Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3898–3902, June 1988
Evolution

Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged

(gymnosperms/evolution/chloroplast DNA/restriction maps/gene mapping)

STEVEN H. STRAUSS*†, JEFFREY D. PALMER‡, GLEN T. HOWE*, AND ALLAN H. DOERKSEN*

*Department of Forest Science, Oregon State University, Corvallis, OR 97331; and ‡Department of Biology, University of Michigan, Ann Arbor, MI 48109

Communicated by R. W. Allard, February 8, 1988 (received for review December 8, 1987)

ABSTRACT **Chloroplast genomes of Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] and radiata (Monterey) pine [*Pinus radiata* D. Don], two conifers from the widespread Pinaceae, were mapped and their genomes were compared to other land plants. Douglas-fir and radiata pine lack the large (20–25 kilobases) inverted repeat that characterizes most land plants. To our knowledge, this is only the second recorded loss of this ancient and highly conserved inverted repeat among all lineages of land plants thus far examined. Loss of the repeat largely accounts for the small size of the conifer genome, 120 kilobases, versus 140–160 kilobases in most land plants. Douglas-fir possesses a major inversion of 40–50 kilobases relative to radiata pine and nonconiferous plants. Nucleotide sequence differentiation between Douglas-fir and radiata pine was estimated to be 3.8%. Both conifer genomes possess a number of rearrangements relative to *Osmunda*, a fern, *Ginkgo*, a gymnosperm, and *Petunia*, an angiosperm. Among land plants, structural changes of this degree have occurred primarily within tribes of the legume family (Fabaceae) that have also lost the inverted repeat. These results support the hypothesis that the presence of the large inverted repeat stabilizes the chloroplast genome against major structural rearrangements.**

Chloroplast genome structure has been studied in a great variety of plants (reviews in refs. 1–3). Among land plants it generally exists as a small, circular molecule, usually between 135 and 160 kilobases (kb) in length. It is highly conserved in size and gene arrangement compared to plant nuclear and mitochondrial genomes. Most land plants, including members of all families of angiosperms examined, one gymnosperm, three ferns, and two bryophytes, possess a major inverted repeat of roughly 10–76 kb that contains the rRNA genes and adjacent DNA. Species lacking this repeat are known only in a single group of land plants consisting of several allied tribes in the subfamily Papilonoideae of the legume family (Fabaceae) (4–6). Gene order is highly conserved among most species that possess the inverted repeat. Large numbers of chloroplast DNA (cpDNA) inversions have been documented only in those legumes lacking the inverted repeat and in *Pelargonium* (7); other rearranged genomes differ by only one or two inversions (1, 2, 4–6, 8–11).

Although chloroplast genome structure has been studied in at least 78 genera and 200 species of angiosperms (2), only a single gymnosperm, *Ginkgo biloba* L., has been studied (12). The phylogenetic diversity and antiquity of gymnosperms, and the ecologic and economic importance of the Pinaceae, justify a closer analysis. We used Southern blotting and filter hybridization to construct restriction site and gene maps of two of the most economically important forest tree species worldwide, Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] and radiata (Monterey) pine [*Pinus radiata* D. Don]. We also compared their genome structure to that of angiosperms. We report that the chloroplast genomes of both conifers differ considerably in structure from the vast majority of land plants and from the monospecific order (13) of gymnosperms represented by *Ginkgo biloba*. The two conifer genomes lack an inverted repeat common to most land plants, contain a number of inversions not present in *Ginkgo* or angiosperm cpDNA, and are differentiated from each other by a major inversion of 40–50 kb.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

cpDNA was isolated from needles of radiata pine and Douglas-fir by using a sucrose-gradient method (14) (additions to the extraction buffer were 10% polyethylene glycol 8000, 0.1% polyvinylpyrrolidone, and 0.5% 2-mercaptoethanol; an addition to the wash buffer was 0.5% 2-mercaptoethanol). Methods for agarose gel electrophoresis, bidirectional Southern transfer, hybridization, and isolation of plasmid DNA were as described (14). cpDNA fragments from a Douglas-fir *Pvu* II digest were cut out of gels, electroeluted in dialysis chambers, extracted with phenol/chloroform and ethyl ether, and precipitated with ethanol. Recombinant plasmids and gel-isolated DNA were radioactively labeled with ^{32}P by nick-translation (15). Filters were washed in 300 mM NaCl/30 mM trisodium citrate and 0.1% or 0.5% sodium dodecyl sulfate for four 30-min periods at 65°C prior to autoradiography. The sources and characteristics of the gene fragments (4, 8, 12), *Petunia*/mung bean fragments (12), and tobacco fragments (16) used as probes have been described. Additional details on hybridizations are given elsewhere (17).

RESULTS

Six of 38 gene probes failed to give detectable hybridization to either Douglas-fir or radiata pine cpDNA, despite their use in at least two separate hybridization experiments. All of the gene probes that gave undetectable hybridization code for ribosomal proteins (1.8-kb 3′ rps7, 0.5-kb 3′ rp122–5′ rps3, 1.0-kb 3′ rp116–rp114–5′ rps8, and 1.9-kb rps16 fragments from tobacco; and 0.7-kb 3′ rps8–infA and 0.6-kb rps11 fragments from spinach) (8). These results suggest that at least some ribosomal protein genes are either deleted or diverge more rapidly in sequence than genes for photosynthetic proteins and rRNAs. However, there is considerable variation in the size and proportion of coding region in the various gene probes used, making inferences based on hybridization efficiency tentative.

Chloroplast genome size is about 121 kb in Douglas-fir and 120 kb in radiata pine based on restriction fragment analysis and restriction site mapping (Fig. 1). The genomes are circular and lack any large repeated sequences. In particular,

Abbreviation: cpDNA, chloroplast DNA.
†To whom reprint requests should be addressed.

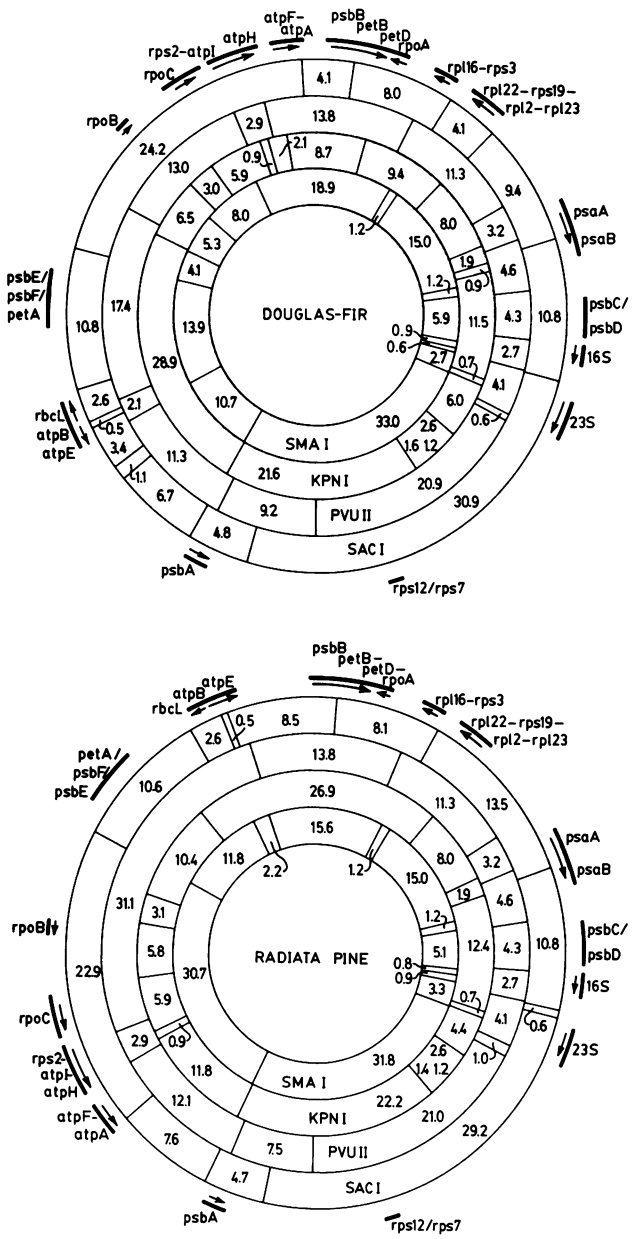

FIG. 1. Circular restriction site and gene maps of Douglas-fir and radiata pine chloroplast genomes. The order of genes separated by dashes has been inferred from their arrangement in tobacco (16). The linear order of genes separated by slashes was also determined by analogy to tobacco, but the orientation of the gene cluster with respect to the remainder of the genome could not be determined. Transcriptional directions of other genes and the location and transcriptional directions of 16S and 23S rRNA have been assigned by analogy to tobacco (16). Dark bands represent the lengths of gene probes used in hybridizations, rather than the lengths of the genes themselves, except for 16S and 23S rRNA.

they contain only a single set of rRNA genes and lack the large inverted repeat found in most land plants. The absence of this typically 25-kb duplication largely accounts for their small size (120 kb) compared to most other land plants (140–160 kb). A single major inversion of 40–50 kb distinguishes Douglas-fir from radiata pine (Fig. 2). This inversion spans the region of the conifer genomes between *atpA* and *atpE*, which corresponds to a major portion of the large single-copy region of those genomes that possess an inverted repeat. Aside from this inversion, the two conifer genomes appear to be colinear.

We estimated the extent of nucleotide sequence differentiation between radiata pine and Douglas-fir by analyzing their restriction site differences (Table 1). This calculation assumes that site gains or losses are due to nucleotide substitutions within enzyme recognition sequences. Sites at or near the junctions of the major inversion were ignored, as were small fragment size differences that likely result from insertions and deletions within fragments and from fragment size estimation error. Estimates from the four restriction enzymes studied were fairly consistent, ranging from 2.1% for *Sma* I to 4.8% for *Kpn* I, with a mean of 3.8%.

We used a set of cpDNA clones from *Petunia* and mung bean to represent the ancestral, or consensus, vascular plant cpDNA genome arrangement (12). Hybridization of these clones to radiata pine and Douglas-fir indicated that a number of rearrangements have taken place since the Pinaceae diverged from its common ancestor with angiosperms (Fig. 3). Radiata pine is most similar to *Petunia*/mung bean, whereas Douglas-fir has additional rearrangements owing to the major inversion that distinguishes it from radiata pine. Nonetheless, a number of rearrangements must be postulated to derive radiata pine cpDNA from the ancestral vascular plant genome represented by *Petunia*/mung bean (12). Allowing for only deletions and inversions, six events are required, two deletions and four inversions (Fig. 4). Step 1 results in shrinkage of the inverted repeat, shown as a deletion of part of the left repeat. This results in an inverted repeat similar in size to that in *Ginkgo* (12), another gymnosperm. Small inverted repeats like those in *Ginkgo*, however, may be the ancestral condition for land plants (12); thus, step 1 may not have been necessary were the true ancestral vascular plant cpDNA, rather than *Petunia*/mung bean cpDNA, used as a starting point. Step 2 results in loss of an inverted repeat, and steps 3–6 invert sections of the genome to achieve colinearity of homologous sections.

DISCUSSION

The estimated degree of cpDNA sequence divergence between Douglas-fir and radiata pine, 3.8% on average, fits within the range of values reported for other fairly closely related species. By using restriction site differences, estimates from intrageneric species comparisons have ranged from 0 to 13.8%; intergeneric estimates roughly span the range reported for species within genera, 0.22–10.8% (ref. 19; reviewed in ref. 8). Comparison of coding sequences for *atpB* and *rbcL* in maize and barley, taxa that are comparable to *Pinus* and *Pseudotsuga* in apparent time of divergence, show 5.6% substitution averaged over proteins and codon positions (20). The modest degree of sequence differentiation between Douglas-fir and radiata pine probably reflects a slow rate of sequence evolution and their close relationship as conifers. A recent study of seed protein immunology confirmed their placement in the same section of the Pinaceae (Pinoid group) (21). Douglas-fir may have arisen from a pine-like ancestor 50 million years ago, the time when *Pseudotsuga* megafossils (seed-bearing cones) have been first observed in the fossil record (22).

The chloroplast genomes of Douglas-fir and radiata pine are highly rearranged in comparison to the genomes of most other land plants that have been studied. At least three factors may be responsible for this. First, conifers are ancient. The conifer line is thought to have diverged from that of *Ginkgo*, their closest living relative (23, 24), roughly 300 million years ago (24). Thus, even a small change in the factors that constrain cpDNA rearrangement could have a substantial net impact. However, other groups of land plants, such as ferns and *Ginkgo* itself, are extremely ancient but differ little from the typical angiosperm cpDNA structure (12). Thus, evolutionary antiquity alone is an insufficient explanation.

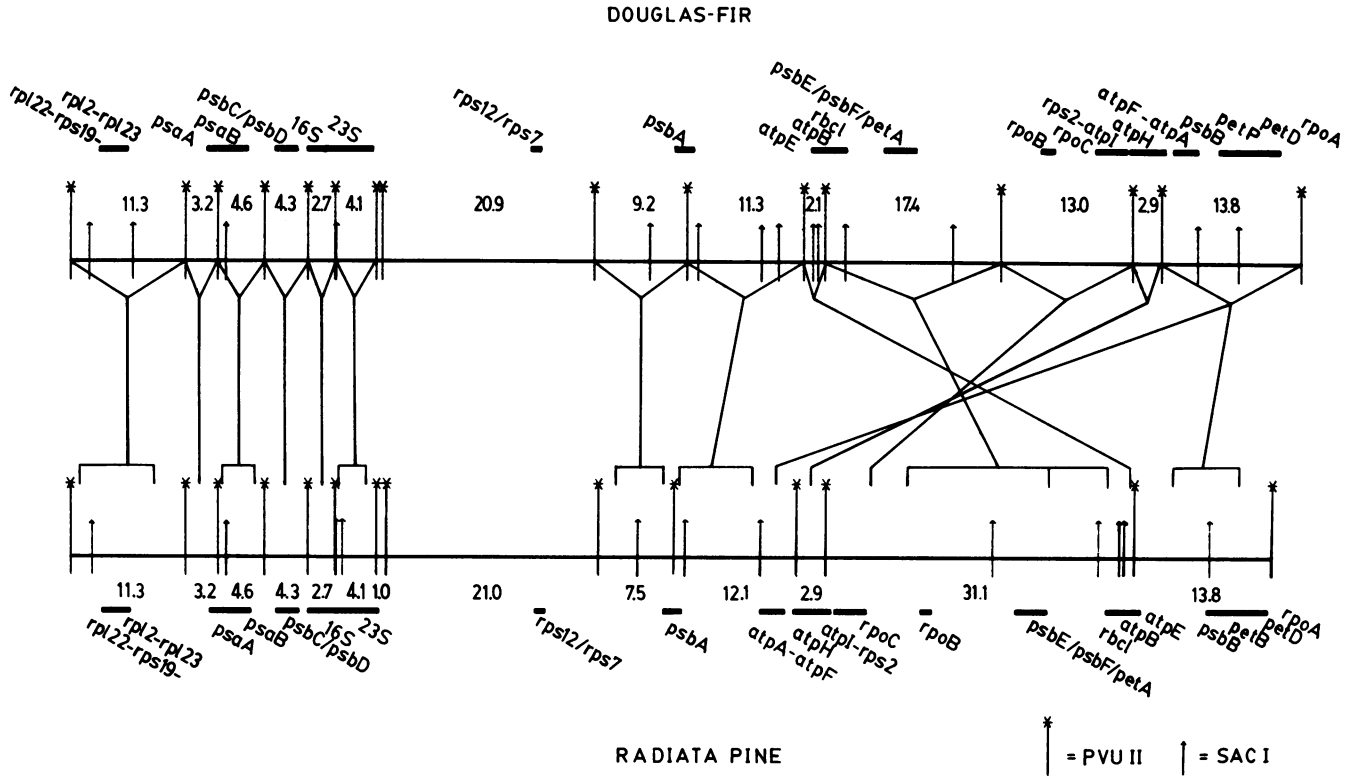

FIG. 2. Arrangement of homologous sequences in the chloroplast genomes of Douglas-fir and radiata pine. Douglas-fir *Pvu* II fragments were cut from gels and used to probe filters containing cpDNA restriction fragments of Douglas-fir and radiata pine. Lines between the two linearized maps connect cross-hybridizing Douglas-fir and radiata pine fragments. *Pvu* II fragment sizes are given in kb. The *rpl16* and *rps3* genes, which overlap the left and right ends of the maps, are not shown here but were mapped and are shown in Figs. 1 and 3.

Second, loss of the inverted repeat may have permitted the genome to undergo more rapid rearrangement (4). Among legumes (Fabaceae) of the subfamily Papilonoideae, the only other known group of land plants to have lost the inverted repeat, a number of cpDNA rearrangements were shown to have occurred after the repeat was lost (5). In genera of the subfamily retaining the inverted repeat, little or no rearrangement occurred. However, not all genera that lost the repeat showed cpDNA rearrangement. Thus, loss of the repeat may lessen constraints on rearrangement but not actually induce it. The possible mechanisms by which the repeat may constrain rearrangement have been discussed elsewhere (2, 25, 26). Briefly, they include the following. (*i*) Homologous recombination-promoting enzymes may be titrated by the lengthy repeat units when the repeats are paired, similar in principle to the inhibition of plasmid replication due to sequestration of plasmid-P1 replication proteins by *incA* sequences (27). Flip-flop heterogeneity of cpDNA (2), the existence of equimolar isomeric forms of cpDNA with the single-copy regions in opposing orientations, suggests that there are ample recombination enzymes to mediate exchange between homologous regions within the repeat; in so doing, most recombination enzymes may be bound, decreasing their availability elsewhere in the genome. (*ii*) The extensive region capable of homologous pairing and recombination may force a coiled cpDNA molecule into a conformation that physically precludes recombination between other loci. (*iii*) The repeat may effectively suppress inversions that either span a repeat unit or have one end in the repeat and the other in a single-copy region because recombination between the resulting direct repeats will lead to deletion of a major portion of the cpDNA molecule. Such deleted chloroplast molecules have been observed in heterotrophic plantlets derived from anther culture in wheat (26).

Table 1. Sequence divergence between Douglas-fir and radiata pine cpDNAs

Enzyme	Restriction sites,* no. Radiata pine	Douglas-fir	Shared sites, no.	Sequence divergence,† %
Sac I	11	14	10	3.73
Pvu II	13	13	10	4.39
Kpn I	16	16	12	4.81
Sma I	11	14	11	2.13
Mean	12.75	14.25	10.75	3.76

*Restriction sites near the ends of the large inversion were not included. Small differences in size of otherwise comparable fragments in radiata pine and Douglas-fir were assumed to be due to within-fragment insertions and deletions and were thus ignored.

†Percent sequence divergence (PSD) was calculated from equations 9 and 10 of Nei and Li (18): $PSD = -(100)(3/2)\ln[(4s^{1/2r} - 1)/3]$, where $s = 2n_{xy}/(n_x + n_y)$, r = number of nucleotides recognized by enzyme, n_{xy} = number of sites shared by taxa x and y, and n_x and n_y = total number of sites in taxa x and y, respectively.

A third factor limiting cpDNA rearrangement may be the paucity of dispersed repeats, potential loci of homologous recombination (1). Dispersed repeats larger than about 10 to 20 base pairs are virtually absent from cpDNA in the majority of land plants. However, Douglas-fir, like wheat (28), appears to possess several families of repetitive sequences that are dispersed throughout the genome (C.-H. Tsai and S.H.S., unpublished data). The importance of dispersed repeats to cpDNA rearrangement is suggested by their presence in two highly rearranged cpDNAs, that of subclover (*Trifolium subterraneum*) (5) and geranium (*Pelargonium hortorum*) (7), though other rearranged legume genomes appear to lack repeats (5). Dispersed repeats coincide with certain inversion endpoints in wheat (10, 11, 29). Moreover, inversion endpoints seem to occur in the same intergenic spacers in a number of taxa, despite a considerable number of other

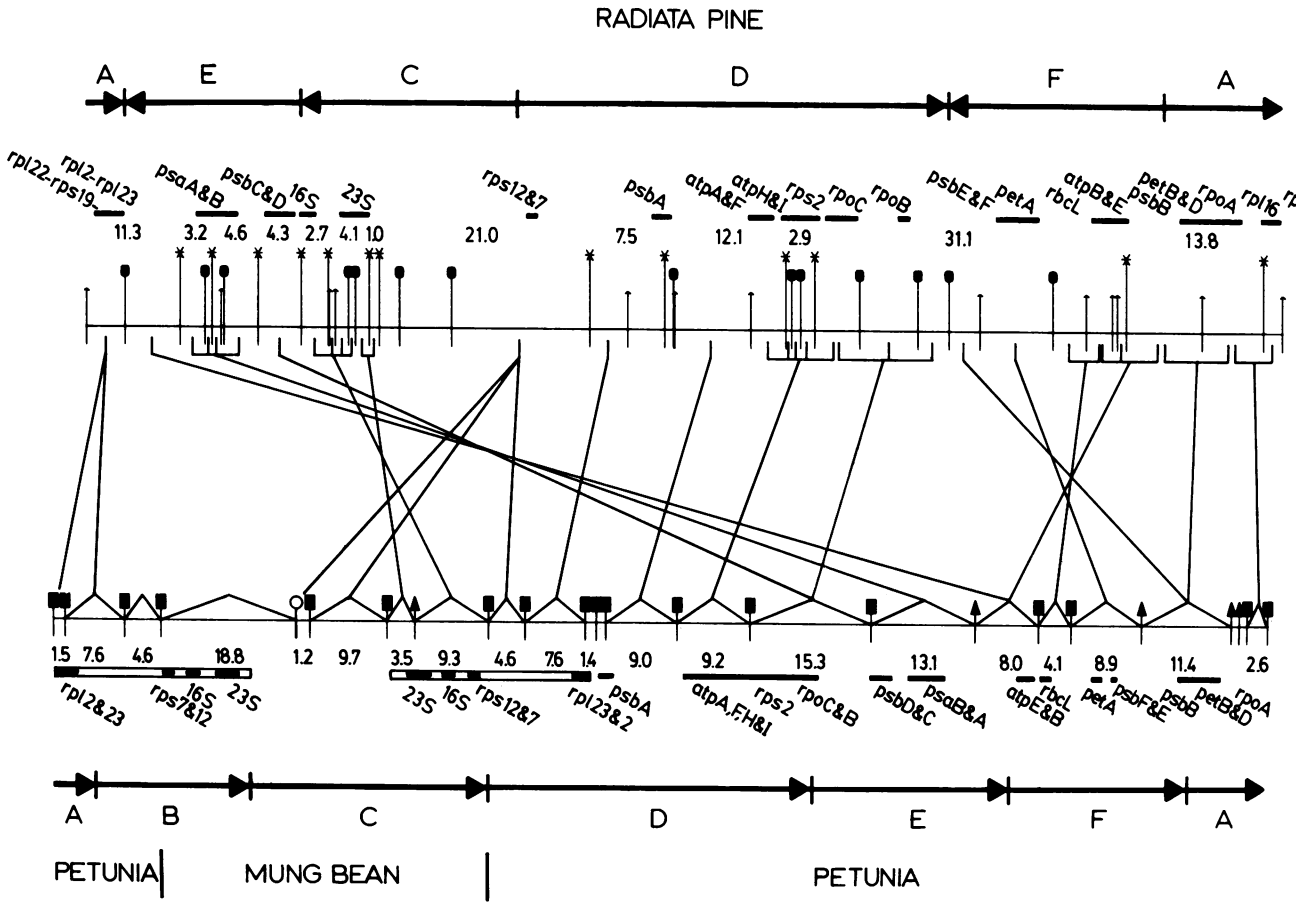

FIG. 3. Arrangement of homologous sequences in radiata pine cpDNA and an ancestral vascular plant chloroplast genome, here represented by a combination of portions of the *Petunia* and mung bean genomes. Cloned *Petunia* and mung bean fragments were hybridized to radiata pine fragments immobilized on filters. Scales differ slightly for the two maps. Lettered arrows above and below maps indicate fragment blocks involved in hypothesized evolutionary rearrangements of a *Petunia*–mung bean-like ancestral cpDNA that led to the present-day radiata pine cpDNA structure (Fig. 4). Block endpoints are approximate; they are located solely with respect to *Petunia* and mung bean fragment hybridizations and not with respect to positions of individual chloroplast genes. Block A is continuous but is shown in two sections in this linearized map. To aid interpretation, hybridizations of angiosperm inverted repeat fragments are shown for only one of the two repeated segments for each location in radiata pine; hybridizations displayed were chosen for consistency with the model of genome evolution in Fig. 4. The hybridization pattern for the 9.3-kb mung bean fragment was inferred from hybridization of 18.8-, 1.2-, 9.7-, and 3.5-kb mung bean fragments. Five *Bam*HI clones (Ba5, Ba6, Ba11, Ba14, and Ba22) (16), from the inverted repeat of tobacco, and two subclones, from the 9.7-kb *Pst* I fragment of the small single-copy region of mung bean cpDNAs (5), were also used as probes but were excluded from the figure for clarity. Results from these hybridizations supported the relationships presented. *Petunia* gene locations for rps19, rpl2, 16S, 23S, psbA, atpA, atpH, atpB, and rbcL are from Palmer and Stein (12); others were determined by analogy to the highly colinear tobacco chloroplast genome (16). Restriction site symbols are as in Fig. 2 and ↑ = *Sal* I, ↓ = *Pst* I, ♀ = *Xba* I, and ♦ = *Kpn* I.

intergenic spacers where such inversions might logically occur [e.g., in tobacco (16)]. One endpoint of the inversion

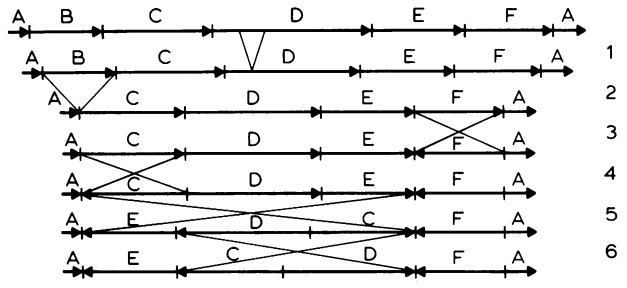

FIG. 4. Hypothesized deletions and inversions during evolution of the radiata pine chloroplast genome from a *Petunia*–mung bean-like ancestral genome shown in Fig. 3. Step 1 is deletion of a part of one repeat, similar to that seen in *Ginkgo* (12), the sole member of a different gymnosperm order. The evolutionary direction of the deletion is not clear (12), whereas the other five mutations shown are all clearly derived in a conifer-specific lineage. Step 2 is deletion of the inverted repeat. Steps 3–6 are inversions. The sequence of rearrangements that occurred during conifer evolution may differ from that presented.

that distinguishes Douglas-fir and radiata pine maps close to an endpoint of another inversion that putatively occurred during conifer evolution (between *atpE* and *psbB* in radiata pine, Fig. 3; junction of sections F and A, Fig. 4). The other endpoint lies downstream of *atpF–atpA*, similar to endpoints of inversions in pea, mung bean, *Oenothera*, lettuce, and wheat (reviewed in ref. 8). Putative inversion endpoints for conifer cpDNA also coincided for steps 4 and 5 (Fig. 4; region between *rpl23* and *psaA* in radiata pine, Fig. 3) and steps 5 and both 6 (Fig. 4; region between *rpoB* and *psbE* in radiata pine, Fig. 3) and 3 (Fig. 4; region between *psaA* and *atpE* in *Petunia*). Sequencing of repeats, particularly those found at endpoints of inversions, followed by hybridization to other taxa and parts of the genome, would shed light on their roles in restructuring chloroplast genomes.

Studies of chloroplast genomes of conifers may provide a number of insights into the factors controlling chloroplast genome evolution. Conifers are unusual in that they display predominantly paternal inheritance of cpDNA (29–32), whereas chloroplasts are inherited maternally, or, in a number of cases, biparentally, among angiosperms that have been studied (33). Paternal inheritance may contribute in as yet

unknown ways toward the production or establishment of cpDNA rearrangements. An evolutionary mapping survey of conifers might reveal whether loss of the inverted repeat occurred prior to rearrangement of the genome, as in legumes (5). If so, this would add support to the hypothesis that the large inverted repeat stabilizes the genome. The rate of rearrangement may differ widely among genera or families and may be related to the presence of the large inverted repeat and/or short dispersed repeats. Cloning and sequencing of cpDNA fragments containing dispersed repeats would shed light on their origin and perhaps on the causes of rearrangements and apparent insertion/deletion hot spots (31) in chloroplast genomes.

We thank M. Sugiura for providing some of the gene clones used in the study and C.-H. Tsai for technical assistance. This research was supported by a Forestry Opportunity Grant to S.H.S. from the College of Forestry at Oregon State University and grants from the National Institutes of Health to S.H.S. (GM 36367) and J.D.P. (GM 35087).

1. Palmer, J. D. (1985) *Annu. Rev. Genet.* **19,** 325–354.
2. Palmer, J. D. (1985) in *Monographs in Evolutionary Biology: Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 131–240.
3. Crouse, E. J., Schmitt, J. M. & Bohnert, H.-J. (1985) *Plant Mol. Biol. Rep.* **3,** 43–89.
4. Palmer, J. D. & Thompson, W. F. (1982) *Cell* **29,** 537–550.
5. Palmer, J. D., Osorio, B., Aldrich, J. & Thompson, W. F. (1987) *Curr. Genet.* **11,** 275–286.
6. Crouse, E. J., Mubumbila, M., Stummann, B. M., Michalowski, C., Bohnert, H. J., Weil, J.-H. & Henningsen, K. W. (1986) *Plant Mol. Biol.* **7,** 143–149.
7. Palmer, J. D., Nugent, J. M. & Herbon, L. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 769–773.
8. Jansen, R. K. & Palmer, J. D. (1987) *Curr. Genet.* **11,** 553–564.
9. Whitfield, P. R. & Bottomley, W. (1983) *Annu. Rev. Plant Physiol.* **34,** 279–310.
10. Howe, C. J. (1985) *Curr. Genet.* **10,** 139–145.
11. Quigley, F. & Weil, J. H. (1985) *Curr. Genet.* **9,** 495–503.
12. Palmer, J. D. & Stein, D. B. (1986) *Curr. Genet.* **10,** 823–833.
13. Benson, L. (1979) *Plant Classification* (Heath, Lexington, MA).
14. Palmer, J. D. (1986) *Methods Enzymol.* **118,** 167–186.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Sugiura, M., Shinozaki, K., Zaita, N., Kusuda, M. & Kumano, M. (1986) *Plant Sci.* **44,** 211–216.
17. Howe, G., Strauss, S. H., Palmer, J. P. & Doerksen, A. (1988) *Proceedings of the Second Tree Molecular Biology Workshop, International Union of Forestry Research Organizations* (Petawawa Natl. Forestry Inst., Chalk River, Ontario), in press.
18. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5269–5273.
19. Coates, D. & Cullis, C. A. (1987) *Am. J. Bot.* **74,** 260–268.
20. Zurawski, G. & Clegg, M. T. (1987) *Annu. Rev. Plant Physiol.* **38,** 391–418.
21. Price, R. A. & Olson-Stojkovich, J. (1987) *Syst. Bot.* **12,** 91–97.
22. Hermann, R. K. (1985) *The Genus Pseudotsuga: Ancestral History and Past Distribution* (Forest Research Lab., II Oregon State Univ., Corvallis, OR), Spec. Publ. 2b.
23. Bremer, K. (1985) *Cladistics* **1,** 369–385.
24. Doyle, J. A. & Donoghue, M. J. (1986) *Bot. Rev.* **52,** 321-431.
25. Palmer, J. D., Osorio, B., Watson, J. C., Edwards, H., Dodd, J. & Thompson, W. F. (1984) in *Biosynthesis of the Photosynthetic Apparatus: Molecular Biology, Development and Regulation*, eds. Thornber, J. P., Staehelin, L. A. & Hallick, R. B. (Liss, New York), New Series, Vol. 14, pp. 273–283.
26. Day, A. & Ellis, T. H. N. (1984) *Cell* **39,** 359–368.
27. Chattoraj, D., Cordes, K. & Abeles, A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6456–6460.
28. Bowman, C. M. & Dyer, T. A. (1986) *Curr. Genet.* **10,** 931–941.
29. Neale, D. B., Wheeler, N. C. & Allard, R. W. (1986) *Can. J. For. Res.* **16,** 1152–1154.
30. Szmidt, A. E., Alden, T. & Hallgren, J.-E. (1987) *Plant Mol. Biol.* **9,** 59–64.
31. Wagner, D. B., Furnier, G. R., Saghai-Maroof, M. A., Williams, S. M., Dancik, B. P. & Allard, R. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2097–2100.
32. Ohba, K., Iwakawa, M., Okada, Y. & Murai, M. (1971) *Silvae Genet.* **20,** 101–107.
33. Sears, B. B. (1983) *Stadler Genet. Symp.* **15,** 119–139.