AN ABSTRACT OF THE THESIS OF

Valerie D. Hipkins for the degree of Doctor of Philosophy
in Genetics and Forest Science presented on June 3, 1993

Title: Repeated Sequences Associated with Inversions and Length Mutations in the Chloroplast Genomes of *Pseudotsuga* and *Pinus*

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

To help understand the mechanisms of conifer chloroplast genome evolution, we characterized the nucleotide sequences of Douglas-fir (*Pseudotsuga menziesii*) and Monterey pine (*Pinus radiata*) chloroplast DNA that are associated with a length mutation hotspot and inversion. The chloroplast DNA of these two conifers are characterized by shared large inversions that are not found in angiosperms or non-coniferous gymnosperms, and are distinguished from each other by an additional 40-50 kb inversion. We cloned and sequenced common endpoints of the 40-50 kb inversion from Monterey pine and Douglas-fir, and a hypervariable chloroplast DNA region from two individuals of Douglas-fir and one individual of its relative, *Pseudotsuga japonica*.

We found repeated sequences to be associated with all the observed structural changes in these genomes. The locus of polymorphism in the *Pseudotsuga* hotspot is characterized by hundreds of base pairs of imperfect, tandem direct repeats flanked by a partially duplicated and intact *trn*Y-GUA gene in direct orientation. Repetitive DNA was likely generated at the time of the partial *trn*Y gene duplication by slipped-strand

mispairing and these sequences expanded further by unequal crossing-over. At inversion borders, we observed the presence of inverted repeats that were hundreds of base pairs in length, adjacent tRNA genes, and pseudo-tRNA genes. We propose that homologous recombination between tRNA genes, with the concomitant pseudo-tRNA gene formation, inversion, and creation of repeated sequences at inversion borders, was the cause of this rearrangement.

Repeated Sequences Associated with Inversions and Length Mutations in the Chloroplast Genomes of *Pseudotsuga* and *Pinus*

by

Valerie D. Hipkins

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed June 3, 1993

Commencement June 1994

TABLE OF CONTENTS

Page

INTRODUCTION 1

GOALS AND OBJECTIVES OF THESIS 1

ORGANIZATION OF THESIS 1

OVERVIEW OF RESEARCH 2

LITERATURE REVIEW 4

THE CHLOROPLAST 4
Structure and Function 4
Size, Shape, and Number 5
Evolutionary History 5

THE CHLOROPLAST GENOME 6
Size and Structure 6
Genome Number 8
Gene Order and Content 9
Gene Expression 11
Replication 14
Inheritance 16

CHLOROPLAST GENOME EVOLUTION 17
Recombination Mechanisms 18
Genome Rearrangements 22
Recombination 22
Transposition 31
Repeated Sequences 32
cpDNA Sequence Evolution 37

THE GENERA *PSEUDOTSUGA* AND *PINUS* 37

CHAPTER 1: A Mutation Hotspot in the Chloroplast Genome of *Pseudotsuga* is Caused by Variability in the Number of Direct Repeats Derived from a Partially Duplicated tRNA Gene 39

ABSTRACT 39

INTRODUCTION 40

MATERIAL AND METHODS 41

RESULTS 47

DISCUSSION 61

REFERENCES 66

CHAPTER 2: Inverted Repeats and tRNA Genes are Associated with Endpoints of a Large Chloroplast DNA Inversion Within Conifers 70

ABSTRACT 70

INTRODUCTION 71

MATERIAL AND METHODS 72

RESULTS 74

DISCUSSION 85

REFERENCES 90

CONCLUSIONS 93

BIBLIOGRAPHY 96

APPENDICES

APPENDIX A. ADDITIONAL DATA 112

APPENDIX B. SEQUENCE DATA 118

APPENDIX C. PROTOCOLS 139

LIST OF FIGURES

Figure		Page
1.	Models for homologous recombination.	20
2.	Structure of hotspot-containing clone PMCO (a 2.735 kb Douglas-fir cpDNA fragment).	43
3.	Examples of a length mutation hotspot in the genus *Pseudotsuga*.	48
4.	Identification of the source of length variability within PMCO using the Polymerase Chain Reaction (PCR).	50
5.	Regression analysis used to study relationships of length variability on Southern blots to that observed with PCR.	51
6.	A comparison of the *Pseudotsuga* cpDNA hotspot relative to the homologous area of *Pinus radiata.*	53
7.	Repeat structure and DNA sequence of a *Pseudotsuga menziesii* length mutation hotspot (clone PMCO).	55
8.	Example of the hierarchical structure of repeated units in the *Pseudotsuga* hotspot.	57
9.	Repeat structure of the chloroplast DNA hotspot in three individuals of *Pseudotsuga* representing two species.	58
10.	Histogram showing percent sequence similarity between the *P. japonica* and *P. menziesii* length variable regions.	60
11.	Restriction enzyme maps of sections of the Douglas-fir and Monterey pine chloroplast genomes (modified from Strauss et al (1988)).	75
12.	Endpoints of a 40-50 kb inversion that distinguishes the chloroplast genomes of Douglas-fir and Monterey pine.	77
13.	Comparison of a dispersed repeat family in the chloroplast genomes of Douglas-fir and Monterey pine.	78
14.	Sequence alignment of dispersed repeat family members.	80
15.	Orientation of repeat family members at inversion borders in Douglas-fir and Monterey pine, and at the Douglas-fir hotspot.	83
16.	Partial digests of *Pseudotsuga menziesii* and *P. japonica* hotspot-containing PCR amplified cpDNA	117

LIST OF TABLES

Table | Page

1. Name, type, and source of hotspot-containing cpDNA clone material. 44

REPEATED SEQUENCES ASSOCIATED WITH INVERSIONS AND LENGTH MUTATIONS IN THE CHLOROPLAST GENOMES OF *PSEUDOTSUGA* AND *PINUS*

INTRODUCTION

GOALS AND OBJECTIVES OF THESIS

The objective of this thesis is to gain insight into the factors controlling chloroplast genome evolution by investigating some of the unusual features of conifer chloroplast DNA. Studying exceptional parts of the genome may shed light on the constrained nature of most other genomic regions, and on its high conservation in most angiosperms. The goal of my research is to characterize nucleotide sequences in conifer chloroplast DNA that are associated with mutation hotspots and genome rearrangements. These studies should allow me to infer some of the mechanisms of recombination that are responsible for structural evolution of chloroplast genomes.

ORGANIZATION OF THESIS

This thesis is organized into two chapters that correspond to manuscripts that will be submitted for publication. Chapter 1, entitled "A mutation hotspot in the chloroplast genome of *Pseudotsuga* is caused by variability in the number of direct repeats derived from a partially duplicated tRNA gene", and chapter 2, "Inverted repeats and tRNA genes are associated with endpoints of a large chloroplast DNA inversion within conifers", will be submitted to Molecular and General Genetics and Nucleic Acids Research, respectively, with major revisions. The manuscripts are preceded by a literature review in which I briefly summarize chloroplast structure, function, and evolutionary history. In more detail I cover chloroplast genome evolution, particularly genome rearrangements, repeated sequences, and recombination mechanisms. Chapters

1 and 2 are followed by main conclusions drawn from the entire thesis research and a bibliography containing all references cited within the thesis, including manuscripts. Appendices include a section containing additional data not part of the two manuscripts but relevent to the content of the thesis, a second section containing laboratory protocols in sufficient detail to repeat all experiments, and a third section including nucleotide sequence data referenced in the manuscripts. This thesis was written to partially fulfill the requirements for the Ph.D. degree in two majors: Forest Science and Genetics.

OVERVIEW OF RESEARCH

Many studies of DNA polymorphism in trees have focused on chloroplast genomes. This is a consequence of their repetitive nature, which facilitates RFLP analysis, and unique characteristics compared to nuclear DNA. They typically display uniparental, paternal inheritance in Pinaceous conifers; evolve conservatively in structure and sequence, facilitating phylogenetic analysis; and have been largely unstudied from a population genetic viewpoint prior to the advent of molecular techniques.

Studies to date have largely confirmed that cpDNA is highly conserved in conifers. They have also, however, shown that specific regions of the genome can possess moderate to high levels of variability. The ease of cpDNA analysis, coupled with the discrete nature of polymorphism, provide an excellent system for dissecting the mutational events giving rise to highly polymorphic gene regions and areas of rearrangement.

Douglas-fir and Monterey pine chloroplast genomes, despite their slow rate of sequence evolution and their close relationship as genera of Pinaceae, are distinguished from each other by the presence of a length mutation hotspot in Pseudotsuga, which is absent in Pinus, and a 40-50 kb inversion. In the first chapter, we characterize the

nucleotide sequences of the length mutation hotspots in Douglas-fir and a related species of Pseudotsuga. We discovered this highly variable region during genome mapping and phylogenetic analysis. We decided to investigate its nature by DNA sequencing and the polymerase chain reaction (PCR)--which allows specific DNA regions to be selectively amplified for study. We show that the region possesses a partially duplicated tRNA gene, separated from the intact gene by a long stretch of imperfect, direct repeats which are the locus of polymorphism. In the second chapter, we cloned and sequenced the common endpoints of the 40-50 kb inversion from Monterey pine and Douglas-fir. Here we show the presence of inverted repeats, adjacent tRNA genes, and pseudogenes at inversion borders. We propose a mechanism of homologous recombination between tRNA genes to explain the rearrangement.

LITERATURE REVIEW

THE CHLOROPLAST

STRUCTURE AND FUNCTION

Chloroplasts are subcellular organelles found in all eukaryotic algae and land plants (reviewed in Campbell 1987, Hoober 1984). One function of chloroplasts is to carry out photosynthsis. In this process, light energy is transformed into electrical energy and then into chemical energy through the fixation of carbon dioxide into carbohydrates with the release of oxygen. Contained within the soluble stroma matrix of the organelle is a system of internal thylakoid membranes embedded with the green pigment chlorophyll, which absorbs light and drives the intermolecular transfer of electrons. The resulting ionic gradient that forms across the membrane is used to drive the synthesis of high-energy chemical compounds including ATP and NADPH. Although the primary purpose of chloroplasts is photosynthesis, it is also the site of other biosynthetic pathways including fatty acid synthesis, nitrogen and sulphur assimilation , and amino acid biosynthesis (Sears 1983, Halliwell 1984).

SIZE, SHAPE, AND NUMBER

The size, shape, and number of chloroplasts found in a cell are highly variable. In higher plants, chloroplasts are lens-shaped and measure approximately 1-3 µm by 5-7 µm in dimension (Hoober 1984). Eukaryotic algae exhibit shapes ranging from lobed to lens-shaped and cup-shaped to long and spiraled. Photosynthetic prokaryotes such as *Chlorella* do not possess chloroplast organelles but instead have thylakoids which lie within the cytoplasm (Goodwin and Mercer 1983). The number of chloroplasts per cell vary with plant species and cell type but generally increase with increasing cell size (Hoober 1984). Ho and Rayburn (1991) report a positive

correlation between genome size and chloroplast number. They observed 4.7 to 7 chloroplasts per guard cell in 15 populations of maize. While *Chlamydomonas* has a single, large chloroplast organelle (Birky 1978), reports of chloroplast number in land plants range from 50-300 plastids per cell; young sections of leaves contain fewer chloroplasts than do mature regions (Hoober 1984, Goodwin and Mercer 1983).

EVOLUTIONARY HISTORY

Chloroplasts contain a genetic system that shares many features with prokaryotic organisms. The most widely promoted theory to explain the predominance of prokaryotic features within the chloroplast organelle is the endosymbiont hypothesis. It hypothesizes that the organelles, both chloroplast and mitochondria, evolved from bacteria-like progenitors that were incorporated into nucleated host cells over one billion years ago (reviewed in Margulis 1981, Gray and Doolittle 1982, Gray 1989, Gray 1991). Endosymbiosis was followed by a progressive loss of genetic information to the host nucleus, depriving the endosymbiont of functions required for autonomous existence and increasing its dependence on host nuclear genes (Weeden 1981, Wallace 1982). Chloroplasts appear to be the reduced remnant of a cyanobacterium-like endosymbiont (Gray 1989, Morden and Golden 1989), whereas the closest contemporary to the mitochondrion is purple photosynthetic bacteria (Gray 1989). Since most chloroplast genomes differ little in size (see section: The Chloroplast Genome, Size and Structure) and are twenty to thirty times smaller than their cyanobacterial ancestor, the genome size reduction most likely took place during a brief period shortly after their origin (Palmer 1985a). It is unclear, however, whether the organelles evolved due to single or multiple endosymbiotic events (Gray 1989, Turner et al 1989).

THE CHLOROPLAST GENOME

SIZE AND STRUCTURE

Typical chloroplast DNAs (cpDNAs) of green algae and land plants are single, circular, double-stranded molecules (reviewed in Palmer 1985a, Wallace 1982, Umesono and Ozeki 1987, Palmer 1985b, Crouse et al 1985, Palmer 1991). The chloroplast genome of land plants generally varies between 120-160 kb, with just a few genomes approaching 200 kb (*Nicotiana accuminata*, 171 kb; *Spirodela oligorhiza*, 180 kb (duckweed); *Pelargonium hortorum*, 217 kb (geranium) (Palmer 1985a)). A distinguishing feature found in the genome of almost all land plants and several major lineages of algae is the presence of a 10-76 kb inverted repeat (IR) that divides the molecule into two unequal regions of single copy DNA of approximately 20 and 80 kb. Each inverted repeat contains a complete set of rRNA genes that are organized and transcribed in the same order as the *E. coli* rRNA operon, and tRNA genes in the 16S to 23S spacer region (Wallace 1982). Except for *Pelargonium*, the rRNA genes are always transcribed toward the small single copy region (Palmer 1985a). The inverted repeat varies in its position around the genome, resulting in single copy regions of variable size-- and thus variation in gene content of the repeat.

Much of the cpDNA size variation is due to the shrinking and spreading of the repeat into the single copy regions (Palmer 1985a, Umesono and Ozeki 1987). In some instances, one copy of the repeat has been lost entirely, resulting in correspondingly smaller chloroplast genomes and only one copy of the rRNA genes. This loss has been reported in five separate lineages of vascular plants: conifers (genome size of about 120kb) (Strauss et al 1988, Lidholm et al 1988, White 1990, Raubeson and Jansen 1992), one genus in each of the families Scrophulariaceae and Orobanchaceae of the order Scrophulariales, two genera of Geraniaceae (order

Geraniales), and six tribes and *Wisteria* of the legume family Fabaceae (order Fabales) (Liston 1993, Downie and Palmer 1992, Lavin et al 1990). In conifer cpDNA, the loss of the IR is thought to occur by deletion of one part of a IR copy and then loss of an entire IR segment. Tsudzuki et al (1992) reported the presence of a 495 bp inverted repeat in black pine (*Pinus thunbergii*) that contains the 3' end of *psb*A and the *trn*I-CAU gene, but no rRNA genes. The authors believe this small repeat represents an incomplete loss of the IR (rather than duplication resulting in repetitive gene sequences) because truncated genes like *psb*A in black pine are frequently found at one end of the IR in other chloroplast genomes; *psb*A genes are often observed at IR junctions; and similar gene organization between black pine and tobacco exists downstream of the inverted repeat. This would suggest that the loss of the IR may not be complete in other conifers, or the black pine inverted repeat represents a recent gain of duplicated sequence. The authors do not discuss these implications.

Whereas the chloroplast genome of plants are comparatively uniform in size, shape, and gene arrangement, the cpDNAs of unicellular eukaryotes are more variable, even among closely related species. In green algae, the chloroplast genome ranges in size between 85-2000 kb. *Acetabularia* has a genome equal to a moderately sized bacterial genome (2000 kb) (Palmer 1985a, Wallace 1982), and *Chlamydomonas* sp. contain only 1/27th of the coding capacity of *Acetabularia* in its 190 to 292 kb cpDNAs (Palmer 1985a, Rochaix 1987). Many algal species contain the inverted repeat structure in their genomes. However, in *Euglena gracilis* cpDNA, instead of containing inverted repeats, there are three tandem repeats, each coding for an rRNA operon (Gray and Hallick 1978). In the *Chlamydomonas* inverted repeat, additional 3S and 7S RNA species exists and the 23S rRNA gene contains an intron (Wallace 1982).

Although cpDNA usually exists as single circular molecules, it has also been shown to occur in linear and variously rearranged forms. Bendich and Smith (1990) found long, linear DNAs in the chloroplasts of watermelon and pea. These molecules were one to four times the total size of the circular genome and are likely to be replicative intermediates. Kolodner and Tewari (1975c) reported that between 1.9% and 3.6% of circular DNA molecules exist as dimers and concatenated dimers. In spinach, the chloroplast genome exists as monomers, dimers, trimers, and tetramers, likely formed through intra- and intermolecular recombination (Deng et al 1989). Linear cpDNAs adopting hairpin structures have been found in the chloroplasts of barley cultivars (Collin and Ellis 1991) and albino barley plants derived from anther cultures (Ellis and Day 1986). A continuous size range of molecules was reported by Collin and Ellis (1991) and are thought to be involved in cpDNA replication. Common to almost all species with IR-containing cpDNAs are two equimolar populations of circular molecules that differ in the relative orientation of their single copy sequences (reviewed in Palmer 1985a, b). The two isomers are produced through intramolecular recombination very near, or within, the inverted repeat sequences (Stein et al 1986, Palmer 1991).

GENOME NUMBER

The chloroplast genome, unlike the nuclear genome, is not complexed with histones or other proteins and is, at least occasionally, attached to organelle membranes. cpDNA appears to be localized in discrete regions throughout the stroma called nucleoids (Birky 1978, Palmer 1987, Gillham et al 1991). Each chloroplast contains ten to several hundred copies of the genome, and an entire cell carries 500-1500 copies of the cpDNA molecule, though genome number has been found to vary in response to developmental and environmental cues (Hoober 1984, Umesono and

Ozeki 1987, Bendich 1987, Birky 1978). This results in cells which are highly polyploid in their cpDNA content. Bendich (1987) offers two explanations for the high genome copy number per chloroplast: (1) multiple genome copies ensure daughter cells receive at least one genome at organelle division, and (2) multiple copies enhance gene expression control by providing increased dosages of organellar genes for transcription.

GENE ORDER AND CONTENT

Chloroplast genome restriction endonuclease cleavage site maps are available for over 350 species of land plants (Downie and Palmer 1992), and the genomes of rice (Hiratsuka et al 1989), tobacco (Shinozaki et al 1986), and liverwort (Ohyama et al 1986b) have been completely sequenced and compared (Shimada and Sugiura 1991). Generally, comparisons among land plants have shown a gene order and a gene content that is highly conserved (Palmer 1987, Palmer 1991). The chloroplast genome of tobacco (*Nicotiana tabacum*) is considered to be similar to the genome of the common ancestor of seed plants because it shares its gene order with almost all angiosperms, ferns, and *Ginko* (Downie and Palmer 1992, Palmer 1985a, Palmer and Stein 1986). In genomes where gene order differences are found, these can be derived from one to a few inversions from the ancestral type. More extensive rearrangements are found in a small number of taxa including conifers (Strauss et al 1988) and the families Fabaceae, Geraniaceae, and Campanulaceae, many of which lack the large inverted repeat (Downie and Palmer 1992, Palmer 1985a, Sugiura 1989, Palmer et al 1987a, Palmer and Thompson 1982). Although not essential for chloroplast genome function, the presence of the IR does seem to confer evolutionary stability on the molecule.

Most vascular plants contain 120-130 distinct chloroplast genes, including four rRNA genes, 30 tRNA genes, about 90 protein genes, and dozens of open reading frames (ORFs) (Wolfe and Sharp 1988, Palmer 1991, Ohyama et al 1986a) not yet identified (reviewed in Whitfeld and Bottomley 1983, Umesono and Ozeki 1987, Crouse et al 1985, Palmer 1985a, Sugiura 1989). One half of the gene component is involved in gene expression in the chloroplast. Besides the rRNA and tRNA genes, the chloroplast also encodes RNA polymerase subunits and some ribosomal proteins. Most of the ribosomal proteins used in the chloroplast are encoded by nuclear genes, synthesized in the cytoplasm, and imported into the organelle (Sugiura 1989).

Over 20 genes encode components of the photosynthetic and electron transport machinery. Included are genes coding for subunits of photosystems I and II, H+-ATP synthase, and the cytochrome b/f complex. The large subunit of ribulose bisphosphate carboxylase, the enzyme that fixes CO_2, is encoded by a chloroplast gene (*rbc*L) while the small subunit of the enzyme is encoded for by a nuclear gene and post-translationally imported into the chloroplast (Umesono and Ozeki 1987). A chlororespiration pathway appears to exist in land plants based on the homology of chloroplast *ndh* genes to the mitochondrial respiratory-chain NADH dehydrogenase complex. The *ndh* genes could encode a functional chloroplast NAD(P)H plastoquinone oxidoreductase which is involved in the mechanism of O_2 uptake (Bennoun 1982). It is also possible that these are mitochondrial genes which have been inserted into the chloroplast genome and are, therefore, pseudogenes (Shinozaki et al 1986). The parasitic flowering plant *Epifagus virginiana* (beechdrop) contains no chlorophyll and does not carry out photosynthesis. It has a very small genome (71 kb) that contains typical-sized inverted repeats with a full complement of genes. However, the single copy regions have been greatly reduced in size as a result of ten deletions. Beechdrop cpDNAs have lost most of their photosynthetic protein genes,

ndh genes, open reading frames, RNA polymerase genes, and a few tRNA genes. The chloroplast has retained ribosomal RNA and ribosomal protein genes, indicating the continued expression of products for organelle function (dePamphilis and Palmer 1990, Wolfe et al 1992).

Unlike prokaryotes, many chloroplast genes contain introns. The number of introns in a given genome is highly variable among species (Palmer 1991). Approximately 20 introns have been found in each of the completely sequenced cpDNAs of tobacco, rice, and liverwort (summarized in Palmer 1991). Six of these occur in tRNA genes, five in ribosomal protein genes, five in protein coding genes, one in an RNA polymerase gene, and several in open reading frames. All but one of the introns is shared between at least one of the angiosperms and liverwort. Large introns (300-2500 bp) interrupt several tRNA and protein genes of tobacco and liverwort, and require post-transcriptional RNA splicing for their expression. The gene that codes for the 30S ribosomal protein S12 (*rps*12) in tobacco consists of three exons that requires trans-splicing. Two of the three exons are encoded by one strand of cpDNA while the third exon lies about 30 kb away on the opposite DNA strand (Umesono and Ozeki 1987, Koller and Delius 1984).

GENE EXPRESSION

The organization and sequence of many cpDNA genes are highly homologous to that of bacterial genes. Chloroplast genes are often organized into polycistronic transcription units that resemble bacterial operons; their transcription and translation signals are similar to those of *E. coli.*; and the amino acid coding sequences retain the universal code present in bacteria (reviewed in Gruissem 1989a, Mullet 1988, Zurawski and Clegg 1987). Specific prokaryotic features of chloroplast gene structure include -35 and -10 promoter sequences, initiation codons preceded by Shine-

Dalgarno ribosome-binding sites, overlapping translation stop/start signals, and 3' untranslated regions that form terminator stem-loop structures (Whitfeld and Bottomley 1983, Bonham-Smith and Bourque 1989).

Chloroplasts contain their own genetic system comprised of DNA, RNA, DNA- and RNA-polymerase, ribosomes, and translation factors (Subramanian et al 1991). Many of the components of the protein synthesis machinery are composed of nuclear encoded gene products (Rochaix 1987) that are synthesized in the cytosol and transported into the organelle (Keegstra 1989, Berry-Lowe and Schmidt 1991). One or more of the RNA-polymerase subunits, and some of the ribosomal proteins and translation factors, are encoded by nuclear genes (Rochaix 1987).

Gene expression is controlled at various levels in the chloroplast. Aside from the transcriptional and translational controls (see also Gruissem 1989b, Sugiura 1991), mRNA stability plays a large role in gene expression. Located in the 3' untranslated regions of protein coding genes are inverted repeats. Their role appears to be one of mRNA processing and transcript stabilization (Akada et al 1990, Gruissem 1989b, Bogorad 1991, Delp and Kossel 1991, Marechal-Drouard et al 1991).

Methylation appears to mediate expression of at least some chloroplast genes (reviewed in Cedar (1988), Dynan (1989), and Ohta et al (1991)). Reduced expression of nuclear genes is associated with increased levels of methylation. Chloroplast DNA of higher plants have much lower total levels of methylation than nuclear DNA. However, a significant number of methylated cytosine residues have been found in the cpDNAs of *Chlamydomonas*, *Pisum sativum*, and the mesophyll cells of maize. In the cpDNA of *P. sativum*, Ohta et al (1991) found that 0.5% of cytosine residues were methylated at C-5. Most of the C*C*(A/T)GG sites and some of the GAT*C* sites in *rbc*L, *atp*B, and *psb*DC were partially methylated. Many of these same sites were also methylated in amyloplast DNA-- which would not be expected if methylation of DNA

was suppressing gene expression in the chloroplast. Instead, cpDNA would be expected to be un-methylated while amyloplast DNA would be methylated. Methylation may, therefore, be functioning in other, as yet unknown, mechanisms in chloroplasts.

Chloroplast development and gene expression are also regulated by light and other environmental factors. Receptors to these exogenous signals that control chloroplast development and nuclear and plastid gene expression, especially to light, have been widely studied (reviewed in Tobin and Silverthorne (1985), Klein and Mullet (1990), and Rodermel and Bogorad (1985)). The most widely studied photoreceptor is phytochrome, a chromophorylated polypeptide that undergoes conformational change upon absorption of red light to an active form capable of affecting transcriptional events (Quail 1991). The expression of *cab* and *rbc*S, both nuclear genes whose products are synthesized in the cytoplasm and transported to the chloroplast, are positively regulated by phytochrome in angiosperms (Tobin and Silverthorne 1985). However, *cab* and *rbc*S are expressed in the absence of light in gymnosperms. Transcripts of both genes were found at substantial levels in dark-grown seedlings of Douglas-fir (Alosi and Neale 1992) and *Pinus thunbergii* (Yamamoto et al 1991). Alosi and Neale (1992) also reported change in steady-state mRNA levels upon irradience with red light indicating phytochrome-mediated gene expression in Douglas-fir. Yamamato et al (1991) propose that light is not required for expression of *cab* and *rbc*S while Alosi and Neale (1992) hypothesize substantial constituative mRNA levels in the dark. Chloroplast protein levels in dark-grown conifer seedlings appear to vary among species, although light-independent expression of photosynthetic genes is common to all species examined (*Podocarpus*, *Abies*, *Tsuga*, *Picea*, *Pinus*, *Cryptomeria*, and *Thuja*) (Mukai et al 1992).

REPLICATION

Most of our knowledge of replication enzymology is gained from phages, plasmids, or animal viruses, and very little is known, or has been directly studied, about the replication machinery of chloroplast DNA. Each replication system is comprised of an array of proteins functioning to duplicate the DNA template (reviewed in Kues and Stahl (1989), Stayton et al (1983), Marians (1984), Newlon (1988), and Kornberg and Baker (1992)). Replication of cpDNA takes place randomly throughout the cell cycle, is not coupled to the nuclear S phase, and all essential replication machinery is nuclear-encoded (Sugiura 1989).

Chloroplast DNA from land plants and algae replicate via Cairns-type and rolling circle mechanisms (Kolodner and Tewari 1975a, Cairns 1963, Kornberg and Baker 1992). As studied in corn and pea (Kolodner and Tewari 1975b), replication begins using the Cairns mechanism: two displacement loops form at the origins of replication (ORIs) and expand toward each other forming theta structures (Bogorad 1991). One round of replication is completed approximately 180° from the ORIs. At, or near, the Cairn termination site a rolling circle mechanism is initiated and replication continues by this method.

The physical location on the DNA molecule at which replication begins is termed the origin of replication. A displacement loop (D-loop) can be observed forming at an ORI under electon microscopy. While there are multiple origin of replication sites along eukaryotic chromosomes (Held and Heintz 1992), bacterial DNA usually contains only one. At a replication origin, an initiator protein binds to the site and duplex DNA of adjacent direct repeats open. Direct repeats usually number from two to five and are between 6 and 16 bp long. Plasmids, phage, bacteria, and organelle DNA origins all share a common organization though not necessarily sequence homology (Eckdahl and Anderson 1990). These sequences are highly A+T

rich regions of variable size, and are associated with tandem and inverted repeat sequences that are capable of forming complex stem-loop structures. Intraspecifically, polymorphic regions are common, and are composed of variable numbers of the tandem direct repeats. The direct repeats are found within boundaries of anti-bent domains composed of oligo (dA) tracts that lie in a 6-8 bp periodicity. The site of initiator protein binding lies in an area of DNA bending caused by 3-6 bp oligo (dA) tracts spaced every 10-11 bp (Eckdahl et al 1989, Eckdahl and Anderson 1990, Hsieh et al 1991).

Another common feature of replication origins is the presence of an autonomously replicating element (ARS) 11 bp core consensus sequence (A/TTTTATG/ATTTA/T). ARS elements are DNA sequences that allow the plasmid into which they are inserted to autonomously replicate in yeast. The 11 bp core consensus sequence in the element is essential for replication of the plasmid (Laskey et al 1989, Marunouchi et al 1987), and is the likely site of initiator protein binding (Held and Heintz 1992, Eckdahl and Anderson 1990). Although ARS elements were likely candidates for replication origins (as is the case in *Saccharomyces cerevisiae*) (Huberman et al 1987, Brewer and Fangman 1987), many ARS elements are not used as origins (Vallet and Rochaix 1985).

Chloroplast DNA replication origins have been mapped from several species. *Chlamydomonas* cpDNA contains two origins that map 6.5 kb apart, 10 kb upstream of the 16S rRNA gene (Waddell et al 1984). Pea and corn cpDNA also each contain two replication origins, approximately 7 kb apart, near the rRNA genes (Meeker et al 1988). *Euglena* cpDNA, however, has only one origin that lies near the 16S rRNA gene (Ravel-Chapuis et al 1982). Sequence data of cpDNA replication origins is available from *Euglena*, *Chlamydomonas*, and maize. The *Euglena* origin of replication maps near a polymorphic region that varies in size by about 800 bp (Koller

and Delius 1982). The polymorphic region is composed of a variable number of tandem repeats (87% A+T rich), and is located between two 96 bp inverted repeats, approximately 5 kb from the 16S rRNA gene. Also associated with this region is a pseudo-*trn*W gene, and intact *trn*Y and *trn*E genes (Schlunegger and Stutz 1984). Wu et al (1986) sequenced a *Chlamydomonas reinhardtii* replication origin and found the region to also be A+T rich (70%) and to contain four stem-loop structures, numerous direct repeats, and an ARS consensus sequence. A putative maize replication origin (Gold et al 1987) was shown to be A+T rich, capable of forming stem-loop structures, and to share some sequence homology to the *Chlamydomonas* origin. However, the authors have not yet shown that D-loops can form in this region--a prerequisite for proving that it is a true ORI. Replication in *Oenothera* is initiated at two pairs of D-loops, each pair located 4 kb apart in the large inverted repeat bracketing the 16S rRNA genes. One site is associated with small length mutations (Chiu and Sears 1992). Tobacco origins have also been mapped to the IR, however, they appear singly in each repeat close to the 23S rRNA gene (Takeda et al 1992).

INHERITANCE

Chloroplast DNA is usually inherited from one parent and then segregates vegetatively during mitotic cell divisions resulting in predominantly homoplastic plants (Birky 1988, Birky et al 1989, Birky 1983, Vaughn 1981, Birky 1978). Lemieux et al (1990) state that biparentally inherited *Chlamydomonas* cpDNA requires only 20 mitotic cell divisions to produce homoplasmic cells. In angiosperms chloroplast DNA is inherited maternally in over 66% of the plant genera and biparentally in the remaining 33% (Sears 1980). However, cpDNA has been shown to be paternally inherited in those conifers studied, all but one of which (*Sequoia*) are members of the Pinaceae: *Pseudotsuga* (Neale et al 1986), *Picea* (Stine et al 1989,

Sutton et al 1991, Szmidt et al 1988), *Pinus* (Wagner et al 1987, Wagner et al 1989, Neale and Sederoff 1989), *Sequoia* (Neale et al 1989), and *Larix* (Szmidt et al 1987). Chloroplast transmission in conifers occurs by the paternal plastids entering the egg cell, followed by the degeneration of maternal plastids (Whatley 1982).

Although it is generally assumed that an individual plant is homoplastic for a unique chloroplast DNA molecule, infrequent cpDNA heteroplasmy has been observed in single plants of rice (Moon et al 1987), cotton (Lax et al 1987), alfalfa (Johnson and Palmer 1989), and some conifers (Wagner et al 1987, Szmidt et al 1987, Govindaraju et al 1988, Govindaraju et al 1989). In the conifers, maternal leakage may provide an important source of genetic variability. Recombination between maternal and paternal cpDNAs, during at least a temporarily heteroplasmic condition, would result in even greater levels of polymorphism. Two types of heteroplasmy are commonly detected: (1) restriction site mutations and (2) insertions and deletions in a hotspot region (Soltis, Soltis, and Milligan 1992). In contrast to the relatively rare heteroplasmy observed in cpDNA, the mitochondrial DNA of animals is highly heteroplastic, showing nucleotide site changes, large insertions/deletions, and hotspot heteroplasmy (Solignac et al 1987).

CHLOROPLAST GENOME EVOLUTION

Although the chloroplast has been evolving for over 400 million years, it is conserved relative to the plant mitochondrial and nuclear genomes. This conservation may be the result of several factors. First, cpDNA structural mutations are rarely polymorphic in populations (Birky 1988, Palmer 1985b). Because most of the chloroplast genome is transcribed, there is little spacer sequence available that could tolerate rearrangements, and rearrangements would, therefore, tend to disrupt gene function and be detrimental to the individual (Palmer and Stein 1986, Birky 1988,

Palmer 1987). Second, there are apparent structural and recombinational constraints imposed by the large inverted repeat (IR) which appears to confer stability on the molecule (discussed in section: The Chloroplast Genome, Gene Order and Content). However, this does not explain why some species, even some of those containing the IR, tolerate many cpDNA rearrangements (Palmer and Stein 1986, Palmer and Herbon 1988, Strauss et al 1988, Palmer and Thompson 1982, Palmer 1983, Downie and Palmer 1992). Third, there is a general lack of dispersed repeated structures throughout the chloroplast genome. Repeated sequences could serve as substrates for recombination and exist in many highly rearranged genomes (Palmer and Stein 1986, Palmer 1987). Last, it is possible that chloroplast structure is conserved because the mechanisms of structural evolution (transposition, transfer of sequence (particularly importation), and recombination) are considered to be rare (Birky 1988). Although this appears to be true for transposition and transfer, there are many cases that indicate recombination does occur in chloroplast genomes. The role of recombination and transposition in genome evolution will be discussed following a summary of recombination mechanisms.

RECOMBINATION MECHANISMS

A number of recombination mechanisms appear to be active in cpDNA. Homologous recombination is a frequent process, including intra- and intermolecular recombination involving both reciprocal and non-reciprocal exchange (Lemieux et al 1990). In homologous recombination (reviewed in Kornberg and Baker 1992, Radding 1982, Smith 1988a, Lonsdale et al 1988, Singer 1988, Fincham and Oliver 1989) strand exchange between homologous segments initiates at any point where the segments exceeds a minimum length. The process is enzymatically driven and an accepted pathway for the event begins with a 3'-OH end of single-stranded DNA

invading a second double-stranded DNA molecule forming a Holliday junction and being processed into one or two recombinant types (Figure 1).

Non-reciprocal recombination, or gene conversion, can be either an intra- or intermolecular event and is the most common form of homologous recombination in the chloroplast genome (Newman et al 1992). Fincham and Oliver (1989) state that gene conversion is the "unilateral transfer of information from one allele to the other," and is often "accompanied by nearby crossing over." A heteroduplex is formed and then resolved, possibly by mismatch repair (Vedel and Delseny 1987, Lemieux et al 1990). Gene conversion is the mechanism responsible for preserving sequence identity between IRs in the chloroplast genome. When gene conversion within the IR is intramolecular, head-to-head dimers are formed (Palmer 1985a) which leads to copy correction of the IR (Lemieux et al 1990).

The IR also participates in reciprocal recombination events (equal and unequal cross-overs) (Bowman et al 1988). Intramolecular reciprocal recombination within the IR results in an equimolar ratio of chloroplast molecules that exist that differ only in the relative orientation of their single copy regions (Palmer 1985a). Reciprocal intramolecular crossing over can also result in deletions and inversions, while intermolecular unequal crossing over yields deletions and insertions (Bowman et al 1988). Unequal crossing over, according to Vedel and Delseny (1987), "occurs when two arrays of tandem repeats pair and recombine out of phase" resulting in "recombination between repeats and variation in copy number." The biolistic transformation of *Chlamydomonas* occurs by an intermolecular homologous recombination mechanism. Persistent heteroplasmic individuals for one of two types of RFLPs could not be found suggesting the operation of an efficient copy correction mechanism and rapid segregation of the genomes (Newman et al 1990). Unequal crossing over and gene conversion are caused by homologous recombination between

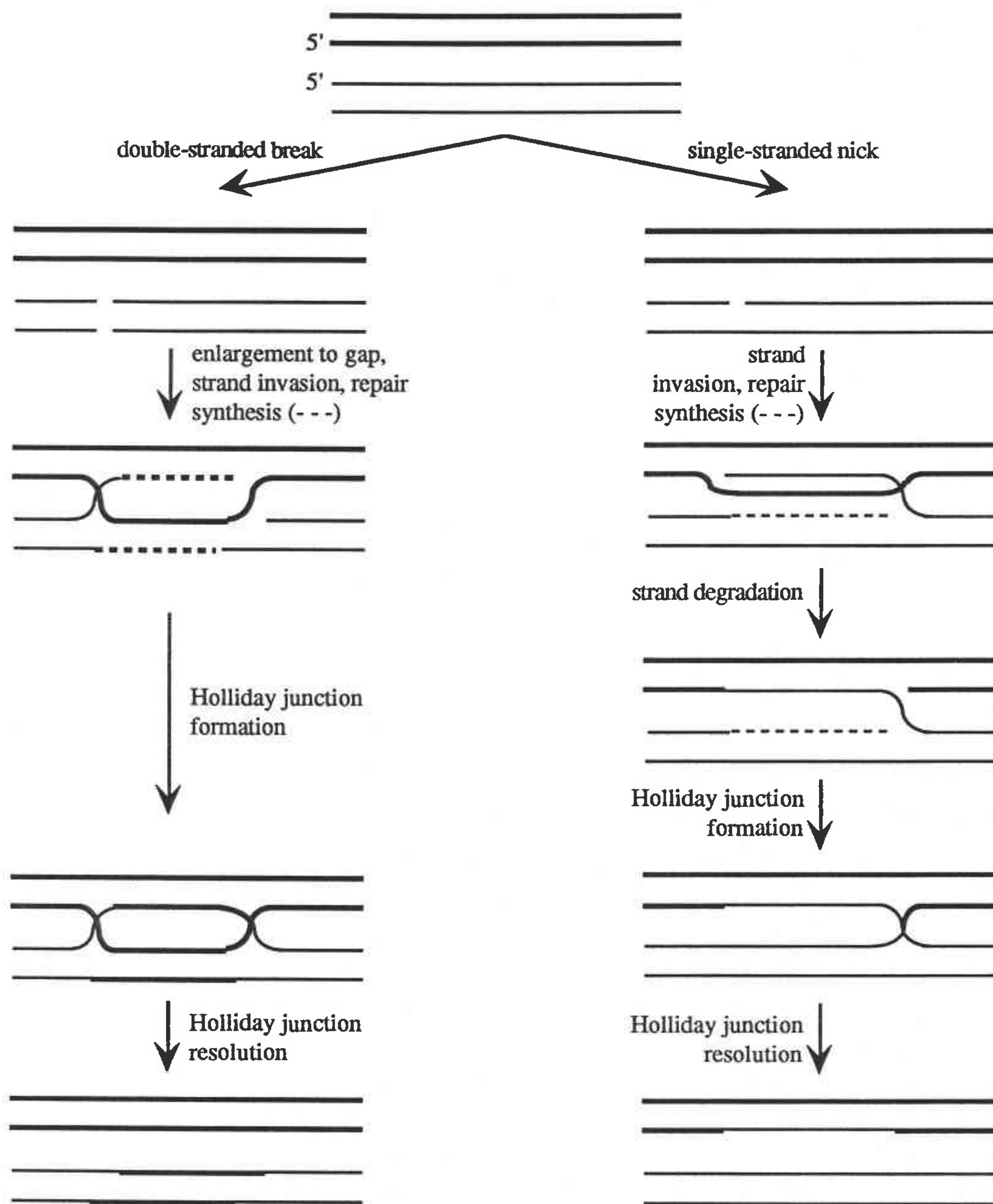


Figure 1. Models for homologous recombination: (left) double-stranded break model; (right) single-stranded break model. Modified from Fincham and Oliver (1989).

repeated sequences (Bowman 1988) that can cause deletions, duplications, inversions, and translocations (Singer 1988). If recombination is responsible for a given structural change, such as an inversion or deletion, there should be homologous sequences at their endpoints (Birky 1988).

Length mutations are also due to illegitimate recombination (defined as events that are not homologous, site-specific, or transpositional (Kornberg and Baker 1992)) and slipped-strand mispairing between repeats (Ogihara et al 1992). Slipped-strand mispairing is a mutational process due to replication error. In the accepted model, "deletions result from slippage at the 3' end of the growing DNA strand and failure to copy part of the DNA template" (Weston-Hafer and Berg 1989). Slippage could occur during DNA replication or repair (Walker et al 1985) and cause the initial formation of short repeated sequences, and also larger insertions and deletions (Weston-Hafer and Berg 1989). These sequences may expand further by unequal crossing-over because of their propensity to mispair (Levinson and Gutman 1987).

Transposition is the process by which a specific DNA sequence moves to a new location in the chromosome (reviewed in Grindley and Reed 1985, Fedoroff 1983, Kornberg and Baker 1992), by three general mechanisms: (1) replicative transposition, (2) nonreplicative transposition, and (3) retrotransposition. If transposition is occurring in cpDNA, it may be happening through a replicative mechanism where the DNA is duplicated and the copy is inserted at its new location via cointegrate formation (due to the presence of multiple copies of transposon-like sequences in some cpDNA molecules).

GENOME REARRANGEMENTS

Recombination

The major cause of chromosomal rearrangements in nuclear genes is the crossing over between repeated sequences, or recombination. Plant mtDNA experiences a high frequency of intra- and intermolecular recombination, and recombination with mitochondrial plasmids (Birky 1988). Recombination is much more limited in the chloroplast genome. In a study of cpDNA recombination in the offspring from crosses of different chloroplast mutants of *Oenothera* (evening primrose), it was concluded that intermolecular recombination was a rare event. Out of over 7500 progeny examined, no wild-type recombinant was recovered (Chiu and Sears 1985). Intermolecular recombination between markers has been observed in *Chlamydomonas* (Harris 1989) along with intramolecular gene conversion between the large inverted repeats (Gillham 1978, Rochaix 1987). *Chlamydomonas* cpDNA recombines at a relatively high frequency though recombination occurs at a 5-fold higher frequency in the large inverted repeat sequences compared to the single copy regions (Lemieux et al 1990). cpDNA recombination between IRs of a single molecule has been shown in *Chlamydomonas* and higher plants (Aldrich et al 1985), and between the direct ribosomal RNA operon-containing repeats in *Euglena gracilis* (Bowman and Dyer 1986).

Length mutations result through either inter- or intramolecular unequal recombination (Nicolas et al 1985). Equimolar isomers of cpDNA single copy regions likely exist due to intramolecular recombination across the large inverted repeat, as documented in bean (*Phaseolus vulgaris*) (Palmer 1983) and *Chlamydomonas reinhardii* (Aldrich et al 1985, Rochaix 1987). Other lines of evidence for recombination in the chloroplast genome are the formation of circular dimers possibly formed by intermolecular recombination between the large inverted repeats (Kolodner

and Tewari 1979), and the many taxa with rearranged genomes previously discussed. Overall, intermolecular recombination appears to be rare in chloroplast DNA of angiosperms (Birky 1988), with an intermolecular recombination frequency of less than 0.05% (no detectable recombinant cpDNAs among over 7500 progeny of crosses of *Oenothera* (Chiu and Sears 1985).

Inversions

Chloroplast genomes contain structural rearrangements (inversions, translocations, and insertion/deletions) that occur non-randomly throughout the genome (Palmer 1985a). Inversions are the most common form of rearrangement. The tobacco chloroplast genome (which is colinear to that of spinach (*Spinacia*), fern (*Osmunda*), and *Ginko*) is considered to be ancestral to other types (Downie and Palmer 1992). Altered genomes usually differ from tobacco cpDNA by only one or two large inversions (Palmer 1987, Downie and Palmer 1992). This is true of lettuce (*Lactuca sativa*), mungbean (*Vigna radiata*), *Oenothera* (Palmer 1985a), liverwort (*Marchantia polymorpha*), and moss (*Physcomitrella patens*) (Downie and Palmer 1992, Palmer 1990). The chloroplast genomes of the grasses (ie. wheat (*Triticum*), maize (*Zea*), and rice (*Oryza*)) are colinear to each other and differ from the ancestral-type genome by three inversions (Sugiura 1989, Downie and Palmer 1992, Palmer 1990, Palmer and Thompson 1982). Several species exhibit genomes with extensive rearrangements including *Pelargonium* (which has the largest known IR) (Palmer et al 1987b), and taxa from conifers, Fabaceae, Campanulaceae, and Lobeliaceae (Downie and Palmer 1992, Palmer 1985a). The loss of the IR has been associated with highly rearranged chloroplast genomes as is the case in conifers and many legumes (most notably subclover (*Trifolium subterraneum*), pea (*Pisum*), and broad bean (*Vicia faba*)). However, two legumes, alfalfa (*Medicago sativum*) and *Wisteria*, though

having lost the IR in their cpDNA, are not rearranged (Palmer 1985a, Palmer et al 1987b). Palmer et al (1987b) hypothesize that the rearrangements observed in many legumes were preceded by the loss of the inverted repeat. Also, highly rearranged species can be found that still possess the cpDNA IR (*Pelargonium*, soybean, *Oenothera*, and grasses) (Sugiura 1989, Sears 1983, Birky 1988, Palmer et al 1987b). Some highly rearranged genomes contain large dispersed repeat families (several 100 bps in length) such as geranium (Palmer et al 1987a), subclover (Birky 1988, Palmer 1987, Milligan et al 1989), and Douglas-fir (Tsai and Strauss 1989). However, the chloroplast genomes of pea and broad bean do not contain a detectable repeat family and are still highly rearranged (Palmer 1985a, Palmer et al 1987b). None of these characterized rearrangements disrupt genes that are polycistronically linked (Palmer 1987).

The chloroplast DNA inversions of many species have been investigated in more detail, often at the nucleotide sequence level. Sunflower (*Helianthus annus*) cpDNA contains a 23.5 kb inversion relative to tobacco, that lies in the large single copy region of its genome. The inversion borders were sequenced and the location of the endpoints determined. One end of the inversion lies between *trn*E and *trn*T, and the other inversion end between *trn*S and *trn*G (Heyraud et al 1987). Two examples of rare intraspecific inversions were examined, one in pea (*Pisum*) and the other in *Salix*. A population of *Pisum humile* was found by Palmer et al (1985) to differ from all other *Pisum* accessions by a small 2.2-5.2 kb inversion. Soltis et al (1992) describe *Salix melanopsis* populations that differ from others in the species by a small inversion in the range of 1.5-6.5 kb. The two *Pisum* inversion borders mapped close to *atp*A on one end and *pet*A on the other. One inversion endpoint also is associated with a 50-1200 bp insertion/deletion hotspot in the *Pisum* genome. The sequence near *atp*A serves as an inversion endpoint in other species of legumes. A 22 kb inversion shared

by 57 genera in the Asteraceae (the sunflower family) border *atp*A on one side and *rpo*B on the other (Jansen and Palmer 1987). Douglas-fir (*Pseudotsuga menziesii*) and Monterey pine (*Pinus radiata*) differ by a 40-50 kb inversion, one end of which lies near *atp*A, the other end near *trn*G-UCC (Strauss et al 1988). The grass family (Poaceae) possess three inversions relative to the ancestral gene order, the largest being 28 kb (Doyle et al 1992). One end of this inversion lies between *trn*G-UCC and *trn*R-UCU (near *atp*A) and the other end between *trn*fM and *rps*14. A smaller ~6 kb inversion, which presumably occurred after the first larger inversion, spans the *trn*G-UCC border (between *trn*S-GCU and *trn*G-UCC) and the other end lies near *psb*DC. The smallest inversion (2.8 kb) has endpoints between *trn*E and *trn*T (as in sunflower) and again near *trn*G-UCC, effectively inverting *trn*T. Both mungbean and *Oenothera* contain a 50 kb inversion and share one border with the grass 28 kb inversion (Howe 1985). The 50 kb inversion is specific to legumes (Palmer and Thompson 1982).

The inversions of wheat have been characterized in the most detail relative to other species. As a member of the grass family, wheat contains the three previously mentioned inversions relative to an ancestral type genome (Sugiura 1989) (though the largest inversion in wheat is usually reported as being 20 kb instead of 28 kb). The endpoints of the two largest wheat inversions have been sequenced (Quigley and Weil 1985, Howe 1985, Howe et al 1988). The wheat 20 kb inversion borders contain a 70 bp repeat that is 85% similar to each other. Each repeat contains a sequence homologous to the 15 bp bacteriophage lambda *att*-site (Howe 1985), which is where lambda precisely integrates into the *E. coli* chromosome (Landy and Ross 1977). Howe (1985) demonstrated that lambda will integrate at this cpDNA *att* site *in vitro*. Also contained within the 70 bp repeat is a pseudo-tRNA gene. This chimeric gene is comprised of the 3' sequence from *trn*G-UCC and the 5' sequence from *trn*fM. A

tRNA pseudogene is also found in maize cpDNA (Rodermel et al 1987). The two smaller inversions also contain repeats at their borders, though they are much smaller (<20 bp). Common to endpoints of all three wheat inversions is the presence of (1) repeated sequences, (2) at least one adjacent tRNA gene, and (3) partially duplicated tRNA genes (Howe et al 1988). tRNA genes at inversion borders have also been found in *Marchantia* (Ohyama et al 1986b), lettuce (Jansen and Palmer 1987), and rice (the grass family) (Hiratsuka et al 1989), though not all inversion borders are near tRNA genes (Howe et al 1988).

Several possible mechanisms have been proposed to explain the rearrangements seen in the chloroplast genome of wheat (and other grasses). Howe (1985) states that the 70 bp repeats at the large inversion endpoints may have mediated the inversion via homologous recombination. The recombination event may have led to the duplication of the pseudo-tRNA gene and may have involved gene conversion. The repeats are now in direct orientation, but this is caused by a second inversion which overlaps one end of the first. Later, Howe et al (1988) question that homologous recombination between repeats is solely responsible for the rearrangements; they hypothesize that the repeats at the borders of the two smaller inversions may be too short to act as substrates. Instead they raise the possibility that tRNA genes are involved in the rearrangements via either homologous or non-homologous mechanisms. Hiratsuka et al (1989) expands on this to explain the rearrangements seen in the rice chloroplast genome. They hypothesize that intermolecular recombination between a 14 bp homologous region of two different tRNA genes gave rise to the chimeric tRNA pseudogene and the resulting inversion. Illegitimate intermolecular recombination between tRNA genes can account for the origin of pseudogenes, inversions, and the creation of repeated sequences near the inversion endpoints (Sugiura 1989).

Length Mutations

Length mutations are responsible for much of the observed variability in chloroplast genome complexity. Besides the large change in genome size caused by contraction, expansion, or loss of the large inverted repeat, small insertions/deletions of 1-10 bp in length occur frequently in non-coding regions of the genome. These mutations are often close to, or are flanked by, very short direct repeats and may be the result of slip-strand mispairing during DNA replication and repair (Palmer 1985a). Larger length mutations (50-1200 bp) are also detectable but occur much less frequently and are the result of such recombinational mechanisms as intermolecular unequal crossing-over and intramolecular recombination between short direct repeats (Palmer 1991). In a study of chloroplast DNA restriction fragment length polymorphisms in *Pseudotsuga menziesii*, *Sequoia sempervirens*, *Calocedrus decurrens* and *Pinus taeda*, Ali et al (1991) found most mutations were small insertions or deletions. Often these will cluster in hotspots at the ends of the large single copy region (Palmer 1987, Palmer 1985a). Although very rare, Palmer (1987) also cites examples of length mutations thousands of kilobases long. One of these is a 7-9 kb addition in *Nicotiana acuminata*, and the other is a 13 kb length mutation in a group of *Linum* sp.

Hotspots

Recombination hotspots have been detected in both prokaryotic and eukaryotic DNA. In prokaryotic DNA, two tertiary replication origins of phage T4 are located at hotspots in the genome (Kornberg and Baker 1992, Yap and Kreuzer 1991). The origins appear to be necessary for the generation of a hotspot. Deletion of an origin eliminates the hotspot and the insertion of an origin into a recombinational coldspot increases recombination in the area (Yap and Kreuzer 1991). Recombination hotspots also include the *E. coli* Chi sites (the site that is nicked by the *Rec*BCD enzyme during

homologous recombination) (Smith 1988b), and the phage lambda excision-integration system (site-specific recombination) (Landy and Ross 1977). DNA that adopt the Z-conformation are hotspots for spontaneous deletions in *E. coli* plasmids. Z-DNA is formed in areas of $(CG)_n$ (ie. GC and CA/GT) (Freund et al 1989). Most cases of discrete length variation and heteroplasmy in animal mitochondrial DNA result from variation in the copy number of short, tandemly repeated sequences (Moritz et al 1987). mtDNA size variation in crickets (*Gryllus firmus*) is due to differences in the number of tandemly repeated 220 bp sequences (Rand and Harrison 1989). The locus of polymorphism is also associated with a replication origin. Variation likely arose through a replication slippage mechanism or intramolecular recombination (Moritz et al 1987, Rand and Harrison 1989). Eight species of *Drosophila* contain a length mutation hotspot in their mtDNA that varies in size by several 1000 bp. Length variation is due to different number of copies of a 470 bp A+T rich repeat. The presence of heteroplasmic individuals suggests a high mutation rate (Birky 1989). In maize nuclear DNA, the A1 locus shows a two orders of magnitude greater recombination rate that for the genome as a whole, supporting the idea that the genome contains areas of high recombinogenic activity (Brown and Sundareson 1991). Repeated sequences have been shown to act as hotspots that undergo both inter- and intramolecular recombination (Newman et al 1992). The hotspot sequences could be serving as recognition sites or increasing availability of the region to the recombinational machinery (Yap and Kreuzer 1991).

Hotspots found in chloroplast DNA are produced by frequent insertions or deletion (Birky 1989). Length mutations of 50-1200 bp are infrequently observed in cpDNA, but when present tend to cluster in hotspots that are usually located at the two ends of the single copy region (Palmer 1985a). Kung et al (1982) observed a cpDNA hotspot in *Nicotiana*, showing a frequent loss and gain of restriction sites, and

deletions/insertions. It was mapped to the large single copy region, near the right hand border of the large inverted repeat (Tassopulu and Kung 1984). Among seven species of *Nicotiana*, deletions of 0.5-11.0 kb were observed (Tassopulu and Kung 1984). A hotspot of length variability was also detected close to one end of the large single copy region in alfalfa and petunia. This region was located between the *trn*H and *psb*A genes, flanked by short, almost identical AT-rich direct repeats (Aldrich et al 1988a). Chloroplast biolistic transformation exchange events in *Chlamydomonas* occurred in a preferential A+T rich 700 bp region at the 3' end of *psb*A that contained 18 to 37 bp direct repeats. Exchange between the chloroplast and donor plasmid occurred by homologous recombination (Newman et al 1992).

Hotspots are usually found at the replication origins of animal mtDNA (Birky et al 1989) and also of cpDNA. The cpDNA replication origin of *Euglena gracilis* maps near a length hypervariable region that is 80% A+T rich. The polymorphic region is located between two short inverted repeats and varies by 800 bp in length between different *E. gracilis* strains. Length variability is attributed to a probable difference in number of direct repeats (Schlunegger et al 1983).

Wheat (*Triticum*) and *Aegilops* cpDNA contains a length mutation hotspot between *rbc*L and *pet*A genes (Ogihara and Tsunewaki 1988, Birky 1989). Ogihara et al (1988), sequenced the hotspot region in *Triticum aestivum*, *Aegilops crassa*, and *Aegilops squarrosa*, and found short, direct repeats that border deletion endpoints. The size difference observed between the largest and smallest variant is 1.1 kb. It is likely that intramolecular recombination is occurring between the short direct repeats, resulting in deletions (Birky 1989, Ogihara et al 1988). Also associated with the hotspot is the non-reciprocal translocation of a *rpl*23 pseudogene flanked by short repeated sequences into the region (Ogihara et al 1991), A+T rich sequences (80%) (Ogihara et al 1992), and many small dispersed direct and inverted sequences (Ogihara

et al 1988) at mutation points (Ogihara et al 1992). A+T rich regions are often associated with genome instability caused by hotspots (Hyrien et al 1987). The majority of the length variation observed in wheat and grass cpDNA involve only several base pairs and is therefore likely due to DNA replication slippage. Illegitimate recombination mediated by short repeated sequences is probably responsible for the translocated *rpl*23 pseudogene into the area (Ogihara et al 1992).

vom Stein and Hachtel (1988) sequenced a variable cpDNA region of *Oenothera odorata* and *O. berteriana*, that lies in the large single copy region upstream of *trn*L-UAA in the spacer between *trn*T-UGU. The regions differ by one 136 bp insertion/deletion flanked by a short inverted repeat containing a sequence similar to the lambda chromosomal attachment site, and one single base pair insertion/deletion. Blasko et al (1988) also characterized a hotspot in the chloroplast genome of *Oenothera hookeri*. The hotspot mapped the the large inverted repeat and variability was due to a difference in copy number (2, 4, or 7 copies) of a 24 bp tandem direct repeat. The authors state that recombination between imprecisely aligned inverted repeats and subsequent copy correction of the hotspot region would result in the addition or deletion of a 24 bp sequence. *Oenothera* cpDNA hotspots were also detected in the intergenic spacers separating *rpl*16, *rpl*14, and *rps*8 (Wolfson et al 1991). Observed features include a 29 bp tandem duplication and two variable length poly-A stretches. Replication slippage was proposed as the mechanism responsible for the length mutations.

Four different regions in the chloroplast genome of pea (*Pisum*) were shown to contain length mutations by Palmer et al (1985). Although no region was sequenced, two of the regions appear as hotspots and size differences are attributed to small deletion/insertions. One hotspot is located at an inversion border in *Pisum humile*. Hypervariable regions were also detected in *Medicago* accessions. Johnson and

Palmer (1989) examined nine accessions and found the size of the hotspot region to vary by about 350 bp.

Conifer cpDNA has also been shown to contain hotspots of length variability. Two polymorphic regions exist in *Pinus contorta* cpDNA (Lidholm and Gustafsson 1991a) that are characterized by variable numbers of tandem repeats. One region contains a 124 bp repeat unit and individuals vary in size by ~250 bp. The other region contains a larger repeat unit (150 bp) and the size of the area varies by ~450 bp. These areas are located near a *psb*A duplication in the genome. A highly polymorphic region was located in the chloroplast genome of Douglas-fir (Ali et al 1991). Ali et al (1991) surveyed 24 individuals and found seven size classes for this region that varied by 280 bp. *Sequoia sempervirens* (coast redwood) appears to possess this same polymorphic region in its genome.

Transposition

Transposition plays a large role in structural evolution of some genomes. Transposable elements can cause structural rearrangements including deletions, inversions, and the cointegration and resolution of circular molecules (Birky 1988, Sears 1983). However, there is little evidence for transposition occurring in the chloroplast genome, unlike that for plasmids and fungal mitochondrial DNA (Palmer 1987, Palmer 1991), and no active transposon has been isolated from any chloroplast genome (Palmer 1991).

DNA sequence evidence for cpDNA transposons are limited, however, several workers have speculated on their roles in cpDNA rearrangements. Near the endpoints of a 30 kb inversion in liverwort relative of tobacco, there are 8 bp direct repeats that are flanked by 10 bp indirect repeats (Zhou et al 1988). The authors postulate that the "repeats may result from a transposon-mediated insertion which would have facilitated

the subsequent inversion." In the chloroplast genome of the conifer Douglas-fir, a dispersed repeat family of several 100 bp also is associated with a transposon-like footprint sequence, and is postulated to having been created and spread through transposition. The repeat is associated with the endpoints of a 40-50 kb inversion relative to *Pinus radiata*, which was likely mediated by recombination between members of the repeat family (Tsai and Strauss 1989). The highly rearranged chloroplast genome of subclover also contains a dispersed repeat family of several hundred base pairs. It is possible these repeats are transposable elements that became active in an ancestral genome and caused the subsequent rearrangements either while transposing or by providing the repeated sequences necessary to undergo recombination (Birky 1988, Milligan et al 1989).

REPEATED SEQUENCES

A general feature of DNA rearrangements are their association with repeated sequences. Repeated sequences play an important role in both non-homologous and homologous recombination mechanisms. Reciprocal and non-reciprocal homologous recombination acts on repeated sequences in nuclear and mitochondrial DNA, and is important for the evolution of these genomes. Through non-homologous recombination mechanisms, repeated sequences are an important part of transposable elements and insertion sequences that are known to mediate DNA rearrangements (reviewed in Bowman and Dyer 1986). *E. coli* DNA undergoes spontaneous deletions, ranging from 9-123 bp, that occur between short 5-8 bp repeats, and 700-1000 bp deletions between repeated sequences up to 17 bp in size. Deletions of this kind have also been demonstrated also in phage and human DNA and may be due to a slipped-strand mispairing mechanism (Albertini et al 1982).

Higher plant nuclear DNA (like other eukaryotic genomes) contains a high percentage of repeated sequences that occur as dispersed repeat families, multigene families, and long blocks of tandem 50-350 bp repeats of satellite DNA (Vedel and Delseny 1987, Ohtsubo et al 1991). Repeats are often G+C rich (65%) (Ohtsubo et al 1991) and, generally, the larger the genome, the more repeated DNA (up to 75%) (Vedel and Delseny 1987). Dispersed repeated sequences are frequently interspersed with unique sequences and tandemly repeated sequences. Repeats evolve by a variety of mechanisms including amplification, deletion, concerted evolution, and transposition (especially dispersed repeats) (Vedel and Delseny 1987). In various cultivars of rice, Ohtsubo et al (1991) report the presence of 540-6000 copies of a tandem 360 bp repeat. Unequal crossing-over between repeats expanded the area and gene conversion then took place within the repeated sequences. Jarman and Wells (1989) call the formation of tandem repeats a two-step process. Beginning with random DNA, there is a chance duplication. The duplicated DNA then misaligns and undergoes unequal crossing-over which expands and forms the tandem array.

Animal mitochondrial DNA also contains clusters of tandem repeats. Tandemly repeated sequences vary intraspecifically in length in scallops (Bivalvia:Pectinidae) and are contained in the genomes of many organisms including insects, reptiles, mammals, fish, and birds. Repeat units vary in size from less than 100 bp to more than 3000 bp, and length variations up to 10,000 bp have been observed (Gjetvaj et al 1992). The mitochondrial DNA of plants, however, contains many short dispersed repeats. These appear to serve as sites of homologous recombination (Palmer 1990) and mediate intragenomic reciprocal recombination that leads to subgenomic circular molecules from the master DNA circle (Vedel and Delseny 1987). The plant mtDNA genome is likely tolerant of the high frequency of structural rearrangements observed because large intergenic spacers exist that can

accept mutations (unlike cpDNA) and mtDNA does not contain a structural element like the cpDNA IR that generally prohibits inversions in the chloroplast genome (Palmer 1990).

Generally, there are few repeated sequences in chloroplast DNA. When they are present, they occur in high frequency and are associated with recombinational events (Palmer 1985a). The highly rearranged genomes of subclover and geranium contain dispersed repeat families. Geranium cpDNA has two short dispersed repeat families made up of eight and nine members, respectively. It is thought that recombination between the repeats may cause inversions (Palmer et al 1987b). At least six copies of a dispersed 0.5 to 1 kb repeat family exists in the subclover (*Trifolium subterraneum*) chloroplast genome, several of which are associated with genome rearrangements. The repeat family may be a transposable element (Milligan et al 1989). The dispersed repeat family of *Chlamydomonas* is scattered throughout the genome and consists of 24 to 40 copies of 100-300 bp inverted repeats (Bowman and Dyer 1986) that may be involved in illegitimate intramolecular recombination (Newman et al 1992). Wheat cpDNA, unlike the genomes of subclover and *Chlamydomonas*, contains 12 different sets of small repeats non-randomly distributed throughout the genome. Thirty-two sequences made up the 12 unrelated families, but no repeat was detected as multiple copies though many were repeated three times (Bowman and Dyer 1986). Eight of the 12 repeat families have a member in one copy of the IR (Bowman and Dyer 1986) and two of the 12 code for functional and nonfunctional ribosomal protein genes *rpl*2 and *rpl*23 and are located at *trn*H (one end of the large single copy region) and at the length mutation hotspot near *rbc*L (Bowman et al 1988). Dispersed repeated sequences have also been identified in the cpDNA of a conifer. Tsai and Strauss (1989) found six families that clustered in four regions of the genome. Members of one family have short transposon-like sequences and are

associated with inversion endpoints (Tsai and Strauss 1989) and a hotspot of length variation (Ali et al 1991). Although not of a dispersed repeat nature, *Pinus contorta* and *P. banksiana* cpDNA does contain variable numbers of tandem repeats of 124 and 150 bp in size. These repeats map to a polymorphic region where the *psb*A gene has been duplicated along with other duplications and rearrangements in the *trn*K-*psb*A region. It is also possible that illegitimate recombination is occurring in this area between very small (3, 4, 7, and 9 bp) direct repeats (Lidholm and Gustafsson 1991b). Other repeat families found in cpDNA include (1) the large inverted repeat present in most genomes (Palmer 1985a), (2) a five member, 10 kb tandem repeat in *Acetabularia* (Palmer 1985a), and (3) the ribosomal RNA operon containing tandem direct repeats (between 1 and 4 copies) in *Euglena* that undergo either inter- or intramolecular unequal crossing over (Nicolas et al 1985).

Another class of repeats in cpDNA are short repeats less than 100 bp in size (Palmer 1985a). Seventy base pair repeats are associated with inversion endpoints in wheat (Howe et al 1988) and rice (Shimada and Sugiura 1989). The repeats flank a *trn*M gene, and one repeat lies between the *trn*M gene and the second exon and partial intron of *trn*G (Howe 1985). Recombination between small direct repeats at the junction of the large inverted repeat and large single copy region may be responsible for the shrinking of the inverted repeat in petunia relative to the spinach chloroplast genome by causing deletions between the short repeats. Repeated sequences are located in the IR and junction, causing IRs to contract following recombination. Spreading of the IR involves recombination between short inverted repeats at the IR border and junction which pair, recombine, and undergo copy correction, thus allowing the large single copy sequence to become part of the IR sequence (Aldrich et al 1988b). In pea cpDNA, the IR has been deleted, yet a 200 bp region containing duplications of parts of *psb*A and *rbc*L remains. The same sort of short duplications at

rearrangement endpoints have been observed in wheat and *Vicia faba*, but the duplications involve sequences at or within tRNA genes (Wolfe 1988).

tRNA genes are involved in many DNA rearrangements. In eukaryotic DNA, tRNA genes are associated with repetitive elements (Hofmann et al 1991) and may also act as landmarks for the integration of transposable elements in lower eukaryotes (Marschalek et al 1989). tRNA genes have also been shown to be involved in rearrangements of animal mtDNA. These genes are typically dispersed throughout the genome and the location changes frequently (although in sea urchins, tRNA genes cluster near the replication origin) (Birky 1989). Knowledge of the role of tRNA in plant mtDNA is limited. In the mitochondrial DNA of wheat, the *trn*P gene has been duplicated. Multiple intragenomic, site-specific rearrangements have occurred in the area of the repeated gene followed by amplification, fixation, and sequence divergence (Joyce et al 1988). In chloroplast DNA of higher plants, tRNA genes are dispersed throughout the genome (Quigley and Weil 1985) and are associated not only with contraction and expansion of the IR, but also with inversion endpoints (as discussed in section: Chloroplast Genome Evolution, Inversions). Commonly associated with these rearrangements are tRNA pseudogenes which are probably the result of intermolecular recombination between tRNA genes (Marechal-Drouard et al 1991). In the rearranged chloroplast genome of rice, eight tRNA or ribosomal protein pseudogenes are present in the large single copy region, six of which cluster near inversion endpoints. These pseudogenes also exist in wheat, whereas tobacco cpDNA only has one pseudogene (Shimada and Sugiura 1989). Primary or secondary structural similarity between tRNA genes may serve as substrates for recombination during replication via strand matching or breakage and ligation (Rand and Harrison 1989).

CPDNA SEQUENCE EVOLUTION

Chloroplast genes generally evolve more slowly than mammalian mitochondrial and nuclear genes, and more quickly than plant mtDNA (Birky 1988, Palmer 1987, Curtis and Clegg 1984, Banks and Birky 1985). Silent substitution rates in protein genes are lowest in plant mtDNA, 4-fold higher in cpDNA, and 4-fold higher still in the plant and mammalian nucleus (Palmer 1991, Birky 1989). Substitution rates differ among different chloroplast genes (Palmer et al 1988) and there is a 4-fold lower rate in the large inverted repeat than in either single copy region (Birky 1989). Most substitutions are silent changes at the third codon position whereas missense substitutions occur at both ends of genes (Palmer 1987). Primarily using *rbc*L and *atp*B/E sequence data, transitions occur more often than transversions by a factor of less than two (Zurawski et al 1984, Palmer 1987, Palmer 1991, Palmer et al 1988). Chloroplast gene sequences have proven useful for phylogenetic and evolutionary analyses of plant genes and sequences (Ritland and Clegg 1987).

THE GENERA *PSEUDOTSUGA* AND *PINUS*

Gymnosperms constitute a paraphyletic group within the monophyletic seed plants based on cytoplasmic rRNA sequence and morphological data (Zimmer et al 1989). The conifers (order Coniferae) include seven families (Pinaceae, Cupressaceae, Taxodiaceae, Araucariaceae, Podocarpaceae, Taxaceae, and Cephalotaceae), 60-63 genera, and 500-600 species. It is by far the largest and most diverse group of extant gymnosperms. Conifers are monophyletic, with Pincaceae acting as the sister group to the other living conifers (Hart 1987). Pinaceae represents the largest family of conifers and includes 10 genera and over 200 species which are restricted to the Northern Hemisphere (reviewed in Price et al 1987). Based on shared morphological features, karyotypes, and radioimmunoassay of seed proteins (Price et

al 1987), Pincaceae can be divided into two groups: *Abies*, *Keteleeria*, *Cedrus*, *Tsuga*, and *Pseudolarix* in one group; *Pinus*, *Picea*, *Larix*, *Pseudotsuga*, and *Cathaya* in another group.

The most important timber producing taxon of conifers is *Pinus*, which includes 90 to 100 species (Harlow et al 1978, Mirov 1967). *Pinus radiata* D. Don (Monterey pine) grows on a strip of coast south of San Francisco, California, in Monterey County, and on the islands of Guadeloupe and Santa Barbara Channel off the coast of southern California (Shaw 1958, Harlow et al 1978). Monterey pine is of little or no commercial value within its range, though it is widely planted in New Zealand, Australia, South Africa, Spain, and Chile for saw timber (Harlow et al 1978).

The genus *Pseudotsuga* includes eight species, two of which are found in the western United States, and the remaining six species is eastern Asia (Hermann 1982). *Pseudotsuga menziesii* (Mirb.) Franco (Douglas-fir) is the only economically important member of the genus, comprising over 50% of standing timber in western North American forests. It exists as two distinct varieties: an interior Rocky Mountain form (var. *glauca*) and a coastal Pacific slope form (var. *menziesii*) (Harlow et al 1978). *P. macrocarpa* (Torr.) Mayr (bigcone Douglas-fir) is smaller in stature than Douglas-fir and is restricted to the mountains of southern California and Baja California (Harlow et al 1978). The six Asian species include: *P. wilsoniana* Hayata (Taiwanese Douglas-fir), *P. japonica* (Shiras.) Beissn. (Japanese Douglas-fir), and from China, *P. sinensis* Dode (Chinese Douglas-fir), *P. forrestii* Craib. (Forrest's Douglas-fir), *P. gaussenii* Flous, and *P. brevifolia* Chang et L.K. Fu (Farjon 1990). It is likely that *Pseudotsuga* and its close relative *Larix* originated in North America from a broadly pine-like ancestor at least 50 million years ago and subsequently migrated to, and diversified in, Asia (Strauss et al 1990).

CHAPTER 1

A MUTATION HOTSPOT IN THE CHLOROPLAST GENOME OF *PSEUDOTSUGA* IS CAUSED BY VARIABILITY IN THE NUMBER OF DIRECT REPEATS DERIVED FROM A PARTIALLY DUPLICATED tRNA GENE

ABSTRACT

We used DNA sequencing and the polymerase chain reaction (PCR) to investigate the nature of a polymorphic region in the chloroplast genome of Douglas-fir and related species in the genus *Pseudotsuga*. Southern blotting and hybridization with cloned cpDNA fragments indicated that a length mutation hotspot resided within a 2.7 kb *Xba*I Douglas-fir fragment. In those individuals surveyed, size variants differed by as much as 1000 bp interspecifically, and by 200 bp within *P. menziesii*. The entire sequence of this 2.7 kb hotspot-containing fragment from Douglas-fir was obtained, and the source of length mutation within the region pinpointed via PCR. We also cloned and sequenced the hotspot from a second individual of Douglas-fir and a single individual of *P. japonica*, and compared these to the homologous non-polymorphic region in *Pinus radiata*.

The locus of polymorphism is characterized by hundreds of base pairs of imperfect, tandem direct repeats flanked by a partially duplicated and intact *trn*Y-GUA gene. The duplication is direct in orientation and consists of 43 bp of the 3' end of *trn*Y and 25 bp of its 3' flanking sequence. Tandem repeats show a high sequence similarity to a 27 bp region of the *trn*Y gene that overlaps one end of the duplication. Repetitive DNA in the *Pseudotsuga* cpDNA hotspot was likely generated at the time of the partial *trn*Y gene duplication by slipped-strand mispairing and these sequences expanded further by unequal crossing-over.

INTRODUCTION

Chloroplast DNA mutation hotspots have been identified in a large number of species (Palmer et al 1985, Johnson and Palmer 1989, Kung et al 1982, Palmer 1985a), and are frequently near the ends of the single copy regions (Tassopulu and Kung 1984, Aldrich et al 1988). Length mutations have been associated with direct repeats (Ogihara et al 1988, Ogihara et al 1991, Ogihara et al 1992, Aldrich et al 1988, Schlunegger et al 1983), inverted repeats (vom Stein and Hachtel 1988, Ogihara et al 1988), and A+T-rich sequences (Aldrich et al 1988, Schlunegger et al 1983, Ogihara et al 1992, and Wolfson et al 1991). In several cases, the hotspot region itself is composed of direct repeats and length variability is attributed to differences in the number of repeat copies (Schlunegger et al 1983, Blasko et al 1988, Wolfson et al 1991, Lidholm and Gustafsson 1991a). However, not all length mutations are flanked by direct repeats, nor are their variable regions composed of tandem repeats (Doebley et al 1987).

In this paper we analyze a length mutation hotspot in the chloroplast genome of the coniferous tree Douglas-fir (Ali et al 1991) and other species in the genus *Pseudotsuga*. Using PCR and Southern analysis, we show that length variation between species exceeds that observed within Douglas-fir by five-fold. By comparing DNA sequences of the hotspot region from two individuals of *P. menziesii* and one individual of *P. japonica* to that of a non-variable homologous region in *Pinus radiata* D. Don (Monterey pine), we identify the source of size variation as variable numbers of imperfect tandem repeats that are flanked by, and resemble, a partially duplicated and intact *trn*Y gene.

MATERIAL AND METHODS

Genomic DNA Isolation and Southern Analysis

Genomic DNA was isolated from needles of 2-3 individuals from each of four species of *Pseudotsuga* (*P. macrocarpa* (Torr.) Mayr (bigcone Douglas-fir), *P. japonica* (Shiras.) Beissn. (Japanese Douglas-fir), *P. wilsoniana* Hayata (Taiwanese Douglas-fir), and *P. sinensis* Dode (Chinese Douglas-fir): see Strauss et al (1990) for plant material source), and 9 individuals of *P. menziesii* (from British Columbia, Washington, Oregon, California, Wyoming, Utah, Colorado, New Mexico, Arizona, and Mexico: see Strauss and Tsai (1988) for plant material source) with a modified CTAB method after Strauss et al (1990). Four to eight micrograms of cesium chloride purified DNA was digested with *Eco*RI according to suppliers instructions and electrophoresed in a 0.8% agarose gel in TAE buffer (80 mM Tris, 16.6 mM sodium acetate, 2 mM EDTA, adjusted to pH 8.1 with glacial acetic acid). DNA was blotted in an alkaline denaturation solution (Reed and Mann 1985) onto Zetabind nylon membranes (Cuno Inc., Meriden, CT) and hybridized with a cloned 2.7 kb *Xba*I fragment from *P. menziesii* cpDNA (Tsai and Strauss 1989) labeled with ^{32}P nucleotides by primer extension with random hexamers (Feinberg and Vogelstein 1983). Filters were washed at low stringency twice for 30 min. (65°C, 2X SSC: 0.3 M NaCl, 0.03 M citric acid, 0.1% SDS (Maniatis et al 1982)), twice again but with 0.5% SDS, and developed.

PCR

Four PCR primers were designed to anneal to the hotspot region based upon preliminary sequence data of the 2.7 kb *Xba*I cpDNA clone from *P. menziesii*. Primers were synthesized at the Central Services Lab, Center for Gene Research and Biotechnology, Oregon State University. Primer pair 1 and 4

(5'TCTAGAAAGGCACTGGCTATCGATC3' and 5'TGATAATTCTAGGCTTTCTAGTTCA3', respectively) amplify the variable region and flanking sequences (the entire 2.7 kb clone), primer pair 1 and 2 (primer 2: 5'ATGCCTACGCTGGTTCAAA3') amplify the flanking sequence 5' of the variable region, and pair 3 and 4 (primer 3: 5'TGCCTCCTTGAAAGAGAGATGTCCT3') amplify the flanking sequence 3' of the variable region (Figure 2). PCR reactions using each of the three primer pairs were carried out on genomic DNA from five individuals of *P. menziesii* and single individuals of *P. japonica*, *P. macrocarpa*, *P. sinensis*, and *P. wilsoniana*. The polymerase chain reactions were performed as specified by the GeneAmp kit protocol (Perkin Elmer Cetus) using 2.5 U of *Taq* polymerase (Promega, Madison, WI), reaction buffer supplied by the manufacturer (Promega), 200 ng template DNA, and 25 pmoles of each primer in a final volume of 100 µl. PCR reactions were carried out in a DNA Thermal Cycler (Perkin Elmer Cetus) using 40 cycles of denaturation for 1 min. @ 94°C, annealing for 1 min. @ 55°C, ramping the temperature to 72°C over 1.5 min., and extension for 2.5 min. at 72°C. Between 10 and 30 µl of the PCR products were visualized in 1% agarose gels stained with ethidium bromide (Maniatis et al 1982).

cpDNA Hotspot Isolation

Four sources of DNA were used for isolating the conifer chloroplast hotspot (Table 1): (1) the cloned 2.735 kb *Xba*I cpDNA fragment from *P. menziesii* (Tsai and Strauss 1989) (this individual of Douglas-fir is from Corvallis, OR and the clone designated as PMCO), (2) a cloned 2.762 kb *Xba*I cpDNA fragment from a second individual of Douglas-fir originating from Coos Bay, OR, (3) cloned PCR amplified hotspot-containing DNA (1.868 kb) from *P. japonica*, and (4) cloned *Kpn*I cpDNA fragments from *Pinus radiata*.

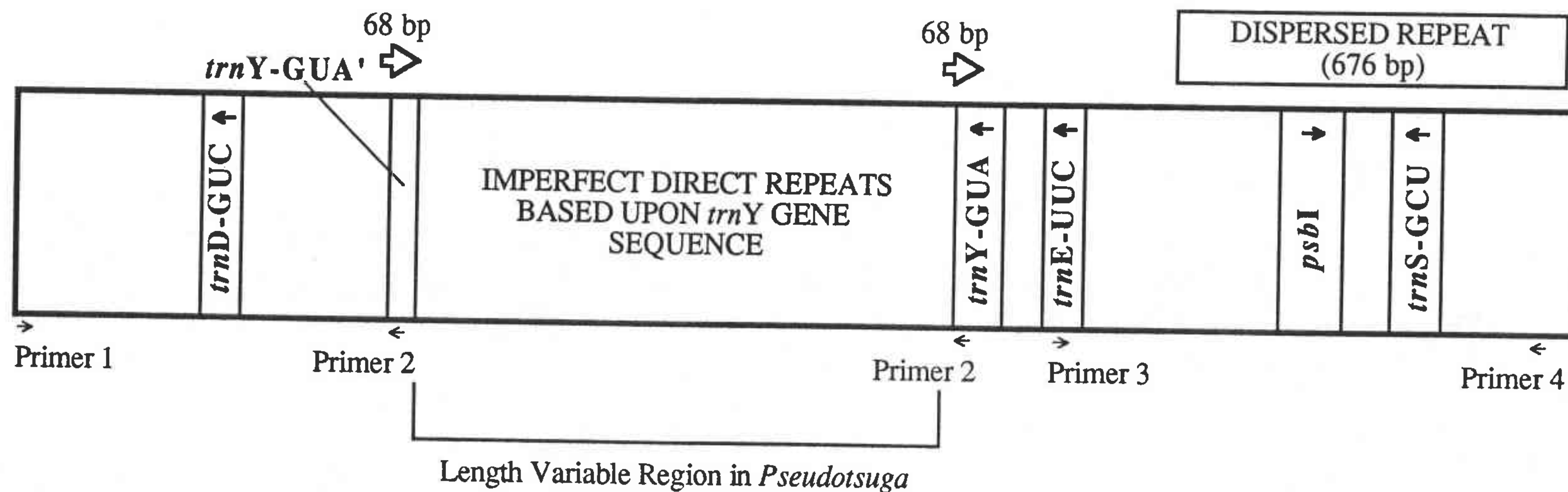


Figure 2. Structure of hotspot-containing clone PMCO (a 2.735 kb Douglas-fir cpDNA fragment). Small arrows under figure indicate sites of PCR primer hybridization. The 676 bp dispersed sequence is repeated twice throughout the Douglas-fir chloroplast genome. Open arrows indicate a 68 bp repeat duplicating the 3' portion of the *trn*Y gene and 25 bp of flanking sequence. Direction of transcription is indicated by arrows in the upper portion of each gene.

Table 1. Name, type, and source of hotspot-containing cpDNA clone material.

CLONE NAME	CLONE TYPE	INSERT SIZE	PLANT SOURCE
PMCO	*Xba*I cpDNA fragment in pUC19	2.735 kb	*Pseudotsuga menziesii*; Corvallis, OR
PMCB	*Xba*I cpDNA fragment in pUC19	2.762 kb	*P. menziesii*; Coos Bay, OR
PJ1,2, and 3	three cpDNA PCR clones in TA-Cloning vectors	all 1.868 kb	*P. japonica*; Cowichan Lake Arbor., B.C. Canada, sample tree #40/1
PR5.8 and PR10.4	*Kpn*I cpDNA fragments in pUC19	5.8 kb and 10.4 kb, respectively	*Pinus radiata*; Corvallis, OR

We isolated chloroplast DNA from the foliage of a single individual of *Pinus radiata* as described in Strauss et al (1988). Chloroplast DNA was digested with *Kpn*I according to manufacturers suggestions, and separated by electrophoresis in a 0.8% agarose gel in TAE buffer. The 5.8 kb and 10.4 kb *Kpn*I cpDNA restriction fragments (Strauss et al 1988) were recovered from the gel under long-wave UV light, eluted with the Centriluter micro-electroeluter (Amicon, Beverely, MA), ligated into a dephosphorylated pUC19 plasmid vector, and transformed into *E. coli* DH5α (BRL) (Maniatis et al 1982). The two clones are designated as PR5.8 and PR10.4. We found that relative positions of *Kpn*I restriction fragments 3.1 kb and 5.8 kb reported in Strauss et al (1988) should be reversed, making the fragment order in that region 5.9kb - 3.1kb - 5.8kb - 10.4kb.

Genomic DNA was isolated from needles of the Coos Bay, OR Douglas-fir using a modified CTAB protocol (Murray and Thompson 1980). DNA was digested with *Xba*I and electrophoresed as was the *Pinus radiata* cpDNA. An ~2.7 kb agarose slice was removed from the gel under long-wave UV light and the DNA purified with Geneclean (Bio 101, La Jolla, CA). The purified DNA fragment was ligated into a dephosphorylated pUC19 vector and transformed into DH5α competent cells. Colonies were screened with a ^{32}P-labelled PMCO clone (Feinberg and Vogelstein 1983) which served as a probe. A 2.762 kb hotspot-containing fragment was cloned and is designated as PMCB.

PCR primer pair 1 and 4 were used in a reaction to amplify the hotspot containing fragment from *P. japonica*. PCR conditions were as described in the preceeding section. The PCR product was recovered from a 1.5% agarose gel under long-wave UV light and eluted with the Centrilutor micro-electroeluter. Fifty nanograms of amplified DNA was ligated into the TA Cloning System Kit (Invitrogen) vector (1.5 to 1 ratio of insert to vector) following manufacturers instructions, and

transformed into TA Cloning System competent cells conferring kanamycin resistence. Three hotspot-containing clones were obtained (PJ1, 2, and 3).

Subcloning

Deletion subclones were obtained for PMCO, PMCB, PJ1, PJ2, PJ3, PR5.8, and PR10.4 using the Erase-A-Base System Kit (Promega). Subclones containing progressive 200 bp deletions were recovered following manufacturers suggestions. PMCO, PJ2, and PJ3 were also subcloned into pUC19 using combinations of *Eco*RV, *Bam*HI, *Hind*III, *Sma*I, *Xba*I, *Pst*I, *Sph*I, and *Pvu*II restriction enzymes (Maniatis et al 1982).

DNA Sequencing

Cloned cpDNA fragments were sequenced using two methods: (1) the dideoxy chain termination method of Sanger et al (1977) by chemically denaturing the plasmid (Chen and Seeburg 1985) and using the Sequenase Version 2.0 sequencing protocol and kit reagents (United States Biochemical Corporation, Cleveland, Ohio), and (2) automated plasmid sequencing using the Applied Biosystem Model 373 DNA Sequencer with dye-on primer technology in the Center for Gene Research and Biotechnology at Oregon State University. All sequencing was done with universal and reverse primers. Clones PMCO, PMCB, PJ1, PJ2, and PJ3 were sequenced entirely while the *Pinus radiata* clones (PR5.8 and PR10.4) were partially sequenced.

Sequence Analysis

Sequence data was entered into the Intelligenetics Suite programs via the computational molecular biology lab of the Center for Gene Research and Biotechnology. Consensus sequences were assembled using the Gel program and aligned with Genalign and Align. GenBank and EMBL databases were searched using IFIND.

RESULTS

A Hypervariable Region Exists in the Chloroplast Genomes of Both Asian and North American Species in the Genus *Pseudotsuga*

We detected cpDNA length variation among five species of *Pseudotsuga* (*P. menziesii*, *P. macrocarpa*, *P. sinesis*, *P. wilsoniana*, and *P. japonica*) by Southern analysis when probing with a 2.7 kb fragment (Tsai and Strauss 1989) from the Douglas-fir chloroplast genome (clone PMCO). The variable chloroplast DNA bands (Figure 3A) range in size from 4.9 to 5.9 kb while the other two bands are invariant. Intraspecific variation is also present in all species except *P. sinensis*. In a broader intraspecific study via Southern analysis, we surveyed 9 individuals throughout the range of *P. menziesii* for variability in the hotspot region (Figure 3B). Variable cpDNA bands range in size by only 200 bp (5.6 to 5.8 kb) compared to the observed 1 kb size range among species. Two other invariant bands are again present. We did not detect length variation in this region when surveying multiple individuals of three species of pines: *Pinus radiata*, *P. muricata* D. Don (bishop pine), and *P. attenuata* Lemm. (knobcone pine).

Length Variability Observed on Southern Blots Originates from a Region with Large Numbers of Tandem Repeats

Nucleotide sequence data shows PMCO to contain four tRNA genes (*trn*D-GUC, *trn*Y-GUA, *trn*E-UUC, and *trn*S-GCU), one gene for a subunit of photosystem II (*psb*I), and a partially duplicated *trn*Y gene (Figure 2) (see Appendix B.i for PMCO sequence data). A 676 bp sequence containing *trn*S-GCU and *psb*I is repeated twice throughout the *P. menziesii* chloroplast genome (Tsai and Strauss 1989) and is found at the 3' end of PMCO. Forty-three basepairs of the 3' end of *trn*Y-GUA together with

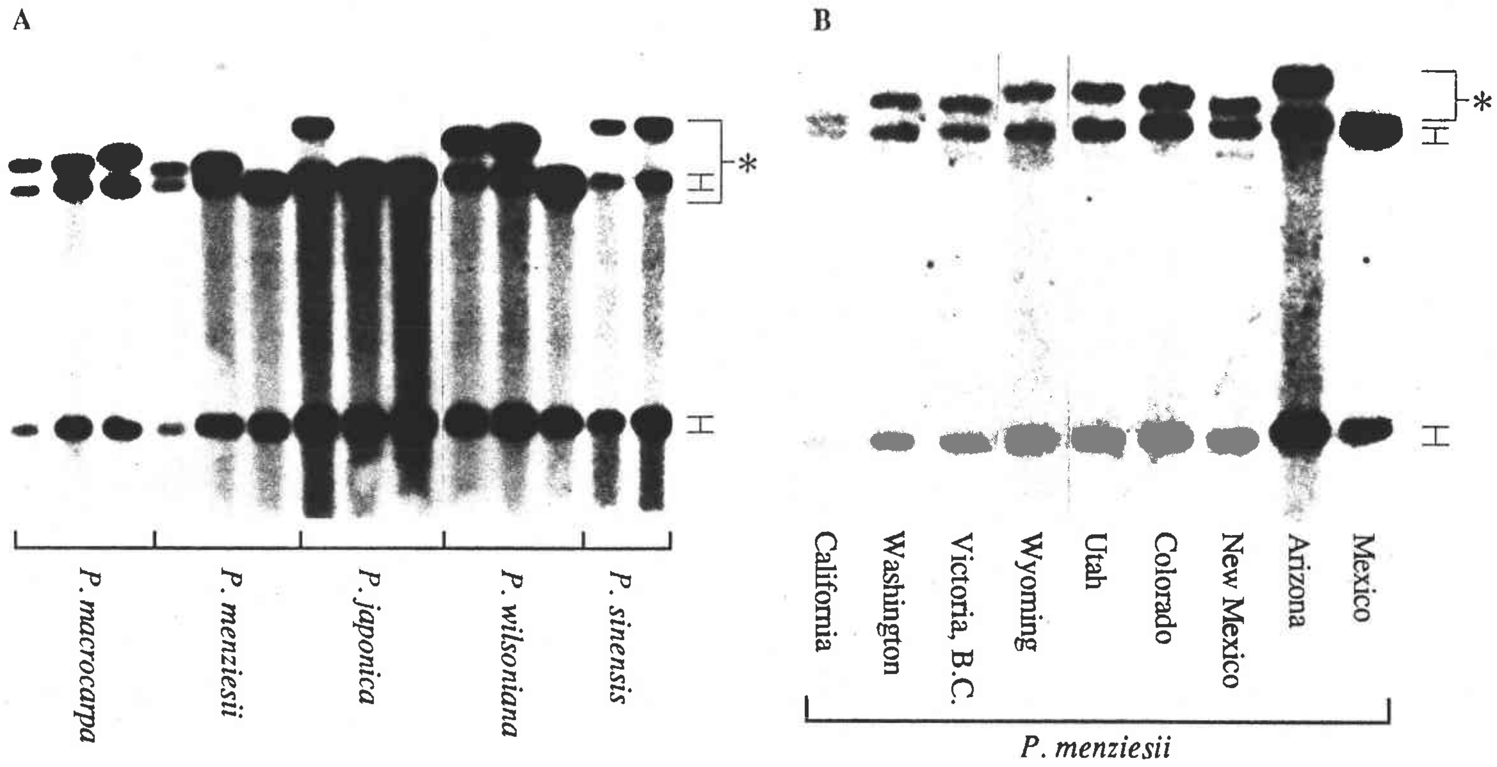


Figure 3. Examples of a length mutation hotspot in the genus *Pseudotsuga*. Genomic DNA was digested with *Eco*RI and probed with a 2.7 kb fragment from Douglas-fir cpDNA (clone PMCO). (A) Among species of *Pseudotsuga*. Asterisks (*) and brackets indicate variable bands that range in size by up to 1 kb. (B) Among individuals of *P. menziesii*. Variable bands differ in size by up to 200 bp. Two rows of invariant bands are present in both experiments (⌶).

25 bp of 3' flanking sequence (68 bp total length) is duplicated to yield the pseudo-*trn*Y gene (*trn*Y'). The 68 bp duplication (showing 96% similarity) is separated by 932 bp of imperfect tandem direct repeats.

To identify the source of length variability within PMCO, four oligonucleotides were synthesized and used as primers to amplify specific parts of the fragment (Figure 2). The primers as pairs amplify (1) the 3' end of the fragment characterized by the presence of a dispersed repeat (primers 3 and 4), (2) the 5' end of the fragment (primers 1 and 2), or (3) the entire fragment (primers 1 and 4). PCR amplified DNA of three samples of five species of *Pseudotsuga* (Figure 4A) showed little or no length variation in the 3' fragment end, and no length variation in the 5' end of three of the five species, but novel mutations in *P. wilsoniana* and *P. sinensis*-- whose banding patterns are highly dissimilar to the others.

DNA amplified from five individuals of *P. menziesii* (Figure 4B) showed almost no length variation in either fragment end (the first and second set of samples). The majority of length variation in the *Pseudotsuga* hotspot region appears to occur in the tandem repeat section, as indicated by the presence of length variation in the total fragment (the third set of samples in each experiment) which was not observed within either fragment end.

The length variation observed on the Southern blots is quantitatively accounted for by the length variation that occurs in the direct repeat region of PMCO as measured by PCR. The size of the variant bands in the Southern analysis were regressed against the size of the PCR amplified fragments containing the direct repeat region to examine their association (Figure 5). For the intraspecific study, a slope of 0.809 and r^2 of 0.956 was found (n=5); for the interspecific study, the slope was 0.957 and $r^2 = 0.991$ (n=5). This shows that the vast majority of length polymorphism observed on

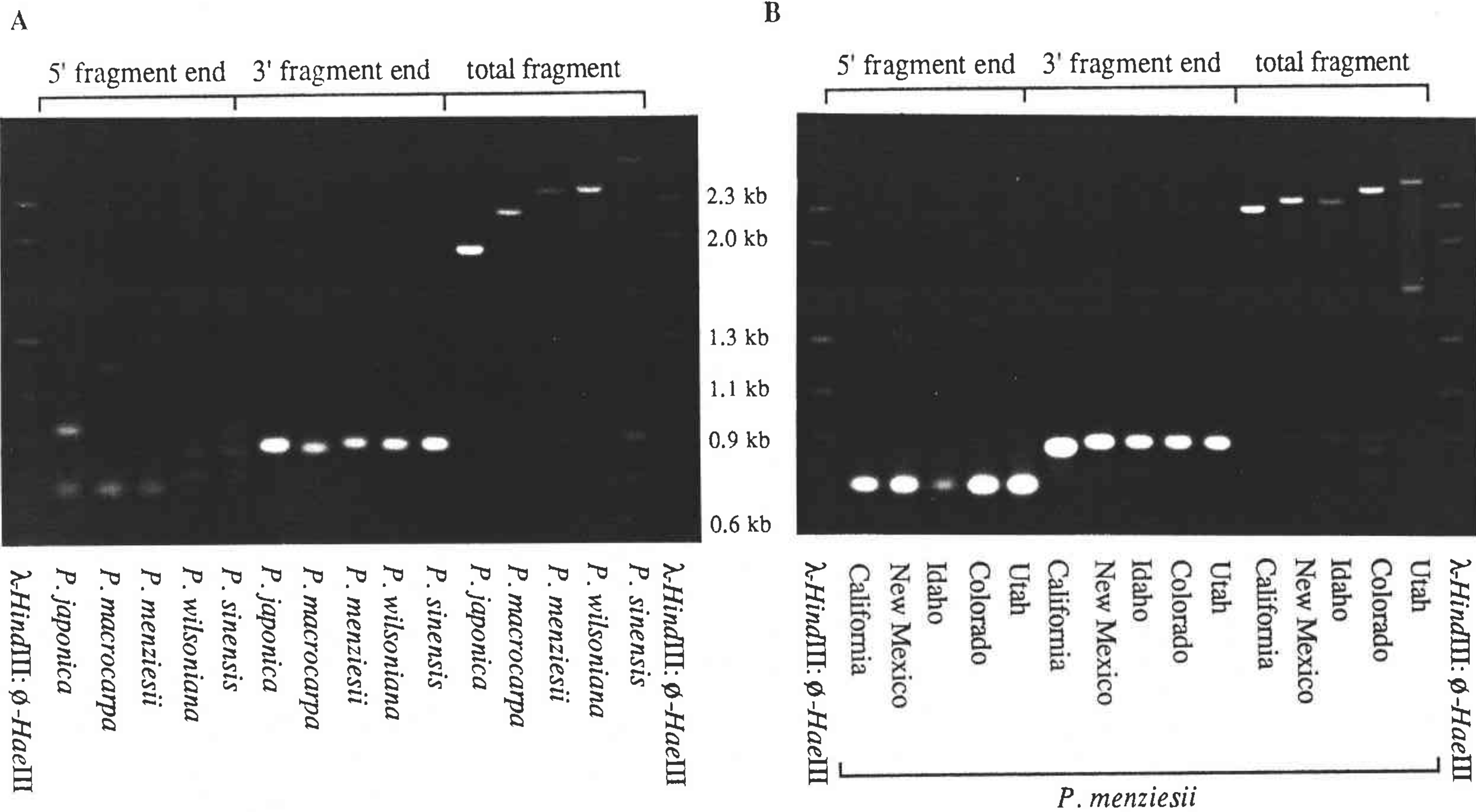


Figure 4. Identification of the source of length variability within PMCO using the Polymerase Chain Reaction. Genomic DNAs from five species of *Pseudotsuga* (A), and five individuals of *P. menziesii* (B), were amplified via PCR with pairs of primers and visualized in 1% agarose gels.

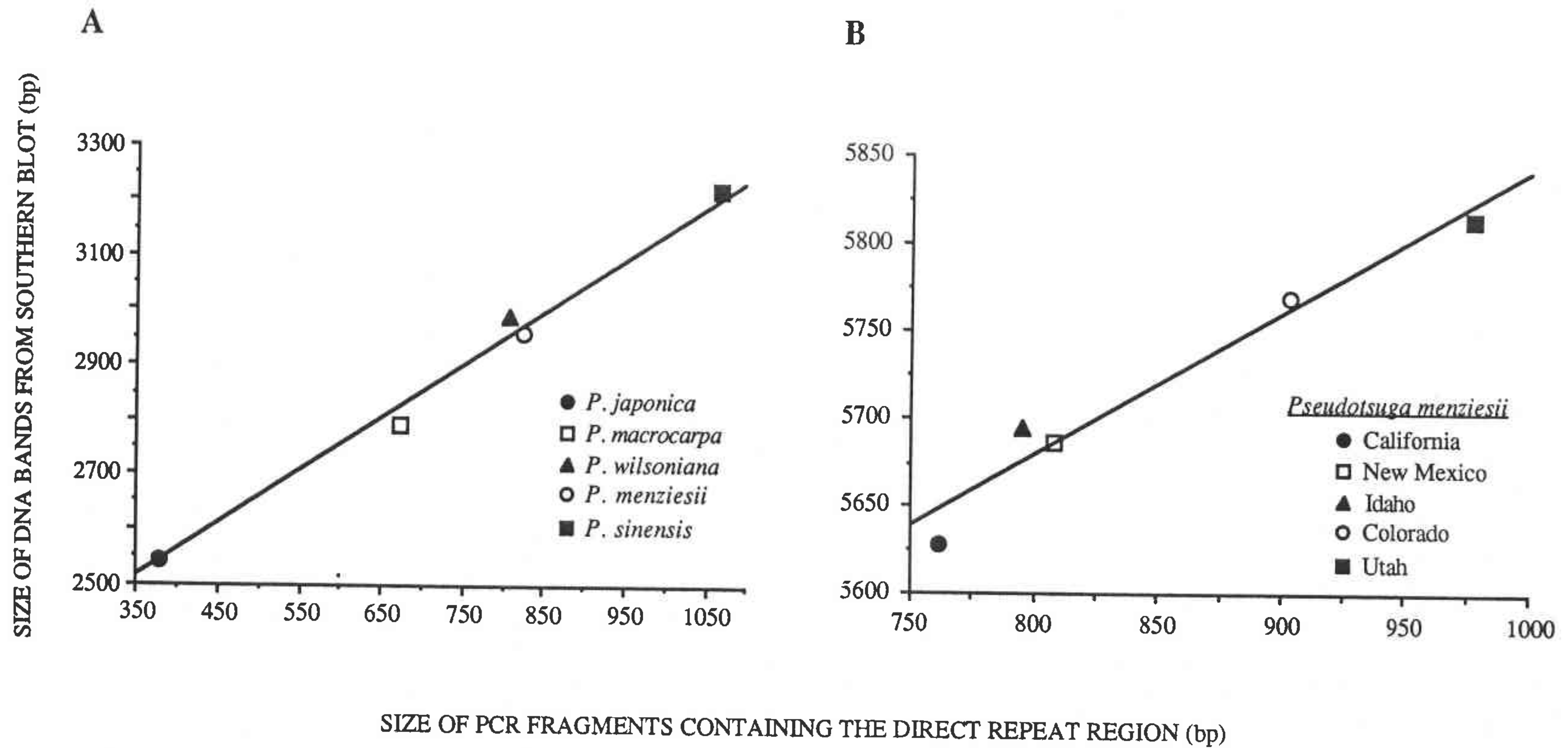


Figure 5. Regression analysis used to study relationships of length variability on Southern blots to that observed with PCR.

Southerns originates in the imperfect, tandem direct repeat region of PMCO and not from elsewhere inside or outside the cloned region.

The Length Variable Repetitive Region Resides Between a Partially Duplicated and an Intact *trn*Y Gene

The conifer cpDNA hotspot was sequenced from two individuals of *P. menziesii* (PMCO and PMCB: see Appendix B.ii for PMCB sequence data), and single individuals of *P. japonica* and *Pinus radiata*. Because *P. japonica* hotspot DNA was derived via PCR, we cloned and sequenced three separate 1.868 kb amplified fragments. We found an error rate of 1 in 1,868 bases (0.05% or 5.3×10^{-4}) with two transitions and one transversion occurring among the samples. Clone PJ1 contained one transition at position 1594 relative to the other two PCR clones, and PJ3 contained a transversion and a transition (positions 1520 and 1630, respectively) relative to PJ1 and PJ2. The consensus sequence (PJ) between the three PCR clones was used in the analysis (see Appendix B.iii for *P. japonica* DNA sequence). In *Pinus radiata*, the region homologous to the *Pseudotsuga* hotspot is located at one end of the *Kpn*I 5.8 kb cpDNA fragment (clone PR5.8) and shows no length variation. The *Kpn*I restriction site between the 5.8 and 10.4 kb fragments (see Strauss et al (1988) for restriction map) lies at the 3' border of the *trn*E gene, therefore including *trn*E at the end of the PR10.4 clone. The reported sequence (PR) consists of data from *trn*D through *trn*E (see Appendix B.iv for PR sequence data) and is used in the analysis.

We compared the sequence from two *P. menziesii* individuals and one individual of *P. japonica* to *Pinus radiata* (Figure 6). In all individuals of *Pseudotsuga* examined, *trn*Y has been partially duplicated and is separated from the intact gene by a length variable region of imperfect tandem direct repeats. No such duplication has occurred in this region in *Pinus radiata* and subsequently no length variability was observed. The

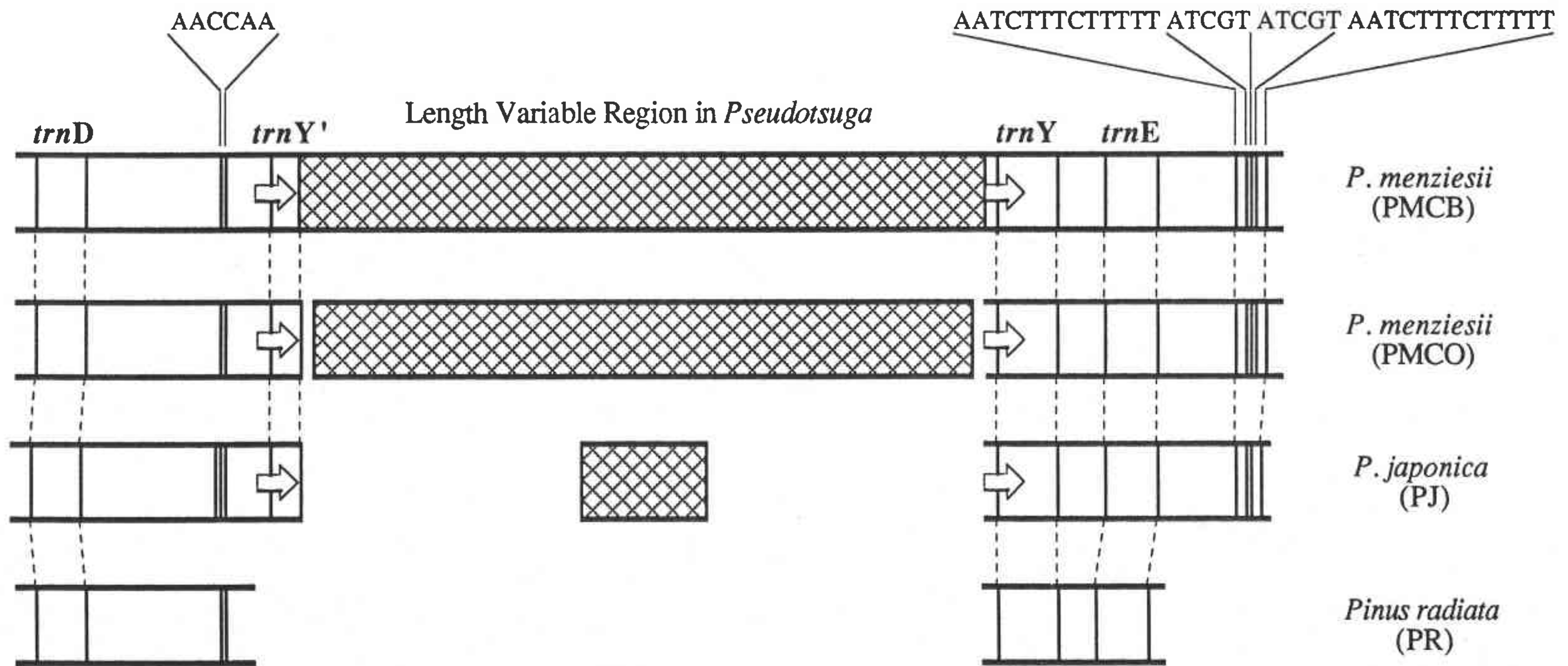


Figure 6. A comparison of the *Pseudotsuga* cpDNA hotspot relative to the homologous area of *Pinus radiata*. The top two sequences are from two individuals of *Pseudotsuga menziesii* (PMCB and PMCO, respectively), the third sequence from *P. japonica*, and the bottom sequence from *Pinus radiata*. The 68 bp repeat duplicating part of the *trn*Y gene is indicated by open arrows. Cross-hatched areas represent regions of length variability.

DNA sequence from the two individuals of *P. menziesii* are completely identical for the entire 2.735 and 2.762 kb length, respectively, except for a 27 bp insertion in the PMCB length variable region relative to that of PMCO. The variable region in *P. japonica* is 168 bp in length, 764 bp shorter than in PMCO. Point mutations and small insertion/deletions in *P. japonica* compared to *P. menziesii* are also observed in the areas flanking the hotspot. In the area 5' of the variable region, *P. japonica* varies from *P. menziesii* by two transitions, a one base pair deletion, and one 6 bp addition (figure 6) in the form of a tandem duplication (5'AACCAA3'). Downstream of the variable region, from *trn*Y through the dispersed repeat, *P. japonica* contains five transitions, four transversions, a 2 bp deletion, and a 5 bp deletion relative to *P. menziesii*. The 5 bp deletion (5'ATCGT3'), indicated in Figure 6, occurs as a tandem repeat flanked by 13 bp direct repeats in *P. menziesii*. All genes sequenced share 100% similarity among the three individuals of *Pseudotsuga* and point mutations and deletions/insertions occur only in intergenic regions. Sequence homology between *Pseudotsuga* and *Pinus radiata* *trn*D, *trn*Y, and *trn*E genes are 97%, 98% and 96%, respectively. Intergenic regions between the genera share less sequence homology: 76.4% between *trn*D and *trn*Y, 75.9% between *trn*Y and *trn*E, 33% 315 bp downstream of *trn*D, and 47.5% 178 bp upstream of *trn*E.

Hotspot Tandem Repeats are Similar to *trn*Y Gene Sequence and Hierarchical

The length-variable region in *Pseudotsuga* is comprised of imperfect tandem direct repeats based on *trn*Y gene sequence. The hotspot in PMCO contains 40 imperfect direct repeats representing 19 repeat units ranging in size from 14 to 30 bp (Figure 7). The location of repeat units along the DNA strand are indicated at the top of the figure and their respective sequences listed. Repeat units contain a centrally located core sequence (ACGGATTT) and fall into two basic groups based on two base pairs

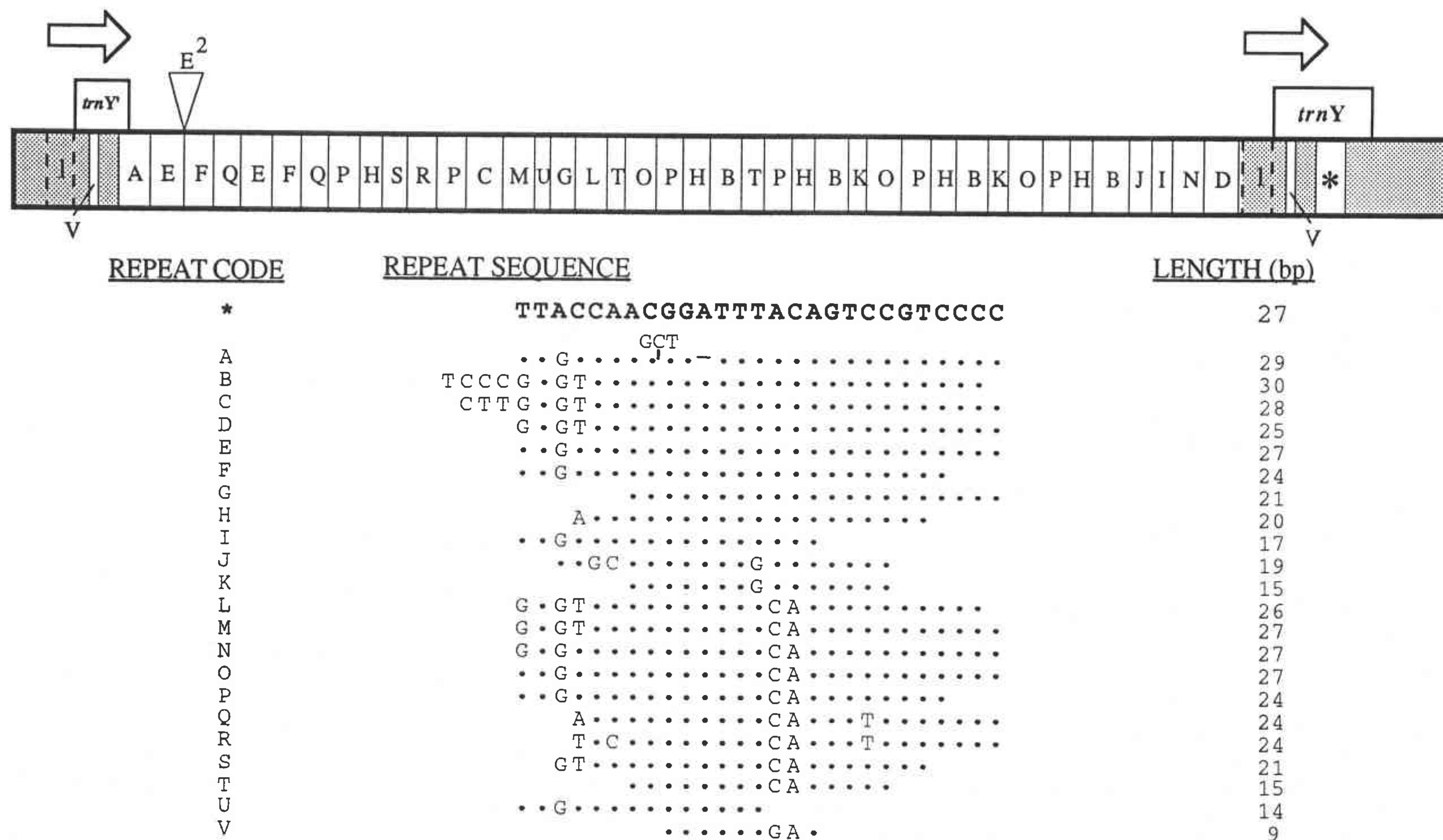


Figure 7. Repeat structure and DNA sequence of a *Pseudotsuga menziesii* length mutation hotspot (clone PMCO). Repeat location is indicated in the upper portion of the figure and corresponding sequences listed below. The asterisk (*) indicates the location and sequence of the *trn*Y gene region upon which repeat sequences are based. Repeat A has a 3bp insertion and a 1bp deletion relative to the tRNA sequence. Shaded areas show no sequence similarity to repeats. [1] Highly A+T-rich 25 bp region: 84% preceeding *trn*Y', 88% preceeding *trn*Y. [2]A repeat E duplication occurs in the hotspot of *P. menziesii*, Coos Bay OR (PMCB) relative to *P. menziesii*, Corvallis OR (PMCO).

following the core (AC or CA). Repeats usually terminate at a run of C nucleotides. Repeat units differ at thirteen out of a possible 30 positions with repeat A containing a 3 bp insertion and a 1 bp deletion relative to the other repeats. Grouping of the repeats is based upon similarity and not on their location along the hotspot. Sequences of the repeats resemble a 27 bp section of the *trn*Y gene. The 27 bp sequence is located toward the center of the gene, overlaps the right side of the *trn*Y duplication, and shares at least 67% similarity to the repeat units. The flanking sequence duplicated with part of *trn*Y is highly A+T-rich (>84% among *Pseudotsuga* individuals) and bears no sequence similarity to repeat units.

A complex hierarchical structure exists to the organization of the repeat units within the hotspot (Figure 8). Sequence 'P' is repeated six times throughout the hotspot, four copies of which are followed by repeats 'H' and 'B', and two of these followed by repeats 'K' and 'O'. Another possible repeat level nested on top of this are repeats 'O-P-H-B' (sharing the third, fifth, and sixth copy of 'P' from the left in the figure), overlapping portions of the 'P-H-B-K-O' repeat cluster.

The PMCB hotspot is 27 bp larger than that of PMCO as determined by DNA sequencing. This insertion takes the form of a repeat 'E' duplication 46 bp from *trn*Y' (Figure 7). The remainder of the sequence is identical between the two individuals of *P. menziesii*. A large deletion has occurred in the hotspot of *P. japonica* relative to PMCB and PMCO (Figure 9). There is 99.7% sequence similarity from repeat 'N' to the end of the PJ sequence with a single base change occurring in the A+T rich region compared to *P. menziesii*. Sequence similarity between *P. japonica* and *P. menziesii* is also high (100%) upstream of the hotspot. However, the region in the *P. japonica* hotspot between the partial *trn*Y duplication and repeat 'N' shows no clear homology to the *P. menziesii* repeat organization although the area still clearly resembles *trn*Y gene sequence and shows strong similarity to *P. menziesii* repeat unit sequences. In fact,

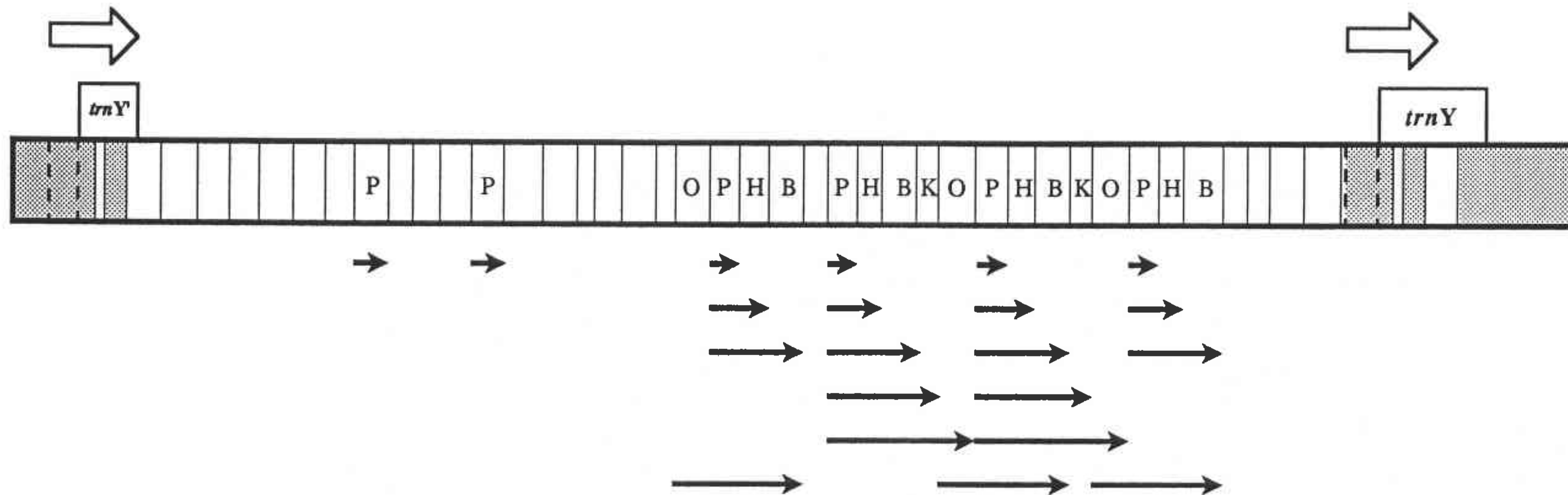


Figure 8. Example of the hierarchical structure of repeated units in the *Pseudotsuga* hotspot. The variable region from clone PMCO is shown with some of several possible nested repeat structures indicated by closed arrows below the DNA strand.

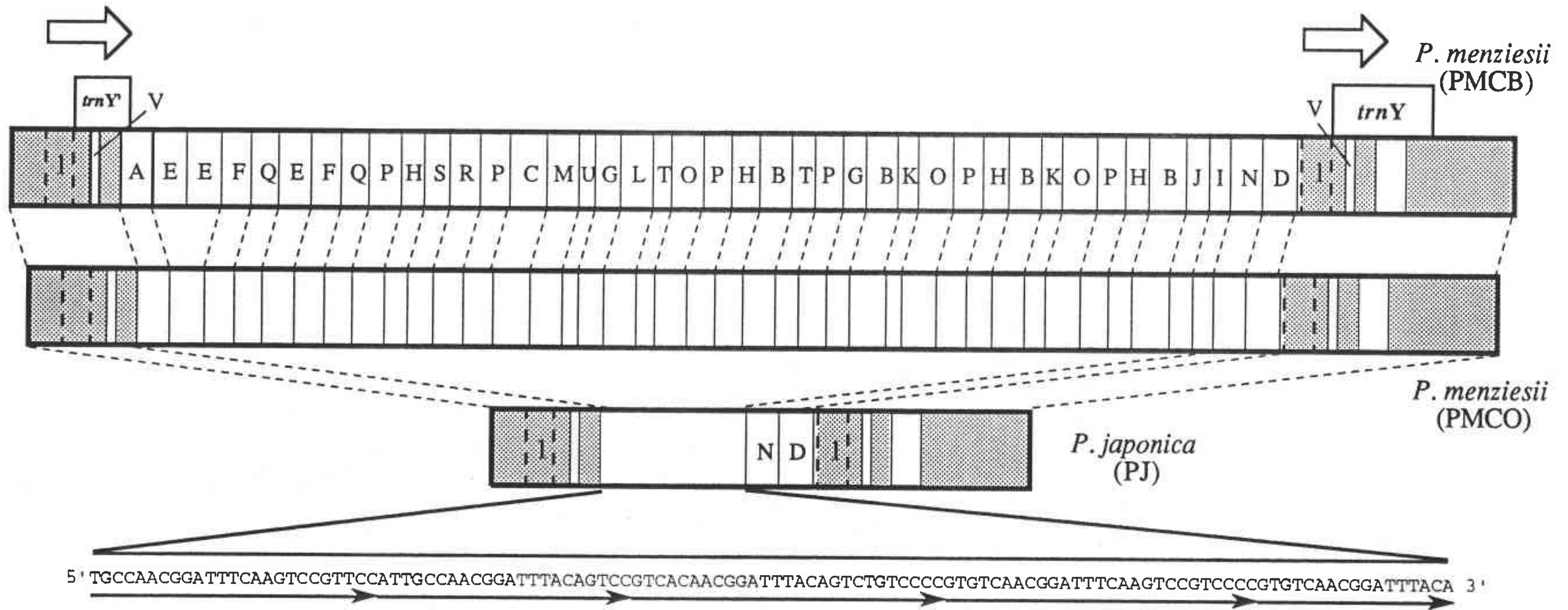


Figure 9. Repeat structure of the chloroplast DNA hotspot in three individuals of *Pseudotsuga* representing two species. The top two samples are from two individuals of *P. menziesii* (PMCB and PMCO, respectively) and the third from *P. japonica*. [1] See Figure 8 legend. Arrows under *P. japonica* sequence data represent imperfect direct repeats, related in sequence to the *P. menziesii* repeat units.

although all three individuals of *Pseudotsuga* share 'N-V' repeats at the 3' end of their hotspots, these sequences may not be homologous and instead be the result of independent rearrangement events. In an attempt to ascertain homology between the *P. japonica* and *P. menziesii* hotspots, sequence similarity was determined between *trn*Y' and the 3' N repeat of the two species (117 bp in *P. japonica* and 880 bp in PMCO) (Figure 10). The 117 bp *P. japonica* region was compared at each PMCO repeat for percent similarity keeping both regions intact. There is no clear region in *P. japonica* which is homologous to *P. menziesii* and, in fact, sequence similarity is above 80% in five different regions along the *P. menziesii* hotspot. Apparent, however, is a strong periodicity in repeat organization along the hotspot.

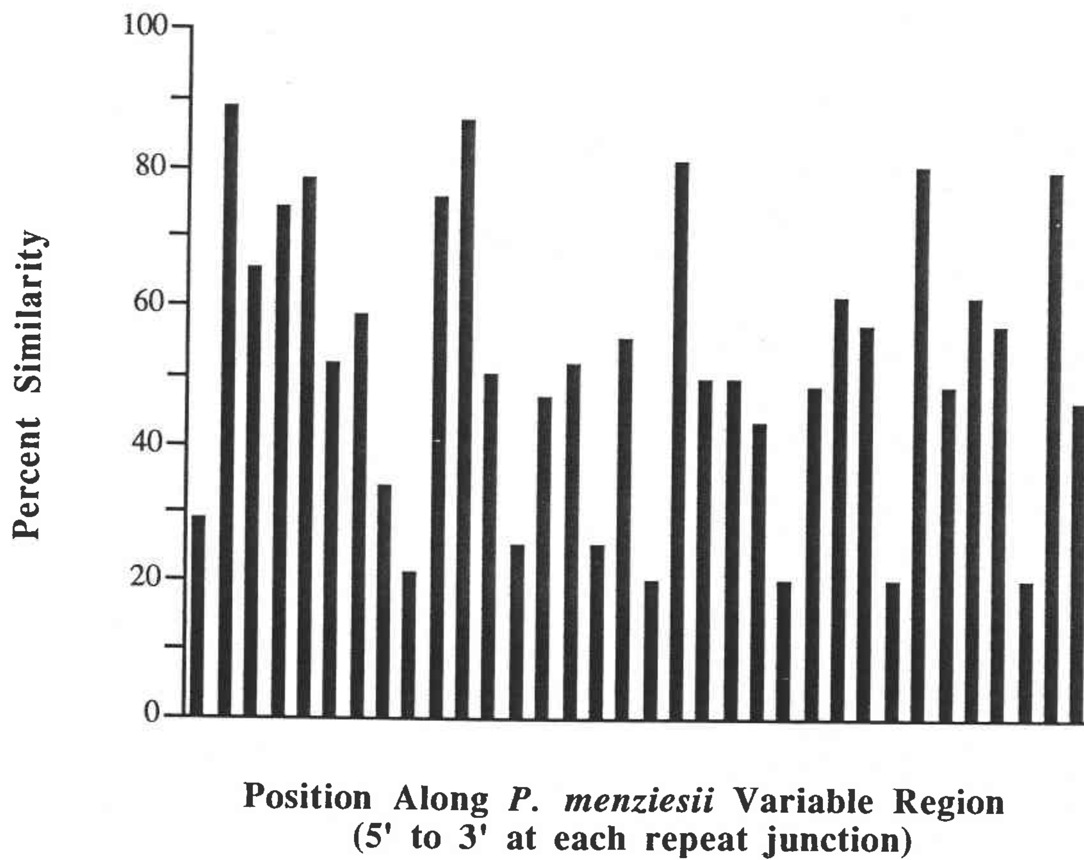


Figure 10. Histogram showing percent sequence similarity between the *P. japonica* and *P. menziesii* length variable regions. 117bp of the *P. japonica* hotspot was kept intact and moved along the *P. menziesii* hotspot at each repeat junction to determine similarity.

DISCUSSION

The *Pseudotsuga* cpDNA hotspot is located in an area 180° from the ribosomal RNA genes on the Douglas-fir cpDNA restriction site map of Strauss et al (1988). Length variability is the result of variable numbers of imperfect, direct tandem repeats. Hotspot regions sequenced in this paper were 959 bp, 932 bp, and 168 bp in length in the two individuals of *P. menziesii* and *P. japonica*, respectively. Variable regions are flanked by a 68 bp duplication in direct orientation containing 43 bp of the 3' end of *trn*Y-GUA and 25 bp of 3' flanking sequence. The hotspot is, therefore, bounded by a partially duplicated and intact *trn*Y gene. The duplicated 3' flanking sequence is highly A+T-rich (>84%). A+T-rich regions of DNA are often associated with genome instability caused by hotspots (Hyrien et al 1987) and small length mutations have been noted to occur preferentially at these locations (Zurawski et al 1984). Tandem repeats in the hotspot vary in length between 14 and 30 bp and their sequences resemble an internal 27 bp region of *trn*Y that overlaps the partial gene duplication (see figure 7). We observed no intra- or interspecific length variation in the homologous region of *Pinus radiata* cpDNA (data not shown), and upon sequencing the region in *P. radiata*, found no *trn*Y gene duplication and the absence of tandem direct repeats and length variability.

From our PCR experiments, one or more repeated primer hybridization sites was indicated by the presence of multiple bands in the first and third sets of samples (Figure 4). When amplifying with primer pair 1 and 2 (first set of samples in each experiment), two fragments are obtained due to dual primer 2 hybridization sites. Primer 2 hybridized to both the intact and pseudo-*trn*Y genes (see figures 2 and 4), amplifying (1) the fragment end plus the tandem repeat region and (2) the 5' fragment end, respectively. The result can be seen in extra variable bands in the first set of

samples which mimics the length variation observed in the third sample set. Unexpected small bands present within the 'total fragment' amplifications are probably artifactual, resulting from the opportunity for amplification of misprimed products given the large target fragments amplified, and the repetitive priming site of primer 4.

Two events appear to have taken place in the evolution of the *Pseudotsuga* hotspot. One is the partial duplication of *trn*Y and flanking sequence (which is absent in *Pinus*) and the other is insertion/deletion events that create the length variable region. Homologous recombination between repeated sequences is a frequent process in chloroplast DNA and includes both reciprocal (intramolecular crossing-over and intermolecular unequal crossing-over) and non-reciprocal (intra-and intermolecular gene conversion) exchanges (Bowman et al 1988). Newman et al (1992) consider gene conversion to be the most common form of homologous recombination in the chloroplast genome and is the mechanism responsible for preserving sequence identity between the large inverted repeats. Repeated sequences are often associated with length mutations and it is thought that they serve as substrates for the inter- (Palmer 1985a, Palmer 1991) and intra- (Palmer 1991, Ogihara et al 1988) molecular recombination mechanisms, or as sites for slipped-strand mispairing during DNA replication and repair (Wolfson et al 1991, Ogihara et al 1992, Dover 1986, Weston-Hafer and Berg 1989).

Size differences observed in a cpDNA hotspot of wheat and *Aegilops* were attributed to intramolecular recombination between short direct repeats resulting in deletions (Ogihara et al 1988) and DNA replication slippage involving several basepairs yielding small length mutations (Ogihara et al 1992). Replication slippage was also proposed as the mechanism responsible for length mutations observed in *Oenothera* cpDNA (Wolfson et al 1991). However, a hotspot which mapped to the large inverted repeat in the chloroplast genome of *Oenothera hookeri*, which differs in copy number (2, 4, or 7 copies) of a 24 bp tandem direct repeat, appears to be the result of

recombination between imprecisely aligned large inverted repeats and subsequent copy correction of the hotspot region (Blasko et al 1988). Very small direct repeats (3-9 bp) may also serve as substrates for recombination as is the case in the cpDNA of *Pinus contorta* and *P. banksiana*. In these genomes illegitimate recombination between these small direct repeats has been proposed to explain the presence of variable numbers of 124 and 150 bp tandem repeats (Lidholm and Gustafsson 1991b).

The observed length variability in *Pseudotsuga* can be explained by several of these mechanisms. Length mutations can result from intra- and intermolecular recombination between direct repeats. Repeats can be spread by intermolecular recombination, also between direct repeats, and replication slippage (Palmer 1985b). In the *Pseudotsuga* hotspot it is possible that DNA slippage caused the initial formation of short repeated sequences and these sequences expanded further by unequal crossing-over because of their propensity to mispair. Initial repeat formation could have also occurred at the time of the *trn*Y duplication. This may be the case since the tandem repeat sequences are highly homologous to a *trn*Y gene region which overlaps the partial duplication.

It is much more difficult to explain the formation of the partial *trn*Y duplication. In chloroplast DNA of higher plants, tRNA genes are dispersed throughout the genome (Quigley and Weil 1985) and are associated with many DNA rearrangements. Intact tRNA genes, and dispersed repeats that are segments of tRNA sequences, have been associated with contraction and expansion of the large inverted repeat (Wolfe 1988) and with inversion endpoints (Howe et al 1988, Shimada and Sugiura 1989), although not all inversion borders are near tRNA genes (Howe et al 1988). tRNA-pseudogene formation may be the result of intermolecular recombination between tRNA genes (Marechal-Drouard et al 1991), creating the gene duplication at the time of rearrangement. In some chloroplast genomes, tRNA genes show a tendency to change

locations and cluster (Mubumbila et al 1984). This is apparent in the cpDNA of *Vicia faba* where two *trn*L(-CAA and -UAA) genes are 443 bp apart compared to being widely separated in all other higher plant chloroplast genomes studied (Bonnard et al 1985). A partial *trn*L-UAA duplication exists in the intergenic region between the two genes that consists of 100 bp of the 5' flanking sequence, 35 bp of the 5' exon and 42 bp of the intron. The duplication is separated by a 23 bp, 78% A+T-rich sequence. Also present are several short direct repeats. The authors suggest that this region may have been involved in rearrangements at the time when the *trn*L duplication occurred.

The *Pseudotsuga* hotspot also is located near an area of past rearrangement. Adjacent to the hotspot is the border of a conifer/Petunia inversion (Strauss et al 1988). The *trn*Y duplication may have, as proposed for *Vicia faba*, occurred at the time of the inversion event through intra- or intermolecular recombination between tRNA genes (Hiratsuka et al 1989). *Pinus* cpDNA does not contain a *trn*Y duplication even though also sharing the Petunia inversion endpoint. It is possible that recombination between the duplicated and intact gene deleted the region subsequent to the divergence of the lineages leading to the current genera. More light may be shed on this by investigating this region in other members of the Pinaceae. Although sequence data is not available, the hotspot appears to exist in the homologous region of *Sequoia sempervirens* (D. Don) Endl. (coast redwood) (Ali et al 1991) and *Thuja plicata* Donn ex D. Don (Western redcedar) (unpublished results). This would support the contention that the partially duplicated *trn*Y and thus potential hotspot was lost in the *Pinus radiata* chloroplast genome.

The usefulness of the hotspot for phylogenetic purposes and intraspecific studies is very doubtful because of very rapid cpDNA alterations in the region. There is a high potential for convergent evolution and similar phenotypes (ie. same total length of a variable region) with different repeat configurations. Also, shared configurations, such

as the 3' end of the variable regions of *P. japonica* and *P. menziesii*, could be misleading since they may actually be the result of different, independent, rearrangement events.

REFERENCES

Aldrich, J., B.W. Cherney, E. Merlin, and L. Christopherson. 1988. The role of insertion/deletions in the evolution of the intergenic region between *psb*A and *trn*H in the chloroplast genome. Curr. Genet. 14:137-146.

Ali, I.F., D.B. Neale, and K.A. Marshall. 1991. Chloroplast DNA restriction fragment length polymorphism in *Sequoia sempervirens* D. Don Endl., *Pseudotsuga menziesii* (Mirb.) Franco, *Calocedrus decurrens* (Torr.), and *Pinus taeda* L. Theor. Appl. Genet. 81:83-89.

Blasko, K., S.A. Kaplan, K.G. Higgins, R. Wolfson, and B.B. Sears. 1988. Variation in copy number of a 24-base pair tandem repeat in the chloroplast DNA of *Oenothera hookeri* strain Johansen. Curr. Genet. 14:287-292.

Bonnard, G., J.-H. Weil, and A. Steinmetz. 1985. The intergenic region between the *Vicia faba* chloroplast tRNA-Leu (CAA) and tRNA-Leu (UAA) genes contains a partial copy of the split tRNA-Leu (UAA) gene. Curr. Genet. 9:417-422.

Bowman, C.M., R.F. Barker, and T.A. Dyer. 1988. In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination. Curr. Genet. 14:127-136.

Chen, E.Y. and P.H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165-170.

Doebley, J.F., D.P. Ma, and W.T. Renfroe. 1987. Insertion/deletion mutations in the *Zea* chloroplast genome. Curr. Genet. 11:617-624.

Dover, G.A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet. 2:159-165.

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

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.-R. Sun, B.Y. Meng, Y.-Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid inversion during the evolution of the cereals. Mol. Gen. Genet. 217:185-194.

Howe, C.J., R.F. Barker, C.M. Bowman, and T.A. Dyer. 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. 13:343-349.

Hyrien, O., M. Debatisse, G. Buttin, and B.R. de Saint Vincent. 1987. A hotspot for novel amplification joints in a mosaic of *ALU*-like repeats and palindromic A+T-rich DNA. EMBO J. 6:2401-2408.

Johnson, L.B. and J.D. Palmer. 1989. Heteroplasmy of chloroplast DNA in *Medicago*. Plant Mol. Biol. 12:3-11.

Kung, S.D., Y.S. Zhu, and G.F. Shen. 1982. *Nicotiana* chloroplast genome. III. Chloroplast DNA evolution. Theor. Appl. Genet. 61:73-79.

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

Lidholm, J. and P. Gustafsson. 1991b. A three-step model for the rearrangement of the chloroplast *trn*K-*psb*A region of the gymnosperm *Pinus contorta*. Nucl. Acids Res. 19:2881-2887.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marechal-Drouard, L. M. Kuntz, and J.H. Weil. 1991. tRNAs and tRNA genes of plastids. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 169-189.

Mubumbila, M., E.J. Crouse, and J.H. Weil. 1984. Transfer RNAs and tRNA genes of *Vicia faba* chloroplasts. Curr. Genet. 8:379-385.

Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res. 8:4321-4325.

Newman, S.M., E.H. Harris, A.M. Johnson, J.E. Boynton, and N.W. Gillham. 1992. Nonrandom distribution of chloroplast recombination events in *Chlamydomonas reinhardtii*: evidence for a hotspot and an adjacent cold region. Genetics 132:413-429.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. USA 85:8573-8577.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1991. Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. Genetics 129:873-884.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1992. Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. Curr. Genet. 22:251-258.

Palmer, J.D. 1985a. Comparative organization of chloroplast genomes. Ann. Rev. Genet. 19:325-354.

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

Palmer, J.D. 1991. Plastid chromosomes: structure and evolution. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 5-53.

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.

Quigley, F. and J.H. Weil. 1985. Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr. Genet. 9:495-503.

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

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schlunegger, B., M. Fasnacht, E. Stutz, B. Koller, and H. Delius. 1983. Analysis of a polymorphic region of the *Euglena gracilis* chloroplast genome. Biochim. Biophys. Acta 739:114-121.

Shimada, H. and M. Sugiura. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr. Genet. 16:293-301.

Strauss, S.H. and C.-H. Tsai. 1988. Ribosomal gene number variability in Douglas-fir. J. Hered. 79:453-458.

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. 68:1502-1510.

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.

Tassopula, D. and S.D. Kung. 1984. *Nicotiana* chloroplast genome. 6. Deletion and hot spot--a proposed origin of the inverted repeats. Theor. Appl. Genet. 67:185-193.

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

vom Stein, J. and W. Hachtel. 1988. Deletions/insertions, short inverted repeats, sequences resembling *att*-lambda, and frame shift mutated open reading frames are involved in chloroplast DNA differences in the genus *Oenothera* subsection *Munzia*. Mol. Gen. Genet. 213:513-518.

Weston-Hafer, K. and D.E. Berg. 1989. Palindromy and the location of deletion endpoints in *Escherichia coli*. Genetics 121:651-658.

Wolfe, K.H. 1988. The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. Curr. Genet. 13:97-99.

Wolfson, R., K.G. Higgins, and B.B. Sears. 1991. Evidence for replication slippage in the evolution of *Oenothera* chloroplast DNA. Mol. Biol. Evol. 8:709-720.

Zurawski, G., M.T. Clegg, and A.H.D. Brown. 1984. Tne nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735-749.

CHAPTER 2

INVERTED REPEATS AND tRNA GENES ARE ASSOCIATED WITH ENDPOINTS OF A LARGE CHLOROPLAST DNA INVERSION WITHIN CONIFERS

ABSTRACT

We cloned and sequenced the borders of a 40-50 kb inversion that distinguishes the chloroplast genomes of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) and Monterey pine (*Pinus radiata* D. Don). The borders of the inversion lie near *atp*A on one end and *trn*G-UCC on the other, and contain large inverted repeat sequences. A 480 bp repeat at the Douglas-fir inversion borders contain intact copies of *psb*I and *trn*S-GCU. Related Monterey pine inversion border repeats are approximately 840 bp in length and contain *trn*S-GCU and non-coding sequence. Also associated with inversion borders are other adjacent tRNA genes and a partially duplicated *trn*G-UCC pseudogene in Monterey pine. We propose that homologous recombination between tRNA genes, with concomitant pseudo-tRNA gene formation, inversion, and creation of repeated sequences at inversion borders, caused the observed DNA rearrangements.

INTRODUCTION

Conifer chloroplast DNA (cpDNA) is characterized by several unusual features compared to most angiosperms. The genome lacks one copy of the large inverted repeat (Strauss et al 1988, Lidholm et al 1988, Raubeson and Jansen 1992), possesses dispersed repetitive DNA (Tsai and Strauss 1989), and is highly rearranged (Strauss et al 1988). Four other separate lineages of vascular plants lack the large inverted repeat, including several tribes of the legume family (Downie and Palmer 1992, Lavin et al 1990).

Most rearranged genomes differ from the ancestral land plant genome by only one or two large inversions (Downie and Palmer 1992, Palmer 1987). The mechanisms responsible for large inversions in chloroplast DNA are poorly known. Three large monocot inversions have been most intensively studied (Quigley and Weil 1985, Sugiura 1989, Downie and Palmer 1992). Common to the inversion endpoints are repeated sequences, at least one adjacent tRNA gene, and partially duplicated tRNA genes (Howe et al 1988). tRNA genes at inversion borders have also been found in *Marchantia* (Ohyama et al 1986), lettuce (*Lactuca sativa*) (Jansen and Palmer 1987), rice (*Oryza sativa*) (Hiratsuka et al 1989), and sunflower (Heyraud et al 1987). When dispersed repetitive DNA is observed in cpDNA, it tends to be most abundant in highly rearranged genomes (Palmer and Stein 1986, Palmer 1987). Tsai and Strauss (1989) studied dispersed repetitive DNA in the conifer Douglas-fir and found four repeat families to be associated with inversion endpoints. The goal of this study was to characterize one of the large conifer inversions in more detail to allow its mechanism and association with dispersed repetitive DNA to be clarified.

MATERIAL AND METHODS

Chloroplast DNA Isolation and Cloning

Chloroplast DNA was isolated from needles of single individuals of *Pseudotsuga menziesii* and *Pinus radiata* after Strauss et al (1988). DNA was digested with *Sst*I according to manufacturers suggestions and electrophoresed on 0.8% agarose gels in TAE buffer. Fragments containing three inversion borders (6.7 kb from *Pseudotsuga menziesii*; 8.5 kb and 7.6 kb from *Pinus radiata*) were excised from the gel under long wave UV light and eluted with the Centriluter micro-electroeluter (Amicon, Beverley, MA), ligated into dephosphorylated pUC-19 plasmid vector, and transformed into *E. coli* DH5α library efficiency competent cells (BRL) (Maniatis et al 1982). Clones are designated as PM6.7, PR8.5, and PR7.6. The fourth inversion border was contained within a previously cloned *Xba*I 3.8 kb Douglas-fir cpDNA fragment (clone PM3.8) (Tsai and Strauss 1989). Deletion subclones containing progressive 200 bp deletions were obtained of each clone using the Erase-A-Base System Kit (Promega).

DNA Sequencing

Deletion subclones surrounding the approximate location of each inversion endpoint were sequenced using two methods: the dideoxy chain termination method of Sanger et al (1977) applied to chemically denatured plasmid DNA (Chen and Seeburg 1985) using the Sequenase Version 2.0 sequencing protocol and kit reagents (United States Biochemical Corporation, Cleveland, Ohio); and automated plasmid sequencing using the Applied Biosystem Model 373 DNA Sequencer with dye-on primer technology at the Center for Gene Research and Biotechnology at Oregon State University. All sequencing was done with universal and reverse primers.

Sequence Analysis

Sequence data was entered into the Intelligenetics Suite programs via the computational molecular biology lab of the Center for Gene Research and Biotechnology. Consensus sequences were assembled using the Gel program and aligned with Genalign and Align. GenBank and EMBL databases were searched using IFIND.

RESULTS

A Dispersed Repeat Family is Associated with the Borders of a 40-50 kb cpDNA Inversion that Distinguishes Douglas-fir and Monterey pine

A three member dispersed repeat family was reported in the chloroplast genome of Douglas-fir (Tsai and Strauss 1989). They classified this group as family 1, and mapped its members to the 1.4, 2.7, and 3.8 kb *Xba*I fragments. We hybridized the 2.7 kb *Xba*I fragment to restriction enzyme digested Monterey pine cpDNA (data not shown) and found three areas of hybridization in its genome (to the 8.5, 22.9, and 7.6 kb *Sst*I fragments) (Figure 11). Four of the repeat members map to the endpoints of a 40-50 kb inversion that distinguishes the two genomes (identified by Strauss et al (1988)). The two Douglas-fir inversion borders (DFleft and right) reside in the 6.7 kb *Sst*I and 3.8 kb *Xba*I fragments, respectively, while Monterey pine inversion borders (MPleft and right) are located in the *Sst*I 7.6 and 8.5 kb fragments, respectively. We pinpointed the location of the inversion endpoints within these fragments by careful measurements of restriction maps of both genomes (see Strauss et al (1988) for restriction maps) and concentrated our DNA sequencing efforts in these areas. A repeat member also hybridizes near the center of the inversion in both genomes; we cloned and sequenced this repeat in Douglas-fir and found it was located next to the length mutation hotspot in *Pseudotsuga* (see Chapter 1).

Inversion Borders are Associated with Inverted Repeats and Gene Duplications

We sequenced 1,347 bp surrounding the left inversion border of Douglas-fir (from clone PM6.7) (see Appendix Bv for sequence); 3,184 bp surrounding the right inversion border (from clone PM3.8) (Appendix Bvi); and 2,801 bp (from clone

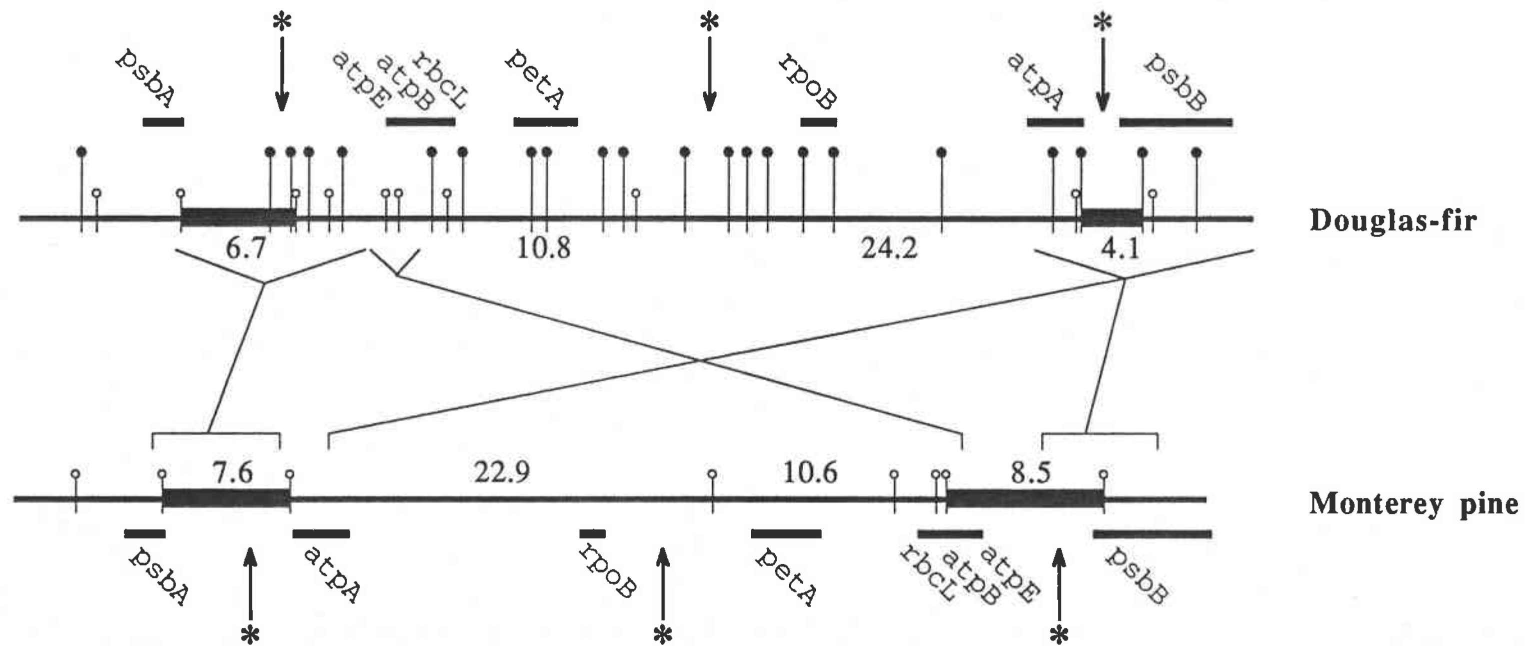


Figure 11. Restriction enzyme maps of sections of the Douglas-fir and Monterey pine chloroplast genomes (modified from Strauss et al (1988)). Cross-hybridizing regions between genomes are indicated by connecting lines. Asterisks (*) denote locations of a dispersed repeat family; solid boxed fragments were determined to contain inversion borders and cloned. Sizes (kb) are given for some *Sst*I fragments. ⫯ = *sst*I, ⫯ = *xba*I.

PR7.6) and 2,982 bp (from clone PR8.5) surrounding the left and right inversion borders, respectively, of Monterey pine (Appendices Bvii and Bviii, respectively).

The inversion endpoints in Douglas-fir cpDNA are associated with 480 bp inverted repeats that contain *psbI* and *trnS*-GCU and share 98.9% sequence similarity with one another (Figure 12). The Douglas-fir repeats map directly adjacent to the inversion borders (where continuous homology between Douglas-fir and Monterey pine cpDNA is broken).

Inversion endpoints in Monterey pine cpDNA are associated with a larger version of the repeat family, also in inverted orientation. Repeats are 844 and 838 bp in length and contain only *trnS*-GCU and intergenic sequence. The 6 bp difference in size is attributed to one and two base pair insertions in the left repeat occurring approximately 225 bp upstream of *trnS* (near the inversion border). *psbI* was not duplicated at the right inversion border of Monterey pine as it was in Douglas-fir. However, a portion of the split *trnG*-UCC gene was duplicated inside the Monterey pine right inversion border repeat. Exon 1 at MPleft shares 100% sequence similarity to the tobacco *trnG*-UCC gene while exon 1 at MPright shares only 82.6%. About 40 bp of the intron was duplicated at MPright and there is little sequence similarity between either intron sequence and the tobacco gene.

Dispersed Repeat Members Share High Sequence Similarity

The dispersed repeat family members near the four inversion borders and the Douglas-fir length mutation hotspot (Chapter 1) share high sequence similarity (Figure 13A). The Douglas-fir inversion border repeats (at DFleft and right) are 482 bp in length and share 98.9% sequence similarity (Figure 13B). The repeat at DFleft shares an additional 5' 148 bp with the repeat near the Douglas-fir hotspot (DFmiddle) and 98.6% sequence similarity; the repeat at DFright shares instead an extra 3' 44 bp with

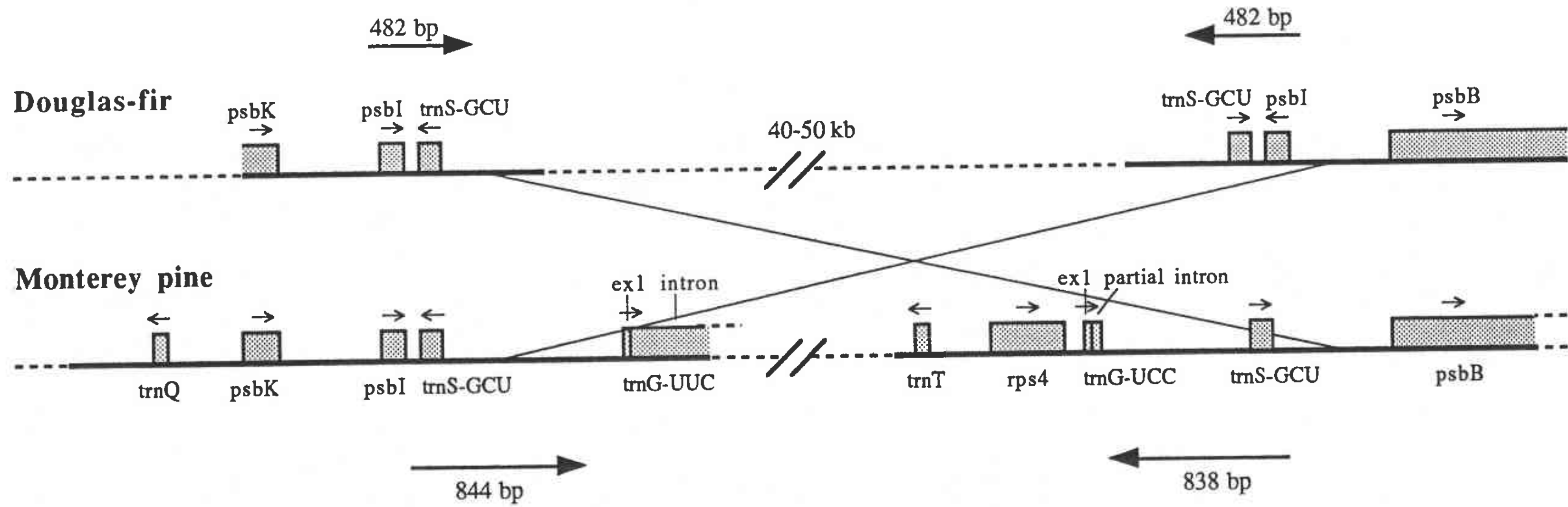
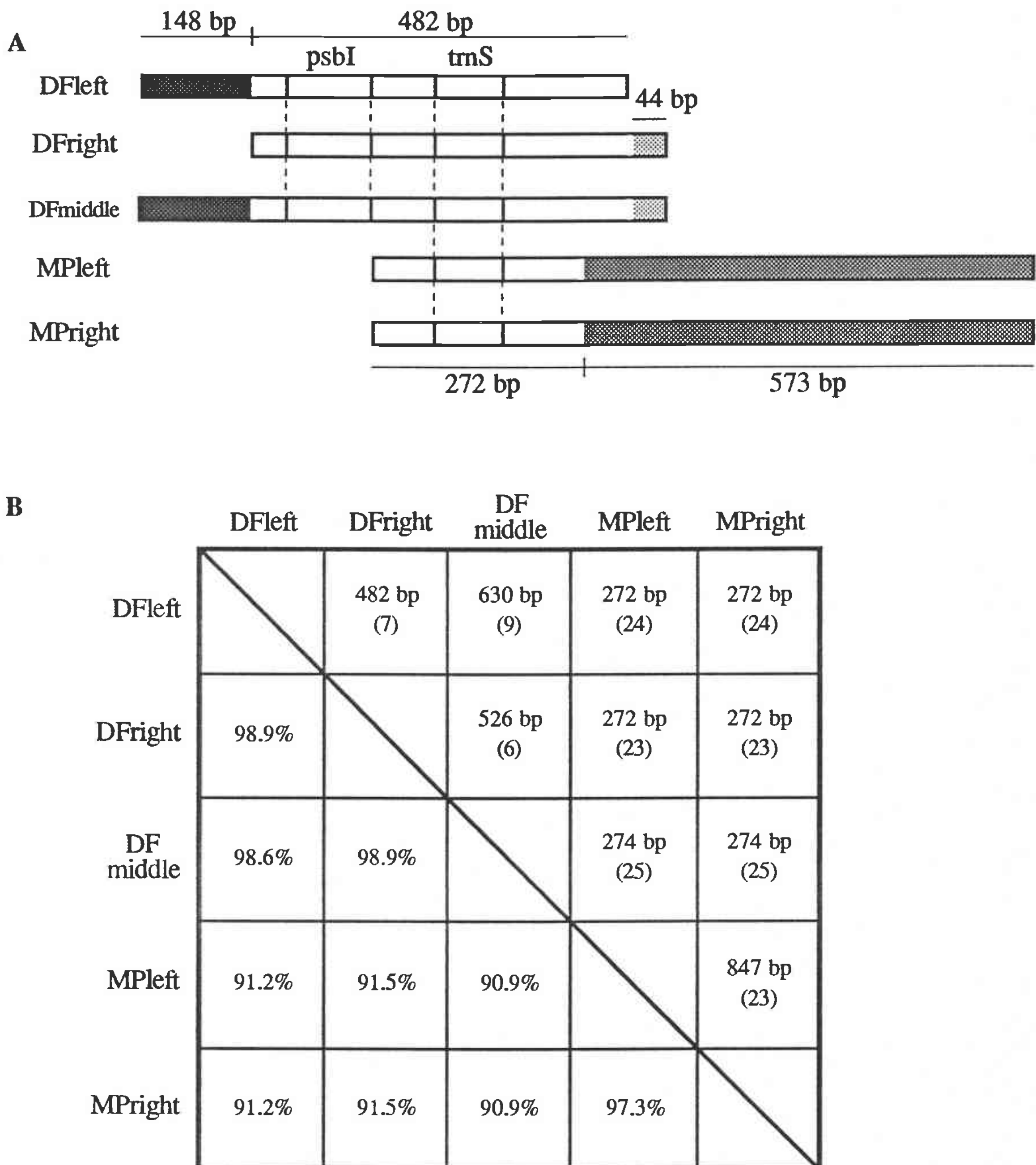


Figure 12. Endpoints of a 40-50 kb inversion that distinguishes the chloroplast genomes of Douglas-fir and Monterey pine (indicated by connecting crossed lines between the genomes). Arrows represent the location and length of a dispersed repeat family associated with inversion endpoints. Dashed lines indicate unsequenced regions. Small arrows above genes indicates direction of transcription.



	DFleft	DFright	DF middle	MPleft	MPright
DFleft		482 bp (7)	630 bp (9)	272 bp (24)	272 bp (24)
DFright	98.9%		526 bp (6)	272 bp (23)	272 bp (23)
DF middle	98.6%	98.9%		274 bp (25)	274 bp (25)
MPleft	91.2%	91.5%	90.9%		847 bp (23)
MPright	91.2%	91.5%	90.9%	97.3%	

Figure 13. Comparison of a dispersed repeat family in the chloroplast genomes of Douglas-fir and Monterey pine. (A) Five of the six repeat family members are represented (DF = Douglas-fir, MP = Monterey pine, left = dispersed repeat near left inversion border, right = dispersed repeat near right inversion border, middle = dispersed repeat in the middle of the inversion near the length mutation hotspot). Shaded and unshaded areas indicate regions of homology whose lengths are indicated. (B) Above diagonal: number of base pairs used in calculating sequence similarities between repeats. Number of base pair differences are in parentheses. Below diagonal: percent sequence similarity between homologous regions of repeat family members.

the DFmiddle repeat (98.9% similarity). Neither Monterey pine repeat contains the 44 bp region and, in fact, share between themselves a 3' 573 bp sequence. Sequence similarity was calculated between homologous regions of each repeat (Figure 13B). For example, when comparing MPleft to DFleft, only the unshaded areas shared by both repeats (272 bp) was considered. The Douglas-fir repeats share over 98.5% sequence similarity among each other, the Monterey pine repeats are over 97.3% similar, and the Monterey pine repeats compared to the Douglas-fir repeats are between 90.9 and 91.5% similar.

Sequence similarity of the genic regions between dispersed repeat members was high (Figure 14). *psb*I shares 98.2% similarity between the inversion border copies and 99.1% between each border repeat and the hotspot repeat. *trn*S is 100% similar between Douglas-fir repeats and between Monterey pine repeats, but 98.9% similar between the two genomes. A possible transposon-footprint sequence is located at the ends of the repeat at DFleft (position 28-43; 638-654) and at DFmiddle (28-43; 640-649), and was first reported in Tsai and Strauss (1989). However, the DFmiddle repeat is missing one of the flanking direct repeats, the DFright repeat does not contain either 5' repeat, and both DFmiddle and DFright repeats are missing the 3' direct repeat. Neither Monterey pine repeat contains transposon-like sequences at their ends. Near the inversion junction (near position 350 in MPleft and right where the Monterey pine repeats lose homology with their Douglas-fir counterparts) is a 38 bp inverted repeat. A number of 3-4 bp direct repeats or tracts of single nucleotides flanking repeat ends are also present.

Repeats are in inverted orientation at inversion borders (Figure 15). Repeat copies at the right borders in both genomes lie inside the inversion endpoints; the Douglas-fir repeat copy at the left border lies outside the endpoint; and the Monterey pine repeat copy at the left border overlaps the inversion endpoint. The dispersed repeat

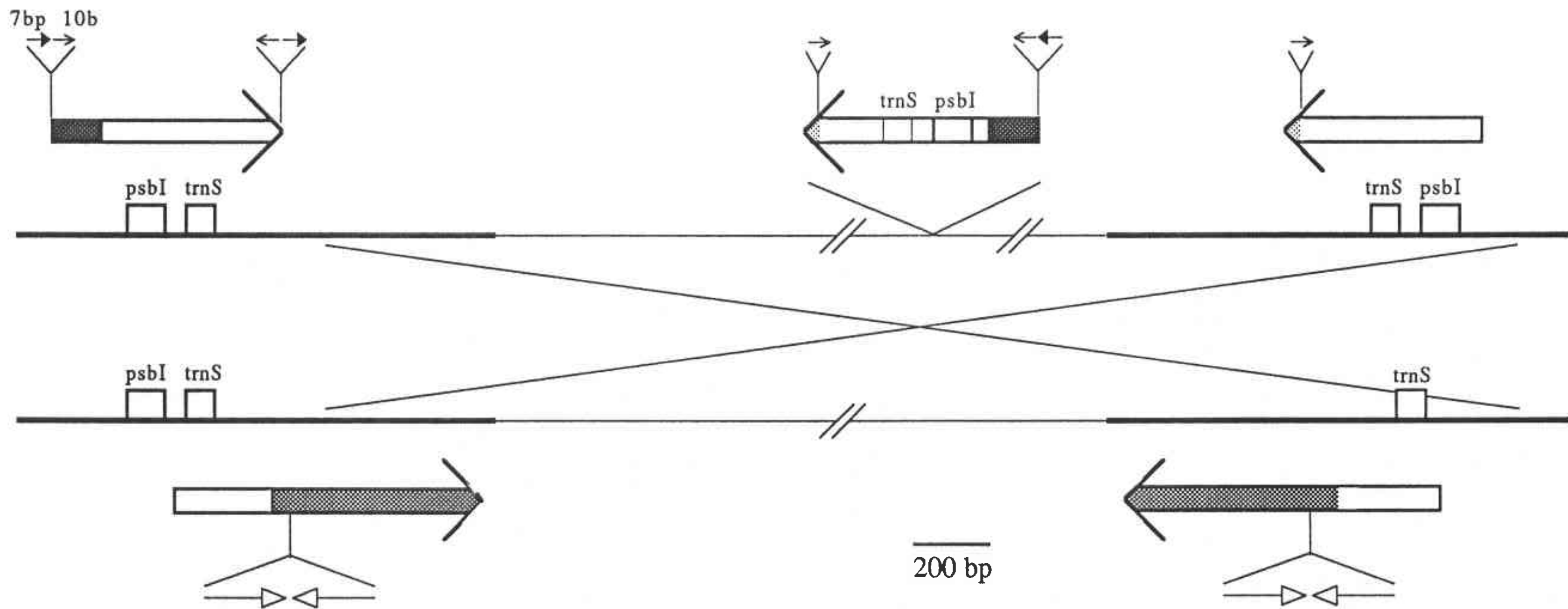
```
DFleft     1                                          atccaaaaaagaagaagggaagaaagaCCaTT
DFmiddle   1                                          tatcgtaatctttctttttattggtttCC TT

DFleft    33 TTG AAAACAAAGGGATAAgTTATCTCCTTCTTTCCAATTTCTTTTCACACGCACGTGATC
DFmiddle  32 TTGgAAAACAAAGGGATAAaTTATCTCCTTCTTTCCAATTTCTTTTCACACGCACGTGATC

DFleft    93 T GAGAAATAATTTCGTGATTTGTATGAATCATACTATTGCTTGGTATTCAAGTATCCATA
DFmiddle  93 TgGAGAAATAATTTCGTGATTTGTATGAATCATACTATTGCTTGGTATTCAAGTATCCATA

DFleft   154 TAcGgTACAAAGATTGATGATCTATTCTGTTGTACTTATAATCAGGATCCTGGAGATTACG
DFmiddle 154 TAtGaTACAAAGATTGATGATCTATTCTGTTGTACTTATAATCAGGATCCTGGAGATTACG
DFright    1                ccatttcttcatatTTCTGTTGTACTTAgAATCAGGATCCTaGAGATTACG

                                              psbI -->
DFleft   215 TAATGCTTACgCTTAAGCTGTTCGTTTACGCAGTAGTGATATTTTTCATTTCTCTTTTTAT
DFmiddle 215 TAATGCTTACTCTTAAGCTGTTCGTTTACGCAGTAGTGATATTTTTCATTTCTCTTTTTAT
DFright   52 TAATGCTTACTCTTAAGtTGTTCGTTTACGCAGTAGTGATATTTTTCATTTCTCTTTTTAT

DFleft   276 CTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGcGAAAAAAT
DFmiddle 276 CTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGTGAAAAAAT
DFright  113 CTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGTGAAAAAAT
MPright    1                                  tgaaaaaaaaggtcaaaataggAAGTATCT
                                                  psbI
MPleft     1                                  cggacgtaaagaatagtgaaaaAAGTATCT

DFleft   327 AGGTTAAgTAGTCTTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAAT
DFmiddle 327 AGGTTAATTAGTCTTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAAT
DFright  174 AGGTTAATTAGTCTTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAAT
MPright   31 AGGTTAATAGGTCTTTTCCGTTCCGTAGAAAGATTCGGGGTTATTCGTTTTCAGGATCAAT
MPleft    31 AGGTTAATAGGTCTTTTCCGTTCCGTAGAAAGATTCGGGGTTATTCGTTTTCgGGATCAAT

                                         <-- trnS-GCU
DFleft   388 AGTGACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTACTACGGATT
DFmiddle 388 AGTGACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTACTACGGATT
DFright  235 AGTGACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTACTACGGATT
MPright   92 AGTGGACGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATGATCCGTACTACGGATT
MPleft    92 AGTGGACGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATGATCCGTACTACGGATT
```

```
DFleft    449 AGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAAC
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DFmiddle  449 AGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAAC
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DFright   296 AGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAAC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
MPright   153 AGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTATAAC
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MPleft    153 AGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTATAAC

DFleft    510 AAAATGAATGGTGGAGTGAAGGTGTATACCATAGCATGTATGG  ATTGTATCGACAATGT
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
DFmiddle  510 AAAATGAATGGTGGAGTGAAGGTGTATACCATAGCATGTATGGggATTGTATCGACAATGT
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
DFright   357 AAAATGAATGGTGGAGTGAAGGTGTATACCATAGCATGTATGG  ATTGTATCGACAATGT
              |||||||||||||||| |||||||||||||||| |||||   |  ||||||||| |||| |
MPright   214 AAAATGAATGGTGGAGTAAAGGTGTATACCATAGTATGTACAG  ATTGTATCGGCAATAT
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
MPleft    214 AAAATGAATGGTGGAGTAAAGGTGTATACCATAGTATGTACAG  ATTGTATCGGCAATAT

DFleft    569 AATGAATAGGTCAATTATTTAGAGaAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTT
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DFmiddle  571 AATGAATAGGTCAATTATTTAGAGaAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTT
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
DFright   416 AATGAATAGGTCAATTATTTAGAGGAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTT
              ||||||||   |||||||| ||  | | |    | |  | | ||         | ||| ||
MPright   273 AATGAATATTGCAATTATTCAGTTGGATAAAGAACAATCCAGTCAATCATATTGTTCATTT
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MPleft    273 AATGAATATTGCAATTATTCAGTTGGATAAAGAACAATCCAGTCAATCATATTGTTCATTT

                        <--------  ---->
DFleft    630 CAAAATAATTCTTTTTCCcctatttcttct
              ||||||||||||||||||
DFmiddle  632 CAAAATAATTCTTTTTCCTGAACTAGAAAGCCTAGAATTATCATAACTTCTTTTTCAATGA
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DFright   477 CAAAATAATTCTTTTTCCTGAACTAGAAAGCCTAGAATTATCATAACTTCTTTTTCAATGA
              |   | ||  |  |  |  |        |||             |  ||
MPright   334 CGTTAAAAATAGTTCTGTATGACTGATTTTTTCTGCTTTTCTTGgcCTGGCCGATGGC AG
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||| ||
MPleft    334 CGTTAAAAATAGTTCTGTATGACTGATTTTTTCTGCTTTTCTTG  CTGGCCGATGGCcAG
                             ------------------------------------->   <----

DFmiddle  693 Agatgaaaaaaaaaagaaaaaaatg
              |
DFright   556 Aaaagaaatccttttcatatcgatt

MPright   394   CAAGAAAAGC G  AAAATC GtCATaC GTGCTTGACCTAATTTGATACCTAGAAAAA
                ||||||||| |   |||||| | ||| | ||||||||||||||||||||||||||||||
MPleft    393 gcCAAGAAAAGCaGaaAAAATCaGgCATgCaGTGCTTGACCTAATTTGATACCTAGAAAAA
              ------------------------------------

MPright   448 CTGCTGTAAAAGCAAGgcAAgcttGCTATCAAAAATTGGACTTCTATTGCCATATCTTCAT
              ||||||||||||||||  ||    |||||||||||||||||||||||||||||||||||||
MPleft    453 CTGCTGTAAAAGCAAG aAAagaaGCTATCAAAAATTGGACTTCTATTGCCATATCTTCAT

MPright   509 TCCCTCCCCAATCAGTTTGATTAAATGCGTTACATGGATTAGTCCATTTATTTATCTaacG
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |
MPleft    513 TCCCTCCCCAATCAGTTTGATTAAATGCGTTACATGGATTAGTCCATTTATTTATCTccaG

MPright   570 TATCCAATTTTATTATCTAGATATTGAAGGGTTCTCTATCTATTTAGGGTTCTCTATCTA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
MPleft    574 TATCCAATTTTATTATCTAGATATTGAAGGGTTCTCTATCTATTTAGGGTTCTCTATCTAt
```

```
MPright  630 TTTATGTATTATTGTAAATATATCAGTTGCTCAACGCCATAGGTTCCTGATCGAAACTACA
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 MPleft  635 TTTATGTATTATTGTAAATATATCAGTTGCTCAACGCCATAGGTTCCTGATCGAAACTACA


MPright  691 CCAATGGGTAGGAGTCCGAAGAAGACAAAATAGAAGAAAAGTGATTGATCCCGACAACATT
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 MPleft  696 CCAATGGGTAGGAGTCCGAAGAAGACAAAATAGAAGAAAAGTGATTGATCCCGACAACATT


MPright  752 TTATTCATACATTCAGTCAATGGAGGGTGAAAGAAAACCAAATGGATCTAGAAGTTATTGC
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 MPleft  757 TTATTCATACATTCAGTCAATGGAGGGTGAAAGAAAACCAAATGGATCTAGAAGTTATTGC


                                                          ---> ---> --->
MPright  823 GCAGCTCACTGTTCTGACTCTGATGGTTGTATttGGCCCTTcAGTTATagatagatagagc
             ||||||||||||||||||||||||||||||||  ||||||| ||||||
 MPleft  828 GCAGCTCACTGTTCTGACTCTGATGGTTGTATcgGGCCCTTtAGTTATtgttttatcagca
                                                     --->    --->
```

Figure 14. Sequence alignment of dispersed repeat family members. Genes are bracketed, with direction of transcription indicated by arrows; possible transposon-footprints are boxed; a 38 bp inverted repeat MPleft and MPright is underlined with open arrows; and small direct repeats flanking repeat ends are indicated with arrows. Flanking sequence is presented in lower case, repeat sequences in upper case.

Douglas-fir



Monterey pine

Figure 15. Orientation of repeat family members at inversion borders in Douglas-fir (top) and Monterey pine (bottom) and at the Douglas-fir hotspot. Connecting lines between genomes indicate the approximate location of a 40-50 kb inversion. Transposon-footprint like sequences proposed by Tsai and Strauss (1989) in some Douglas-fir repeats are indicated with arrows (➝: 8bp direct repeats, →: 10bp inverted repeat). Thirty-eight bp inverted repeats within the Monterey pine dispersed repeat is represented with open arrows (—▷).

member near the middle of the inversion in Douglas-fir is located approximately 450 bp 'upstream' of the hotspot (Chapter 1). The orientation of the Monterey pine repeat located in the region homologous to the DFmiddle repeat is not known.

DISCUSSION

We found the borders of a 40-50 kb inversion distinguishing the chloroplast genomes of Douglas-fir and Monterey pine to share several features with endpoints of other chloroplast inversions. These characteristics include the association of inversion borders with repeated sequences and intact or pseudo-tRNA genes. Inversion border repeats in Douglas-fir are 480 bp in length, inverted in orientation, and contain a protein coding gene (*psbI*) and tRNA gene (*trn*S-GCU). The repeat at the Monterey pine borders are longer (approximately 840 bp), carry only *trn*S-GCU and non-coding sequence, and are also in inverted orientation. The Monterey pine repeats are directly adjacent to *trn*G-UCC at one border and a partially duplicated *trn*G-UCC at the other border.

Douglas-fir repeats are >98.6% similar to one another, while Monterey pine repeats share 97.3% sequence similarity. It appears that an active gene conversion mechanism exists that is maintaining sequence identity between repeats within a molecule. If gene conversion is operating in these genomes, we expect that repeats within a molecule (the repeats at the left and right inversion borders of Douglas-fir, for example) would share greater similarity than homologous repeats between species (ie. the repeat at the left inversion border of Monterey pine and the right inversion border of Douglas-fir). This is what we observe. Sequence similarity between the Douglas-fir inversion border repeats is 98.9%, and similarity between the left Monterey pine repeat and right Douglas-fir repeat is 91.5%. Similarity between any Monterey pine and Douglas-fir repeat is no greater than 91.5%. Gene conversion has also been invoked to explain the conservation and maintenance of a *rpl*23 pseudogene in wheat (*Triticum aestivum*) and maize (*Zea mays*) cpDNA (Bowman et al 1988, Ogihara et al 1991), and

the preservation of sequence identity between the large inverted cpDNA repeats (Lemieux et al 1990).

Characterized repeated sequences in higher plant cpDNA usually include simple or tandem repeat sequences (Zurawski et al 1984, see also Chapter 1), partial tRNA genes (Bonnard et al 1985), and non-coding sequence (Howe 1985). Very rarely are repeated protein coding sequences noted. Exceptions are the *psb*I duplications in Douglas-fir, and non-functional dispersed copies of the ribosomal protein genes *rpl*2 and *rpl*23 in the chloroplast genome of wheat (Bowman et al 1988).

Repeated sequences are often associated with inversion endpoints (Howe et al 1988, Tsai and Strauss 1989, Shimada and Sugiura 1989), as is the presence of partially duplicated and adjacent tRNA genes (Marechal-Drouard et al 1991). Repeated sequences have been found at the inversion borders of wheat (Quigley and Weil 1985, Howe 1985, Howe et al 1988), though they are much smaller than those we observe in conifers. A 70 bp repeat associated with the endpoints of a 20 kb inversion contain a chimeric pseudo-tRNA gene comprised of the 3' sequence from *trn*G-UCC and the 5' sequence from *trn*fM. Two smaller inversions (approximately 6 and 3 kb) also contain repeats at their borders, though they are much smaller (<20 bp) (Howe et al 1988). However, not all inversion borders are near tRNA genes or repeated sequences (Howe et al 1988).

A possible mechanism of rearrangement proposed in the wheat genome is homologous recombination between the 70 bp repeats. This recombination event may have also led to the duplication of the pseudo-tRNA gene (Howe 1985). Another possibility is that tRNA genes are directly involved in the rearrangements (Howe et al 1988). In the chloroplast genome of rice, an inversion was explained by an intermolecular recombination event between 14 bp homologous regions of two different tRNA genes which gave rise to a chimeric tRNA pseudogene and the resulting inversion

(Hiratsuka et al 1989). In the rearranged rice genome, eight tRNA or ribosomal protein pseudogenes are present in the large single copy region, six of which cluster near inversion endpoints. These pseudogenes also exist in wheat whereas tobacco cpDNA only has one pseudogene (Shimada and Sugiura 1989).

tRNA genes have also been involved in rearrangements in mitochondrial and nuclear genomes. In the mitochondrial DNA of wheat, the *trn*P gene has been duplicated. The repeated gene appears to be derived from multiple intragenomic, site-specific rearrangements followed by amplification, fixation, and sequence divergence (Joyce et al 1988). tRNA genes have been shown to take part in rearrangements of animal mtDNA. These genes are typically dispersed throughout the genome, but their locations differ among orders (although in sea urchins, tRNA genes cluster near the replication origin) (Birky 1989). In eukaryotic DNA, tRNA genes are associated with repetitive elements (Hofmann et al 1991), and appear to act as landmarks for the integration of transposable elements in lower eukaryotes (Marschalek et al 1989).

Several mechanisms can be proposed to explain our 40-50 kb conifer inversion. In doing so, we have to account for both the spread of the repeat throughout the genome and the inversion itself. The repeat is duplicated three times in both the Douglas-fir and Monterey pine genomes; once at each inversion border and once in the middle of the inversion. The centrally located region in Douglas-fir is adjacent to a *Pseudotsuga* length mutation hotspot (see Chapter 1) and also served as a conifer/Petunia inversion border (Strauss et al 1988). Tsai and Strauss (1989) partially sequenced the repeat at one Douglas-fir inversion border (DFleft) and at the hotspot and found both ends to have a transposon-like combination of short imperfect short direct (8 bp) and inverted (10 bp) repeats, although the dispersed repeat near the hotspot was missing one of the transposon-like direct repeats. They proposed the spread of the repeat occurred by transposition which then mediated the inversions via homologous recombination. Upon

sequencing the other conifer inversion borders, we found that the second Douglas-fir border repeat contained only one of the transposon-like inverted repeats and that these sequences did not always occur at repeat ends. The Monterey pine repeats contain no noticeable transposon-footprint sequences at their ends. It is possible that, if the repeat did arise through transposition events, is was subsequently rearranged and lost much of its initial structure.

A common feature of most repeat ends are tracts of single nucleotides or several 3-4 bp direct repeats (see figure 14). This may be the result of slipped-strand mispairing during DNA replication or repair at the time of repeat spread or inversion. A 38 bp inverted repeat exists near the center of each Monterey pine inversion border repeat. Their location in the repeats is near where the Monterey pine and Douglas-fir repeats lose homology. It is unlikely that the inverted repeats are acting as target sites for recombination based on their internal placement with the Monterey pine repeats and their absence in the Douglas-fir repeats. However, they may be a signal for replication slippage because of their ability to fold into stable stem-loop structures. Cruciform structures play a role in the initiation of DNA replication. They may have instead been formed at the time of rearrangement.

Many chloroplast inversions in other plant genomes appear to share common borders. The sequence near *atp*A serves as an endpoint for a small 2.5 - 5.2 kb *Pisum* inversion (Palmer et al 1985), a 22 kb inversion shared by 57 genera in the Asteraceae (Jansen and Palmer 1987), a 28 kb grass family (Poaceae) inversion (Doyle et al 1992), a 50 kb inversion shared by mungbean and *Oenothera* (Howe 1985, Palmer and Thompson 1982), and one end of the Douglas-fir/Monterey pine inversion (Strauss et al 1988). *atp*A is located near several tRNA genes (*trn*R, *trn*G-UCC, *trn*S-GCU, and *trn*Q-UUG) in the ancestral-like chloroplast genome. A 23.5 kb inversion in soybean lies between *trn*E and *trn*T on one side, and *trn*S-GUC and *trn*G-UCC on the other

(Heyraud et al 1987). *trn*S-GUC has been duplicated in each conifer repeat (100% similarity of *trn*S-GUC within species, 98.9% between species, and 89.9 and 90.9% between tobacco and Monterey pine and tobacco and Douglas-fir, respectively) and *trn*G-UCC borders the Monterey pine inversion-- at one end in pseudogene form. *trn*E lies between the *Pseudotsuga* length mutation hotspot (see Chapter 1) and one copy of the dispersed repeat. We also observed the presence of *trn*T at one Monterey pine inversion border. Two of the largest wheat inversions have one border which is also near *trn*G-UCC and one 6 kb inversion border between *trn*G-UCC and *trn*S-GCU. The smallest wheat inversion borders *trn*E on one side and *trn*T on the other (also common to soybean and conifers, as discussed).

A common mechanism may be responsible for many of the characterized inversions sharing borders. If transposition were involved (ie. the repeat was spread to these specific locations via transposition and inversion was mediated via homologous recombination), we would expect to see similar repeat structures at inversion borders. In fact, similar repeat structures (ie. homologous repeats) are not observed at borders. However, if these locations represent highly recombinogenic areas, we may expect to observe similar characteristics, instead of specific repeats, at inversion borders (such as the observed proximity to tRNA genes, tRNA pseudogenes, and some apparently unrelated repeated sequences). It may be that homologous recombination between tRNA genes is responsible for most rearrangements in the chloroplast genome. Intermolecular recombination between tRNA genes can account for the origin of pseudogenes, inversions, and the creation of repeated sequences near the inversion endpoints (Sugiura 1989). It is possible in the case of conifers cpDNA evolution that duplicated sequences are being created and spread at the time of inversion by a common mechanism involving tRNA genes at these highly recombinogenic areas in the genome.

REFERENCES

Birky, C.W.Jr. 1989. Organelle evolution. Genome 31:1095-1097.

Bonnard, G., J.-H. Weil, and A. Steinmetz. 1985. The intergenic region between the *Vicia faba* chloroplast tRNA-Leu (CAA) and tRNA-Leu (UAA) genes contains a partial copy of the split tRNA-Leu (UAA) gene. Curr. Genet. 9:417-422.

Bowman, C.M., R.F. Barker, and T.A. Dyer. 1988. In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination. Curr. Genet. 14:127-136.

Chen, E.Y. and P.H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165-170.

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

Doyle, J.J., J.I. Davis, R.J. Soreng, D. Garvin, and M.J. Anderson. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc. Natl. Acad. Sci. USA 89:7722-7726.

Heyraud, F., P. Serror, M. Kuntz, A. Steinmetz, and P. Heizmann. 1987. Physical map and gene localization on sunflower (*Helianthus annuus*) chloroplast DNA: evidence for an inversion of a 23.5-Kbp segment in the large single copy region. Plant Mol. Biol. 9:485-496.

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.-R. Sun, B.Y. Meng, Y.-Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid inversion during the evolution of the cereals. Mol. Gen. Genet. 217:185-194.

Hofmann, J., G. Schumann, G. Borschet, R. Gosseringer, M. Bach, W.M. Bertling, R. Marschalek, and T. Dingermann. 1991. Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5' and 3'-flanking regions. J. Mol. Biol. 222:537-552.

Howe, C.J. 1985. The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to *att*-lambda. Curr. Genet. 10:139-145.

Howe, C.J., R.F. Barker, C.M. Bowman, and T.A. Dyer. 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. 13:343-349.

Jansen, R.K. and J.D. Palmer. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Acad. Sci. USA 84:5818-5822.

Joyce, P.B.M., D.F. Spencer, and M.W. Gray. 1988. Multiple sequence rearrangements accompanying the duplication of a tRNA-Pro gene in wheat mitochondrial DNA. Plant Mol. Biol. 11:833-843.

Lavin, M., J.J. Doyle, and J.D. Palmer. 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution 44:390-402.

Lemieux, B., M. Turmel, and C. Lemieux. 1990. Recombination of *Chlamydomonas* chloroplast DNA occurs more frequently in the large inverted repeat sequence than in the single-copy regions. Theor. Appl. Genet. 79:17-27.

Lidholm, J., A.E. Szmidt, J.-E. Hallgren, and P. Gustafsson. 1988. The chloroplast genome of conifers lack one of the rRNA-encoding inverted repeats. Mol. Gen. Genet. 212:6-10.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marechal-Drouard, L. M. Kuntz, and J.H. Weil. 1991. tRNAs and tRNA genes of plastids. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 169-189.

Marschalek, R., T. Brechner, E. Amon-Bohm, and T. Dingermann. 1989. Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science 244:1493-1496.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1991. Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. Genetics 129:873-884.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki. 1986. Complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Plant Mol. Biol. Rep. 4:148-175.

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

Palmer, J.D. and D.B. Stein. 1986. Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 10:823-833.

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

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.

Quigley, F. and J.H. Weil. 1985. Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr. Genet. 9:495-503.

Raubeson, L.A. and R.K. Jansen. 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. Biochem. Syst. Ecol. 20:17-24.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Shimada, H. and M. Sugiura. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr. Genet. 16:293-301.

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.

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Annu. Rev. Cell Biol. 5:51-70.

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

Zurawski, G., M.T. Clegg, and A.H.D. Brown. 1984. Tne nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735-749.

CONCLUSIONS

The main findings of this thesis include:

- an inversion and length mutation hotspot in the chloroplast genomes of Douglas-fir and Monterey pine are associated with repeated sequences and tRNA genes,

- a dispersed repeat hundreds of basepairs in length exists in three copies in each genome; at the borders of the 40-50 kb inversion, and adjacent to a length mutation hotspot in *Pseudotsuga*,

- the dispersed repeat contains *trn*S-GCU in Monterey pine, and *trn*S-GCU and *psb*I in Douglas-fir,

- the *Pseudotsuga* length mutation hotspot consists of a partial *trn*Y gene duplication and variable numbers of imperfect, tandem direct repeats resembling *trn*Y gene sequence, and

- Douglas-fir and Monterey pine inversion endpoints are associated with inverted repeats, adjacent tRNA genes, and partial tRNA genes.

In many species, similar regions of the chloroplast genome have been involved in independent rearrangements. Many inversion borders in land plant cpDNA, including those characterized in Douglas-fir and Monterey pine, are located in specific, shared regions, suggesting the existence of recombinational hotspots around the genome. Associated with the majority of inversion borders are repeated sequences and adjacent tRNA genes. Repeated sequences are usually tens of basepairs in length when present, unlike the repeats in Douglas-fir and Monterey pine which are several hundred

basepairs in size. More common to these regions is the presence of intact tRNA genes and partially duplicated tRNA genes (usually 20-70 bp). These sequences are often found adjacent to inversion endpoints and other genome rearrangements. tRNA duplications are not found in chloroplast genomes retaining the ancestral form, which suggests the duplication itself is involved in the rearrangement.

The tRNA duplication observed at the *Pseudotsuga* hotspot is not found in other studied land plant cpDNAs. The hotspot is located directly downstream of the intact *trn*Y gene. *trn*Y is cotranscribed with *trn*E in cpDNA, and thus is relatively conserved. The *trn*S-GCU/*psb*I containing dispersed repeat is located upstream of the *trn*E/*trn*Y genes by several hundred basepairs. *trn*S-GCU has been duplicated three times in each conifer genome studied; adjacent to the *Pseudotsuga* hotspot, and the Monterey pine and Douglas-fir inversion borders. *trn*S-GCU borders the cpDNA inversion endpoints of many species. These conifer repeats do not appear to exist in species containing ancestral-type genomes, suggesting that the duplication took place at the time of rearrangements. Duplications of intact cpDNA genes is rarely reported.

The role tRNA genes may be playing in genome rearrangements is speculative. tRNA genes have limited primary sequence homology (mostly around the -GTTC- sequence present in all tRNA genes), and contain no apparent recombinogenic sequence such as Chi sites. It is possible that the tRNA genes are providing a common secondary structure recognized by recombination enzymes. In animal mtDNA, the association of rearrangements with tRNA genes, their intragenomic mobility relative to other genes, and their similar secondary structures suggests they may facilitate rearrangements. tRNA genes may move by processes similar to splicing, which could rearrange adjacent parts of the genome in the process. Movement of tRNA genes toward clustering also occurs in land plant chloroplast DNA.

Closer examination of rearrangements is needed to discern the role that tRNA genes are playing in chloroplast genome evolution. Sequence information of additional length mutations and inversions, especially of more closely related species, would be useful to verify common characteristics of these sites and the mechanisms acting upon them. Chloroplast transformation will be useful to test the recombinogenic potential of various sequences. A better grasp of mechanisms governing cpDNA recombination will aid in our understanding of chloroplast genome evolution.

BIBLIOGRAPHY

Akada, S., Y.Q. Xu, H. Machii, and S.D. Kung. 1990. Stem-loop structures at the 3' end of tobacco Rubisco large subunit mRNA. Gene 94:195-199.

Albertini, A.M., M. Hofer, M.P. Calos, and J.H. Miller. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29:319-328.

Aldrich, J., B.W. Cherney, E. Merlin, and L. Christopherson. 1988a. The role of insertion/deletions in the evolution of the intergenic region between *psb*A and *trn*H in the chloroplast genome. Curr. Genet. 14:137-146.

Aldrich, J., B.W. Cherney, C. Williams, and E. Merlin. 1988b. Sequence analysis of the junction of the large single copy region and the large inverted repeat in the petunia chloroplast genome. Curr. Genet. 14:487-492.

Aldrich, J., B. Cherney, E. Merlin, C. Williams, and L. Mets. 1985. Recombination within the inverted repeat sequences of the *Chlamydomonas reinhardii* chloroplast genome produces two orientation isomers. Curr. Genet. 9:233-238.

Ali, I.F., D.B. Neale, and K.A. Marshall. 1991. Chloroplast DNA restriction fragment length polymorphism in *Sequoia sempervirens* D. Don Endl., *Pseudotsuga menziesii* (Mirb.) Franco, *Calocedrus decurrens* (Torr.), and *Pinus taeda* L. Theor. Appl. Genet. 81:83-89.

Alosi, M.C. and D.B. Neale. 1992. Light- and phytochrome-mediated gene expression in Douglas-fir seedlings. Physiol. Plant. (in press).

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.

Bendich, A.J. 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays 6:279-282.

Bendich, A.J. and S.B. Smith. 1990. Moving pictures and pulsed-field gel electrophoresis show linear DNA molecules from chloroplasts and mitochondria. Curr. Genet. 17:421-425.

Bennoun, P. 1982. Evidence for a respiratory chain in the chloroplast. Proc. Natl. Acad. Sci. USA 79:4352-4356.

Berry-Lowe, S.L. and G.W. Schmidt. 1991. Chloroplast protein transport. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 257-302.

Birky, C.W.Jr. 1978. Transmission genetics of mitochondria and chloroplasts. Ann. Rev. Genet. 12:471-512.

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. *In* Plant Evolutionary Biology, eds. L.D. Gottlieb and S.K. Jain. Chapman and Hall, New York, pp. 23-53.

Birky, C.W.Jr. 1989. Organelle evolution. Genome 31:1095-1097.

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

Blasko, K., S.A. Kaplan, K.G. Higgins, R. Wolfson, and B.B. Sears. 1988. Variation in copy number of a 24-base pair tandem repeat in the chloroplast DNA of *Oenothera hookeri* strain Johansen. Curr. Genet. 14:287-292.

Bogorad, L. 1991. Replication and transcription of plastid DNA. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 93-124.

Bonham-Smith, P.C. and D.P. Bourque. 1989. Translation of chloroplast-encoded mRNA: potential initiation and termination signals. Nucl. Acids Res. 17:2057-2080.

Bonnard, G., J.-H. Weil, and A. Steinmetz. 1985. The intergenic region between the *Vicia faba* chloroplast tRNA-Leu (CAA) and tRNA-Leu (UAA) genes contains a partial copy of the split tRNA-Leu (UAA) gene. Curr. Genet. 9:417-422.

Bowman, C.M. and T.A. Dyer. 1986. The location and possible evolutionary significance of small dispersed repeats in wheat ctDNA. Curr. Genet. 10:931-941.

Bowman, C.M., R.F. Barker, and T.A. Dyer. 1988. In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination. Curr. Genet. 14:127-136.

Brewer, B.J. and W.L. Fangman. 1987. The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell 51:463-471.

Brown, J. and V. Sundaresan. 1991. A recombination hotspot in the maize A1 intragenic region. Theor. Appl. Genet. 81:185-188.

Cairns, J. 1963. The chromosome of *Escherichia coli*. Cold Spring Harbor Symp. Quant. Biol. 28:43-46.

Campbell, N.A. 1987. Biology. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Cedar, H. 1988. DNA methylation and gene activity. Cell 53:3-4.

Chen, E.Y. and P.H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4:165-170.

Chiu, W.-L. and B.B. Sears. 1985. Recombination between chloroplast DNAs does not occur in sexual crosses of *Oenothera*. Mol. Gen. Genet. 198:525-528.

Chiu, W.-L. and B.B. Sears. 1992. Electron microscopic localization of replication origins in *Oenothera* chloroplast DNA. Mol. Gen. Genet. 232:33-39.

Collin, S. and T.H.N. Ellis. 1991. Evidence for the presence of hairpin chloroplast DNA molecules in barley cultivars. Curr. Genet. 20:253-258.

Crouse, E.J., J.M. Schmitt, and H.-J. Bohnert. 1985. Chloroplast and cyanobacterial genomes, genes and RNAs: a compilation. Plant Mol. Biol. Reporter 3:43-89.

Curtis, S.E. and M.T. Clegg. 1984. Molecular evolution of chloroplast DNA sequences. Mol. Biol. Evol. 1:291-301.

Delp, G. and H. Kossel. 1991. rRNAs and rRNA genes of plastids. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 139-167.

Deng, X.-W., R.A. Wing, and W. Gruissem. 1989. The chloroplast genome exists in multimeric forms. Proc. Natl. Acad. Sci. USA 86:4156-4160.

dePamphilis, C.W. and J.D. Palmer. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337-339.

Doebley, J.F., D.P. Ma, and W.T. Renfroe. 1987. Insertion/deletion mutations in the *Zea* chloroplast genome. Curr. Genet. 11:617-624.

Dover, G.A. 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet. 2:159-165.

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

Doyle, J.J., J.I. Davis, R.J. Soreng, D. Garvin, and M.J. Anderson. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc. Natl. Acad. Sci. USA 89:7722-7726.

Dynan, W.S. 1989. Understanding the molecular mechanism by which methylation influences gene expression. Trends Genet. 5:35-36.

Eckdahl, T.T. and J.N. Anderson. 1990. Conserved DNA structures in origins of replication. Nucl. Acids Res. 18:1609-1612.

Eckhahl, T.T., J.L. Bennetzen, and J.N. Anderson. 1989. DNA structures associated with autonomously replicating sequences from plants. Plant Mol. Biol. 12:507-516.

Ellis, T.H.N. and A. Day. 1986. A hairpin plastid genome in barley. EMBO J. 5:2769-2774.

Farjon, A. 1990. Pinaceae: drawings and discription of the genera *Abies, Cedrus, Pseudolarix, Keteleeria, Nothotsuga, Tsuga, Cathaya, Pseudotsuga, Larix,* and *Picea.* Koeltz Scientific Books, Konigstein.

Fedoroff, N.V. 1983. Controlling elements in maize. *In* Mobile Genetic Elements, ed. J.A. Shapiro. Academic Press, New York. pp. 1-63.

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

Fincham, J.R.S. and P. Oliver. 1989. Initiation of recombination. Nature 338:14-15.

Freund, A.-M., M. Bichara, and R.P.P. Fuchs. 1989. Z-DNA-forming sequences are spontaneous deletion hot spots. Proc. Natl. Acad. Sci. USA 86:7465-7469.

Gillham, N.W. 1978. Organelle Heredity. Raven Press, New York.

Gillham, N.W., J.E. Boynton, and E.H. Harris. 1991. Transmission of plastid genes. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 55-92.

Gjetvaj, B., D.I. Cook, and E. Zouros. 1992. Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (Bivalvia: Pectinidae). Mol. Biol. Evol. 9:106-124.

Gold, B., N. Carrillo, K.K. Tewari, and L. Bogorad. 1987. Nucleotide sequence of a preferred maize chloroplast genome template for *in vitro* DNA synthesis. Proc. Natl. Acad. Sci. USA 84:194-198.

Golub, E.I. 1988. 'One minute' transformation of competent *E. coli* by plasmid DNA. Nucl. Acids Res. 16:1641.

Goodwin, T.W. and E.I. Mercer. 1983. Introduction to Plant Biochemistry, second edition. Pergamon Press, Oxford.

Govindaraju, D.R., B.P. Dancik, and D.B. Wagner. 1989. Novel chloroplast DNA polymorphism in a sympatric region of two pines. J. Evol. Biol. 2:49-59.

Govindaraju, D.R., D.B. Wagner, G.P. Smith, and B.P. Dancik. 1988. Chloroplast DNA variation within individual trees of a *Pinus banksiana-Pinus contorta* sympatric region. Can. J. For. Res. 18:1347-1350.

Gray, M.W. 1989. The evolutionary origins of organelles. Trends Genet. 5:294-299.

Gray, M.W. 1991. Origin and evolution of plastid genomes and genes. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 303-330.

Gray, M.W. and W.F. Doolittle. 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. 46:1-42.

Gray, P.W. and R.B. Hallick. 1978. Physical mapping of the *Euglena gracilis* chloroplast DNA and ribosomal RNA gene region. Biochem. 17:284-289.

Grindley, N.D.F. and R.R. Reed. 1985. Transpositional recombination in prokaryotes. Ann. Rev. Biochem. 54:863-896.

Gruissem, W. 1989a. Chloroplast RNA: transcription and processing. *In* The Biochemistry of Plants: A Comprehensive Treatise, Vol. 15, ed. A. Marcus. Academic Press, New York, pp. 151-191.

Gruissem, W. 1989b. Chloroplast gene expression: how plants turn their plastids on. Cell 56:161-170.

Halliwell, B. 1984. Chloroplast Metabolism. Oxford University Press, New York.

Harlow, W.M., E. S. Harrar, and F.M. White. 1978. Textbook of Dendrology. McGraw-Hill Book Company, New York.

Harris, E.H. 1989. The *Chlamydomonas* Sourcebook. Academic Press, San Diego.

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

Held, P.G. and N.H. Heintz. 1992. Review: eukaryotic replication origins. Biochim. Biophys. Acta 1130:235-246.

Hermann, R.K. 1982. The Genus *Pseudotsuga*: Historical Records and Nomenclature. Forest Research Laboratory, Oregon State University, Corvallis, Oregon.

Heyraud, F., P. Serror, M. Kuntz, A. Steinmetz, and P. Heizmann. 1987. Physical map and gene localization on sunflower (*Helianthus annuus*) chloroplast DNA: evidence for an inversion of a 23.5-Kbp segment in the large single copy region. Plant Mol. Biol. 9:485-496.

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.-R. Sun, B.Y. Meng, Y.-Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid inversion during the evolution of the cereals. Mol. Gen. Genet. 217:185-194.

Ho, I. and A.L. Rayburn. 1991. The relationship between chloroplast number and genome size in *Zea mays* ssp. *mays*. Plant Sci. 74:255-260.

Hofmann, J., G. Schumann, G. Borschet, R. Gosseringer, M. Bach, W.M. Bertling, R. Marschalek, and T. Dingermann. 1991. Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5' and 3'-flanking regions. J. Mol. Biol. 222:537-552.

Hoober, J.K. 1984. Chloroplasts. Plenum Press, New York.

Howe, C.J. 1985. The endpoints of an inversion in wheat chloroplast DNA are associated with short repeated sequences containing homology to *att*-lambda. Curr. Genet. 10:139-145.

Howe, C.J., R.F. Barker, C.M. Bowman, and T.A. Dyer. 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. 13:343-349.

Hsieh, C.-H., M. Wu, and J. Yang. 1991. The sequence-directed bent DNA detected in the replication origin of *Chlamydomonas reinhardtii* chloroplast DNA is important for the replication function. Mol. Gen. Genet. 225:25-32.

Huberman, J.A., L.D. Spotila, K.A. Nawotka, S.M. El-Assouli, and L.R. Davis. 1987. The in vivo replication origin of the yeast 2µm plasmid. Cell 51:473-481.

Hyrien, O., M. Debatisse, G. Buttin, and B.R. de Saint Vincent. 1987. A hotspot for novel amplification joints in a mosaic of *ALU*-like repeats and palindromic A+T-rich DNA. EMBO J. 6:2401-2408.

Jansen, R.K. and J.D. Palmer. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Acad. Sci. USA 84:5818-5822.

Jarman, A.P. and R.A. Wells. 1989. Hypervariable minisatellites: recombinators or innocent bystanders? Trends Genet. 511:367-371.

Johnson, L.B. and J.D. Palmer. 1989. Heteroplasmy of chloroplast DNA in *Medicago*. Plant Mol. Biol. 12:3-11.

Joyce, P.B.M., D.F. Spencer, and M.W. Gray. 1988. Multiple sequence rearrangements accompanying the duplication of a tRNA-Pro gene in wheat mitochondrial DNA. Plant Mol. Biol. 11:833-843.

Keegstra, K. 1989. Transport and routing of proteins into chloroplasts. Cell 56:247-253.

Klein, R.R. and J.E. Mullet. 1990. Light-induced transcription of chloroplast genes. J. Biol. Chem. 265:1895-1902.

Koller, B. and H. Delius. 1982. Origin of replication in chloroplast DNA of *Euglena gracilis* located close to the region of variable size. EMBO J. 1:995-998.

Koller, B. and H. Delius. 1984. Intervening sequences in chloroplast genomes. Cell 36:613-622.

Kolodner, R.D. and K.K. Tewari. 1975a. Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. Nature 256:708-711.

Kolodner, R. and K.K. Tewari. 1975b. Presence of displacement loops in the covalently closed circular chloroplast deoxyribonucleic acid from higher plants. J. Biol. Chem. 250:8840-8847.

Kolodner, R. and K.K. Tewari. 1975c. The molecular size and conformation of the chloroplast DNA from higher plants. Biochim. Biophys. Acta 402:372-390.

Kolodner, R. and K.K. Tewari. 1979. Inverted repeats in chloroplast DNA from higher plants. Proc. Natl. Acad. Sci. USA 76:41-45.

Kornberg, A. and T.A. Baker. 1992. DNA Replication. W.H. Freeman and Company, New York.

Kues, U. and U. Stahl. 1989. Replication of plasmids in gram-negative bacteria. Microbiol. Rev. 53:491-516.

Kung, S.D., Y.S. Zhu, and G.F. Shen. 1982. *Nicotiana* chloroplast genome. III. Chloroplast DNA evolution. Theor. Appl. Genet. 61:73-79.

Landy, A. and W. Ross. 1977. Viral integration and excision: structure of the lambda *att* sites. Science 197:1147-1160.

Laskey, R.A., M.P. Fairman, and J.J. Blow. 1989. S phase of the cell cycle. Science 246:609-614.

Lavin, M., J.J. Doyle, and J.D. Palmer. 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution 44:390-402.

Lax, A.R., K.C. Vaughn, S.O. Duke, and J.E. Endrizzi. 1987. Structural and physiological studies of a plastome cotton mutant with slow sorting out. J. Hered. 78:147-152.

Lemieux, B., M. Turmel, and C. Lemieux. 1990. Recombination of *Chlamydomonas* chloroplast DNA occurs more frequently in the large inverted repeat sequence than in the single-copy regions. Theor. Appl. Genet. 79:17-27.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4:203-221.

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

Lidholm, J. and P. Gustafsson. 1991b. A three-step model for the rearrangement of the chloroplast *trn*K-*psb*A region of the gymnosperm *Pinus contorta*. Nucl. Acids Res. 19:2881-2887.

Lidholm, J., A.E. Szmidt, J.-E. Hallgren, and P. Gustafsson. 1988. The chloroplast genome of conifers lack one of the rRNA-encoding inverted repeats. Mol. Gen. Genet. 212:6-10.

Liston, A. 1993. Use of the polymerase chain reaction to survey for the loss of the inverted repeat in the legume chloroplast genome. *In* Advances in Legume Systematics, Vol. 7, eds. J.J. Doyle and M.D. Crisp. (In press).

Lonsdale, D.M., T. Brears, T.P. Hodge, S.E. Melville, and W.H. Rottmann. 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Phil. Trans. R. Soc. Lond. B 319:149-163.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marechal-Drouard, L. M. Kuntz, and J.H. Weil. 1991. tRNAs and tRNA genes of plastids. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 169-189.

Margulis, L. 1981. Symbiosis in Cell Evolution. W.H. Freeman, San Francisco.

Marians, K.J. 1984. Enzymology of DNA replication in prokaryotes. CRC Crit. Rev. Biochem. 17:153-216.

Marschalek, R., T. Brechner, E. Amon-Bohm, and T. Dingermann. 1989. Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science 244:1493-1496.

Marunouchi, T., Y. Matsumoto, H. Hosoya, and K. Okabayashi. 1987. In addition to the ARS core, the ARS box is necessary for autonomously replicating sequences. Mol. Gen. Genet. 206:60-65.

Meeker, R., B. Nielsen, and K.K. Tewari. 1988. Localization or replication origins in pea chloroplast DNA. Mol. Cell. Biol. 8:1216-1223.

Milligan, B.G., J.N. Hampton, and J.D. Palmer. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol. Biol. Evol. 6:355-368.

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

Moon, E., T.-H. Kao, and R. Wu. 1987. Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes. Nucl. Acids Res. 15:611-630.

Morden, C.W. and S.S. Golden. 1989. *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature 337:382-385.

Moritz, C., T.E. Dowling, and W.M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18:269-292.

Mubumbila, M., E.J. Crouse, and J.H. Weil. 1984. Transfer RNAs and tRNA genes of *Vicia faba* chloroplasts. Curr. Genet. 8:379-385.

Mukai, Y., K. Tazaki, T. Fujii, and N. Yamamoto. 1992. Light-independent expression of three photosynthetic genes, *cab*, *rbc*S and *rbc*L, in coniferous plants. Plant Cell Physiol. 33:859-866.

Mullet, J.E. 1988. Chloroplast development and gene expression. Ann. Rev. Plant Physiol. Plant Mol. Biol. 39:475-502.

Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res. 8:4321-4325.

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

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

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

Newlon, C.S. 1988. Yeast chromosome replication and segregation. Microbiol. Rev. 52:568-601.

Newman, S.M., E.H. Harris, A.M. Johnson, J.E. Boynton, and N.W. Gillham. 1992. Nonrandom distribution of chloroplast recombination events in *Chlamydomonas reinhardtii*: evidence for a hotspot and an adjacent cold region. Genetics 132:413-429.

Newman, S.M., J.E. Boyton, N.W. Gillham, B.L. Randolph-Anderson, A.M. Johnson, and E.H. Harris. 1990. Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: molecular and genetic characterization of integration events. Genetics 126:875-888.

Nicolas, P., P. Ravel-Chapuis, P. Heizmann, and V.-M. Nigon. 1985. Unequal recombination between tandemly repeated chloroplast ribosomal operons. Curr. Genet. 10:229-238.

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.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc. Natl. Acad. Sci. USA 85:8573-8577.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1991. Molecular analysis of the hot spot region related to length mutations in wheat chloroplast DNAs. I. Nucleotide divergence of genes and intergenic spacer regions located in the hot spot region. Genetics 129:873-884.

Ogihara, Y., T. Terachi, and T. Sasakuma. 1992. Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. Curr. Genet. 22:251-258.

Ohta, N., N. Sato, S. Kawano, and T. Kuroiwa. 1991. Methylation of DNA in the chloroplasts and amyloplasts of the pea, *Pisum sativum*. Plant Sci. 78:33-42.

Ohtsubo, H., M. Umeda, and E. Ohtsubo. 1991. Organization of DNA sequences highly repeated in tandem in rice genomes. Jpn. J. Genet. 66:241-254.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki. 1986a. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572-574.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki. 1986b. Complete nucleotide sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Plant Mol. Biol. Rep. 4:148-175.

Palmer, J.D. 1983. Chloroplast DNA exists in two orientations. Nature 301:92-93.

Palmer, J.D. 1985a. Comparative organization of chloroplast genomes. Ann. Rev. Genet. 19:325-354.

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

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

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

Palmer, J.D. 1991. Plastid chromosomes: structure and evolution. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 5-53.

Palmer, J.D. and L.A. Herbon. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. 28:87-97.

Palmer, J.D. and D.B. Stein. 1986. Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 10:823-833.

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

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., J.M. Nugent, and L.A. Herbon. 1987a. Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc. Natl. Acad. Sci. USA 84:769-773.

Palmer, J.D., B. Osorio, J. Aldrich, and W.F. Thompson. 1987b. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr. Genet. 11:275-286.

Palmer, J.D., R.K. Jansen, H.J. Michaels, M.W. Chase, and J.R. Manhart. 1988. Chloroplast DNA variation and plant phylogeny. Ann. Missouri Bot. Gard. 75:1180-1206.

Price, R.A., J. Olsen-Stojkovich, and J.M. Lowenstein. 1987. Relationships among the genera of Pinaceae: an immunological comparison. Syst. Bot. 12:91-97.

Quail, P.H. 1991. Phytochrome: a light-activated molecular switch that regulates plant gene expression. Ann. Rev. Genet. 25:389-409.

Quigley, F. and J.H. Weil. 1985. Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr. Genet. 9:495-503.

Radding, C.M. 1982. Homologous pairing and strand exchange in genetic recombination. Ann. Rev. Genet. 16:405-437.

Rand, D.M. and R.G. Harrison. 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551-569.

Raubeson, L.A. and R.K. Jansen. 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. Biochem. Syst. Ecol. 20:17-24.

Ravel-Chapuis, P., P. Heizmann, and V. Nigon. 1982. Electron microscopic localization of the replication origin of *Euglena gracilis* chloroplast DNA. Nature 300:78-81.

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

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

Rochaix, J.-D. 1987. Molecular genetics of chloroplasts and mitochondria in the unicellular green alga *Chlamydomonas*. Microbiology Reviews 46:13-34.

Rodermel, S., P. Orlin, and L. Bogorad. 1987. The transcription termination region between two convergently-transcribed photoregulated operons in the maize plastid chromosome contains *rps*14, *trn*R(UCU) and a putative *trn*fM pseudogene. Nucl. Acids Res. 15:5493.

Rodermel, S.R. and L. Bogorad. 1985. Maize plastid photogenes: mapping and photoregulation of transcript levels during light-induced development. J. Cell Biol. 100:463-476.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning, A Laboratory Manual, 2nd edition. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schlunegger, B. and E. Stutz. 1984. The *Euglena gracilis* chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication. Curr. Genet. 8:629-634.

Schlunegger, B., M. Fasnacht, E. Stutz, B. Koller, and H. Delius. 1983. Analysis of a polymorphic region of the *Euglena gracilis* chloroplast genome. Biochim. Biophys. Acta 739:114-121.

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

Sears, B.B. 1983. Genetics and evolution of the chloroplast. *In* Stadler Symposium, Vol 15, University of Missouri, Columbia, pp.119-139.

Shaw, G.R. 1958. The Genus *Pinus*. The Murray Printing Company, Massachusetts.

Shimada, H. and M. Sugiura. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr. Genet. 16:293-301.

Shimada, H. and M. Sugiura. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucl. Acids Res. 19:983-995.

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. Kamagoshira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Suguira. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its organization and expression. EMBO J. 5:2043-2049.

Singer, B.S. 1988. On the role of homologous sequences in chromosomal rearrangements. Genes Develop. 2:1800-1811.

Smith, G.R. 1988a. Homologous recombination in procaryotes. Microbiol. Rev. 52:1-28.

Smith, G.R. 1988b. Homologous recombination sites and their recognition. *In* The Recombination of Genetic Material, ed. K.B. Low. Academic Press, New York. pp. 115-154.

Solignac, M., J. Genermont, M. Monnerot, and J.-C. Mounolou. 1987. *Drosophila* mitochondrial genetics: evolution of heteroplasmy through germ line cell divisions. Genetics 117:687-696.

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

Stayton, M.M., L. Bertsch, S. Biswas, P. Burgers, N. Dixon, J.E. Flynn, Jr., R. Fuller, J. Kaguni, J. Kobori, M. Kokaira, R. Low, and A. Kornberg. 1983. Enzymatic recognition of DNA replication origins. Cold Spring Harbor Symp. Quant. Biol. 47:693-700.

Stein, D.B., J.D. Palmer, and W.F. Thompson. 1986. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. Curr. Genet. 10:835-841.

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 C.-H. Tsai. 1988. Ribosomal gene number variability in Douglas-fir. J. Hered. 79:453-458.

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. 68:1502-1510.

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.

Subramanian, A.R., D. Stahl, and A. Prombona. 1991. Ribosomal proteins, ribosomes, and translation in plastids. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 191-215.

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Annu. Rev. Cell Biol. 5:51-70.

Sugiura, M. 1991. Transcript processing in plastids: trimming, cutting, splicing. *In* Cell Culture and Somatic Cell Genetics in Plants, Vol. 7, The Molecular Biology of Plastids, eds. L. Bogorad and I.K. Vasil. Academic Press, San Diego, pp. 125-137.

Sutton, B.C.S., D.J. Flanagan, J.R. Gawley, C.H. Newton, D.T. Lester, and Y.A. El-Kassaby. 1991. Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. Theor. Appl. Genet. 82:242-248.

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

Szmidt, A.E., Y.A. El-Kassaby, A. Sigurgeirsson, T. Alden, D. Lindgren, and J.-E. Hallgren. 1988. Classifying seedlots of *Picea sitchensis* and *P. glauca* in zones of introgression using restriction analysis of chloroplast DNA. Theor. Appl. Genet. 76:841-845.

Takeda, Y., H. Hirokawa, and T. Nagata. 1992. The replication origin of proplastid DNA in cultured cells of tobacco. Mol. Gen. Genet. 232:191-198.

Tassopula, D. and S.D. Kung. 1984. *Nicotiana* chloroplast genome. 6. Deletion and hot spot--a proposed origin of the inverted repeats. Theor. Appl. Genet. 67:185-193.

Tobin, E.M. and J. Silverthorne. 1985. Light regulation of gene expression in higher plants. Ann. Rev. Plant Physiol. 36:569-593.

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

Tsudzuki, J., K. Nakashima, T. Tsudzuki, J. Hiratsuka, M. Shibata, T. Wakasugi, and M. Sugiura. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of rps16. Mol. Gen. Genet. 232:206-214.

Turner, S., T. Burger-Wiersma, S.J. Giovannoni, L.R. Mur, and N.R. Pace. 1989. The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. Nature 337:380-382.

Umesono, K. and H. Ozeki. 1987. Chloroplast gene organization in plants. Trends Genet. 3:281-287.

Vallet, J.-M. and J.-D. Rochaix. 1985. Chloroplast origins of DNA replication are distinct from chloroplast ARS sequences in two green algae. Curr. Genet. 9:321-324.

Vaughn, K.C. 1981. Organelle transmission in higher plants: organelle alteration vs. physical exclusion. J. Hered. 72:335-337.

Vedel, F. and M. Delseny. 1987. Repetitivity and variability of higher plant genomes. Plant Physiol. Biochem. 25:191-210.

vom Stein, J. and W. Hachtel. 1988. Deletions/insertions, short inverted repeats, sequences resembling *att*-lambda, and frame shift mutated open reading frames are involved in chloroplast DNA differences in the genus *Oenothera* subsection *Munzia*. Mol. Gen. Genet. 213:513-518.

Waddell, J., X.-M. Wang, and M. Wu. 1984. Electron microscopic localization of the chloroplast DNA replicative origins in *Chlamydomonas reinhardii*. Nucl. Acids Res. 12:3843-3872.

Wagner, D.B., D.R. Govindaraju, C.W. Yeatman, and J.A. Pitel. 1989. Paternal chloroplast DNA inheritance in a diallel cross of jack pine (*Pinus banksiana* Lamb.). J. Hered. 80:483-485.

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 and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA 84:2097-2100.

Walker, G.C., L. Marsh, and L.A. Dodson. 1985. Genetic analysis of DNA repair: inference and extrapolation. Ann. Rev. Genet. 19:103-126.

Wallace, D.C. 1982. Structure and evolution of organelle genomes. Microbiol. Rev. 46:208-240.

Weeden, N.F. 1981. Genetic and biochemical implicatiions of the endosymbiotic origin of the chloroplast. J. Mol. Evol. 17:133-139.

Weston-Hafer, K. and D.E. Berg. 1989. Palindromy and the location of deletion endpoints in *Escherichia coli*. Genetics 121:651-658.

Whatley, J.M. 1982. Ultrastructure of plastid inheritance: green algae to angiosperms. Biol. Rev. 57:527-569.

White, E.E. 1990. Chloroplast DNA in *Pinus monticola*. I. physical map. Theor. Appl. Genet. 79:119-124.

Whitfeld, P.R. and W. Bottomley. 1983. Organization and structure of chloroplast genes. Ann. Rev. Plant Physiol. 34:279-310.

Wolfe, K.H. 1988. The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. Curr. Genet. 13:97-99.

Wolfe, K.H. and P.M. Sharp. 1988. Identification of functional open reading frames in chloroplast genomes. Gene 66:215-222.

Wolfe, K.H., C.W. Morden, and J.D. Palmer. 1992. Small single-copy region of plastid DNA in the non-photosynthetic angiosperm *Epifagus virginiana* contains only two genes. J. Mol. Biol. 223:95-104.

Wolfson, R., K.G. Higgins, and B.B. Sears. 1991. Evidence for replication slippage in the evolution of *Oenothera* chloroplast DNA. Mol. Biol. Evol. 8:709-720.

Wu, M., J.K. Lou, D.Y. Chang, C.H. Chang, and Z.Q. Nie. 1986. Structure and function of a chloroplast DNA replication origin of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA 83:6761-6765.

Yamamoto, N. Y. Mukai, M. Matsuoka, Y. Kano-Murakami, Y. Tanaka, Y. Ohashi, Y. Ozeki, and K. Odani. 1991. Light-independent expression of *cab* and *rbc*S genes in dark-grown pine seedlings. Plant Physiol. 95:379-383.

Yap, W.Y. and K.N. Kreuzer. 1991. Recombination hotspots in bacteriophage T4 are dependent on replication origins. Proc. Natl. Acad. Sci. USA 88:6043-6047.

Zhou, D.X., O. Massenet, F. Quigley, M.J. Marion, F. Moneger, P. Huber, and R. Mache. 1988. Characterization of a large inversion in the spinach chloroplast genome relative to *Marchantia*: a possible transposon-mediated origin. Curr. Genet. 13:433-439.

Zimmer, E.A., R.K. Hamby, M.L. Arnold, D.A. Leblanc, and E.C. Theriot. 1989. Ribosomal RNA phylogenies and flowering plant evolution. *In*: Fernholm, B., K. Bremer, and H. Jornvall, eds. The Hierarchy of Life. Elsevier Science Publishers, Amsterdam, pp. 205-214.

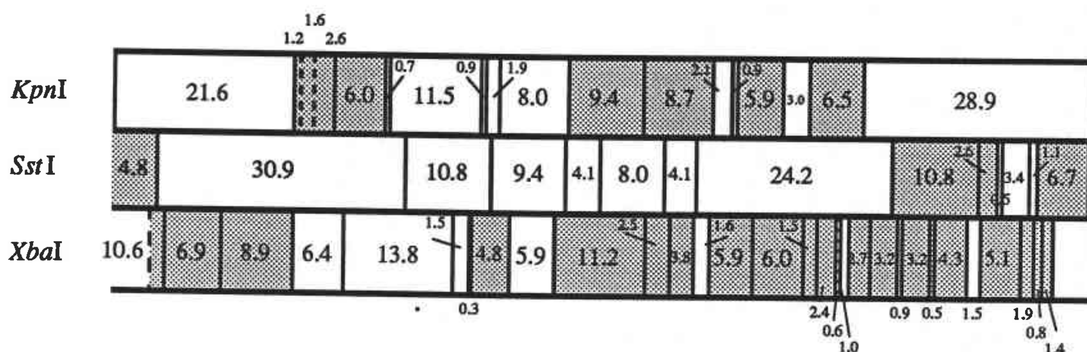
Zurawski, G. and M.T. Clegg. 1987. Evolution of higher-plant chloroplast DNA-encoded genes: implications for structure-function and phylogenetic studies. Ann. Rev. Plant Physiol. 38:391-418.

Zurawski, G., M.T. Clegg, and A.H.D. Brown. 1984. Tne nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics 106:735-749.

APPENDICES

APPENDIX A. ADDITIONAL DATA

i. Douglas-fir cpDNA Restriction Site Map



Douglas-fir restriction site map for three endonucleases (*Kpn*I, *Sst*I, *Xba*I). *Kpn*I and *Sst*I information is from Strauss et al (1988); *Xba*I map information in from Tsai and Strauss (1989). *Xba*I fragments were cloned by Tsai and Strauss; *Kpn*I and *Sst*I fragments by Hipkins following the protocol in Appendices Ci,ii,v,vi. Cloned fragments are indicated by shaded areas. Eighty seven percent of the Douglas-fir chloroplast genome has been cloned.

Corrections to map published in Tsai and Strauss (1989) (above is corrected version): (1) the left end of the *Xba*I 1.4 kb fragment lies 23 bp inside the left side of the *Sst*I 6.7 kb fragment (instead of the ~500 bp indicated); (2) there is a *Xba*I restriction site approximately 900 bp inside the right border of *Sst*I 4.8 kb (it is unknown whether the *Xba*I 6.9 kb fragment is actually 8.8 kb in size (6.9 + 0.9 kb), the *Xba*I 6.9 kb fragment should shift to the left by 900 bp, or a new 900 bp *Xba*I fragment is overlapping the *Sst*I 4.8 kb / *Sst*I 30.9 kb restriction site); and (3) the cloned 0.8 kb *Xba*I fragment available in the lab does not hybridize to the indicated map region, nor does the DNA sequence match sequence from the 6.7 kb *Sst*I fragment.

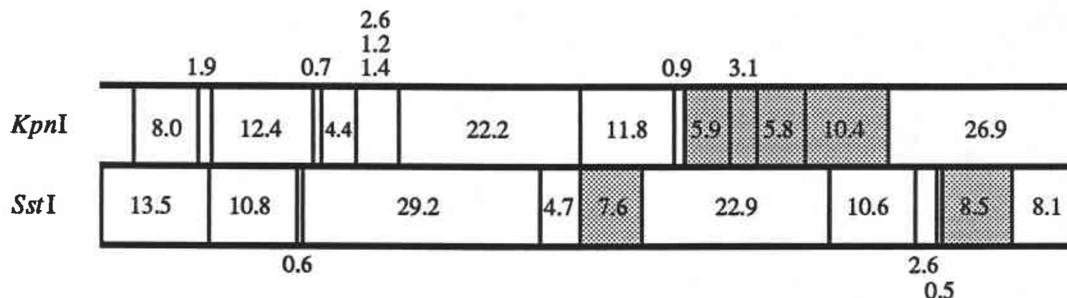
The goal of cloning *Kpn*I and *Sst*I fragments was to obtain a complete Douglas-fir cpDNA plasmid library. One region of approximately 10kb remained uncloned. This region (*Kpn*I 11.5 kb) contains the ribosomal RNA genes and may be difficult to clone due to secondary structure of the DNA. A second uncloned region is the *Xba*I 5.9 kb fragment-- between the 4.8 kb and 11.2 kb fragments. This was erroneously indicated in Tsai and Strauss (1989) as cloned (therefore no *Sst*I or *Kpn*I fragments were cloned in that area).

ii. Douglas-fir cpDNA Plasmid Library

All cpDNA fragments have been cloned into pUC19 and are stored as 50% glycerol stocks at -80°C. Clone names are those of samples stored in the lab.

Clone Name	Cloning Enzyme	Fragment Size (kb)
KpnI1.2DF	KpnI	1.2
KpnI1.6DF	KpnI	1.6
KpnI2.6DF	KpnI	2.6
KpnI6.0DF	KpnI	6.0
KpnI0.7DF	KpnI	0.7
KpnI9.4DF	KpnI	9.4
KpnI8.7DF	KpnI	8.7
KpnI5.9DF	KpnI	5.9
KpnI6.5DF	KpnI	6.5
SstI4.8DF	SstI	4.8
SstI10.8DF	SstI	10.8
SstI2.6DF	SstI	2.6
SstI6.7DF	SstI	6.7
Xba9	XbaI	6.9
Xba10	XbaI	8.9
Xba6	XbaI	4.8
Xba12	XbaI	11.2
XbaB	XbaI	2.5
Xba4	XbaI	3.8
Xba8-B1	XbaI	5.9
Xba8-A	XbaI	6.0
Xba1b	XbaI	1.5
XbaA	XbaI	2.4
Xba0.6	XbaI	0.6
Xba2	XbaI	2.7
Xba3A	XbaI	3.2
unknown	XbaI	0.9
Xba3B	XbaI	3.2
Xba0.5	XbaI	0.5
Xba5	XbaI	4.3
Xba7	XbaI	5.1
XbaC	XbaI	1.9
Xba0.7	XbaI	0.8
Xba1.4	XbaI	1.4

iii. Monterey Pine cpDNA Restriction Site Map



Monterey pine cpDNA restriction site map for two endonucleases (*Sst*I and *Kpn*I) (from Strauss et al 1988). Cloned fragments are indicated by cross-hatching and were cloned by Hipkins following the protocol in Appendices Ci, ii, v, vi.

The objective of cloning was to obtain fragments containing the Douglas-fir/Monterey pine inversion endpoints (the two *Sst*I fragments) and the area in Monterey pine homologous to the Douglas-fir hotspot (*Kpn*I fragments).

Correction compared to published map in Strauss et al (1988) (above map is correct version): the position of *Kpn*I fragments 3.1 kb and 5.8 kb have been reversed.

iv. Monterey Pine cpDNA Plasmid Library

All cpDNA fragments have been cloned into pUC19 and are stored as 50% glycerin stocks at -80°C. Clone names are those of samples stored in the lab.

Clone Name	Cloning Enzyme	Fragment Size (kb)
SstI8.5MP	SstI	8.5
SstI7.6MP	SstI	7.6
MP10.4KpnI	KpnI	10.4
MP5.8KpnI	KpnI	5.8
MP3.1KpnI	KpnI	3.1
MP5.9KpnI	KpnI	5.9

v. Douglas-fir cpDNA Origin of Replication (ORI)

Because the Douglas-fir cpDNA hotspot structurally resembles known chloroplast genome origins of replication, two experimental approaches were used (Southern analysis and *in vitro* replication) to investigate the possibility that the hotspot is functioning in this capacity.

(1) Goal: the pea cpDNA origins of replication have been mapped to a 12.7 kb *Pst*I fragment. If the hotspot was in fact the Douglas-fir cpDNA ori, it may hybridize to the pea 12.7 kb fragment.

Chloroplast DNA was extracted from garden pea seedlings (*Pisum sativum* cv Alaska) and digested with *Sal*I, *Sma*I, *Pst*I, and *SmaI/Pst*I. DNA was separated on a 0.8% agarose gel, Southern transferred to Zetabind membrane, and probed with the ^{32}P hexamer labeled Douglas-fir *Xba*2.7kb hotspot containing clone (see Appendices Ci, xi, xii for protocols).

A 12.7 kb *Pst*I pea cpDNA fragment, along with several other fragments, did hybridize faintly to the probe. A *Pst*I 5.7kb fragment hybridized strongly to the probe (100 fold difference in intensity). Although I did obtain faint hybridization between the Douglas-fir hotspot-containing probe and the pea 12.7 kb fragment containing the origin of replication, several other pea fragments also hybridized faintly to the probe. The strongly hybridizing 5.7 kb pea cpDNA fragment lies 1.1 kb from the 12.7 kb pea fragment. Hybridization results did not clearly identify homology between the Douglas-fir hotspot and the pea cpDNA ORI, although faint hybridization did occur.

(2) Goal: if the Douglas-fir cpDNA hotspot region was functioning as an origin of replication, the cpDNA fragment should be able to replicate in an *in vitro* system.

*Xba*2.7 kb (the hotspot containing plasmid) and *Xba*2.5 kb (a cloned control fragment) were cesium chloride purified to obtain supercoiled plasmid samples. Plasmids were sent to Dr. Brent Nielsen (Auburn University, Auburn, Alabama), where they were placed in an angiosperm *in vitro* replication system. Dr. Nielson observed that no replication was initiated from the Douglas-fir cpDNA fragments. Either: (a) the *Xba*2.7 kb hotspot containing fragment is not an origin of replication, (2) the *Xba*2.7 kb fragment contained only part of the origin of replication and therefore was not capable of *in vitro* replication, or (3) an angiosperm *in vitro* replication system is not sufficient for conifer replication.

The experimental evidence is inconclusive to prove that the Douglas-fir hotspot region is a chloroplast origin of replication.

vi. Douglas-fir cpDNA Fingerprinting

The hotspot region in *Pseudotsuga* cpDNA varies both inter- and intra-specifically. The goal of this work was to determine if this region could be used as a DNA fingerprinting marker for *Pseudotsuga* and ultimately Douglas-fir. The approach used was to partially digest PCR-amplified DNA containing the hotspot from *Pseudotsuga japonica* and *P. menziesii* (PCR fragment size 1.9 kb and 2.7 kb, respectively),with a restriction enzyme that cuts within the variable region, separate amplified DNA on an agarose gel, blot the gel, and probe with the hexamer labeled 5'-PCR primer. Upon film exposure, a ladder of bands would be visuallized, each sample resulting in variable ladders.

P. japonica and *P. menziesii* were chosen because of the large difference in the size of their hotspot regions. PCR-amplified DNAs (using hotspot primers #1 and #4) were separated on a 0.8% agarose gel (1X TAE) and bands were exised from the gel under long wave UV light. DNA was eluted from the gel slices with the Amicon microelectroeluter (see Appendix Cii for detail). Five hundred nanograms of DNA from each sample was digested with 0.5U of *Hinc*II. Seventy ng of DNA was removed from the reaction at 1, 2, 4, 8, 16, 30, and 60 minute intervals and 0.5M EDTA added to stop the reaction. DNA samples were loaded on a 1.5% agarose gel (0.8X TAE) and run for 20 hours. The gel was blotted to Zetabind membrane (see Appendix Cxi) and hybridized with 100 ng ^{32}P-hexamer labeled primer #1 (see Appendix Cxii). *P. menziesii* (lanes 1-7) shows a 5 band ladder with the 'best' digestion at 4 to 6 minutes (lanes 3 and 4) (Figure 16). Partial digestion of the *P. japonica* hotspot yields a two banded ladder (lanes 9-15), digestion times of 16 to 30 minutes (lanes 13 and 14) resulting in the most complete ladders. Longer digestion times may have been required for *P. japonica* because a greater amount of DNA was loaded per lane than for *P. menziesii*. The number of observed bands corresponds to expected bands based on known sequence data of these regions (see Chapter 1).

Although partial digestion worked well on this PCR-amplified DNA, the resolving capacity of the 1.5% agarose gel was not sufficient to completely separate the digests. It is most likely a polyacrylamide-agarose mix would be needed to resolve fragments that range in size from 1.3 kb to 0.7 kb, at 20-50 bp increments.

This approach to fingerprinting individuals of *Pseudotsuga* appears to be unnecessary because length variability can be monitored by the insertion/deletions. A fingerprint can, therefore, be simply determined by the total length of the region. If it is just total length that varies between individuals, it is unlikely that there would be enough variability to serve as a fingerprint. A useful follow-up study would be to survey this area among individuals of several populations to determine at what level this variability continues to occur and its possible feasibility as, if not a fingerprint marker, then a population or regional marker.

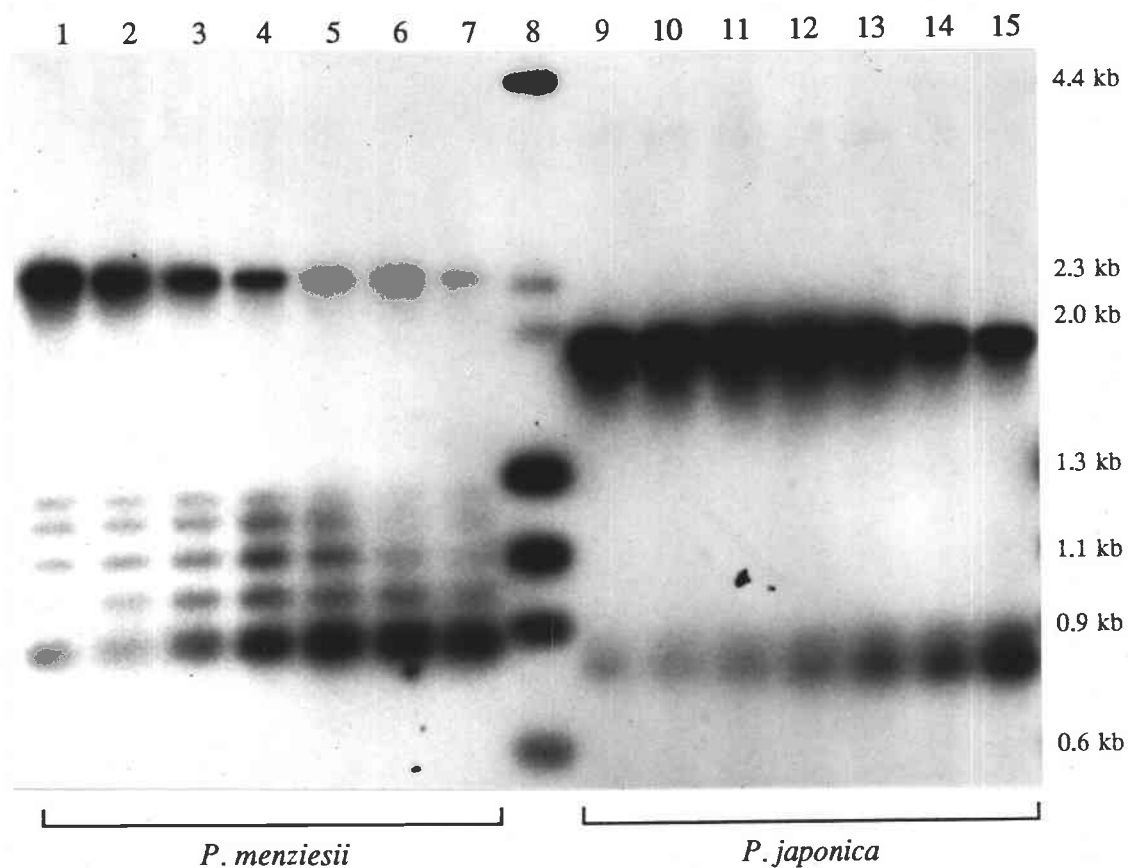


Figure 16. Partial digests of *Pseudotsuga menziesii* and *P. japonica* hotspot-containing PCR amplified cpDNA. Lanes 1-7 contain DNA from *P. menziesii* and lanes 9-15 from *P. japonica*. Lane 8 is a λ-*Hind*III/ø-*Hae*III DNA size marker. Partial digests were terminated at 1 min. (lanes 1,9), 2 min. (lanes 2,10), 4 min. (lanes 3,11), 8 min. (lanes 4, 12), 16 min. (lanes 5, 13), 30 min. (lanes 6, 14), and 60 min. (lanes 7, 15).

APPENDIX B. SEQUENCE DATA

i. DNA Sequence of PMCO

PMCO is the cpDNA hotspot-containing clone from Pseudotsuga menziesii, Corvallis OR. Genes are boxed and direction of transcription indicated with small arrows. PCR primer hybridization sites are underlined. A 68 bp trnY duplication is overlined with solid arrows (positions 626-693 and 1625-1691). Possible transposon-like footprint sequences are boxed (direct repeat: 2028-2033; inverted repeat: 2034-2039 and 2654-2659). Other repeats, discussed in Chapter 1, are underlined with half arrows (1983-2019). The hotspot is located from position 626 to 1624.

```
              10        20        30        40        50
               |         |         |         |         |
  1 TCTAGAAATCCACTGGCTATCGATCATGAAAAAGAAGTATGAAAATGAAA
         primer 1
 51 TAAATATTCATGGAGAAGAAGAGAATTTGATTCTTCGGAGAGGGATGAAT

101 ATTTATTGCAGATTCACTATGATGATTAGATTTTATCCCCGAAAGAAGGG

151 TCTTTTTTTCAAACTGAATTATCGATCTAGTTAATGTATCTAATGGATAG

201 ATATACTAAATATCTATATGAATAGAGATACTAAATATCTAGTATCTTTA

251 TTCAACCCTATTTTCTTTTTCACTCTTCTACGGGGATTCAGAGCTGAATG

301 GATTAACTTATTGGATCGGGACTGACGGGGCTCGAACCCGCAACTTCCGT
                          ←trnD
351 CTTGACAGGGCGGTACTCTAACCAATTGAACTACAATCCCAATACAGTAC

401 AGTTCACTTACTATTGAATCATATTTATTCTATGGTAGGTGCTAGATAGA

451 TCGTATAGATTACGTGAGCGCTAAGTCGATTAAATATCTTATCCTTCTCT

501 TGGATCAAAGTATCAATTCATATGGAATTGGGTACATATCTATATGATAT

551 GAATATATCATAGATATCGGAGTTCAATAACCAATTATCTTTTCATCCAT

601 GATTGGCATGAATATAACCATACCGATAGATTTATATTGATTATATTGGT

651 TGGGTCCAGCTGGATTTGAACCAGCGTAGGCATATTGCCAACGCTGGTTT
                     primer 2
701 ACAGTCCGTCCCCTTGCCAACGGATTTACAGTCCGTCCCCTTGCCAACGG

751 ATTTACAGTCCGTCACAACGGATTTCAAGTTCGTCCCCTTGCCAACGGAT
```

```
                10        20        30        40        50
                 |         |         |         |         |
801  TTACAGTCCGTCCCCTTGCCAACGGATTTACAGTCCGTCACAACGGATTT
851  CAAGTTCGTCCCCTTGCCAACGGATTTCAAGTCCGTCACAACGGATTTAC
901  AGTCCGTGTCAACGGATTTCAAGTCCGTTCCACGGATTTCAAGTTCGTCC
951  CCTTGCCAACGGATTTCAAGTCCGTCCTTGTGTCAACGGATTTACAGTCC
1001 GTCCCCGTGTCAACGGATTTCAAGTCCGTCCCCTTGCCAACGGATTTACG
1051 GATTTACAGTCCGTCCCCGTGTCAACGGATTTCAAGTCCGTCCCACGGAT
1101 TTCAAGTCCTTGCCAACGGATTTCAAGTCCGTCCCCTTGCCAACGGATTT
1151 CAAGTCCGTCACAACGGATTTACAGTCCGTTCCCGTGTCAACGGATTTAC
1201 AGTCCGTCCCACGGATTTCAAGTCCTTGCCAACGGATTTCAAGTCCGTCA
1251 CAACGGATTTACAGTCCGTTCCCGTGTCAACGGATTTACAGTCCGTCCCA
1301 CGGATTGACAGTCCTTGCCAACGGATTTCAAGTCCGTCCCCTTGCCAACG
1351 GATTTCAAGTCCGTCACAACGGATTTACAGTCCGTTCCCGTGTCAACGGA
1401 TTTACAGTCCGTCCCACGGATTGACAGTCCTTGCCAACGGATTTCAAGTC
1451 CGTCCCCTTGCCAACGGATTTCAAGTCCGTCACAACGGATTTACAGTCCG
1501 TTCCCGTGTCAACGGATTTACAGTCCGTCCCACGCACGGATTGACAGTCC
1551 TTGCCAACGGATTTACAGTGCCAACGGATTTCAAGTCCGTCCCCGTGTCA
1601 ACGGATTTACAGTCCGTCCCCCAATAGATTTATATTTATTATATTGGTT
1651 GGGTCGAGCTGGATTTGAACCAGCGTAGGCATATTACCAACGGATTTACA
                    ←trnY
1701 GTCCGTCCCCATTAACCACTCGGGCATCGACCCAGGAACCAGAAAGTAAT
1751 TGAAAGTATTTAGGTTAAGATACCAAACGAATGGATATCCTATTCCATGG
1801 TACCCCTAGGGGAAGTCGAATCCCCGTTGCCTCCTTGAAAGAGAGATGTC
              ←trnE                      primer 3
1851 CTGGTCCACTAGACGATAGGGGCTACCAATCTTCATTATATTCCAGTTCC
1901 CGGGAAGTTATCATAGGGGCTACCAATTTTCATTATATTCAAGTTCCCTG
1951 GAAGTTGTCAATAGTATGGCCAGAATTATTCAGAATCTTTCTTTTTATCG
2001 TATCGTAATCTTTCTTTTTATTGGTTTCCTTTTGGAAAACAAAGGATAA
```

```
                 10        20        30        40        50
                  |         |         |         |         |
2051 ATTATCTCCTTCTTTCCAATTTCTTTTCACACGCACGTGATCTGGAGAAA
2101 TAATTTCGTGATTTGTATGAATCATACTATTGCTTGGTATTCAAGTATCC
2151 ATATATGATACAAAGATTGATGATCTATTCTGTTGTACTTATAATCAGGA
2201 TCCTGGAGATTACGTAATGCTTACTCTTAAGCTGTTCGTTTACGCAGTAG
                              psbI →
2251 TGATATTTTTCATTTCTCTTTTTATCTTTGGATTTCTATCGAACGATCCA
2301 GGACGTAATCCCGGACGTAAAGAATAGTGAAAAAATAGGTTAATTAGTCT
2351 TTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAATAGT
2401 GACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTAC
                              trnS →
2451 TACGGATTAGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGAT
2501 GGAAGAGTTCATGTGTAACAAAATGAATGGTGGAGTGAAGGTGTATACCA
2551 TAGCATGTATGGGGATTGTATCGACAATGTAATGAATAGGTCAATTATTT
2601 AGAGAAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTTCAAAATAAT
          ←
2651 TCTTTTTCCTGAACTAGAAAGCCTAGAATTATCATAACTTCTTTTTCAAT
                    primer 4
2701 GAAGATGAAAAAAAAAAGAAAAAAATGAATCTAGA
```

ii. DNA Sequence of PMCB

PMCB is the cpDNA hotspot-containing clone from Pseudotsuga menziesii, Coos Bay OR. Genes are boxed and direction of transcription indicated with small arrows. PCR primer hybridization sites are underlined. A 68 bp trnY duplication is overlined with solid arrows (positions 626-693 and 1652-1719). Possible transposon-like footprint sequences are boxed (direct repeat: 2055-2060; inverted repeat: 2061-2066 and 2681-2686). Other repeats, discussed in Chapter 1, are underlined with half arrows (2011-2046). The hotspot is located from position 626 to 1651.

```
          10        20        30        40        50
           |         |         |         |         |
  1 TCTAGAAATCCACTGGCTATCGATCATGAAAAAGAAGTATGAAAATGAAA
        primer 1
 51 TAAATATTCATGGAGAAGAAGAGAATTTGATTCTTCGGAGAGGGATGAAT

101 ATTTATTGCAGATTCACTATGATGATTAGATTTTATCCCCGAAAGAAGGG

151 TCTTTTTTTCAAACTGAATTATCGATCTAGTTAATGTATCTAATGGATAG

201 ATATACTAAATATCTATATGAATAGAGATACTAAATATCTAGTATCTTTA

251 TTCAACCCTATTTTCTTTTTCACTCTTCTACGGGGATTCAGAGCTGAATG

301 GATTAACTTATTGGATCGGGACTGACGGGGCTCGAACCCGCAACTTCCGT
                                  <-trnD
351 CTTGACAGGGCGGTACTCTAACCAATTGAACTACAATCCCAATACAGTAC

401 AGTTCACTTACTATTGAATCATATTTATTCTATGGTAGGTGCTAGATAGA

451 TCGTATAGATTACGTGAGCGCTAAGTCGATTAAATATCTTATCCTTCTCT

501 TGGATCAAAGTATCAATTCATATGGAATTGGGTACATATCTATATGATAT

551 GAATATATCATAGATATCGGAGTTCAATAACCAATTATCTTTTCATCCAT

601 GATTGGCATGAATATAACCATACCGATAGATTTATATTGATTATATTGGT

651 TGGGTCCAGCTGGATTTGAACCAGCGTAGGCATATTGCCAACGCTGGTTT
                     primer 2
701 ACAGTCCGTCCCCTTGCCAACGGATTTACAGTCCGTCCCCTTGCCAACGG

751 ATTTACAGTCCGTCCCCTTGCCAACGGATTTACAGTCCGTCACAACGGAT

801 TTCAAGTTCGTCCCCTTGCCAACGGATTTACAGTCCGTCCCCTTGCCAAC

851 GGATTTACAGTCCGTCACAACGGATTTCAAGTTCGTCCCCTTGCCAACGG
```

```
              10        20        30        40        50
               |         |         |         |         |
901  ATTTCAAGTCCGTCACAACGGATTTACAGTCCGTGTCAACGGATTTCAAG
951  TCCGTTCCACGGATTTCAAGTTCGTCCCCTTGCCAACGGATTTCAAGTCC
1001 GTCCTTGTGTCAACGGATTTACAGTCCGTCCCCGTGTCAACGGATTTCAA
1051 GTCCGTCCCCTTGCCAACGGATTTACGGATTTACAGTCCGTCCCCGTGTC
1101 AACGGATTTCAAGTCCGTCCCACGGATTTCAAGTCCTTGCCAACGGATTT
1151 CAAGTCCGTCCCCTTGCCAACGGATTTCAAGTCCGTCACAACGGATTTAC
1201 AGTCCGTTCCCGTGTCAACGGATTTACAGTCCGTCCCACGGATTTCAAGT
1251 CCTTGCCAACGGATTTCAAGTCCGTCACAACGGATTTACAGTCCGTTCCC
1301 GTGTCAACGGATTTACAGTCCGTCCCACGGATTGACAGTCCTTGCCAACG
1351 GATTTCAAGTCCGTCCCCTTGCCAACGGATTTCAAGTCCGTCACAACGGA
1401 TTTACAGTCCGTTCCCGTGTCAACGGATTTACAGTCCGTCCCACGGATTG
1451 ACAGTCCTTGCCAACGGATTTCAAGTCCGTCCCCTTGCCAACGGATTTCA
1501 AGTCCGTCACAACGGATTTACAGTCCGTTCCCGTGTCAACGGATTTACAG
1551 TCCGTCCCACGCACGGATTGACAGTCCTTGCCAACGGATTTACAGTGCCA
1601 ACGGATTTCAAGTCCGTCCCCGTGTCAACGGATTTACAGTCCGTCCCCCC
1651 AATAGATTTATATTTATTATATTGGTTGGGTCGAGCTGGATTTGAACCAG
                                               ← trnY
1701 CGTAGGCATATTACCAACGGATTTACAGTCCGTCCCATTAACCACTCGG
1751 GCATCGACCCAGGAACCAGAAAGTAATTGAAAGTATTTAGGTTAAGATAC
1801 CAAACGAATGGATATCCTATTCCATGGTACCCCTAGGGGAAGTCGAATCC
                                               ← trnE
1851 CCGTTGCCTCCTTGAAAGAGAGATGTCCTGGTCCACTAGACGATAGGGGC
           primer 3
1901 TACCAATCTTCATTATATTCCAGTTCCCGGGAAGTTATCATAGGGGCTAC
1951 CAATTTTCATTATATTCAAGTTCCCTGGAAGTTGTCAATAGTATGGCCAG
2001 AATTATTCAGAATCTTTCTTTTTATCGTATCGTAATCTTTCTTTTTATTG
2051 GTTTCCTTTTGGAAAACAAAGGATAAATTATCTCCTTCTTTCCAATTTC
2101 TTTTCACACGCACGTGATCTGGAGAAATAATTTCGTGATTTGTATGAATC
```

```
           10        20        30        40        50
           |         |         |         |         |
2151 ATACTATTGCTTGGTATTCAAGTATCCATATATGATACAAAGATTGATGA
2201 TCTATTCTGTTGTACTTATAATCAGGATCCTGGAGATTACGTAATGCTTA
2251 CTCTTAAGCTGTTCGTTTACGCAGTAGTGATATTTTTCATTTCTCTTTTT
                         psbI →
2301 ATCTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGA
2351 ATAGTGAAAAAATAGGTTAATTAGTCTTTTACGTTCCGTAGAAAGATTCG
2401 GAGTTATTCGTTTTCAGGATCAATAGTGACCGAACGGAGAGAGAGGGATT
2451 CGAACCCTCGGTACGGATAATCCGTACTACGGATTAGCAATCCGCCGCTT
                ← trnS
2501 TGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAACAAAA
2551 TGAATGGTGGAGTGAAGGTGTATACCATAGCATGTATGGGATTGTATCG
2601 ACAATGTAATGAATAGGTCAATTATTTAGAGAAAAATCAATCTGGCGAAT
                                         ←
2651 CGTATTGTTCATTCCGTTCAAAATAATTCTTTTTCCTGAACTAGAAAGCC
                                               primer 4
2701 TAGAATTATCATAACTTCTTTTTCAATGAAGATGAAAAAAAAAAGAAAAA
2751 AATGAATCTAGA
```

iii. DNA Sequence of PJ

PJ is the consensus cpDNA hotspot-containing PCR clone from Pseudotsuga japonica. Genes are boxed and direction of transcription indicated with small arrows. PCR primer hybridization sites are underlined. A 68 bp trnY duplication is overlined with solid arrows (positions 606-673 and 840-908). Possible transposon-like footprint sequences are boxed (direct repeat: 1239-1244; inverted repeat: 1245-1250 and 1863-1868). Other repeats, discussed in Chapter 1, are underlined with half arrows (1200-1230). The hotspot is located from position 606 to 839.

```
             10        20        30        40        50
              |         |         |         |         |
  1 ATGAAAAAGAAGTATGAAAATGAAATAAATATTCATGGAGAAGAAGAGAA

 51 TTTGATTCTTCGGAGAGGGATGAATATTTATTGCAGATTCACTATGATGA

101 TTAGATTTTATCCCCGAAAGAAGGGTCTTTTTTTCAAACTGAATTATCGA

151 TCTAGTTAATGTATCTAATGGATAGATATACTAAATATCTATATGGATAG

201 AGATACTAAATATCTAGTATCTTTATTCAACCCTATTTTCTTTTTCACTC

251 TTCTACGGGGATTCAGAGCTGAATGGATTAACTTATTGGATCGGGACTGA

301 CGGGGCTCGAACCCGCAACTTCCGTCTTGACAGGGCGGTACTCTAACCAA
      <-trnD
351 TTGAACTACAATCCCAATACAGTACAGTTCACTTACTATTGAATCATATT

401 TATTCTATGGTAGGTGCTAGATAGATCGTATAGATTACGTGAGCGCTAAG

451 TCGATTAAATATCTTATCCTTCTCTTGGATCAAAGTATCAATTCATATGG

501 AATTGGGTACATATCTATATGATATGAATATATATAGATATCGGAGTTCA

551 ATAACCAAAACCAATTATCTTTTCATCCATAATTGGCATGAATATAACCA

601 TACCGATAGATTTATATTGATTATATTGGTTGGGTCCAGCTGGATTTGAA

651 CCAGCGTAGGCATATTGCCAACGGATTTCAAGTCCGTTCCATTGCCAACG
      primer 2
701 GATTTACAGTCCGTCACAACGGATTTACAGTCTGTCCCCGTGTCAACGGA

751 TTTCAAGTCCGTCCCCGTGTCAACGGATTTACAGTGCCAACGGATTTCAA

801 GTCCGTCCCCGTGTCAACGGATTTACAGTCCGTCCCCCCAATAGATTTAT

851 ATTGATTATATTGGTTGGGTCGAGCTGGATTTGAACCAGCGTAGGCATAT
                                        <-trnY
```

```
              10        20        30        40        50
901  TACCAACGGATTTACAGTCCGTCCCCATTAACCACTCGGGCATCGACCCA
951  GGAACCAGAAAGTAATTGAAAGTATTTAGGTTAAGATACCAAACGAATGG
1001 ATATCCTATTTCATGGTACCCCTAGGGGAAGTCGAATCCCCGTTGCCTCC
                                  ← trnE
1051 TTGAAAGAGAGATGTCCTGGTCCACTAGACGATAGGGGCTACCAATCTTC
        primer 3
1101 ATTATATTCCAGTTCCCGGGAAGTTATCATAGGGGCTACCAATTTTCATT
1151 ATATTCAAGTTCCCTGGAAGTTGTCAATAGTATGGTCAGAATTATTCAGA
1201 ATCTTTCTTTTTATCGTAATCTTTCTTTTTCTTGGTTTCCTTTTGGAAAA
1251 CAAAGGGAGAAATTATCTCCTTCTTTCAAATTTCTTTTCACACGCACGTG
1301 ATCTGGAGAAATAATTTCGTGATTTGTATGAATCATACTATTGCTTGGTA
1351 TTCAAGTATCCATATATGATACAAAGATTGATGATCTATTCTGTTGTACT
1401 TATAATCAGGATCCTGGAGATTACGTAATGCTTACTCTTAAGCTGTTCGT
1451 TTACGCAGTAGTGATATTTTTCATTTCTCTTTTTATCTTTGGATTTCTAT
                              psbI →
1501 CGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGTGAAAAAATAGG
1551 TTAATTAGTCTTTTACGTTTCGTAGAAAGATTCGGAGTTATTCGTTTTCA
1601 GGATCAATAGTGACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGG
                                       ← trnS
1651 ATAATCCGTACTACGGATTAGCAATCCGCCGCTTTGGTCCGCTCAGCCAT
1701 CTCTCCAAGATGGAAGAGTTCATGTGTAACAAAATGAATGATGGAATGAA
1751 GGTGTATACCATAGCATGTATGGATTGTATCGACAATGTAATGAATAGGT
1801 CAATTATTTAGAGAAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTT
1851 CAAAATAATTCTTTTTCC
```

iv. DNA Sequence of PR

PR is the region in Monterey pine cpDNA that is homologous to the Pseudotsuga cpDNA hotspot. Sequence data originates from two cloned Monterey pine cpDNA fragments. The first 788 bp are from the 5.8 kb KpnI clone; bases 789-1041 are from the 10.4 kb KpnI clone. Genes are boxed and direction of transcription indicated with small arrows. KpnI restriction site is underlined. Note absence of the trnY duplication found in Pseudotsuga.

```
              10        20        30        40        50
               |         |         |         |         |
  1 AGCTTGCATGCTATCGATCATGAAAAGAAGTATTAGCATGAAATAGTAA

 51 CAATATTCCTGGGGAAGAACAGAAGGAAGATCTCCTAAAAATCATTGGGA

101 ATTATCTATAGGGATCTCTGTGGGATTCTGACTTAGAAGGACTACCTCTG

151 TGTCACCCTAAAATCTATTGATCTTCCTATGTTTTTGATAGAGAAATTTT

201 ATTCTTCGGTGAGGGAAGAATATTTCTTGTAGATAAAATATGATTATTAT

251 ATTTTATCCTCGAAAGAAGGGTCTTTTTCCAAACTGAATTATCGATCTGT

301 GAATGTATCTAATGGATCGGGACTGACGGGACTCGAACCCGCAACTTCCG
                          <--trnD
351 TCTTGACAGGGCGGTACTCTAACCAATTGAGCTACAATCCCAATACAGTA

401 CAGTTCACCTACTATTGGATAATATTTATTCCATGATAGGTGCTAGATAG

451 GTCATATAGATTATGCGAGTGGCTAGGTCGATTAAATATCTTAATCTTCT

501 CTTTCATTTTTGAAATGTATCGATTCATACGGAATCGGGCATCTACGATA

551 TGAATAGATATCGATGCCGGGGTTCAATAACCAATTATCTTTTCATTCAT

601 GATTAGCATGAATATAACCATACCGATAGATTGGTATTGATGATATTGGT

651 TGGGTCGAGCTGGATTTGAACCAGCGTAGGCATATTGCCAACGGATTTAC
                 <--trnY
701 AGTCCGTCCTCATTAACCACTCGGGCATCGACCCAGGAACAAGACAGTAA

751 TTGAAAATTATTTAGGATACCTAACGAATGGATCATGGTACCCCCAGGGG
                                         KpnI
801 AAGTTGAATCCCCGTTGCCTCCTTGAAAGAGAGATGTCCTGATCCACTAG
    <--trnE
851 ACGATAGGGGCCACCAATCTATTCTTTATAATATGAAAGTTATCGGGAAG

901 TTGTCAATAGTATGACCAGAATTCTTGGTTTCCTCAAGTTTTTTTTTCAA
```

```
                10        20        30        40        50
                 |         |         |         |         |
 951 TAAACTTGCCTATCGATAAAATGGAGTCCCTTCAGAAATTACTTATTGCA

1001 ACTAAAACAACTCTAAAGTAATTTCCGGAATCTCTATTTGT
```

v. DNA Sequence of DFleft

1,347 bp of Douglas-fir cpDNA sequence data surrounding the left border of the 40-50 kb Douglas-fir/Monterey pine inversion. Sequence is a portion of the PM6.7 clone. Genes are boxed and direction of transcription indicated with small arrows. Inversion junction (the approximate location where continuous homology between Douglas-fir and Monterey pine is broken) is at position 1140.

```
             10        20        30        40        50
              |         |         |         |         |
  1 TCTAGATGATGCCTTTATTCATTCAAATAATCCCTTTTTTGGAAAATTAC
                               psbK -->
 51 CTGAGGCTTATGCGATTTCCGATCCAATTGTCGATGTAATGCCAATTATT

101 CCCGTTCTCTCTTTTCTTTTAGCCTTTGTTTGGCAAGCTGCTGTAAGTTT

151 TCGATAAAAAGTATCCCCTTTTTTCCTTTTTCAAGTTTTTGTGTCGCTGT

201 CATTTATCTAATTTTTGTATCACTCTTTCCATTTTTTGTCGCAGAAGTTT

251 TATCCTTGCTCTACCCGACAATACCAGATCGAGATACCTCATCTGCTCTC

301 GACTAAAAAGCTTTTTAACTCACCTTCGTCAATTCCTTCCGATCTCATCG

351 CTCACTTTGGATCGGGCTATTTGGTCACGTATTTATACGAATGACATATT

401 TTCATAAATATTTGATAAATATCTGGTTGATCCAAAAAAGAAGAAGGAA

451 GAAAGACCATTTTGAAAACAAAGGGATAAGTTATCTCCTTCTTTCCAATT

501 TCTTTTCACACGCACGTGATCTGAGAAATAATTTCGTGATTTGTATGAAT

551 CATACTATTGCTTGGTATTCAAGTATCCATATACGGTACAAAGATTGATG

601 ATCTATTCTGTTGTACTTATAATCAGGATCCTGGAGATTACGTAATGCTT

651 ACGCTTAAGCTGTTCGTTTACGCAGTAGTGATATTTTTCATTTCTCTTTT
                               psbI -->
701 TATCTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAG

751 AATAGCGAAAAAATAGGTTAAGTAGTCTTTTACGTTCCGTAGAAAGATTC

801 GGAGTTATTCGTTTTCAGGATCAATAGTGACCGAACGGAGAGAGAGGGAT

851 TCGAACCCTCGGTACGGATAATCCGTACTACGGATTAGCAATCCGCCGCT
             <-- trnS
901 TTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAACAAA

951 ATGAATGGTGGAGTGAAGGTGTATACCATAGCATGTATGGATTGTATCGA
```

```
          10        20        30        40        50
           |         |         |         |         |
1001 CAATGTAATGAATAGGTCAATTATTTAGAGAAAAATCAATCTGGCGAATC
1051 GTATTGTTCATTCCGTTCAAAATAATTCTTTTTCCCCTATTTCTTCTGAC
1101 CTCTGCCGGTGGCCAGGCCAGGCCAAGAAAAACAAAAAAGAATTCATGCA
1151 TCAGACAATGCGTTAGCTAATCGGTAAGCGAAAAAAGTGGTTGTAACGGT
1201 AAGAAAAAACAGACCGAAAAAAAAATAGAACAGATTGAACATCTAGTGTC
1251 ATCTTTTTATTCTCTCCCTAATAATTTTCAATAAGTTAGTTACATGGAAT
1301 GGATTAGTCCATTTATTTCTCTCCAGTATAAAATTTCAATATCTAGA
```

vi. DNA sequence of DFright

3,184 bp of Douglas-fir cpDNA sequence data surrounding the right border of the 40-50 kb Douglas-fir/Monterey pine inversion. Sequence is a portion of the PM3.8 clone. Genes are boxed and direction of transcription indicated with small arrows. Inversion junction (the approximate location where continuous homology between Douglas-fir and Monterey pine is broken) is at position 2204.

```
                10        20        30        40        50
                 |         |         |         |         |
  1 ATGATATTGGATCACCTAATAATGAAACTGCAACCTTAGTCGCCATCTCC
                                      psbN -->
 51 ATATCTTGTTTACTTGTGAGCTTTACTGGTTATGCCCTATACACCGCCTT
101 TGGGCAACCCTCTGAACAACTTAGGGATCCTTTTGAGGATCACGAGGACT
151 AATAGAGAGAATGACCTTCCCCTATAGGGAAGGTCATTCTCTCTTTGATG
201 GGAGAGAAAGAATGAGGATTTGGAGAAGCAAAATTATTTTCCCCCTTTAC
251 CTGGAATTTTGGGTGGTTCTCGGAAGAAAATAGCGAAAAAGATTATCCCT
301 AGGGTCGATACCAACAGGAATGTATAAACCAATGCTTCCATAGATTCGAT
351 CGTAATTTACAATTATGGAGATAGCTTTCGTACTAATATTGATTAGAAAA
401 GAAATAAGTTGATTAGAGAAGAAATAAGAGCAATATCATACCACTTGTCT
451 CTTAGTAGTTGGATCTCCCAGTTTTTGGAATGCTCCAAATTCTACTCGAG
501 CATCCAAATCCGGGTCAATACCAGCAAAAACATCTCTGAATAGGGTTCTA
551 GCGCCATGCCAAATGTGTCCAGGAAAGGAAAGGAGGGCAAATGTAGCATG
601 TCCGAAAGTAAACCAACCCCTTGGACTACTACGAAAAACACCATCGGATT
651 TTAGAGTAGCACGAGTCTAATTCAAAAATTTCACCTAATTGAGCACGTCT
                    <-- psbB
701 AGCATATTTTTTGACTATAGCAGGATCACCAAAACTAACCCTGTCAAGTC
751 CACCACCATAGAACTCAACAGTTACACCTACTTGTTCAACACTATACTTT
801 GATTCTGCCCTCCTAAAAGGAACATCGGCTTTCACAATTCCTTCTTTGTC
851 TACCAGAACTACTGGAAATGTTTCGAAAAAGGTAGGCATACGACGTACAA
901 AAAGCTCATTTCCATCCTTATCTTTGAAGATAGGGTGTCCTAACCAACCA
```

```
           10        20        30        40        50
            |         |         |         |         |
951  ACAGCTATTCCATCTCCATTGTCCATCGCACCTGCTCTGAATAACCCCCC
1001 TTTAGCTGGATTATTACCAATGTAATCATAAAAAGCGAGTTTCTCGGGAA
1051 TTTTTGACCAGGCTTCCGATAGGCTCAAATTTTCGGCCAGACCGGCACGA
1101 ACCCGTCGATCTATTTCTTGCTGAAAGTATCCCTGATCCCACTGGTAACG
1151 AGTGGGACCAAATAATTCAACCGGAGTAGTTGCAGAGCCATACCACATGG
1201 TTCCGGCAACAACGAAAGCTGCAAAAAACACAGCAGCAATACTGCTGGAT
1251 AGGACCGTTTCAATATTCCCCATGCGTAATCCTATGTATAAACGTTGGGG
1301 AGGACGAACACTGAGATGAAATAGACCTGCTAATATACCCAATATACCCG
1351 CTGCAATATGATGAGAAGCTATTCCTCCCGGAACAAAAGGATCAAAACCT
1401 TCAGCTCCCCATGCTGGATCCACTGGTTGTATTTTTCCAGTTAGTCCATA
1451 AGGATCAGACACCCATATCCCAGGACCATCAAACCTGTTACATGAAATGC
1501 TCCAAATCCGAAACAAGCTACTCCTGAGAGGAATAAATGAATTCCAAATA
1551 CTTTTGGCAAATCCAAACAAAGTTTTCCCGTACGGTCATCACAGAATAGG
1601 TCCAGATCCCAATATACCCAATGCCAGATAGCTGCCAAAAAGCACAGGCC
1651 AGAAAAGACGATATGTGCCCCGGCCACACCTTCATAACTCCAAATACCAG
1701 GATTAATTACAGTTTCTCCAGTGATGCTCCATCCACTCCATGAATCCTTT
1751 ATTCCCAAACGAGTCATAAAAGGTATAACGAACATACCTTGTCTCCACAT
1801 TGGATCAAGAACAGGATCGGATGGATCAAAAACTGCTAATTCGTACAGAG
1851 CCATTGAACCGGCCCAACCAGAAACTAGAGCTGTATGCATTATATGTACA
1901 GAAATTAACCGGCCAGGATCATTCAATACGACGGTATGAACGCGATACCA
1951 AGGCAAACCCATGCAAACACCCCTTTATCGGAAAAAGTAGACACTATGTA
2001 ACTTTCTCACATTGGATTGGAAGAGATCCTGCTATCGAGTTAAACCCGAA
2051 TCATCTCGAAGGTGAACATCTATTCACTTGGACATTGCTATAGAACAGGT
2101 TCGAAACAATTCTGTTCATACATAATAGCAGGGAGAAGCAAAGATATTTT
2151 ATCGTTTGTATGTACCAATACACAATGTGGGATTGGTCCCATTTATACT
```

```
                10        20        30        40        50
                 |         |         |         |         |
2201 GATCAAACGCATAAAGACTTGTTCATTATTTGAGGAACCTGCAAAAAAAG
2251 AGGAAACTAGATCTTCCTATTAATTCTTCCGTTCGTCCATGAACGAGAAC
2301 GATATCTCCTTTCCTTCCAAGTTATGGACCAGAATATACTTTCCGATAAA
2351 CACGTACCAGAAGTCCATTTCTTCATATTTCTGTTGTACTTAGAATCAGG
2401 ATCCTAGAGATTACGTAATGCTTACTCTTAAGTTGTTCGTTTACGCAGTA
                                  psbI →
2451 GTGATATTTTTCATTTCTCTTTTTATCTTTGGATTTCTATCGAACGATCC
2501 AGGACGTAATCCCGGACGTAAAGAATAGTGAAAAAATAGGTTAATTAGTC
2551 TTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAATAG
2601 TGACCGAACGGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTA
                       ← trnS
2651 CTACGGATTAGCAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGA
2701 TGGAAGAGTTCATGTGTAACAAAATGAATGGTGGAGTGAAGGTGTATACC
2751 ATAGCATGTATGGATTGTATCGACAATGTAATGAATAGGTCAATTATTTA
2801 GAGGAAAATCAATCTGGCGAATCGTATTGTTCATTCCGTTCAAAATAATT
2851 CTTTTTCCTGAACTAGAAAGCCTAGAATTATCATAACTTCTTTTTCAATG
2901 AAAAAGAAATCCTTTTCATATCGATTGGCTTGTTTTCGACATAGCACCGC
2951 TGTACTTTCCATGTTTTCTATTTGCATTTGGCTTTATATCTATTTTATGT
3001 GCTATGATATTGTTTATAGTGACTGAGGGCTATAAGTTCCTGACCGATAA
3051 AATAAACAACAAAAAAAACATAGAAAATGGAAGAAAAGTGATGAATCTTG
3101 AAAACATTTTATTCATACATCCATCAAGTCGATGGGATCATAGGGGTGAA
3151 AGAAAACGAAATGGATCTAGAGTCGACCTGCAGG
```

vii. DNA Sequence of MPleft

2,801 bp of Monterey pine cpDNA sequence data surrounding the left border of the 40-50 kb Douglas-fir/Monterey pine inversion. Sequence is a portion of the PR7.6 clone. Genes are boxed and direction of transcription indicated with small arrows. Inversion junction (the approximate location where continuous homology between Douglas-fir and Monterey pine is broken) is at position 1786.

```
              10        20        30        40        50
               |         |         |         |         |
  1 AAATTGATCCGTATTCCCCATCTTTGGCACCACAGATATTATCCCCTTGA

 51 TCTCTTCCCTATTTTCACATTCTTCCCTTCTGATAAGTATGGTTAAATTA

101 TGATAAAAAGAAAAGTTCCATTGGATCTTTTACGAAAGAAAAAAGGGAAG

151 AGCGAAGGAAGGGAAAACAGGAACCAATGAAATAAGTCTGGACGAAGATT
                                                 <-- trnQ
201 CGAACCTCCGAATAACAGGATCAAAACCTGCTGCCTTACCGCTTGGCCGA

251 CGCCCCGATTCCCATCCTATTTCCCTATTCATATGCTAATGAATACTTAT

301 ATCAAATTTTTTGTCAACCCATTAGTATCCAAGATTCATTAGATCAGATC

351 TCAATTGGAAAAAAATTTTTGTGGATATTTCAACAAAATAGGTTATTGAT

401 GGAATTGACATAATAGGAGAAACAGAAAAAGGACAAGAATATCCATTCAT

451 TGATCATTTTCTATTAGATTGGGTAAAATGTTGTATGTATGAAAGAATCA

501 TCCCTTCCACGACCCCAAGTCCGGTCAAACTTGCCGAAGAGCTTCGATCG

551 ATTCGATCAATCAAAGACTTTTATGGCTTTACCCCGGCCTAATTTTTATT

601 GGAGAATAAAAATGCCAGTTATGTTCAATATTTTTCTAGATGATGCCTTT

651 ATTCATTCAAATAATCCCTTTTTTGGAAAATTACCTGAGGCTTATGCGAT
                                  psbK -->
701 TTCTGATCCAATTGTCGATGTAATGCCAATTATTCCTGTTCTCTCTTTTC

751 TCTTAGCCTTTGTTTGGCAAGCTGCTGTAAGTTTCGATAAAAAGTATCCC

801 CTTTTTTCTTTTTCAAGTTTTTGGGTCGCTGTCATTGATCCAATTTTTGT

851 ATAACTTTTTCCATTTTTTTTTGATAGAAGTTCTATCCTTGCTCCACCCG

901 ACGATACCAGATTCAGATACCTTATCTGCTCCCGGATAAAAACTTTTCCA
```

```
             10        20        30        40        50
              |         |         |         |         |
 951 CTCACCTTCATCAATTCCTTCCGATCCCACCGCTCACTTCGGATCGGGCT
1001 ATTGGGTCACGTATTGATACGATCACCAATGACTTATTTTCATAAATATT
1051 CAATAAATATCTGGTTGATCCAAAAAAATAAGAGGGAAAAAAGACCATTT
1101 TGAAAACAAAGATAAAAATTATATCTTCTTTCAAATTTATTTTCACACGT
1151 GATTCGGAGAAATAAATTCCGTGATTGTAACAATAATACTATTGTTTGGT
1201 ATTCAAGTATCCATATATGGTACAAAGATTTGATGATCTATTCTGTTGTA
1251 CTTATAATCAGGATTCCGGAGATTACGTAATGCTTACTCTTAAGCTGTTC
                                          psbI →
1301 GTTTACGCAGTAGTGGTATTTTTCATTTCTCTTTTTATCTTTGGATTTCT
1351 ATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGTGAAAAAAGT
1401 ATCTAGGTTAATAGGTCTTTTCCGTTCCGTAGAAAGATTCGGGGTTATTC
1451 GTTTTCGGGATCAATAGTGGACGAACGGAGAGAGAGGGATTCGAACCCTC
                                         ← trnS
1501 GGTACGGATGATCCGTACTACGGATTAGCAATCCGCCGCTTTGGTCCGCT
1551 CAGCCATCTCTCCAAGATGGAAGAGTTCATGTATAACAAAATGAATGGTG
1601 GAGTAAAGGTGTATACCATAGTATGTACAGATTGTATCGGCAATATAATG
1651 AATATTGCAATTATTCAGTTGGATAAAGAACAATCCAGTCAATCATATTG
1701 TTCATTTCGTTAAAAATAGTTCTGTATGACTGATTTTTTCTGCTTTTCTT
1751 GCTGGCCGATGGCCAGGCCAAGAAAAGCAGAAAAAATCAGGCATGCAGTG
1801 CTTGACCTAATTTGATACCTAGAAAAACTGCTGTAAAAGCAAGAAAAGAA
1851 GCTATCAAAAATTGGACTTCTATTGCCATATCTTCATTCCCTCCCCAATC
1901 AGTTTGATTAAATGCGTTACATGGATTAGTCCATTTATTTATCTCCAGTA
1951 TCCAATTTTATTATCTAGATATTGAAGGGTTCTCTATCTATTTAGGGTTC
2001 TCTATCTATTTTATGTATTATTGTAAATATATCAGTTGCTCAACGCCATA
2051 GGTTCCTGATCGAAACTACACCAATGGGTAGGAGTCCGAAGAAGACAAAA
2101 TAGAAGAAAAGTGATTGATCCCGACAACATTTTATTCATACATTCAGTCA
2151 ATGGAGGGTGAAAGAAAACCAAATGGATCTAGAAGTTATTGCGCAGCTCA
```

```
              10        20        30        40        50
               |         |         |         |         |
2201 CTGTTCTGACTCTGATGGTTGTATCGGGCCCTTTAGTTATTGTTTTATCA

2251 GCAATTCGCAAAGGTAATCTATAATTACAATGAGCCATCTCCGGAGATGG

2301 CTCATTGTAATGATGAAAACGAGGTAATGATTGATAGAAACTTTCAATAG

2351 AGGTTGATTGATAACTCCTCATCTTCCTATTGGTTGACAAAAGATCGATC

2401 CATATGTTACAATCAGATCGTGGCGGGTATAGTTTAGTGGTAAAAGTGTG
                             exon 1
2451 ATTCGTTACATTATCAACTCGAATAGTTGAAGGATCTTTTACATGGATCT

2501 TTGATCCATCTAAAGCTCTATTTCCAGATAGGTTCAAAACCTTCCCTATG
                        trnG →
2551 AATACCCAGGAATAGCATACCTTCTCGTAATCTGGGTCTGAAATGATGAA

2601 AGAAAGATAAACACATTTTTGGTTCCAAGATAGCGACACAGAATGTTTGA

2651 AAGGTTAGATAAATTACAGATCTATGGGNNAGATCTTTTTTCATCGTGAC
                        intron
2701 TAAACCTTCAACTTATGGATGAAAGGACCCCAGGTTGAAGCCGAATAGCT

2751 ATAGAACTATGACTTTTGTTTACTTGGGTTATCGGSATTTCGTTCGAGCT

2801 C
```

viii. DNA Sequence of MPright

2,982 bp of Monterey pine cpDNA sequence data surrounding the right border of the 40-50 kb Douglas-fir/Monterey pine inversion. Sequence is a portion of the PR8.5 clone. Genes are boxed and direction of transcription indicated with small arrows. Inversion junction (the approximate location where continuous homology between Douglas-fir and Monterey pine is broken) is at position 935.

```
              10        20        30        40        50
               |         |         |         |         |
  1 CGTAATCCTACGTATAAACGTTGGGGAGGACGAACACTGAGATGGAATAG
 51 ACCTGCTAATATACCCAAAATACCCGCTGCTATATGATGAGAAGCTATTC
101 CTCCTGGAACAAAAGGATCAAAACCTTCAGCTCCCCATGCTGGATCCACT
151 GGTTGTATTTTTCCAGTTAGTCCATAAGGATCAGACACCCATATCCCAGG
201 ACCATACAAACCCGTTACATGAAATGCTCCAAATCCGAAACAAGCTACTC
251 CTGAGAGGAATAAATGAATTCCAAATACTTTTGGCAAATCTAAACAACGT
                        <-psbB
301 TTTCCCGTACGTTCATCACAGAATATGTCCAGATCCCAATATACCCAATG
351 CCAGATAGCTGCCAAAAAGCACAGGCCAGAAAACATGATATGTGCCCCGG
401 CCACACCTTCATAACTCCAAATACCGGGATTAATTACAGTTTCTCCGGTG
451 ATGTTCCATCCACTCCATGAATCCTTTATTCCCAAACGAGTCATAAAAGG
501 TATAACGAACATACCTTGTCTCCACATTGGATCAAGAACAGGATCGGATG
551 GTTCAAAAACTGCTAATTCGTACAGAGTCATTGAACCAGCCCAACCAGCA
601 ACTAGAGCTGTATGCATTATATGTACAGAAATTAACCGTCCAGGATCATT
651 CAATACGACGGTATGAACGCGATACCAAGGCAAACCCATGCAAACACCCC
701 TTTATCGGAAAAAGTAGACACTATGTAACTTTCTCACATTGGATTGGAAG
751 AGATCATGTTATCGAGCTAGACCCGGATCATCTCGAAAGTGAACATCTAT
801 TCACTTGGACATTGCTAATGATGGAACAGGTTTGAAACCATTTTGTTCAT
851 ACATAATAGCAGGGAGAGGCAAAGATATTTTATCGTTTGTATGTACCAAT
901 ACACAATGTAGGATTGGTCCCATTTATATTGATAAAAAAAGAGGAAAGA
```

```
               10        20        30        40        50
                |         |         |         |         |
 951 AACGAGATCTTCTAATCCTTCCATTCGTTCATGAACGATACCTTTGCAAT
1001 TTATGGACTAAGTCAGATATAATTTTTGATTAAAAATAGCATTTTTCTAC
1051 TAAACAGTAATACTTTACTGCGGAAAAAATAAAATACTGAGCATAGTTC
1101 AAATGCTCAGTCAAAAATGATAACTTTATCATTTTGTGCTATGCAATGAA
1151 AAAAAAGGTCAAAATAGGAAGTATCTAGGTTAATAGGTCTTTTCCGTTCC
1201 GTAGAAAGATTCGGGGTTATTCGTTTTCAGGATCAATAGTGGACGAACGG
1251 AGAGAGAGGGATTCGAACCCTCGGTACGGATGATCCGTACTACGGATTAG
                   <- trnS
1301 CAATCCGCCGCTTTGGTCCGCTCAGCCATCTCTCCAAGATGGAAGAGTTC
1351 ATGTATAACAAAATGAATGGTGGAGTAAAGGTGTATACCATAGTATGTAC
1401 AGATTGTATCGGCAATATAATGAATATTGCAATTATTCAGTTGGATAAAG
1451 AACAATCCAGTCAATCATATTGTTCATTTCGTTAAAAATAGTTCTGTATG
1501 ACTGATTTTTTCTGCTTTTCTTGGCCTGGCCGATGGCAGCAAGAAAAGCG
1551 AAAATCGTCATACGTGCTTGACCTAATTTGATACCTAGAAAAACTGCTGT
1601 AAAAGCAAGGCAAGCTTGCTATCAAAAATTGGACTTCTATTGCCATATCT
1651 TCATTCCCTCCCCAATCAGTTTGATTAAATGCGTTACATGGATTAGTCCA
1701 TTTATTTATCTAACGTATCCAATTTTATTATCTAGATATTGAAGGGTTCT
1751 CTATCTATTTAGGGTTCTCTATCTATTTATGTATTATTGTAAATATATCA
1801 GTTGCTCAACGCCATAGGTTCCTGATCGAAACTACACCAATGGGTAGGAG
1851 TCCGAAGAAGACAAAATAGAAGAAAAGTGATTGATCCCGACAACATTTTA
1901 TTCATACATTCAGTCAATGGAGGGTGAAAGAAAACCAAATGGATCTAGAA
1951 GTTATTGCGCAGCTCACTGTTCTGACTCTGATGGTTGTATTTGGCCCTTC
2001 AGTTATAGATAGATAGAGCTTTGGATGGATCAGAGATCCATGTCAAAATA
                 partial intron                exon 1
2051 TCTTTAACTATTTAAGTTGATAATGTAACGAATAAAAACCCACTGTATCA
         <- trnG
2101 CTAAACTATACACGCCACAATCCCATTGACGGATATTCCGTCACTTCTTT
2151 CCTTCCACATTTGAATCCAAATTCCGGAATTCCTTTCTTCTCTCTCGGAA
```

```
          10        20        30        40        50
           |         |         |         |         |
2201 ATAGAAGAGAGAGAAGAAAGGATTCTATCCATTCTAGTTTTTAGAAATTC
2251 TAGGTTGTTCTTTCCCTTTATCAAGGAATTTTTTGCTCAACTCCTCTAAT
2301 CTTCTAATTGAATTCTTTCTAATTCATTTTAGAAAGATCTTCCTCTTCCA
                                ←rps4
2351 GGAAGAAAAAGAGCCATAAACTGGAATGGCTCCATATTCTTAGCTCTCAA
2401 TCTTTGTTGATCTCTTATAGTAATCACATCTTGAGGTTTGCAAGGGTAAC
2451 TTGGTATATCTACCATATGGTCATTGACCCGGATATGTCATGATTCACTA
2501 ATTGTCTAGCTCCGGGAATAGTGGGAGCCATACCCCAATCGAAAAAGAAT
2551 TATTATCCAAACACATTTCAAGTAATTGCAATTAGGACCTGACCCGTTTG
2601 AGCTTTGGCTCTTCTAGCAATACGAACATATTTCAGCAAATGATCTCAGG
2651 CATAACTTGTCTCCATTCGCTTCGTATCAGCTCAGCCCGGAGTTTCCAGT
2701 ATAGCGATCCTTACCCTACGAGCTACTATAATGTTCTCAACTTGATATCT
2751 ATAAAGATAGATGGATCATCTATATCCAATTAATTTGTTGTCCATCTACC
2801 ATAGTAAACTCACTCATTCATATATGATGAGGGGGCCCTTTTAACTCAGC
                                              trnT→
2851 GGTAGAGTAACGCCATGGTAAGGCGTAAGTCATCGGTTCAAGTCGATAAA
2901 GGGCTCATGTTTCCAGGAGAAAGAGGGTGGGGTGTCATTTATATTAAAAG
2951 ATAGAAATATTTCACGAATATGAATGAAAGAT
```

APPENDIX C. PROTOCOLS

i. CPDNA Isolation (from Strauss lab)

1. Collect branches, cut under water, and leave in dark for about 2-5 days. Prefer soft, dark green, but matured new growth (often found on partly shaded foliage of 2- to 10-year-old trees).
2. Use 50g of foliage per 400 ml extraction buffer, applied to three 37ml sucrose gradients.
3. In cold room, homogenize in steel blender with lizuid N_2 until well-powdered; add N_2 to cover needles (2-4 additions of N_2 needed).
4. Let N_2 evaporate, then add 400ml extraction buffer and mix well with spatula. Homogenize for about 1 minute in polytron, small probe, speed setting 4 to 5.
5. Pour into 2 fummels, each with 4 layers of cheesecloth above 1 layer of miracloth. Squeeze cheesecloth, rotate and 'nudge' miracloth to speed filtration. Funnels drain into 250ml GSA tubes in tray of ice.
6. Centrifuge filtrate at 1,000xg (3,250 rpm in GSA rotot) for 15 minutes at 4°C.
7. Resuspend each pellet in 9ml ice-cold wash buffer-I with a soft paint brush.
8. Load resuspended pellet onto a two-step gradient: 13ml of 52% sucrose under 5 ml of 41% sucrose under 7 ml of 30% sucrose (sucrose solutions all 50mM Tris-HCl, pH 8.0, 25 mM EDTA). Mix overlays slightly when making gradients to create diffuse interfaces to chloroplasts are not trapped.
9. Centrifuge gradients at 25,000 rpm for 30 min. at 4°C in an AH-629 swinging bucket rotor.
10. Remove chloroplast back using a wide-bore transfer pipet; avoid dense green 'cap' in top third of gradient--take more diffuse green 'band' (1-3 cm thick) toward middle of gradient. Put in 50 ml oak-ridge style centrifuge tube(s), dilute with 3-10 volumes of wash buffer-II, and centrifuge at 1,500xg (3,900 rpm in SA-600 rotor) for 15 minutes at 4°C.
11. Resuspend pellet in wash beffer-II to a final volume of 2 ml (1 ml/tube if 2 tubes used); combine 1 ml samples.
12. Add one-tenth volume (200 µl) of proteinase K (10 mg/ml) and incubate for 2 minutes at room temperature.
13. Very gently add one-fifth volume lysis buffer and mix in by slowly inverting tube several times over 10- to 15-minute period at room temperature.
14. Centrifuge 10 minutes at room temperature in IEC clinical centrifuge (top speed = setting 7) to pellet residual starch and cell-wall debris.
15. Pour supernatant into 15 ml falcon tube, add 4.48 g freshly powdered CsCl, 119 µl ethidium bromide (10mg/ml), and distilled H_2O to 5.93 ml. Rock on nutator covered with foil until dissolved.
16. Transfer to 6 ml ultracentrifuge tube, top off with H_2O or 75% CsCl solution, balance, and centrifuge in TFT 45.6 rotor at 40,000 rpm, 22°C, overnight.
17. Extract ethidium bromide several times with isopropanol saturated with H_2O and NaCl (until pink color disappears).
18. Dialyze against at least 3 changes of 2 liters of dialysis buffer (1X).
19. Store chloroplast DNA at 4°C for short-term use and -20°C for long-term use.

Extraction Buffer
0.35 M sorbitol
50 mM tris-HCl, pH 8.0
5 mM EDTA
0.1% BSA
0.25% PVP
10% PEG
0.5% spermine
0.5% spermidine
0.5% 2-mercaptoethanol

Lysis Buffer
5% sarkosyl (lauryl sarcosine)
50 mM tris-HCl, pH8.0
25 mM EDTA

Dialysis Buffer (50X)
500 mM tris-HCl, pH 8.0
500 mM NaCl
5 mM EDTA

Wash Buffer-I
0.35 M sorbitol
50 mM tris-HCl, pH 8.0
25mM EDTA
10% PEG

Wash Buffer-II
Wash Buffer-I without PEG

ii. DNA Elution

1. DNA was eluted out of agarose gels using the Centrilutor, Micro-Electroeluter (Amicon, Danvers, MA).
2. Cut DNA bands out of agarose gel under long wave UV light and place gel slice in perforated, flat-bottomed 0.5 ml microfuge tubes
3. Assemble Centrilutor Micro-Electroeluter as specified. Prepare elution buffer to match the running buffer of the agarose gel. Place microfuge tubes into Centricon Microconcentrators-30 (Amicon) and the microconcentrators into the Centrilutor Unit. Fill unit with elution buffer. Remove all bubbles in the Microconcentrators using a glass pipette.
4. Electroelute DNA for 1-2 hours at 100-400 V.
5. Dismantle unit and remove microfuge tubes from Microconcentrators.
6. Place filtrate cup onto Microconcentrator and spin at 4200 rpm for 30 min. in a SM-24 rotor. Add 1.6 ml H_2O to Microconcentrator and spin again. Flip Microconcentrator, add retentate cup and spin at 2500 rpm for 3 minutes. Determine DNA concentration using the DNA Dipstick (Invitrogen, San Diego, CA).

iii. Polymerase Chain Reaction

1. The following uses the GeneAmp DNA Amplification Reagent Kit (Perkin-Elmer Cetus) reagents.
2. In a sterile 0.5 ml microfuge tube add in order:

ddH_2O	to 100 µl
10X Reaction buffer	10 µl
dNTP mix, 1.25mM	16 µl
5' primer	1.0 µM
3' primer	1.0 µM
template DNA	50-200 ng
Taq Polymerase (Promega)	2.5 U

   Incubate samples at 94°C for 1.5 min. before addition of *Taq*. Spin down samples, add *Taq*, and overlay with 70 µl mineral oil.
3. PCR reactions were performed in a Perkin Elmer Cetus DNA Thermal Cycler using various temperature cycles.

iv. PCR Cloning

1. PCR amplify fragment of interest. Separate product on agarose gel and elute DNA fragment out of gel (Appendix x).
2. Prepare a 1:2 ratio of vector to insert DNA. PCR cloning is done with the TA Cloning System Kit (Invitrogen). In a microfuge tube, add PCR amplified insert DNA (~100 ng), 1 µl 10X ligation buffer, 50 ng vector DNA, 1 µl ligase, and H_2O to 11 µl. Incubate at 12°C overnight.
3. Transform plasmid DNA into TA Cloning System competent cells following kit protocol. Spread transformation reactions on X-Gal plates containing kanamycin. Incubate at 37°C overnight.

v. Vector Dephosphorylation (following Sambrook et al 1989)

1. Digest 10 µg of vector DNA (pUC19) with the appropriate restriction enzyme. Run 0.3 µg on a mini-gel to ensure complete digestion. Phenol:chloroform purigy digest and ethanol precipitate. Dissolve DNA pellet in 90 µl 10 mM Tris-HCl, pH 8.3. Remove 20 ng and save at 4°C.
2. To remaining digest add 10 µl 10X CIP Buffer (BRL) and 5.6 µl 1 U/µl CIP (Calf Intestinal Alkaline Phasphatase) (BRL). Incubate at 37°C for 15 minutes. Add an additional 5.6 µl CIP and incubate at 55°C for 45 minutes.
3. Add to the reaction 1.5 µl 0.5 M EDTA pH 8, 3.8 µl 20% SDS, and 35.4 µl H_2O. Mix well and add 2 µl 10 mg/ml proteinase K (to 100 µg/ml) and 48 µl H_2O. Mix and incubate for 30 minutes at 56°C.
4. Cool reaction to room temperature. Extract twice with 200 µl phenol:chloroform.
5. Add 20 µl 3 M sodium acetate pH 7. Mix. Add 440 µl 100% EtOH and precipitate DNA. Resuspend pellet in 32 µl 10:0.1 TE.
6. Check quality of vector by transforming DH5(alpha) competent cells with both unligated and ligated vector reactions (few blue colonies in both transformations should be observed).

vi. Ligation and Transformation

1. Determine 2:1 molar concentration of ends for vector (pUC19) and insert DNA (eluted conifer cpDNA fragments). Linearize vector DNA with appropriate restriction enzyme and dephosphorylate molecules (Appendix X).
2. In a 1.5 ml microfuge tube, combine vector DNA, insert DNA, 5X Ligation Buffer (BRL), 2 U of T4 DNA Ligase (BRL), and H_2O to a final 30 µl volume. Maintain a 2: 1 ratio of vector to insert and a total DNA concentration of 100-200 ng in the reaction. Add all components except the ligase and vortex. Add ligase and mix thoroughly. Incubate at 15°C overnight.
3. Add 1 µl of the ligation reaction to 20 µl of Library Efficiency DH5(alpha) Competent Cells (BRL) on ice. Mix by swirling pipette tip. Incubate 30 minutes on ice. Place reactions at 42°C for 40 seconds and immediately on ice for 2 minutes. In a transfer hood add 80 µl SOC and shake at 225 rpm for 1 hour at 37°C.
4. Spread each transformation on 3-4 X-Gal plates that contain ampicillan.
5. Incubate at 37°C overnight. White colonies contain those plasmids with inserts. Subcloning Efficiency and MAX Efficiency DH5(alpha) Competent Cells (BRL) may also be used.

SOC	
bactotryptone	2 g
yeast extract	0.5 g
NaCl	1 ml of 1 M
KCl	0.25 ml of 1 M
$MgCl_2$,$MgSO_4$	1 ml of 2 M
glucose	1 mo of 2 M
dH_2O	to 100 ml

vii. Quick Transformation (from Golub 1988)

1. In a 1.5 ml microfuge tube on ice, combine 3 µl competent cells (Subcloning Efficiency DH5(alpha)) (BRL) and 1 µl (<10 ng) of plasmid DNA. Mix gently with pipette tip.
2. Transfer tube immediately to 44°C for 1 minute. Place on ice for 0.5-1 minute.
3. Add 100 µl SOC and spread reaction immediately on 2 X-Gal plates.
4. Incubate overnight at 37°C.

viii. ExoIII Deletion Subcloning and Quickscreens

1. Follow procedures and use reagents provided in the Erase-A-Base System Kit (Promega).
2. Double cut 5-10 µg of circular DNA (plasmid + insert) with restriction enzymes that leave a 4-base 3' overhang protecting the primer binding site and a 5' overhang or blunt end adjacent to the insert from which the deletions are to proceed.
3. Extract with 1 volume phenol:chloroform:isoamyl alcohol and then with chloroform:isoamyl alcohol. Precipitate DNA with 0.1 vol. 2 M NaCl and 2 volumes 100% EtOH.
4. Dissolve dried DNA pellet in 1X ExoIII buffer.
5. Fill the appropriate number of 0.5 ml microfuge tubes with 7.5 µl of S1 mix and leave on ice.
6. Warm the DNA tube to 37°C. Add 60-100 U ExoIII / µg DNA. Mix rapidly. Remove 2.5 µl samples at 30 second intervals into the S1 tubes on ice.
7. After all smaples have been taken, move tubes to room temperature for 30 minutes. Add 1 µl of S1 stop buffer and heat at 70°C for 10 minutes.
8. Analyze 2 µl (~40 ng) samples from each time point and load on a 1% agarose gel.
9. Transfer time points to 37°C, add 1 µl Klenow mix to each, and incubate for 3 minutes. Add 1 µl of the dNTP mix and incubate 5 minutes at 37°C.
10. Move time points to room temperature, add 40 µl ligase mix, and incubate at room temperature for 1 hour.
11. See Appendix x for plasmid transformation protocol.

Quick Screens

1. Pick white colonies from each time point and make a single 2 inch streak of each on selective plates. Incubate overnight at 37°C.
2. Using a pipette tip, scrape the cells and suspend in 40 µl of 10 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM EDTA in a 1.5 ml microfuge tube.
3. Add 1 volume phenol:chloroform and vortex for 1 minute. Centrifuge at 10,000xg for 1 minute.
4. Transfer top phase to fresh tube, add standard DNA loading buffer, and load on 0.8% agarose gels using the BRL Supercoiled DNA Marker (~2-16 kb).

ix Boil Lysis Miniprep (from Dr. Theo Drehers lab. OSU)

* Strain dependent procedure--needs negative endonuclease A mutant like DH5(alpha) (not JM101).

1. Prepare 3 ml overnight cultures.
2. Spin down ~1.5 ml of culture in microfuge tube for 3 min. at 5,000 rpm.
3. Pour off supernatant and add other 1.5 ml of culture to microfuge tube. Spin 3 minutes at 5,000 rpm.
4. Pour off all but about 100 µl of supernatant and resuspend bacteria by vortexing.
5. Have a boiling H_2O bath ready.
6. Add 0.3 ml of STET with lysozyme (0.6 mg/ml lysozyme) to resuspended bacteria.
7. Mix and immediately put into boiling H_2O bath for 90 seconds.
8. Put on ice on way to microfuge (leave on ice for about 1 minute). Spin for 20 minutes at top speed.
9. Remove pellet with sterile toothpick.
10. To supernatant add equal volume isopropanol. Mix well. Place on ice for 5 minutes and spin for 20-30 minutes in microfuge.
11. Wash pellet with 70% EtOH, dry, and resuspend in ~30 µl 10:1 TE.
12. Yield varies from about 20-50 µg. This DNA is good for restriction digest and DNA sequencing. For other purposes (ie. cloning), follow with phenol/chloroform.

STET	
8% sucrose	8 g
5% Triton X-100	5 ml
50 mM EDTA, pH 8	10 ml of 0.5 M
50 mM Tris-HCl, pH 8	25 ml of 1 M

Store at 4°C. Add lysozyme before use.

x. Large-Scale Alkaline Lysis Plasmid Prep (from Strauss lab)

1. Centrifuge 35 ml of overnight culture in an Oak Ridge tuve at 5,000 rpm in the SA-600 rotor for 5 min. and discard the supernatant.
2. Add 1 ml of lysis buffer (with 5 mg/ml lysozyme), vortex and let stand 6 minutes.
3. Add 2 ml of SDS/NaOH solution. Invert sharply 10 times and put on ice for 10 minutes.
4. Add 1.5 ml of cold 5 M KAc/HAc. Invert sharply 10 times and put on ice for 10 min. or longer.
5. Centrifuge at 10,000 rpm for 10 minutes.
6. Optional: Transfer supernatant to a 15 ml Falcon tube, add one half volume (2.5 ml) of phenol/chloroform/IAA, vortex, and centrifuge in the IEC or Econospin centrifuge at top speed.
7. Transfer the supernatant to a new Oak Ridge tube and add 2 volumes (9 ml) of 95% EtOH. Let stand for at least 10 minutes.
8. Centrifuge at 14,000 rpm for 10 min. and discard the supernatant and keep the pellet.
9. Wash the pellet with 70% EtOH and vacuum dry.
10. Resuspend the pellet in 520 µl of 10:0.1 TE, add 3 µl of RNase A, and incubate at 37°C for 30 minutes.
11. Optional: Add 5 µl of Proteinase K (10 mg/ml) and 25 µl of 20% SDS, and incubate for 30 min. at 37°C.
12. Transfer to a microfuge tube, add 1/10 volume (55 µl) of NaAc, 600 µl of chloroform/IAA, then mix well.
13. Spin at 14,000 rpm for 5 min. and transfer 500 µl of the aqueous phase to a new microfuge tube.
14. Add 2 volumes (1 ml) of 95% EtOH, then spin at 14,000 rpm in the microfuge for 10 minutes.
15. Discard the supernatant and wash the pellet with 70% EtOH.
16. Dry the pellet and resuspend in 50 µl TE (10:0.1).

SDS/NaOH (20 ml)
1 ml of 20% SDS
10 ml of 0.4 N NaOH
9 ml of H_2O

5M KAc/HAc (15 ml)
9 ml of 5 M KAc
1.7 ml of acetic acid
4.3 ml of H_2O

Lysis Buffer (10 ml)
0.25 ml of 1 M Tris-HCl, pH 8
0.2 ml of 0.5 M EDTA
0.45 ml of 20% glucose
1 ml of 50 mg/ml lysozyme
H_2O to 10 ml

RNase A
DNase-free, boiled, 10 mg/ml

xi. Alkaline Southern Transfer (from Strauss lab after Reed and Mann 1985)

1. After photographing gel, trim away unused portions and notch upper right corner (ie. removing part of wells; looking down at top of gel).
2. If studying large fragments (>10 kb), acid depurinate by soaking gel in several volumes of 0.25 M HCl for 9 minutes with gentle shaking. Drain and rinse once with dH_2O.
3. Denature DNA by soaking gel in several volumes on 0.4 M NaOH, 20-30 minutes, gentle shaking.
4. Wearing gloves, cut nitrocellulose and 5 pieces of Whatman 3 MM filter paper to dimensions about 2 mm greater than gel. Soak Zetabind in H_2O until evenly wet, 2-3 minutes. Notch nitrocellulose like gel.
5. Wet 3 of the Whatman papers in 0.4 M NaOH 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; 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 on top of gel and carefully lay on Zetabind, lining up gel wells with one edge. Use finger to gently force all bubbles out.
9. Soak other 2 Whatman papers with dH_2O and lay on top of Zetabind, avoiding bubbles.
10. Add about 2 inches of paper towels on top, followed by a glass tray and about 0.5 kg of mass.
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 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. Than bake at 80°C for 1.5-2.0 hours. Store in sealed seal-a-meal bag at room temperature until ready to use. Prewash filters in 0.1X SSC, 0.5% SDS, at 65°C for one hour before first use.

<u>0.25 M HCl</u>
40.4 ml conc. HCl (37%)
H_2O to 2 liters

<u>0.4 N NaOH</u>
80.0 g NaOH
H_2O to 5 liters

<u>20X SSC</u>
876.5 g NaCl
441.0 g sodium citrate
adjust pH to 7.0 w/NaOH or HCl
H_2O to 5 liters

xii. Hexamer Labeling (from Strauss lab)

* Random Primed DNA Labeling Kit from Boehringer-Mannheim

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 (in (lambda) marker DNA is also to be labeled, add just before denaturing).
3. Cool samples on ice and quick spin down.
4. Add Solution A, 3 µl per reaction (Solution A contains equal part of dATP, dGTP, and dTTP).
5. Add reaction mixture, 2 µl per reaction (reaction mixture is hexanucleotide mix in 10X reaction buffer).
6. Add (alpha)dCTP[32P], X µl (ie. 4 µl if 2 weeks old).
7. Mix using pipettor.
8. Add Klenow enzyme, 1 µl per reaction (2 units).
9. Mix using pipettor (total volume for one reaction is 20 µl).
10. Incubate at 37°C for 30 minutes to overnight.
11. Stop the reaction with 2 µl stop buffer (0.2 M EDTA).
12. Heat to 65°C for 10 minutes.
13. Spin down.
14. Spin through Sephadex G-50 column for 10 minutes 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. Add Sephadex G-50, fill column, spin 1 minute at setting 6 in IEC clinical centrifuge (put column in 15 ml falcon tube).
d. Fill and spin until packed Sephadex is at 0.9 ml level.
e. Add 100 µl TE 10:1 and spin 4 minutes 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. Denature in heat block at 95°C for 10 minutes (put hole in upper part of tube--not cap--otherwise cap may blow off).
17. Put on ice, spin down.
18. Dilute 5 µl of sample in 995 µl of TE. 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 along with 1 ml EDTA (10mM final volume).

Prehybridizing Solution (for 50 ml)

H_2O	27.75 ml
20X SSC	15.00 ml
20% SDS	1.25 ml
Denhardts Solution	5.00 ml
Salmon Sperm DNA	0.50 ml (denature 10 min. at 95°C)

Washing Blots (low stringency)
Wash blots in 2X SSC, 0.1% SDS for 30 min. at 65°C, agitating gently. Repeat wash. Wash blots in 2X SSC, O.5% SDS for 30 min. at 65°C, agitating gently. Repeat wash. Blot membranes dry on paper towels. Wrap in saran wrap. Expose to X-Ray film.

xiii. Double-Stranded DNA Sequencing

Template Labeling

1. Dissolve 2µg of plasmid DNA in 0.2 N NaOH, 0.2mM EDTA to denature (Chen and Seeburg 1985). Incubate at 37°C for 30 min. (reference?). Neutralize solution by adding 2µl of 2 M ammonium acetate-acetic acid, pH 4.5 and precipitate DNA with 2 volumes of ethanol. Dry DNA under vaccuum and use immediately for sequencing.
2. Resuspend DNA in 7 µl dH_2O. Following the Sequenase Version 2.0 sequencing protocol and using kit reagents (United States Biochemical Corporation, Cleveland, Ohio), add 0.5 pmol primer, 2 µl 5X Sequenase annealing buffer, and H_2O to 10µl.
3. Incubate sample at 65°C for 2 min. Gradually decrease temperature to below 35°C over 30-45 min.
4. To the annealed primer-template, add 1µl 0.1 M DTT, 2 µl Labeling Nucleotide Mix, 5 µCi (0.5 µl) ['alpha'-^{35}S]dATP, and 3 U (2µl) diluted Sequenase Version 2.0. Incubate for 2-5 min. at room-temperature.
5. Prepare 4 tubes labeled G, A, T, and C for each reaction sample. Add 2.5 µl of the appropriate termination mixture to the tubes. Pre-warm tubes to 37°C. Transfer 3.5µl of the completed labeling reaction to each termination tube. Incubate at 37°C for 2-5 min. Add 4 µl Stop Solution to each tube, mix, and store at -20°C.
6. To load onto gel, heat samples at 95°C for 2 min., place on ice for several minutes, and load 3 µl per lane.

Gel Preparation, Polyacrylamide

1. Thoroughly clean Bio-Rad Sequi-Gen Sequencing Cell (Hercules, CA) glass plates with BonAmi cleanser. Rinse completely with dH_2O, dry with Kimwipes, and rinse with 95% EtOH. Prepare bonded glass plate with RainX following instructions (plate can also be silanized using silane solution).
2. Assemble gel form using side spacers and clamps washed with 95% EtOH.
3. Place foam pad and filter paper strip in casting tray, prepare casting solution, and pour solution into tray. Place gel form into tray, tighten screws, and hold in upright position for 2 minutes.
4. Prepare gel solution. Degas for 10-15 minutes and filter through Whatman #4 filter paper. Add TEMED and ammonium persulfate. Pour gel solution between the 2 glass plates of the gel form. Allow gel to set-up for 1 hour. Pre-run gel for 1 hour in 1X TBE running buffer. Load samples and run for 1.5-16 hours.
5. Dissamble gel form after run. Discard lower buffer chamber solution in radioactive waste. Fix gel in a 10% MeOH, 10% acetic acid solution for 45 min. Place gel on a Whatman #3 filter paper and dry under vaccuum at 80°C for 1.5 hours.
6. Expose dried gel to X-ray film for 1-2 days. Develop film.

10X TBE (for 38 x 50 cm gel)	
Trizma	21.6g
Boric Acid	11.0g
EDTA	1.86g
H_2O	to 200ml

Casting Gel
20ml gel solution
140µl 25% ammonium persulfate
100µl TEMED

Gel Solution (6%)	
H2O	48.8ml
10X TBE	17.0ml
lq acrylamide, 30%	24.3ml
lq bisacrylamide, 30%	25.5ml
urea	71.57g

Sequencing Gel
150ml gel solution
320µl 25% ammonium persulfate
80µl TEMED

Gel Preparation, Hydrolink Long Ranger (AT Biochem, Malvern, PA)

1. Prepare gel form as above using uniform, not wedge spacers (running time will be slower with the wedge spacers used with polyacrylamide gels).
2. Prepare gel solution and filter through Whatman #4 paper (no degassing neccessary).
3. Run gel in 0.6X TBE running buffer. Pre-run gel only 10-15 minutes before loading samples. Long Ranger gels to not need to be fixed after the run. All other procedures are as for polyacrylamide gels.

Gel Solution (5%)	
urea	46.2g
10X TBE	11.0ml
50X long ranger	11.0ml
H_2O	53.8ml

Sequencing Gel
90ml gel solution
220µl 25% ammonium persulfate
55µl TEMED

Automated Sequencing

1. Provide 1.2µg of plasmid DNA template in 7 µl dH_2O. All labeling reactions are carried out by the Central Services Laboratory, Center for Gene Research and Biotechnology, Oregon State University.
2. Samples are run on an Applied Biosystem Model 373 DNA Sequencer at the Central Services Lab, OSU. Readable sequence is between 300-600 bp / sample.

xiv. LB Media and X-Gal Plates

LB Media

tryptone	10 g
yeast	5 g
NaCl	10 g
H_2O	to 900 ml

Adjust pH to 7.5 with 5 N NaOH. Bring volume to 1 liter with dH_2O. Autoclave. Cool to room temperature. Store in sterile cabinet.

X-Gal Plates

tryptone	10 g
yeast	5 g
NaCl	10 g
H_2O	to 900 ml

Adjust pH to 7.5 with 5 N NaOH. Add 15 g bacto-agar. Bring volume to 1 liter. Autoclave. Cool to 60°C. Add 100 mg ampicillan, 40 mg X-Gal, an 8.8 mg IPTG. In transfer hood, pour ~25 ml media per petri dish. Store plates in the dark at 4°C. Plates are good for 1-1.5 months.

xv. Long-Term cpDNA Clone Storage

In cryovial, add 0.9 ml of overnight plasmid culture. Fill remainder of cryovial (to mark on vial) with 100% glycerol. Mix well. This makes a 50% glycerol storage stock. Store at -80°C.