AN ABSTRACT OF THE THESIS OF

Yong-Pyo Hong for the degree of Doctor of Philosophy in Forest Science presented on June 25, 1991.

Title: Chloroplast DNA Variability and Phylogeny in the California Close Cone Pines

Abstract approved by: Signature redacted for privacy.
Steven H. Strauss

Chloroplast DNA (cpDNA) variability and phylogeny were studied via analysis of restriction site mutations and DNA sequencing in a complex of three closely related species of pines: Pinus radiata D. Don, P. attenuata Lemm., and P. muricata D. Don. Genomic DNA from 384 trees representing 20 populations in the complex were digested with 20 restriction enzymes and probed with cloned cpDNA fragments from Douglas-fir that comprise 85% of the chloroplast genome. Using five trees to represent each major genetic group in the complex, 667 bp of the intergenic region between the chloroplast genes rbcL and atpB were amplified via the polymerase chain reaction and directly sequenced.

No variation was observed in the sequenced intergenic region. 313 restriction sites were surveyed, accounting for 1.5% of the chloroplast genome. Twenty-four variable restriction site mutations were observed in the complex. Genetic diversity was nearly confined to differences among species; nucleotide diversity among species was estimated to be 0.3% (±0.09%). Monterey and knobcone pines displayed almost no genetic variation within or among populations.

Bishop pine showed strong population differentiation, a result of differences among three geographic groups, but almost no variability within populations (G_{ST} = 97.7%). This pattern of genetic architecture contrasts with that found in a recent allozyme study of the complex, and that of nuclear gene diversity in outcrossing species generally. Factors contributing to this difference are discussed, and may include smaller effective population sizes enhancing genetic drift, lower mutation rates, and periodic (natural) selection of organelle genes. Regions of the genome subject to length mutations were observed, as well as a heteroplasmic individual and a major genome inversion.

Phylogenetic analysis of restriction site differences by several methods showed that the three species were distinct, and that bishop and Monterey pines were most closely related. Knobcone pine was closest to the outgroup species, P. oocarpa Schiede, and appeared to have diverged earliest. The relationships of the three geographic groups of Bishop pine indicated a south to north migration along coastal California.

Chloroplast DNA Variability and Phylogeny in the California Closed Cone Pines

by

Yong-Pyo Hong

A THESIS

Submitted to

Oregon State University

In partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed June 25, 1991

Commencement June 1992

APPROVED:

Signature redacted for privacy.

Associate Professor of Forest Science in charge of major

Signature redacted for privacy.

Head of Department of Forest Science

Signature redacted for privacy.

Dean of Graduate School

Date thesis is presented June 25, 1991

Typed by Yong-Pyo Hong

ACKNOWLEDGEMENT

My sincere gratitude goes to the members of my graduate committee: Drs. S. H. Strauss, W. T. Adams, S. J. Knapp, K. L. Chambers, A. I. Liston, and P. J. Breen for their guidance during my graduate program and their critical review of my thesis. I express my deep appreciation to my major professor from the bottom of my heart, Dr. S. H. Strauss, for his guidance, encouragement, support, and patience during my graduate studies at Oregon State University. I am grateful to Dr. L. Norris and the Forest Science Department for support during my tenure, and to National Institutes of Health Grant GM 36367 to Dr. S. H. Strauss for salary and supply support.

I would like to thank Dr. J. Bousquet who taught me to use several computer programs for phylogenetic analysis; Dr. G. Furnier who sampled the Santa Ana population of knobcone pine for me, and Dr. M. Lynch who provided the computer program of HAPLO. I would like to acknowledge the help and advice of colleagues in Dr. Strauss' lab; A. H. Doerksen who helped greatly during numerous experiments; G. T. Howe who helped in field collections of samples; V. D. Hipkins who provided use of her Douglas-fir chloroplast DNA clones; and H. Bae, T. D. Giland-Parks, L. Nyers, and A. Krupkin. I also wish to thank my friend, Byung-Oh Cho, who helped to design and run computer programs for estimating population differentiation.

At last, I would like to thank my wife and daughter for their love, encouragement, and patience, and my parents-in-law in Korea for their love and support during my graduate studies. I dedicate this thesis to my late parents who wished me to be....

TABLE OF CONTENTS

LIST OF FIGURES

PAGE

LIST OF TABLES

PREFACE

The thesis is organized to facilitate conversion of its main chapters into manuscripts for publication. After a general background and literature review section, the next two chapters are in the form of manuscripts for submission to journals to the extent allowed by the Oregon State University Graduate Division. Chapter I on cpDNA diversity is intended for Genetics, Chapter II on phylogeny is intended for Systematic Botany. Due to a paucity of genetic differences, Chapter III on intergenic sequencing will not be submitted for publication. The introduction and discussion sections of the manuscripts, as well as their literature cited sections, by necessity contain information which is repeated in the literature review and bibliography sections of the thesis. The appendices contain information too detailed for manuscripts. The last section contains key protocols used in the thesis work, given in sufficient detail to allow most of the work to be repeated without referring to other publications.

Chloroplast DNA Variability and Phylogeny in the California Closed Cone Pines

BACKGROUND AND LITERATURE REVIEW

Systematic and Evolutionary Studies of the California Closed Cone Pines

The California closed cone pines include a montane, interior species, P. attenuata; and two maritime species, P. muricata and P. radiata (Critchfield and Little 1966) (Fig. 1). P. attenuata grows on dry, interior sites of southern Oregon, California, and on one location in northern Baja California, as small, disjunct populations. The two maritime pines occur discontinuously along the coast of California and on four offshore islands, and are sympatric on a part of the Monterey Peninsula. P. attenuata and P. radiata are sympatric in a few stands near Pt. Año Nuevo. Natural hybrids between these two pairs of coexisting species occur within each sympatric area (Plessas and Strauss 1987).

Monterey pine is distributed as three mainland populations (Año Nuevo, Monterey, and Cambria) and two insular populations (Cedros and Guadalupe Islands). Although the three mainland populations differ in morphological traits (Burdon and Bannister 1954, 1973, Fielding 1961, Forde 1964, Guinon et al. 1982),

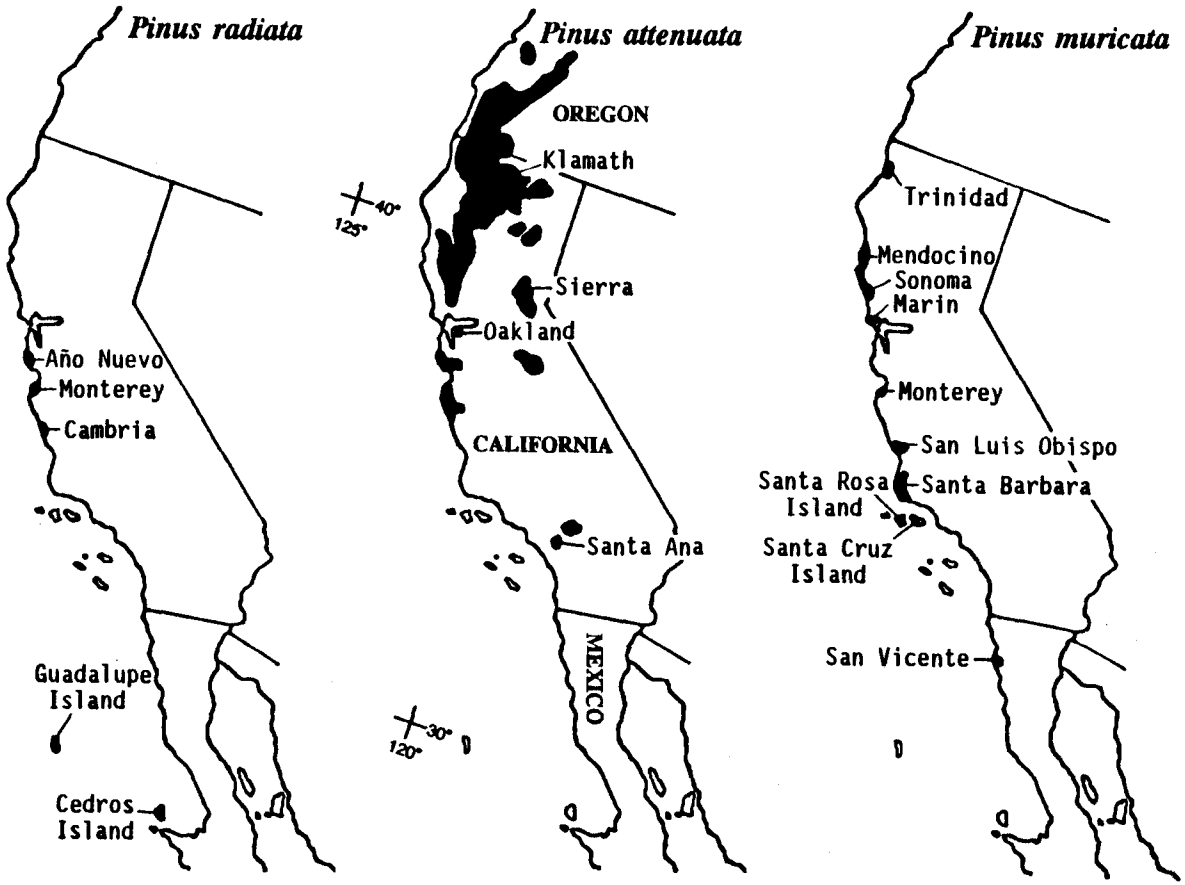


Figure 1. Distribution of P. radiata, P. attenuata, and P. muricata, showing the sampled populations.

interpopulation divergence is relatively small and patterns of resemblances among populations differs among traits. The two island populations, however, differ greatly from one another and from the three mainland populations in many morphological traits (reviewed in Millar 1986). This has resulted in the naming of three different varieties in this species: P. radiata var. radiata - the three mainland populations, P. radiata var. cedrosensis - Cedros Island population, and P. radiata var. binata - Guadalupe Island.

Compared to the two maritime species (Mirov et al. 1966), the few studies of geographic variation in knobcone pine have shown that relatively little variation is present both within and among populations. On the basis of the studies on morphological and growth traits (Brown and Doran 1985, Newcomb 1962), two patterns of variation were identified in knobcone pine: clinal variation with latitude among northern and central populations (Sierra, Cascade, and north coast ranges), and discontinuous variation among southern California and Baja California populations.

In bishop pine, variation in morphological traits such as growth, bark characteristics, stem form, cone abundance, cone serotiny, and resin canal number indicate that populations north of Sonoma differ strongly from populations south of Monterey, while the intervening Sonoma, Marin, and Monterey populations appear intermediate (Everard and Fourt 1974, Fielding 1961, Shelbourne et al.

1982). Several characteristics distinguish the two island populations, Santa Cruz and Santa Rosa, from the mainland populations (Forde and Blight 1964, Mason 1930, Millar et al. 1988, Mirov et al. 1966). The island populations have a high frequency of trees with thin-scaled, symmetric cones (Axelrod 1983, Howell 1941, Linhart 1978, Linhart et al. 1967, Mason 1930), as well as stouter, darker-green foliage, with more resin canals.

There have been considerable changes in the nomenclature of P. muricata since the early 20th century. Three varieties have been named: P. muricata var. borealis for the populations from Monterey northward; P. muricata var. muricata for the mainland populations south of Monterey; and P. muricata var. remorata for the island populations. P. muricata var. borealis is further divided into two discrete genetic groups--"Blue bishop" for the Trinidad and Mendocino populations, and "Green bishop" for the Sonoma, Marin, and Monterey populations. The two groups are distinguished by needle anatomy (sunken waxy stomata), monoterpene composition, phenology, and allozyme frequencies (Duffield 1951, Forde and Blight 1964, Millar 1983, Mirov et al. 1966).

On the basis of turpentine composition, Mirov (1948) concluded that the California closed-cone pines closely resemble one another, but that there is considerable genetic differentiation among populations and species

(Bannister and McDonald 1983, Bannister et al. 1962, Forde and Blight 1964, Mirov et al. 1966). Studies on allozyme diversity (Millar 1983, Moran et al. 1988, Plessas and Strauss 1986, Strauss and Conkle 1986) have shown considerable genetic variation within species, though less than that of more widespread conifers (typically less than 10% of G_{ST}: Hamrick and Godt 1990). Genetic differentiation among populations, however, often exceeds that in the more widespread conifers, at least for bishop and Monterey pines (F_{ST} = 13 to 22%: Millar et al. 1988).

Studies of crossability have indicated that knobcone and Monterey pines hybridize more readily than any other combinations of species (Critchfield 1967, Millar and Critchfield 1988). Surprisingly, strong crossing barriers exist within bishop pine. The Mendocino County populations (P. muricata var. borealis) are nearly unable to cross with the southern populations of this species (P. muricata var. muricata and P. muricata var. remorata), and also cross poorly with knobcone and Monterey pines. There is not a strong crossing barrier, however, among the southern populations of Bishop pine, nor between them and knobcone or Monterey pines (Critchfield 1967, Millar and Critchfield 1988). Almost complete barriers to crossing exist among the California closed cone pines and their putative close relatives of subsection Oocarpae in Latin America (Critchfield 1967).

The phylogenetic relationships of the California closed cone pines have recently been reviewed by Millar (1986) and Millar et al. (1988). This species complex appears to have originated in the Oligocene [26-38 million years before present (Bp)]. Based on floristic and tectonic evidence, Axelrod proposed that Oocarpae, to which the California closed cone pines belong (Critchfield 1967), originated in Mexico in the early-mid Tertiary (about 38 million years Bp). Axelrod also proposed a monophyletic origin for the Californian species from an ancestral Latin American pine that diverged from a P. oocarpa Schiede-like pool in the Oligocene. P. oocarpa is thus a good representative of the common ancestor of the California closed cone pines.

Phylogenetic relationships of the California closed cone pines was studied by analysis of allozyme frequencies (Millar et al. 1988). Despite the complexity of morphological trait relationships, allozymes indicated that all populations clearly belonged within one of the three species, which were distinct. They found roughly equal differentiation of the species from one another, and hypothesized a monophyletic origin of the complex from a P. oocarpa-like ancestor--with similar divergence times and rates.

Chloroplast DNA Studies

Chloroplast DNA (cpDNA) exists as a small, circular molecule, which typically contains two large inverted repeats, comprising about 10 to 70% of the genome (reviewed by Palmer 1990). It is generally homogeneous within individuals and varies little in size among species--ranging from about 120 to 217 kilobase pairs (kb). The gene order and content is highly conserved. The complete cpDNA sequence of tobacco indicated 4 rRNAs, 30 tRNAs, 39 proteins, and 11 other predicted protein coding genes (reviewed by Shinozaki et al. 1986).

The chloroplast genome is generally inherited uniparentally, though rare cases of biparental inheritance accompanied by somatic segregation have been reported (Kirk and Tilney-Bassett 1978, Sears 1980). In most flowering plants, the chloroplast genome is strictly inherited through the maternal parent. In confers, however, the chloroplast genome is generally inherited through the paternal parent (Neale et al. 1986, Szmidt et al. 1987, Wagner et al. 1987, and White 1990) with occasional biparental inheritance (Govindaraju et al. 1988, White 1990).

Rates of nucleotide substitution among chloroplast, mitochondrial, and nuclear DNA in plants was estimated through extensive DNA sequence comparisons (Wolfe et al. 1987). Non-synonymous substitution rates are similar in

chloroplast and mitochondrial genes, but the synonymous substitution rate in chloroplast genes (1-3 X 10^{-9} substitutions/site/year) is almost three times higher than that in mitochondrial genes (0.2-1.0 X 10^{-9}). The synonymous substitution rate in plant nuclear genes (5.0-30 X 10^{-9}), however, appears to be at least two times higher than that in chloroplast genes and five times higher than that in mitochondrial genes. Rate heterogeneity also exists within the chloroplast genome; the silent substitution rate (combined estimate from non-coding regions and synonymous sites) within the large inverted repeat is four times lower than that in single copy regions (Wolfe et al. 1987, 1989).

In contrast to the lower synonymous substitution rates in mitochondrial genes of plants, the mitochondrial genome is far more variable in organization and size than the chloroplast genome (reviewed in Birky 1988, Palmer 1990). The mitochondrial genome (218 to 2,500 kb) is generally larger than the chloroplast genome, and contains abundant short dispersed repeats that appear to serve as substrate for homologous recombination. The chloroplast genome has a relatively compact gene arrangement and few dispersed repeats. Its large inverted repeats also appears to inhibit many kinds of rearrangements (Palmer 1990). Among angiosperms, loss of the large inverted repeat has been found in only certain legumes, where some highly rearranged

genomes occur (Palmer et al. 1987). Strauss et al. (1988) observed the loss of the large inverted repeat in two conifers, Douglas-fir [Pseudotsuga menziesii (Mirb.) Franco] and Monterey pine, which were also highly rearranged.

The development of RFLP methodology has facilitated direct comparisons of homologous DNA segments in large numbers of samples (Strauss et al. 1989). Because of the conservative nature of cpDNA evolution, homologous mutations can be readily identified in species of same genus, and sometimes among genera within a family. The presence of shared mutations provide excellent characters for phylogenetic analysis and allow a direct estimate of DNA sequence-level genetic differentiation (Palmer 1987). Furthermore, several properties make the chloroplast genome particularly suitable for assessing phylogenetic relationships.

The chloroplast genome is generally homogeneous within individuals and small, which permits resolution on a single agarose gel of most fragments produced by the commonly employed restriction enzymes with a six base pair recognition site. In most plant groups, there have been few major genetic rearrangements during chloroplast genome evolution. Such events are common in nuclear and mitochondrial genomes, and can obscure the evolutionary history of DNA segments relative to that of organisms. The

high degree of DNA sequence conservation of chloroplast DNA compared to nuclear DNA (Lamppa and Bendich 1979, Palmer and Thompson 1982, Wolfe et al. 1987), and the extremely low incidence of parallelism and convergence in cpDNA restriction site mutations at the interspecific level (Palmer and Zamir, 1982; Palmer et al., 1983, 1985), promote the use of cladistic methods of phylogenetic inference since gains or losses of specific mutations can be analyzed. Finally, the chloroplast genome is generally inherited uniparentally; chloroplast DNA phylogenies can therefore be interpreted primarily in terms of mutation without the complicating factor of recombination.

Restriction site mutation analyses of cpDNA have proven useful at several taxonomic levels, though particularly for elucidating relationships among congeneric species in flowering plants (Crawford et al. 1990, Doebley et al. 1987, Palmer and Zamir 1982, Palmer et al. 1983, 1985, Sigurgeirsson and Szmidt 1988, Strauss and Doerksen 1990, Strauss et al. 1990, Sytsma and Schaal 1985, Szmidt et al. 1988). Sigurgeirsson and Szmidt (1988) investigated cpDNA variation among North-American Picea species, and its phylogenetic implications. Szmidt et al. (1988) studied phylogenetic relationships among Pinus species based on chloroplast DNA polymorphisms. In these studies, they observed relatively good agreement between taxonomic groupings based on cpDNA analyses and morphological traits,

but moderate to poor agreement between taxonomic groupings based on cpDNA analyses and crossability patterns. Strauss and Doerksen (1990) studied phylogeny in the genus Pinus using restriction fragment analysis of chloroplast, nuclear, and mitochondrial DNA. They found that Subsection Parrya is the most ancestral of extant taxa in Subgenus Strobus, and that the hard pines are strongly differentiated into Sections Pinus and Ternatae. Strauss et al. (1990) studied evolutionary relationships of Douglas-fir and its relatives from DNA restriction fragment analysis of chloroplast, nuclear, and mitochondrial DNA. They found strong differentiation among North American and Asian species, and predicted that the genus is likely to have evolved in North America and then migrated to Asia based on a high proportion of shared mutations with their outgroup, western larch.

Although there have been numerous studies of genetic variability in nuclear DNA in plants through analysis gene products (i.e., allozymes, terpenes, and morphology), little had been learned about variability in organelle DNA until molecular genetic methods were developed in the late 1970's and 1980's. Restriction site mutations among cpDNAs of different populations or species have been detected in several angiosperms, often from survey of few samples per population (Banks and Birky 1985, Crawford et al. 1990, Dang and Pring 1986, Doebley et al. 1987, Jasen and Palmer

1988, Ogihara and Tsunewaki 1988, Palmer et al. 1983, 1985, Scowcroft 1979, Soltis et al. 1989, Sytsma and Gottlieb 1986). The mean pairwise estimates of cpDNA nucleotide divergences in these studies range from 1.55 - 3.24% (mean: 2.4%) among genera, 0.10 - 1.62% (mean: 0.61%) among congeneric species, and 0.007-0.076% (mean: 0.041%) among conspecific populations. Banks and Birky (1985) sampled 100 individuals from 21 populations of Lupinus texensis Hook and found low amounts of cpDNA variation within as well as among populations (nucleotide divergence among populations was 0.007%). Soltis et al. (1989) found no cpDNA polymorphisms within 37 populations of Tolmiea menziesii Toor. & Gray, and nucleotide divergence of 0.076% among populations.

Although chloroplast DNA variation has been studied in woody plants, nucleotide divergences have not been calculated as a consequence of either small population samples, or study of restricted sections of the genome that contain complex genetic polymorphisms (Ali et al. 1991, Keim et al. 1989, Stine et al. 1989, Wagner et al. 1987, White 1990). For example, Wagner et al. (1987) surveyed a large number of trees, but focused on a small region of the genome subject to apparent length mutations and complex rearrangements. Only restriction site mutations can be directly related to nucleotide divergence (Nei 1987).

Neale et al. (1988) detected three cpDNA restriction

site mutations both within and among populations of wild barley (Hordeum vulgare L. ssp. spontaneum Koch), and among populations of cultivated barleys (Hordeum vulgare L. ssp. vulgare). Frequencies of the three multi-site genotypes for 19 populations of wild barley were 32.8, 36.0, and 31.2%. Stine et al. (1989) surveyed a limited number of trees of blue and white spruce (Picea pungens Engelm. and P. glauca (Moench) Voss) in rangewide samples. They found five restriction fragment length polymorphisms, mainly due to fragment size differences, that differentiated the cpDNA of both species. Ali et al. (1991) surveyed four conifers, Sequoia sempervirens D. Don, Pseudotsuga menziesii (Mirb.) Franco, Calocedrus decurrens (Torr.), and Pinus taeda L., and found almost no restriction site variability, but in P. menziesii did identify regions prone to large amounts of length variability--for which several different size variants could be identified. They also identified a highly rearranged genome in a single tree of S. sempervirens. Wagner et al. (1987) and Govindaraju et al. (1988) extensively surveyed length variants in Pinus contorta Loud. and P. banksiana Lamb. and found hotspot-like restriction fragment length variation and heteroplasmic individuals in a sympatric region. White (1990) found heteroplasmy for a restriction site polymorphism in Pinus monticola Dougl.

Length mutations have been frequently observed in

cpDNA studies and must play an important role in chloroplast genome evolution. However, the majority of cpDNA length mutations in both coding and non-coding sequences appear to be small deletions or insertions of 1 - 10 bp in size (Palmer 1985) that are not readily detectable by conventional RFLP analysis. The kinds of length mutations observed in RFLP surveys are larger (ca. 50 - 1,200 bp in size) and generally occur in a few restricted regions of the genome known as hotspots. Restriction site mutations, on the other hand, tend to be distributed relatively uniformly throughout the genome (Palmer et al. 1985, Ogihara and Tsunewaki 1988, Wagner et al. 1987). Length mutation hotspots may reflect the presence of tandemly repeated sequences (Palmer 1985, Hipkins and Strauss, unpubl. data). Length mutations in hotspots might occur as a result of deletions and insertions brought about by unequal crossing-over within tandem repeats. Polymorphic length mutations have been observed within a number of species and populations (Coates and Cullis 1987, Govindaraju et al. 1988, Ogihara and Tsunewaki 1988, Wagner et al. 1988).

Restriction site and length mutations give contrasting kinds of evolutionary insights. First, because of their restricted genomic distribution, length mutations cannot be regarded as providing a random sample of the genome. Site mutations, however, do appear to approximate a random

genomic sample (Wolfe et al. 1987, 1989). Second, as discussed above the rate and mechanism of mutation appears to differ greatly among length and site mutations. Length mutations occur as a consequence of the high mutability and recombination potential of repetitive DNA. Site mutations appears to occur as random point mutations. As a consequence, regions of the genome prone to length mutations are far more polymorphic than are random regions of the genome for site mutations. Length mutations are therefore, at least in theory (see below), better markers for studies of within-population genetic processes. The rarity of site mutations, and their tendency to be fixed within species or populations, makes them more suitable as phylogenetic markers. Finally, small length mutations cannot be as precisely characterized as can site mutations with RFLP analysis--and even with full characterization via sequencing their repetitive nature complicates interpretation of unique evolutionary events. This makes them more prone to biological or experimentally caused phylogenetic convergence (Palmer et al. 1985). Considerations of relative polymorphism aside, therefore, site mutations are superior characters for both phylogenetic and population genetic studies since evolutionary homology among alleles and among phylogenetically informative mutations can be much more confidently inferred.

Mitochondrial DNA (mtDNA) Studies in Animals

There have been a large number of studies of genetic diversity based on mtDNA analyses in animals, which can provide insights for organelle DNA studies of plants. Despite many studies using mtDNA to study introgression, phylogeny, and diversity, relatively few studies have quantified population subdivision (Crease et al. 1990, Davis 1986, Desalle et al. 1987, Palumbi and Wilson 1990, Whittam et al. 1986). In dioecious animals, the degree of population subdivision is expected to be considerably higher for mitochondrial genes compared to nuclear genes. The effective population size is roughly one-quarter that of nuclear genes as a consequence of haploidy and uniparental inheritance through females. Moreover, females often migrate less than males (Birky et al. 1983).

Eestimates of population subdivision based on mtDNA are much higher than those based on nuclear genes. G_{ST} estimates for mtDNA were 3.3 to 80-fold higher than those for allozymes (Crease et al. 1990, Davis 1986, DeSalle et al. 1987). For example, Desalle et al. (1987) found substantial population subdivision for mtDNA haplotypes in natural populations of Drosophila mercatorum ($G_{ST} \approx 0.16 \pm 0.04$), but essentially none for allozymes ($G_{ST} \approx 0.002 \pm 0.016$). Davis (1986) found much higher population differentiation ($G_{ST} \approx 0.31$) for mtDNA haplotypes than for allozymes ($G_{ST} \approx 0.04$) in the gopher Geomys bursarius.

Crease et al. (1990) found high subdivision of mtDNA haplotypes ($G_{ST} \approx 0.20 \pm 0.11$), but little allozyme subdivision ($G_{ST} \approx 0.06 \pm 0.02$) in the cladoceran Daphnia pulex.

It is unclear whether these results, which indicate that organelle DNA shows more subdivision than nuclear DNA, will hold for plants, and particularly for conifers. First, because most conifers are monoecious, organelle DNA can be transmitted to progeny from all sexually mature trees in populations. Thus, unlike dioecious animals, effective population size should be roughly one-half, not one-quarter, that of nuclear DNA. Second, because cpDNA is paternally rather than maternally inherited, differences in migration among the sexes is not likely to enhance subdivision. Paternal gametes effectively disperse twice, as both pollen and seed, whereas maternal gametes disperse only as seed, thus paternal inheritance may in fact reduce subdivision of cpDNA compared to nuclear DNA. Third, the mutation rate for organelle genes in plants is far lower than that for mtDNA in animals. cpDNA mutation was estimated to be at least two-fold lower than for nuclear genes in plants, and at least seven-fold lower than for mtDNA in animals (Wolfe et al. 1987). This should tend to enhance subdivision because allelic variation within populations that is depleted by drift will only slowly be restored by mutation (Banks and Birky 1985). Finally,

natural selection may differ in stringency among genomes. The high rate of genetic polymorphism for mtDNA in animals suggests weak selection among mtDNA variants. In contrast, the low site polymorphism and highly conserved structure of cpDNA nearly throughout the plant kingdom suggests that it may be under stringent natural selection--perhaps as a consequence of its essential role in primary energy production through photosynthesis. Selection could act to enhance subdivision in uniparentally inherited genomes by driving genomes that contain favorable mutant alleles to fixation in some populations, while simultaneously driving to extinction genomes that do not contain such alleles--thus depleting populations of diversity (periodic selection: Maruyama and Birky 1991).

Quantitative Phylogenetic Analysis

DNA sequences as quantitative genetic markers have several advantages over other characters such as allozymes, terpenes, and morphology (reviewed by Strauss et al. 1991). First, most DNA sequence variation occurs in non-coding DNA and synonymous sequences of coding DNA, and is thus very weakly related to fitness. Evolutionary inferences may therefore be nearly free from the constraints imposed by natural selection. Second, it allows a greater variety of evolutionary scales to be studied, ranging from highly polymorphic DNA fingerprints to highly conserved

chloroplast genes. Third, genotypes can be more easily and accurately interpreted from phenotypes--facilitating the separation of evolutionary homology from analogy. Fourth, a much larger number of independent homologous characters (nucleotides or restriction sites) can be identified with DNA analyses than are possible with other methods. Fifth, the use of statistical methods (e.g., bootstrap and maximum likelihood) for assessing reliability of phylogenetic trees can be facilitated by the ability to more clearly infer the mutational mechanism of genetic change.

DNA sequences as genetic markers, however, also have disadvantages. First, molecular markers are not yet widely available in plants and their effectiveness for evolutionary studies within and among closely related species is still being tested. Second, and most important, the added cost of laboratory set up, supplies, and scientific training is a substantial deterrent to their use (Clegg 1989).

Inferring phylogenetic relationships from molecular data requires the selection of an appropriate method from many available techniques. Two commonly used broad categories are phenetic and cladistic approaches. Phenetic methods, often called distance-matrix methods, construct a tree by considering phenotypic similarities of the operational taxonomic units (OTUs) without trying to understand the evolutionary pathways of specific

characters. Simple phenetic methods, such as UPGMA and WPGMA, are better able to construct phylogenetic trees than cladistic methods when the rules of evolutionary change of the characters used are not well understood (Nei 1987). Tree construction is largely a process of grouping organisms based on pairwise similarity or distances among OTUs. Molecular genetic distances are typically measured as nucleotide diversity among OTUs. Such conversions of character state (e.g., nucleotide sequence) data into single numerical summaries of genetic distance (e.g., nucleotide diversity) result in considerable loss of information (Hills 1987, Swofford and Olsen 1990) especially because they are usually based on a simple probability models for evolutionary change. Superior distance methods start from genetic distances whose calculation reflects biological realities of evolutionary change (e.g., transitions vs. transversions; synonymous vs. non-synonymous nucleotide changes) (Ritland and Clegg 1987, Swofford 1990). Phenetic methods can be distinguished on the basis of whether they assume homogeneity of evolutionary rates (e.g., UPGMA, WPGMA, UPGMS, and UPGMC), or allow rates to vary (e.g., neighbor-joining, Fitch-Margoliash, and distance Wagner methods).

In contrast to phenetic methods, cladistic methods consider various paths of evolution by analyzing the gains and losses of individual characters. Methods based on the

principle of maximum parsimony (fewest hypothesized mutations in trees) (Fitch 1977) have been most widely used. The branching patterns and branch lengths in phylogenetic trees indicate an inferred temporal order of evolutionary events (character gains and losses). Possibilities of change can be weighted differentially for individual characters, rather than for broad classes of characters, as for distance-matrix methods. Two disadvantages of cladistic methods are their poor suitability when OTUs differ in frequency, rather than in possession of fixed characteristics; and, when using unweighted algorithms, their lack of consistency when characters evolve rapidly or at heterogeneous rates (Felsenstein 1985b).

Molecular analyses have facilitated the statistical testing of phylogenetic topologies by providing characters whose mutational basis--and thus independence--is known (reviewed in Strauss et al. 1991). The bootstrap is the most popular form of analysis, and has been widely applied to cladistic parsimony methods for estimating significance of branching patterns as well as confidence limits on branch lengths (Ritland and Clegg 1987). Purely phenetic analyses that do not characterize mutational bases of phenotypes--whether or not molecular--cannot properly be subject to such statistical treatments. An example is use of total shared restriction fragments to estimate

phylogenies (Sigurgeirsson and Szmidt. 1988, Szmidt et al. 1988); because individual restriction site mutations produce multiple fragment changes, such estimates are composed of non-independent observations. Moreover, they do not distinguish among site and length mutations--which, as discussed above, differ greatly in their suitability for phylogenetic analysis. Because of the complexity of phylogenetic topologies, and thus the large numbers of phylogenetically informative characters needed for robust estimation, calculation of the statistical confidence is critical. Many sources of error can lead to incorrect phylogenies, including character convergence, incompletely sampled genetic polymorphisms, differences among gene and species genealogies, and experimental errors is assigning homology and interpreting genotypes from phenotypes.

The evolutionary history of individual genes and genomes will not always be congruent with the evolutionary history of whole organisms (Nei 1987). Organelle genomes might be expected to differ considerably from nuclear genomes given their different rates of evolution, degrees of subdivision, mode of inheritance, and functions. Because most aspects of whole organisms, on which species classifications are based, are affected by nuclear genes, organelle genes might often disagree with classical species designations. Nonetheless, because reproductive isolation is the fundamental element of phylogenetic differentiation,

and it affects all genes and genomes, congruence in phylogenetic branching--though not necessarily in evolutionary rate--is usually expected regardless of which gene or genome is studied.

Significant errors in inferring organismal phylogenies from gene phylogenies are most likely among closely related taxa, or among well-differentiated taxa that had diverged from a common ancestor at similar times. In the first case, few genetic differences will be present, and stochastic changes of allele frequency or allele fixation may bias phylogenetic inferences. In the second case, many genetic differences will be present but few will be phylogenetically informative, and stochastic events that occurred soon after speciation will impede correct phylogenetic interpretations. Long-time evolutionary diversification may also erase the few phylogenetically informative characters originally present. The only solution for these situations is to generate a very large number of independent characters to hopefully allow phylogenetically meaningful patterns to emerge above the background "genetic noise".

CHAPTER I

ORGANIZATION OF CHLOROPLAST DNA DIVERSITY IN A SPECIES COMPLEX: THE CALIFORNIA CLOSED CONE PINES

ABSTRACT

The amount, distribution, and mutational nature of chloroplast DNA polymorphisms were studied via analysis of restriction fragment length polymorphisms in a complex of closely related tree species, the California closed cone pines: (*Pinus attenuata* Lemm., *P. muricata* D. Don, and *P. radiata* D. Don). Genomic DNA from 384 trees representing 20 populations in the complex were digested with 20 restriction enzymes and probed with cloned cpDNA fragments from Douglas-fir that comprise 85% of the chloroplast genome. 313 restriction sites were surveyed, and 24 of these were observed to be polymorphic in the complex. Differences among species accounted for the majority of genetic diversity; nucleotide diversity among species was estimated to be 0.3% (±0.09%). *P. attenuata* and *P. radiata* displayed almost no genetic variation within or among populations. *P. muricata* also showed little variability within populations, but did display strong, and apparently

fixed, population differences (G_{ST} = 97.7%), that were a result of three distinctive geographic groups. Intrapopulation polymorphisms were found in only five populations; mean within population nucleotide diversity was estimated to be 0.0017%. This pattern of genetic architecture contrasts strongly with findings from study of nuclear genes (allozymes) in the complex, where most genetic diversity resided within populations, rather than among populations or species. Regions of the genome subject to fragment length mutations were identified, as well as rare individuals that were heteroplasmic, or whose cpDNA contained a major inversion. Estimates of subdivision based on length variant frequencies bore no resemblance to that based on site mutations or allozymes.

* List of abbreviations:

bp, base pair; cpDNA, chloroplast DNA; kb, kilo base pair; mtDNA, mitochondrial DNA; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism.

INTRODUCTION

The chloroplast genome has been the focus of considerable study by plant evolutionists. Its slow rate of change both in structure and sequence has made it the subject of many studies of organismal phylogeny (Palmer 1987). Restriction fragment analyses have been the predominant mode of analysis within genera, whereas sequence analyses, particularly of ribulose bisphosphate 1,5-carboxylase (*rbc*L), has been the predominant mode of analysis of higher taxonomic categories such as families and orders (Clegg and Zurawski 1987, Wolfe et al. 1987).

As a consequence of its slow rate of evolution, there has been relatively little effort to characterize the nature and extent of genetic polymorphism within species. Banks and Birky (1985) found cpDNA variation within as well as between populations from 100 individuals representing 21 populations of *Lupinus texensis* Hook. Their estimate of cpDNA nucleotide divergence among populations was 0.007%. Soltis et al. (1989) found no cpDNA RFLPs within populations from 37 populations (two or more samples per population) of *Tolmiea menziesii* Toor. & Gray, but from restriction site mutations estimated cpDNA diversity among populations to be 0.076%. Neale et al. (1988) detected cpDNA restriction site variation both within and among populations of wild barley (*Hordeum vulgare* L. ssp. *spontaneum* Koch), and among populations of cultivated

barley (*Hordeum vulgare* L. ssp. *vulgare*). Frequencies of the three multi-site genotypes for 19 populations of wild barley were 32.8, 36.0, and 31.2%. Stine et al. (1989) surveyed a limited number of spruce trees of blue and white spruce: *Picea pungens* Engelm. and *P. glauca* (Moench) Voss in rangewide samples. They found five restriction fragment length polymorphisms, mainly due to fragment size differences, that differentiated the cpDNA of both species. Ali et al. (1991) surveyed four conifers, *Sequoia sempervirens* D. Don, *Pseudotsuga menziesii* (Mirb.) Franco, *Calocedrus decurrens* (Torr.), and *Pinus taeda* L., and found almost no restriction site variability, but in *P. menziesii* did identify regions prone to large amounts of length variability--for which several different size variants could be identified. They also identified a highly rearranged genome in a single tree of *S. sempervirens*. Wagner et al. (1987) and Govindaraju et al. (1988) extensively surveyed length variants in *Pinus contorta* Loud. and *P. banksiana* Lamb. and found hotspot-like restriction fragment length variation and heteroplasmic individuals in a sympatric region. White (1990) found heteroplasmy for a restriction site polymorphism in *Pinus monticola* Dougl.

Although these studies have made it clear that cpDNA polymorphism within species is low and often comprised of length rather than site mutations, the extent of population

and species subdivision is unknown. Organelle genomes may have substantially different orgnaizations of genetic diversity than nuclear genes as a consequence of a number of factors (Birky 1988), including their lower mutation rates (1-3 X 10^{-9} substitution/site/year) (Wolfe et al. 1987); their lower effective population sizes as a result of haploidy and, for dioecious species, uniparental inheritance; and the potential for periodic selection to speed fixation of neutral alleles while ridding populations of genetic polymorphism (Maruyama and Birky 1991). Strong subdivision for organelle DNA compared to nuclear DNA have been observed in studies of mitochondrial DNA in animals (Crease et al. 1990, Davis 1986, DeSalle et al. 1987), but we know of no estimates of cpDNA subdivision in plants.

The goal of this study was to examine the nature and architecture of chloroplast DNA genetic diversity in a species complex that had already been well characterized by several other means, including nuclear gene polymorphisms (allozyme frequencies and terpene composition), crossability, morphology, and paleohistory. We report that although levels of within-population polymorphism are extremely low, genetic differences may become fixed in populations and species far more rapidly than is observed for nuclear genes. Rare heteroplasmic individuals indicate occasional biparental inheritance and the potential for genome recombination; and rare highly rearranged (inverted)

genomes indicate that the potential for major changes in genome structure such as have been observed among conifer genera exist within populations.

MATERIALS AND METHODS

Plant Materials

We sampled a total of 384 trees originating from 20 populations (Fig. I.1; Table I.1). All samples of knobcone pine were from trees growing in natural populations. The Klamath population was sampled over an 2 mile transect adjacent to U.S. Interstate 5 north of Shasta Lake Recreation Area, California (latitude 40°50', longtitude 122°45'). The Sierra Nevada population was sampled over an 2 mile transect located adjacent to U.S. Interstate 80 near Auburn, California (latitude 38°54', longtitude 121°08'). The Oakland population was sampled over an 1.5 mile transect along Flicker Ridge, in the hills east of Oakland, California (latitude 37°50', longtitude 122°30'). The Santa Ana population was sampled by Dr. Glenn Furnier from widely spaced trees in the Santa Ana mountains near Corona, California (latitude 33°50', longitude 117°37').

Bishop and Monterey pines were sampled from gene conservation and genetic test plantations in three locations: Albany, California (Gill Tract) and LaFayette, California (Russell Reservation) established by Dr. W.J. Libby of the University of California at Berkeley; and the

Table I.1. Sample origins and sizes.

Species	Populations	No. of Sample Trees	Abbreviation	Origin[1]
P. radiata	Año Nuevo	30	MANO	G
	Cambria	28	MCAM	G
	Cedros Island	27	MCED	G & R
	Guadalupe Island	24	MGUA	G & R
	Monterey	27	MMOT	G
P. attenuata	Klamath	25	KKLA	Nat. Population
	Oakland	25	KOAK	Nat. Population
	Sierra	25	KSIE	Nat. Population
	Santa Ana	25	KSAA	Nat. Population

Table I.1. (continued)

Species	Populations	No. of Sample Trees	Abbreviation	Origin[1]
P. muricata	Mendocino	6	BMEN	C & R
	Trinidad	7	BTRI	R
	Marin	5	BMAR	C & R
	Sonoma	7	BMSO	R
	Monterey	12	BMON	C & R
	San Luis Obispo	12	BSAL	C & R
	San Vicente	20	BSAV	C & R
	Santa Barbara	19	BASB	C & R
	Santa Cruz-bishop	22	BSAC	C & R
	Santa rosa	17	BSAR	C & R
	Santa Cruz-remorata	21	BSCR	C & R
TOTAL	20 Populations	384 Sample trees		

[1] C = Concord plantation, G = Gill Tract, R = Russell Reservation

Figure I.1. Distribution of *P. radiata*, *P. attenuata*, and *P. muricata*, showing the sampled populations.

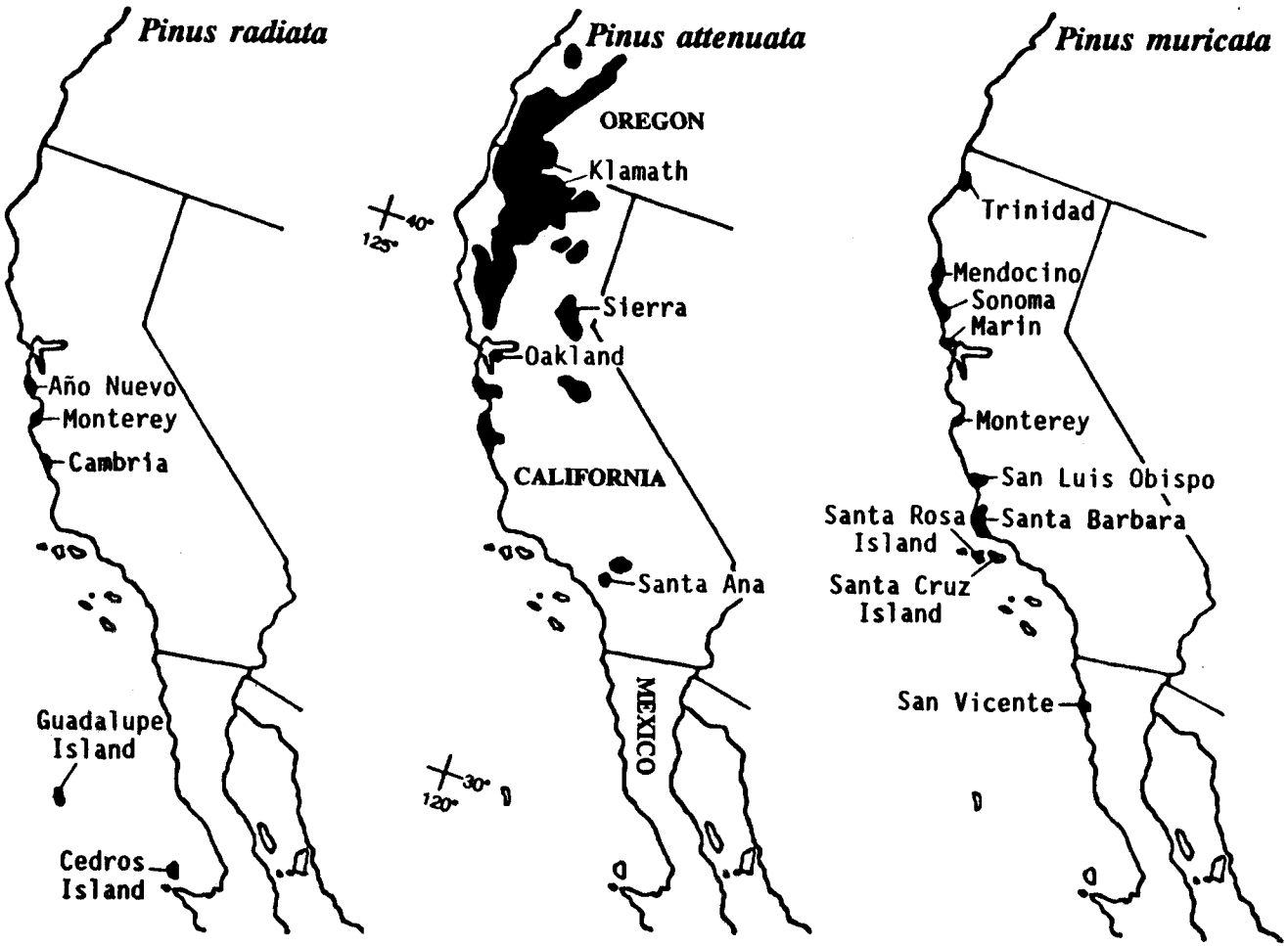

Pinus radiata
Año Nuevo
Monterey
Cambria
Guadalupe Island
Cedros Island
Pinus attenuata
OREGON
Klamath
40°
125°
Sierra
Oakland
CALIFORNIA
Santa Ana
MEXICO
30°
120°
Pinus muricata
Trinidad
Mendocino
Sonoma
Marin
Monterey
San Luis Obispo
Santa Barbara
Santa Rosa Island
Santa Cruz Island
San Vicente

Figure I.1.

U.S. Naval Weapons Research Station in Concord, California established by the U.S. Forest Service, Pacific Southwest Experiment Station in Berkeley, California (maintained by Dr. F. T. Ledig). All of the trees in these plantations were either clonal replicates of trees from natural populations, or derived from seeds collected in natural populations. Fifty-six of the total 284 trees collected in these populations were known to be open-pollinated siblings of another sampled tree (discussed below).

DNA Analysis

Total cellular DNA was extracted from 40 g of sampled needles using a CTAB-based DNA extraction method (Wagner et al. 1987). We modified this method only in that we worked in a cold room; used liquid nitrogen, rather than extraction buffer, for initial grinding of needles; and shook tubes vigorously during denaturation of membranes in sarkosyl because of the high viscosity of our extractions. DNA was then digested with restriction enzymes according to manufacturers suggestions and fractionated according to size by electrophoresis in 0.8% or 1.5% agarose gels in a TAE buffer (80 mM Tris, 16.6 mM sodium acetate, 2 mM EDTA, adjusted to PH 8.1 with glacial acetic acid). The gels were then blotted in an alkaline denaturation solution (Reed and Mann 1985) onto Zetabind nylon membranes (Cuno Inc., Meriden,CT), and hybridized with four mixtures of

equal amounts of cloned Douglas-fir cpDNA fragments (Fig. I.2, Table I.2). The clones were radioactively labelled with ^{32}P by random-hexamer-primer-extension (Feinberg and Vogelstein 1983) and together covered about 85% of the chloroplast genome. Hybridization, washing, and stripping of blots was as described in Strauss and Doerksen (1990).

To veryfy that a putative heteroplasmic tree didnot result from DNA contamination, a 712 bp region of the mitochondrial cytochrome oxidase I gene (*coxI*) from knobcone pine was amplified via the polymerase chain reaction using primers to conserved regions (kindly supplied by Jeffrey C. Glaubitz and Dr. John E. Carlson, University of British Columbia). It was used as a hybridization probe for the same blots as used for the cpDNA study under identical labelling and hybridization conditions.

Strategy for Surveying Polymorphisms

We used a two-stage screening procedure to detect cpDNA polymorphisms. For the initial screen, we randomly chose two trees from each of the 20 sampled populations in the complex, plus a single tree of *Pinus* oo*carpa* as an outgroup [obtained from Botanical Garden in Berkeley, California, accession # 67.1648; native in Valle de Bravo, Mexico (latitude 13°14', longtitude 100°07')]. We then digested these DNAs with 17 restriction enzymes with six bp

Table I.2. List of cpDNA clones used for probes.

Probe set	cpDNA clones[1]	Restriction enzymes showing intra- or interspecific polymorphism
Set A	XbaI-6.9, 8.9, 4.8	BamHI, BclI, DraI, EcoRI
	KpnI-1.2, 1.6, 2.6, 6.0, 0.7	EcoRV, KpnI, XbaI, XmnI
Set B	XbaI-11.2, 2.5, 3.8, 1.6	
	KpnI-8.7	EcoRI
Set C	XbaI-5.9b, 6.0, 2.7, 3.2a	
	KpnI-6.5	EcoRI
Set D	XbaI-5.1, 1.9, 0.8	
	SstI-10.8a, 2.6, 6.7, 4.8	XhoI

[1] Number = fragment size (kb) from map (Fig. 2)

Figure I.2. Douglas-fir cpDNA restriction site map for *Kpn*I, *Sst*I, and *Xba*I. Maps and clones were developed by Strauss et al. (1988), Strauss and Tsai (1988), and ourselves. Dashed lines within probe sets indicate uncloned cpDNA fragments (thus, not included as hybridization probes). See Table I.2 for detailed composition of each probe set. Arrow heads with asterisk denote inversion borders in Douglas-fir relative to Monterey pine (Strauss et al. 1988).

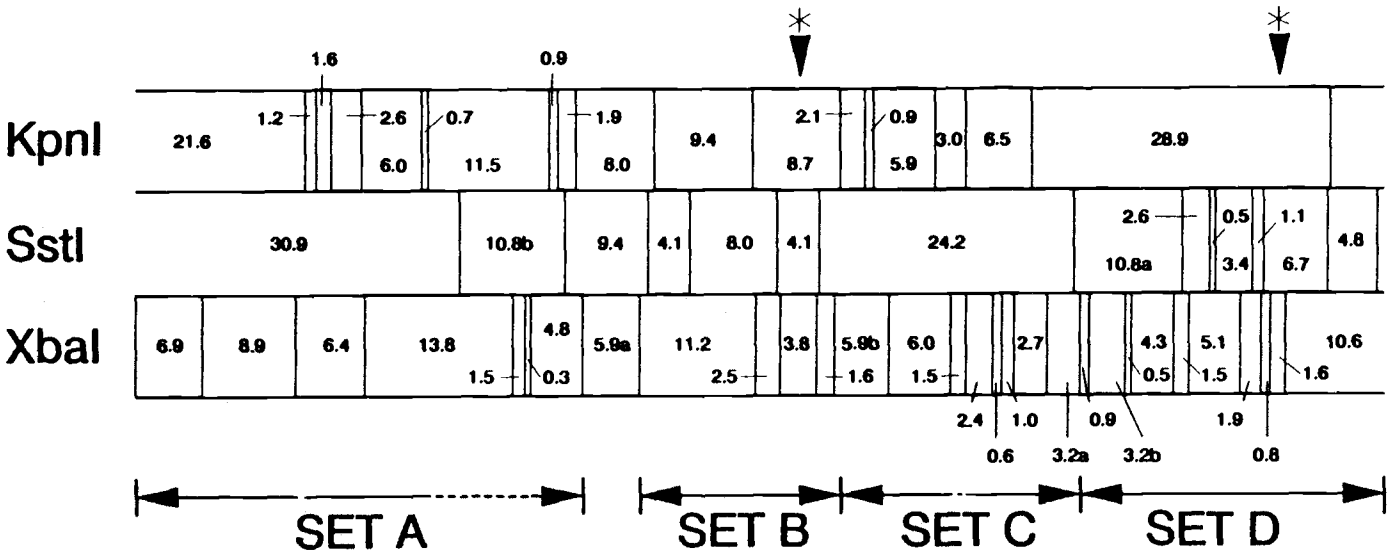


Figure I.2.

recognition sites (*Bam*HI, *Bcl*I, *Bgl*I, *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III, *Kpn*I, *Pst*I, *Sal*I, *Sma*I, *Sst*I, *Stu*I, *Xba*I, *Xho*I, and *Xmn*I) and 3 enzymes with four bp recognition sites (*Hinf*I, *Hpa*II, and *Msp*I), and probed their blotted DNA with our cpDNA probe sets. Only those nine enzymes (*Bam*HI, *Bcl*I, *Dra*I, *Eco*RI, *Eco*RV, *Kpn*I, *Xba*I, *Xho*I, and *Xmn*I) that detected polymorphisms either within or between species in this primary survey were retained for the full survey of all 384 study trees.

Quantitative Analyses

Haplotypes were recognized on the basis of having unique restriction fragment patterns over the various restriction enzyme and probe combinations studied. Only clearly interpretable site mutations were considered in nucleotide diversity analyses; genome polymorphisms and length mutations were excluded. To avoid counting fragments and site mutations more than once, for each enzyme identical blots were hybridized with each probe set and autoradiograms from adjacent probe sets carefully compared. The number of unique fragments observed was assumed equal to the total number of sites surveyed. Because of the gaps in genome coverage provided by our probe sets, however, this probably gives a slight underestimate of the number of sites surveyed. The number of different kinds of haplotypes in each population for

restriction enzymes with four and six bp recognition sites were used to estimate nucleotide diversity and its standard errors among and within populations and species (with the computer program HAPLO: Lynch and Crease 1990).

The relative degree of differentiation at each hierarchical level studied was quantified by four parameters: N_{ST}, G'_{ST}, G_{ST}, and F_{ST}. $N_{ST} = V_b / (V_w + V_b)$, in which V_w is the average number of substitutions per site for random pairs of chloroplast DNA within a hierarchical level (e.g., within populations), and V_b is the average number of substitutions per site for random pairs of chloroplast DNA among hierarchical levels (e.g., among populations) (Lynch and Crease 1990). N_{ST} weights evolutionary distances among pairs of haplotypes based on sequence divergence estimated by maximum likelihood (Nei and Tajima 1983). G'_{ST} is Nei's (1977) fractional gene diversity statistic for population differentiation calculated using haplotype frequencies. N_{ST}, G'_{ST}, and their standard errors were calculated by HAPLO. G_{ST} is the fractional gene diversity statistic among populations calculated from allele (site mutation) rather than haplotype frequencies. F_{ST} is Wright's F-coefficient for population subdivision (Weir and Cockerham 1984) and calculated, along with its jackknife-derived standard errors, with a computer program supplied courtesy of R. Weir (1990).

RESULTS

We assayed 255 restriction sites with 17 six bp recognition enzymes, and 58 sites with three four bp enzymes, giving a total of 313 sites and 1,762 bp assayed per tree. This is equivalent to study of 1.5% of the roughly 120 kb pine chloroplast genome (Strauss et al. 1988). Over the entire species complex, this is nearly equivalent to sequencing of 63 kb in the preliminary screen of 36 trees with 20 enzymes, and 208 kb in the full survey of 384 trees with 9 enzymes (total = 271 kb). Twenty-four of the 313 sites were polymorphic either within or among species (Table I.3). These 24 variable and 289 monomorphic sites could be assigned to nine different haplotypes. When based on the full 20 enzymes included in the primary survey, the haplotypes consisted of 300 to 307 site assignments.

Nucleotide Diversity within Populations

Twenty of the 24 polymorphic sites showed a complete lack of variability within populations. Of the 20 populations surveyed, only 5 contained polymorphisms--two in Monterey pine (Año Nuevo and Guadelupe) and three in bishop pine (San Vicente, Santa Cruz-*remorata*, and Monterey). Frequencies of the mutant sites in the five populations were 0.002, 0.002, 0.011, 0.006, and 0.023, respectively. All 100 individuals from four geographically

Table I.3. Restriction fragment phenotypes used to infer site mutations.

OBSV	ENZY	PROB	MMMMM ACMGC NAOUE OMNAD	KKKK KOSS LAAI AKAE	BB MT ER NI	BBB MSM AOO RNN	BBBBBB SSSSSS AAAAAC RLRVR BCLRA	OO OO	PHEN
1	*Bam*HI	A	XXXXX	OOOO	XX	XXX	XXXXXX	NI	2.84, 2.02+0.62
2	*Bam*HI	A	OOOOO	XXXX	OO	OOO	OOOOOO	O	3.17, 3.10+0.07
3	*Bcl*I	A	OOOOO	OOOO	XX	XXX	OOOOOO	O	7.98, 7.77+0.21
4	*Bcl*I	A	XXXXX	XXXX	OO	OOO	XXXXXX	X	5.09, 4.57+0.52
5	*Dra*I	A	XXXXX	OOOO	XX	XXX	XXXXXX	NI	6.61, 6.09+0.52
6	*Dra*I	A	OOOOO	XXXX	OO	OOO	OOOOOO	NI	7.42, 4.08+3.34
7	*Dra*I	A	XXXXX	XXXX	OO	OOO	XXXXXX	NI	3.34, 2.87+0.47
8	*Eco*RI	B	XXXXX	XXXX	OO	OOO^{1}	OOOOOO	X	4.42, 3.03+1.39
9	*Eco*RI	B	XXXXX	XXXX	OO	OOO	XXXXXX	X	7.71, 4.40+3.31
10	*Eco*RI	B	XXXXX	XXXX	OO	OOO	XXXXXX	X	4.73, 4.21+0.52
11	*Eco*RI	B	XXXXX	XXXX	XX	OOO	OOOO$O^{1}O^{1}$	X	1.93, 1.40+0.53
12	*Eco*RI	C	XXXXX	XXXX	XX	XXX^{1}	OOOO$O^{1}O^{1}$	X	4.67, 3.00+1.67
13	*Eco*RI	A	$X^{1}X$ $X^{1}X$ X	OOOO	XX	XXX	XXXXXX	X	10.48, 8.64+1.84

Table I.3. (continued)

OBSV	ENZY	PROB	MMMMM AACMGC NAOUE OMNAD	KKKK KOSS LAAI AKAE	BB MT ER NI	BBB MSM AOO RNN	BBBBBB SSSSSS AAAAAC BCLRVR	O O O	PHEN
14	*EcoRV*	A	X X X X X	O O O O	X X	X X X	X X X X X X	NI	5.15, 3.37+1.78
15	*EcoRV*	A	X X X X X	X X X X	O O	O O O	X X X X X X	NI	5.15, 4.70+0.45
16	*EcoRV*	A	X X X X X	X X X X	O O	X X X	X X X X X X	NI	4.25, NO
17	*Kpn*I	A	X X X X X	O O O O	X X	X X X	X X X X X X	X	22.10, 19.8+2.3
18	*Xba*I	A	O O O O O	X X X X	O O	O O O	O O O O O O	O	19.22, 11.24+6.19+1.79[2]
19	*Xba*I	A	O O O O O	X X X X	O O	O O O	O O O O O O	X	5.19, 4.68+0.51
20	*Xho*I	D	O O O O O	X X X X	O O	O O O	X X X X X X	NI	5.52, 4.25+1.27
21	*Xmn*I	A	X X X X X	O O O O	X X	X X X	X X X X X X	NI	8.25, 6.65+1.60
22	*Xmn*I	A	X X X X X	X X X X	O O	O O O	X X X X X X	NI	8.25, 7.60+0.65
23	*Xmn*I	A	O O O O O	X X X X	O O	O O O	O O O O O O	O	2.79, 2.70+0.70

Abbreviations: OBSV, observation number; ENZY, restriction enzyme; PROB, probe-set number; PHEN, fragment phenotypes observed and used to infer mutation (kb); O, restriction site present; X, restriction site absent; NI, not interpretable as single point mutation due to

distant populations of knobcone pine were monomorphic. Mean nucleotide diversity within populations (V_w) was 0.0036% for bishop pine, 0.0008% for Monterey pine, zero for knobcone pine, and 0.0015% for the species complex (Table I.4).

Nucleotide Diversity among Populations

Nucleotide diversity among populations was substantially higher than that within populations. Due to a complete lack of site polymorphism, no population differentiation was detected in knobcone pine. In Monterey pine, due to very low levels of polymorphism within populations and a lack of fixed differences among populations, subdivision was very low with all genetic parameters--ranging from 0 (negative estimates) to 2.14%. In contrast, in bishop pine population subdivision was extreme with all genetic parameters. Three of four values exceeded 97%, and G'_{ST} was above 87% (Tables I.5, I.6). This high degree of differentiation resulted from both strong, fixed differences among populations (mean V_b = 0.213%) as well as a paucity of diversity within populations (mean V_w = 0.004%).

Three regional groups of bishop pine populations accounted for the large majority of its subdivision: a northern region composed of Trinidad and Mendocino populations; an intermediate region composed of Marin,

Table I.4. Matrix of cpDNA nucleotide diversity (%) among and within populations and species based on 313 surveyed restriction sites. Interpopulation diversity is above diagonal, mean diversity within populations is on diagonal, and interspecies diversity is below diagonal (± standard error).

Table I.4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	MANO	MCAM	MMON	MGUA	MCED	KKLA	KOAK	KSAA	KSIE	BMEN	BTRI	BMAR	BSON	BMON	BSAB	BSAC	BSAL	BSAR	BSAV	BSCR
		Monterey				Knobcone				Bishop										
1.		0.0000	0.0001	-0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
2.			0.0000	0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
3.			0.0008 (±0.0008)	0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
4.				0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
5.						0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
6.							0.0000	0.0000	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
7.								0.0000	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
8.			0.3910 (±0.1154)					0.0000 (±0.0000)	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
9.										0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4388	0.4447
0.											0.0000	0.0290	0.0290	0.0321	0.3244	0.3244	0.3244	0.3244	0.3129	0.3188
1.												0.0290	0.0290	0.0321	0.3244	0.3244	0.3244	0.3244	0.3129	0.3188
2.													0.0000	0.0031	0.2952	0.2952	0.2952	0.2952	0.2897	0.2924
3.														0.0031	0.2952	0.2952	0.2952	0.2952	0.2897	0.2924
4.															0.2689	0.2689	0.2689	0.2689	0.2644	0.2666
5.			0.0909 (±0.0872)				0.4378 (±0.1601)							0.0036 (±0.0029)		0.0000	0.0000	0.0000	0.0003	0.0000
6.																	0.0000	0.0000	0.0003	0.0000
7.																		0.0000	0.0003	0.0000
8.																			0.0003	0.0000
9.																				-0.0003
0.																				

Table I.5. Hierarchical partitioning of cpDNA variation based on nucleotide diversities among (V_b) and within (V_w) populations (N_{ST}), and on haplotype frequencies (G'_{ST}) (Lynch and Crease 1990).

Level of analysis	Nucleotide diversity[1] (%) (±SE)	N_{ST} (%) (±SE)	G'_{ST} (%) (±SE)
Pooled total complex			
within populations	0.0022 (±0.0019)		
among populations	0.3073 (±0.0755)	99.29 (±0.50)	96.40 (±2.70)
Species			
within species	0.0391 (±0.0388)		
among species	0.3066 (±0.0949)	88.68 (±10.37)	84.14 (±13.20)
Populations within species			
Monterey pine			
within populations	0.0008 (±0.0008)		
among populations	-0.0000 (±0.0025)	-2.24 (±88.15)	-0.89 (undef.)

Table I.5. (continued)

Level of analysis	Nucleotide diversity[1] (%) (±SE)	N_{ST} (%) (±SE)	G'_{ST} (%) (±SE)
knobcone pine			
within populations	0.0000		
among populations	0.0000	undefined	undefined
bishop pine			
within populations	0.0036 (±0.0029)		
among populations	0.2135 (±0.0603)	98.33 (±1.43)	87.39 (±7.10)
Regional - bishop pine			
within regions	0.0054 (±0.0046)		
among regions	0.2103 (±0.0629)	97.49 (±1.95)	84.22 (±10.81)
Northern group[2]			
within populations	0.0000		
among populations	0.0000	undefined	undefined

Table I.5. (continued)

Level of analysis	Nucleotide diversity[1] (%) (±SE)	N_{ST} (%) (±SE)	G'_{ST} (%) (±SE)
Intermediate group[3]			
within populations	0.0076 (±0.0076)		
among populations	0.0031 (±0.0140)	28.76 (±88.41)	33.33 (±4.03)
Southern group[4]			
within populations	0.0028 (±0.0023)		
among populations	0.0002 (±0.0024)	6.61 (±65.80)	2.30 (undef.)

[1] Means (± standard error)

[2] The Northern group includes BMEN and BTRI.

[3] The Intermediate group includes BMAR, BSON, and BMON.

[4] The Southern group includes BSAL, BSAV, BSAB, BSAC, BSAR, and RSAC.

Table I.6. Hierarchical partitioning of gene diversity based on allelic frequencies for 313 surveyed restriction sites.

Level of analysis	Gene diversity (%)	G_{ST} (%)	F_{ST} (%) (±SE)
Pooled total complex			
within populations	0.0203		
among populations	2.6095	99.23	99.27 (±0.36)
Species			
within species	0.4311		
among species	2.0526	82.64	86.31 (±3.76)
Populations within species			
Monterey pine			
within populations	0.0086		
among populations	0.0002	2.14	-1.13 (±1.08)

Table I.6. (continued)

Level of analysis	Gene diversity (%)	G_{ST} (%)	F_{ST} (%) (±SE)
knobcone pine			
within populations	0.0000		
among populations	0.0000	undefined	undefined
bishop pine			
within populations	0.0342		
among populations	1.6417	97.92	97.03 (±1.71)
Regional - bishop pine			
within regions	0.0035		
among regions	0.0719	95.33	97.63 (±1.55)
Northern group[2]			
within populations	0.0000		
among populations	0.0000	undefined	undefined

Table I.6. (continued)

Level of analysis	Gene diversity (%)	G_{ST} (%)	F_{ST} (%) (±SE)
Intermediate group[3]			
within populations	0.0037		
among populations	0.0054	59.79	58.95 (±43.58)
Southern group[4]			
within populations	0.0001		
among populations	0.0000	6.00	1.45 (±0.00)

1 Means (± standard error)

[2] The Northern group includes BMEN and BTRI.

[3] The Intermediate group includes BMAR, BSON, and BMON.

[4] The Southern group includes BSAL, BSAV, BSAB, BSAC, BSAR, and RSAC.

Sonoma, and Monterey populations; and a southern region consisting of all 6 populations south of Monterey. Estimates of subdivision among the regions ranged from 84 to 98% (Tables I.5, I.6). When the species was subdivided into three regional groups and analyzed separately, its high degree of population subdivision disappeared for the northern (monomorphic) and southern groups, but remained high for the intermediate group for all parameters. G'_{ST} and N_{ST} were about 30%, and both G_{ST} and F_{ST} were nearly 60% (Tables I.5, I.6).

Nucleotide Diversity among Species

Nucleotide diversity among species far exceeded that among populations within species except in the case of bishop pine, where it was only marginally greater (V_b of 0.307% vs. 0.213%: table I.5). Relative to diversity within hierarchical levels, however, population subdivision in bishop pine exceeded that among species with all subdivision parameters (Tables I.5, I.6). For example, N_{ST} was 88.7% among species but 98.3% among bishop pine populations. Knobcone pine was most divergent from the other species, differing by an average of 0.414% from bishop and Monterey pines--which differed from one another by only 0.091% (Table I.4). Considering the species complex as a unit, subdivision among populations was very large--ranging from 96.4% to 99.3%--a consequence of the

large number of site mutations that are fixed among species and among populations within species.

Genomic Diversity

We discovered a single heteroplasmic individual in the Santa Cruz Island population of bishop pine. It possesses a combination of restriction fragments typical of both Santa Cruz bishop pine and Monterey pine (Fig. I.3). This combination of fragments from the two species RFLP's were observed with all assayed restriction enzyme/probe-set combinations that showed restriction site mutations between the two species (Fig. I.3). No heteroplasmy, however, was observed when a mitochondrial gene probe amplified from knobcone pine by the polymerase chain reaction was used (*Cox* I, unpubl. data).

We observed a rearrangement in two individuals in the southernmost population of bishop pine from San Vicente, Mexico. We inferred that an inversion was responsible because of hybridization with probe sets B and D gave two new pairs of bands that could not have been the result of simple site mutations, and each of whose sizes were equal when summed (8.2 + 6.9 = 15.1, and 6.2 + 8.9 = 15.1 kb) (Fig. I.4, I.5). The location of the inversion borders were further narrowed by hybridization with Douglas-fir clones (*Sst*I 6.7 & 4.8, and *Xba*I 2.5 & 3.8; Fig. I.2, I.4)

Figure I.3. Autoradiograms showing evidence for a heteroplasmic hybrid individual containing genomes of both Monterey and bishop pines (marked by *). Genomic DNA was cut with *EcoRI*, blotted, and probed with probe set B. (A) Santa Cruz population of bishop pine including a heteroplasmic individual, next to a member of the Año Nuevo population of Monterey pine (B). Note the presence of both Monterey and bishop fragments in the heteroplasmic tree. (C) (Santa Cruz bishop pine) and (D) (Año Nuevo Monterey pine) Genomic DNA from some of the same individuals in (A) and (B) was cut with *Xba*I, blotted, and probed with PCR-amplified mitochondrial *cox*I gene.

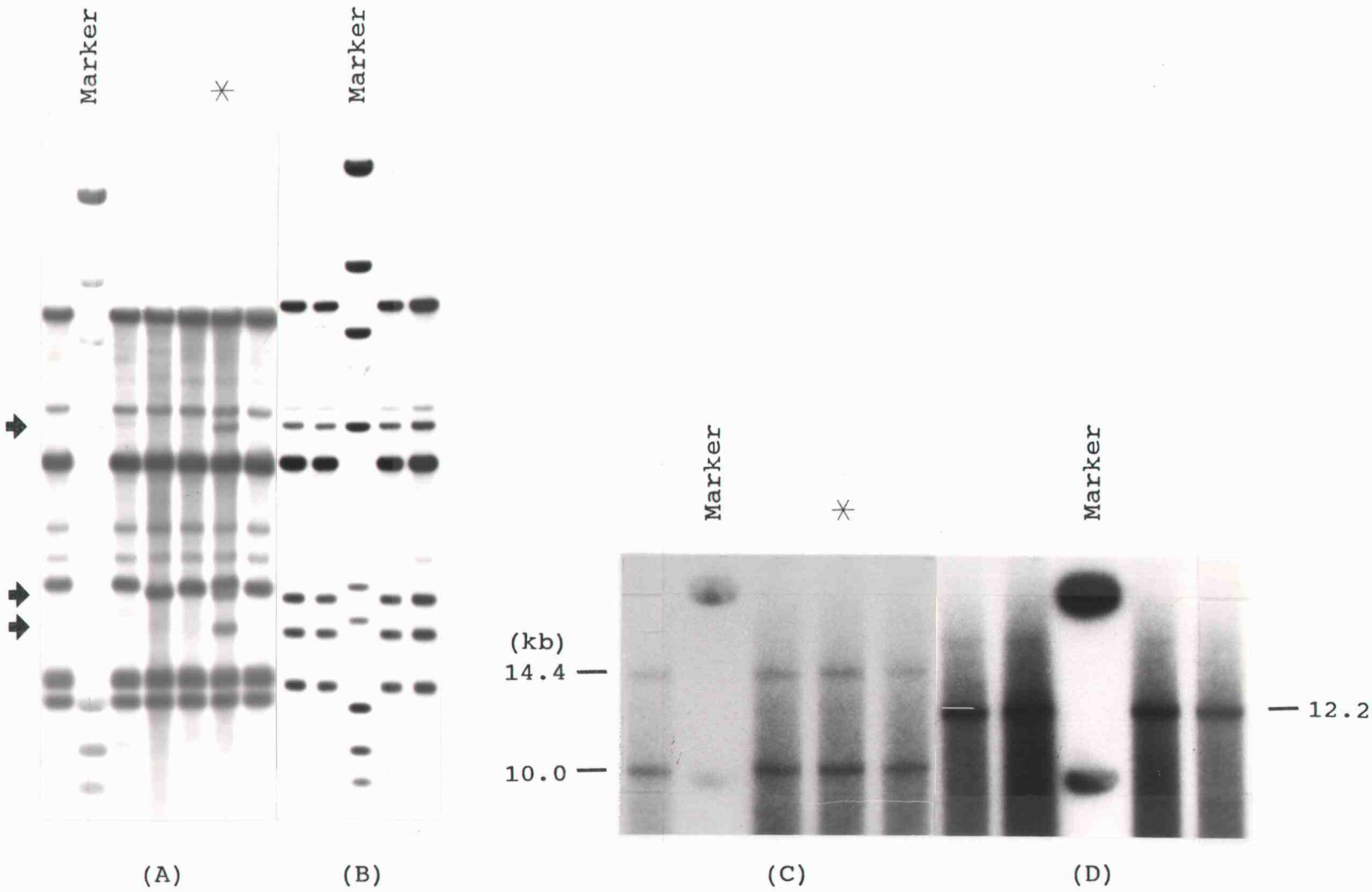


Figure I.3.

Figure I.4. Autoradiograms showing evidence for a cpDNA inversion. (A) Genomic DNA of seven individuals from the San Vicente population of bishop pine was cut with *Xho*I, blotted, and probed with *Xba*I-2.5 and *Xba*I-3.8 cloned fragments. (B) The same blot probed with *Sst*I-4.8 and *Sst*I-6.7 cloned fragments. Arrows indicate newly derived restriction fragments due to inversion (A: 6.2 ⇒ 8.2 kb; B: 8.9 ⇒ 6.9 kb). The sum of the sizes of the two new fragments (8.2 and 6.9 kb) is equal to the sum of the two homologous fragments (6.2 and 8.9 kb). See Fig. I.5 for schematic diagram of inversion. (C) and (D) show further evidence for inversion with a different restriction enzyme. (C) Genomic DNA from the same individuals in (A) was cut with *EcoRV*, blotted, and probed with the same fragments as in (A). (D) The same blot was probed with same fragments in (B). Note fragment stoichiometry again (4.7 + 4.3 = 5.0 + 4.0).

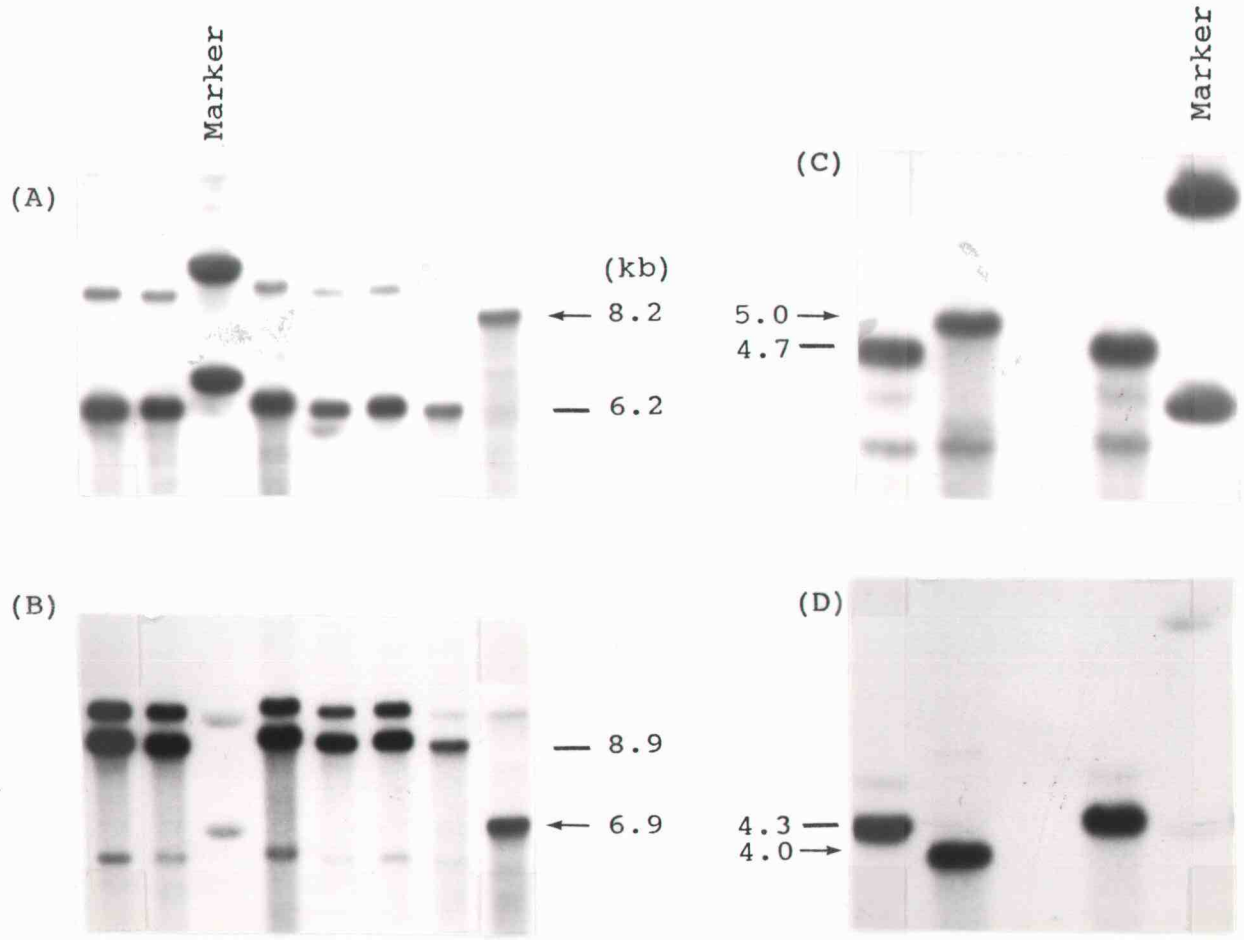


Figure I.4.

Figure I.5. Schematic diagram of cpDNA inversion. (A) Normal chloroplast genome. Arrow heads denote hypothesized inversion borders. A, B, C, and D denote hypothetical restriction sites. (B) Hypothesized "figure" showing putative intragenomic recombination event that gave rise to inversion. (C) Derived chloroplast genome with inversion. Note that the sum of the newly derived restriction fragments (C: AC + BD) should be approximately equal to the sum of fragments in the non-inverted genome (A: AB + CD).

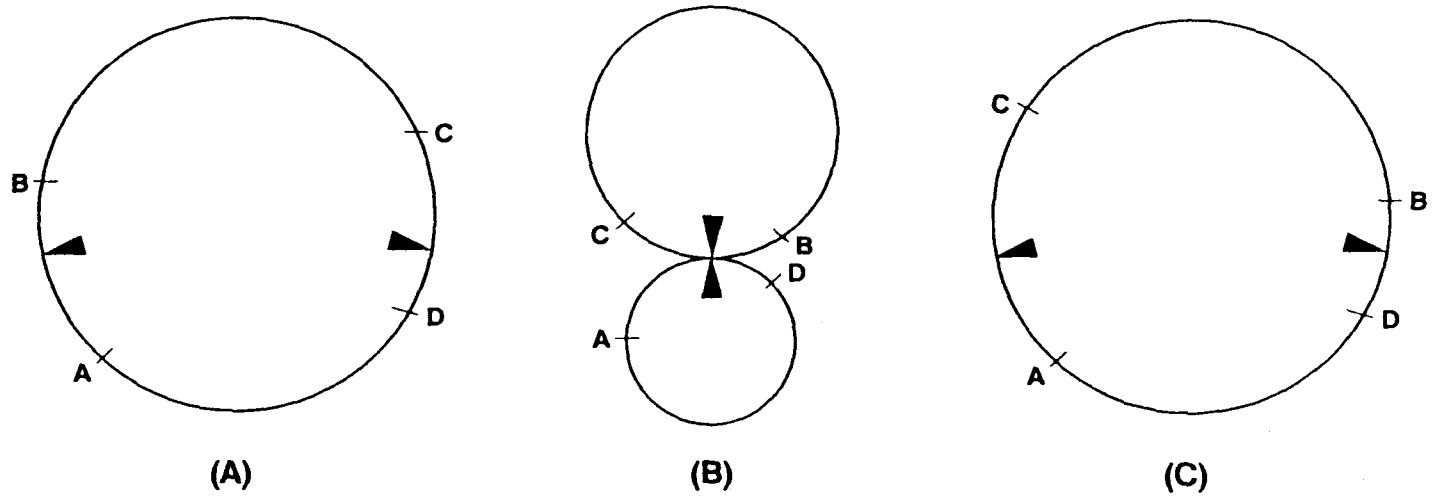


Figure I.5.

that mapped very near to the borders of a nearly 50 kb inversion that differentiates Monterey pine from Douglas-fir (Strauss et al. 1988). At our mapping resolution, the endpoints of this inversion and the San Vicente inversion are identical.

We observed that at least three different parts of the genome were subject to length mutations in our preliminary screen for polymorphism. Based on hybridization with a hotspot-containing clone from Douglas-fir (*Xba*-2.7), none of them coincided with the mutation hotspot observed there (Ali et al. 1991). Two of mutable regions were revealed by hybridizations with probe-set C, and one by probe-set A. Due to small changes in fragment sizes (< 100 bp), however, we were unable to score two of the three with confidence. The length variation revealed by probe-set A appeared to be composed of three distinct phenotypes (5.62, 5.53, and 5.30 kb) (Fig. I.6), and were thus scored in all individuals in our full survey. Five populations of Monterey pine showed polymorphism within populations for all three phenotypes (Table I.7). In knobcone pine, only the Santa Ana population showed polymorphism. All populations of bishop pine appeared to be fixed for the 5.62 kb fragment. The degree of population subdivision estimated from the length mutation frequencies suggested substantial subdivision, but well below that observed for site mutations: knobcone pine (G_{ST} = 63.64%, F_{ST} = 68.18%); Monterey pine (G_{ST} = 9.11%, F_{ST}

Table I.7. Observed frequencies of restriction fragment variants due to length mutations revealed by EcoRI digest and hybridization with probe set IV. Bishop pine was monomorphic for the 5.62 kb and thus is not shown.

	Fragment phenotype (kb)		
Population[1]	5.62	5.53	5.3
MANO	0.75	0.18	0.07
MCAM	0.73	0.23	0.04
MMON	0.46	0.25	0.29
MGUA	0.40	0.45	0.15
MCED	0.35	0.26	0.39
Mean	0.54	0.27	0.19
KKLA	1.00	0.00	0.00
KOAK	1.00	0.00	0.00
KSAA	0.30	0.00	0.70
KSIE	1.00	0.00	0.00
Mean	0.82	0.00	0.17

[1] Population abbreviations are given in Table 1.

Figure I.6. Autoradiogram showing a cpDNA length mutation. Genomic DNA of 16 individuals from the Año Nuevo population of Monterey pine was cut with *EcoRI*, blotted, and probed with probe set A.

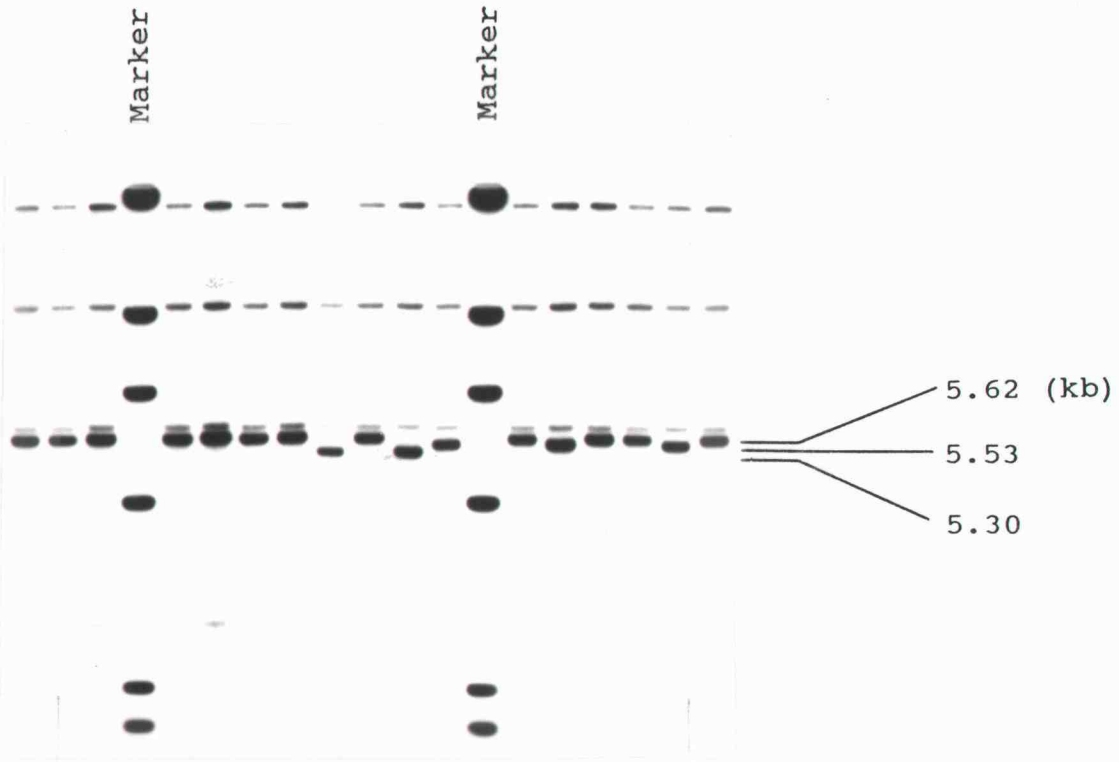


Figure I.6.

= 7.7%); among the three species (G_{ST} = 16.91%; F_{ST} = 23.64%); and among 20 populations of the complex pooled as a unit (G_{ST} = 41.05%; F_{ST} = 34.69).

DISCUSSION

We detected very little cpDNA restriction site polymorphism within or among species of the California closed cone pines. Surveys in herbaceous angiosperm species--often based on a limited number of samples per population--have likewise detected low amounts of cpDNA variation. Available estimates range from 1.55 to 3.24% among genera (mean of 2.4%: Dang and Pring 1986, Jansen and Palmer 1988), 0.1 to 1.62% among congeneric species (mean of 0.61: Crawford et al. 1990, Doebley et al. 1987, Jasen and Palmer 1988, Ogihara and Tsunewaki 1988, Palmer et al. 1983, 1985, Systma and Gottlieb 1986), and 0.007 to 0.076% among populations within species (mean of 0.041%: Banks and Birky 1985, Doebley et al. 1987, Soltis et al. 1987). Although cpDNA restriction fragment polymorphism has been studied in several woody plants, nucleotide diversities have not to our knowledge been calculated (Ali et al. 1991, Keim et al. 1989, Stine et al. 1989, Wagner et al. 1987, White 1990).

Our estimates of cpDNA nucleotide divergence among species (mean of 0.31%) are at the lower range of estimates from angiosperms. Our estimates of nucleotide diversity

among populations within species, however, which range from zero to 0.213% (mean of 0.071%), are at the upper range of estimates from angiosperms. The relatively high value of intraspecific diversity that we observed is mainly due to the high differentiation among geographic groups in bishop pine, which appear to be fixed for several cpDNA mutations.

The distribution of cpDNA nucleotide diversity that we observed among species--and among populations in bishop pine--contrasts markedly with that typically seen in long-lived woody species, and with that observed in a recent allozyme survey of the complex. Long-lived woody species usually maintain less than 10% of their genetic diversity among populations (Hamrick and Godt 1990). In a recent allozyme survey of the complex, Millar et al. (1988) reported that 12 to 22% of the total diversity resided among populations within species. Although these values are above-average compared to what is usually observed for long-lived perennials, they are several-fold less than our estimates for bishop pine (> 87%). Likewise, Millar et al. (1988) reported that 24% of the total diversity resided among species--again several-fold less than our estimates (> 84%). These results indicate that the relative strength of evolutionary forces shaping the organization of genetic diversity in chloroplast genomes can differ substantially from those affecting nuclear genomes.

As a consequence of the limited numbers of trees

surviving in the plantations sampled, a modest total number of trees were available for the intermediate (24) and northern (13) population groups of bishop pine (Table I.1). This problem was slightly compounded by the sampling of 2 trees in the northern group and 4 trees in the intermediate group that were half-siblings (open pollinated progeny) to another tree. Though paternal inheritance makes use of such siblings of less concern than would be the case for biparentally or maternally inherited genomes, such limited samples do not allow us to conclude whether alleles are indeed fixed in populations. They would, however, allow us to detect variants at low to moderate frequencies. For example, with a sample size of 11 (i.e., sampled trees from distinct families in the northern group), we can conclude with 95% confidence that an allele that is extremely rare or absent in other populations of bishop pine is present in frequencies over 72.0% in the northern group (Steel and Torrie 1980). Likewise, with a sample size of 20 (i.e., sampled trees from distinct families in the intermediate group), we can be 95% confident that alleles undetected elsewhere are present in frequencies over 83.1% in the intermediate group. Thus, the marked differences in cpDNA variant frequencies among population groups in bishop pine--while not necessarily due to complete fixation--cannot be regarded as an artifact of small samples. Nor can it explain the near fixation we observed in the southern group

of bishop pine (N = 111), or among species.

The small samples for the northern and intermediate population groups of bishop pine are a result of their undergoing substantially greater mortality than did the southern group of bishop pine in the hot, arid conditions of the Concord plantation. This could possibly have affected the cpDNA variation present. This is unlikely to have been important, however, as the samples of knobcone and Monterey pine, which had no such mortality, contained about the same or less diversity within populations than did these populations of bishop pine. Moreover, the intermediate group in bishop pine had more diversity than did the well-sampled, and better surviving, southern group.

The factors responsible for the differences in organization of genetic diversity between nuclear and chloroplast genes are unclear, but are likely to involve some combination of genetic drift, natural selection, and mutation. As a consequence of haploidy and predominant fixation within individuals (Birky 1983), the effective population size for organelle genes is expected to be one-half that of nuclear genes in monoecious species. Given equivalent rates of migration (see below), this is expected to result in roughly twice the genetic subdivision for neutral organelle genes than for nuclear genes at equilibrium [$F_{ST} \approx 1/(1 + 4Nm)$ for nuclear genes: Slatkin 1987]. Thus, effective population size alone does not seem

able to account for the more than three- to four-fold greater subdivision in bishop pine, nor for the nearly complete fixation of variants within populations and species.

Because of paternal inheritance of chloroplasts and monoecy in pines, the rate of migration is not likely to differ greatly between organelle and nuclear genomes. Paternal gametes are effectively dispersed twice--once as pollen and again as embryos within seeds. Thus, the average migration distance for paternally inherited genomes must be at least equal to, and usually greater, than that for maternally or biparentally inherited genomes. Moreover, because paternal gametes (pollen) are likely to on average disperse much farther than seeds, this would also tend to diminish relative subdivision for paternally inherited genomes. Thus, the different mode of inheritance of cpDNA compared to nuclear DNA does not help to explain the great subdivision we observed for cpDNA in pines.

Even for neutral variants, natural selection may play a much larger role in the genetic architecture of organelle genes than it does for nuclear genes. Because of the rarity of biparental inheritance and the lack of a sexual process to enable recombination among unlike genomes, linkage disequilibrium is expected to be very high in organelle genomes. This was illustrated by the presence of only five common and four rare haplotypes in our survey of

384 trees, despite observing 24 different mutations. This lack of recombination potentiates periodic selection--where selection for an advantageous mutation purges a population of polymorphism (Maruyama and Birky 1991). Intensive natural selection might be expected in species such as the closed cone pines which are adapting to cool climates as they migrate northward, but whose progenitors are subtropical in origin. The subsection *Oocarpa*e, and *P. oocarpa* in particular, are believed to represent the ancestors of the complex, and are subtropical (Mexican, Central American) in present distribution.

This does not explain, however, why only bishop pine, and not knobcone or Monterey pine, shows fixation among populations. This difference may result from a greater amount of historical gene flow among populations in knobcone and Monterey pine compared to bishop pine. Bishop pine shows almost twice as much population differentiation for allozymes than does either of the other species (22% vs. 12 and 13%, respectively), yet nearly equal amounts of allelic and gene diversity within populations (Millar et al. 1988). In the northernmost group of bishop pine--which appears to have migrated furthest--one population (Trinidad) has substantially reduced genetic diversity (expected heterozygosity of 7%), but the other does not (Mendocino: expected heterozygosity of 12%) when compared to the species average of 11.8% (Millar et al. 1988). Yet

both appear to be fixed for new cpDNA variants. The three species have similar levels of gene diversity within populations, ranging from 12 to 14% (Millar et al. 1988). These observations suggest that bottleneck events have not been important--and thus cannot explain--the apparent fixation of cpDNA variants among populations of bishop pine and among species in the complex.

Based on estimates of F_{ST} (Millar et al. 1988), the product of effective population size and migration rate (Nm) is less than one in bishop pine (0.89), which should permit substantial local differentiation, but greater than one in knobcone and Monterey pines (1.7 - 1.8), which is expected to inhibit differentiation [Slatkin 1987: $Nm \approx \{(1/F_{ST})-1\}/4]$. The lack of restricted diversity within populations, however, suggests that genetic drift has not been important, and therefore that effective population sizes large; the low value of Nm in bishop pine must therefore reflect extremely low migration. This is not surprising given that most populations are distant from one another and that the southern group of bishop pine is sexually isolated from the intermediate and northern groups (Critchfield 1967, Millar and Critchfield 1988). Thus, bishop pine populations--particularly the three geographic groups--are likely to have been less subject to the homogenizing effects of gene flow during their history than populations in the other species, perhaps facilitating the

fixation of favorable variants by periodic selection.

Mutation rates differ widely among cpDNA and nuclear genes, and may contribute to the greater subdivision observed for cpDNA. Birky et al. (1989) derived an equation for G_{ST} in organelle genomes that included mutation rate: $\hat{G}_{ST} \approx 1 / [1 + 2N_{eo}\{L/(L-1)\}(m_e + u)]$, where N_{eo} is the total population size for organelle genes, L is the number of subpopulations, m_e is the migration rates, and u is the mutation rate. This equation is also suitable for estimating subdivision for nuclear genes if $4N_e$ is substituted for $2N_{eo}$. Assuming migration is effectively zero among geographic groups of bishop pine (see above) and among species, the expression reduces to $\hat{G}_{ST} \approx 1 / [1 + 2N_{eo}\{L/(L-1)\}u]$ indicating that at equilibrium subdivision among isolated populations is a function of population number, size, and mutation rate. When $\hat{G}_{ST}$ was evaluated for a variety of population sizes using this equation and published cpDNA and allozyme mutation rates, nearly complete subdivision was found for cpDNA at all population sizes, but intermediate values of $\hat{G}_{ST}$ for allozymes when population sizes were in the vicinity of 150,000 (Fig. I.7). Although these estimates may be off by several-fold as a consequence of non-equilibrium conditions, it demonstrates that mutation rate may be a cause of the high subdivision we observed for cpDNA. Failure to reach equilibrium would likely results in a lower level of G_{ST} at

Figure I.7. Hypothetical estimates of G_{ST} for cpDNA and allozymes at equilibrium with different total population sizes. Calculations assume a mutation rate of 10^{-9} substitutions/site/year for cpDNA (Wolfe et al. 1987) and 2 X 10^{-6} for allozymes (Mukai and Cockerham 1977), and three populations.

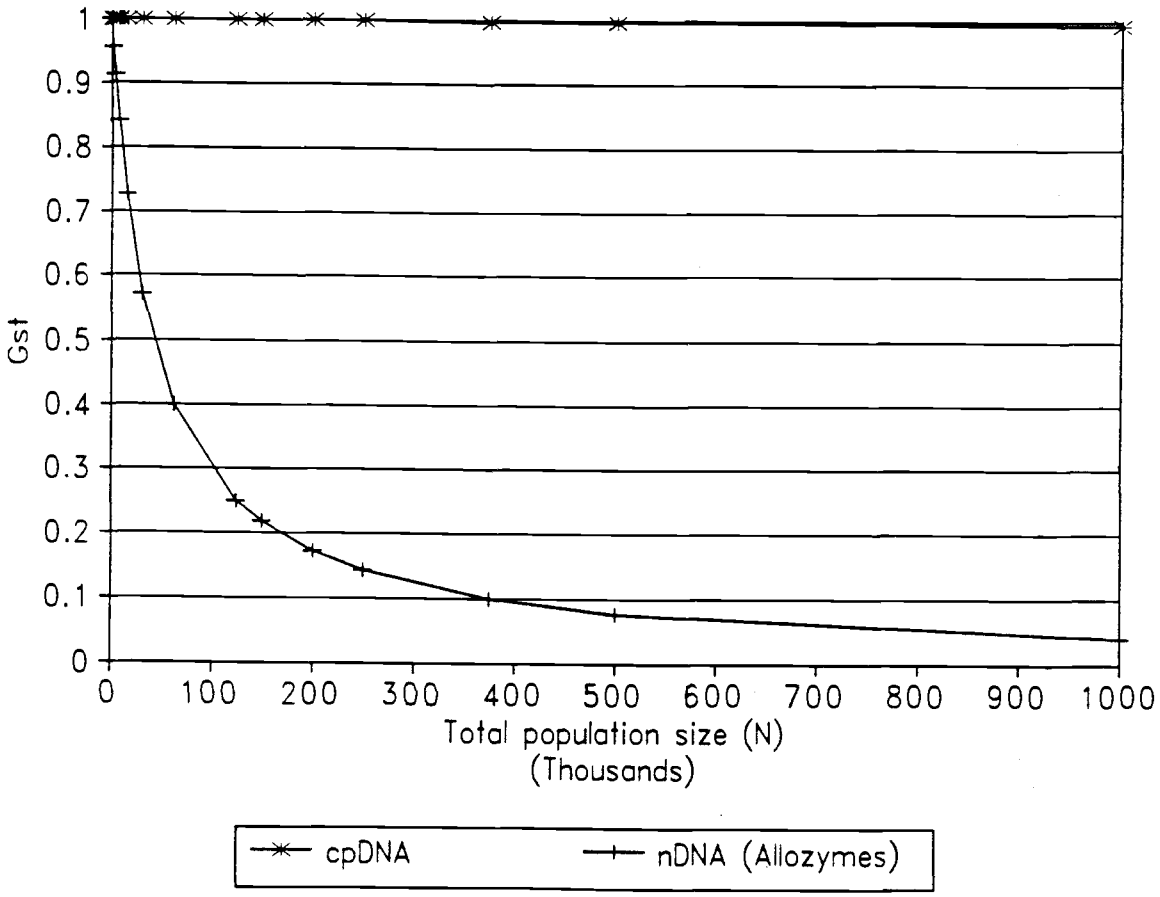

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Gst
0
100
200
300
400
500
600
700
800
900
1000
Total population size (N)
(Thousands)
cpDNA
nDNA (Allozymes)

Figure I.7.

a given population size--giving results in better agreement with realistic population sizes and estimates of G_{ST} for allozymes that were observed. The lack of subdivision in knobcone and Monterey pine is largely a consequence of their near lack of polymorphism--supporting the contention that the species are far short of subdivision-equilibrium. Although no quantitative theory is available to model non-equilibrium conditions, it appears that the low mutation rate for cpDNA can contribute to its high amount of subdivision by causing a slow rate of increase of within population diversity once other forces, such as drift and periodic selection, differentiate and deplete them of diversity.

We observed a single heteroplasmic individual in bishop pine that appears to be the result of a hybridization event--followed by biparental inheritance--with a Monterey pine pollen-parent. Monterey pine and southern bishop pine are interfertile in controlled crosses (Critchfield 1967), and occasional biparental inheritance following hybridization appears to be common in conifers (Govindaraju et al. 1988, Wagner et al. 1987, White 1990)--indicating a breakdown of the mechanism causing normal paternal inheritance. Although no natural populations of Monterey pine occur on Santa Cruz island, it is widely planted throughout California, and used extensively as a Christmas tree. Alternatively, it is possible that

Monterey pine chloroplast genomes are naturally present in the Santa Cruz population of bishop pine. Santa Cruz appears to be the most ancestral, and closely related to Monterey pine, of all bishop pine populations (Millar et al. 1988). Though otherwise undetected in our survey, it is conceivable that some remnant Monterey pine genomes are naturally present in low frequency in the Santa Cruz population.

We observed three regions of the genome subject to length mutations. One region contained insertion/deletions of sufficient length to enable their consistent recognition. The organization of genetic diversity for these variants bore no obvious resemblance to that observed from study of site mutations. Although we provided estimates of allele frequencies and subdivision, we recognize that homology among length variants cannot be confidently assessed with the resolution provided by our agarose gels. Our estimates must therefore be considered highly tentative.

The presence, and degree of polymorphism, of such length mutation-prone loci differs widely among species (Ali et al. 1990)--probably related to the presence and extent of repetitive DNA (Hipkins and Strauss, unpubl. data). Such regions are therefore likely to represent an unusual genetic locus where mutation rate--rather than other factors--predominates as an evolutionary force.

Depending on the presence/absence, number, and size of repeats in such regions, the mutation and reversion rates--and the ability to detect mutations (size variants)--could vary dramatically among populations, species, and over time. In addition to possible difficulties with determining allelic homologies, this may be a major cause of the lack of concordance with estimates of subdivision from site mutations and allozymes (Millar et al. 1988).

We observed two trees from different open-pollinated families in the San Vicente population of bishop pine that contained an identical major genome rearrangement. Although the rearrangement was not mapped in detail, it appeared to result from an inversion of about 50 kb similar to another inversion observed among genera of conifers (Strauss et al. 1988). Thus, as suggested by Tsai and Strauss (1990), convergent rearrangements may occur in conifers, and rearrangements should therefore be used with caution as phylogenetic markers. Ali et al. (1990) also observed a single redwood with a major, unmapped cpDNA rearrangement. The major changes of genome structure observed in "macroevolutionary" studies of conifer evolution can therefore also be seen during "microevolutionary" surveys--suggesting that no novel evolutionary events need to be postulated to explain their origin. They exist as infrequent forms in natural populations, which could then proceed to fixation during or

after speciation.

The cpDNA genetic patterns discovered in the closed cone pines raise questions that need to be addressed by comparable studies of other taxa and genomes. First, are these chloroplast genomes somehow unusual, or will relatively high degrees of subdivision be found in other conifers and plants? This is of interest because the closed cone pines, and bishop pine in particular, exists in a number of disjunct populations and shows higher than average subdivision for a conifer. How will continuously distributed plant species with low levels of nuclear gene subdivision compare? Second, are the results found for cpDNA generalizable for organelle DNA? A survey of mitochondrial DNA would help to answer this question. Such a study would help to evaluate the causes of the high cpDNA subdivision observed. If periodic selection, rather than genetic drift or mutation rate, is important, little correlation among cpDNA and mitochondrial DNA patterns of subdivision would be expected. Finally, analysis of nuclear genomes at the DNA rather than the gene product level would facilitate comparisons with organelle DNA. This could be accomplished using either standard RFLP methods, or the new random amplified polymorphic DNA (RAPD) technique (Williams et al. 1990). Polymorphic nuclear probes and RAPD primers are now becoming widely available as a result of genome mapping efforts.

ACKNOWLEDGEMENTS

We thank Dr. G. Furnier who sampled the Santa Ana population of knobcone pine for us, Dr. M. Lynch who provided the computer program of HAPLO, and G. T. Howe who helped in field collections of samples. We are also grateful to Dr. W.J. Libby, Dr. F.T. Ledig, and U.S. Forest Service, Pacific Southwest Experiment Station in Berkeley, California, for the use of plantations. This study was supported by National Institutes of Health Grant GM 36367.

LITERATURE CITED

Ali, I. F., D. B. Neale, and K. A. Marshall. 1991. Chloroplast DNA restriction fragment length polymorphism in coast redwood, Douglas-fir, incense-cedar, and loblolly pine. Theor. Appl. Genet. 81: 83-89.

Banks, J. A. and C. W. Birky, Jr. 1985. Chloroplast DNA diversity is low in a wild plant, Lupinus texensis. Proc. Natl. Acad. Sci. USA 82: 6950-6954.

Birky, C. W. Jr. 1983. Relaxed cellular controls and organelle heredity. Science 222: 468-475.

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

Coates, D. and C. A. Cullis. 1987. Chloroplast DNA variability among Linum species. Amer. J. Bot. 74: 260-268.

Crawford, D. J., J. D. Palmer, and M. Kobayashi. 1990. Chloroplast DNA restriction site variation and the phylogeny of Coreopsis section Coreopsis (Asteraceae). Amer. J. Bot 77: 552-558.

Critchfield, W. B. 1967. Crossability and relationships of the California closed cone pines. Silvae Genet. 16: 89-97.

Dang, L. H. and D. R. Pring. 1986. A physical map of the sorghum chloroplast genome. Plant Mol. Biol. 6: 119-123.

Doebley, J., W. Renfroe, and A. Blanton. 1987. Restriction site variation in the Zea chloroplast genome. Genetics 117: 139-147.

Govindaraju, D. R., B. P. Dancik, and D. B. Wagner. 1988. Novel chloroplast DNA polymorphism in a sympatric region of Two pines. J. Evol. Biol. 2: 49-59.

Jansen, R. K. and J. D. Palmer. 1988. Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). Amer. J. Bot. 75: 753-766.

Keim, P., K. N. Paige, T. G. Whitham, and K. G. Lark. 1989. Genetic analysis of an interspecific hybrid swarm of Populus: occurrence of unidirectional introgression. Genetics 123: 557-565.

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

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.

Millar, C. I. 1989. Allozyme variation of bishop pine associated with pygmy-forest soils in northern California. Can. J. For. Res. 19: 870-879.

Millar, C. I., S. H. Strauss, M. T. Conkle, and R. Westfall. 1988. Allozyme differentiation and biosystematics of the Californian closed-cone pines. Syst. Bot. 13: 351-370.

Neale, D. B., M. A. Saghai-Maroof, R. W. Allard, Q. Zhang, and R.A. Jorgensen. 1988. Chloroplast DNA diversity in populations wild and cultivated barley. Genetics 120: 1105-1110.

Nei, M. 1977. F-statistics analysis of gene diversity in subdivided populations. Ann. Hum. Genet., Lond 41: 225-233.

Nei, M. and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105: 207-217.

Ogihara, Y. and K. Tsunewaki. 1988. Diversity and evolution of chloroplast DNA in Triticum and Aegilops as revealed by restriction fragment analysis. Theor. Appl. Genet. 76: 321-332.

Palmer, J. D., C. R. Shields, D. B. Cohen, and T. J. Orton. 1983. Chloroplast DNA evolution and the origin of amphidiploid Brassica species. Theor. Appl. Genet. 65: 181-189.

Palmer, J. D., R. A. Jorgensen, and W. F. Thompson. 1985. Chloroplast DNA variation and evolution in Pisum : Patterns of change and phylogenetic analysis. Genetics 109: 195-213.

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

Reed, K. C. and D.A. Mann. 1985. Rapid DNA transfer from agarose gels to nylon membranes. Nucleic Acids Res. 13: 7207-7221

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science 236: 787-792.

Soltis, D. E., P. S. Soltis, T. A. Ranker, and B. D. Ness. 1988. Chloroplast DNA variation in a wild plant, Tolmiea menziesii. Genetics 121: 819-826.

Steel, R. G. D. and J. H. Torrie. 1980. Principles and procedures of Statistics, a biometrical approach (second ed.). McFraw-Hill Book Co. New York.

Stine, M., B. B. Sears, and D. E. Keathley. 1989. Inheritance of plastids in interspecific hybrids of blue spruce and white spruce. Theor. Appl. Genet. 78: 768-774.

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

Strauss, S. H., A. H. Doerksen, and J. R. Byrne, 1990 Evolutionary relationships of Douglas-fir and its relatives (genus Pseudotsuga) from DNA restriction fragment analysis. Can. J. Bot. 68: 1052-1510.

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.

Sytsma, K. J. and L. D. Gottlieb. 1986. Chloroplast DNA evolution and phylogenetic relationships in Clarkia sect. Peripetasma (Onagraceae). Evolution 40: 1248-1261.

Tsai, C.-H. and S. H. Strauss. 1989. Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr. Genet. 16: 211-218.

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.

Weir, B. S. 1990. Intraspecific differentiation. pp. 373-410 in Molecular Systematics, edited by D.M. Hills and C. Moritz. Sinauer Associates, Inc. Massachusetts.

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

White, E.E. 1990. Chloroplast DNA in Pinus monticola - 2. Survey of within-species variability and detection of heteroplasmic individuals. Theor. Appl. Genet. 79: 251-255.

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 Acies Res. 18: 6531-6535.

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 88: 9054-9058.

Zurawski, G. and M. T. Clegg. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Ann. Rev. Plant Physiol. 38: 391-418.

CHAPTER II

PHYLOGENETIC RELATIONSHIPS IN THE CALIFORNIA CLOSED CONE PINES BASED ON CHLOROPLAST DNA RESTRICTION SITE ANALYSIS

ABSTRACT

We studied phylogenetic relationships among populations and species via chloroplast DNA restriction site analysis in the California closed cone pines (Pinus radiata D. Don, P. attenuata Lemm., and P. muricata D. Don). Genomic DNAs from 384 trees derived from 20 populations in the species complex, and from a single tree of an outgroup species, P. oocarpa Schiede, were digested with 20 restriction enzymes, blotted, and probed with cloned chloroplast DNA fragments from Pseudotsuga menziesii (Mirb.) Franco. Phylogenetic trees were constructed by bootstrapped Wagner and Dollo parsimony analyses of site gains and losses, and by three distance-matrix analyses of estimated nucleotide differentiation.

Despite very little chloroplast DNA polymorphism in the complex, a number of phylogenetically informative mutations were discovered at the population and species levels. Of 313 restriction sites surveyed, 24 were polymorphic in the complex, and 23 of these were phylogenetically informative (partially shared) among

populations or species. Parsimony analysis aligned all of the populations with their taxonomically designated species, and revealed three statistically significant ($P < 0.05$) clades: the species complex relative to the outgroup *P. oocarpa*; a *P. radiata*-*P. muricata* clade separate from *P. attenuata* (Dollo only); and the intermediate and northern populations of *P. muricata* separate, and derived from, its southern populations. Neighbor-joining and Fitch-Margoliash trees corroborated the parsimony results, indicating a close relationship of *P. radiata* and *P. muricata*--particularly with the southern *P. muricata* populations--and a substantial distance of these species from *P. attenuata*. They also suggested heterogeneous rates of evolution in the complex: *P. attenuata* evolved two-fold more rapidly than *P. radiata*, followed by northern and intermediate populations of *P. muricata*, which evolved three-fold more rapidly than *P. radiata*. UPGMA appeared to give misleading results, clustering the slowly evolving populations within *P. muricata* and *P. radiata* as a single clade.

* List of abbreviations:

bp, base pair; BP, before present; cpDNA, chloroplast DNA; kb, kilo base pair; OTU, operating taxonomic unit; RFLP, restriction fragment length polymorphism.

INTRODUCTION

The California closed cone pines are a complex of closely related species within subsection Oocarpae (Critchfield and Little 1966), and include bishop (P. muricata D. Don), Monterey (P. radiata D. Don), and knobcone pines (P. attenuata Lemm.). Bishop and Monterey pines are maritime species, rarely are found more than 12 km from the ocean, whereas knobcone pine is largely an interior montane species, occupying a variety of elevations in the Sierra Nevada, Klamath Mountains, San Bernardino Mountain, Santa Ana Mountain, and Coast Range of California (Fig. II.1). Systematic botanists have studied the complex for nearly two centuries, employing a wide variety of traits--including morphology, crossability, secondary compound chemistry, and allozyme frequencies (reviewed in Millar 1986, Millar et al. 1988). The complex has been characterized as a group of three or more species, often with high amounts of genetic diversity among its many disjunct populations.

Despite a number of phylogenetic analyses in the complex, many questions remain about relationships among populations and species. Distribution and habitat considerations suggest that Monterey and bishop pine are most closely related, however, based on crossability knobcone and Monterey pine are most closely related. Bishop pine is one of the very few conifer species known

Figure II.1. Distribution of P. radiata, P. attenuata, and P. muricata, showing the sampled populations.

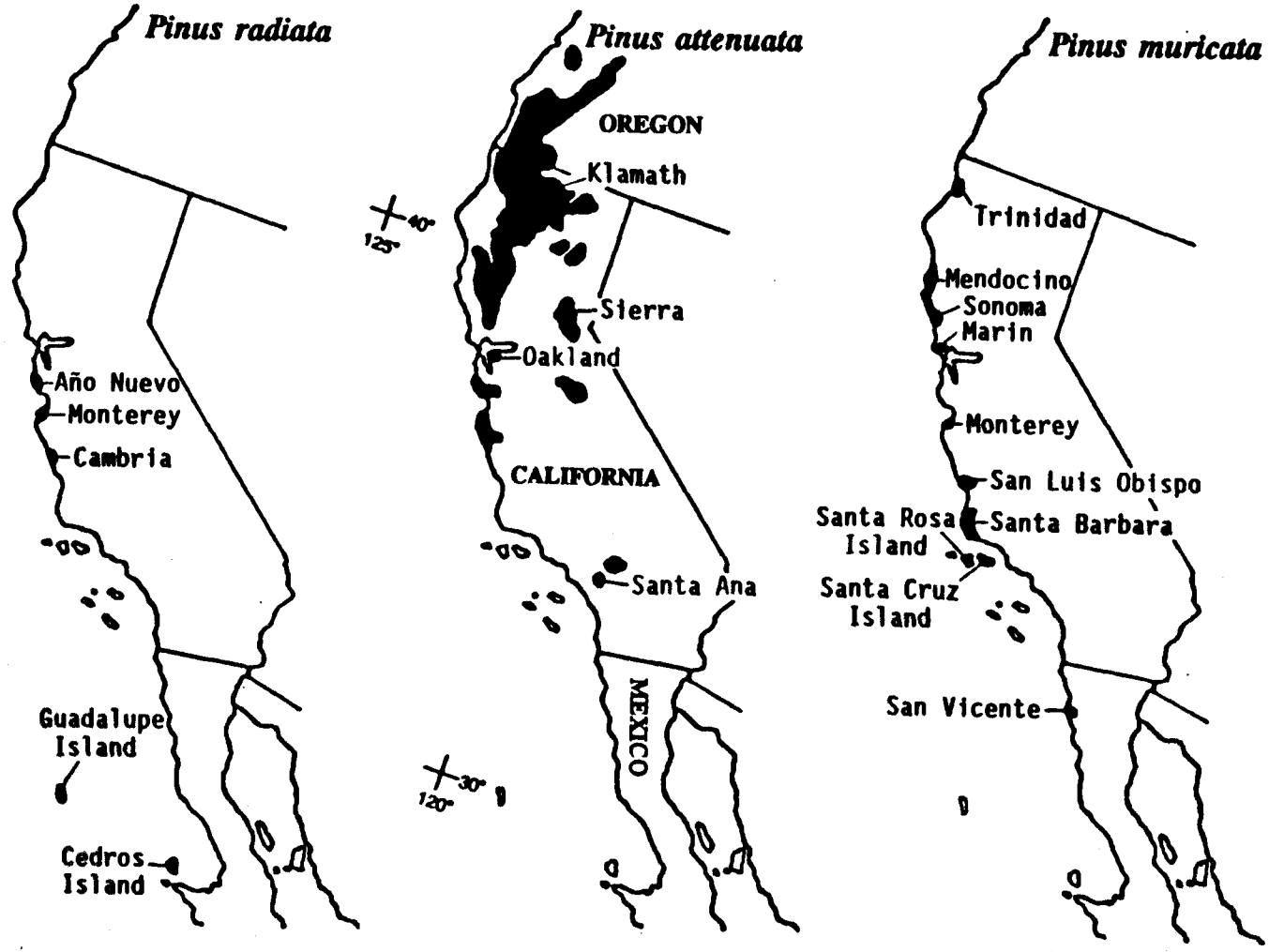

Pinus radiata
Año Nuevo
Monterey
Cambria
Guadalupe Island
Cedros Island
Pinus attenuata
OREGON
Klamath
40°
125°
Sierra
Oakland
CALIFORNIA
Santa Ana
MEXICO
30°
120°
Pinus muricata
Trinidad
Mendocino
Sonoma
Marin
Monterey
San Luis Obispo
Santa Barbara
Santa Rosa Island
Santa Cruz Island
San Vicente

Figure II.1.

where strong barriers to hybridization have developed within species, but incomplete barriers exist among species; its Mendocino populations cross very poorly with the southern populations, though the southern populations cross readily with knobcone and Monterey pines (Critchfield 1967, Millar and Critchfield 1988).

Morphological and biochemical traits have sometimes given conflicting pictures of relationships--both among themselves and with one another. For example, in Monterey pine, allozyme frequencies and terpene composition differed from cone and bark morphology in aligning the mainland populations (reviewed in Plessas and Strauss 1986). Allozyme analyses of relationships among populations or species differed depending on the method of phylogenetic tree construction (Millar et al. 1988) and loci studied (cf. Plessas and Strauss 1986, Moran et al. 1988).

The difficulties with phylogenetic analyses of the complex result, at least partly, from the characteristics of the traits available for study. First, they have often been highly polymorphic within populations and species. This means that inadequate sampling can bias results unless sample sizes are large. Such sampling-related problems have been observed when comparing independent allozyme surveys--both in the loci sampled and the allele frequencies derived when common loci were studied (cf. Plessas and Strauss 1986, Moran et al. 1987, Millar et al.

1988). The problem of high polymorphism is compounded by the relative youth of the complex, whose common ancestor is thought to have arisen in the Oligocene (26-38 million years before present) (reviewed in Millar et al. 1988). Populations and species arose much later, and are therefore weakly differentiated; the majority of genetic diversity resides within populations in any one species (Millar et al. 1988). Second, morphological traits such as cone morphology have been emphasized in taxonomic studies (Axelrod 1983, Howell 1941, Linhart 1978, Linhart et al. 1967, Mason 1930). Such traits are often strongly subject to natural selection, which can result in convergent evolution and variability in evolutionary rates. Plessas and Strauss (1986) discuss a case where adaptation to drought may have accelerated evolution of cone morphology in a population of Monterey pine, confounding phylogenetic interpretations.

Study of chloroplast DNA should give new phylogenetic insights (reviewed in Strauss et al. 1991). Because of its conservative nature--evolving slowly both in sequence and genome structure--homologous mutations can be readily identified in divergent populations and species (Palmer 1990). This results in extremely low incidences of parallelism and convergence at the interspecific level (Palmer and Zamir 1982, Palmer et al. 1983, 1985), and substantially lessens the need for large samples to

characterize taxa. Because of the rareness of biparental inheritance and recombination among diverged genomes (Kirk and Tilney-Bassett 1978, Sears 1980), chloroplast DNA phylogenies can be interpreted without the complications of recombination and gene conversion that affect nuclear genes. Finally, its slow rate of evolution facilitates interpretation of restriction fragment differences in terms of individual mutations, facilitating cladistic and statistical analyses of phylogenetic patterns. We demonstrate that analyses of chloroplast DNA provide a number of new insights that substantially extend systematic knowledge of this well-studied species complex.

MATERIALS AND METHODS

Plant Materials

We sampled a total of 384 trees originating from 20 populations (Fig. II.1; Table II.1). All samples of knobcone pine were from trees growing in natural populations. The Klamath population was sampled over an 2 mile transect adjacent to U.S. Interstate 5 north of Shasta Lake Recreation Area, California (latitude 40°50', longtitude 122°45'). The Sierra Nevada population was sampled over an 2 mile transect located adjacent to U.S. Interstate 80 near Auburn, California (latitude 38°54', longtitude 121°08'). The Oakland population was sampled over an 1.5 mile transect along Flicker Ridge, in the hills

Table II.1. Sample origins and sizes.

Species	Populations	No. of Sample Trees	Abbreviation	Origin[1]
P. radiata	Año Nuevo	30	MANO	G
	Cambria	28	MCAM	G
	Cedros Island	27	MCED	G & R
	Guadalupe Island	24	MGUA	G & R
	Monterey	27	MMOT	G
P. attenuata	Klamath	25	KKLA	Nat. Population
	Oakland	25	KOAK	Nat. Population
	Sierra	25	KSIE	Nat. Population
	Santa Ana	25	KSAA	Nat. Population

Table II.1. (continued)

Species	Populations	No. of Sample Trees	Abbreviation	Origin[1]
P. muricata	Mendocino	6	BMEN	C & R
	Trinidad	7	BTRI	R
	Marin	5	BMAR	C & R
	Sonoma	7	BMSO	R
	Monterey	12	BMON	C & R
	San Luis Obispo	12	BSAL	C & R
	San Vicente	20	BSAV	C & R
	Santa Barbara	19	BASB	C & R
	Santa Cruz-bishop	22	BSAC	C & R
	Santa rosa	17	BSAR	C & R
	Santa Cruz-remorata	21	BSCR	C & R
TOTAL	20 Populations	384 Sample trees		

[1] C = Concord plantation, G = Gill Tract, R = Russell Reservation

east of Oakland, California (latitude 37°50', longtitude 122°30'). The Santa Ana population was sampled by Dr. Glenn Furnier from widely spaced trees in the Santa Ana mountains near Corona, California (latitude 33°50', longitude 117°37').

Bishop and Monterey pines were sampled from gene conservation and genetic test plantations in three locations: Albany, California (Gill Tract) and LaFayette, California (Russell Reservation) established by Dr. W.J. Libby of the University of California at Berkeley; and the U.S. Naval Weapons Research Station in Concord, California established by the U.S. Forest Service, Pacific Southwest Experiment Station in Berkeley, California (maintained by Dr. F.T. Ledig). All of the trees in these plantations were either clonal replicates of trees from natural populations, or derived from seeds collected in natural populations. Fifty-six of the total 284 trees collected in these populations were known to be open-pollinated siblings of another sampled tree (discussed below).

DNA Analysis

Total cellular DNA was extracted from 40 g of sampled needles using a CTAB-based DNA extraction method (Wagner et al. 1987). We modified this method only in that we worked in a cold room; used liquid nitrogen, rather than extraction buffer, for initial grinding of needles; and

shook tubes vigorously during denaturation of membranes in sarkosyl because of the high viscosity of our extractions. DNA was then digested with restriction enzymes according to manufacturers suggestions and fractionated according to size by electrophoresis in 0.8% or 1.5% agarose gels in a TAE buffer (80 mM Tris, 16.6 mM sodium acetate, 2 mM EDTA, adjusted to PH 8.1 with glacial acetic acid). The gels were then blotted in an alkaline denaturation solution (Reed and Mann 1985) onto Zetabind nylon membranes (Cuno Inc., Meriden,CT), and hybridized with four mixtures of equal amounts of cloned Douglas-fir cpDNA fragments (Fig. II.2). The clones were radioactively labelled with ^{32}P by random-hexamer-primer-extension (Feinberg and Vogelstein 1983) and together covered about 85% of the chloroplast genome. Hybridization, washing, and stripping of blots was as described in Strauss and Doerksen (1990).

Strategy for Surveying Polymorphisms

We used a two-stage screening procedure to detect cpDNA polymorphisms. For the initial screen, we randomly chose two trees from each of the 20 sampled populations in the complex, plus a single tree of Pinus oocarpa as an outgroup [obtained from Botanical Garden in Berkeley, California, Accession # 67.164; native in Valle de Bravo, Mexico (latitude 13°14', longtitude 100°07')]. We then digested these DNAs with 17 restriction enzymes with six bp

Table II.2. Restriction fragment phenotypes used to infer site mutations.

OBSV	ENZY	PROB	MMMMM ACMGC NAOUE OMNAD	KKKK KOSS LAAI AKAE	BB MT ER NI	BBB MSM AOO RNN	BBBBBB SSSSSS AAAAAC BCLRVR	OO CA	PHEN
1	*Bam*HI	A	X X X X X	O O O O	X X	X X X	X X X X X X	NI	2.84, 2.02+0.62
2	*Bam*HI	A	O O O O O	X X X X	O O	O O O	O O O O O O	O	3.17, 3.10+0.07
3	*Bcl*I	A	O O O O O	O O O O	X X	X X X	O O O O O O	O	7.98, 7.77+0.21
4	*Bcl*I	A	X X X X X	X X X X	O O	O O O	X X X X X X	X	5.09, 4.57+0.52
5	*Dra*I	A	X X X X X	O O O O	X X	X X X	X X X X X X	NI	6.61, 6.09+0.52
6	*Dra*I	A	O O O O O	X X X X	O O	O O O	O O O O O O	NI	7.42, 4.08+3.34
7	*Dra*I	A	X X X X X	X X X X	O O	O O O	X X X X X X	NI	3.34, 2.87+0.47
8	*Eco*RI	B	X X X X X	X X X X	O O	O O O^1	O O O O O O	X	4.42, 3.03+1.39
9	*Eco*RI	B	X X X X X	X X X X	O O	O O O	X X X X X X	X	7.71, 4.40+3.31
10	*Eco*RI	B	X X X X X	X X X X	O O	O O O	X X X X X X	X	4.73, 4.21+0.52
11	*Eco*RI	B	X X X X X	X X X X	X X	O O O	O O O O O^1 O^1	X	1.93, 1.40+0.53
12	*Eco*RI	C	X X X X X	X X X X	X X	X X X^1	O O O O O^1 O^1	X	4.67, 3.00+1.67
13	*Eco*RI	A	X^1X X^1X X	O O O O	X X	X X X	X X X X X X	X	10.48, 8.64+1.84

Table II.2. (continued)

OBSV	ENZY	PROB	MANO	MCAM	MMON	MGUA	MCED	KKLA	KOAK	KSAA	KSIE	BMEN	BTRI	BMAR	BSON	BMON	BSAB	BSAC	BSAL	BSAR	BSAV	BSCR	OOCA	PHEN
14	EcoRV	A	X	X	X	X	X	O	O	O	O	X	X	X	X	X	X	X	X	X	X	X	NI	5.15, 3.37+1.78
15	EcoRV	A	X	X	X	X	X	X	X	X	X	O	O	O	O	O	X	X	X	X	X	X	NI	5.15, 4.70+0.45
16	EcoRV	A	X	X	X	X	X	X	X	X	X	O	O	X	X	X	X	X	X	X	X	X	NI	4.25, NO
17	KpnI	A	X	X	X	X	X	O	O	O	O	X	X	X	X	X	X	X	X	X	X	X	X	22.10, 19.8+2.3
18	XbaI	A	O	O	O	O	O	X	X	X	X	O	O	O	O	O	O	O	O	O	O	O	O	19.22, 11.24+6.19+1.79[2]
19	XbaI	A	O	O	O	O	O	X	X	X	X	O	O	O	O	O	O	O	O	O	O	O	X	5.19, 4.68+0.51
20	XhoI	D	O	O	O	O	O	X	X	X	X	O	O	O	O	O	X	X	X	X	X	X	NI	5.52, 4.25+1.27
21	XmnI	A	X	X	X	X	X	O	O	O	O	X	X	X	X	X	X	X	X	X	X	X	NI	8.25, 6.65+1.60
22	XmnI	A	X	X	X	X	X	X	X	X	X	O	O	O	O	O	X	X	X	X	X	X	NI	8.25, 7.60+0.65
23	XmnI	A	O	O	O	O	O	X	X	X	X	O	O	O	O	O	O	O	O	O	O	O	O	2.79, 2.70+0.70

Abbreviations: OBSV, observation number; ENZY, restriction enzyme; PROB, probe-set number; PHEN, fragment phenotypes observed and used to infer mutation (kb); O, restriction site present; X, restriction site absent; NI, not interpretable as single point mutation due to

Figure II.2. Autoradiogram showing evidence for cpDNA restriction site and length mutations within the Año Nuevo population of Monterey pine. All trees shown are from the Año Nuevo population. Arrows indicate the new fragments resulting from the restriction site mutation. The asterisk indicates the position of a fragment containing a small deletion/insertion.

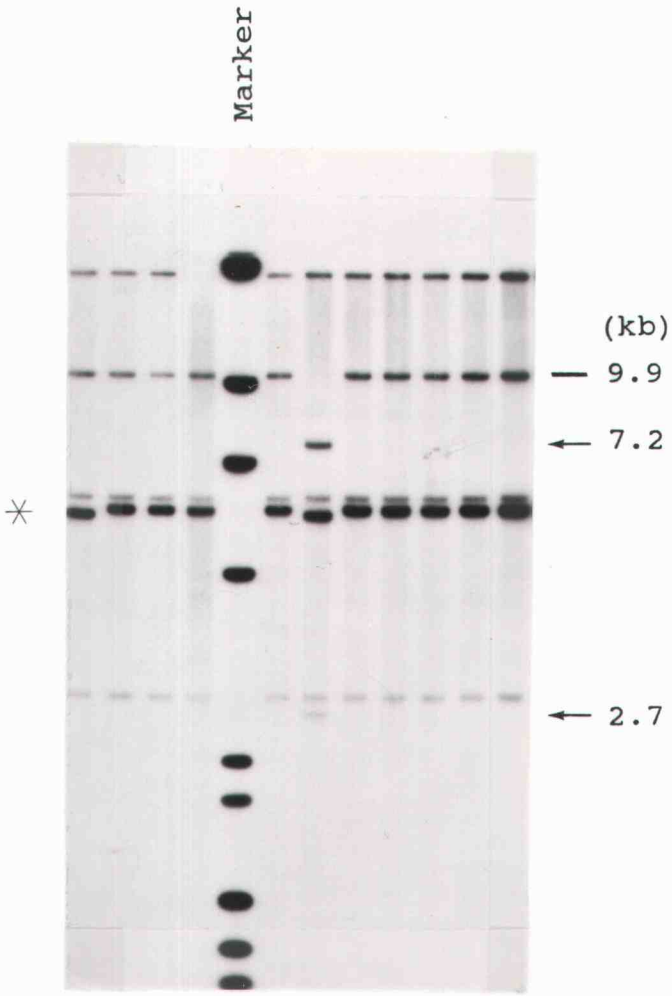


Figure II.2.

recognition sites (BamHI, BclI, BglI, BglII, DraI, EcoRI, EcoRV, HindIII, KpnI, PstI, SalI, SmaI, SstI, StuI, XbaI, XhoI, and XmnI) and 3 enzymes with four bp recognition sites (HinfI, HpaII, and MspI), and probed their blotted DNA with our cpDNA probe sets. Only those nine enzymes (BamHI, BclI, DraI, EcoRI, EcoRV, KpnI, XbaI, XhoI, and XmnI) that detected polymorphisms either within or between species in this primary survey were retained for the full survey of all 384 study trees.

Interpretation of RFLPs

Although we observed RFLPs that resulted from at least four causes--site mutation, length mutation, inversion, and heteroplasmy--we used only site mutations in our phylogenetic analyses. This is because homology among length mutations that give rise to similarly sized fragments is difficult to establish with confidence under the resolution provided by agarose gels. Inversion and heteroplasmy were too rare to be of use in phylogenetic analyses. Operationally, length mutations were defined as small (< 1 kb) changes in the length of individual fragments that could be observed with more than one restriction enzyme, and which were typically polymorphic within populations. Conversely, point mutations were observed with only a single enzyme, and gave rise to typically large, stoichiometric changes in sizes of several

fragments. Restriction site mutation, and the fragment phenotypes on which they are based, are given in Table II.2.

Quantitative Analyses

Phylogenetic trees were constructed from character state data on restriction site presence and absence by Wagner and Dollo parsimony with global branch swapping. Rare within-population polymorphisms (Table II.2) were excluded from parsimony analyses, but not from distance-matrix analyses (see below). Wagner parsimony attempts to build a dendrogram that minimizes the total number of character state changes in a tree without regard to whether they are gains or losses. The Dollo method assumes that characters can arise only once, but losses can occur multiple times, and thus attempts to minimize the total number of character losses. Dollo trees were rooted at the midpoint of the branch connecting the most divergent taxa. This method is particularly suitable for restriction site data because far more mutational events can result in loss than in gain of a specific site. It is also well suited to studies of closely related species because homoplasious site gains are expected to be very rare.

Wagner trees were rooted with a single individual of another member of subsection Oocarpae, P. oocarpa. It is thought to closely represent the common ancestor of the

California closed cone pines (reviewed by Millar 1986), a concept corroborated by recent molecular analyses (Strauss and Doerksen 1990). Confidence limits on phylogenetic topologies were examined via 100 bootstrap samples (Felsenstein 1985a). In bootstrapping, random samples of characters drawn with replacement are used to iteratively construct phylogenetic trees. The most frequently obtained tree is presented, and the frequency with which clades (monophyletic groups of species) appear are recorded. Clades appearing in more than 95% of the trees are judged to be significant at the 5% confidence level. Analyses were implemented with Phylip version 3.1 (Felsenstein 1988), while PAUP version 3.0 (Swofford 1990) was used to calculate homoplasy and consistency indices. The homoplasy index is the percent of total character state changes, which do not exist in a common ancestor, that occur independently on different taxa sharing a common ancestor. Whereas the consistency index is the percent of characters, that occur only once on a phylogenetic tree, relative to the number of changes invoked on a tree.

The frequency of restriction site mutations among and within the 20 populations studied were used to estimate nucleotide differentiation with the program HAPLO (Table II.3) (Lynch and Crease 1990). The nucleotide divergences were then used to construct phylogenetic trees via the neighbor-joining (Saitou and Nei 1987) and unweighted pair-

group (UPGMA) methods using the program RESTSITE (Nei and Millar 1990), and to construct Fitch-Margoliash trees using Phylip version 3.1 (Felsenstein 1988).

We used the Wagner parsimony tree as a basis to study variability in evolutionary rates using the concept of phylogenetically-based contrasts (Felsenstein 1985). The amount of genetic differentiation observed in clades below a reference node were contrasted either directly (Wagner tree) or via comparison to a reference outgroup identified from the phylogenetic topologies. The significance of differences in number of mutations invoked on the Wagner tree were tested via χ^2 analyses with one degree of freedom.

RESULTS

We observed that cpDNA restriction site polymorphism within populations was extremely infrequent, concordant with observations in other species (Chapter I). In three cases, however, the observed polymorphisms were phylogenetically informative. (1) The bishop pine population at Monterey had two cpDNA site mutations at frequencies of 16.7% (±10.8%) and 33.3% (±13.6%) that were identical to those fixed in the southern populations. (2) The bishop pine populations at San Vicente, Mexico and Santa Cruz island (P. muricata var. remorata) shared two rare polymorphic restriction site mutations at frequencies of 9.5% (±6.4%) and 4.8% (±4.6%), respectively, that were

Table II.3. Matrix of cpDNA nucleotide diversity (%) among and within populations and species based on 313 surveyed restriction sites. Interpopulation diversity is above diagonal, mean diversity within populations is on diagonal, and interspecies diversity is below diagonal (±standard error).

Table II.3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	MANO	MCAM	MMON	MGUA	MCED	KKLA	KOAK	KSAA	KSIE	BMEN	BTRI	BMAR	BSON	BMON	BSAB	BSAC	BSAL	BSAR	BSAV	BSCR
	Monterey					Knobcone				Bishop										
1.		0.0000	0.0001	-0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
2.			0.0000	0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
3.			0.0008 (±0.0008)	0.0000	0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
4.					0.0000	0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
5.						0.3910	0.3910	0.3910	0.3910	0.2659	0.2659	0.2366	0.2366	0.2202	0.1187	0.1187	0.1187	0.1187	0.1072	0.1131
6.							0.0000	0.0000	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
7.								0.0000	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
8.			0.3910 (±0.1154)					0.0000 (±0.0000)	0.0000	0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4383	0.4447
9.										0.6586	0.6586	0.6292	0.6292	0.6128	0.4503	0.4503	0.4503	0.4503	0.4388	0.4447
0.											0.0000	0.0290	0.0290	0.0321	0.3244	0.3244	0.3244	0.3244	0.3129	0.3188
1.												0.0290	0.0290	0.0321	0.3244	0.3244	0.3244	0.3244	0.3129	0.3188
2.													0.0000	0.0031	0.2952	0.2952	0.2952	0.2952	0.2897	0.2924
3.														0.0031	0.2952	0.2952	0.2952	0.2952	0.2897	0.2924
4.															0.2689	0.2689	0.2689	0.2689	0.2644	0.2666
5.			0.0909 (±0.0872)				0.4378 (±0.1601)							0.0036 (±0.0029)		0.0000	0.0000	0.0000	0.0003	0.0000
6.																	0.0000	0.0000	0.0003	0.0000
7.																		0.0000	0.0003	0.0000
8.																			0.0003	0.0000
9.																				-0.0003
0.																				

also found in knobcone and Monterey pines, and in P. oocarpa. And, (3) the two mainland populations (Año Nuevo and Monterey) of Monterey pine had a common rare polymorphism with a frequency of 3.3% (±3.3%) and 3.7% (±3.6%) (Fig. II.2). No polymorphisms for site mutations of any kind were observed in knobcone pine. Despite exclusion of these rare polymorphisms from parsimony analyses, results were very similar to that of the distance-matrix analyses which included them (discussed below).

Twenty phylogenetically informative site mutations were nearly fixed among populations of bishop pine and among species. Bishop pine could be subdivided into three distinct groups of populations with little polymorphism within the groups. The northern group consisted of the Trinidad and Mendocino populations; the intermediate group consisted of Marin, Sonoma, and Monterey populations; and the southern group consisted of San Luis Obispo, San Vicente, Santa Barbara, Santa Cruz-bishop, Santa Rosa, and Santa Cruz-remorata populations. The northern group differed by two mutations from the intermediate group, and by 11 mutations from the southern group; the intermediate and southern groups differed by nine mutations (Table II.2). Monterey and the southern group of bishop pine were most closely related, differing by four mutations, whereas these species were differentiated from knobcone pine by 12

and 14 mutations, respectively. Surprisingly, more mutations separated the southern group of bishop pine from its other conspecific populations, than separated the southern group and Monterey pine. Thus, intraspecific cpDNA genetic differentiation exceeded interspecific differentiation.

Wagner parsimony analysis gave rise to a single most parsimonious phylogenetic tree that contained only a single parallel restriction site loss (homoplasy index of 4.2%; consistency index of 92.3%; Swofford 1990) (Fig. II.3). Significant at the 5% level was the species complex itself, and the clade of populations in bishop pine that included the intermediate and northern groups. When the consensus branching patterns in 80% or more of the bootstrap samples were examined (Fig. II.4), knobcone pine was distinct from the two maritime species--separated from the Monterey and bishop pine clade by 13 character state changes. It also appeared to be the earliest to diverge of the species in the complex. The branching pattern of the southern group of populations of bishop pine were not resolved from Monterey pine--from which they differed by only four mutations. Dollo parsimony also gave rise to a single most parsimonious tree and, in contrast to the Wagner tree, it indicated that the clade leading to bishop and Monterey pine was highly statistically significant (Fig. II.5).

The Fitch-Margoliash and neighbor-joining trees gave

Figure II.3. The single most parsimonious tree derived from Wagner parsimony analysis. Numbers given along the branches are the number of restriction site gains or losses along each branch; numbers within circles at the nodes are percentages of 100 bootstrap replicates in which the species below this node occurred as a monophyletic group.

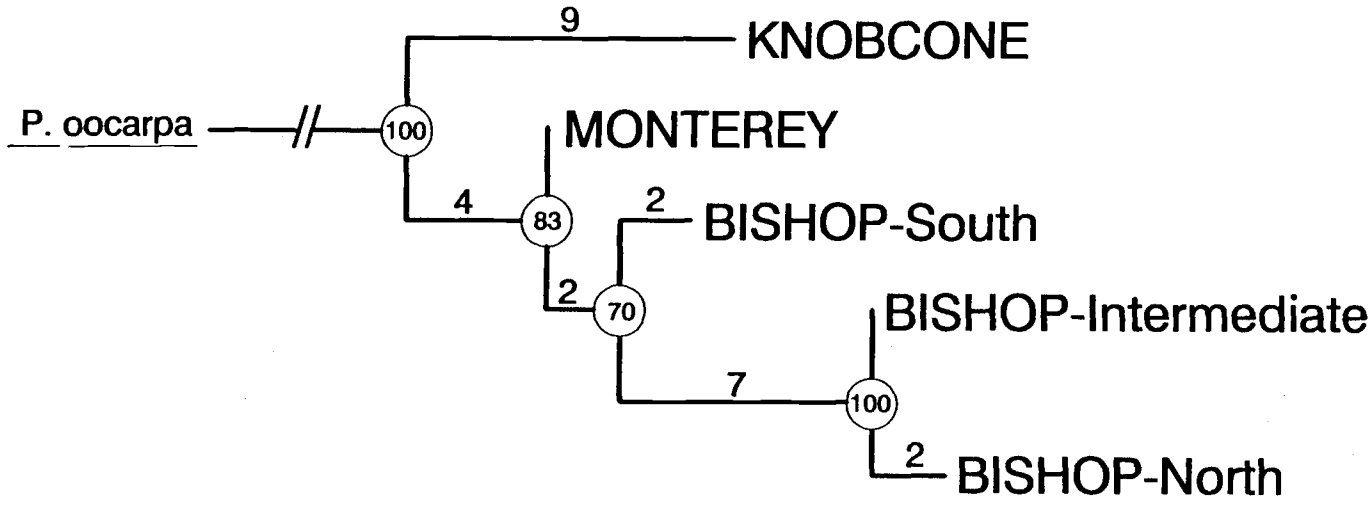

9
KNOBCONE
P. oocarpa
100
MONTEREY
4
83
2
BISHOP-South
2
70
BISHOP-Intermediate
7
100
2
BISHOP-North

Figure II.3.

Figure II.4. Bootstrap 80 % majority-rule consensus tree generated from 100 bootstrap samples. Only the branching pattern, not branch length, is meaningful.

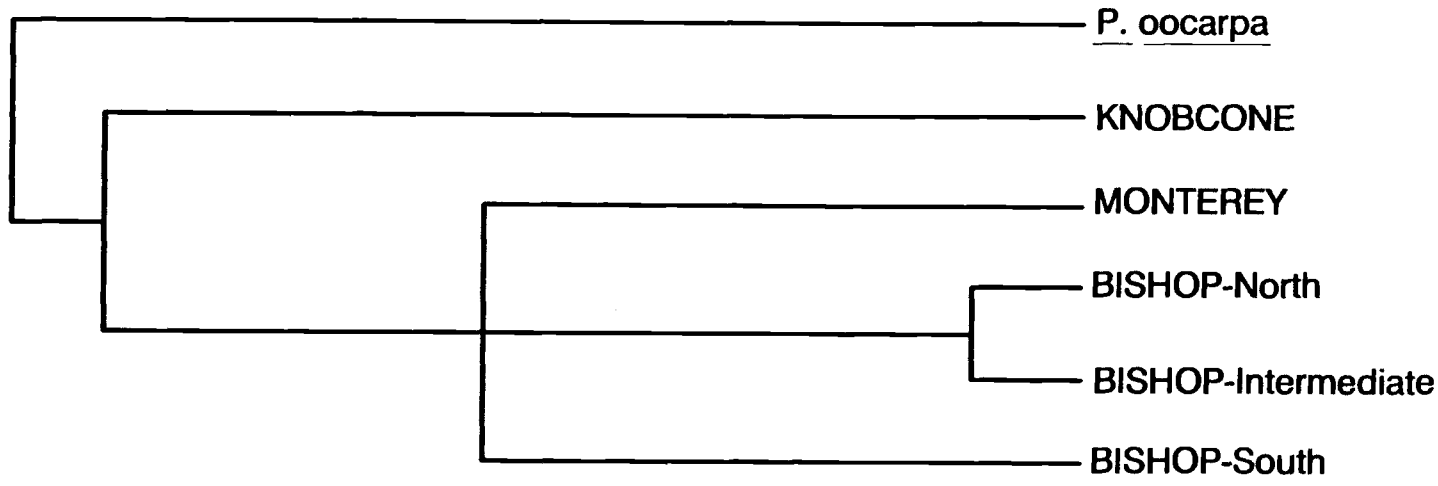

P. oocarpa
KNOBCONE
MONTEREY
BISHOP-North
BISHOP-Intermediate
BISHOP-South

Figure II.4.

Figure II.5. Phylogenetic tree derived from Dollo parsimony analysis. See Fig. II.3 for explanation.

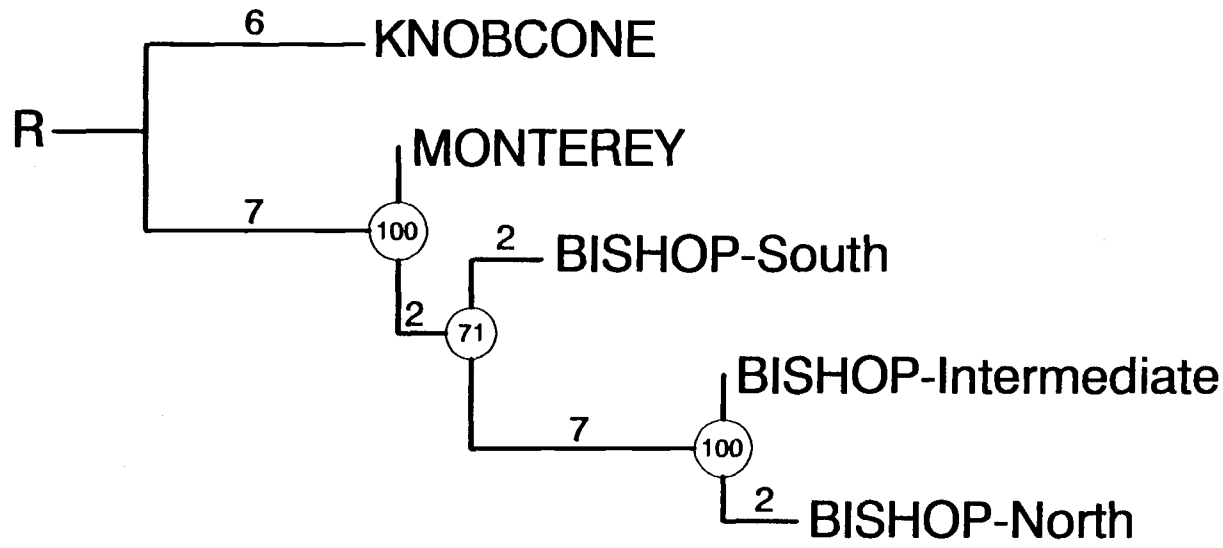

6
KNOBCONE
R
MONTEREY
7
100
2
BISHOP-South
2
71
BISHOP-Intermediate
7
100
2
BISHOP-North

Figure II.5.

results that agreed well with the results of the parsimony analyses and agreed well with one another (Fig. II.6a, 6b). The three species occupied distinct clades, with knobcone pine splitting early from the clade leading to bishop and Monterey pines. The southern group of bishop pine was again more closely related to Monterey pine than to other populations of bishop pine, and was ancestral to the intermediate and northern groups in bishop pine.

The trees suggest a great deal of variability in rate of cpDNA evolution. Based on the matrix of nucleotide diversities among taxa when compared to putative outgroup reference taxa, relative rates varied from 1.14 for the intermediate vs. northern populations of bishop pine compared to the southern population, to 2.25-fold for the southern and northern populations of bishop pine compared to Monterey pine (Table II.4). Based on number of mutations observed on the Wagner tree, the southern populations of bishop pine and knobcone pine had nearly equal numbers of mutations, whereas Monterey pine had no mutations compared 11 for the northern populations of bishop pine--a result that was highly statistically significant.

The UPGMA tree gave results that were at odds with the other phylogenetic trees. It suggested that Monterey pine split from within one of two distinct clades of bishop pine. It also indicated that the northern and intermediate

Figure II.6. Phylogenetic trees derived from 3 distance-matrix methods. (a) Results of neighbor-joining method. (b) Results of Fitch-Margoliash method. (c) Results of UPGMA method. Abbreviations: K, knobcone pine; M, Monterey pine; B, bishop pine.

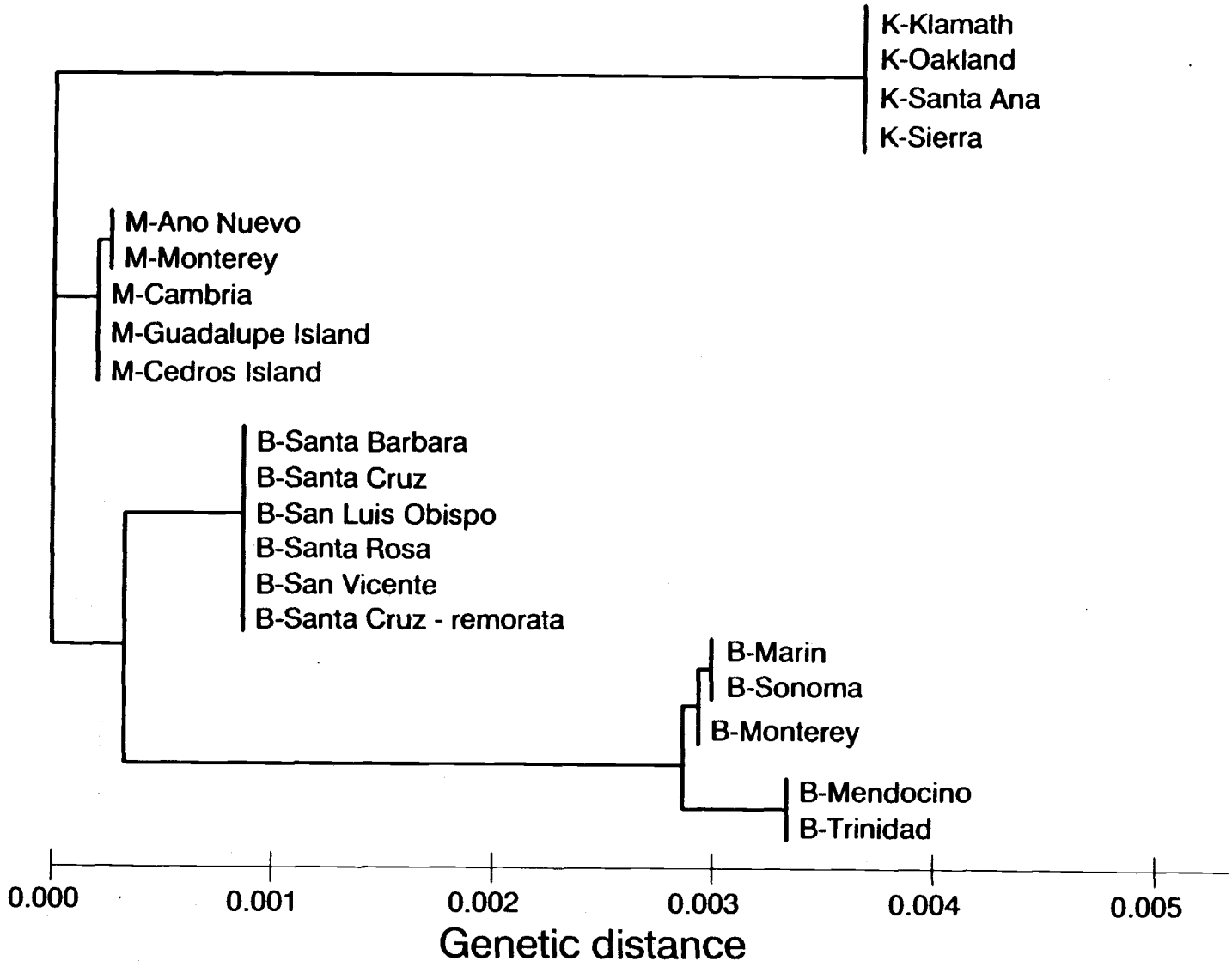

K-Klamath
K-Oakland
K-Santa Ana
K-Sierra
M-Ano Nuevo
M-Monterey
M-Cambria
M-Guadalupe Island
M-Cedros Island
B-Santa Barbara
B-Santa Cruz
B-San Luis Obispo
B-Santa Rosa
B-San Vicente
B-Santa Cruz - remorata
B-Marin
B-Sonoma
B-Monterey
B-Mendocino
B-Trinidad
0.000
0.001
0.002
0.003
0.004
0.005
Genetic distance

Figure II.6.(a)

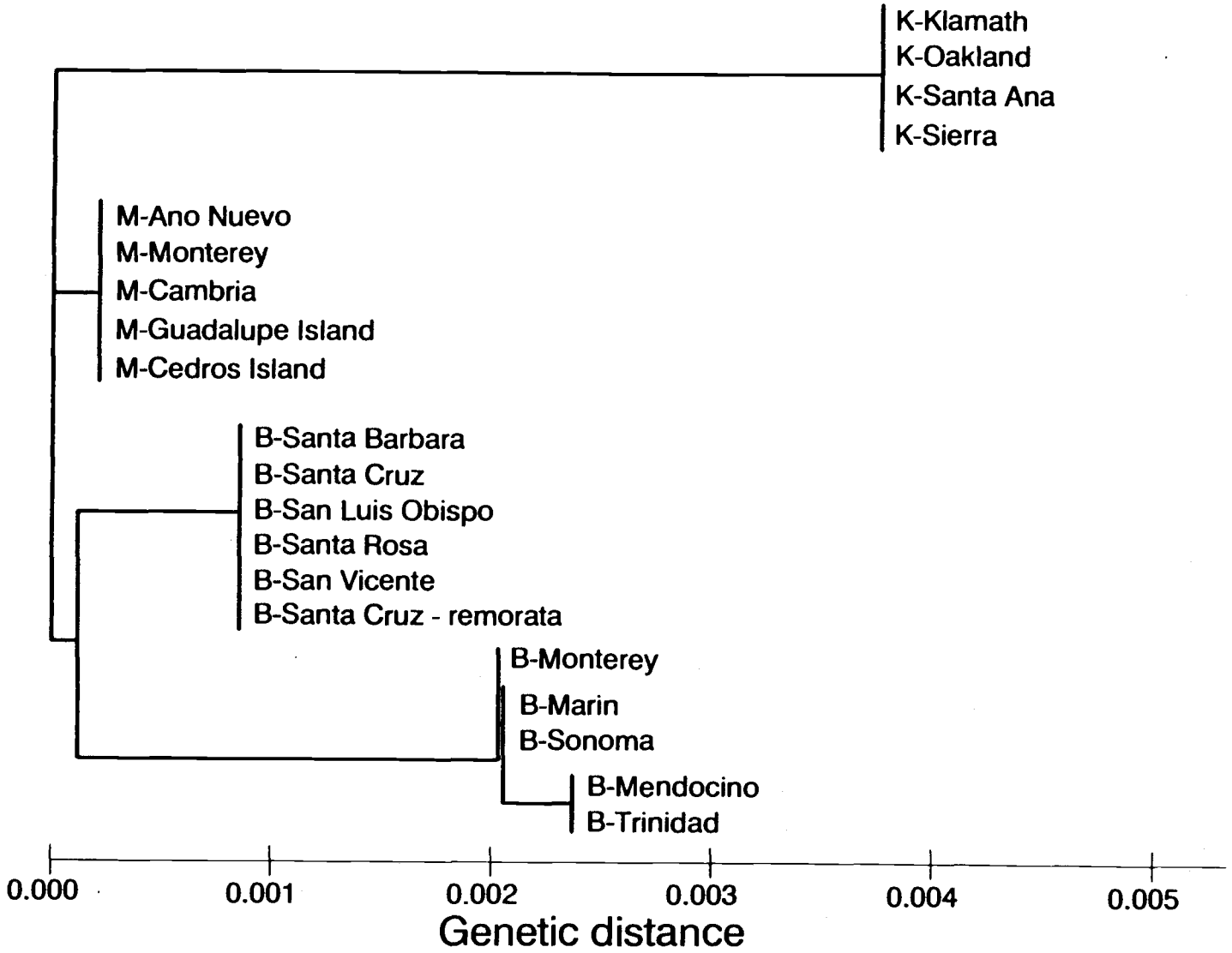

K-Klamath
K-Oakland
K-Santa Ana
K-Sierra
M-Ano Nuevo
M-Monterey
M-Cambria
M-Guadalupe Island
M-Cedros Island
B-Santa Barbara
B-Santa Cruz
B-San Luis Obispo
B-Santa Rosa
B-San Vicente
B-Santa Cruz - remorata
B-Monterey
B-Marin
B-Sonoma
B-Mendocino
B-Trinidad
0.000
0.001
0.002
0.003
0.004
0.005
Genetic distance

Figure II.6.(b)

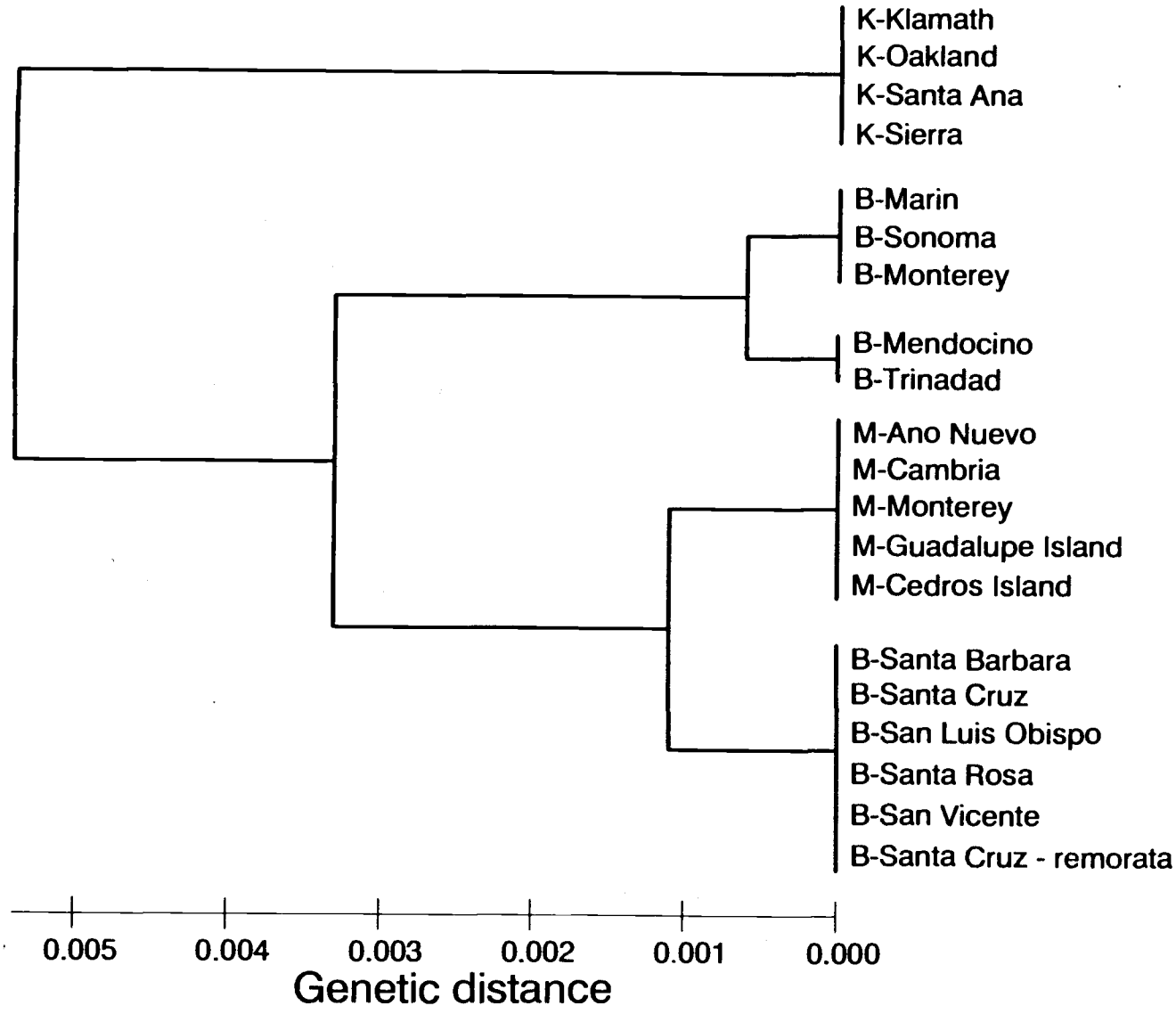

K-Klamath
K-Oakland
K-Santa Ana
K-Sierra
B-Marin
B-Sonoma
B-Monterey
B-Mendocino
B-Trinadad
M-Ano Nuevo
M-Cambria
M-Monterey
M-Guadalupe Island
M-Cedros Island
B-Santa Barbara
B-Santa Cruz
B-San Luis Obispo
B-Santa Rosa
B-San Vicente
B-Santa Cruz - remorata
0.005
0.004
0.003
0.002
0.001
0.000
Genetic distance

Figure II.6.(c)

Table II.4. Relative rates of evolution.

a. Based on nucleotide diversity matrix.

Contrast	Reference taxon	Distance to reference taxon	Relative rate
Monterey:Bishop-N	Knobcone	0.0039:0.0066	1.69
Monterey:Bishop-I	Knobcone	0.0039:0.0062	1.59
Monterey:Bishop-S	Knobcone	0.0039:0.0045	1.15
Bishop-S:Bishop-N	Monterey	0.0012:0.0027	2.25
Bishop-S:Bishop-I	Monterey	0.0012:0.0023	1.92
Bishop-I:Bishop-N	Bishop-S	0.0028:0.0032	1.14

Table II.4. (continued)

b. Based on Wagner parsimony tree.

Contrast	No. mutations from common progenitor	Relative rate[1]	χ^2
Knobcone:Bishop-N	9:15	1.67	1.50
Knobcone:Bishop-I	9:13	1.44	0.73
Knobcone:Bishop-S	9:8	1.13	0.06
Knobcone:Monterey	9:4	2.25	1.92
Monterey:Bishop-N	0:11	--	11.00**
Monterey:Bishop-I	0:9	--	NT
Monterey:Bishop-S	0:4	--	NT
Bishop-S:Bishop-N	2:9	4.50	NT
Bishop-S:Bishop-I	2:7	3.50	NT
Bishop-I:Bishop-N	0:2	--	NT

[1] ratio of larger to smaller number of mutations

** significant at 1% level

NT, not tested

population groups in bishop pine diverged earliest in the Monterey-bishop clade, and that the southern bishop andMonterey pine clades diverged later.

DISCUSSION

Despite very low levels of overall cpDNA variability, we found that the polymorphisms present were informative both among and within species. This echoes the conclusion of Soltis et al. (1988) that cpDNA polymorphisms are useful at low taxonomic levels. Two key advantages of the use of rare fixed cpDNA polymorphisms are that they show little homoplasy, and extensive within-taxon samples are not necessary. Our homoplasy index from Wagner parsimony was 4.2%, the result of a single parallel restriction site loss among 24 polymorphic sites in the complex. Similar homoplasy indexes have been reported for intraspecific cpDNA studies in other groups (range of 0 to 4.9%: Doebley et al. 1987). Because of the near lack of polymorphism within two of the three species studied, and within most populations, we would have obtained nearly identical results had we sampled a total of only 5 trees: a single individual each from Monterey and knobcone pines, and three individuals that span the range in bishop pine. These results support the common practice of using very small samples for cpDNA-based phylogenetic studies. A safe guideline would be to attempt to sample a single individual

from each of a very few distinctive and isolated populations in each species. The results also show that chloroplast DNA restriction site mutations would be highly effective genetic markers for identifying seed sources in bishop pine as a consequence its geographic group-specific mutations.

Our results indicate that cpDNA provides unique insights into evolutionary relationships and genetic diversity that are not necessarily provided by other traits such as allozymes. Distinctive taxa based on cpDNA may not be differentiated based on allozyme frequencies; and conversely, distinctive taxa based on allozyme frequencies may not be differentiated for cpDNA. Examples include (1) the distinct northern and intermediate groups of populations in bishop pine identified in cpDNA analyses; these units were not differentiated based on allozyme frequencies (Millar et al. 1988). (2) The Cedros island population of Monterey pine, which is highly divergent from the other island and mainland populations based on allozyme frequencies (Moran et al. 1987), but was identical to the other populations in our cpDNA sample. (3) Knobcone pine was distinct and highly diverged from the bishop-Monterey pine clade based on cpDNA, whereas the three species were nearly equidistant based on allozyme frequencies (Millar et al. 1988). And, (4) the closer relationship we found among the southern populations of bishop pine relative to

Monterey pine, than with its conspecific intermediate and northern populations. Although allozymes also identified the southern bishop group as most ancestral in the species, they were clearly more similar to other conspecific populations than to either of the other species (Millar et al. 1988). Rates of evolution for nuclear DNA, and effects of speciation, are clearly uncoupled from that for chloroplast DNA. Organelle DNA similarity has also been observed to transgress species boundaries in many studies of mtDNA and cpDNA in plants and animials (reviewed in Rieseberg et al. 1988).

The division of bishop pine into three geographic groups agrees well with most other traits studied, including morphology (Fielding 1961, Shelbourne et al. 1982), monoterpene composition (Forde and Blight 1964, Bannister and McDonald 1983), crossability (Critchfield 1967, Millar and Critchfield 1988), phenology, and frequencies at specific allozyme loci (Millar 1983, Millar et al. 1988). The northern group of two populations represents the "blue" foliage race of bishop pine, which differs in several traits from the "green" bishop populations (Forde and Blight 1964, Millar 1989, Shelbourne et al. 1982). Shelbourne et al. (1982) reported that the "blue" foliage race were much more frost-tolerant than the "green" foliage race. Its distinctness was also reflected in its cpDNA; we detected two mutations that appeared to be

fixed (95% confidence interval was greater than 0.76) in these populations, but that were very rare or absent elsewhere (Chapter I).

The phylogenetic topologies and relative rate comparisons suggested substantial variance in cpDNA evolutionary rates. For relative rate comparisons we incorporated Felsenstein's (1985b) suggestion of using a phylogenetic tree to ensure that data are independent of phylogenetic constraint. Unfortunately, we were forced to use trees constructed with the same data under test--thereby employing a non-independent framework. We do not, however, believe this to be an important problem. The trees constructed were either congruous, or not in conflict, with other data, and the trees contained a very low number of homoplasious mutations. Nonetheless, because of this non-independence the rate tests and comparisons must be treated as tentative.

Variance in rate of evolution appears to have distorted the results of the UPGMA tree, which assumes homogeneous rates of evolution, compared to the other trees. The Monterey and southern bishop groups clustered first, likely on account of their slow rate of evolution and thus low level of divergence from one another. This made it appear that they were derived from a progenitor that resembled the more rapidly evolving intermediate and northern bishop pine populations, contrary to

paleobotanical evidence and the southerly location of the putative ancestral Mexican relatives of the complex (reviewed in Millar et al. 1988).

This effect was also seen with UPGMA when applied to allozyme data (Millar et al. 1988). Rapidly evolving northern populations in bishop pine, and northern and coastal populations in knobcone pine, clustered later than the southern populations, making them appear ancestral. However, contrary to these results were the genetic distance-matrix, a distance-Wagner analysis (which does not assume homogeneous rates), affinities to the outgroup P. oocarpa, and affinities among populations--which indicated a south to north migration. Thus, the assumption of rate homogeneity appears to be seriously violated in the both nuclear DNA and cpDNA in the California closed cone pines, giving misleading results when methods of analysis such as UPGMA that assume rate constancy are used.

The genetic differentiation among species and populations in bishop pine can be used to make a rough estimate of times of divergence. Li and Graur (1991) showed that the divergence time for neutral mutations (T) is a simple function of mutation rate (r) and number of substitutions per site (K):

$$T = K/2r$$

The mutation rate for cpDNA was estimated to be 1-3 X 10^{-9} substitutions per site per year (Wolfe et al. 1987).

Assuming the lower bound of these estimates is correct for conifers based on the observed relatively slow rate of rbcL evolution (Strauss, unpubl. data), this suggests that the knobcone and Monterey-bishop lineages split 1.9 million years ago, and Monterey and bishop (southern group) 0.6 million years ago. We assume that the great divergence of the intermediate and northern bishop groups compared to southern bishop results from deviations from neutrality such as from selection (discussed in Chapter 1). These estimates are much more recent than might be expected given Axelrod's suggestion (reviewed in Millar et al. 1988) that the common ancestor of the complex arose in the Oligocene (26-38 million years ago). Either the ancestor of the complex existed for many millions of years before speciation occurred, or the rate of cpDNA evolution is much slower than expected based on mutation rate.

Acknowledgements

We thank Dr. G. Furnier who sampled the Santa Ana population of knobcone pine for us, Dr. M. Lynch who provided the computer program of HAPLO, Dr. J. Miller who provided the computer program of RESTSITE, and G. T. Howe who helped in field collections of samples. We are also grateful to Dr. W.J. Libby, Dr. F.T. Ledig, and U.S. Forest Service, Pacific Southwest Experiment Station in Berkeley, California, for the use of plantations. This study was

supported by National Institutes of Health Grant GM 36367.

LITERATURE CITED

Axelrod, D. I. 1983. New Pleistocene conifer records, coastal California. Univ. Calif. Publ. Geol. Sci. 127: 1-108.

Bannister, M. H.. and J. R. McDonald. 1983. Turpentine composition of the pines of Guadalupe and Cedros Islands Baja California. N. Z. Jour. Bot. 21: 373-377.

Critchfield, W. B. 1967. Crossability and relationships of the California closed cone pines. Silvae Genet. 16: 89-97.

Critchfield, W. B. and E. L. Little. 1966. Geographic distribution of pines of the world. USDA Misc. Publ. 991. pp. 97.

Doebley, J., W. Renfroe, and A. Blanton. 1987 Restriction site variation in the Zea chloroplast genome. Genetics 117: 139-147.

Feinberg, A. P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Felsenstein, J. 1985a. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.

Felsenstein, J. 1985b. Phylogenies and the comparative method. Am. Nat. 125: 1-15.

Felsenstein, J. 1988. PHYLIP: Phylogeny Inference Package. Version 3.1 Univ. of Washington, Seattle, WA.

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

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

Howell, J. T. 1941. The closed-cone pines of insular California. Leaflets of Western Botany 3: 1-24.

Kirk, J. T. O. and R. A. E. Tilney-Bassett. 1978. The plastids. Their chemistry, structure, growth, and inheritance. Second edition. Elsevier, Amsterdam.

Li, W.-H. and D. Graur. 1991. Fundamentals of Molecular evolution. Sinauer Associates, Inc. Sunderland, Massachusetts.

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: 29-40.

Linhart, Y. B., B. Burr, and M. T. Conkle. 1967. The closed-cone pines of the northern Channel Islands. In: R.B. Philbrick (ed.), Proceedings of the symposium on the biology of the California islands pp. 151-177. Santa Barbara Bot. Garden, CA.

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

Mason, J. L. 1930. The Santa Cruz Island pine. Madroño 2: 10.

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

Millar, C. I. 1986. The Californian 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: 39-53.

Millar, C. I., S. H. Strauss, M. T. Conkle, and R. Westfall. 1988. Allozyme differentiation and biosystematics of the Californian 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 the conservation of the five natural populations of Pinus radiata. Can. J. For. Res. 18: 506-514.

Nei, M. and J. C. Miller. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125: 873-879.

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

Palmer, J. D. and D. Zamir. 1982. Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. Proc. Natl. Acad. Sci. USA 79: 5006-5010.

Palmer, J. D., C. R. Shields, D. B. Cohen, and T. J. Orton. 1983. Chloroplast DNA evolution and the origin of amphidiploid Brassica species. Theor. Appl. Genet. 65: 181-189.

Palmer, J. D., R. A. Jorgensen, and W. F. Thompson. 1985. Chloroplast DNA variation and evolution in Pisum: Patterns of change and phylogenetic analysis. Genetics. 109: 195-213.

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

Reed, K. C. and D. A. Mann. 1985. Rapid DNA transfer from agarose gels to nylon membranes. Nucleic Acids Res. 13: 7207-7221.

Rieseberg, L. H., D. E. Soltis, and J. D. Palmer. 1988. A molecular reexamination of introgression between Helianthus annuus and H. bolanderi (Compositae). Evolution 42: 227-238.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Sears, B. B. 1980. Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. Plasmid 4: 233-255.

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. Jour.For. 27(1): 50-66.

Soltis, D. E., P. S. Soltis, T. A. Ranker, and B. D. Ness. 1988. Chloroplast DNA variation in a wild plant, Tolmiea menziesii. Genetics 121: 819-826.

Strauss, S. H., J. Bousquet, V. D. Hipkins, and Y.-P. Hong, 1991. Contributions of biochemical and molecular genetic markers to biosystematic studies of forest trees. New Forests (in press).

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

Swofford, D. L. 1990. PAUP: Phylogenetic analysis using parsimony, user's manual, version 3.0. Illinois Natural History Survey, Champaign, IL.

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.

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 88: 9054-9058.

CHAPTER III

SEQUENCE OF THE INTERGENIC REGION BETWEEN rbcL AND atpB GENES OF CHLOROPLAST DNA FROM THE CALIFORNIA CLOSED CONE PINES

ABSTRACT

667 bp of the chloroplast intergenic sequences between rbcL and atpB genes were sequenced from five individuals of the California closed cone pines by direct sequencing of polymerase chain reaction (PCR) amplified DNA. No variation was observed among sequences of the five individuals, which represented three species and three geographic regions of bishop pine. The upstream sequences (251 bp) of the rbcL gene from Douglas-fir were aligned with part of our intergenic sequences, and showed about 88.4% similarity.

INTRODUCTION

The chloroplast intergenic sequences between rbcL and atpB genes were completely sequenced from one individual of each Monterey, knobcone, and three geographic groups of bishop pines. This cpDNA region was chosen because we expected that it would evolve more rapidly than average cpDNA, which includes coding regions, and because of the ready availability of primers from both flanking genes. We

hoped to identify small length or point mutations that would be useful phylogenetic characters.

MATERIALS AND METHODS

Genomic DNA samples, extracted by a modified CTAB method (Wagner et al. 1987), were randomly selected from representatives of each of the five distinct cpDNA groups in the complex (species and bishop pine geographic groups) (Chapter 1). The intergenic region was amplified using rbcL-358r (5'-CCCACAATGGAAGTGAACAAGTTAGTAAC) and atpB-190r [5'-CCTAATAATTGC(&T)TGTACC(&T)TCACA] primers (Fig. III.1). To obtain single stranded DNA template suitable for sequencing with several internal primers, one primer of each pair was phosphorylated while keeping the other one unphosphorylated before amplification, and the amplified DNA treated with λ-exonuclease (which digests only the phosphorylated strand at the 5'-end) (Higuchi and Ochman 1989).

PCR conditions were as follows: total volume of reaction 100 μl, 100 ng of template genomic DNA, 50 pmoles of each primer, 50 mM $MgCl_2$, 0.01 % gelatin, 0.1 % Triton X-100, and 2.5 units Taq polymerase (Promega). Amplification was conducted in a DNA thermal cycler (Perkin Elmer Cetus) with 35 cycles, each consisting of a denaturing step of 1 min. at 94°C, followed by an annealing step of 1 min. at 50°C, and an extension step of 2.5 min. at 72°C. The last

Figure III.1. Location of the primers used for amplification and sequencing. * and ** denote the primers which were used for cpDNA amplification.

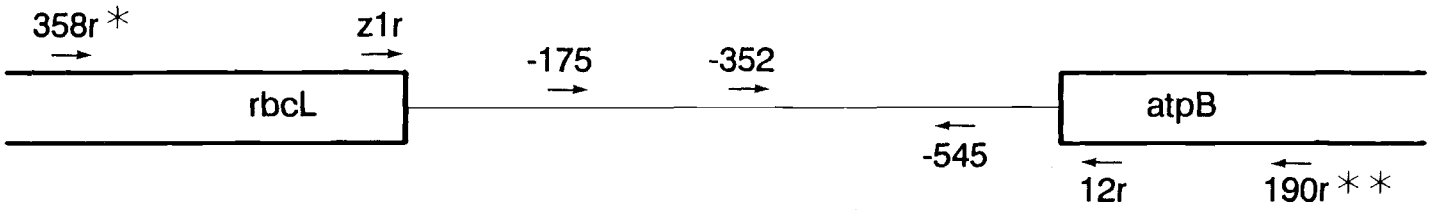


Figure III.1.

20 extension steps were extended by 5 sec. for every step to ensure full extension. The last cycle of amplification was followed by a period of 10 min. at 72°C to allow extension reactions to proceed to completion.

The alternative single strand DNAs were obtained by modified λ-exonuclease treatment (Higuchi and Ochman 1989). The phosphorylated DNA strands at 5'-end were digested at 37°C for 30 min. using 4 units of λ-exonuclease (BRL) in a 100 μl reaction volume. After digestion of the phosphorylated DNA strands, the unphosphorylated single strands were separated from the remaining undigested double strands by gel electrophoresis (1.5% agarose) and purified in Centricon-30 microconcentrators. DNA was then extracted with phenol-chloroform and chloroform, and ethanol precipitated before being resuspended in 10 μl TE (10:0.1). Enough DNA was obtained for four different labelling reactions from a single amplification.

Four additional internal primers were sequentially designed based on determined sequences with the aid of a primer side identification program (Primer: Lowe et al. 1990) (Fig. III.1). For sequencing the upstream region of the rbcL gene, internal primers annealing to positions -1 (z1r: 5'-ACTTGCTTTAGTTTCTGTTTGTGGTGACAT), -175 (5'-TATGACGCAACCCAATCTTTGC), and -352 (5'-TCAATCAGAGAAAGGTCATG) relative to the rbcL translation initiation site were used. The opposite strand was primed by annealing two

internal primers to positions +12 (5'-CGGAAACCCCAAGAACAAGAG G) and -545 (5'-AAGAAATGGAAGTTACCACC) relative to the atpB translation initiation site. The atpB-190r primer, used for amplification, was also used for sequencing to determine the sequence for designing the atpB-12r primer. Sequencing was conducted using 6% polyacrylamide-7M urea gels according to the manufacturer's recommendations (Bio-Rad). X-ray film were exposed from 24 hours to several days. The developed autorads were read with aid of a digitizer (Graf/Bar MarkII: Science Accessories Corporation) and sequences compared using the Intelligenetics analysis package (PC/gene: Bairoch 1989).

RESULTS AND DISCUSSION

No variation was observed among intergenic sequences of the five individuals (Fig. III.2). This extreme conservation may be due to the presence of regulatory sequences of two genes, which extend about 250-300 bp from each coding sequence (Hipkins et al. 1990). The close relationship of the sampled species, combined with the low mutation rate for cpDNA (1-3 X 10^{-9} substitution/site/year: Wolfe et al. 1987) led us to expect few mutations. However, given that the species complex is believed to be Oligocene in age (26 to 38 million years old), we expected at least a few mutations. Were the maximum divergence 2 million years and the yearly mutation rate 10^{-9}, 1.3

Figure III.2. Complete nucleotide sequence of the intergenic region between rbcL and atpB genes in the California closed cone pines. N denotes ambiguous bases which could not be read with confidence.

SEQUENCE 667 BP:

226 (33.9%) A; 110 (16.5%) C; 113 (16.9%) G; 211 (31.6%) T;

7 (1.1%) OTHER;

3'- rbcL coding sequences <---

```
  1 - TTCAGGGAGG AAATAATTAT TAATAAAGAG CGTTCCTGCT CCAGACGAGC

 51 - TGTACCTGGC TTGCATTTGT TAGTTAAAAC TTTCTTTATA GGATGTTTTT

101 - CAGCAGGCAA TAACCCCACG ATCTATGTGA GAGTAACATA NNCAAGAAAT

151 - ACATACTGCG TTGGGTTAGA AACGAGAACT TCAAGAAGGA GGTACCTAAC

201 - TGGGCTCTTG GAAAGTCACC AATTTGATCA AGTACTTACT TTAGTTCCTT

251 - GGCTTAACTG GAAACTGGTA NTACCACGCT TAACAGGCTT TTATGTATAT

301 - GTCTACACGC ATANTTCTCT TCTCTATTAG TAGTTAGTCT CTTTCCAGTA

351 - CTTATAGTTC AAAGTATAAA GTGTCTACTA GACCTGTATT AAAACTTCAT

401 - AAAAGCCAAA ATCAATACCT CTTTAACCAA GATCAATATC TATTTAGCTT

451 - CTGAAGGAGT ACTAGGAATA NGTATAGGTA GATGGATGAA GTATAAGGTT

501 - TATACCTACT TTAAATTTGA AAAGTTATTT CAACCTAAAA TGGTTCCACC

551 - ATTGAAGGTA AAGAATAAAT GACCAGACTA GCTGGGTTTT ACCTTGTAAA

601 - AANNACTAAC TACTATAACC GTTTTAGTAT AAATATATAA TAAGTACTTT

651 - TATTAAAGGT AAAAACA -5'
```

Figure III.2.

substitutions would be expected; at 20 million years, 13 substitutions.

Possible -35 (TTGCGT) and -10 (TACAAT) regions of prokaryotic-like consensus promoter sequences for the rbcL gene in the closed cone pines are located at positions 134 to 139 and 157 to 162, respectively (underlined in Fig. III.3). The -35 sequence in the closed cone pines differs in one nucleotide position from that identified in Douglas-fir (TTGCGT instead of TTGCAT) (Fig. III.3), and differs in one nucleotides from angiosperm sequences (TTGCGT instead of TTGCGC) (Hipkins et al. 1990). The putative ribosome binding site (GGAGG) is identified at positions six to ten (underlined in Fig. III.3).

We compared the upstream sequence (251 bp) from the rbcL gene to those determined from Douglas-fir (Fig. III.3). Only 224 of the 251 bp (88.4%) could be aligned with Douglas-fir, a result of several length mutations in the region. We also compared the upstream sequence (91-165 bp) from rbcL gene to those determined from five pine species (Strauss and Doerksen, unpubl. data), in which the identity of compared sequences ranged from 87.3 to 91.2% (mean of 89.3%). The presence of site and length variants in these comparisons suggests that the intergenic region may be informative for studying phylogeny among more highly divergent congeneric species.

Figure III.3. Sequence comparison between the California closed cone pines (CCCP) and Douglas-fir (D-fir) from upstream (5') region of the rbcL gene. Possible prokaryotic-like -35 and -10 promoter regions, and the ribosome binding site, are underlined.

rbcL coding sequences <---

```
CCCP   - TTCAGGGAGGAAATAATTATTAATAAAGAGCGTTCCTGCTCCAGACGAGC -50
         ::::::::::::::::::::::::::::::::::::::::::::::::::
D-fir  - TTCAGGGAGGAAATAATTATTAATAAAGAGCGTTCCTGCTCCAGACGAGC -50

CCCP   - TGTACCTGGCTTGCATTTGTTAGTTAAAACTTTCTTTATAGGATGTTTTT -100
         ::::::: ::::::::::::::: ::::::: ::::::::::::::::::
D-fir  - TGTACCTAGCTTGCATTTGTTAGCTAAAACTGTCTTTATAGGATGTTTTT -100

CCCP   - -CAGCAGGCAATAACCCC---------ACGATCTATGTGAGAGTAACATA -140
          ::::::::::::::::          : ::::::::::::::::::::
D-fir  - TCAGCAGGCAATAACCTGTTATTCTAAATGATCTATGTGAGAGTAACATA -150

CCCP   - NNCAAGAAATACATACTGCGTTGGGTTAGAAACGAGAACTTCAAGAAGGA -190
             : :::::::::: :::::::::::: ::::::::::::::::
D-fir  - TTACAAAAATACATACTACGTTGGGTTAGAA-CGAGAACTTCAAGAAG-- -197

CCCP   - GGTACCTAACTGGGCTCTTGGAAAGTCACCAATTTGATCAAGTACTTACT -240
              :::::::::: ::::::::  :::::::::::: :: :::::
D-fir  - ------TAACTGGGCTCGTGGAAAGTCGACAATTTGATCAATTAATTACT -241

CCCP   - TTAGTTCCTTG -251
         : : ::::::
D-fir  - TAAATTCCTT  -251
```

Figure III.3.

LITERATURE CITED

Bairoch, A. 1989. PC/gene user and reference manuals (release 6.01). IntelliGenetics, Inc. Geneva, Switzerland.

Higuchi, R.G. and H. Ochman. 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. 17: 5865.

Hipkins, D. V., C.-H. Tsai, and S. H. Strauss. 1990. Sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from a gymnosperm, Douglas fir. Plant Mol. Biol. 15: 505-507.

Lowe, T., J. Sharefkin, S. Q. Yang, and C. W. Dieffenbach. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reactions. Nucleic Acids Res. 18: 1757-1761.

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.

Winship, P. R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res. 17: 1266.

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 88: 9054-9058.

SUMMARY OF KEY RESULTS

A Total of 313 restriction sites (representing 1.5% of the chloroplast genome) were assayed from 384 trees drawn from 20 populations in the California closed cone pines. Twenty-four of 313 assayed restriction sites were polymorphic among populations or species. Twenty of the 24 polymorphic restriction sites showed no polymorphism within populations.

The major results from this study are:

1. Virtually no chloroplast DNA restriction site mutations were found among or within populations in Monterey and knobcone pines; nucleotide diversity was estimated as 0.0008% and zero, respectively.
2. Several restriction site mutations appeared to be fixed among populations of Bishop pine, a result of three distinct geographic groups of populations.
3. The majority of cpDNA diversity resided among species (N_{ST} = 89%).
4. Monterey and knobcone pines showed almost no population subdivision (N_{ST} = -2.24% and undefined, respectively), but bishop pine showed almost complete population subdivision (N_{ST} = 98%).
5. The strong cpDNA subdivision among species and populations of bishop pine contrast with strikingly

that of nuclear DNA, where a recent allozyme survey of the species complex found only 36% of the gene diversity among populations and species.

6. The topologies generated from phylogenetic analyses indicated that the California closed cone pines originated monophyletically from a P. oocarpa-like pool and evolved at different rates.
7. Knobcone pine diverged earliest of the California closed cone pines.
8. The geographic distribution of diversity in Bishop pine, and the topology generated from phylogenetic analyses, suggests that bishop pine can be treated as a complex of two to three sibling species, rather than as a single species.
9. 667 bp of chloroplast intergenic sequences between rbcL and atpB genes were determined, but no variation was observed among sequences from five individuals that represented all of the major genetic divisions in the species complex.
10. One individual from the Santa Cruz Island population of bishop pine was heteroplasmic, containing cpDNA genomes from both Monterey and bishop pine.
11. An inversion of about 50 kb was observed in two individuals in the southernmost population of bishop pine from San Vicente, Mexico. Its borders mapped very near to those of a large inversion that differentiates

Monterey pine and Douglas-fir.

12. We observed at least three regions of the genome subject to length mutations. One of them was studied in detail and showed three distinct phenotypes (5.62, 5.53, and 5.3 kb), and patterns of diversity uncorrelated with that observed for site mutations.

CONCLUSIONS

Our results indicate that chloroplast DNA markers provides unique insights into genetic diversity and evolutionary relationships that are not necessarily provided by products of nuclear DNA such as allozymes and morphological traits.

The major conclusions from this study are:

1. The unusual distribution of genetic diversity among species, and among populations of bishop pine, suggests that the relative strength of the evolutionary forces shaping genetic diversity in chloroplast genomes differs substantially from those affecting nuclear genomes. The extremely low cpDNA nucleotide diversity within populations and species, and the high diversity among species and populations of bishop pine, contrasts markedly with that typically seen in long-lived woody species in allozyme surveys--where much higher levels of genetic diversity typically reside within rather than among populations.
2. A combination of several evolutionary forces--genetic drift, gene flow, natural selection, and mutation--appear to play a role in causing different organizations of genetic diversity in nuclear and chloroplast genomes. We suspect, however, that two factors are most important: periodic selection and the

low mutation rate in cpDNA.

a) Genetic drift will act differently on cpDNA than on nuclear DNA because of haploidy and predominant fixation of cpDNA within individuals, which results in about one-half the effective population size for chloroplast compared to nuclear genes. This leads us to expect roughly twice the degree of genetic subdivision for neutral chloroplast genes than for nuclear genes in monoecious plants. We observed, however, that estimates of subdivision for cpDNA were more than four times higher than for nuclear genes (allozymes).

b) Natural selection may play a much larger role in the genetic architecture of organelle genes than it does for nuclear genes because of its high degree of linkage disequilibrium. This results from the haploid nature of chloroplast genomes, the rarity of heteroplasmy and associated biparental inheritance, intracellular genetic drift during vegetative segregation, and the lack of a sexual process to enable recombination among unlike genomes. The lack of effective recombination potentiates periodic selection--which is an extreme form of genetic hitchhiking; natural selection for a single favorable mutation will purge populations of variability.

c) The lower mutation rate for cpDNA (ca. 500 to 2,000 times lower than for allozymes) will act to slow accumulation of genetic variability in populations. Combined with drift or periodic selection, this will result in a high degree of population subdivision--especially where migration among populations is limited.

3. The genetic patterns discovered in bishop pine cpDNA, compared to allozymes, may not apply for other conifers. Bishop pine has several unique characteristics compared to other conifers. It has developed intraspecific crossing barriers, and most populations are physically distant from one another. This will inhibit gene flow and thus enhance development of population subdivision. There might also have been strong selection for cpDNA variants during bishop pines northward migration. For example there is evidence of much higher frost-tolerance in northern "blue" foliage race than in the "green" foliage race. Perhaps selection for different photosynthetic characteristics associated with this altered foliage resulted in periodic selection for mutant cpDNA genomes. Because of a lack of recombination, this would also drive neutral variants to fixation along with the selected variants.

4. The low rate of homoplasy in our phylogenetic

topologies suggests that cpDNA is well suited for phylogenetic analyses. Biochemical markers such as allozymes or morphological traits, which have traditionally been emphasized in taxonomic studies, seem more prone to problems of sampling bias resulting from high polymorphism within species or populations, convergent evolution as a consequence of natural selection (morphological traits), and heterogeneity of evolutionary rates. Nonetheless, despite a low degree of convergence, substantial rate variability was observed for cpDNA--apparently confounding the commonly used UPGMA method of phylogenetic analysis.

5. Chloroplast DNA restriction site mutations would be highly effective genetic markers for identifying seed sources of bishop pine as a consequence of geographic group-specific mutations. This may be of use in tree improvement studies to identify seed of unknown origin.
6. The near lack of cpDNA diversity within Monterey and knobcone pine, and within geographic groups of bishop pine, supports the common practice of using very small samples for cpDNA-based phylogenetic studies. A safe guideline would be to attempt to sample a single individual from each of a very few distinctive and isolated populations in each species.

7. Future work could help to test the hypotheses generated in this study by:

   a) Analysis of cpDNA length mutations at the DNA sequence level. This would allow homology among "length-alleles" to be verified, and thus contrasting organization of genetic diversity to that of site mutations to be better understood.

   b) Analysis of organization of genetic diversity for mitochondrial DNA. This will enable a test of whether the results we found for cpDNA are general for organelles. It also allows an even greater contrast among genomic and site mutation genetic diversity--since the former evolves slower and the latter faster in mtDNA than in cpDNA.

   c) Analysis of nuclear genomes at the DNA rather than the gene product level--thereby facilitating comparisons with organelle DNA. This could readily be done using either standard RFLP methods, or the new random amplified polymorphic DNA (RAPD) technique. Polymorphic nuclear probes and RAPD primers are now becoming widely available as a result of genome mapping efforts.

BIBLIOGRAPHY

Ali, I. F., D. B. Neale, and K. A. Marshall. 1990. Chloroplast DNA restriction fragment length polymorphism in coast redwood, Douglas-fir, incense-cedar, and loblolly pine. Theor. Appl. Genet. 81: 83-89.

Axelrod, D. I. 1983. New Pleistocene conifer records, coastal California. Univ. Cal. Publ. Geol. Sci. 127: 1-108.

Bairoch, A. 1989. Pc/gene user and reference manuals (release 6.01). IntelliGenetics, Inc. Geneva, Switzerland.

Banks, J. A. and C. W. Birky, Jr. 1985. Chloroplast DNA diversity is low in a wild plant, Lupinus texensis. Proc. Natl. Acad. Sci. USA 82: 6950-6954.

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. Jour. Sci. 5: 486-495.

Bannister, M. H. and J. R. McDonald. 1983. Turpentine composition of the pines of Guadalupe and Cedros Islands Baja California. N. Z. Jour. Bot. 21: 373-377.

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

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

Brown, A. G. and J. C. Doran. 1985. Variation in growth and branching characteristics of Pinus attenuata. Silvae Genet. 34(2-3): 100-104.

Clegg, M. T. 1989. Molecular diversity in plant populations, pp 98-105. In: Brown, A. H. D., M. T. Clegg, A. L. Kahler, and B. S. Weir, (Eds). Plant population Genetics, Breeding, and Genetic Resources. Sinaur Associates, Inc. Sunderland, Massachusetts.

Coates, D. and C. A. Cullis. 1987. Chloroplast DNA variability among Linum species. Amer. J. Bot. 74: 260-268.

Crawford, D. J., J. D. Palmer, and M. Kobayashi. 1990. Chloroplast DNA restriction site variation and the phylogeny of Coreopsis section Coreopsis (Asteraceae). Amer. J. Bot 77: 552-558

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 Genet. 16(3): 89-97.

Dang, L. H. and D. R. Pring. 1986. A physical map of the sorghum chloroplast genome. Plant Mol. Biol. 6: 119-123.

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

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.

Doebley, J., W. Renfroe, and A. Blanton. 1987. Restriction site variation in the Zea chloroplast genome. Genetics 117: 139-147.

Duffield, H. W. 1951. Interrelationships of California closed cone pines with special references to Pinus muricata D. Don. Ph.D. Dissertation, Univ. Calif. Berkeley. 114 pp.

Feinberg, A. P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Felsenstein, J. 1985a. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783-791.

Felsenstein, J. 1985b. Phylogenies and the comparative method. Am. Nat. 125: 1-15.

Felsenstein, J. 1988. PHYLIP: Phylogeny Inference Package. Version 3.1 Univ. of Washington, Seattle, WA.

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

Fitch, W. M. 1977. On the problem of discovering the most parsimonious tree. Am. Nat. 111:223-257.

Forde, M. B. 1964. Variation in natural populations of Pinus radiata in California. Part 1-4. N. Z. Jour. Bot. 2: 213-359;459-501.

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

Govindaraju, D. R., B. P. Dancik, and D. B. Wagner. 1988. Novel chloroplast DNA polymorphism in a sympatric region of Two pines. J. Evol. Biol. 2: 49-59.

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

Higuchi, R.G. and H. Ochman. 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. 17: 5865.

Hills, D. M. 1987. Molecular versus morphological approaches to systematics. Ann. Rev. Ecol. Syst. 18: 23-42.

Hipkins, D. V., C.-H. Tsai, and S. H. Strauss. 1990. Sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from a gymnosperm, Douglas fir. Plant Mol. Biol. 15: 505-507.

Howell, J. T. 1941. The closed-cone pines of insular California. Leaflets of Western Botany 3: 1-24.

Jansen, R. K. and J. D. Palmer. 1988. Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). Amer. J. Bot. 75: 753-766.

Keim, P., K. N. Paige, T. G. Whitham, and K. G. Lark. 1989. Genetic analysis of an interspecific hybrid swarm of Populus: occurrence of unidirectional introgression. Genetics 123: 557-565.

Kirk, J. T. O. and R. A. E. Tilney-Bassett. 1978. The plastids. Their chemistry, structure, growth, and inheritance. Second edition. Elsevier, Amsterdam.

Lamppa, G. K. and A. J. Bendich. 1979. Chloroplast DNA sequence homologies among vascular plants. Plant Physiol. 63: 660-668.

Li, W.-H. and D. Graur. 1991. Fundamentals of Molecular evolution. Sinauer Associates, Inc. Sunderland, Massachusetts.

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.

Linhart, Y. B., B. Burr, and M. T. Conkle. 1967. The closed-cone pines of the northern Channel Islands. In: R.B. Philbrick (ed.), Proceedings of the symposium on the biology of the California islands pp. 151-177. Santa Barbara Bot. Garden.

Lowe, T., J. Sharefkin, S. Q. Yang, and C. W. Dieffenbach. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reactions. Nucleic Acids Research 18: 1757-1761.

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

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, J. L. 1930. The Santa Cruz Island pine. Madroño 2: 10.

Millar, C. I. 1989. Allozyme variation of bishop pine associated with pygmy-forest soils in northern California. Can. J. For. Res. 19: 870-879.

Millar, C. I. 1986. The Californian closed cone pines (subsection Oocarpae Little and Critchfield): A taxonomic history and review. Taxon 35: 657-670.

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

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 Californian 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 the conservation of the five natural populations of Pinus radiata. Can. J. For. Res. 18: 506-514.

Mukai, T. and C. C. Cockerham. 1977. Spontaneous mutation rates at enzyme loci in Drosophila melanogaster. Proc. Natl. Acad. Sci. 74: 514-2517.

Murai, K. and K. Tsunewaki. 1987. Chloroplast genome evolution in the genus Avena. Genetics 116: 613-621.

Neale, D. B., M. A. Saghai-Maroof, R. W. Allard, Q. Zhang, and R. A. Jorgensen. 1988. Chloroplast DNA diversity in populations wild and cultivated barley. Genetics 120: 1105-1110.

Neale, D. B., N. C., Wheeler, and R. W. Allard. 1986. Paternal inheritance of chloroplast DNA in Pseudotsuga menziesii Can. J. For. Res. 16: 1152-1154.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. 41: 225-233.

Nei, M. and J. C. Miller. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125: 873-879.

Nei, M. and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105: 207-217.

Newcomb, G. B. 1962. Geographical variation in Pinus attenuata Lemmon. Ph.D, Thesis, Univ. Calif., Berkeley. 191 pp.

Ogihara, Y. and K. Tsunewaki. 1988. Diversity and evolution of chloroplast DNA in Triticum and Aegilops as revealed by restriction fragment analysis. Theor. Appl. Genet. 76: 321-332.

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

Palmer, J. D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am. Nat. 130: S6-S29.

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

Palmer, J. D., R. A. Jorgensen, and W. F. Thompson. 1985. Chloroplast DNA variation and evolution in Pisum : Patterns of change and phylogenetic analysis. Genetics. 109: 195-213.

Palmer, J. D., C. R. Shields, D. B. Cohen, and T. J. Orton. 1983. Chloroplast DNA evolution and the origin of amphidiploid Brassica species. Theor. Appl. Genet. 65: 181-189.

Palmer, J. D. and W. F. Thompson. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat is lost. Cell 29: 537-550.

Palmer, J. D. and D. Zamir. 1982. Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. Proc. Natl. Acad. Sci. USA 79: 5006-5010.

Palumbi, S. R. and A. C. Wilson. 1990. Mitochondrial DNA diversity in the sea urchins Strongylocentrotus purpuratus and S. droebachiensis. Evolution 44: 403-415.

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

Reed, K. C. and D. A. Mann. 1985. Rapid DNA transfer from agarose gels to nylon membranes. Nucleic Acids Res. 13: 7207-7221.

Rieseberg, L. H., D. E. Soltis, and J. D. Palmer. 1988. A molecular reexamination of introgression between Helianthus annuus and H. bolanderi (Compositae). Evolution 42: 227-238.

Ritland, K. and M. T. Clegg. 1987. Evolutionary analysis of plant DNA sequences. Am. Nat. 130: S74-S100.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Scowcroft, W.R. 1979. Nucleotide polymorphism in chloroplast DNA of Nicotiana debneyi. Theor. Appl. Genet. 55: 133-137

Sears, B. B. 1980. Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. Plasmid 4:233-255.

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. Jour.For. 27(1): 50-66.

Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, G. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO 5: 2043-2049.

Sigurgeirsson, A. and A. Szmidt. 1988. Chloroplast DNA variation among North-American Picea species, and its phylogenetic implications. Proceedings of the frans Kempe symposium in Umeå, Sweden. 49-65.

Slatkin, M. 1987. Gene flow and the Geographic structure of natural populations. Science 236: 787-792.

Soltis, D. E., P. S. Soltis, T. A. Ranker, and B. D. Ness. 1988. Chloroplast DNA variation in a wild plant, Tolmiea menziesii. Genetics 121: 819-826.

Stine, M., B. B. Sears, and D. E. Keathley. 1989. Inheritance of plastids in interspecific hybrids of blue spruce and white spruce. Theor. Appl. Genet. 78: 768-774.

Strauss, S. H., J. Bousquet, V. D. Hipkins, and Y.-P. Hong, 1991. Contributions of biochemical and molecular genetic markers to biosystematic studies of forest trees. New Forests (in press).

Strauss, S. H. and M. T. Conkle. 1986. Segregation, linkage, and diversity of allozymes in knobcone pine. Theor. Allo. Genet. 72: 483-493.

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

Strauss, S. H., A. H. Doerksen, and J.R. Byrne. 1990. Evolutionary relationships of Douglas-fir and its relatives (genus Pseudotsuga) from DNA restriction fragment analysis. Can. J. Bot. 68: 1052-1510.

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.

Swofford, D. L. 1990. PAUP: Phylogenetic analysis using parsimony, user's manual, version 3.0. Illinois Natural History Survey, Champaign, IL.

Swofford, D. L. and G. J. Olsen. 1990. Phylogeny reconstruction, pp 411-501. In: Hills D.M. and C. Moritz. (Eds). Molecular Systematics. Sinauer Associates, Inc. Massachusetts.

Sytsma, K. J. and L. D. Gottlieb. 1986. Chloroplast DNA evolution and phylogenetic relationships in Clarkia sect. Peripetasma (Onagraceae). Evolution 40: 1248-1261.

Sytsma, K. J. and B. A. Schaal. 1985. Phylogenetics of the Lisianthius skinneri (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 39: 594-608.

Szmidt, A. E., T. Alden, and J.-E. Hallgren. 1987. Paternal inheritance of chloroplast DNA in Larix. Plant Molec. Biol. 9: 59-64.

Szmidt, A. E., A. Sigurgeirsson, X.-R. Wang. J.-E. Hallgren, and D. Lindgren. 1988. Genetic relationships among Pinus species based on chloroplast DNA polymorphism. Proceedings of the frans Kempe symposium in Umeå, Sweden. 33-47.

Tsai, C.-H. and S. H. Strauss. 1989. Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr. Genet. 16: 211-218.

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.

Weir, B. S. 1990. Intraspecific differentiation, pp 373-410 In: Hills D.M. and C. Moritz. (Eds). Molecular Systematics. Sinauer Associates, Inc. Massachusetts.

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

White, E.E. 1990. Chloroplast DNA in Pinus monticola - 2. Survey of within-species variability and detection of heteroplasmic individuals. Theor. Appl. Genet. 79: 251-255.

Whittam, T. S., A. G. Clark, M. Stoneking, R. L. Cann, and A. C. Wilson. 1986. Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA 83: 9611-9615.

Winship, P. R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res. 17: 1266.

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.

Wolfe, K. H., M. Gouy, Y.-W. Yang, P. M. Sharp, and W.-H. Li. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA 86: 6201-6205.

Zurawski, G. and M. T. Clegg. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Ann. Rev. Plant Physiol. 38: 391-418.

Appendices

Appendix I. Supplementary data.

A. Sample origins

Pinus muricata

Concord Weapons Station

Population[1]	Accession No.	Row/Column
MMa	164002	C-23/80A
MMe	093003	C-23/77A
MMo	071001	C-21/73A
MMo	071002	C-22/83
MMo	071005	C-19/72
MMo	071005	C-19/85A
MMo	071012	C-23/68
MMo	071016	C-22/67A
MMo	071017	C-22/78
MSb	052008	C-22/83A
MSb	052008	C-23/60
MSb	052009	C-23/73A
MSb	052009	C-23/78
MSb	053004	C-21/59A
MSb	053004	C-23/76
MSb	054001	C-19/20
MSb	054002	C-19/76A
MSb	055001	C-22/75A
MSb	055002	C-21/78A
MSb	055002	C-19/73
MSb	056003	C-19/67
MSb	056003	C-21/69A
MSb	056004	C-22/75
MSb	056004	C-20/77A
MSc	112001	C-19/69A
MSc	121001	C-19/68
MSc	121002	C-23/57A
MSc	121007	C-23/69A
MSc	121010	C-20/86
MSc	121010	C-19/65A
MSc	122058	C-22/74
MSc	122058	C-20/69A
MSc	122064	C-23/68A
MSc	122064	C-23/59
MSc	123112	C-22/82
MSc	123113	C-22/72
MSc	123113	C-19/81A
MSc	123138	C-22/58
MSc	123138	C-22/63A
MSc	123143	C-22/76A

Population[1]	Accession No.	Row/Column
MScb	131006	C-23/63A
MScb	131008	C-23/57
MScb	131008	C-23/59A
MScb	131017	C-23/70A
MScb	131017	C-19/75
MScb	131023	C-19/80
MScb	131023	C-20/73A
MScb	132069	C-20/83
MScb	132076	C-19/79
MScb	132076	C-22/55a
MScb	133102	C-22/86
MScb	133102	C-20/68a
MScb	133135	C-21/61
MScb	133139	C-21/80
MScb	133142	C-23/75A
MSlo	061002	C-19/59
MSlo	061003	C-23/58
MSlo	061006	C-23/81A
MSlo	061007	C-22/62A
MSlo	061007	C-19/85
MSlo	064001	C-23/60A
MSlo	064006	C-19/56A
MSr	111021	C-20/71
MSr	111022	C-22/73A
MSr	111022	C-22/69
MSr	112002	C-22/69
MSr	112003	C-19/57
MSr	112005	C-23/79A
MSr	112005	C-23/63
MSr	112006	C-19/82
MSr	112009	C-19/70A
MSr	112010	C-22/81
MSr	112017	C-23/64
MSr	112017	C-20/55A
MSr	112018	C-20/59
MSr	112018	C-22/66
MSv	040001	C-21/78
MSv	041013	C-21/66
MSv	042023	C-22/55
MSv	042023	C-20/59A
MSv	042025	C-22/61
MSv	042025	C-20/80A
MSv	045053	C-23/75
MSv	045053	C-20/58A
MSv	045054	C-20/58

Russell Reservation

Population[1]	Accession No.	Row/Column
MMa	161006	7/3
MMa	161002	1/12
MMa	161004	6/3
MMa	165001	21/39
MMe	091003	23/38
MMe	093001	31/31
MMe	094003	1/21
MMe	094004	3/5
MMe	095003	6/6
MMo	071001	17/31
MMo	071002	5/21
MMo	071003	8/3
MMo	071014	8/2
MMo	071014	18/31
MMo	071020	27/31
MMo	073005	7/2
MSb	052008	19/31
MSb	053003	30/37
MSb	054002	26/31
MSb	054007	7/6
MSc	121001	5/6
MSc	121005	7/11
MSc	121002	22/39
MSc	122064	27/37
MSc	122077	8/11
MSc	123138	29/31
MScb	131008	12/31
MScb	131013	8/9
MScb	131139	1/13
MScb	132065	7/10
MScb	133103	7/9
MScb	133142	25/37
MSlo	061006	32/31
MSlo	061007	24/31
MSlo	063003	13/35
MSlo	064001	25/38
MSlo	064006	1/8
MSo	081002	14/42
MSo	082005	6/4
MSo	083001	14/34
MSo	084003	6/13
MSo	084005	5/4
MSo	085001	2/21
MSo	085002	6/20
MSr	111021	20/31
MSr	112004	8/12

Population[1]	Accession No.	Row/Column
MSr	112008	8/13
MSv	040001	6/8
MSv	041012	34/31
MSv	042023	38/31
MSv	042023	1/20
MSv	042025	11/43
MSv	043033	11/41
MSv	043033	14/31
MSv	043035	25/31
MSv	045053	20/40
MSv	045053	11/35
MSv	045069	1/18
MTr	101002	23/31
MTr	101002	12/39
MTr	102003	38/34
MTr	101005	2/7
MTr	103005	21/31
MTr	103005	12/37
MTr	104003	11/38
MTr	104005	1/7
MTr	105002	12/36

[1] Population abbreviations: MMa=*muricata* Marin, MMe=*muricata* Mendocino, MMo=*muricata* Monterey, MSb=*muricata* Santa Barbara, MSc=*muricata* var. *remorata* Santa Cruz, MScb=*muricata* Santa Cruz, MSlo=*muricata* San Luis Obispo, MSr=*muricata* Santa Rosa, MSv=*muricata* San Vicente, MTr=*muricata* Trinidad

Pinus radiata

Gill hedge clone bank collection

Año Nuevo	Cambria	Monterey	Guadalupe
A01	C02	M02	L06b
A02	C03	M03	L08a
A03	C04	M04	L19b
A05	C11	M05	L21a
A11	C13	M11	L21b
A12	C14	M12	L28a
A16	C21	M13	L35a
A17	C22	M14	L59a
A22	C24	M15	L66a
A24	C31	M21	L72b
A26	C32	M23	L74b
A32	C33	M24	L76a
A35	C34	M31	L77a
A36	C42	M32	L81b
A43	C43	M33	L86b
A46	C44	M34	L89b
A52	C52	M35	
A53	C54	M41	
A62	C61	M52	
A63	C62	M61	
A67	C63	M62	
A72	C71	M72	
A74	C72	M81	
A77	C81	M82	
A82	C83	M83	
A85	C91	M91	
A87	C92	M92	
A92	C93		
A94			
A97			

Russell Reservation

Population	Accession No.	Row/Column
Guadalupe L	141054	36/31
L	141056	12/31
L	143003	38/33
L	143065	7/15
L	144024	19/40
L	144080	8/15
L	145068	6/21
L	145068	17/41

Population		Accession No.	Row/Column
Cedros	T	151030	8/14
	T	152032	22/38
	T	152032	11/32
	T	152037	22/31
	T	152082	7/14
	T	152088	12/40
	T	152090	31/35
	T	152090	13/43
	T	02	46/72c
	T	08	47/67
	T	09	46/64c
	T	10	47/77c
	T	11	31/61
	T	11	42/60
	T	28	31/59
	T	34	44/62
	T	35	38/58
	T	41	46/70
	T	41	46/80c
	T	45	46/78
	T	55	47/73
	T	68	47/71
	T	75	47/65
	T	86	46/76c
	T	90	40/58
	T	91	43/61
	T	97	41/59

B. Computer program for estimating Nei's G_{ST} with allelic frequencies.

```
      Program for estimating Nei's G_ST
      By Yong-Pyo Hong and Byung-Oh Cho
      dimension
      p(50,50,10),HI(50,50),HT(50),PBAR(50,10),hsi(50)
     $  ,kmax(50),X(1500)
      CHARACTER*20 popname(50)
      INTEGER GROUP
c     dimension p(TOT.NO.POPNS., NO.LOCI, kMAX)
c     GROUP=SPECIES OR GROUP OF POPULATIONS
999   READ(5,*,END=1000) GROUP,N,LOCI
      read(5,*) (kmax(i),i=1,loci)
      do 44 m=1,N
      KSUM=0
      DO 13 I=1,LOCI
13    KSUM=KSUM+KMAX(I)
      READ(5,11) POPNAME(M)
11    FORMAT(1X,A20)
      READ(5,*) (X(J),J=1,KSUM)
      J=0
      DO 44 II=1,LOCI
      DO 44 K=1,KMAX(II)
      J=J+1
44    P(M,II,K)=X(J)
```

```
C
      do 78 m=1,n
     write(6,11) POPNAME(M)
     WRITE(6,14) ((P(M,I,K),K=1,KMAX(I)),I=1,LOCI)
78    continue
14    FORMAT(50(5X,10F7.4,/))
C
      do 34 m=1,n
      DO 20 I=1,LOCI
       sumpik=0.
      DO 25 K=1,KMAX(I)
25    SUMPIK=SUMPIK+P(m,I,K)**2
20    HI(M,I)=1.-SUMPIK
34    continue
C
      DO 35 I=1,LOCI
      SUMHI=0.
      DO 30 IN=1,N
30    SUMHI=SUMHI+HI(IN,I)
      HSI(I)=SUMHI/FLOAT(N)
35    CONTINUE
C
      WRITE(*,*) 'HSI(I)'
      WRITE(*,120) (HSI(I),I=1,LOCI)
120   FORMAT(11e10.3,50(/,10e10.3))
      HSBAR=0.
```

```
      DO 40 I=1,LOCI
40    HSBAR=HSBAR+HSI(I)
      HSBAR=HSBAR/LOCI
C
      DO 48 I=1,LOCI
      do 49 k=1,kmax(I)
       sumpik=0.
      do 53 m=1,n
53     SUMPIK=SUMPIK+P(m,I,K)
      PBAR(I,K)=SUMPIK/FLOAT(N)
49    CONTINUE
48    CONTINUE
C
      WRITE(*,*) 'PBAR(I,K)'
      WRITE(*,121) ((PBAR(I,K),K=1,3),I=1,LOCI)
121   FORMAT(3e10.3)
      DO 70 I=1,LOCI
       SUMPBAR=0.
      DO 69 K=1,KMAX(I)
69     SUMPBAR=SUMPBAR+PBAR(I,K)**2
70    HT(I)=1.-SUMPBAR
C
      WRITE(*,*) 'HT(I)'
      WRITE(*,122) (HT(I),I=1,LOCI)
122   FORMAT(11e10.3,50(/,10e10.3))
      HTBAR=0.
```

```
      DO 80 I=1,LOCI
80    HTBAR=HTBAR+HT(I)
      HTBAR=HTBAR/LOCI
C
      WRITE(*,*) 'HSBAR=', HSBAR
      WRITE(*,*) 'HTBAR=', HTBAR
      DST=HTBAR-HSBAR
      WRITE(*,*) 'DST=', DST
      GST=1.-HSBAR/HTBAR
      WRITE(*,*) 'GST=', GST
      WRITE(*,*) '    '
      GOTO 999
1000  STOP
      end
```

1) Example of input data.

1 3 21 **(group number, No. of populations, No. of loci)**

2 2 2 2 2 2 2 2 2 2 2 3 3 2 2 2 2 2 3 2 **(No. of alleles observed in each locus)**

MANO **(population name could be longer than 10 characters)**

1.0 0. 0. 1.0 0. 1.0 1.0 0. 1.0 0.

0. 1.0 1.0 0. 1.0 0. 1.0 0. 1.0 0.

1.0 0. 1.0 0. 0.967 0. 0.033 1.0 0. 0.

1.0 0. 1.0 0. 0. 1.0 0. 1.0 0. 1.0

1.0 0. 0. 0. 1.0 **(observed frequencies of each allele in each locus) (type data within seven columns)**

MCAM **(input data should be typed from the sixth column)**

1.0 0. 0. 1.0 0. 1.0 1.0 0. 1.0 0.

0. 1.0 1.0 0. 1.0 0. 1.0 0. 1.0 0.

1.0 0. 1.0 0. 1.0 0. 0. 1.0 0. 0.

1.0 0. 1.0 0. 0. 1.0 0. 1.0 0. 1.0

1.0 0. 0. 0. 1.

MMON

1.0 0. 0. 1.0 0. 1.0 1.0 0. 1.0 0.

0. 1.0 1.0 0. 1.0 0. 1.0 0. 1.0 0.

1.0 0. 1.0 0. 0.963 0. 0.037 1.0 0. 0.

1.0 0. 1.0 0. 0. 1.0 0. 1.0 0. 1.0

1.0 0. 0. 0. 1.

C. Example of input data for Weir's program for F_{ST} based on haploid data.

5 19 3 **(# of populations, # of loci, max. # of alleles in a locus)**

1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
1	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
2	1	2	1	2	1	2	2	2	2	2	1	1	1	3	2	1	2	2	2
3	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
3	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
3	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
3	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
3	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
4	1	2	1	2	1	2	2	2	2	2	2	1	1	3	1	1	2	2	2
5	1	2	1	2	1	2	1	2	2	2	2	1	1	3	1	1	2	2	2
5	1	2	1	2	1	2	1	2	2	2	2	1	1	3	1	1	2	2	2
5	1	2	1	2	1	2	1	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	1	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2
5	1	2	1	2	1	2	2	2	2	2	2	2	1	3	1	1	2	2	2

↑ **Type of allele in locus 1** ↑ **Type of allele in locus 11**

↑

Individuals in populations: 6 individuals in population 1
7 individuals in population 2..

D. Character states matrices used for parsimony analyses.

1) Character states for Wagner parsimony analysis.

```
 6 29 O    : (No. of OTUs, No. of characters,
             Option=outgroup)
 O 6       : outgroup=6th OTU
Radiata    00110010000000000000111100001
Knobcone   10010100000000110001000001000
Bs-North   00101010111110000101011110010 1
Bs-Interm  00101010111111000100011110010 1
Bs-South   00110010010011000000111000001
Oocarpa    11110011000000011110110111111
```

2) Character states for Dollo parsimony analysis.

```
5 24       : (No. of OTUs, No. of characters)
Radiata    01100100000000000111100 1
Knobcone   10101000000011001000010 0
Bs-North   01010111110000110111101 1
Bs-Interm  01010111111000100111101 1
Bs-South   01100101001100000111000 1
```

E. Input data for computer program - HAPLO.

1) DAG : **Population matrix**

"M1" "M2" "M3" "M4" "M5" "K1" "K2" "K3" "K4" "B1" "B2" "B3"
"B4" "B5" "B6" "B7" "B8" "B9" "B10" "B11" - **population name should be put on the first line.**

29 28 26 24 27 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 25 25 25 25 0 0 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 6 7 0 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 5 7 6 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0

0 0 0 0 0 0 0 0 0 0 0 0 0 0 19 22 12 17 18 20

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 1

- **each horizontal row gives the number of a specific haplotype observed in the respective population.**

2) ILLMT : **Number of shared sites between haplotypes**

```
"I" "II" "III" "IV" "V" "VI" "VII" "VIII" "IX"
242 242 235 241 241 241 241 241 241
243 235 241 241 241 241 241 241
241 234 234 234 234 235 235
249 248 247 248 241 241
248 247 248 241 240
247 247 241 240
249 242 240
244 242
242
58 58 58 58 58 58 58 58 58
58 58 58 58 58 58 58 58
58 58 58 58 58 58 58
58 58 58 58 58 58
58 58 58 58 58
58 58 58 58
58 58 58
58 58
58
```

- **the first line gives the labels for the nine haplotypes.**
- **in this example, there are two types of restriction enzymes, six and four base-pair cutters. Each triangular file gives the total data associated with the two types.**

- **each number is the number of sites shared between two haplotypes.**

Ex) Row one indicates haplotype I has 242 sites recognized by six base-pair cutters, 242 of which are shared with haplotypws II, 235 of which are shared with haplotype III.

1X1	**1X2**	**1X3**	**1X4**	**1X5**	**1X6**	**1X7**	**1X8**	**1X9**
242	242	235	241	241	241	241	241	241

2X2	**2X3**	**2X4**	**2X5**	**2X6**	**2X7**	**2X8**	**2X9**
243	235	241	241	241	241	241	241

.
.
.
.

8X8	**8X9**
244	242

9X9
242

3) DAPHUP : **Haplotype distance-matrix**

```
1 .00030 .00391 .00266 .00237 .00207 .00266 .00119 .00059I
2 .00421 .00295 .00266 .00237 .00295 .00148 .00089II
3 .00659 .00629 .00600 .00659 .00450 .00391III
4 .00029 .00058 .00058 .00324 .00266IV
5 .00029 .00029 .00295 .00296V
6 .00058 .00266 .00267VI
7 .00265 .00326VII
8 .00059VIII
9
```

- **this is the upper triangular part of the fitted distance matrix between all pairs of haplotypes.**
- **one way to get this is to run HAPLO, answering the question "Do you want to compute a haplotype distance matrix?" with a Y (yes). Then save an output distance matrix file, which serves as input to the program HAPLO.**

F. Example of input data for distance-matrix methods based on cpDNA diversity estimated by HAPLO.

MONTEREY : 1

KNOBCONE : 2

BISHOP-NORTH : 3

BISHOP-INTERMEDIATE : 4

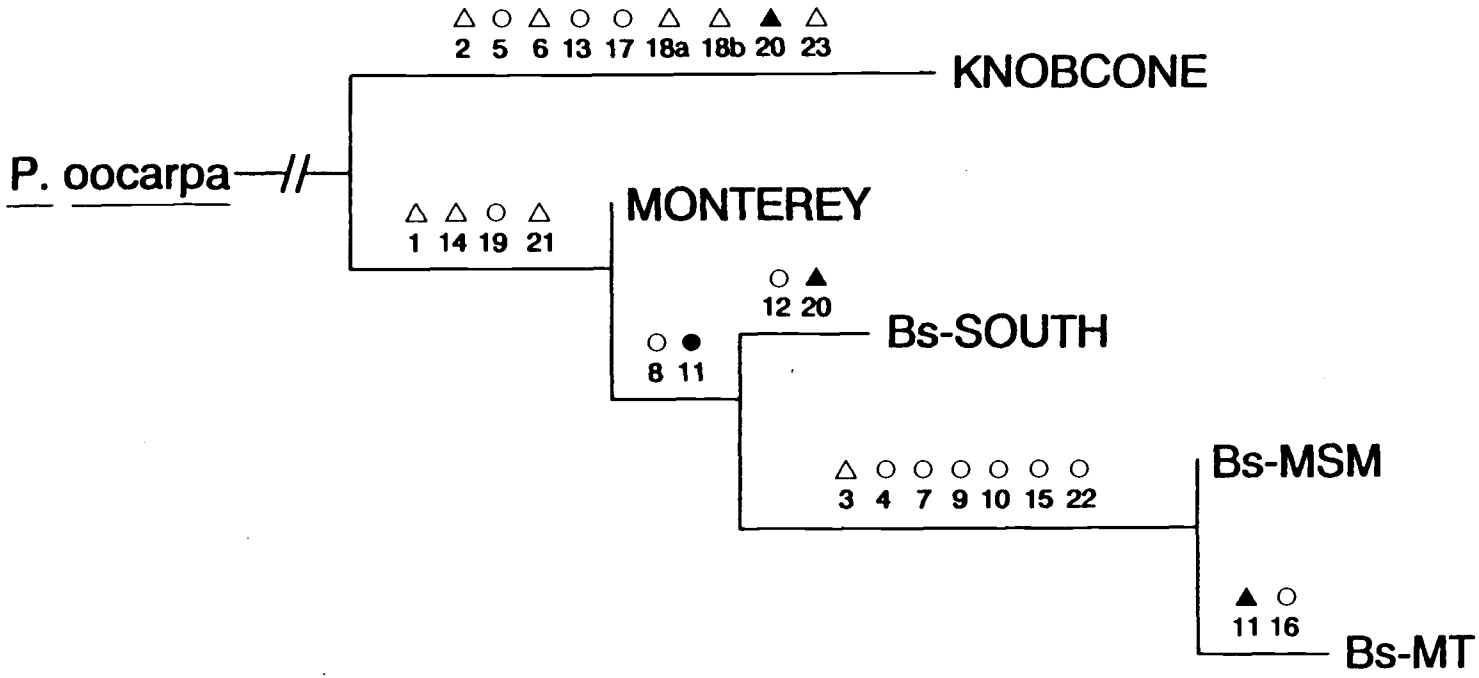
BISHOP-SOUTH : 5

1 x 2	**1 x 3**	**1 x 4**	**1 x 5**	
0.003910	0.002659	0.002276	0.001155	→ **cpDNA diversity**

	2 x 3	**2 x 4**	**2 x 5**
	0.006586	0.006202	0.004471
		3 x 4	**3 x 5**
		0.000298	0.003213
			4 x 5
			0.002798

- this is upper right triangle part of the fitted distance matrix between all pairs of operating taxanomic units (OTUs).

G. Phylogenetic trees showing character state changes on branches for Wagner parsimony analysis. Numbers on the branch denote the observation number in Table II.2. (▲ = parallel site loss, △ = site losses, ○ = site gains)


△ ○ △ ○ ○ △ △ ▲ △
2 5 6 13 17 18a 18b 20 23
KNOBCONE
P. oocarpa
△ △ ○ △
1 14 19 21
MONTEREY
○ ▲
12 20
Bs-SOUTH
○ ●
8 11
△ ○ ○ ○ ○ ○ ○
3 4 7 9 10 15 22
Bs-MSM
▲ ○
11 16
Bs-MT

Appendix II. Supplementary detail on methods.

A. Procedures for modified CTAB DNA peparation.

1. Start with 40 g of fresh needles. Avoid brown needles, fascicles, twigs. Pull from branches 1-7 days before use. For pines or broadleaved plants, chop foliage into short pieces (ca. 1 cm long or 1 cm^2 in area, respectively). Store frozen, chopped, weighed needles in labelled, sealed plastic bag.

2. Cut cheesecloth (4 layers) and miracloth (1 layer) for each sample and fold miracloth into quarters for use in funnel. Place miracloth below cheesecloth in funnel.

3. Take frozen needles and immediately place under ice for transport to cold room for grinding. Do not let needles thaw before grinding.

5. In cold room, put frozen needles into large Waring blender and add liquid N_2 to same depth as needles. Grind at low setting for 10-15 seconds and then check needles. Continue grinding at high setting for 20 seconds to 1 minute (until ice crystals all form frost on outside of container). Needles should be a fine powder. If not, add more liquid N_2 and grind at high setting. Pour powdered sample into 1,000 ml tri-pour beaker, add 170 ml extraction buffer (make sure sample is not frozen against beaker wall), and grind with polytron (setting 5, small generator probe) until sample flows smoothly. Do not rinse blender between samples but do rinse polytron between samples.

6. Pour homogenate into funnel through cloths. Rinse beaker with 30ml extraction buffer and pour into funnel. Squeeze cheesecloth firmly to remove all juice. Let homogenate filter through miracloth into 1,000 ml tri-pour beakers (surrounded by ice in dishpan) while doing all other samples. If necessary, shift miracloth to unclogged areas and squeeze very gently.

7. Transfer to lab in ice inside dishpan with centrifuge caps closed. Balance pairs of tubes to ± 0.1 g by adding H_2O to low samples. Spin in cold GSA rotor at 9,000 rpm for 10 minutes, 4° C.

8. Pour off supernatant and resuspend pellets by rapid brushing in 10 ml of ice cold Wash buffer.

9. When pellet is resuspended, transfer to labelled 50 ml centrifuge tube, add 1/5 volume of 5% sarkosyl (or 1/10 volume of 10% sarkosyl), and shake vigorously. Keep tube on ice until all samples are resuspended in sarkosyl.

10. Bring to room temperature by sitting 15 to 30 minutes in room temperature water in beaker.

11. Add 1/7 volume of 5 M NaCl and shake vigorously.

12. Add 1/10 volume 8.6% CTAB, 0.7 M NaCl and shake vigorously.

13. Incubate at 60°C for 10 minutes.

14. Add about 20 ml of chloroform/octanol (raise volume to 35 ml total) and shake vigorously.

15. Centrifuge at speed setting 7 in IEC clinical centrifuge at room temperature for 10 minutes (if emulsion does not compress, transfer top layer and emulsion to 15 ml Falcon tube(s) and spin at 6,500 rpm for 10 minutes in SS-34 rotor with adapters in superspeed centrifuge at 4 to 25°C).

16. As soon as sample comes out of centrifuge, carefully transfer upper aqueous phase to second labelled 50 ml centrifuge tube avoiding all material at interface.

17. Add 2/3 volume of isopropanol, mix several times by inversion and then swirl to aid DNA to clot.

18. Hook out clot of precipitate with hooked, labelled Pasteur pipette and let as much solution run from precipitate as possible (press against side of tube).

19. Transfer the precipitate to labelled 50 ml centrifuge tube containing about 20 ml 76% EtOH/10mM NH_4Ac. Set for 20 minutes to overnight.

20. Let as much solution run from precipitate as possible again by pressing against side of tube and then invert labelled pipettes in microtube rack. Let dry in laminar flow hood for 10 to 30 minutes, or if no hood is available, put in vacuum chamber for about 3 minutes and then transfer to 1.5 ml microtube (sterile) containing from 200-1,000 μl of TE (10:0.1)--depending on size of precipitate. Remove DNA from pipette after about 10 minutes by gently lifting and spinning pipette in tube. Close tube and place in refrigerator for

several days. Use cut-tip pipettes, or 60°C incubation, to help resuspend DNA if it does not resuspend during this time.

21. Label tubes clearly with time tape and store in refrigerator.

Extraction Buffer
0.1% BSA* **
0.35 M Sorbitol
50 mM Tris pH 8.0
5 mM EDTA
10% PEG 4000*
0.5% 2-mercaptoethanol*
0.1% spermine tetrachloride
0.1% spermidine trihydrochloride

Wash Buffer
.35 M Sorbitol
50 mM tris pH 8.0
25 mM EDTA
0.1% 2-mercaptoethanol*

*Add within about 1 week of use **Add first

B. Protocol for alkaline Southern Transfers.

1. After photographing gel, trim away unused portions and notch upper right corner (i.e., removing upper portion of wells; looking down at top of gel).

2. To study large fragments (> 10 Kb), depurinate with acid by soaking gel in several volumes of 0.25 M HCl for 9 minutes with gentle shaking. Drain and rinse once with tap deionized H_2O.

3. Denature DNA by soaking gel in several volumes of 0.4 M NaOH with gentle shaking for 20-30 minutes.

4. Cut nylon filter (Zetabind) and 5 pieces of Whatman 3MM filter paper to dimensions about 2 mm greater than gel with wearing gloves. Soak Zetabind in H_2O until evenly wet for 2-3 minutes. Notch Zetabind upper left corner.

5. Wet 3 of the Whatman papers in 0.4 M NaOH (= transfer medium) and place on saran wrap on table top. Wet thoroughly; be sure there are no bubbles.

6. Carefully invert gel using a pair of plexiglass plates so bottom faces up for transfer and lay down on Whatman paper soaked in 0.4 M NaOH. Remove any bubbles with fingertips.

7. If needed, surround gel with saran wrap and/or spacers to insure transfer is not short-circuited.

8. Pipette several drops of dH_2O onto top of gel and carefully lay on Zetabind by lining up gel wells with one edge. Use finger to force all bubbles out gently.

9. Soak other 2 Whatman papers with dH_2O and lay on top of Zetabind with avoiding bubbles.

10. Add about 2 inches of paper towels on top. Put a plexiglass plate and about 1/2 Kg of mass on top.

11. Let transfer proceed 6-48 hours (usually overnight). Turn on vacuum oven so it is hot (80°C) when transfer is complete.

12. Put on gloves and remove blotting papers and Zetabind. Keep DNA side up and avoid touching it. Label Zetabind at the edge of DNA side with a pencil. Soak in 6X SSC

with gentle shaking for about 20 minutes.

13. Place wet filter in Whatman 3 MM folder cut about 2 cm larger than the filter and let it dry for about 1 hour in incubator or at room temperature. Then bake at 80°C for 1-1/2 to 2 hours. Store in sealed bag at room temperature until ready to use. Prewash filters in .1x SSC, .5% SDS, at 65° for one hour before first use.

0.25 M HCl	0.4 M NaOH	20X SSC
-40.4 ml conc. HCl (37%)	-80.0 g NaOH	-876.5 g NaCl
-H_2O to 2 l	-H_2O to 5 l.	-441.0 g sodium citrate
		-adjust pH to 7.0 with NaOH or HCl
		-H_2O to 5 l

to get	20X to use
6 x	3 of 10
2 x	1 of 10
1 x	.5 of 10
.5x	.25 of 10
.1x	.05 of 10

C. Protocol for prehybridization.

1. New blot: wash blot for 1 hr. in 0.1x SSC/0.5% SDS at 65°C with shaking at speed 2.
   Old blot: wash blot for 30 min in 0.4 N NaOH at 42°C, 40 rpm.
   wash blot for 30 min in washing solution at 42°C, 40 rpm.

2. Prepare prehybridization solution.

3. Place blot (DNA side in) into roller tube.

4. Remove any washing solution (remove any major bubbles).

5. Put in prehybridization solution.

6. Put on lid and vacuum seal.

7. Place tube on roller, spout on left, and push right side up close to wall.

8. Check tube occasionally.

9. Prehybridize for 6 hrs. to overnight.

Preparation

* 0.1x SSC/0.5% SDS

5	ml	20X SSC
25	ml	20% SDS
970	ml	H_2O
1,000	ml	

* Pre-hybridization solution (place in 65°C incubator)

27.25	ml	H_2O
15	ml	20x SSC
1.25	ml	20% SDS
5	ml	50x Denhardt's solution
0.5	ml	salmon sperm DNA (before adding, denature for 10 min. at 95°C)
49	ml	(add 1ml of 0.5 M EDTA when add probe)

* Washing solution

0.1x SSC	5 ml of 20x SSC
0.5% SDS	25 ml of 20% SDS

0.2 M Tris (pH 7.5)	200 ml of 1 M Tris-HCl 770 ml of H_2O

D. Protocol for probe labeling and hybridization.

1. DNA must be linear. Cut DNA for one hour with appropriate restriction enzyme (usually EcoRI).
2. Stop reaction and denature by heating for 10 min. at 95°C (if marker DNAs, λ-HindIII and ϕ-HaeIII, are also to be labeled, add just before denaturing).
3. Cool samples on ice and spin down.
4. Add solution A, 3 μl per reaction (solution A is mixture of equal parts of dATP, dGTP, dTTP).
5. Add reaction mixture, 2 μl per reaction (reaction mixture is hexanucleotide mix in 10X reaction buffer).
6. Add ^{32}P, X μl (volume depends on age of ^{32}p).
7. Mix using pipettor.
8. Add Klenow enzyme, 1 μl per reaction.
9. Mix using pipettor. Total volume for one reaction is 20 μl.
10. Incubate at 37°C for 30 min. to overnight.
11. Stop the reaction with 2 μl stop buffer (0.2 M EDTA).
12. Heat to 65°C for 10 min..
13. Spin down.
14. Spin through Sephadex G-50 column for 10 min. at setting 6 (IEC centrifuge).

Column preparation:

a. Remove plunger and tip cover from 1 ml tuberculin syringe.

b. Wet glass wool with TE (10:1) and fill bottom of tube to the 0.2 ml mark.

c. Fill column with Sephadex-G-50 and put column in 15 ml falcon tube. Spin 1 min. at setting 6 in IEC clinical

centrifuge.

d. Fill and spin until packed Sephadex is at 0.9 ml level.

e. Add 100 µl TE (10:1) and spin 4 min. at setting 6.

f. Column is ready for Hexamer labeled sample.

g. Place 1.5 ml microfuge tube (with cap removed) in falcon tube to catch the labeled sample after it passes through column.

15. Bring volume of sample to 200 µl with TE (10:1).

16. Put hole in upper part of tube and denature in heat block at 95°C for 10 min.

17. Put tube on ice and then quick spin down.

18. Dilute 5 µl of sample in 995 µl of TE (10:1). Vortex this 1000 µl sample and aliquot 10 µl onto glass filter, dry filter and measure counts in scintillation counter.

19. Add remaining 195 µl sample to prehybridization solution with 1 ml of 0.5 M EDTA, pH 8.0, for each 50 ml of prehybridization solution.

20. Incubate at 65°C for overnight.

E. Protocol for washing blots.

1. Make up 2x SSC/0.1% SDS, and 2x SSC/0.5% SDS, and warm to 65°C.

2. Open tube and pipette out hybridization solution.

   - Pour in 12 ml of 2x SSC/0.1% SDS (two times).
   - Put lid on.
   - Hand turn.
   - Pipette out washing solution.

3. Place each blot between two screens in container.

   - Pour in 500 ml of 2x SSC/0.1% SDS.
   - Shake at 65°C for 30 min.

4. Pour washing solution into waste bottle, repeat 2x SSC/0.1% SDS washing.

5. Wash blots in 500 ml of 2x SSC/0.5% SDS for 30 min at 65°C.

6. Repeat 2x SSC/0.5% SDS washing.

7. Dry blot with paper towels.

   - Wrap with saran wrap.
   - Expose with intensifying screen.
   - Store in -80°C freezer.

PREPARATION

* 2x SSC/0.1% SDS (2,000 ml)

200	ml	20x SSC
10	ml	20% SDS
1790	ml	H_2O

* 2x SSC/0.5% SDS (2,000 ml)

200	ml	20x SSC
50	ml	20% SDS
1750	ml	H_2O

F. Protocol for producing single-stranded DNA strands from double-stranded PCR DNA by digesting one strand with λ-exonuclease.

1. Phosphorylate the 5'end of one of the primer pair.

A) For each 100 pmol of primer:

1)	Primer (50 pmol/μl)	2μl
2)	10X Kinase buffer	2μl
3)	10mM ATP	4μl
4)	Polynucleotide Kinase (10 units/μl)	2μl
5)	H_2O up to 20 μl	10μl

B) Incubate at 37°C for 30 min.

C) Stop reaction with 0.5M EDTA 0.8μl

D) Ethanol precipitation.

1) Add 1/10 volume of 3M KoAc, pH 4.8 2.1μl

2) Add 2 volume of 100% ETOH 46.0μl

3) Leave in ultra cold freezer for 30 min.

4) Spin at 12,000 rpm for 30 min (cold centrifuge).

5) Wash with 70% ETOH and vacuum dry.

6) Resuspend with H_2O 1.5μl

2. DNA amplification by polymerase chain reaction (PCR) using 5'end phosphorylated and 3'end unphosphorylated primer pair.

3. λ-exonuclease treatment:

A) Ethanol precipitation (1/10 volume of 2M AmAc plus 2 volume of 100% ETOH).

B) Resuspend in 50 μl of λ-exonuclease buffer.

C) Add 1μl (4 units) λ-exonuclease.

D) Incubate at 37 for 30 min.

E) Extract with phenol/chloroform/IAA (25:24:1).

F) Extract with chloroform/IAA (24:1).

G) Ethanol precipitation (1/10 volume of 2M AmAc plus 2 volume of 100% ETOH).

H) Resuspend in TE (10:0.1) and store at 4 C.

4. DNA purification by gel electrophoresis and ultrafiltration in Centricon-30.

10X Kinase buffer (PNK buffer)

0.5M Tris-HCl, pH 7.6
0.1 M MgCl2
50 mM dithiothreitol
1 mM Spermidine
1 mM EDTA, pH 8.0
Filter sterilize

λ-exonuclease buffer

67mM glycine-NaOH, pH 7.4
2.5 mM MgCl2
Filter sterilize

G. Protocol for sequencing of PCR-amplified DNA.

1. Make 10X TBE (should be prepared no longer than 24 hours).

	Large gel	Small gel
Trizma	21.6 g	12.96 g
Boric acid	11.0 g	6.60 g
EDTA	1.86 g	1.12 g
H_2O	to 200 ml	to 120 ml

2. Mix gel components.

	Large gel	Small gel
H_2O	48.8 ml	31.57 ml
10X TBE	17.0 ml	11.0 ml
Liquid acrylamide	24.3 ml	15.73 ml
Liquid bis	25.5 ml	16.5 ml
Urea	71.57 g	46.31 g

3. Make 25% ammonium persulfate.

Ammonium persulfate	0.15 g
H_2O	600 μl

4. Clean the backplate.

   A) Carefully place the backplate into the sink.

   B) Use dish detergent solution and Kimwipes to clean the plate.

   C) Use a circular motion to scrub the entire surface.

   D) Wash off the detergent with deionized water. Make sure all of the detergent is removed.

   E) Clean with 70% ethanol after washing.

5. Siliconize the surface of the backplate by applying Rain-X. Apply Rain-X every 5-10 sequencing electrophoreses by following manufacture recommendation.

6. Cleaning the front plate.

   A) Clean with BonAmi and Kimwipes (soaking overnight in a dish detergent solution helps clean the plate).

   B) Rinse with water and dry with Kimwipes--both surfaces.

   C) Clean with 95% ethanol.

7. Assemble gel form.

   A) Clean spacers and side clamps with 95% ethanol.

   B) Lay down the back plate and place spacers along the sides. Thick side of the spacers should be placed toward the bottom side of the gel form.

   C) Lay the front plate on top of spacers.

   D) Force on the side clamps and make sure the plates are even at the bottom.

8. Casting tray.

   A) Unscrew the tightening screws on the casting tray.

   B) Place a foam pad to the bottom of the tray and lay a filter paper strip on the pad. Filter paper should be cut about half inch shorter than the pad because filter paper will be extended by gel solution.

   C) Combine the casting gel components:

      1) 20 ml acrylamide gel solution.
      2) 140 μl ammonium persulfate.
      3) 100 μl Temed (gel hardens in about 20 seconds).

   D) Mix and pour onto the filter strip.

   E) Put the gel form on this gel solution and tighten the screws against the form side-clamps.

   F) Hold the form in an upright position for 2 minutes.

   G) Lay the form on the wall in upright position.

9. Pouring the gel.

   A) Degas the remaining 150 ml of gel--use the vacuum pump and evacuate for 10-15 minutes. Break vacuum before turning off the vacuum pump. After degassing, do not allow any bubbles into the gel.

   B) Put gel form front side down over the ice cube containers that should be located upper portion of the gel form to give an angle. Lay couple of paper towels underneath the ice cube containers.

   C) Filter through a Whatman #4 filter--allow the filtrate to run down spatula to prevent bubbles.

   D) Add together the following: (will polymerize in about 15 minutes).

Large gel		Small gel	
Gel solution	150 ml	Gel solution	90 ml
Ammonium persulfate	320 μl	Ammonium persulfate	192 μl
Temed	80 μl	Temed	48 μl

   E) After mixing without bubbles, pour the gel solution very gently along the spatula. Start to fill the bottom right first and then along the right side by lifting the gel form to give an angle. While pouring the gel solution, move the gel form up and down slightly to fill the gel form evenly from bottom right to upper left. Gel solution should be poured within ten minutes.

   F) Pour gel solution until gel form is filled and lay down the gel form over the ice cube container to keep an angle. Place combs between glass plates with teeth facing outward. Put combs slowly into the plates to prevent bubbles until teeth is lined up with the edge of the front plate. Put two large clamps over the plates to secure the combs. Keep remaining gel solution in beaker to see whether gel solution gets well polymerized or not.

   G) Wait 45 min. to allow polymerization (check beaker).

10. Dilute remaining 10X TBE buffer to 1X buffer by adding 1647 ml of H_2O. Heat 1X TBE buffer up to 55°C.

11. Mounting the apparatus.

   A) Pour about 300-400 ml 1X TBE buffer into base

tank.

B) Separate gel form from casting tray and get rid of filter paper strip.

C) Put gel form into the base chamber and then assemble parts together.

D) Pour the rest of buffer over the top after removing the two clamps holding the combs. Keep about 200 ml 1X TBE buffer for washing micropipette tip during loading samples.

E) Pull combs out carefully to avoid distortion of the gel and wash the excess of urea on top of gel before putting combs back by disposable plastic pump.

F) Put combs in inverted position (teeth facing downward) until teeth touches the gel about 1mm depth.

G) Connect the electric cords and set power supply for maximum as follows: 200 Watts, 200 Amp., and 2000 Volts.

H) Let run at 1700-1800 Volts to heat to 60°C for 60 minutes. Check watts frequently because that may increase with time.

12. Loading samples.

A) Decrease voltage to 1500-1600 Volts just before loading.

B) Check gel temperature by reading strip thermometer on the front plate and maintain temperature above 50°C.

C) Mark loading zone with marker by dividing entire area (maximum 48 wells plus one spacer between combs. by four wells per sample, which makes it possible to load 12 samples.

D) Turn voltage down and turn off power supply.

E) Wash out urea off the wells with disposable plastic pump.

F) Denature samples at 85°C for three min. in heat block.

G) Load 3 μl of each sample using preset micropipette. Rinse micropipette tip by five shots of 1X TBE buffer into small beaker taking from beaker with fresh buffer between samples. Keep 3 μl of 1X TBE buffer in micropipette tip until loading next sample.

H) Load samples by group of 8 or 12. After loading each group, let the samples get into the gel at 1000 Volts, then run at 1300 Volts for one min. to allow the dyes to separate, and then run at 1700 Volts to get temperature kept above 50°C.

I) After finishing loading, run samples at 1500-1600 Volts for two to three hours until the front dye is migrated to about one inch from the bottom.

J) Keep gel temperature around 50°C.

K) Bring three pounds of dry ice and make 1200 ml of gel fixation solution (10% methanol and 10% acetic acid) during running.

13. Dismounting gel form.

A) Turn off power supply and disconnect electric cords gently.

B) Take gel form to sink, wash bottom portion, pour off buffer from chamber, and wash chamber with deionized water.

C) Put gel form front side down on counter, disassemble by pulling off the side clamps gently (be careful not to squeeze plates), and take combs away.

D) Take the backplate off gently by lifting upper right corner (check air coming from the upper right corner).

E) Put front plate with gel on it in tray and add fixation solution gently with chicken pump toavoid lifting gel from plate. Circulate solution by pumping over the gel every 15 min. for 45 min. or until bottom track dye turns yellowish.

F) Discard 1X TBE buffer in base chamber into S^{35} liquid waste bottle.

14. Dry gel.

A) Cut a piece of Whatman #3 little bit smaller than gel width.

B) Take away fixation solution with pump.

C) Put filter paper on top of gel carefully to avoid forming air bubbles between them.

D) Cut the excess of gel around filter paper with a scalpel.

E) Lift the gel with paper, reverse them, put on extended arm, and transfer on clean counter.

F) Cover gel with two pieces of Saran wrap carefully to avoid forming air bubbles. Cut the excess of Saran Wrap with a razor.

G) Put another piece of precut filter paper on gel dryer, transfer gel to gel dryer using a thin plastic plate, and put gel on precut filter paper.

H) Set vacuum timer at 1 1/2 hour. Set temperature at 80°C for 1 3/4 hour.

I) Start vacuum pump and check for air leaks.

15. Cleaning gel apparatus.

16. Exposing and developing film.

A) After 1 1/2 hour, stop dryer by taking away the vacuum and then turning off the vacuum pump (change vacuum pump oil).

B) Get rid of Saran Wrap delicately. For better results, leave it at room temperature for overnight to help drying.

C) Take gel stuck on Whatman paper into dark room.

D) Put a piece of X-ray film at the bottom of cassette and then put gel upside down on top ofthe film.

E) Leave for 24 hours to several days.

F) Develop film.
1) 5 min. in developer without agitating.

2) 5 min. in fixer with occasional agitating.
3) 10 min. in water with occasional agitating.