals of the Missouri Botanical Garden 82: 247-277.

Avise, J. C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Annual Review of Genetics 25: 45-69.

Avise, J. C. 1994. Molecular markers, Natural History and Evolution. Chapman and Hall, New York.

Axelord, D. I. 1980. History of the Maritime Closed-Cone Pines, Alta and Baja California. Univ. Calif. Press. 143pp.

Axelord, D. I. 1983. New Pleistocene Conifer Records, Coastal California. Univ. Cal. Publ. Geol. Sci. 127: 1-108

Backeljau, T. L. D. Bruyn, H. D. Wolf, K. Jordaens, S. V. Dongen, R. Verhagen, and B. Winnepenninckx. 1995. Random amplified polymorphic DNA (RAPD) and parsimony methods. Cladistics 11: 119-130.

Bannister, M. H., A. Williams, I. McDonald, and M. B. Forde. 1962. Variation of turpentine composition in five population samples of *Pinus radiata*. N. Z. J. Sci. 5: 486-495.

Baruffi, L., G. Damian, C. R. Guglielmino, C. Bandis, A. R. Malacrida, and G. Gasperi. 1995. Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. Heredity 74: 425-437.

Begun, D. J. and C. F. Aquadro. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365: 548-550.

Belhassen, E., A. Atlan, D. Couvet, P. H. Gouyon, and F. Quetier. 1993. Mitochondrial genome of *Thymus vulgaris* L. is highly polymorphic between and among natural populations. Heredity 71: 462-472.

Birky, C. W. Jr. 1988. Evolution and variation in plant chloroplast and mitochondrial genomes. In L. D. Gottlieb, and S. K. Jain (eds.). Plant Evolutionary Biology, pp. 23-53. Chapman and Hall, London.

Birky, C. W. Jr., P. Furest, and T. Maruyama. 1989. Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121: 613-627.

Black IV W. C., N. M. DuTeau, G. J. Puterka, M. J. Nechols, and J. M. Pettorini. 1992. Use of the random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) to detect DNA polymorphism in aphids (*Homoptera*: Aphididae). Bull. Entomol. Res. 82: 151-159.

Black IV W. C. 1996. RAPDDIST 1.0, RAPDFST 4.0, RAPDPLOT 2.4. Department of Microbiology, Colorado State University, Fort Collins, Colorado (ftp lamar.colostate.edu in pubwcb4).

Boscherini, G., M. Morgante, P. Rossi and G. G. Vandramin. 1994. Allozyme and chloroplast DNA variation in Italian and Greek population of *Pinus leucodermis*. Heredity 72: 284-290.

Bradshaw, H. D., S. M. Wilbert, K. G. Otto and D. W. Schemske. 1995. Genome mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). Nature 376: 762-765.

Brown, A. H. D., and D. J. Schoen. 1992. Plant population genetic structure and genetic conservation. In O. T. Sandlund, K. Hindar, and A. H. D. Brown (eds.). Biodiversity for Sustainable Development, pp. 88-104. Scandanavian University Press, Oslo.

Bucci, G. and P. Menozzi. 1993. Segregation analysis of random amplified polymorphic DNA (RAPD) markers in *Picea abies* Karst. Mol. Ecol. 2: 227-232.

Buren, R. V., K. T. Harper, W. R. Andersen, D. J. Stanton, S. Seyoum and J. L. England. 1994. Evaluating the relationship of autum buttercup (*Ranunculus acriformis var. aestivalis*) to some close congeners using random amplified polymorphic DNA.

Byrne, M. and G. F. Moran. 1994. Population divergence in the chloroplast genome of *Eucalyptus nitens*. Heredity 73: 18-28.

Cabot, E. L., P. Doshi, M. L. Wu and C. I. Wu. 1993. Population genetics of tandem repeats in centrometric heterchromatin - unequal crossing over and chromosomal divergence at the responder locus of *Drosophila melanogaster*. Genetics 135: 477-487.

Chalmers, K. J., R. Waugh, J. I. Sprent, A. J. Simons and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity 69: 465-472.

Charlesworth, B. P. Sniegowski and W. Stephen. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215-220.

Chase, M., R. Kessel and K. Bawa. 1996. Microsatellite markers for population and conservation genetics of tropical trees. American Journal of Botany 83(1): 51-57.

le Corre, V., S. Dumolin-Lapègue and A. Kremer. 1997. Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. Mol. Ecol. 6: 519-529.

Crease, T. J., M. Lynch and K. Spitze. 1990. Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of *Daphnia pulex*. Mol. Biol. Evol. 7: 444-458.

Critchfield, W. B. 1967. Crossability and relationships of the California Closed-Cone Pines. Silvae Genetica 16(3): 89-97.

Critchfield, W. B. and E. L. Little. 1966. Geographic Distribution of Pines of the World. USDA Misc. Publ. 991. 97pp.

Crowley, J. J. 1974. An analysis of nine-year growth data on the California Closed-Cone Pines. U. C. Berkely Dept. Forestry Honor's Thesis. 104pp.

Davis, S. K. 1986. Population structure and patterns of speciation in *Geomys* (*Rodentia*: Geomyidae): an analysis using mitochondrial and ribosomal DNA. Ph.D. dissertation, Washington University, St. Louis.

Davis, T. M., H. Yu, K. M. Haigis, and P. J. McGowan. 1995. Template mixing: a method for enhancing detection and interpretation of codominant markers. Theor. Appl. Genet. 91: 582-588.

Dawson, I. K., A. J. Simons, R. Waugh, and W. Powell. 1995. Diversity and genetic differentiation among subpopulations of *Gliricidia sepium* revealed by PCR-based markers.

Dean, R. and J. Arnold. 1996. Cytonuclear disequilibria in hybrid zones using RAPD markers. Evolution 50: 1702-1705.

Desalle, R., A. Templeton, I. Mori, S. Pletscher, and J. S. Johnston. 1987. Temporal and spatial heterogeneity of mtDNA polymorphisms in natural populations of *Drosophila mercatorum*. Genetics 116: 215-223.

Deu, M., P. Hamon, P. Dufour, A. D'hont, and C. Lanaud. 1995. Mitochondrial DNA diversity in wild and cultivated sorghum. Genome 38: 635-645.

Dong, J. and D. B. Wagner 1993. Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta*. Theor. Appl. Genet. 86: 573-578.

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

Doran, J. C. 1974. Variation in growth of *Pinus muricata* provenances and comparison with *Pinus radiata*. Aust. For. Res. 6(3): 19-24.

Dowling, T. E., C. Moritz, J. D. Palmer, and L. H. Rieseberg. 1996. Nucleic Acids III: analysis of fragments and restriction sites. In D. M. Hillis, B. K. Mable and C. Mortiz (eds.). Molecular Systematics, pp. 249-320. Sinauer, Sunderland, MA.

Downie, S. R. and J. D. Palmer. 1992. Use of chloroplast DNA rearrangements in reconstructuring plant phylogeny. In: P. S. Soltis, D. E. Soltis and J. J. Doyle (eds.). Molecular Systematics of Plants, pp. 14-35. Chapman and Hall, New York.

Doyle, J. J. 1992. Gene trees and species: molecular systematics as one-character taxonomy. Systematic Botany 17: 144-163.

Duffield, J. W. Interrelationships of the California Closed-cone Pines with special references to *Pinus muricata* D. Don. Ph. D. dissertation, Univ. Calif. Berkely. 114pp.

Dumolin-Lapègue, S., B. Demesure, S. Fineschi, V. le Corre and R. J. Petit. 1997. Phylogenetic structure of white oaks throughout the European continent. Genetics 146: 1475-1487

Ellsworth, D. L., K. D. Rittenhouse and R. L. Honeycutt. 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. Biotechniques 14: 214-217.

Everard, J.E. and D. F. Fourt. 1974. Monterey pine and bishop pine as plantation trees in southern Britain. Quat. Jour. For. 68(2): 111-125.

Excoffier, L., P. E. Smouse, J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 146: 1475-1487.

Excoffier, L. 1993. *WINAMOVA*. Genetics and Biometry Laboratory, University of Geneva, Carouge, Switzerland (ftp:129.194.113.13ftpcompwin or ftp:anthropologie.unige.chpubcompwin).

Felsenstein, J. 1995. PHYLIP (Phylogeny Inference Package) version 3.57c. Department of Genetics, University of Washington, Seattle, Washington.

Fielding, J.M. 1961. Provenances of monterey and bishop pines. Aust. For. Timb. Bureau Bull. 38: 1-30.

Forde, M. B. 1964. Variation in natural populations of *Pinus radiata* in California. Parts 1-4. N. Z. J. Bot. 2: 213-259; 459-501.

Forde, M. B. and Blight, M. M. 1964. Geographic variation in the turpentine of bishop pine. N. Z. J. Bot. 2: 44-52.

Furman, B. J., W. S. Dvorak, R. R. Sederoff, and D. M. O'Malley. 1996. Molecular markers as diagnostic tools to identify species, hybrids and introgression: a study of Central American and Mexican pines. In M. J. Dieters, A. C. Matheson, D. G. Nikles, C. E. Harwood, and S. M. Walker (eds.). Tree Improvement for Sustainable Tropical Forestry. Proc. OFRI-IUFRO Conf., Caloundra, Qqeensland, Australia.

Gilbert, D. A., N. Lehman, S. J. O'Brien and R. K. Wayne. 1990. Genetic fingerprinting reflects population differentiation in the California Channel Island fox. Nature 344: 764-767.

Grabau, E. A., W. H. Davis, N. D. Phelps and B. G. Gengenbach. 1992. Classification of soybean cultivars based on mitochondrial DNA restriction fragment length polymorphisms. Crop Science 32: 271-274.

Graham, J. and R. J. McNicol. 1995. An examination of the ability of RAPD markers to determine the relationships within and between *Rubus* species. Theor. Appl. Genet. 90: 1128-1132.

Gray, M. W. 1989. Origin and evolution of mitochondrail DNA. Annu. Rev. Cell. Biol. 5: 25-50.

Guinon, M., J. V. Hood and W. J. Libby. 1982. A clonal study of intraspecific variability in radiata pine. II. Growth and form. Austr. For. Res. 12: 191-201.

Haig, S. M., J. M. Rhymer, and D. G. Heckel. 1994. Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. Mol. Ecol. 3: 581-595.

Hamrick, J. L. and Godt, M. J. W. 1990. Allozyme diversity in plant species. In: A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir (eds.). Plant Population Genetics, Breeding and Genetic Resources, pp. 43-63. Sinauer Associates, Inc., Sunderland, Massachusetts.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution 4: 6-11.

Heun, M. and T. Helentjaris. 1993. Inheritance of RAPDs in F_1 hybrids of corn. Theor. Appl. Genet. 85: 961-968.

Hiesel, R., B. Combettes and A. Brennicke, 1994 Evidence for RNA editing in mitochondria of all major groups of land plants except the *Bryophyta*. Proc. Natl. Acad. Sci. USA 91: 629-633.

Hipkins, V. D. 1993. Repeated sequences associated with inversions and length mutations in the chloroplast genomes of *Pinus* and *Pseudotsuga*. Ph.D. disseration, Oregon State University, Corvallis, Oregon.

Hipkins, V. D., K. V. Krutovskii, and S. H. Strauss. 1994. Organelle genomes in conifers: structure, evolution, and diversity. Forest Genetics 4: 179-189.

Hong, Y. P. 1991. Chloroplast DNA variability and phylogeny in the California Closed-Cone Pines. Ph.D. dissertation, Oregon State University, Corvallis, Oregon.

Hong, Y. P., V. D. Hipkins and S. H. Strauss. 1993a. Chloroplast DNA diversity among trees, populations and species in the California Closed-Cone Pines. Genetics 135: 1187-1196.

Hong, Y. P., A. B. Krupkin and S. H. Strauss 1993b. Chloroplast DNA transgresses species boundaries and evolves at variable rates in the California Closed-Cone Pines (*Pinus radiata, P. muricata, and P. attenuata*). Molecular Phylogenetics and Evolution 2(4): 322-329.

Hong, Y. P., B. Ponoy, and J. E. Carlson. 1995. Genetic diversity and phylogeny in Douglas-fir based on RFLP and RAPD (DAF) analysis of nuclear, chloroplast, and mitochondrial genomes. In: P. Baradat, W. T. Adams, and G. Muller-Starck (eds.). Population Genetics and Genetic Conservation of Forest Trees, pp. 247-266. SPB Academic Publishing, Amsterdam, The Netherlands.

Hood, J. V. and W. J. Libby 1980. A clonal study of intraspecific variability in radiata pine. I. Cold and animal damage. Austr. For. Res. 10: 9-20.

Hsiao, J. Y. and L. H. Rieseberg. 1994. Population genetic structure of *Yushania niitakayamensis* (Bambusoideae, Poaceae) in Taiwan. Mol. Ecol. 3: 201-208.

Huff, D. R., R. Peakall. and P. E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dachtyloides* (Nutt.) Engelm.] Mol. Ecol. 4: 135-147.

Innan, H., R. Terauchi and N. T. Miyashita. 1997. Microsatellite polymorphism in matural populations of the wild plant *Arabidopsis thaliana*. Genetics 146: 1441-1452.

Isabel, N., J. Beaulieu, and J. Bousequet. 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc. Natl. Acad. Sci. USA 92: 6369-6373.

Jelinski, D. E., and W. M. Cheliak. 1992. Genetic diversity and spatial subdivision of *Populus tremuloides* (Salicaceae) in a heterogeneous landscape. American Journal of Botany. 79: 728-736.

Karl, S. A. and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256: 100-102.

Kazan, K., J. M. Manners, and D. F. Cameron. 1993. Inheritance of random amplified polymorphic DNA markers in an interspecific cross in the genus *Srylosanthes*. Genome 36: 50-56.

Kremer, A. and R. J. Petit. 1993 Gene diversity in natural populations of oak species. Ann. Sci. Forest. 50 (Suppl. 1): 186s-202s.

Lannér-Herrera, C., M. Gustafsson, A. S. Fält and T. Bryngelsson. 1996a. Diversity of wild *Brassica oleraceae* as estimated by isozyme and RAPD analysis. Genetic Resources and Crop Evolution 43: 13-23.

Lannér-Herrera, C., T. Bryngelsson and M. Gustafsson. 1996b. Genetic validity of RAPD markers at the intra- and inter-specific level in wild *Brassica* species with n = 9. Theor. Appl. Genet. 93: 9-14.

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

Lewis, P. O. 1994. GeneStat-PC 3.3. Department of Statistics, North Carolina State University, Raleigh, North Carolina.

Lewis, P. O. and D. Zaykin. Genetic Data Analysis (GDA). Department of Statistics, North Carolina State University, Raleigh, North Carolina.

Li, W. H. and D. Graur. 1991. Fundamentals of Molecular Evolution. Sinauer Association, Inc. Sunderland, Massachusetts.

Lidholm, J. and P. Gustafsson. 1991. The chloroplast genome of the gymnosperm *Pinus contorta*: a physical map and a complete collection of overlapping clones. Curr. Genet. 20: 161-166.

Lifante, Z. D. and I. Aguinagalde. 1996. The use of random amplified polymorphic DNA (RAPD) markers for the study of taxonomical relationships among species of *Asphodelus* sect. *Verinea* (Asphodelaceae). American Journal of Botany 83: 949-953.

Linhart, Y. B. 1978. Maintenance of variation in cone morphology in California Closed-Cone Pines: The role of fires, squirrels and seed output. Southwest Nat. 23(1): 29-40.

Liu, Z. and G. R. Furnier. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling and bigtooth aspen. Theor. Appl. Genet. 87: 97-105.

Lonsdale, D. M. 1989. The plant mitochondrial geneome. In A. Marcus (ed.). Molecular Biology, pp. 230-295. Academic Press, New York.

Lorenz, M., A. Weihe and T. Borner. 1994. DNA fragments of organellar origin in random amplified polymorphic DNA (RAPD) patterns of sugar beet (*Beta vulgaris* L.). Theor. Appl. Genet. 88: 775-779.

Lu, M. Z., A. E. Szmidt, X. R. Wang. 1995. Inheritance of RAPD fragments in haploid and diploid tissues of *Pinus sylvestris* (L.). Heredity 74: 582-589.

Luo, H., Coppenolle, B. V., Seguin, M., and Boutry, M. 1995. Mitochondrial DNA polymorphism and phylogenetic relationships in *Hevea brasiliensis*. Molecular Breeding 1: 51-63.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Mol. Biol. Evol. 7: 478-484.

Lynch, M., and T.J. Crease. 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. 7: 377-394.

Lynch, M. and B. G. Milligan. 1994. Analysis of population structure with RAPD markers. Mol. Ecol. 3: 91-99.

Marillia, E. F. and G. J. Scoles. 1996. The use of RAPD markers in *Hordeum* phylogeny. Genome 39: 646-654.

Maruyama, T. and C. W. Birky, Jr. 1991. Effect of periodic selection on gene diversity in organelle genomes and other systems without recombination. Genetics 127: 449-451.

Mason, H. L. 1930. The Santa Cruz Island pine. Madrono 2: 10.

Mason, H. L. 1932. A phylogenetic series of the California Cone Pines suggested by the fossil record. Madrono 2: 49-56.

Mason-Gamer, R. J., K. E. Holsinger, and R. K. Jansen. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). Mol. Biol. Evol. 12(3): 371-381.

McCauley, D. E., J. E. Stevens, P. A. Peroni, and J. A. Raveill. 1996. The spatial distribution of chloroplast DNA and allozyme polymorphism within a population of *Silene alba* (Caryophyllaceae). American Journal of Botany 83: 727-731.

Millar, C. I. 1983. A steep cline in *Pinus muricata*. Evolution 37: 311-319.

Millar C. I. 1986. The California Closed-Cone Pines (Subsection *Oocarpae* Little and Critchfield): A taxonomic history and review. Taxon 35: 657-670.

Millar, C. I. and W. B. Critchfield. 1988. Crossability and relationships of bishop pine. Madroño 35(1): 39-53.

Millar, C. I., S. H. Strauss, M. T. Conkle and R. Westfall. 1988. Allozyme differentiation and biosystematics of the California Closed-Cone Pines. Syst. Bot. 13: 351-370.

Mirov, N. T., E. Zavarin, K. Snajberk and K. Costello. 1966. Further studies of *Pinus muricata* in relation to its taxonomy. Phytochemistry 5: 343-355.

Moran, G. F., J. C. Bell and K. G. Eldridge 1988. The genetic structure and conservation of the five natural populations of *Pinus radiata*. Can. J. For. Res. 18: 506-514.

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Mol. Ecol. 3: 401-411.

Mosseler, A, K. N. Egger, and G. Hughes. 1992. Lack of genetic diversity in *Pinus resinosa* confirmed by random amplified polymorphic markers. Can. J. For. Res. 22: 1332-1337.

Murphy, T. M. 1981. Immunochemical comparisons of seed proteins from populations of *Pinus radiata*. Am. J. Bot. 68(2): 254-259.

Neale, D. B. and R. R. Sederoff. 1988. Inheritance and evolution of conifer organelle genomes. In: J. T. Hanover and D. E. Keathley (eds.). Genetic Manipulation of Woody Plants, pp. 251-254. Plenum Press, New York.

Neale, D. B. and R. 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., K. A. Marshall and R. R. Sederoff. 1989. Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. Proc. Natl. Acad. Sci. USA 86: 9347-9349.

Neale, D. B., K. A. Marshall and D. E. Harry 1991. Inheritance of chloroplast and mitochondrial DNA in incence-cedar (*Calocedrus decurrens*). Can. J. For. Res. 16: 1152-1154.

Nei, M. 1972. Genetic distance between populations. The American Naturalist 106: 283-292.

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

Nei, M. and W. H. Li. 1985. Mathematical model for studying genetic variation in terms of restriction endonuclease. Proc. Natl. Acad. Sci. USA 76: 5269-5273.

Nei, M. 1986. Definition and estimation of fixation indices. Evolution 40: 643-645.

Nesbitt, K. A., B. M. Potts, R. E. Vaillancourt, A. K. West, and J. B. Reid. 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). Heredity 74: 628-637.

Newcomb, G. B. 1962. Geographic variation in *Pinus attenuata* Lemmon. Ph.D. dissertation, University of California at Berkeley, California.

Oda, K., K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, Y. Orura, T. Kohchi and T. Ohyama. Gene organization deduced from the compelete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. Journal of Molecular Biology 223: 1-7.

Olmstead, R. G. and J. D. Palmer. 1994. Chloroplast DNA systematics: a review of methods and data analysis. American Journal of Botany 81: 1205-1224.

Owens, J. N. and Morris, S. J. 1990. Cytological basis for cytoplasmic inheritance in *Pseudotsuga menziesii*. I. Pollen tube and archegonial development. American Journal of Botany 77: 433-445.

Owens, J. N. and Morris, S. J. 1991. Cytological basis for cytoplasmic inheritance in *Pseudotsuga menziesii*. II. Fertilization and proembryo development. American Journal of Botany 78: 1515-1527.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358 (http:taxonomy.zoology.gla.ac.ukrodtreeview.html).

Palmer J.D. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae. In: R. J. MacIntyre (ed.). Molecular Evolutionary Genetics, pp.131-240. Plenum Press. New York and London.

Palmer J. D. 1990. Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. 6: 115-120.

Palmer J. D. 1992a. Mitochondrial DNA in plant systematics: applications and limitations. In: P. S. Soltis, D. E. Soltis and J. J. Doyle (eds.). Molecular Systematics of Plants, pp. 36-49. Chapman and Hall, New York.

Palmer J. D. 1992b. Comparison of chloroplast and mitochondrial genome evolution in plants. In: R. G. Herrmann (ed.). Cell Organelles, pp. 100-133. Springer-Verlag Wien New York.

Peakall R., P. E. Smouse and D. R. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dachtyloides*. Mol. Ecol. 4: 135-147.

Perron, M., A. G. Gordon and J. Bousquet. Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. Theor. Appl. Genet. 91: 142-149.

Perrotta, G., T. M. R. Regina, L. R. Ceci and C. Quagliariello. 1996. Conservation of the organization of the mitochondrial *nad3* and *rps12* genes in evolutionary distant angiosperms. Molecular and General Genetics. 251(3): 326-337.

Petit, R. J., A. Kremer, and D. B. Wagner. 1993a. Finite island model for organelle and nuclear genes in plants. Heredity 71: 630-641.

Petit, R. J., A. Kremer, and D. B. Wagner. 1993b. Geographic structure of chloroplast DNA polymorphisms in Europe oaks. Theor. Appl. Genet. 87: 122-128.

Plessas, M., and S. H. Strauss. 1986. Allozyme differentiation among populations, stands and cohorts in Monterey pine. Can. J. For. Res. 16: 1155-1164.

Pring, D. R. and D. M. Lonsdale. 1985. Molecular biology of higher plant mitochondrial DNA. Int Rev Cytol 97: 1-46.

Rajora, O. P., J. W. Barrett, B. P. Dancik and C. Strobeck. 1992. Maternal transmission of mitochondrial DNA in interspecific hybrids of *Populus*. Current Genetics 22: 141-145.

Rieseberg, L. H. 1996. Homology among RAPD fragments in interspecific comparisons. Mol. Ecol. 5: 99-105.

Rieseberg, L. H., H. Choi, R. Chan and C. Spore. 1993. Genomic map of a diploid hybrid species. Heredity 70: 285-293.

Rieseberg, L. H. and Soltis, D. E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants 5: 65-84.

Schardl, C. L., D. M. Lonsdale, D. R. Pring and K. R. Rose. 1984. Linearization of maize mitochondrial chromosomes by recombination with linear episomes. Nature 310: 292-296.

Schnell, R. J., Ronning, C. M., and R. J. Knight, Jr. 1995. Identification of cultivars and validation of genetic relationships in *Mangifera indica* L. using RAPD markers. Theor. Appl. Genet. 90: 269-274.

Sederoff, R. R. 1987. Molecular mechanisms of mitochondrial genome evolution in higher plants. Am. Nat. 130: S30-S45.

Shelbourne, C .J. A., M. H. Bannister and M. D. Wilcox. 1982. Early results of provenance studies on *Pinus muricata* in New Zealand. N. Z. J. Forestry 27(1): 50-66.

Smith, J. F., J. S. Scott-Craig, J. R. Leadbetter, G. L. Bush, D. L. Roberts and D. W. Fulbright. 1994. Characterization of random amplified polymorphic DNA (RAPD)

products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. Molecular Phylogenetics and Evolution 3: 135-145.

Smith, J. F., C. C. Burke and W. L. Wagner. 1996. Interspecific hybridization in natural populations of *Cyrtandra* (Gesneriaceae) on the Hawaiian Islands: evidence from RAPD markers. Plant Systematics and Evolution 200: 61-77.

Sokal R. R. and F. J. Rohlf. 1981. Biometry, second edition. W. H. Freeman and Co., New York.

Soltis, D. E., P. S. Soltis, and B. G. Milligan. 1992. Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: P. S. Soltis, D. E. Soltis and J. J. Doyle (eds.). Molecular Systematics of Plants, pp. 117-150. Chapman and Hall, New York.

Strauss, S. H., J. D. Palmer, G. T. Howe and A. H. Doerksen. 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc. Natl. Acad. Sci. USA 85: 3898-3902.

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

Strauss, S. H., Y. P. Hong, and V. D. Hipkins. 1993. High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata, P. muricata,* and *P. attenuata*. Theor. Appl. Genet. 86: 605-611.

Swofford, D. L. and R. B. Selander. 1989. BIOSYS-1.7: A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Illinois Natural History Survey, Champaign. IL.

Swofford, D. L. and G. J. Olsen. 1990. Phylogenetic reconstruction. In: D. M. Hillis and C. Moritz (eds.). Molecular Systematics, pp. 411-501. Sinauer, Sunderland, MA.

Swofford, D. L., G. J. Olsen, P. J. Waddell and D. M. Hillis. 1996. Phylogenetic inference. In: D. M. Hillis, B. K. Mable, and C. Moritz (eds.). Molecular Systematics, second edition, pp. 407-514. Sinauer, Sunderland, MA.

Szmidt, A. E., X. Wang, and M. Lu. 1996. Empirical assessment of allozyme and RAPD variation in: *Pinus sylvestris* (L.) using haploid tissue analysis. Heredity 76: 412-420.

Thormann, C. E., M. E. Ferreira, L. E. A. Camargo, J. G. Tivang, and T. C. Osborn. 1994. Comparison of RFLP and RAPD markers for estimating genetic relationships within and among cruciferous species. Theor. Appl. Genet. 88: 973-980.

Tulsieram, L. K., J. C. Glaubitz, G. Kiss, and J. E. Carlson. 1992. Single tree genetic linkage mapping conifers using haploid DNA from megagametophytes. BioTechnology 10: 686-690.

Van de Zande, L. and R. Bijlsma. 1995. Limitations of the RAPD technique in phylogeny reconstruction in *Drosophila*. Journal of Evolutionary Biology 8: 645-656.

Van Heusden, A. W. and K. Bachmann. 1992a. Genetic differentiation of *Microseris pygmaea* (Asteraceae, Lactuceae) studied with DNA amplification from arbitrary primers (RAPDs). Acta Botanica Neerlandica 41: 385-395.

Van Heusden, A. W. and K. Bachmann. 1992b Genotypic relationships in *Microseris elegans* (Asteraceae, Lactuceae) revealed by DNA amplification from arbitrary primers (RAPDs). Plant Systematics and Evolution 179: 221-233.

Van Heusden, A. W. and K. Bachmann. 1992c Nuclear DNA polymorphism among strains of *Microseris bigelovii* (Asteraceae, Lactuceae) amplified from arbitrary primers. Botanica Acta 105: 331-336.

Vicario, F., G. G. Vendramin, P. Rossi, P. Lio and R. Gianini. 1995. Allozyme, chloroplast DNA and RAPD markers for determining genetic relationships between *Abies alba* and the relic population of *Abies nebrodensis*. Theor. Appl. Genet. 90: 1012-1018.

Virk, P. S., B. V. Ford-Lloyd, M. T. Jackson, and H. J. Newbury. 1995. Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74: 170-179.

Wagner D.B., G. R. Furnier, M. A. Saghai-Maroof, S. M. Williams, B. P. Dancik, and R. W. Allard. 1987. Chloroplast DNA polymorphisms in lodgepole pines and their hybrids. Proc. Natl. Acad. Sci. USA 84: 2097-2100.

Wagner D. B., J. Dong, M. R. Carlson and A. D. Yanchuk. 1991. Paternal leakage of mitochondrial DNA in *Pinus*. Theor. Appl. Genet. 82: 510-514.

Wagner D. B., W. L. Nance, C. D. Nelson, T. Li, R. N. Patel and D. R. Govindaraju. 1992. Taxonomic pattern and inheritance of chloroplast DNA variation in a survey of *Pinus echinata, P. elliottii, P. palustris* and *P. taeda*. Can. J. For. Res. 22: 683-689.

Wallace, D. C. 1994. Mitochondrial DNA sequence variation in human evolution and disease. Proc. Natl. Acad. Sci. USA 91: 8739-8746.

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

Weising, K., H. Nybom, K. Wolff and W. Meyer. 1995. DNA fingerprinting in plants and fungi. CRC Press, Boca Raton, FL.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213-7218.

Wheeler, N. C. and R. P. Guries. 1982. Population structure, genic diversity, and morphological variation in *Pinus contorta* Dougl. Can. J. For. Res. 12: 595-606.

Whitkus, R., J. Doebley, and J. F. Wendel. 1994. Nuclear DNA markers in systematics and evolution. In: R. L. Philips and I. K. Vasil (eds.). DNA-based Markers in Plants, pp. 116-141. Kluwer Academic Publishers, Netherlands.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 18: 6531-6535.

Williams J. G. K., M. K. Hanafey, J. A. Rafalski, and S. V. Tingey. 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology 218: 704-740.

Wolfe, A. D. and A. Liston. 1997. Contribution of the polymerase reaction to plant systematics. In: P. S. Soltis, D. E. Soltis and J. J. Doyle (eds.). Molecular Systematics of Plants. Chapman and Hall, New York.

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

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

Wright, S. 1978. Evolution and genetics of populations, vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago.

Wu, J. Y., K. V. Krutovskii and S. H. Strauss. 1998. Abundant mitochondrial genome diversity population differentiation, and convergent evolution in pines. Genetics (in press).

Yeh, F. C., D. K. X. Chong and R. C. Yang. 1995. RAPD variation within and among natural populations of trembling aspen (*Populus tremuloides* Michx.) from Alberta. Journal of Heredity 86: 454-460.

APPENDICES

Appendix A Restriction fragment phenotypes of 28 observed mtDNA haplotypes obtained from 22 probe-enzyme combinations

Probe-Enzyme	#[a]	Size, kb[b]	Haplotype																											
			a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	bb
*atp1-Bam*HI	1	2.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*atp1-Xba*I	1	3.4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*atp6-Bam*HI	1	12.8	+[c]				+	+						+		+											+			+
	2	11.8		+	+	+			+	+	+	+	+		+		+	+	+	+	+	+	+	+	+	+		+	+	
*atp6-Xba*I	1	22.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*cob-Bam*HI	1	6.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	2	5.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*cob-Xba*I	1	12.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	2	2.6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*cox1-Bam*HI	1	17.2								+						+														
	2	11.3		+	+	+			+	+	+	+	+		+		+	+	+	+	+	+	+	+						

Appendix A (continued)

	3	9.0	+				+	+						+		+										+	+		+
	4	7.4			+																			+	+		+	+	
	5	6.0																						+			+	+	
	6	5.6																		+	+	+	+	+					
	7	5.0															+												
*cox1-Xba*I	1	21.0								+																			
	2	19.6	+				+	+						+		+													+
	3	15.2								+							+							+			+	+	
	4	13.0		+	+	+			+	+	+	+	+		+			+	+	+	+	+	+	+			+		
	5	10.8			+																			+	+		+	+	
	6	8.8																							+				
*cox2-Bam*HI	1	11.4					+							+															+
	2	8.2						+																					
	3	6.2		+	+	+			+	+	+	+	+		+			+	+					+	+		+		
	4	5.6								+																		+	
	5	5.0														+				+	+	+	+	+					
	6	3.4	+														+												
	7	2.8																								+			
	8	1.9								+																		+	
	9	1.7								+	+		+																

Appendix A (continued)

*cox2-Xba*I	1	24.0																							+	+		+	+		
	2	18.8		+	+				+	+	+	+	+				+								+	+		+	+		
	3	16.8													+																
	4	15.2																+													
	5	14.0																									+				
	6	11.2				+																									
	7	8.0					+							+														+			+
	8	5.4								+																			+		
	9	5.0								+	+	+	+																+		
	10	3.6	+													+															
	11	2.8						+																							
	12	2.3								+																			+		
	13	2.0															+		+	+	+	+	+								
*cox3-Bam*HI	1	12.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*cox3-Xba*I	1	11.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*nad3-Bam*HI	1	2.0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*nad3-Xba*I	1	25.0															+		+	+	+		+								

Appendix A (continued)

	2	24.0					+	+						+				+									+			+
	3	22.8	+										+		+	+															
	4	21.8		+	+	+			+	+	+	+						+							+	+		+	+	
	5	20.5								+																					
	6	19.8																					+								
nad4a-BamHI	1	3.6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad4a-XbaI	1	24.2	+				+	+						+		+											+			+
	2	22.2		+	+	+			+	+	+	+	+		+		+	+	+		+	+	+		+	+		+	+	
	3	20.0																		+	+		+	+						
nad5a-BamHI	1	8.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad5a-XbaI	1	22.5															+			+	+	+	+	+							
	2	10.8				+																									
	3	10.2		+	+		+	+	+	+	+	+	+		+			+	+						+	+			+	
	4	7.8	+											+		+											+	+		+
	5	6.6								+																					

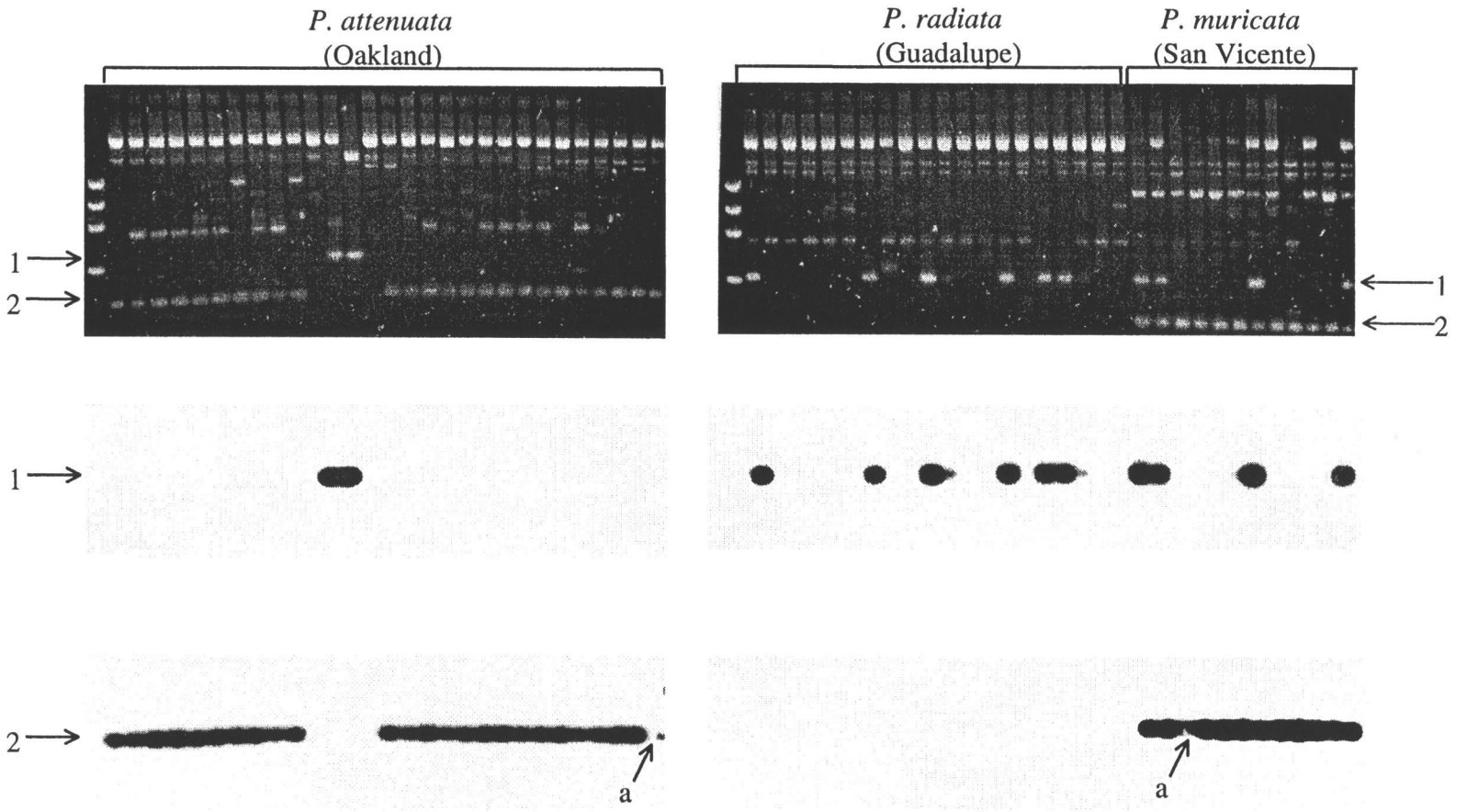
Appendix A (continued)

nad5d-BamHI	1	13.4		+	+	+			+	+	+		+	+	+		+	+	+	+	+	+	+	+	+	+		+	+	
	2	12.9					+	+				+															+			+
	3	8.4	+													+														
nad5d-XbaI	1	12.6																						+						
	2	11.6					+	+						+													+			+
	3	8.4		+	+	+				+							+		+	+	+	+		+	+			+	+	+
	4	7.9	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+
	5	4.4	+													+														
rps14-BamHI	1	5.0								+																				
	2	4.4		+	+	+			+	+	+	+	+		+	+		+	+	+	+	+	+	+	+	+		+	+	+
	3	3.9	+				+	+						+			+										+	+		
	4	1.0	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps14-XbaI	1	13.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Total	76		26	26	28	26	26	26	25	38	27	26	27	25	25	25	25	25	25	26	27	28	28	27	29	28	26	32	33	27

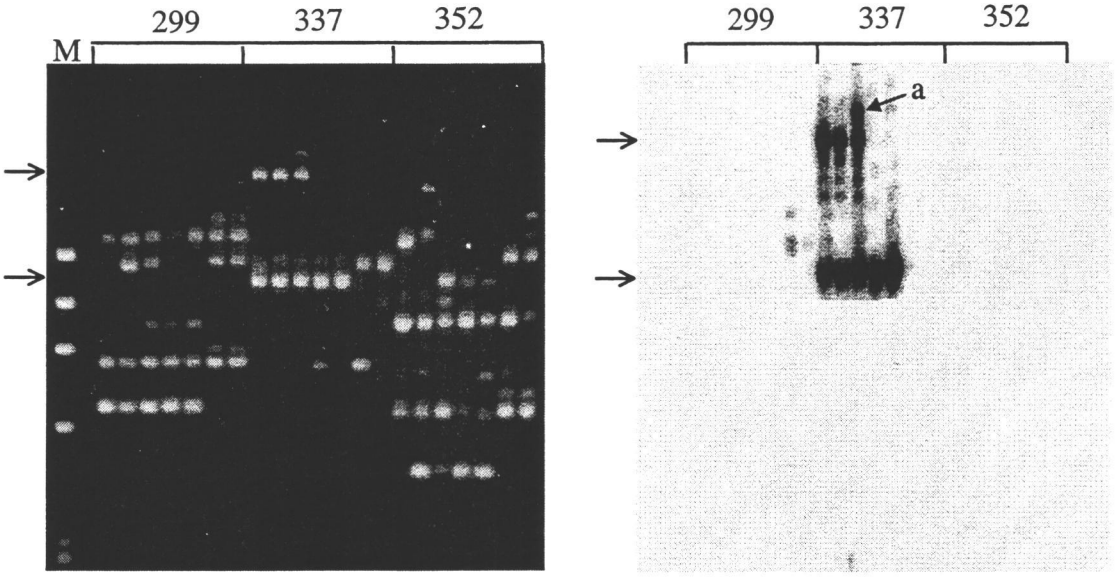
Appendix A (continued)

[a] Restriction fragment number.
[b] Restriction fragment size.
[c] Restriction fragment presence

Appendix B Sequence homology tests of RAPD markers. Panels show the RAPD gel profiles and the corresponding autoradiograms where the two indicated RAPD fragments were used separately as hybridization probes under high stringency to confirm RAPD fragment homology among species. a: DNA was excised from gel and purified to make a probe for fragment 2. The probe DNA for fragment 1 was excised from other gels.



Appendix C Southern blot of a RAPD gel hybrized with chloroplast DNA. Three RAPD primers (UBC 299, 377 and 352) were used to amplify RAPDs from seven trees of three California Closed-Cone Pines. RAPDs were separated on an agarose gel (A) and blotted for Southern hybridization. The blot was probed with cloned chloroplast DNAs. Two indicated fragments (arrows) were identified as RAPD products of chloroplast origin based on their strong hybridization signal and lack of background. a: a potential chloroplast DNA marker. M: a DNA size marker with fragment lengths of 1.3, 1.1, 0.9, and 0.6 kb (ΦX174 DNA/*HaeIII* digest).



A. Agarose electrophoresis **B. Southern hybridization**

Appendix D Genetic diversity within populations (H_S) and species (H_T), and population differentiation (G_{ST}, Nei 1986), based on codominant multiallelic allozymes and simulated dominant, biallelic loci at different sample sizes (N).

Species	N	H_S (SD)		H_T (SD)		G_{ST} (SD)	
		codominant	dominant	codominant	dominant	codominant	dominant
Knobcone	10	0.1004	0.0570	0.1258	0.0805	0.2512	0.3532
		(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)
	25	0.1006	0.0642	0.1259	0.0851	0.2509	0.3017
		(0.00)	(0.00)	(0.00)	(0.01)	(0.02)	(0.02)
	40	0.1010	0.0671	0.1262	0.0871	0.2497	0.2837
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	55	0.1009	0.0680	0.1261	0.0877	0.2501	0.2781
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	70	0.1008	0.0691	0.1263	0.0885	0.2519	0.2723
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	85	0.1009	0.0696	0.1262	0.0887	0.2505	0.2672
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	100	0.1009	0.0700	0.1262	0.0891	0.2512	0.2661
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Monterey	10	0.1323	0.0805	0.1472	0.0949	0.1436	0.2094
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.05)
	25	0.1318	0.0838	0.1469	0.0969	0.1464	0.1892
		(0.00)	(0.01)	(0.01)	0.01)	(0.01)	(0.02)
	40	0.1322	0.0855	0.1473	0.0987	0.1464	0.1878
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	55	0.1322	0.0861	0.1474	0.0991	0.1474	0.1849
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	70	0.1320	0.0864	0.1470	0.0991	0.1463	0.1814
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
	85	0.1318	0.0867	0.1469	0.0994	0.1466	0.1803
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
	100	0.1321	0.0868	0.1471	0.0994	0.1458	0.1783
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Bishop	10	0.1062	0.0497	0.1504	0.0876	0.3425	0.4873
		(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.05)

25	0.1060	0.0544	0.1502	0.0886	0.3427	0.4395
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	0.03)
40	0.1062	0.0557	0.1507	0.0895	0.3438	0.4310
	(0.00)	(0.00)	0.00)	(0.00)	(0.01)	(0.03)
55	0.1060	0.0567	0.1504	0.0898	0.3438	0.4214
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.03)
70	0.1059	0.0574	0.1502	0.0903	0.3432	0.4168
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
85	0.1061	0.0584	0.1504	0.0908	0.3432	0.4098
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)
10	0.1061	0.0584	0.1504	0.0908	0.3432	0.4098
0	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)

APPENDIX E: DNA Sample Storage

DNA samples corresponding to the individuals and populations of three species studied in Chapter 2 and 3 have been stored in -80°C freezer in Strauss Lab (FSL075). These samples were collected by Hong et al. in 1987 and Wu et al. in 1994. The samples representing same populations from two collections are combined into the same box. There are two different sets of DNA samples: unpurified and purified (three additional phenol/chloroform purifications after DNA extraction). Each box label includes the species name, population name, "purified or unpurified". Samples within each box are labelled with population names followed by the individual number. Individuas represent total genomic DNA extracted from needles. DNAs extracted from megagametophytes of seeds for segregation study are also stored in the same freezer.

APPENDIX F: Location of Electronic Files and Southern x-ray Films

Data files

The data files used for this study are stored in the DATA subdirectory of the P:\TGERC\THESIS\Wu directory of the Oregon State University College of Forestry computer network. There are three sub-subdirectories under DATA subdirectory: Allozyme, MtDNA and RAPD. Under each of the three directories, each program for our data analysis (e.g. GeneStat, Biosys, GDA, Winamova) is given a subdirectory. All data source files required for these programs to run are stored there. These data files are also stored in a floppy disk labelled as "Thesis data files for Junyuan Wu (1999)".

Thesis

An electronic copy of the entire thesis including all text, tables, and figures is stored in the THESIS subdirectory of the P:\TGERC\THESIS\Wu directory of the Oregon State University College of Forestry computer network as the file "thesis.doc" formatted for Microsoft Word 7.0

Southern x-ray Films

Three boxes are used to store Southern x-ray films for mtDNA study. A sheet is attached to each box to indicate the blot numbers, restriction enzymes used, and the number of sampes used for each population in each blot. Inside the boxes, blots used for the same probes are combined together and each blot is marked with the blot number and probe name. Boxes have been placed on the top shelf over the information desk at Strauss Lab FSL 075.