AN ABSTRACT OF THE THESIS OF

Ching-Hsiu Tsai for the degree of Master of Science in Genetics presented on Feb. 10, 1989

Title: Dispersed Repetitive Sequences in the Chloroplast Genome of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco)

Redacted for privacy

Abstract approved: Steven H. Strauss

Chloroplast genomes of conifers are unusual in a number of respects, including the possession of much repetitive DNA. By studying the distribution and sequences of dispersed repetitive DNA, I hoped to gain insight into the mechanisms of chloroplast genome evolution.

I used restriction mapping and DNA sequencing to characterize dispersed repetitive DNA in the chloroplast genome of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco). A chloroplast DNA clone bank was constructed in a pUC19 plasmid vector. High stringency hybridization among these recombinant plasmids, and to restricted chloroplast DNA, was used to map repeat families. Dispersed repeats were grouped into three classes: long repeat families (around 600 bp), medium repeat families, and short repeat families (10-20 bp). The long and medium repeats were classified into a total of six families and placed on a

restriction site map.

Sequence analysis of one repeat family shared among three restriction fragments revealed the presence a 633 bp inverted repeat. The repeat contains a complete transfer RNA-Serine (GCU) gene and an open reading frame (ORF36). This open reading frame is highly homologous with chloroplast sequences in tobacco and liverwort. Both ends of the repeat possess transposon-like structures. Two copies of the dispersed repeats are located near the endpoints of a large inversion that distinguishes Douglas-fir and radiata pine. The third copy of the repeat is also located near the endpoint of an inversion that occurred during conifer evolution, and is associated with a length mutation hotspot, and short inverted repeats; the latter appear to be capable of forming numerous stem-loop structures.

Sequence comparisons with tobacco cpDNA revealed that one of the Douglas-fir chloroplast DNA clones is homologous to three tobacco chloroplast DNA segments. These regions in tobacco appear to have been juxtaposed in Douglas-fir by a number of inversions and a small deletion. An 8 bp (CATCTTTT) direct repeat in tobacco is sandwiched between two inverted sections in Douglas-fir. This repeat may be a target sequence for homologous recombination.

Dispersed Repetitive Sequences in the Chloroplast Genome of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco)

by

Ching-Hsiu Tsai

A Thesis

submitted to

Oregon State University

in partial fulfillment of the
requirements for the degree of
Master of Science

Completed February 10, 1989
Commencement June, 1989

APPROVED:

Redacted for privacy

Assistant Professor of Forest Science in charge of major

Redacted for privacy

Head of department of Genetics

Redacted for privacy

Dean of Graduate School

Date thesis is presented Feb. 19, 1989

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Steve Strauss, who provided me with the opportunity to study in his laboratory. I thank Dr. Jeff Palmer, who provided both helpful advice and DNA clones. During this two year study period, Allan Doerksen helped a great deal in the laboratory, so did my fellow graduate students in Dr. Strauss' lab, Glenn Howe, Barry Goldfarb, Valerie Hipkins, and Yong-Pyo Hong.

TABLE OF CONTENTS

	PAGE
INTRODUCTION AND ORGANIZATION OF THESIS	1
LITERATURE REVIEW	3
Chloroplast Genome Organization and Evolution	3
Dispersed Repetitive DNA in Chloroplast Genome	8
DISPERSED REPETITIVE SEQUENCES IN THE CHLOROPLAST GENOME OF DOUGLAS-FIR (*Pseudotsuga menziesii* (Mirb.) Franco)	15
Summary	16
Introduction	17
Material and Methods	18
Results	20
Discussion	24
Figures	29
References	40
CONCLUSIONS	42
BIBLIOGRAPHY	44
APPENDICES	52
I. Supplementary data.	52
A. XbaI/PvuII/SstI Douglas-fir restriction fragment sizes	52
B. Hybridization data from Douglas-fir cpDNA restriction enzyme mapping	55
C. Hybridization data from dispersed repeat mapping experiments	60

D. Additional Douglas-fir chloroplast DNA sequences 63
II. Supplementary detail on methods. 72
A. Procedures for chloroplast DNA isolation 72
B. Protocol for chloroplast DNA cloning 76
a. Vector preparation 76
b. Ligation reaction 78
c. Transformation 80
C. Protocol for plasmid isolation 81
a. Minipreparation of plasmid DNA 81
b. Large scale preparation of plasmid DNA 83
D. Protocol for plasmid sequencing 85
III. Supplementary detail on results. 88
A. Effect of agarose on transformation efficiency 88
B. Effect of long and short wave UV light on cloning efficiency 90
C. Determination of the chloroplast DNA restriction fragment size 92
D. Douglas-fir chloroplast DNA cloning 94

LIST OF FIGURES

FIGURE PAGE

1. Douglas-fir cpDNA (chloroplast DNA) restriction site map for SstI, PvuII, and XbaI. 29
2. Mapping of dispersed repeats and cloned fragments of Douglas-fir cpDNA. 31
3. Autoradiograms from hybridization of cloned XbaI fragments containing dispersed repeats to Douglas-fir cpDNA mapping blots. 33
4. Sequence of the 633 bp dispersed repeat family. 35
5. Sequences from the 2.7 kb XbaI fragment showing several imperfect direct repeats. 37
6. Aligment of the DNA sequences between the 1.4 kb XbaI fragment of Douglas-fir and tobacco cpDNA. 38
7. A hypothetical model for inversions and deletions that occurred during evolution of a section of the Douglas-fir chloroplast genome from a tobacco-like ancestral genome. 39
8. Restriction digest of Douglas-fir cpDNA and autoradiogram from hybridization of subcloned dispersed repeat fragments to blotted cpDNA. 55
9. Autoradiograms from Douglas-fir restriction site mapping using cloned Douglas-fir cpDNA probes. 56

10. Autoradiograms from Douglas-fir restriction site mapping using labelled cpDNA fragments cut from a gel. 57

11. Electrophoresis and a hybridization with the Douglas-fir XbaI-cpDNA clone bank. 60

12. Autoradiogram from dideoxy chain termination DNA sequencing using ^{32}P and ^{35}S-labelled nucleotides. 63

13. The DNA sequence of the 1.4 kb XbaI cpDNA clone. 64

14. The restriction map of the 2.7 kb XbaI cpDNA clone. 65

15. Sequence of the Xba2kpn fragment (a subclone of the 2.7 kb XbaI cloned fragment). 66

16. Sequence of the Xba2pvu fragment (a subclone of the 2.7 kb XbaI cloned fragment). 67

17. Possible secondary structures of the Xba2pvu fragment. 68

18. Aligned sequences for tRNA-Serine (GCU) genes for Douglas-fir, tobacco, and wheat. 69

19. Comparision of ORF36 cpDNA sequences from Douglas-fir, tobacco, and liverwort. 70

20. The DNA sequence of a 0.5 kb XbaI cloned cpDNA fragment. 71

LIST OF APPENDIX TABLES

TABLE	PAGE
1. Sizes of Douglas-fir cpDNA restriction fragments generated by XbaI, PvuII and SstI.	52
2. Sizes of Douglas-fir cpDNA restriction fragments generated by PvuII/XbaI and SstI/XbaI double digests.	53
3. Summary of hybridizations of cpDNA clones for restriction site mapping.	58
4. Summary of hybridizations of cloned cpDNA fragments (inserts) for mapping of dispersed repeats.	61

LIST OF TABLES

TABLE PAGE

1. Sizes of Douglas-fir cpDNA restriction fragments generated by XbaI, PvuII and SstI. 52
2. Sizes of Douglas-fir cpDNA restriction fragments generated by PvuII/XbaI and SstI/XbaI double digests. 53
3. Summary of hybridizations of cpDNA clones for restriction site mapping. 58
4. Summary of hybridizations of cloned cpDNA fragments (inserts) for mapping of dispersed repeats. 61

Dispersed Repetitive Sequences in the Chloroplast Genome of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco)

INTRODUCTION AND ORGANIZATION OF THESIS

Molecular evolution of chloroplast DNA (cpDNA) has received considerable study. Most available information is from work on angiosperms; gymnosperms have received little attention. The major sources of mutation in chloroplast genomes are nucleotide substitutions and length mutations. However, the cpDNA of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) and other conifers show evidence of a number of major rearrangements. They lack the large inverted repeat that is present in the cpDNA of nearly all land plants, and possess many large inversions (Strauss et al., 1988). Inversions and deletions may be potentiated by dispersed repetitive DNA, which can act as substrates for homologous recombination. Thus, study of dispersed repetitive DNA sequences may provide insight into the unusual evolutionary history of conifer cpDNA.

This thesis is written in manuscript format, following the Oregon State University thesis preparation guidelines. The text part of the thesis contains six sections, introduction, literature review, manuscript, conclusions, bibliography, and appendices. There is a short review of chloroplast genome evolution and dispersed repetitive cpDNA

in the literature review section. The manuscript section follows guidelines of the journal Current Genetics and thus differs from other parts of the thesis. The materials and methods portion of the manuscript is brief; much more detailed techniques and protocols are listed in the appendices section. The references in the manuscript do not include titles, however, full citations can be found in bibliography section of the thesis. The appendices section also contains detailed data which are only briefly referred to in the manuscript.

LITERATURE REVIEW

Chloroplast Genome Organization and Evolution

With the exception of primitive amoeba Pelomyxa, a proto-eukaryote which lacks mitochondria (Whatley, 1979), all eukaryotic cells are thought to harbor two physically and functionally distinct genomes, the nucleus and the mitochondrion. Photosynthetic eucaryotes harbor a third genome within their plastids. Two hypotheses have been proposed to explain the origin of the extranuclear genomes. (i) The endosymbiont or xenogenous origin hypothesis: nuclear genomes and organellar genomes initially inhabited different sorts of cells. (ii) The direct filiation or autogenous origin hypothesis: nuclear and organellar genomes became physically compartmentalized and functionally specialized within a single sort of cell (Gray and Doolittle, 1982). Recent studies strongly support the first hypothesis; chloroplasts in plants are the descendants of a photosynthetic procaryote.

Each chloroplast contains tens to hundreds of copies of its circular genome, and there may be as many as 20-50 chloroplasts per cell. Chloroplast DNA (cpDNA) of land plants ranges in size from 120 kilobases (kb) to 217 kb (see reviews, Whitfeld and Bottomley, 1983; Palmer, 1985a; Palmer, 1987a). This size variation is primarily associated with variation in size of the large inverted repeat, which

is present in almost all land plant chloroplast genomes, and which separates the circular cpDNA molecule into large and small single-copy DNA regions (Whitfeld and Bottomley, 1983). In general, the repeat region contains ribosomal RNA genes (rDNA) (16S-23S-5S), some transfer RNA (tRNA) genes, and some protein-coding genes. The largest land-plant cpDNA, the 217 kb genome of Pelargonium hortorum, contains the largest known inverted repeat, 76 kb (Palmer et al., 1987a). Some of the smallest genomes (120-140 kb) among angiosperms have been found in a group of legumes whose genomes have lost one entire copy of the repeat. Recently, a small genome (120 kb) and loss of the inverted repeat has also been found in several conifers, including Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) and radiata pine (Pinus radiata D. Don) (Strauss et al., 1988; Lidholm et al., 1988).

Chloroplast genomes are rather uniform in gene organization. In general, chloroplast genomes contain a complete set of rDNA and tRNA genes, and about 50 protein encoding genes. Three cpDNAs have been completely sequenced, liverwort (Marchantia polymorpha) (Ohyama et al., 1986), tobacco (Nicotiana tabacum) (Shinozaki et al., 1986), and rice (Oryza sativa) (Hira et al., 1988). The total number of genes encoded by the chloroplast genome was estimated to be 128 in liverwort (Ohyama et al., 1986) and 82 in tobacco (Shinozaki et al., 1986). A dox-matrix

comparison of tobacco and liverwort cpDNA showed that there are 74 homologous open reading frames (ORFs) conserved in length and amino acid sequence (Wolfe and Sharp, 1988). Because most chloroplast genomes encode the same set of genes, it has been suggested that most of the chloroplast progenitor's genes were deleted and/or transferred to the nucleus soon after the establishment of endosymbiosis (Palmer, 1985a; Palmer, 1987b.).

Chloroplast gene order sometimes differs among species (Palmer, 1985a), however, a number of lines of evidence indicate that there are severe constraints on chloroplast genome rearrangement. (1) All well-characterized rearrangements are inversions, though some recent evidence indicates transposition might also occur (Zhou et al., 1988; Milligan et al., 1988). (2) Most cases of rearrangement are simple, involving one or two discrete inversions. (3) Two highly rearranged kinds of genomes, those of Pelargonium and some legumes, have a highly rearranged or deleted inverted repeat, respectively, suggesting that it somehow constrains rearrangement (rev. in Strauss et al., 1988). (4) Endpoints of the inversions from wheat and lettuce are located entirely within noncoding regions; compared to plant nuclear and mitochondrial genomes, non-coding regions comprise a relatively small portion of total cpDNA. (5) Some of the highly rearranged genomes contain families of large dispersed repeats (of several hundred bp) (Palmer,

1987a,b); such repeats appear to be lacking in most chloroplast genomes (Palmer, 1985a). (6) Only one of the rearrangements that have been characterized to date disrupt a co-transcribed group of genes (Milligan, 1989).

The stable sizes of chloroplast genomes contrast with the great variation in size of both of nuclear (10^5-10^8 Kb) and mitochondrial (200-2400 kb) genomes of plants (Sederoff, 1987). DNA in plant mitochondria is a heterogeneous mixture of linear and circular molecules of varying sizes (Sederoff, 1987). Plant mitochondrial DNA (mtDNA) changes rapidly in size and structure, but slowly in primary sequence (Palmer, 1985b; Wolfe et al., 1987). This may be a consequence of the ability of plant mitochondria, but not chloroplasts, to fuse -- providing an opportunity for recombination between genomes in different mitochondria. Enhanced recombination would favor both rearrangements and copy correction, thereby explaining the combination of rapid structural change and slow sequence evolution (Lonsdale et al., 1988). Gene order in chloroplast genomes are similar in the majority of land plants, the differences among species being largely due to a limited number of inversions as described previously. Gene order of plant mtDNA is highly variable as a result of a high rate of both inversion and length mutation. In contrast to cpDNA, plant mtDNA usually contains for more DNA than is needed to encode genes, and for more repetitive

DNA, thus providing more opportunities for recombination.

Dispersed Repetitive DNA in Chloroplast Genomes

Besides the large inverted repeats found in cpDNA, large tandem repeats have been described in two algal species. Acetabularia mediterranea contains five copies of a 10 kb tandem repeat, and Euglena gracilis contains between one and five tandemly arrayed copies of a 6.2 kb repeat. Dispersed repeats also have been reported in Chlamydomonas reinhardtii and C. smithii which contain 25-40 short (100-300 bp) inverted repeats dispersed throughout the genome, and in subclover (Trifolium subterraneum), which contains at least six copies of a 1 kb repeat dispersed throughout the genome (Milligan et al., 1988).

Dispersed repeats have been mapped in wheat (Triticum aestivum) by using low-stringency hybridizations with repeat-containing cloned cpDNA fragments. There were thirty-two repeated DNA segments dispersed through the genome, which could be grouped into 12 families (Bowman and Dyer, 1986).

cpDNA repeats may be substrates for homologous recombination, and thus play a role in genome rearrangement. Recombination among repeated sequences has been proposed as an important mechanism for the restructuring and evolution of nuclear and mitochondrial genomes (Flavell, 1986). In prokaryotic genomes, site-specific DNA rearrangements are involved in a variety of biological functions including DNA transposition, faithful

partitioning of extrachromosomal elements, gene regulation, and the promotion and regulation of genetic divisity (Simon and Herskowitz, 1985). There are a number of well characterized systems that mediate site-specific recombination. (1) The *lox*-*cre* system in bacteriophage P1 requires two 34 bp recombination sites (*lox*) and a single polypeptide recombinase encoded by the *cre* gene (Abremski and Hoess, 1984). (2) Resolvase is associated with the Tn3 transposon that mediates site-specific resolution (deletion of a specific DNA segment between two direct repeats) (Krasnow and Cozzarelli, 1983). (3) Site-specific integration by bacteriophage lambda integrase (Int-*att*) requires a complex site (*att*P) and a host factor to initiate the integration reaction (Weisberg and Landy, 1983). (4) A group of recombinase families include the products of the *hin*, *gin*, *cin*, and *pin* genes. They mediate site-specific inversion and they are all found in different genetic contexts. The *hin* gene regulates inversion of a 996 bp segment of DNA that includes the promoter controlling transcription of a flagellin gene in *Salmonella*. It requires two 26 bp recombination sites in inverted configuration (Johnson and Simon, 1985), a 60 bp sequence that increases the recombination rate 150 fold, a host protein (Factor II), and one of the major histone-like proteins of *E. coli* HU (Johnson et al., 1986). The Gin function of bacteriophage Mu catalyzes inversion of the G

DNA segment, which results in switching of the host range of Mu phage particles. This site-specific recombination takes place between inverted repeat sequences (IR) that border the G segment (Kahmann et al., 1985). The crossover sites for site-specific C inversion consists of imperfect 12 bp inverted repeats. The phage P1 Cin recombinase can act at a mismatched position within a 2 bp sequence (Iida and Hiestand-Nauer, 1986).

All the well characterized site-specific recombination systems in prokaryotes require not only repetitive DNA sequences (direct or inverted), but also at least one protein. The proteins can be either from the phage alone, or from both phages and hosts. Recombination systems have not to date been characterized in plant chloroplast genomes. Therefore, the recombination related proteins could be among the many uncharacterized ORFs in chloroplast genomes, or encoded from the nuclear genome and transported into the chloroplast genome.

Repeated sequences have been found at inversion endpoints of cpDNA in several species. The endpoints of an inversion in wheat chloroplast DNA are associated with copies of a short (70 bp) repeated sequence (Howe, 1985). This repeated sequence contains an element homologous to the core of the bacteriophage lambda att-site which is the locus of the recombination between phage and bacterial genomes. Sequence comparisons of inversion junction regions among Marchantia, tobacco, and spinach showed that

inversions may have been proceeded by transposition events. Two 8 bp direct repeats flank two 10 bp indirect repeats near the endpoints of a 30 kb inversion. These repeats may be "footprints" of a transposon-mediated insertion, which may then have facilitated a subsequent inversion via homologous recombination (Zhou et al., 1988).

Additional evidence of transposition in cpDNA has been reported in _Trifolium_. There are at least six copies of 1 kb sequence present in the subclover (_Trifolium subterraneum_) chloroplast genome. Because one of these repeats is inserted within a normally highly conserved, co-transcribed group of genes, these repeats have been proposed to be the result of transpositions. The repeats presumably then mediated the subsequent _T. subterraneum_ cpDNA gene rearrangements (Milligan et al., 1989).

Transfer RNA (tRNA) genes have also been associated with inversions. A 119 bp duplication located at an inversion junction region has been reported in the wheat chloroplast genome, and this duplication contains the $tRNA^{Met}$-CAU gene (Quigley and Weil, 1985). Two of three inversions that occurred during the evolution of wheat chloroplast DNA are not associated with short repeats, although there are repeats present just beyond the endpoints of the inversion. However, all of the three inversions are adjacent to at least one tRNA gene, and three of the tRNA genes have been partially duplicated,

possibly at the time of inversion (Howe et al., 1988).

Gene duplication of tRNA also has been observed in the *Vicia faba* chloroplast genome. The intergenic region of two $tRNA^{Leu}$ genes contains a series of short repeats and a partial copy of the $tRNA^{Leu}$(UAA) gene. This duplication may have occurred during the rearrangement of the two $tRNA^{Leu}$ genes in broad bean, or at the time of deletion of one copy of the large inverted repeat (Bonnard et al., 1985).

An association of tRNA genes with genome rearrangements has also been observed in mitochondrial genomes. Repeated $tRNA^{Met}$ genes may have been the target sequences of recombination that generated chromosome heteromorphism in *Neurospora crassa* (Gross et al., 1984). In mitochondria of *Podospora anserina*, tRNAs are associated with the excision-junction sites of mitochondrial excision-amplification plasmids. It was proposed that these tRNA genes might form secondary structure to bring the excision site regions into close proximity (Turker et al., 1987).

Dispersed repetitive DNA sequences in chloroplast genomes may result from several causes. First, transposition may result in the production of short inverted repeats and/or direct repeats at excision sites, and result in duplication of the entire element at various positions in the genome (Zhou et al.,1988; Milligan et al., 1989). Second, duplications may have occurred at the time of deletion of the large inverted repeat. In pea (*Pisum sativum*), which lacks an inverted repeat, a roughly 200 bp

DNA sequence includes two regions that were derived by duplication from portions of the psbA and rbcL genes (Wolfe, 1988; Bonnard et al. (1985). Third, tRNA may play a role in inversions by as yet unknown mechanisms. The extensive secondary structure of stem-loop regions of tRNA genes possible during replication might predispose these regions to cleavage, perhaps by a process similar to intron splicing (Cech, 1987). The result may be a high frequency of single stranded regions which may facilitate recombination, leading in some cases to inversions. After ligation and filling in of these inverted single-strand regions, complete or partial tRNA duplications that flank inversion endpoints would result. Associations of tRNA duplication and inversion endpoints have been noted in bean (Bonnard et al., 1985), and wheat cpDNA (Quigley and Weil, 1985; Howe et al., 1988).

Study of conifer cpDNA may provide a number of insights into the mechanisms of cpDNA evolution. Douglas-fir and radiata pine cpDNA possesses a number of rearrangements relative to the ancestral land plant chloroplast genome. These rearrangements appear to have resulted from inversions and deletion of the large inverted repeat (Strauss et al., 1988). Loss of the large inverted repeat appears to potentiate genome rearrangements (Palmer, 1982), and accumulation of dispersed repetitive DNA (Milligan et al., 1989). Conifer cpDNA possesses regions

prone to a high frequency of length mutations (Wagner et al., 1987), and these appear to be associated with dispersed repetitive DNA (unpublished data). Therefore, by studying conifer cpDNA, which is highly rearranged, lacks a large inverted repeat, contains dispersed repetitive DNA, and possesses mutation hotspots, we may gain insight into several of the factors controlling chloroplast genome evolution.

DISPERSED REPETITIVE SEQUENCES IN THE CHLOROPLAST GENOME OF DOUGLAS-FIR (*Pseudotsuga menziesii* (Mirb.) Franco)

Ching-Hsiu Tsai

and

Steven H. Strauss

Department of Forest Science

Peavy Hall 154

Oregon State University

Corvallis, Oregon 97331-5705

U. S. A.

Running Head: Dispersed repeats in cpDNA

For submission to: Current Genetics

Summary

We used restriction mapping and DNA sequencing to characterize dispersed repetitive DNA in the chloroplast genome of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco). To map repeat families, a cpDNA clone bank was constructed in a plasmid vector and hybridized at high stringency to one another and to restricted cpDNA. Short repetitive DNA sequences of about 10-20 bp are dispersed throughout the genome. Longer repeats are clustered in 4 regions of the genome and comprise six families. Sequence analysis of one repeat family shared among three fragments indicated the presence of a 630 bp inverted repeat which contains a complete tRNA-Serine (GCU) gene and a highly conserved open reading frame (ORF 36). Both ends of this 630 bp dispersed repeat have a transposon-like combination of short direct and inverted repeats. One clone which contains one of the endpoints of a major inversion is homologous to three tobacco cpDNA segments that appear to have been juxtaposed in Douglas-fir by inversions and a small deletion. An 8 bp (CATCTTTT) direct repeat in tobacco is sandwiched between two inverted sections in Douglas-fir; it may be a target sequence for homologous recombination.

Key words: Chloroplast DNA - Repetitive DNA - Inversion - Recombination - Transposition

Introduction

The chloroplast genome is highly conserved in size and gene arrangement among the majority of land plants (reviewed in Palmer 1985). However, conifers (Strauss et al. 1988; Lidholm et al 1988) and a subfamily of legumes (Palmer et al. 1987) are unusual in lacking the large inverted repeat. Some species in these groups also show considerable rearrangement, and possess much dispersed repetitive DNA (Milligan et al. 1989). The loss of the large inverted repeat, and the accumulation of dispersed repetitive DNA, have been proposed to accelerate chloroplast DNA (cpDNA) rearrangement (Palmer and Thompson 1982; Palmer 1985).

A number of families of dispersed repetitive DNA were found during the cloning and mapping of the Douglas-fir chloroplast genome. The goals of this paper are to describe their distribution in the genome, and to study in detail one pair of repeat-containing fragments that are associated with inversions that occurred during conifer evolution (Strauss et al. 1988).

Materials and Methods

Isolation and cloning of chloroplast DNA. cpDNA was isolated from needles of Douglas-fir by sucrose- and cesium chloride-gradient methods as described by Palmer (1986) and modified by Strauss et al. (1988). (Detailed protocol given in Appendix IIA)

cpDNA was digested with XbaI and separated by electrophoresis in low melting point agarose (Boehringer Mannheim) gels (Maniatis et al. 1982). DNA fragments were cut out from gels under long-wave UV light and ligated directly (in gel) into a dephosphorylated pUC19 plasmid vector and transformed into *E. coli* (DH5 α). Small fragments which were not visualized via ethidium bromide staining of gels were cloned via shotgun cloning of unfractionated cpDNA into the same vector and host as above. (Detailed cpDNA cloning procedures are given in Appendix IIB)

Dispersed repeat identification and mapping. Partially purified cpDNA was digested with several enzymes, and separated by gel electrophoresis, and blotted onto nylon membranes (Zetabind, AMF-Cuno). Blots of cpDNA were stripped and reprobed for restriction site mapping. Blots of cpDNA clones were used only once, however, to insure that no carry over of signals confounded repeat identification. Methods of hybridization and alkaline

plasmid DNA preparation were as described (Maniatis et al. 1982; Tsai 1989). (A detailed protocol for plasmid isolation is given in Appendix IIC)

Dispersed repeats were identified by high stringency hybridization of cloned Douglas-fir cpDNA fragments to one another: blots were washed four times for 30 minutes, twice with 2X SSC (0.3M NaCl, 0.03M citric acid), 0.1% SDS (sodium dodecylsulfate), and twice with 0.1X SSC, 0.5% SDS at 65°C (Maniatis et al., 1982). The cloned fragments (inserts) were cut from gels and directly labeled with ^{32}P via primer extension using random hexamer as primers (Feinberg and Vogelstein 1983).

Dispersed repetitive DNA sequencing. Sequences were determined by the dideoxy chain termination method of Sanger et al. (1977). Plasmids were directly sequenced following Zhang et al. (1988). (A detailed protocol is given in Appendix IID)

Sequence analysis. Douglas-fir cpDNA sequences were compared to the complete cpDNA sequence of tobacco using GenBank Release 56 accessed via the computational molecular biology lab of the Center for Gene Research and Biotechnology at Oregon State University. Similarity searches were also done at the computational molecular biology lab using the FASTA program (Pearson and Lipman 1988).

Results

<u>Restriction endonuclease analysis and cloning of Douglas-fir cpDNA</u>. At least thirty discernible fragments are generated when Douglas-fir cpDNA is digested with XbaI. All putative clones were screened both by size and by hybridization with corn cpDNA. Because our cpDNA preparations from Douglas-fir contained significant nuclear DNA contamination, the latter step was necessary to insure that our clones were indeed cpDNA. Twenty-one of at least 30 total XbaI fragments were cloned (Fig. 1); fragments of 13.8, 10.6, 5.9, 1.5 kb and some other small fragments which were identified by detailed mapping remain uncloned (Appendix IIIC). A restriction map shows the position of XbaI sites relative to PvuII and SstI (Fig. 1); the latter were mapped with respect to SmaI, KpnI, and a number of chloroplast genes by Strauss et al. (1988).

<u>Detection and mapping of dispersed repetitive DNA sequence</u>.

There are at least six dispersed repetitive families that were detected by high stringency hybridization (Figs. 2 and 3). These repeats were placed on an XbaI restriction site map (Fig. 1). Based on relative signal strength on autoradiograms, dispersed repeats could be classified into three arbitrary categories: long dispersed repeat families (strong signal), medium dispersed repeat families, and short dispersed repeat families (faint signal). Based on

sequencing (results below), the long repeat families contain around 600 bp and the short repeat families about 10-20 bp. Short repeats appear to be dispersed throughout the genome, and are thus difficult to group into families (Details given in Appendix I, Table 4). We grouped the strong and medium signals into six families. The long repeat families include the 1.4, 2.7 and 3.8 kb XbaI fragments; 0.8 and 3.2(a) kb fragments; 4.3 and 4.8 kb fragments; and, 2.5 and 3.2(b) kb fragments. The medium signal families are 1.4, 1.5(a), 3.8 and 4.3 kb fragments; and 2.5, 3.2(a) and 3.2(b) kb fragments (Fig.1).

Sequencing of dispersed repeats. The 3.8, 2.7 and 1.4 kb XbaI fragments comprise dispersed repeat family 1 (Fig. 1). The 1.4 kb clone was completely sequenced, and the 2.7 kb clone was partially sequenced. Comparison of these two sequences revealed a 633 bp near-identical repeat (97.9% similarity) with an inverted orientation (Fig. 4). There is an open reading frame (ORF36) and a tRNA-Serine (GCU) gene within this dispersed repeat. The ORF36 fragment in Douglas-fir is highly homologous with tobacco cpDNA ORF36a (8398-8508) and liverwort cpDNA ORF36a (23107-22997). There is 84.7% similarity (94/111) between Douglas-fir and tobacco, 86.5% (96/111) between Douglas-fir and liverwort, and 85.6% (95/111) between tobacco and liverwort. (Appendix I, Fig. 19)

Transposon-like insertion sequences at both ends of the dispersed repeats. At both ends of the 633 bp dispersed repeat there is a 6 bp complementary inverted repeat GGAAAA at the 5' end (one bp missing in the 1.4 kb fragment) and TTTTCC at the 3' end. There is an apparent direct repeat flanking the indirect repeat: CCATTTT at 5' end and CCTATTT at 3' end (1.4 kb fragment only). In the 1.4 kb fragment there is an added base pair in the CCTTTT 5' repeat, and in the 2.7 kb fragment the direct repeat at the 3' end is missing (Fig. 4).

Short direct repeats near to the dispersed repeat. We subcloned a 0.9 kb XbaI-KpnI double digest fragment from the 2.7 kb XbaI fragment containing the 633 bp repeat. Sequences near to the dispersed repeat contain several different short repeats (Fig. 5). Two fifteen bp direct repeats (AATCTTTCTTTTTAT) flank an 8 bp section that contains two 3 bp (CGT) direct repeats. A pair of direct, tandem repeats farther from the 633 bp dispersed repeat is about 44 bp long and 88.6% homologous. Additional tandem repeats are present in the adjoining unsequenced region (Hipkins and Strauss, pers. comm.).

Sequence of an inversion endpoint. Hybridizations to restricted Douglas-fir and radiata pine cpDNA indicated that the 1.4 kb XbaI fragment spans an endpoint of a 40-50 kb inversion that distinguishes these species (hybridization data not shown; Strauss et al. 1988). This

inversion is also lacking in petunia and most other angiosperms, including tobacco (Strauss et al. 1988).

The 1,348 bp fragment is homologous to three tobacco cpDNA segments, 7881-8736, 51455-51502, and 72747-72804 (numbered according to Shinozaki et al. 1986) (Fig. 6). A 142 bp section has 75.4% similarity with the tobacco 7881-8022 segment, and is within ORF98. A 214 bp segment has 79.1% similarity to the tobacco 8310-8521 segment and includes ORF36a (Wolfe and Sharp 1988). A 115 bp segment has 84.3% similarity with the tobacco 8623-8736 segment and contains tRNA-Serine (GCU) (8719-8623) (Fig. 6). Both the 214 and 115 bp segments are within the 633 bp dispersed repeat.

A 94 bp segment of the 1.4 kb XbaI fragment is similar to two tobacco segments, 70.8% with 51455-51502 and 63.8% with 72747-72804. The 51455-51502 segment is a noncoding region between ORF158 and psbG, and the 72747-72804 segment is within an intron region between exon 2 and exon 3 of ORF203/196 in the tobacco genome (Wolfe and Sharp 1988). Both segments contain an 8 bp direct repeat CATCTTTT, however, there is only one repeat CATCTTTTTA shown in this region in Douglas-fir. The 8 bp direct repeat in tobacco is sandwiched between the two segments that are inverted in Douglas-fir (Fig. 6).

Discussion

We mapped the location of dispersed repeats in Douglas-fir cpDNA using high stringency hybridizations therefore relative differences in band intensity on autoradiograms depend largely on repeat length, rather than on degree of sequence similarity. Short repeats are dispersed throughout the genome, and often did not show reciprocal hybridization (i.e., one clone hybridized to another only when blotted or used as a probe). This may result from several causes, such as Southern transfer efficiency, radioactive labeling efficiency, and probe size relative to repeat size. Therefore, the short dispersed repeats are difficult to group into families. However, the long and medium sized repeats were clearly interpretable and showed reciprocal hybridization.

Although several XbaI fragments were uncloned, including ones of 10.6 and 13.8 kb, hybridizations of the other clones to cpDNA blots showed that they did not possess repeats from any of the six families identified. The only repeat families missed would be those confined to the uncloned fragments.

Two of the three clones from repeat family 1 (Fig. 1) were sequenced. However, hybridization with a 345 bp subclone of the 2.7 kb XbaI fragment [that was almost entirely comprised of a part of the 633 bp segment (78.0%)], indicated that the 633 bp repeat is present on

all three cloned fragments (Appendix I, Fig. 8). In the 1.4 kb XbaI fragment, the 633 bp repeat appears to flank one endpoint of a 40-50 kb inversion that distinguishes Douglas-fir from both tobacco and radiata pine (P. radiata D. Don) (Strauss et al. 1988). The dispersed repeat in the 3.8 kb XbaI fragment is located adjacent to, but does not include, the other end of the inversion. The 2.7 kb XbaI fragment lies in the middle of this large inversion, but maps very close to the endpoint of a roughly 23 kb inversion shared by both Douglas-fir and radiata pine (Fig. 3-4, section F of Strauss et al. 1988).

Duplications that occurred during inversions cannot account for the present distribution of dispersed repeats in Douglas-fir. Assuming that conifer cpDNA was derived from a tobacco-like ancestral genome (Palmer and Stein 1986), only the 40-50 kb inversion has an endpoint near to the ancestral position of the tRNA-serine/ORF36 region, and thus could have brought it from its ancestral position to the second of its three locations in Douglas-fir (other end of 40-50 kb inversion). This does not, however, explain its location at the endpoints of the 23 kb inversion (one of the endpoints of the 23 kb inversion is within the 40-50 kb inverted section, and the other coincides with one of the endpoints of the 40-50 kb inverted section). The 23 kb inversion is shared by both Douglas-fir and radiata pine, and thus probably preceded the 40-50 kb inversion, which

is present only in Douglas-fir. Thus, the chronological sequence of inversions does not coincide with the distribution of repeats. Either the 40-50 kb inversion occurred twice, which is unlikely, or the repeats spread through the genome by some other means. Given the large size of the repeats, and their transposon-like ends, it seems likely that transposition, rather than duplication during the inversion, it is unlikely that the inversions occurred at the time of transposition (e.g., via formation of a intramolecular cointegrate: Grindley and Reed 1985). Instead, the repeats probably dispersed first and the mediated the associated inversions via homologous recombination.

The 1.4 kb XbaI fragment contains three different homologous fragments of the tobacco chloroplast genome. These regions appear to have been juxtaposed in Douglas-fir by inversions (Strauss et al. 1988) (Fig. 7). Two short fragments in tobacco separated by about 20 kb each contain an 8 bp direct repeat (CATCTTTT), which may have been brought together in Douglas-fir via three inversions and a short deletion. The full sequence of this short repeat in Douglas-fir is CATCTTTTTATT, which is very similar to the repeats near to the endpoints of a major inversion in Marchantia (CATTTTTTTATT) (Zhou et al. 1988) that have been proposed as part of transposon-associated insertion repeats. It also resembles some of the cpDNA sequences involved with recombination in Oenothera (3'-CATCATTTT, 5'-

ATCTTTT: vom Stein and Hachtel 1988), Aegilops (CATTTTTTT: Ogihara et al. 1988, wheat (CTTTTTTA: Howe 1985), and λ-att (3'-CATATTTTTT: Landy and Ross 1977).

The high degree of homology (98%) between the dispersed repeat segments is surprisingly high. Because the repeats are present in the same positions in pine and Douglas-fir, they are probably over 50 to 130 million years old [first observations of distinct Pseudotsuga (Hermann 1985) and Pinus (Axelrod 1986) in fossil record respectively]. This suggests that a copy correction mechanism may be operating in conifer cpDNA, as was suggested to explain the homology of two dispersed ribosomal protein gene segments in wheat (Bowman et al. 1988).

Transposon-like insertion sequences, two 6 bp inverted repeats flanked by two 7 bp direct repeats, appear to be present in the 633 bp dispersed repeat. The imperfect homology of these repeats is probably due to deletion and base substitution that occurred during of subsequent to an ancient transposition event. This observation, plus our other evidence for transposon movement discussed above, lend credence to recent observations that transposable elements have impacted the structural evolution of chloroplast genomes (Zhou et al. 1988; Milligan et al. 1989).

Acknowledgments. We thank Jeff Palmer for providing both helpful advice and DNA clones; Al Doerksen for assistance in the laboratory; Glenn Howe for many thoughtful discussions on data interpretation; and Jack Kramer for help in analyzing DNA sequences in the Oregon State University Center for Gene Research and Biotechnology. This research was supported by a grants to S.H.S. from the College of Forestry at Oregon State University and the National Institutes of Health (GM36367).

Fig. 1. Douglas-fir cpDNA restriction site map for SstI, PvuII, and XbaI. The numbers on the map are the fragment sizes in kb. The thick lines below the map show the XbaI fragments that were cloned and used to map the dispersed repetitive DNA sequences. Dispersed repeat families (1-6) are aligned and placed on the restriction map. Open boxes indicate repeat-containing XbaI fragments.

SstI 30·9 10·8 9·4 4·1 8·0 4·1 24·1 10·8 0·5 2·6 1·1 3·4 6·7 4·8

PvuII 9·2 20·9 0·6 4·1 2·7 4·3 4·6 3·2 11·3 13·8 2·9 13·0 17·4 2·1 11·3

XbaI 6·9 8·9 6·4? 13·8 1·5a 0·3 4·8 5·9 11·2 2·5 3·8 1·6 5·9 6·0 1·5b 0·6 2·4 1·0 2·7 3·2 a 0·9 3·2 b 0·5 4·3 1·5c 5·1 0·8 ·9 1·4 10·6

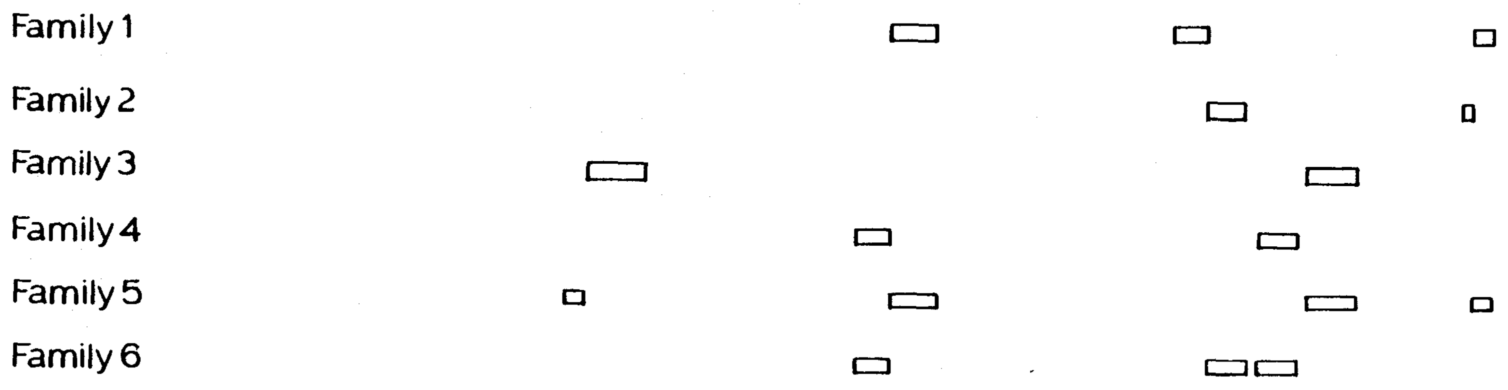


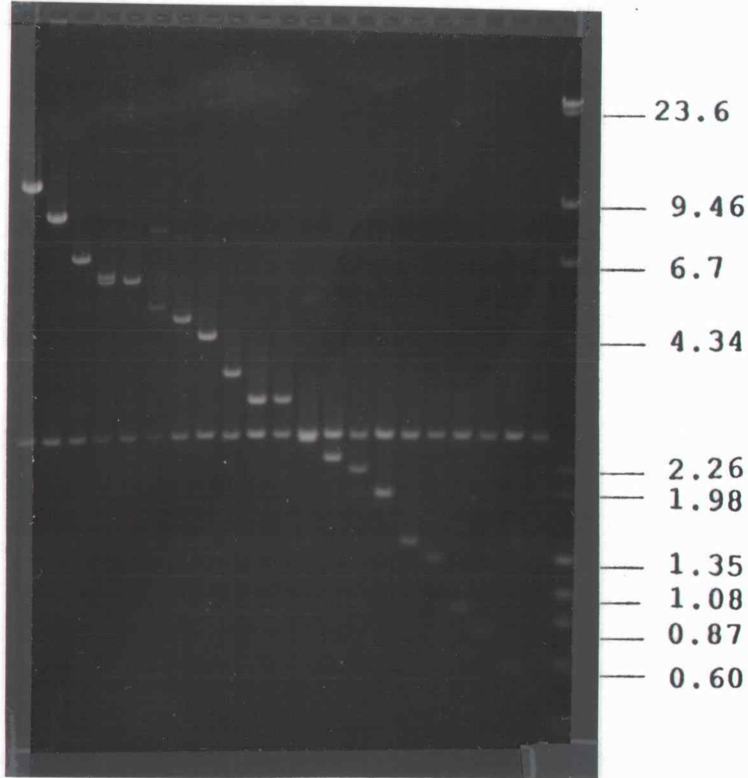
Fig. 2. Mapping of dispersed repeats in Douglas-fir cpDNA.

A: All XbaI clones were digested with XbaI, electrophoresed on a 0.8% agarose gel, and stained with ethidium bromide.

B: An example of autoradiography after hybridization with an insert of a cloned cpDNA fragment which was cut from a low melting point agarose gel and labelled with ^{32}P. The 3.2(a) kb XbaI cloned fragment was used as a probe; it hybridized strongly to itself (lane 10) and to the 0.8 kb XbaI cloned fragment (lane 19), with moderate intensity to the 2.5 (lane 13) and 3.2(b) kb (lane 11) fragments, and weakly to several other fragments. Excision of inserts from gels was not entirely successful in removing vector DNA, which is evident in all lanes. The "chain" of bands in lane 17-21 (1.3 to 2.0 kb) are spurious; they are in positions devoid of cloned fragments.

A.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21



B.

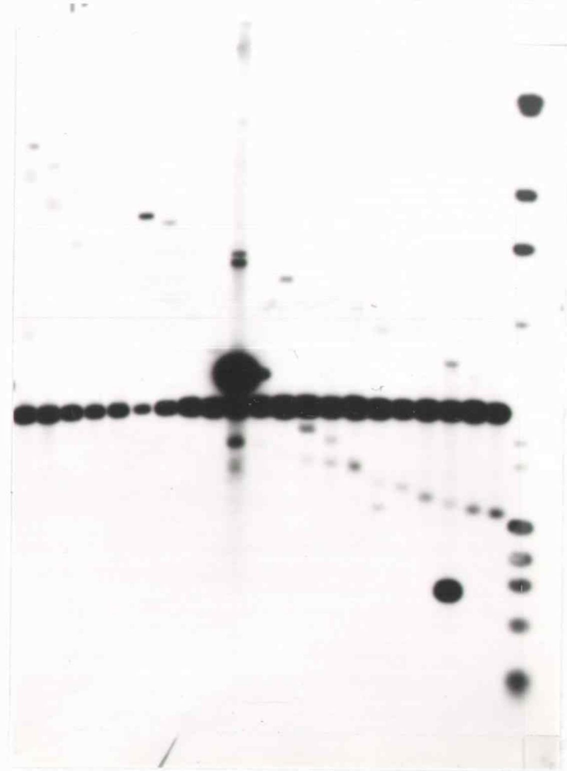


Figure 2.

Fig. 3. Hybridization of cloned XbaI fragments of Douglas-fir cpDNA to mapping blots. Partially purified chloroplast DNA was digested with restriction enzymes (PvuII (P), PvuII and XbaI (PX), XbaI (X), XbaI and SstI (XS), and SstI (S)) and blotted.
A: Hybridization with the 2.7 kb cloned XbaI fragment; lane X shows that it not only hybridized with itself (arrow), but also hybridized with several other fragments, indicative of dispersed repetitive DNA.
B: The same blot when hybridized with the 1.4 kb cloned XbaI fragment, giving a pattern similar to that of the 2.7 kb fragment (both are members of family 1, Fig. 1).

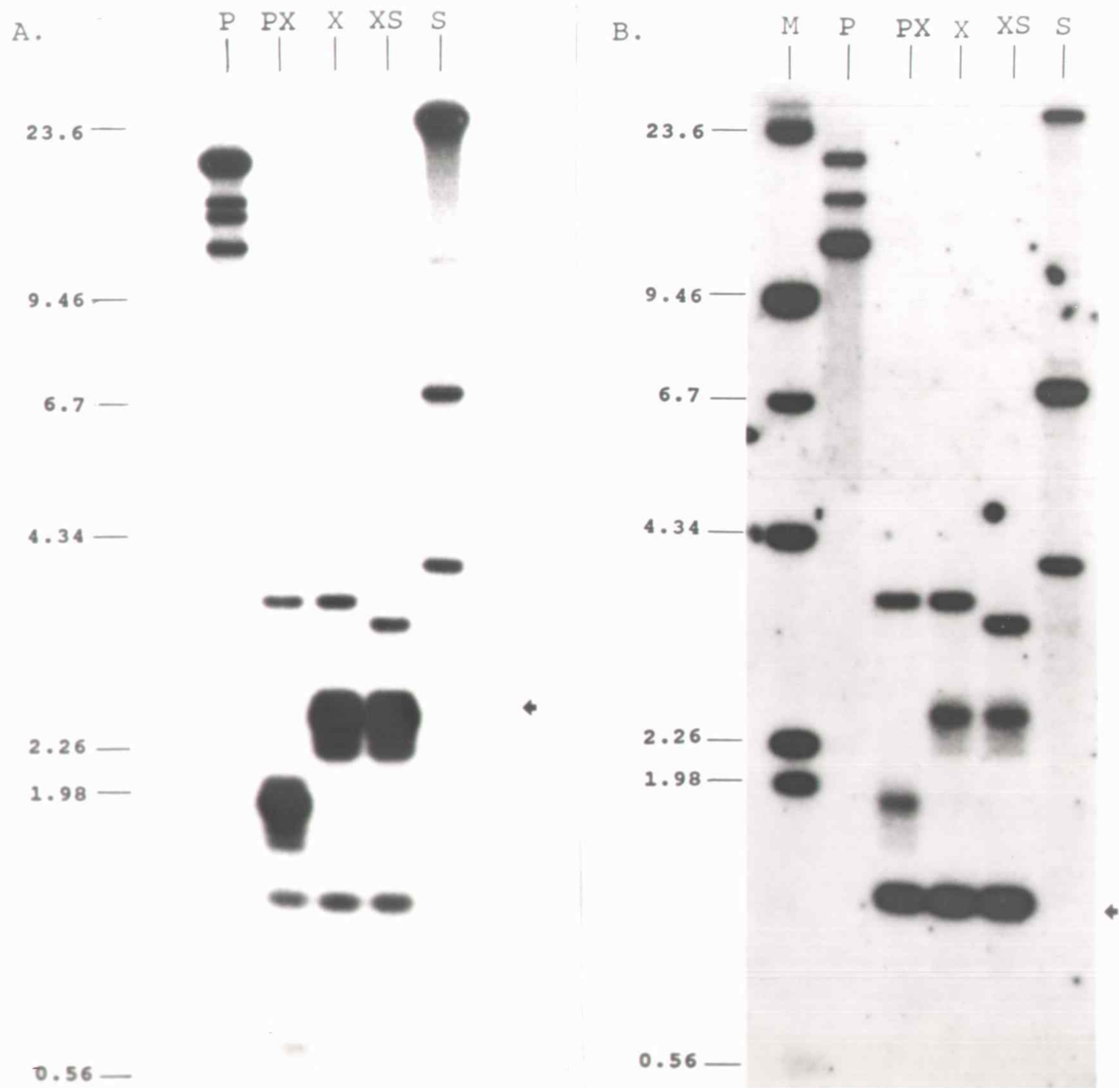
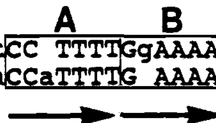
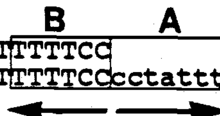


Figure 3.

Fig. 4. Sequence of one of the dispersed repeat families. Capital letters indicate homology between the cloned fragments. The top sequence is from the 2.7 kb XbaI fragment and the bottom sequence is from the 1.4 kb XbaI fragment. The total size of the repeat, including deletions, is 633 bp. The sequence within the box indicates the putative transposon-like endpoints (A = direct repeats, B = complementary inverted repeats). Within the dispersed repeat, a tRNA-serine gene is underlined and an arrow shows the direction of transcription. The open reading frame (ORF 36) is also underlined and the direction of transcription shown. The similarity between these repeats is 97.9% (number of identical nucleotides divided by the total number of nucleotides including gaps).




```
tctttttatcgtatcgtaatctttctttttattggtttCC TTTTGgAAAACAAAGGATAAaTTATCTC
atatctggttgatccaaaaaagaagaagggaagaaagaCCaTTTTG AAAACAAAGGATAAgTTATCTC

CTTCTTTCCAATTTCTTTTCA  CGCACGTGATCTGgAGAAATAATTTCGTGATTTGTATGAATCATACT
CTTCTTTCCAATTTCTTTTCAcaCGCACGTGATCTG AGAAATAATTTCGTGATTTGTATGAATCATACT

ATTGCTTGGTATTCAAGTATCCATATAtGaTACAAAGATTGATGATCTATTCTGTTGTACTTATAATCAG
ATTGCTTGGTATTCAAGTATCCATATAcGgTACAAAGATTGATGATCTATTCTGTTGTACTTATAATCAG

GATCCTGGAGATTACGTAATGCTTACcCTTAAGCTGtTCGTTTACGCAGTAGTGATATTTTTCATTTCTC
GATCCTGGAGATTACGTAATGCTTACgCTTAAGCTGaTCGTTTACGCAGTAGTGATATTTTTCATTTCTC
                    ORF 36 ——→

TTTTTATCTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGtGAAAAAATAG
TTTTTATCTTTGGATTTCTATCGAACGATCCAGGACGTAATCCCGGACGTAAAGAATAGcGAAAAAATAG

GTTAAtTAGTCTTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAATAGTGACCGAAC
GTTAAgTAGTCTTTTACGTTCCGTAGAAAGATTCGGAGTTATTCGTTTTCAGGATCAATAGTGACCGAAC

GGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTACTACGGATTAGCAATCCGCCGCTTTGGTC
GGAGAGAGAGGGATTCGAACCCTCGGTACGGATAATCCGTACTACGGATTAGCAATCCGCCGCTTTGGTC

CGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAACAAAATGAATGGTGGAGTGAAGGTGTATAC
CGCTCAGCCATCTCTCCAAGATGGAAGAGTTCATGTGTAACAAAATGAATGGTGGAGTGAAGGTGTATAC
←—— tRNA-serine

CATAGCATGTATGGggATTGTATCGACAATaTAATGAATAGGTCAATTATTTAGAGAAAAATCAATCTGG
CATAGCATGTATGG  ATTGTATCGACAATgTAATGAATAGGTCAATTATTTAGAGAAAAATCAATCTGG
```




```
CGAATCGTATTGTTCATTCCGTTCAAAATAATTCTTTTTCC       tgaactagaaagcctagaa
CGAATCGTATTGTTCATTCCGTTCAAAATAATTCTTTTTCCcctatttcttctgacctctgccggtg
```

A.

```
TTAGTAGATACAAACGAATGATATCCTATCCATGGTACCCTAGGGAAGTCGAATCCCGTTGCCTCCTTGA

AAGAGAGATGTCCTGGTCCACTAGACGATAGGGCATACCAATCTTCATTATATTCCAGTTCCCGGGAAGT
                           A1 -->

TATC ATAGGGGTTACCAATTTTCATTATATTCAAGTTCCCTGGAAGTTGTCAATAGTATGGCCAGAATT
     A2 -->

ATTCAGAATCTTTCTTTTTAT CGTATCGT AATCTTTCTTTTTATTGGTTTCCTTTTGGAAAACAAA
      B1 -->          C1   C2  B2 -->                DISPERSED REPEAT
```

B.

```
     A1: ATAGGGcaTACCAATcTTCATTATATTCcAGTTCCCgGGAAGTTaTC
(1)      ||||||  ||||||| |||||||||||| ||||||| ||||||| ||
     A2: ATAGGGgtTACCAATtTTCATTATATTCaAGTTCCCtGGAAGTTgTC


     B1: AATCTTTCTTTTTAT
(2)      |||||||||||||||
     B2: AATCTTTCTTTTTAT
```

Fig. 5. The 2.7 kb XbaI fragment possesses several imperfect direct repeats, which are located close to the region of the 633 bp dispersed repeat (Fig. 4).

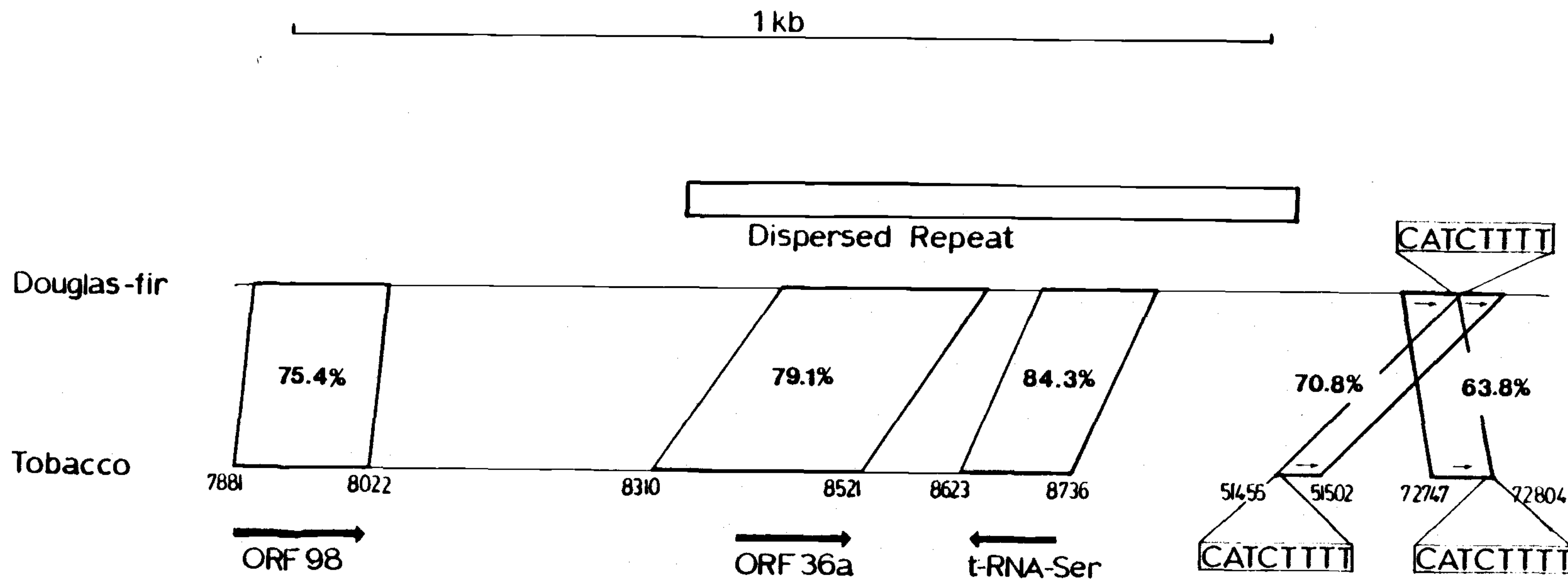


Fig. 6. Alignment of DNA sequences between the 1.4 kb XbaI fragment of Douglas-fir and tobacco cpDNA. Percent nucleotide homology between regions of Douglas-fir and tobacco are shown between fragments. The numbering of the nucleotides in tobacco cpDNA (not drawn to scale) is given at bottom (Shinozaki et al. 1986).

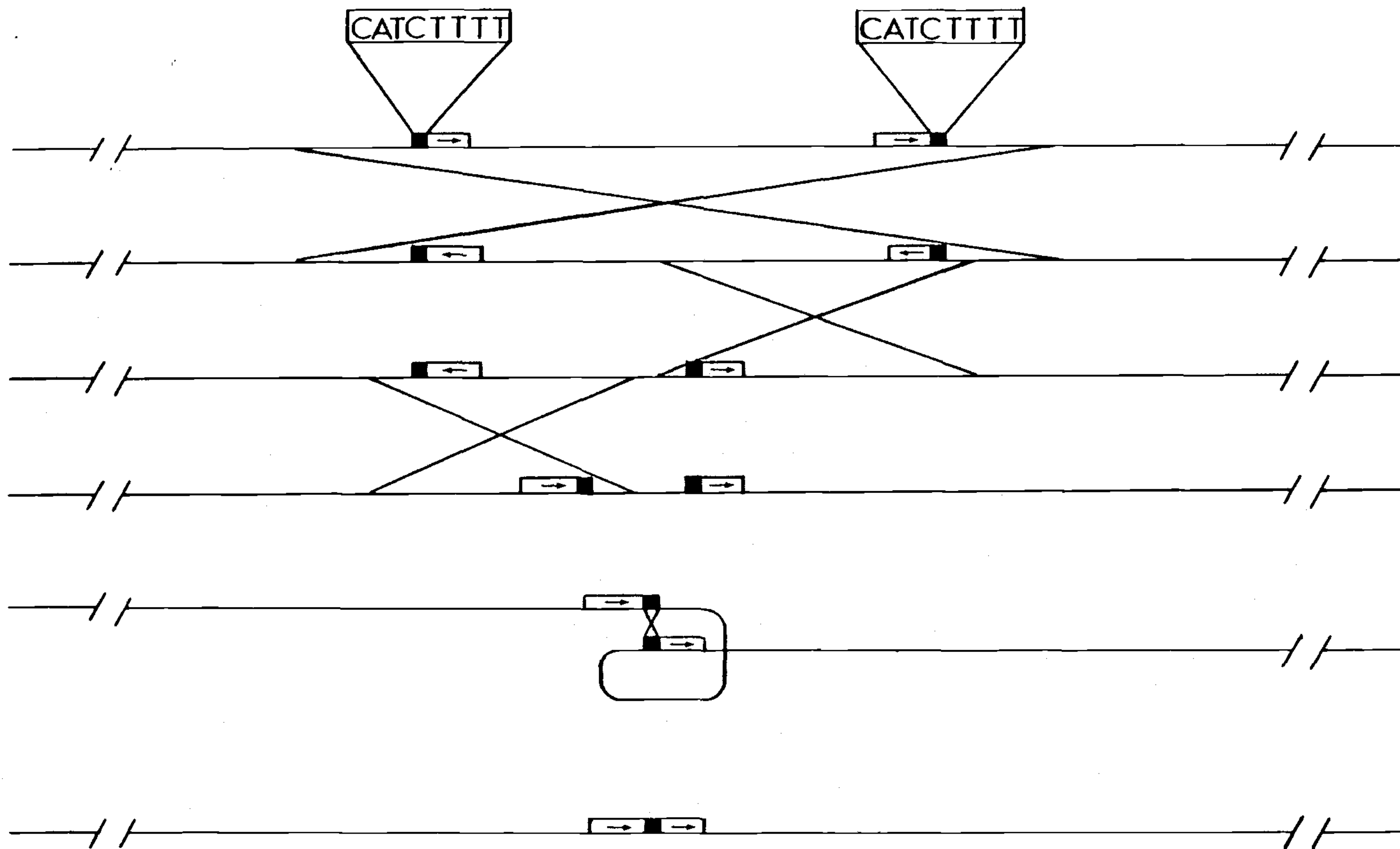


Fig. 7. A hypothetical model for inversions and a deletion during evolution of the Douglas-fir chloroplast genome from a tobacco-like ancestral genome.

References

Axelrod DI (1986) Ann Missouri Bot Gard 73:565-641

Bowman CM, Dyer TA (1986) Curr Genet 10:931-941

Bowman CM, Barker RF, Dyer TA (1988) Curr Genet 14:127-136

Doebley JF, Ma DP, Renfroe WT (1987) Curr Genet 11:617-624

Feinberg AP, Vogelstein B (1984) Anal Biochem 137:266

Grindley NDF, Reed RR (1985) Ann Rev Biochem 54:863-96

Hermann RK (1985) The Genus Pseudotsuga: Ancestral History and Past Distribution (Forest Research Lab., II Oregon State Univ Corvallis OR) Spec Pulb 26

Howe CJ (1985) Curr Genet 10:139-145

Howe CJ, Barker RF, Bowman CM, Dyer TA (1988) Curr Genet 13:343-349

Landy A, Ross W (1977) Science 197:1147-1160

Lidholm J, Szmidt AE, Hallgren J-E, Gustafsson P (1988) Mol Gen Genet 212:6-10

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York

Milligan BG, Hampton JN, Palmer JD (1989) Mol Biol Evol (In press)

Ogihara Y, Terachi T, Sasakuma T (1988) Proc Natl Acad Sci USA 85:8573-8577

Palmer JD, Thompson WF (1982) Cell 29:537-550

Palmer JD (1985) Ann Rev Genet 19:325-354

Palmer JD (1986) Methods Enzymol 118:167-186

Palmer JD, Stein DB (1896) Curr Genet 11:823-834

Palmer JD, Osorio B, Aldrich J, Thompson WF (1987) Curr Genet 11:275-286

Pearson WR, Lipman DJ (1988) Proc Natl Acad Sci USA 85:2444-2448

Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463-5467

Strauss SH, Palmer JD, Howe GT, Doerksen AH (1988) Proc Natl Acad Sci USA 85:3898-3902

Tsai C-H (1989) M.S. Thesis, Oregon State University, Corvallis, Oregon, USA

vom Stein J, Hachtel W (1988) Mol Gen Genet 213:513-518

Wolfe KH, Sharp PM (1988) Gene 66:215-222

Zhang H, Scholl R, Browse J, Somerville C (1988) Nucleic Acids Res 16:1220

Zhou DX, Massenet O, Quigley F, Marion MJ, Moneger F, Huber P, Mache R (1988) Curr Genet 13:433-439

CONCLUSIONS

The unusual structure of Douglas-fir chloroplast DNA makes it a good model for studying cpDNA evolution. Unlike most other land plants, its cpDNA lacks a large inverted repeat, possesses much repetitive DNA, and has undergone several major genome rearrangements. Understanding the causes of the unusual chloroplast genomes of conifers such as Douglas-fir should help to elucidate the factors that constrain cpDNA evolution in plants.

The major conclusions from this study of dispersed repetitive DNA in Douglas-fir cpDNA are:

1. Repeats are non-randomly distributed throughout the genome, and cover a wide variety of sizes; they range from less than 10 bp to over 600 bp in length.
2. Long repeats appear to be associated with endpoints of major inversions.
3. One dispersed repeat is associated with a length mutation hotspot.
4. As in other reports, tRNA gene duplications are associated with inversions.
5. Transposable elements appear to have played a role in cpDNA evolution. They may have initiated the formation of dispersed repeats, thereby potentiating genome rearrangement.
6. Future work could help to test the hypotheses generated above by:

a. Examining the structure of other dispersed repeats.
b. Determining the role of repeats in cpDNA length mutation hotspots.
c. Sequencing other inversion endpoints to help elucidate the causes of cpDNA inversions.
e. Searching for additional evidence concerning the sources and roles of transposable elements in chloroplast genomes.

APPENDICES

BIBLIOGRAPHY

Axelord, D. I. 1986. Cenozoic history of some Western American Pines. Ann. Missouri Bot. Gard. 73:565-641.

Abremski, K. and Hoess, R. 1984. Bacteriophage P1 site-specific recombination: purification and properties of the Cre recombinase protein. J. Biol. Chem. 259:1509-1514.

Bonnard, G., Weil, J. -H. and Steinmetz, A. 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 Dyer, T. A. 1986. The location and possible evolutionary significance of small dispersed repeats in wheat ctDNA. Curr. Genet. 10:931-941.

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

Cech, T. R. 1987. The chemistry of self-splicing RNA and RNA enzymes. Science 236:1532-1539.

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

Doebley, J. F., Ma, D. P. and Renfroe, W. T. 1987. Insertion/deletions mutations in the Zea chloroplast

genome. Curr. Genet. 11:617-624.

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

Grindley, D. F. G., Reed, R. R. 1985. Transpositional recombination in prokaryotes. Ann. Rev. Biochem. 54:863-96.

Gross, S. R., Hsieh. T. -s. and Levine, P. H. 1984. Intramolecular recombination as a source of mitochondrial chromosome heteromorphism in Neurospora. Cell 38:233-239.

Hermann, R. K. 1985. The Genus Pseudotsuga: Ancestral History and Past Distribution (Forest Research Lab., II Oregon State Univ., Corvallis, OR), Spec. Publ. 26.

Hira, A., Kanno, A., Iwahashi, M., Nishizawa, Y., Hiratsuka, J. and Sugiura, M. 1988. The organization of chloroplast and mitochondrial genomes in rice. Genome 30 (Suppl. 1):320.

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

Howe, G., Strauss, S. H., Palmer, J. P. and Doerksen, A. 1988. Gene and restriction site maps of Douglas-fir

(*Pseudotsuga menziesii* [Mirb.] Franco) and radiata pine (*Pinus radiata* D. Don) chloroplast genomes. p54-66. In: Molecular genetics of forest trees. W. M. Cheliak and A. C. Yapa, eds. Petawawa National Forestry Institute Information Report PI-X-80, Chalk River, Ontario, Canada.

Iida, S. and Heistand-Nauer, R. (1986). Localized conversion at the crossover sequences in the site-specific DNA inversion system of bacteriophage P1. Cell. 45:71-79.

Johnson, R. C. and Simon, M. I. (1985). Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. Cell. 41:781-791.

Johnson, R. C., Bruist, M. F. and Simon, M. I. (1986). Host protein requirements for *in vitro* site-specific DNA inversion. Cell. 46:531-539.

Kahmann, R., Rudt, F., Koch, C. and Mertens, G. (1985). G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. Cell. 41:771-780.

Krasnow, M. A. and Cozzarelli, N. R. (1983). Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented *res* sites. Cell. 32:1313-1342.

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

197:1147-1160.

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

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

Milligan, B. G., Hampton, J. N. and Palmer, J. D. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol. Biol. Evol. (in press).

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

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

Palmer, D. J. and Thompson, W. F. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted

repeat sequence is lost. Cell 29:537-550.

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

Palmer, D. J. 1985b. Evolution of chloroplast and mitochondrial DNA in plants and algae. Pages 131-240 in R. J. MacIntyre, ed. Monographs in evolutionary biology: molecular evolutionary genetics. Plenum, New York.

Palmer, J. D. 1986. Isolation and structural analysis of chloroplast DNA. Methods Enzymol. 118:167-186.

Palmer, J. D. and Stein, D. B. 1986. Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 11:823-834.

Palmer, J. D., Nugent, J. M. and Herbon, L. A. 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. 1987b. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am. Nat. 130:S6-S29.

Pearson, W. J. and Lipman, D. J. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA. 85:2444-2448.

Quigley, F. and Weil, J. H. 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.

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

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

Simon, M. and Herskowitz, I. eds. (1985). Genome rearrangement. (Alan R. Liss, Inc., New York).

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

Turker, M. S., Domenico, J. M. and Cummings, D. J. 1987. Excision-amplification of mitochondrial DNA during senescence in Podospora anserina - A potential role for an 11 base-pair consensus sequence in the excision process. J. Mol. Biol. 198:171-185.

Umesono, K. and Ozeki, H. 1987. Chloroplast gene

organization in plants. Trends Genet. 3: 281-287.

vom Stein, J. and Hachtel, W. 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.

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

Whatley, J. M., John, P. and Whatley, F. R. 1979. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. Proc. R. Soc. Lond. B. 204:165-187.

Whitfeld, P. R. and Bottomley, W. 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 Sharp, P. M. 1988. Identification of functional open reading frames in chloroplast genomes. Gene. 66: 215-222.

Yanisch-Perron, C., Vieira, J. and Messing, J. 1985. Improved M13 phage cloning vectors and host strains:

nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Zhang, H., Scholl, R., Browse, J. and Somerville, C. 1988. Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res. 16:1220.

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

Appendix I. Supplementary data

A. XbaI/PvuII/SstI Douglas-fir restriction fragment sizes

Table 1. Sizes of Douglas-fir cpDNA restriction fragments generated by XbaI, PvuII and SstI.

Fragment Name	(kb) XbaI	Fragment Name	(kb) PvuII	Fragment Name	(kb) SstI
13*	13.8 (kb)	14	21 (kb)	12	30.9 (kb)
12	11.2	13	17.5	11	24.2
11*	10.6	12	14.0	10	10.8
10	8.9	11	13.2	10	10.8
9	6.9	10a	11.6	9	9.4
8a	6.0	10b	11.6	8	8.0
8b-1	5.9	9	9.2	7	6.7
8b-2*	5.9	8	4.6	6	4.8
7	5.1	7	4.3	5a	4.1
6	4.8	6	4.1	5b	4.1
5	4.3	5	3.2	4	3.4
4	3.8	4	2.9	3	2.6
3a	3.2	3	2.7	2	1.1
3b	3.2	2	2.1	1	0.5
2	2.7	1	0.6		
B	2.5				
A	2.4				
C	1.9				
1a	1.5				
1b*	1.5				
1.4	1.4				
0.9	0.9				
0.8	0.8				
0.6	0.6				
0.5	0.5				
	6.4+				
	1.6+				
	1.5+				
	1.0+				
	0.3+				
total	121.1 kb		122.1 kb		119.8 kb

* Uncloned XbaI fragment
+ Postulated fragments based on the single/double digest restriction enzyme map

Table 2. Sizes of Douglas-fir cpDNA restriction fragments generated by PvuII/XbaI and SstI/XbaI double digests. Fragment sizes are based on gels run with molecular weight standards, then adjusted based on calculations from multiple-enzyme restriction maps.

PvuII/XbaI (kb)	SstI/XbaI (kb)
8.9	8.9
7.5	7.6
6.4*	6.9
6.0	6.4*
5.8	6.2
4.8	6.0
4.4	5.9
4.3	5.9
4.1	5.6
3.8	4.8
3.8	4.8
3.7	4.3
3.7	4.1
3.4	3.8
3.2	3.2
3.2	2.8
3.0	2.7
2.9	2.6
2.7	2.4
2.5	2.4
2.5	2.3
2.4	2.3
2.2	2.0
2.1	1.5
2.1	1.5
2.0	1.5
1.9	1.4
1.6	1.4*
1.6*	1.2*
1.5*	1.1
1.5	1.0*
1.4	0.9
1.4	0.8
1.1	0.6
1.0*	0.6
0.9	0.5

(continued)

Table 2. (continued)

PvuII/XbaI (kb)	SstI/XbaI (kb)
0.9	0.5
0.8	0.5
0.8*	0.3
0.7	0.3*
0.6*	0.3
0.6*	0.2*
0.6	0.1*
0.5	
0.3*	
121.1	120.4

* These fragments were not directly observed on gels or autoradiograms, but are predicted to exist based on gaps found in restriction map construction.

B. Hybridization data from Douglas-fir cpDNA restriction site mapping

A. B.

M P PX X

23.6
9.46
6.7
4.34
2.26
1.98
1.35
1.08
0.87
0.60

Fig 8. Restriction fragments of Douglas-fir cpDNA cut with P (PvuII), P/X (PvuII/XbaI), and X (XbaI). Electrophoresis in 0.8% agarose gel (panel A), and a hybridization with a subclone from 2.7 kb XbaI fragment containing part of 633 bp dispersed repeat (panel B).

Fig. 9. Autoradiograms from Douglas-fir restriction site mapping using cloned Douglas-fir cpDNA probes. Partially purified cpDNA was digested with PvuII (P), PvuII/XbaI (XP), XbaI (X), SstI/XbaI (XS), and SstI restriction enzymes, electrophoresed in 0.8% agarose gel and Southern transferred to Zetabind nylon membranes. The probes were 1.5(a) kb XbaI clone (panel A) and 4.8 kb XbaI clone (panel B). The numbers indicate the size in kb.

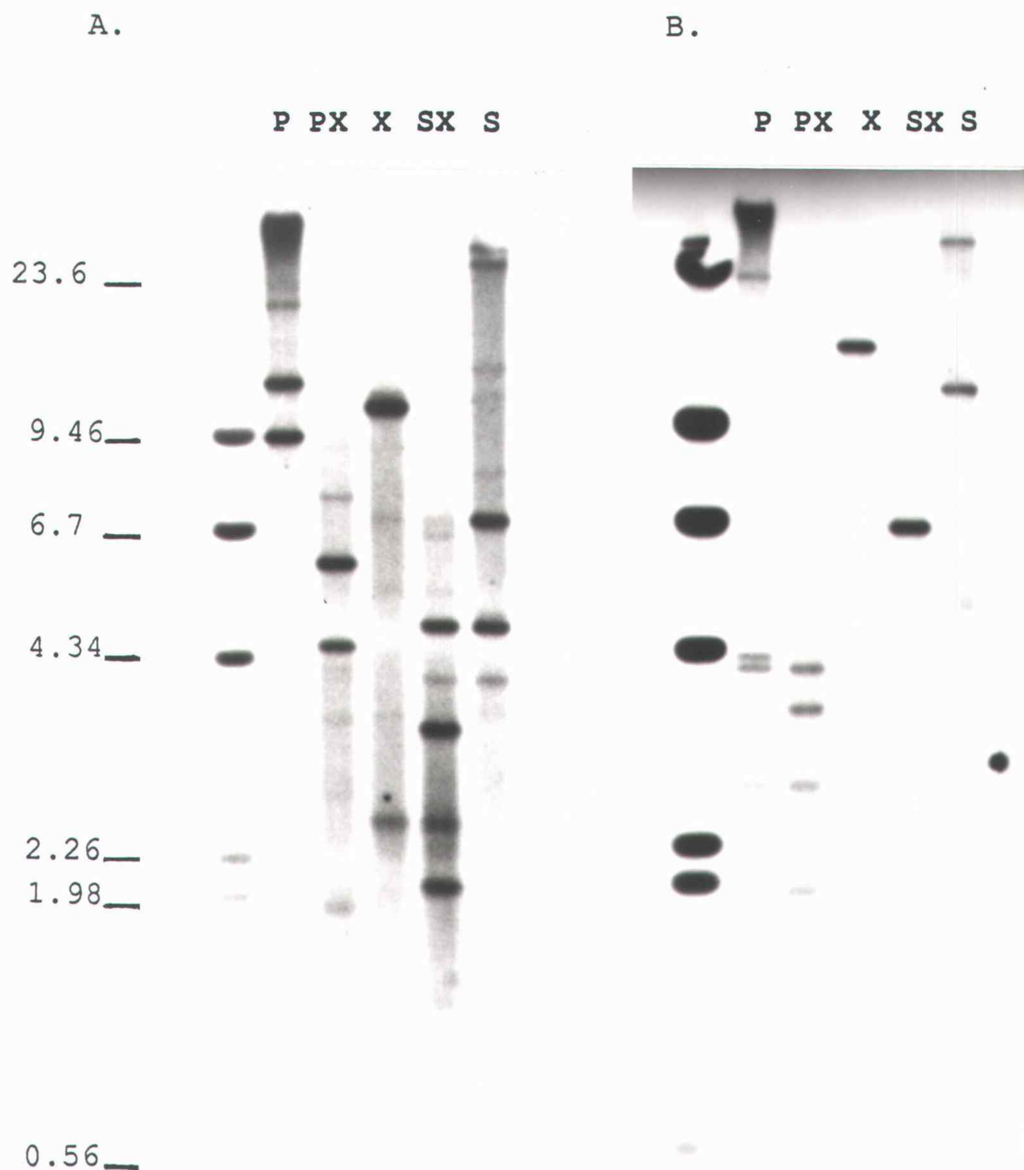


Fig. 10. Autoradiograms from Douglas-fir restriction site mapping using labelled cpDNA fragments cut from gel. The Probes were two uncloned fragments, 10.6 kb XbaI fragment (panel A) and 13.8 kb XbaI fragment (panel B) which were labelled directly within the low melting point agarose gel.

Table 3. Summary of hybridizations of cpDNA clones for restriction site mapping.

	Fragments Hybridized to (kb)				
Clone*	PvuII	PvuII/XbaI	XbaI	XbaI/SstI	SstI
Xba-13.8	20.9 4.3 4.1 2.7 0.6	4.1 3.4 3.0 2.7 0.6	13.8	7.6 6.2	30.9 10.8
Xba-11.2	13.8 11.3	7.5 3.7	11.2	5.6 4.1 1.5	9.4 8.0 4.1
Xba-10.6	11.3 9.2	5.8 4.8	10.6	4.8 4.8 1.0	30.9 6.7 4.8
xba-8.9	20.9	8.9	8.9	8.9	30.9
Xba-6.9	20.9 9.2	4.4 2.5	6.9	6.9	30.9
Xba-6.0	13.0	6.0	6.0	6.0	24.1
Xba-5.9 (8b-1)	13.8 13.0 2.9	2.9 2.2 0.8	5.9	5.9	24.1
Xba-5.9 (8b-2)	11.3 3.2	3.8 2.1	5.9	5.9	9.4
Xba-5.1	17.4 11.3 2.1	2.1 1.6 1.4	5.1	2.3 2.3 0.5	3.4 2.6 0.5
Xba-4.8	4.6 3.2	3.7 1.1	4.8	2.8 2.0	10.8 9.4
Xba-4.3	17.4	4.3	4.3	4.3	10.8
Xba-3.8	13.8	3.8	3.8	3.8	4.1

(Continued)

Table 3. (continued)

Fragments Hybridized to (kb)

Clone*	PvuII	PvuII/XbaI	XbaI	XbaI/SstI	SstI
Xba-3.2a	17.4	3.2	3.2	2.6 0.6	24.1 10.8
Xba-32.b	17.4	3.2	3.2	3.2	10.8
Xba-2.7	17.4 13.0	2.0 0.7	2.7	2.7	24.1
Xba-2.5	13.8	2.5	2.5	2.4 0.1	8.0 4.1
Xba-2.4	13.0	2.4	2.4	2.4	24.1
Xba-1.9	11.3	1.9	1.9	1.1 0.8	3.4 1.1
Xba-1.5b	13.0	1.5	1.5	1.5	24.1
Xba-1.5a	4.6 4.3	0.9 0.6	1.5	1.5	10.8
Xba-1.4	11.3	1.4	1.4	1.4	6.7
Xba-0.9	17.4	0.9	0.9	0.9	24.1
Xba-0.8	11.3	0.8	0.8	0.5 0.3	6.7 1.1
Xba-0.6	13.0	0.6	0.6	0.6	24.1

* Clones are designated by sizes of cloned fragments: Xba-0.5 indicates a 0.5 kb XbaI fragment is the insert.

C. Hybridization data from dispersed repeat mapping

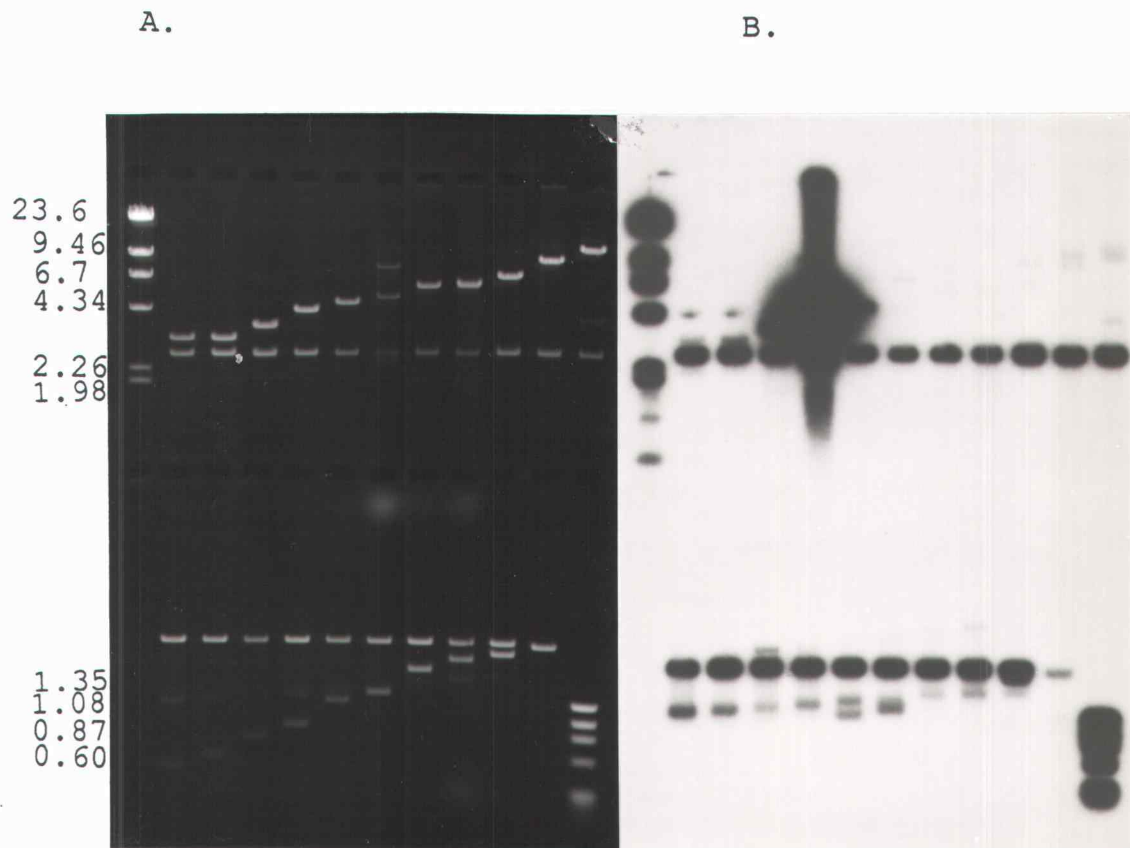


Fig. 11. Electrophoresis and a hybridization with the Douglas-fir XbaI-cpDNA clone bank.

A: Gel electrophoresis of all the clones stained with ethidium bromide.

B: Autoradiography after hybridization with the 3.2(a) kb XbaI fragment which was cut from a gel and labelled directly.

Table 4. Summary of hybridizations of cloned cpDNA fragments (inserts) for mapping of dispersed repeats.

Cloned fragments	Strong signal	Medium signal	Weak signal
Xba-0.5			Xba-0.6, 1.4 1.5a, 2.7 3.8, 6.9 8.9, 11.2
Xba-0.6			Xba-0.5
Xba-0.8	Xba-3.2a		Xba-6.9
Xba-0.9			Xba-2.5
Xba-1.4	Xba-2.7 3.8		Xba-0.8, 0.9 1.5a, 1.9 3.2a
Xba-1.5a		Xba-1.4	Xba-2.7, 3.8
Xba-1.9			Xba-1.5a, 2.4 6.9, 8.9
Xba-2.4			Xba-1.9, 4.3 4.8, 5.1 5.9-1, 6.9 8.9
Xba-2.5		Xba-3.2b	Xba-1.5a, 3.2a 3.8, 4.3 4.8, 5.1 8.9, 11.2
Xba-2.7	Xba-1.4 3.8		
Xba-3.2a	Xba-0.8	Xba-3.2b 2.5	Xba-1.5a, 1.9 2.4, 3.8 4.3, 4.8 5.1, 6.0 5.9-1, 6.9 8.9, 11.2
Xba-3.2b	Xba-2.5	Xba-3.2a	Xba-1.5a, 2.4

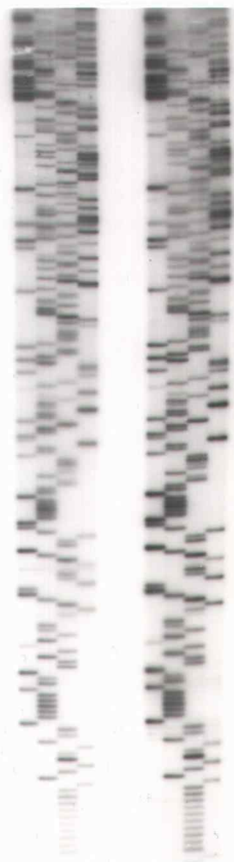
(continued)

Table 4. (continued)

Cloned fragment	Strong signal	Medium signal	Weak signal
Xba-3.8	Xba-1.4 2.7		Xba-1.5a, 2.4 2.5, 3.2a 4.3, 4.8 8.9, 11.2
Xba-4.3	Xba-1.5a 4.8	Xba-1.4 3.8	Xba-1.9, 2.4 2.5, 2.7 3.2a, 3.2b 5.1, 5.9-1 6.0, 6.9 8.9, 11.2
Xba-4.8	Xba-4.3	Xba-3.2a 3.8	Xba-5.9-1, 6.9 8.9, 11.2
Xba-5.1			Xba-1.4, 2.7 3.8
Xba-5.9-1			Xba-1.4, 1.9 2.4, 2.7 3.8, 4.8 6.9, 8.9 11.2
Xba-6.0		Xba-4.3	Xba-1.5, 2.5 2.7, 3.2a 3.2b, 3.8 4.3, 4.8 5.1, 11.2
Xba-6.9			
Xba-8.9			Xba-6.9
Xba-11.2			

D. Additional Douglas-fir cpDNA sequences

A.



B.

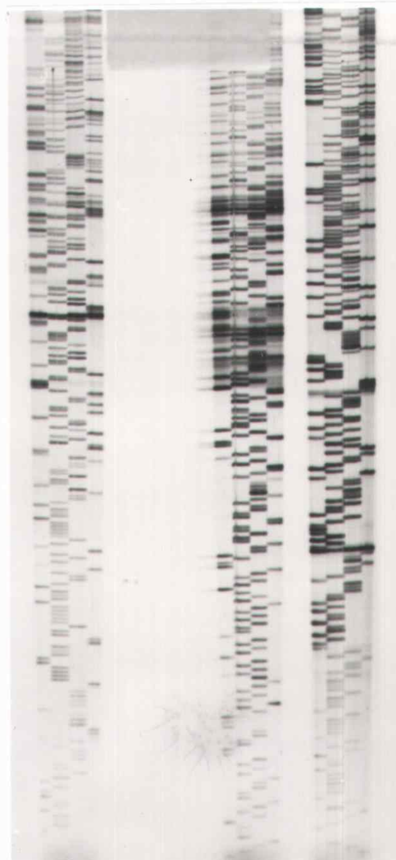


Fig. 12. Autoradiogram from dideoxy chain termination DNA sequencing using ^{32}P and ^{35}S-labelled nucleotides.

A: Douglas-fir cpDNA was sequenced by labelled ^{32}P-alpha dCTP.

B: Douglas-fir cpDNA was sequenced by labelled ^{35}S-alpha dATP.

```
        10         20         30         40         50         60
TCTAGATGAT GCCTTTATTC ATTCAAATAA TCCCTTTTTT GGAAAATTAC CTGAGGCTTA
        70         80         90        100        110        120
TGCGATTTCC GATCCAATTG TCGATGTAAT GCCAATTATT CCCGTTCTCT CTTTTCTTTT
       130        140        150        160        170        180
AGCCTTTGTT TGGCAAGCTG CTGTAAGTTT TCGATAAAAA GTATCCCCTT TTTTCCTTTT
       190        200        210        220        230        240
TCAAGTTTTT GTGTCGCTGT CATTTATCTA ATTTTTGTAT CACTCTTTCC ATTTTTTGTC
       250        260        270        280        290        300
GCAGAAGTTT TATCCTTGCT CTACCCGACA ATACCAGATC GAGATACCTC ATCTGCTCTC
       310        320        330        340        350        360
GACTAAAAAG CTTTTTAACT CACCTTCGTC AATTCCTTCC GATCTCATCG CTCACTTTGG
       370        380        390        400        410        420
ATCGGGCTAT TTGGTCACGT ATTTATACGA ATGACATATT TTCATAAATA TTTGATAAAT
       430        440        450        460        470        480
ATCTGGTTGA TCCAAAAAAG AAGAAGGGAA GAAAGACCAT TTTGAAAACA AAGGGATAAG
       490        500        510        520        530        540
TTATCTCCTT CTTTCCAATT TCTTTTCACA CGCACGTGAT CTGAGAAATA ATTTCGTGAT
       550        560        570        580        590        600
TTGTATGAAT CATACTATTG CTTGGTATTC AAGTATCCAT ATACGGTACA AAGATTGATG
       610        620        630        640        650        660
ATCTATTCTG TTGTACTTAT AATCAGGATC CTGGAGATTA CGTAATGCTT ACGCTTAAGC
       670        680        690        700        710        720
TGTTCGTTTA CGCAGTAGTG ATATTTTTCA TTTCTCTTTT TATCTTTGGA TTTCTATCGA
       730        740        750        760        770        780
ACGATCCAGG ACGTAATCCC GGACGTAAAG AATAGCGAAA AAATAGGTTA AGTAGTCTTT
       790        800        810        820        830        840
TACGTTCCGT AGAAAGATTC GGAGTTATTC GTTTTCAGGA TCAATAGTGA CCGAACGGAG
       850        860        870        880        890        900
AGAGAGGGAT TCGAACCCTC GGTACGGATA ATCCGTACTA CGGATTAGCA ATCCGCCGCT
       910        920        930        940        950        960
TTGGTCCGCT CAGCCATCTC TCCAAGATGG AAGAGTTCAT GTGTAACAAA ATGAATGGTG
       970        980        990       1000       1010       1020
GAGTGAAGGT GTATACCATA GCATGTATGG ATTGTATCGA CAATGTAATG AATAGGTCAA
      1030       1040       1050       1060       1070       1080
TTATTTAGAG AAAAATCAAT CTGGCGAATC GTATTGTTCA TTCCGTTCAA AATAATTCTT
      1090       1100       1110       1120       1130       1140
TTTCCCCTAT TTCTTCTGAC CTCTGCCGGT GGCCAGGCCA GGCCAAGAAA AACAAAAAAG
      1150       1160       1170       1180       1190       1200
AATTCATGCA TCAGACAATG CGTTAGCTAA TCGGTAAGCG AAAAAAGTGG TTGTAACGGT
      1210       1220       1230       1240       1250       1260
AAGAAAAAAC AGACCGAAAA AAAAATAGAA CAGATTGAAC ATCTAGTGTC ATCTTTTTAT
      1270       1280       1290       1300       1310       1320
TCTCTCCCTA ATAATTTTCA ATAAGTTAGT TACATGGAAT GGATTAGTCC ATTTATTTCT
      1330       1340
CTCCAGTATA AAATTTCAAT ATCTAGA
```

Fig. 13. The DNA sequence of the 1.4 kb XbaI cpDNA clone.

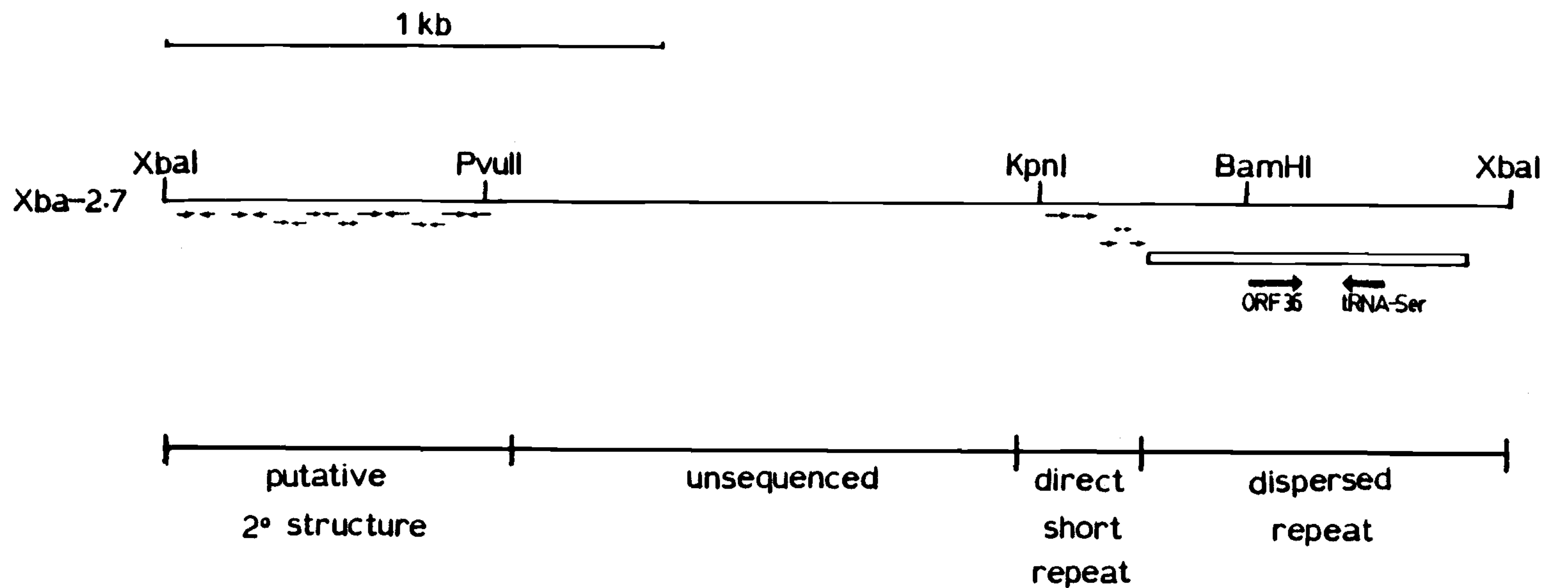


Fig. 14. The restriction map of the 2.7 kb XbaI cpDNA clone.

```
        10         20         30         40         50
TTAGTAGATA CAAACGAATG ATATCCTATC CATGGTACCC TAGGGAAGTC
        60         70         80         90        100
GAATCCCGTT GCCTCCTTGA AAGAGAGATG TCCTGGTCCA CTAGACGATA
       110        120        130        140        150
GGGCATACCA ATCTTCATTA TATTCCAGTT CCCGGGAAGT TATCATAGGG
       160        170        180        190        200
GTTACCAATT TTCATTATAT TCAAGTTCCC TGGAAGTTGT CAATAGTATG
       210        220        230        240        250
GCCAGAATTA TTCAGAATCT TTCTTTTTAT CGTATCGTAA TCTTTCTTTT
       260        270        280        290        300
TATTGGTTTC CTTTTGGAAA ACAAAGGGAT AAATTATCTC CTTCTTTCCA
       310        320        330        340        350
ATTTCTTTTC ACGCACGTGA TCTGGAGAAA TAATTTCGTG ATTTGTATGA
       360        370        380        390        400
ATCATACTAT TGCTTGGTAT TCAAGTATCC ATATATGATA CAAAGATTGA
       410        420        430        440        450
TGATCTATTC TGTTGTACTT ATAATCAGGA TCCTGGAGAT TACGTAATGC
       460        470        480        490        500
TTACCCTTAA GCTGTTCGTT TACGCAGTAG TGATATTTTT CATTTCTCTT
       510        520        530        540        550
TTTATCTTTG GATTTCTATC GAACGATCCA GGACGTAATC CCGGACGTAA
       560        570        580        590        600
AGAATAGTGA AAAAATAGGT TAATTAGTCT TTTACGTTCC GTAGAAAGAT
       610        620        630        640        650
TCGGAGTTAT TCGTTTTCAG GATCAATAGT GACCGAACGG AGAGAGAGGG
       660        670        680        690        700
ATTCGAACCC TCGGTACGGA TAATCCGTAC TACGGATTAG CAATCCGCCG
       710        720        730        740        750
CTTTGGTCCG CTCAGCCATC TCTCCAAGAT GGAAGAGTTC ATGTGTAACA
       760        770        780        790        800
AAATGAATGG TGGAGTGAAG GTGTATACCA TAGCATGTAT GGGGATTGTA
       810        820        830        840        850
TCGACAATAT AATGAATAGG TCAATTATTT AGAGAAAAAT CAATCTGGCG
       860        870        880        890        900
AATCGTATTG TTCATTCCGT TCAAAATAAT TCTTTTTCCT GAACTAGAAA
       910        920        930        940        950
GCCTAGAATT ATCATAACTT CTTTTTCAAT GAAGATGAAA AAAAAAAGAA
       960
AAAAATGAAT CTAGA
```

Fig. 15. Sequence of the Xba2kpn fragment (a subclone of the 2.7 kb XbaI cloned fragment).

```
        10         20         30         40         50
TCTAGAAAGG CACTGGCTAT CGATCATGAA AAAGAAGTAT GAAAATGAAA

        60         70         80         90        100
TAAATATTCA TGGAGAAGAA GAGAATTTGA TTCGGCGGAG AGGGATGAAT

       110        120        130        140        150
GTTTATTGCA GATTCACTAT GATGATTAGA TTTTATCCCC GAAAGAAGGG

       160        170        180        190        200
TCTTTTTTTC AAACTGAATT ATCGATCTAG TTAATGTATC TAATGGATAG

       210        220        230        240        250
ATATACTAAA TATCTATATG AATAGAGATA CTAAATATCT AGTATCTTTA

       260        270        280        290        300
TTCAACCCTA TTTCTTTTCA CTCTTCTACG GGATTCAGAG CTGAATGGAT

       310        320        330        340        350
TAACTTATTG GATCGGGACT GACGGGGCTC GAACCGCAAC TTCGCTTGAC

       360        370        380        390        400
AGGGCGTACT CTAACCAATT GAACCTACAT CCAATACAGT ACAGTCACTT

       410        420        430        440        450
ACTATGATCA TATTATCTAT GGTAGGTCCT AGATAGATCG AATGATACTA

       460        470        480        490        500
GCCTAAGACG ATTAAATCTA CTCCTGGATC AAAGTATCAA TTCATATGGA

       510        520        530        540        550
ATTGGGTACA TATCTATATG ATATGAATAT ATCATAGATA TCGGAGGAGT

       560        570        580        590        600
TCAATAACCA ATTATCAAAA CATCCATGAT TGGCATGAAT ATAACCATAC

       610        620        630        640        650
CGATAGATTT ATTTTGATTA TTTTGGTTGG GTCCAGCTGG ATTTGAACCA

       660
GCGTAGGCAT ATC
```

Fig. 16. Sequence of the Xba2pvu fragment (a subclone of the 2.7 kb XbaI cloned fragment).

1.

2.

3.

Fig. 17. Possible secondary structures of the Xba2pvu fragment.

A.

```
Douglas-fir  GGAGAGATGGCTGAGCGGACCAAAGCGGCGGATTGCTAATCCGT
             ||||||||||||||| |||| |||||||||||||||||||||||
Tobacco      GGAGAGATGGCTGAGTGGACTAAAGCGGCGGATTGCTAATCCGT
             ||||||||||||||||||||||||||||||||||||||||||||
Wheat        GGAGAGATGGCTGAGTGGACTAAAGCGGCGGATTGCTAATCCGT

Douglas-fir  AGTACGGATTATCCGTACCGAGGGTTCGAATCCCTCTCTCTCCG
              |||||  |||  |||||||||||||||||||||||||| ||||
Tobacco      TGTACGAGTTAATCGTACCGAGGGTTCGAATCCCTCTCTTTCCG
             ||||| | ||  | ||||||||||||||||||||||||||||||
Wheat        TGTACAATTTTTTTGTACCGAGGGTTCGAATCCCTCTCTTTCCG
```

B.

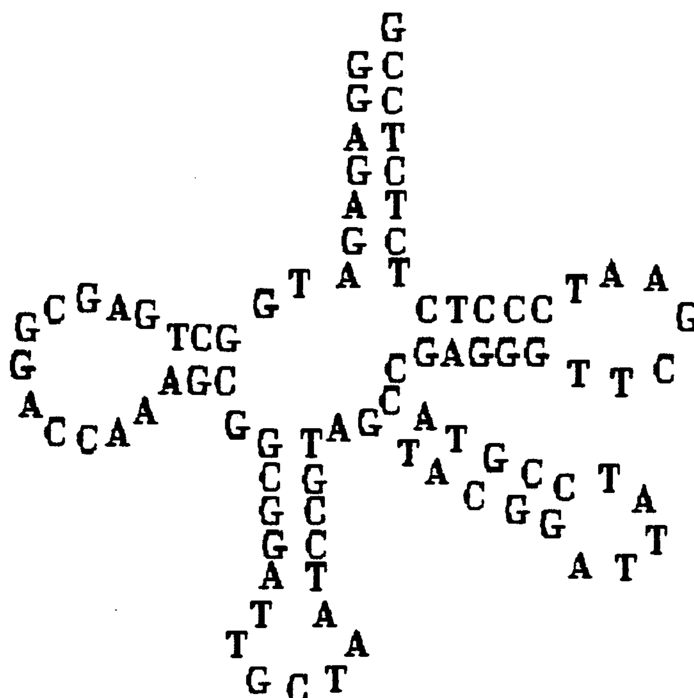


Fig. 18. Aligned sequences for tRNA-Serine (GCU) genes for Douglas-fir, tobacco, and wheat.

A. The nucleotide sequences were aligned.

B. Putative secondary structure of Douglas-fir cpDNA tRNA-serine (GCU) gene.

```
Douglas-fir   ATGCTTACcCTTAAGCTGTTCGTTTACgCAGTAGTGATATTTTTcaTTTCTCTTTT
              |||||||| ||||||||| |||||||| ||||||||||||||||  ||||||||||
Douglas-fir   ATGCTTACgCTTAAGCTGaTCGTTTACgCAGTAGTGATATTTTTcaTTTCTCTTTT
              |||||||| || || ||  |||||||| ||||||||||||||||  ||||||| ||
Tobacco       ATGCTTACtCTcAAaCTcTTCGTTTACaCAGTAGTGATATTTTTtgTTTCTCTcTT
              |||||||| || || || || |||||  |||| || || |||||  ||||||| ||
Liverwort     ATGCTTACtCTTAAGCTGTTtGTTTAtaCAGTgGTtATtTTTTTtgTTTCTCTTTT


Douglas-fir   TATCTTTGGATTTCTATCgAAcGATCCAGGACGTAATCCcGGACGTAAAGAATAg
              |||||||||||||||||| || ||||||||||||||||| ||||||||||||||
Douglas-fir   TATCTTTGGATTTCTATCgAAcGATCCAGGACGTAATCCcGGACGTAAAGAATAg
               ||||||||||| ||||| || ||||| ||||||||||| |||||| |||||||
Tobacco       cATCTTTGGATTcCTATCtAAtGATCCcGGACGTAATCCtGGACGTgAAGAATAa
                | ||||||||  |||| || ||||| ||||||||||| |||||| |||||||
Liverwort     TgTtTTTGGATTTtTATCtAAtGATCCAGGACGTAATCCtGGACGTAAAGAATAa
```

Fig. 19. Comparison of ORF36 cpDNA sequences from Douglas-fir, tobacco, and liverwort. Capital letters indicate the identical nucleotides among them.

```
        10         20         30         40         50
TCTAGATAAT AGACCCTACA TGCATGTATA AAGTCTTTTC ACTCACGAGA

        60         70         80         90        100
CCGAAGAGAA ATGAAGGTGG AAAAATGTGG GAAACAGATT TTTAATTCCA

       110        120        130        140        150
TAAAATAACA TTGTCTAATG ATTAGAAGGG ATGTAGCGCA GTTGGTAGCG

       160        170        180        190        200
TGTTTGTTTT GGGTACAAAA TGTCGCAGGT TCAAATCTGT CGACCCTACC

       210        220        230        240        250
TTTCCCTTCT TTTCTATGGA AAGAGAGAGG AACGAAAGAT CATTTGATAT

       260        270        280        290        300
CGATTCGAAT GGAACATCAA ATAATTGAAT CATATCAGAT GCGAAAGTGT

       310        320        330        340        350
TTTACTCTAA GAGTATAAAC GAAAGGTATT TTCCCTTCCC TTTATCTCCG

       360        370        380        390        400
GATGTTAAAG AAAGCGCTCT TAGTTCAGTT CGGTAGAACG TAGGTCTCCA

       410        420        430        440        450
AAACCTGATG CCGTAGGTTC AAATCCTACA GAGCGTGATT CTGTTCCTAT

       460
CAAATCCAAC CCTCTAGA
```

Fig. 20. The DNA sequence of a 0.5 kb XbaI cloned cpDNA fragment.

Appendix II. Supplementary detail on methods.

A. Procedures for chloroplast DNA isolation (modified from Palmer 1986).

1. Branches were collected, cut under water, and left in dark for 2-5 days. Soft, dark green, but matured new growth (often found on partly shaded foliage of 2- to 10-year-old trees) seems to give best results.
2. 50 gm of foliage and 400 ml of extraction buffer were prepared for each sample and were applied to three 37 ml sucrose gradients.
3. In a cold room, homogenize in steel blender with liquid 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 400 ml extraction buffer and mix well with spatula. Homogenize for about 1 minute in polytron, small probe, speed setting 4 to 5 (on scale of 1 to 10).
5. Pour into 2 funnels, each with 4 layers of cheesecloth above 1 layer of miracloth. Squeeze cheesecloth, rotate and "nudge" miracloth to speed filtration. Funnels drain into 250 ml GSA (Sorvall) tubes in tray of ice.
6. Centrifuge filtrate at 1,000 g (3,250 rpm in GSA rotor) for 15 minutes at 4°C.
7. Resuspend each pellet in 9 ml ice-cold wash buffer-I

with soft paint brush.

8. Load resuspended pellet onto a two-step gradient: 13 ml of 52% sucrose over 5 ml 41% sucrose and over 7 ml 30% sucrose (sucrose solutions all 50 mM tris-HCl, pH 8.0, 25 mM EDTA). Mix overlays slightly when making gradients to create diffuse interfaces so chloroplasts are not trapped.
9. Centrifuge gradients at 25,000 rpm for half hour at 4°C in AH-629 (Sorvall) swinging bucket rotor.
10. Remove chloroplast band using a wide-bore transfer pipette; avoid dense green "cap" in top third of gradient -- take more diffuse green "band" (1-3 ml thick) toward middle of gradient. Put in 50 ml oak-ridge style (Sorvall) centrifuge tube(s), dilute with 3-10 volumes of wash buffer-II, and centrifuge at 1,500 g (3,900 rpm in SA-600 rotor-Sorvall) for 15 minutes at 4°C.
11. Resuspend pellet in wash buffer-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 ul) 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; about 1,500g) to pellet residual starch and cell-wall debris.

15. Pour supernatant into 15 ml Falcon tube, add 4.48 gm freshly powdered CsCl, 119 ul ethidium bromide (10 mg/ml), and distilled water to 5.93 ml. Rock on nutator covered with foil until dissolved.
16. Transfer to 6 ml ultracentrifuge tube, top off with water or 75% CsCl solution, balance, and centrifuge in TFT 45.6 rotor (Sorvall) at 40,000 rpm, 22°C, overnight.
17. Extract ethidium bromide several times with isopropanol saturated with water 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.

<u>Extraction buffer</u>

0.35	M	Sorbitol
50	mM	tris-HCl (pH 8.0)
5	mM	EDTA
0.1	%	BSA
0.25	%	PVP
10.0	%	PEG
0.5	%	spermidine
0.5	%	spermine
0.5	%	2-mercaptoethanol

<u>Wash Buffer-I</u>

0.35	M	sorbitol
50	mM	tris-HCl (pH 8.0)
25	mM	EDTA
10	%	PEG

<u>Wash Buffer-II</u>

Wash Buffer-I without PEG

Lysis Buffer

5 %	sarkosyl (N-Lauroylsarcosine)
50 mM	tris-HCl (pH 8.0)
25 mM	EDTA

Dialysis Buffer

500 mM	tris-HCl (pH 8.0)
500 mM	NaCl
5 mM	EDTA

B. Protocol for cpDNA cloning

a. Vector preparation (Modified from Maniatis et al., 1982).

1. Prepare 10 ug pUC19 plasmid for reaction.
2. Add 2 ul restriction enzyme 10X buffer, 2 ug RNase A, 20 units XbaI, and water up to a total volume of 20 ul.
3. Incubate at 37°C for 4 hours and add 180 ul water for phenol/chloroform extraction.
4. Add 200 ul phenol/chloroform/isoamyl-alcohol (25:24:1) to the tube, mix by vortexing. Centrifuge for 5 minutes at 15,000 rpm in a microfuge at room temperature, and transfer supernatant to a new tube.
5. Repeat step 4.
6. Add 200 ul chloroform/isoamyl-alcohol (24:1) and vortex, centrifuge for 5 minutes as in step 4.
7. Repeat step 6. After centrifuge, take only 180 ul of supernatant (to prevent taking material from near interface layer).
8. Add 90 ul 7.5 M ammonium acetate and 540 ul (2 volumes) of 100% ethanol, mix by inverting gently and incubate at 4°C for 2-4 hours.
9. Centrifuge at 4°C in an eppendorf microfuge for 30 minutes; aspirate the supernatant and wash the pellet with ice-cold 70% ethanol.

10. Vacuum dry and resuspend the pellet in 48 ul 100 mM tris-HCl (pH 9.0) buffer.
11. Add 1 ul CIP (1 unit/ul) and incubate at 37°C for 30 minutes (CIP: Calf intestinal alkaline phosphatase, Boehringer Mannheim, 20 units/ul stock, diluted to 1 unit/ul with TE (10:.1) buffer before use).
12. Add 1 ul CIP solution and incubate at 37°C for 30 minutes again.
13. Stop the reaction by incubating at 68°C.
14. Add 150 ul water and bring the total volume to 200 ul for phenol/chloroform extraction.
15. Repeat steps 4-9.
16. Vacuum dry and resuspend the DNA pellet in TE (10:1) buffer; store at 4°C for future use.

    Note: If the restriction enzyme digest produces blunt ends (ex. SmaI), steps 11 and 12 should be changed as below:

    11. Add 1 ul CIP (1 unit/ul) and incubate at 37°C for 20 minutes and transfer to 58°C for 15 minutes.
    12. Repeat step 11.

b. Ligation Reaction (modified from Maniatis et al., 1982)

1. Measure the cpDNA concentration by fluorometry (Hoefer Science instruments. TKO 100 DNA Mini-fluorometer), purified pUC19 plasmid DNA is used as a standard.
2. Digest 3 ug cpDNA with XbaI overnight.
3. Separate the XbaI restriction fragments by gel electrophoresis on 0.6% low melting point agarose (Boehringer Mannheim product for best result, see Appendix II. A.).
4. Cut out bands after staining with ethidium bromide, visualized with long-wave UV light (see Appendix II. B. for comparison of cloning efficiency under long and short wave UV light).
5. Melt the gel containing the suitable fragment at 68°C and estimate the volume (normally around 50 ul).
6. Take 10 ul (after calculation the concentration showed be near to 1 fM (10^{-15} M)) of gel samples; add 3 fM of vector (ratio 1:3); mix by vortexing and coincubate at 68°C for 5 minutes.
7. Add 6 ul 5X ligation buffer (250 mM Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 25 % (w/v) polyethylene glycol 8000, 5 mM ATP, 5 mM dithiothreitol), 0.5 units T4 ligase, and water to a final volume of 30 ul.

8. After vortexing, incubate overnight at 15°C.

c. Transformation (modified from Maniatis et al., 1982)

1. Prepare competent cells (E. coli DH5 alpha) 200 ul and add 10 ul of sample solution (after ligation described before); mixed by tapping the tube gently.
2. Incubate on ice for 30 minutes.
3. Transfer the tube to 42°C for heat shock 90 seconds.
4. After heat shock, put onto ice and add 800 ul LB medium.
5. Incubate at 37°C shaker (225 rpm) for 1 hour.
6. Plate 200 ul cell on each LB plate containing ampicillin (100 ug/ml); incubate at 37°C overnight.

C. Protocol for plasmid isolation

a. Minipreparation of plasmid DNA (modified from Maniatis et al., 1982)

1. Prepare the overnight culture of bacteria cell in 5 ml LB medium containing ampicillin (100 ng/ml).
2. Take 1.5 ml of the culture into an microfuge tube. Centrifuge at 5,000 rpm for 3 minutes, aspirate the supernatant.
3. Resuspend the pellet in 100 ul of ice-cold lysis buffer containing lysozyme (4 mg/ml) by vortexing.
4. Keep at room temperature for 5 minutes.
5. Add 200 ul of 0.2 N NaCl and 1% SDS solution to each tube. Invert the tube several times, then store on ice for 5 minutes.
6. Add 150 ul of 3M sodium acetate (pH 4.8) to each tube. Invert the tube several times (make sure the solution mixes well), then store the tube on ice for 5 minutes.
7. Centrifuge at 10,000 rpm (Beckman microfuge 11 model) for 10 minutes at room temperature.
8. Transfer the supernatant into a new tube.
9. Add 1 ml of 100% ethanol to each tube, and store at room temperature for at least 2 minutes.
10. Centrifuge in Eppendorf microfuge for 15 minutes at $4^{o}C$.

11. Aspirate the supernatant and wash the pellet with 70% ethanol.
12. Vacuum dry the pellet and dissolve in 50 ul of boiled RNase solution (100 ug/ml), incubate at 37°C for 3 hours.
13. Add 150 ul water to each tube to bring the total volume to 200 ul for phenol/chloroform extraction.
14. Add 200 ul phenol/chloroform/isoamyl-alcohol (25:24:1) solution to each tube, then vortex completely and centrifuge at 10,000 rpm for 5 minutes.
15. Transfer the aqueous phase (upper layer) to a new tube, and add 200 ul chloroform/isoamyl-alcohol (24:1) solution to each tube. Vortex the tube completely and centrifuge at 10,000 rpm for 5 minutes.
16. Transfer 180 ul aqueous solution to a new tube (do not take the interface layer by accident), and then add 90 ul of 7.5 M ammonium acetate and 540 ul of 100% ethanol. Store the tube at 4°C overnight.
17. Centrifuge the tube in Eppendorf microfuge for 15 minutes at 4°C, and aspirate the supernatant. Wash the pellet with 70% ice-cold ethanol and vacuum dry the pellet.
18. Resuspend the DNA pellet with TE (10:1) buffer and store at 4°C for future use.

b. Large scale preparation of plasmid DNA (modified from Maniatis et al., 1982)

1. Centrifuge 25 ml of overnight culture in a 50 ml centrifuge tube at 5,000 rpm for 3 minutes and aspirate the supernatant.
2. Add 1 ml of lysis buffer containing fresh prepared lysozyme (4 mg/ml), and vortex the tube.
3. Store at room temperature for 5 minutes.
4. Add 2 ml of 0.2 N NaOH and 0.1% SDS solution, and invert the tube several times.
5. Store the tube on the ice for 5 minutes.
6. Add 1.5 ml of 3 M sodium acetate (pH 4.8) and invert the tube several times; make sure the solution mixes completely.
7. Store the tube on ice for 5 minutes, and then centrifuge in a SA-600 rotor (Sorvall) 1,000 rpm (14,000 g) for 10 minutes at room temperature.
8. Transfer the supernatant to a new tube (50 ml), and add 10 ml of 100% ethanol; store the tube at room temperature for 5 minutes.
9. Centrifuge in SA-600 rotor 12,000 rpm (20,000 g) for 10 minutes at 4°C.
10. Aspirate the supernatant and wash the plasmid DNA pellet with ice-cold 70% ethanol.
11. Vacuum dry the pellet and resuspend in 1 ml DNase free RNase (boiled RNase A) solution (100 ug/ml);

incubate at $37^{o}C$ for 3 hours.

12. Split the 1 ml DNA solution into two 1.5 ml microfuge tubes (each one contains 500 ul), and add 500 ul phenol/chloroform/isoamyl-alcohol (25:24:1) solution to each tube.
13. Vortex vigorously, centrifuge at microfuge 10,000 rpm for 5 minutes at room temperature. After centrifuging, transfer supernatant to new tubes.
14. Add 500 ul of chloroform/isoamylalcohol (24:1) to each tube and mix by vortexing. After centrifuge for 5 minutes in microfuge at room temperature, transfer supernatant to a new tube.
15. The total volume of recovery from each sample is around 900 ul, then split the solution into 3 tubes (300 ul each) and add 150 ul 7.5 M ammonium acetate and 900 ul 100% ethanol to each tube.
16. Mix by inverting the tube gently, and store at $4^{o}C$ overnight.
17. Centrifuge for 15 minutes in an Eppendorf at $4^{o}C$, and remove the supernatant by aspiration.
18. Wash the pellet with 70% ice-cold ethanol and vacuum dry in a vacuum desicator.
19. Resuspend the pellet in 100 ul TE (10:1) (pH 8.0) buffer and store in refrigerator or freezer.

<u>Lysis Buffer</u>

50 mM glucose — 25 mM tris-HCl (pH 8.0)
10 mM EDTA — (filter sterilize and store at $4^{o}C$)

D. Protocol for plasmid sequencing (modified from Yanisch-Perron et al., 1985; Chen and Seeburg, 1985; and Zhang et al., 1988)

1. Prepare 2 ug of insert-containing plasmid DNA for sequencing; add water to a total volume of 10 ul.
2. Add 10 ul (1 volume) of denaturing solution (0.4 N NaOH and 0.4 mM EDTA); mix and set at room temperature for 5 minutes (final concentration is 0.2 N NaOH and 0.2 mM EDTA).
3. Add 2 ul of 2 M ammonium acetate (pH 4.5) and 44 ul (2 volumes) 100% ethanol; mix by tapping the tube gently to insure complete mixing.
4. Incubate on dry ice for 15 minutes. (Note: If the DNA solution was not mixed well at step 3, there will be a white salt pellet coming out after dry ice incubation).
5. Centrifuge at 4°C with Eppendorf microfuge for 30 minutes.
6. Using a fine pipette, pipette out the supernatant and wash the pellet with 70% ice-cold ethanol.
7. Vacuum dry and resuspend the DNA in 7 ul TE (10:1) or water.
8. Add 1 ul of primer (0.5 pmole/ul) and 2 ul 5X Sequenase buffer (200 mM tris-HCl pH 7.5, 100 mM $MgCl_2$, and 250 mM NaCl); mix well and incubate for 2 minutes in a 65°C heat block. (Sequenase kit can be

purchased from United States Biochemical corporation).

9. After incubation, turn off the heat block and wait until the temperature drops to below 35°C gradually (about 45 minutes) for DNA annealing.
10. Prepare 4 tubes labeled G, A, T, and C for each reaction sample; add 2.5 ul of termination mixture to bottom of the tubes (ddGTP in tube G, ddATP in tube A, ddTTP in tube T, and ddCTP in tube T).
11. When the sample is cooled to 35°C, add 1 ul 0.1 M DTT (Dithiothreitol), 2 ul diluted Labeling Mix (1X, 1.5 uM dGTP, 1.5 uM dCTP, 1.5 uM dTTP), 0.5 ul [alpha-^{35}S]dATP (10 uCi/ul), and 2 ul diluted Sequenase.
12. Mix thoroughly and incubate at room temperature for 5-10 minutes.
13. Prewarm the tubes (at 37°C) which contained the termination mixture at least 1 minute.
14. When the labeling reaction is complete, transfer 3.5 ul to each tube at the inside wall of the tube (do not mix with the termination mixture).
15. Place the tubes into microfuge carefully, and spin down all samples at the same time.
16. Mix the sample by tabbing the tubes and incubate in a 37°C water bath for 5-30 minutes.
17. Stop the reaction by adding 4 ul of Stop Solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) for each tube. Store the samples in

freezer or use directly.

Appendix III. Supplementary detail on results

A. Effect of agarose on transformation efficiency

1. pUC19 plasmid as a sample source for different treatments.
2. Treatments:
   a. DNA digested with XbaI and then ligated.
   b. DNA digested with XbaI and then electrophoresed in BRL (Bethesda Research Labs) (lot #: 50701) low melting point agarose gel (0.6%); cut out the band and ligate directly in gel.
   c. DNA digested with XbaI and then electrophoresed in BM (Boehringer Mannheim) (lot #: 134601/9-88) low melting point agarose gel (0.6%); cut out the band and ligate directly in gel.
   d. DNA digested with XbaI.
   e. DNA no treatment as control.
3. Transformation: calculate the efficiency
   a. Cut and religate: 2×10^6
   b. In BRL gel: 1×10^5
   c. In BM gel: 3×10^6
   d. Cut plasmid: 1×10^2
   e. Uncut plasmid: 5×10^6
4. Conclusion: The more highly purified BM low melting point agarose is more suitable for direct cloning by using ligation and transformation within the gel than

is BRL low melting point agarose. The efficiency for transformation in BM is 30 times better than in BRL.

B. Effect of long and short wave UV light on cloning efficiency.

1. Digest pUC19 plasmid with XbaI.
2. Load 1 ug per lane on 0.6% low melting point agarose gel (Boehringer Mannheim).
3. After gel electrophoresis, the treatments for each lane shown as below (Short-wave treatments on Fotodyne Foto/PrepI transilluminator in preparative mode) (which is not purported to cause photo-damage to DNA with short exposures):
   a. cut band with long-wave UV light
   b. cut band with short-wave UV light exposed for 1 minute
   c. cut band with short-wave UV light exposed for 2 minutes
   d. cut band with short-wave UV light exposed for 4 minutes
   e. cut band with short-wave UV light exposed for 6 minutes
   f. cut band with short-wave UV light exposed for 8 minutes
   g. cut band with short-wave UV light exposed for 10 minutes
   h. control: plasmid not run on gel
4. After ligation at 15°C overnight, measure sample concentration with fluorometer.

5. Take 20 ng for transformation.
6. Results shown below:
   a. LW: 210 colonies (transformation efficiency 10^5 transformants/ug DNA)
   b. SW1: 1 colony
   c. SW2: 6 colonies
   d. SW3: 2 colonies
   e. SW6: 0 colony
   f. SW8: 2 colonies
   g. SW10: 1 colony
   h. control (transformation efficiency 10^6): innumerable colonies.
7. Conclusion: ligation directly within the gel followed by transformation is strongly affected by UV light. Long-wave UV light has a small effect, but even short exposures to short-wave UV light are very inhibitory. This may be due either to photonicking of DNA or production of compounds that inhibit DNA ligase.

C. Determination of cpDNA restriction fragment size

1. Douglas-fir cpDNA digested with PvuII, XbaI, and SstI and double digested with PvuII and XbaI, and SstI and XbaI, after electrophoresis then transferred to nylon membrane.
2. Calculate the distance from the sample loading well origin to the fragments which were shown up on the film after hybridization and exposure (the mapping data film is on Appendix III.).
3. Use the lambda HindIII fragments as marker and also the Douglas-fir cpDNA PvuII fragments (already calculated by Howe G. et al., 1988) as a standard to calculate the XbaI restriction enzyme mapping fragments.
4. The program used for calculating the fragment size is described below. It solves the following equation via least-squares analysis: Log (Molecular Weight) = $B_1 + B_2$ (Migration Distance = MD) + B_3 (MD^2) + B_4 (MD^3).
   1. Known fragment sizes and distances are input (file = FRAGSAS.DAT).
   2. Unknown fragment distances are input (file = DIST.DAT).
   3. The Program (file = FRAG.SAS) is designed to run on the SAS (Statistical Analysis System, Cary, North Carolina) software package.

```
(Input/Output control information on this line)
DATA MOLWT;
  INFILE 'FRAGSAS.DAT';
  INPUT KB DIST;
   LKB=LOG(KB);
   DIST2= DIST**2;
   DIST3= DIST**3;
PROC REG DATA=MOLWT OUTEST=REGOUT;
  MODEL LKB=DIST DIST2 DIST3;
DATA MOLWT1;
  INFILE'DIST.DAT';
  INPUT DIST;
   DIST2=DIST**2;
   DIST3=DIST**3;
PROC SCORE DATA=MOLWT1 SCORE=REGOUT OUT=NEWD
TYPE=PARMS;
    VAR DIST DIST2 DIST3;
DATA FIN; SET NEWD;
   MOLWT=EXP (MODEL1); RUN;
PROC PRINT; VAR DIST MOLWT;
RUN;
```

D. Douglas-fir cpDNA cloning

1. Ligation within the gel and direct transformation:

Cloned Fragments	Uncloned Fragments
Xba-11.2	Xba-13.8
Xba- 8.9	Xba-10.6
Xba- 6.9	Xba- 5.9b
Xba- 6.0	Xba- 1.5b
Xba- 5.9a	
Xba- 5.1	
Xba- 4.8	
Xba- 4.3	
Xba- 3.2a	
Xba- 2.7	
Xba- 1.5a	

2. "Shut-gun" cloning method (unfractionated cpDNA)

Cloned Fragments

Xba-3.8
Xba-3.2b
Xba-2.5
Xba-2.4
Xba-1.9
Xba-1.4
Xba-0.9
Xba-0.8
Xba-0.6
Xba-0.5

3. Based on the XbaI, PvuII, SstI restriction map, there are several postulated fragments (i.e., unobserved and uncloned):

a. 6.4 kb fragment between Xba-8.9 and Xba-13.8

b. 1.6 kb fragment between Xba-3.8 and Xba-5.9

c. 1.0 kb fragment between Xba-0.6 and Xba-2.7

d. 1.5 kb fragment between Xba-4.3 and Xba-5.1

e. 0.3 kb fragment between Xba-1.5 and Xba-4.8

4. Problems encountered during cpDNA cloning:

a. Despite numerous attempts, we were unable to clone two large fragments of 13.8 kb and 10.6 kb. (For the 13.8 kb fragment, I did five cloning experiments and screened around 100 total colonies; for the 10.6 kb fragment, I did four cloning experiments and screened around 80 total colonies) My failure to clone these fragments may be due to several factors:

(1) pUC19 vector capacity may be incapable of maintaining such large inserts; however, we stably cloned an 11.2 kb insert, thus this explanation alone in untenable.

(2) The host strain may be unable to accommodate certain inserts, such as those able to form secondary structures.

(3) Large inserts may be damaged by in gel manipulations during cloning, effectively reducing cloning efficiency.

(4) Transformation frequency was not high enough to clone many large fragments (the average transformation efficiency was 10^6 transformation per microgram of uncut pUC19 plasmid).

b. XbaI digestion of Douglas-fir cpDNA generates at least 3 doublets, one of 5.9, one of 3.2, and one of 1.5 kb. The identification of clones from doublet fragments was difficult.

c. XbaI generates many small (below 2 kb) fragments which are difficult to detect in ethidium bromide-stained gels containing a wide range of cpDNA sizes (100 bp to about 30 kb).