Phylogeny and Divergence Times in Pinaceae: Evidence from Three Genomes

Xiao-Quan Wang,* David C. Tank,† and Tao Sang†

*Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China; and
†Department of Botany and Plant Pathology, Michigan State University

In Pinaceae, the chloroplast, mitochondrial, and nuclear genomes are paternally, maternally, and biparentally inherited, respectively. Examining congruence and incongruence of gene phylogenies among the three genomes should provide insights into phylogenetic relationships within the family. Here we studied intergeneric relationships of Pinaceae using sequences of the chloroplast *matK* gene, the mitochondrial *nad5* gene, and the low-copy nuclear gene 4CL. The 4CL gene may exist as a single copy in some species of Pinaceae, but it constitutes a small gene family with two or three members in others. Duplication and deletion of the 4CL gene occurred at a tempo such that paralogous loci are maintained within but not between genera. Exons of the 4CL gene have diverged approximately twice as fast as the *matK* gene and five times more rapidly than the *nad5* gene. The partition-homogeneity test indicates that the three data sets are homogeneous. A combined analysis of the three gene sequences generated a well-resolved and strongly supported phylogeny. The combined phylogeny, which is topologically congruent with the three individual gene trees based on the Templeton test, is likely to represent the organismal phylogeny of Pinaceae. This phylogeny agrees to a certain extent with previous phylogenetic hypotheses based on morphological, anatomical, and immunological data. Disagreement between the previous hypotheses and the three-genome phylogeny suggests that morphology of both vegetative and reproductive organs has undergone convergent evolution within the pine family. The strongly supported monophyly of *Nothotsuga longibracteata, Tsuga mertensiana,* and *Tsuga canadensis* on all three gene phylogenies provides evidence against previous hypotheses of intergeneric hybrid origins of *N. longibracteata* and *T. mertensiana.* Divergence times of the genera were estimated based on sequence divergence of the *matK* gene, and they correspond well with the fossil record.

Introduction

A plant cell has one nuclear and two organellar (chloroplast and mitochondrial) genomes. Genes from the different genomes may have distinct phylogenies as a result of different inheritance pathways and differential responses to processes such as lineage sorting, gene duplication/deletion, lateral gene transfer, and hybrid speciation (Doyle 1997; Maddison 1997; Wendel and Doyle 1998). Conversely, congruent phylogenies among the three genomes could suggest strongly that the gene trees are also congruent with the single underlying phylogeny—the species phylogeny. Therefore, comparison of gene phylogenies of the three genomes will provide an opportunity for robust reconstructions of complex plant phylogenies (e.g., Qiu and Palmer 1999).

Inheritance pathways of the three genomes in the pine family (Pinaceae) are strikingly different; the chloroplast, mitochondrial, and nuclear genomes are paternally, maternally, and biparentally inherited, respectively (Gillham 1994; Hipkins, Krutovskii, and Strauss 1994; Mogensen 1996). Pinaceae, comprising 11 genera and more than 200 species (Farjon 1998), is the largest extant family of gymnosperms. Many species of the pine family constitute the major forest elements in the northern temperate region. Due to morphological convergence within the family, Pinaceae has been a phylogenetically complex group (Hart 1987; Farjon 1990). Phylogenetic relationships of two monotypic genera, *Cathaya* and

Nothotsuga, and *Tsuga mertensiana* (once recognized as a monotypic genus, *Hesperopeuce*), are particularly controversial because each of them shares morphological features with several other genera (Frankis 1988; Page 1988; Lin, Hu, and Wang 1995; Wang, Han, and Hong 1998*a*). *Nothotsuga longibracteata* and *T. mertensiana* were further hypothesized to be intergeneric hybrids based on their morphological intermediacy (Van Campo-Duplan and Gaussen 1948; Gaussen 1966).

Previous molecular phylogenetic studies of intergeneric relationships in Pinaceae were based primarily on the chloroplast genome. The phylogeny generated from *rbcL* gene sequences was poorly resolved and contradicts the phylogeny generated from PCR restriction fragment length polymorphisms of six chloroplast genes (Tsumura et al. 1995; Wang, Han, and Hong 1998*b*). Each previous molecular phylogenetic study involving Pinaceae based on nuclear ribosomal genes sampled only three or four genera (Chaw et al. 1997; Stefanovic et al. 1998; Gernandt and Liston 1999).

In this study, we included all the extant genera of Pinaceae and compared gene phylogenies of the three genomes in order to clarify intergeneric relationships. We chose a rapidly evolving gene, *matK,* for reconstruction of the phylogeny of the chloroplast genome (Johnson and Soltis 1994, 1995; Steele and Vilgalys 1994). For the mitochondrial genome, which has slow rates of nucleotide substitutions (Hiesel, Haeseler, and Brennicke 1994; Laroche et al. 1997), we sequenced an intron of the *nad5* gene encoding subunit 5 of NADH dehydrogenase. The nuclear genome of conifers is large in size and complex in organization, and genes usually exist in large gene families (Perry and Furnier 1996; Kinlaw and Neale 1997; Murray 1998). The questions of how dynamically gene duplication/deletion occurs in co-

Table 1
Collection localities of Species of Pinaceae and *Cycas* Sampled for DNA Sequencing in this Study

Species	Collection Locality
Abies beshanzuensis Wu..............................	Longquan, Zhejiang, China
Abies firma Sieb. et Zucc...........................	Botanic Garden, Institute of Botany, Beijing
Abies holoyphylla Maxim............................	Botanic Garden, Institute of Botany, Beijing
Cathaya argyrophylla Chun et Kuang..................	Huaping, Guangxi, China
Cedrus atlantica Manetti............................	Michigan State University, East Lansing, Mich.
Keteleeria evelyniana Mast..........................	Botanic Garden, Institute of Botany, Kunming
Larix gmelini (Rupr.) Rupr..........................	Botanic Garden, Institute of Botany, Beijing
Nothotsuga longibracteata Hu ex Page................	Xinning, Hunan, China
Picea smithiana (Wall.) Boiss.......................	Botanic Garden, Institute of Botany, Beijing
Pinus armandi Franch...............................	Botanic Garden, Institute of Botany, Beijing
Pinus banksiana Lamb...............................	Botanic Garden, Institute of Botany, Beijing
Psuedolarix amabilis (Nelson) Rehd..................	Botanic Garden, Institute of Botany, Kunming
Pseudotsuga menziesii (Mirbel) Franco...............	Botanic Garden, Institute of Botany, Beijing
Pseudotsuga sinensis Dode..........................	Botanic Garden, Institute of Botany, Kunming
Tsuga canadensis Carr..............................	Michigan State University, East Lansing, Mich.
Tsuga mertensiana (Bong.) Rydb.....................	Mt. Hood, Oreg.
Cycas panzhihuaensis Zhou et Yang	Panzhihua, Sichuan, China

nifers and how this will affect phylogenetic utility of nuclear genes remain open (Kinlaw and Neale 1997). Single- or low-copy nuclear genes have been increasingly used for phylogenetic studies on angiosperms (e.g., Doyle, Kanazin, and Shoemaker 1996; Gottlieb and Ford 1996; Sang, Donoghue, and Zhang 1997; Mason-Gamer, Weil, and Kellogg 1998; Small et al. 1998). In the present study, we chose the low-copy nuclear gene 4CL, encoding 4-coumarate : coenzyme A ligase in the lignin biosynthetic pathway (Zhang and Chiang 1997), for study of gene duplication/deletion and inference of the phylogeny of Pinaceae from the nuclear genome.

In addition to reconstructing phylogenetic relationships, we estimated divergence times among the genera of Pinaceae using a molecular clock. It has been shown for both plants and animals that divergence times calculated from the molecular clock may not be concordant with those based on the fossil record (e.g., Martin, Gierl, and Saedler 1989; Wolfe et al. 1989; Bromham et al. 1998). The abundant fossil record of the pine family (Florin 1963) offers an excellent opportunity for the comparison of these two approaches to determine divergence times.

Materials and Methods

All 11 recognized genera of Pinaceae were sampled, including *Abies* (fir), *Cathaya, Cedrus* (cedar), *Keteleeria, Larix* (larch), *Nothotsuga, Picea* (spruce), *Pinus* (pine), *Pseudolarix* (golden larch), *Pseudotsuga* (Douglas-fir), and *Tsuga* (hemlock). Sampling localities are given in table 1. Voucher specimens have been deposited in the herbaria of the Institute of Botany, Beijing, and Michigan State University. Total DNA was isolated from fresh leaves using the CTAB method (Doyle and Doyle 1987) and purified with a Wizard DNA Clean-up System (Promega).

Genes were amplified through the following PCR cycles: (1) 70°C, 4 min; (2–5) 94°C, 1 min; 48–55°C, 30 s; 72°C, 2 min; (6–36) 94°C, 20 s; 48–55°C, 30 s; 72°C, 2 min; and (37) 72°C, 5 min. Primers for ampli-

fying the *matK* gene are *trnK*-3914F and *trnK*-2R (Johnson and Soltis 1995) with an additional forward primer, *trnK*F1 (5′-TACTGATCAGAAGTTAA-GAGC). For the *nad5* gene, the forward primer *nad5*-aF (5′-GGAAATGTTTGATGCTTCTTGGG) and the reverse primer *nad5*-bR (5′-CTGATCCAAAAT-CACCTACTCG) are located on exons a and b, respectively. For the 4CL gene, the forward primers 4CLpF2 (5′-AGAGTVGCGGAATTCGCAG) and 4CLpF3 (5′-CCAATCCTTTYTACAAGCCG) are located on exon 1, and the reverse primers 4CLpR2 (5′-TTTGAGCGT-TMCGGACGAC) and 4CLpR3 (5′-CGGGGAARGGCT-YCTTTGC) are located on exon 2.

PCR products of *matK* and *nad5* genes were purified using Genclean (Bio 101). PCR products of the nuclear 4CL gene were cloned with a TA cloning kit (Invitrogen). For each species, 10–30 clones were screened by examining restriction site or sequence (from one primer) variation (Sang, Donoghue, and Zhang 1997). Distinct clones were fully sequenced and included in the phylogenetic analyses. Sequencing was done on an ABI 373 automated DNA sequencer using the Dye Terminator Cycle Sequencing reaction kit (PE Applied Biosystems). Sequences have been deposited in GenBank under accession numbers AF143412–AF143425 (*nad5*), AF143427–AF143441 (*matK*), and AF144499–AF144529 (4CL). Additional sequences obtained from GenBank include *matK* genes of *Pinus thunbergii* (D11467) (Tsudzuki et al. 1992), *Pinus contorta* (X57097) (Lidholm and Gustafsson 1991), *Picea glauca* (AF059341), *Picea rubens* (AF059342), and *Picea mariana* (AF059343) and 4CL genes of *Pinus taeda* (U39404 and U39405) (Zhang and Chiang 1997) and *Arabidopsis thaliana* (U18675) (Lee et al. 1995).

Sequence alignments were made with CLUSTAL W (Thompson, Higgins, and Gibson 1994) and refined manually. A few regions in the 4CL intron could not be aligned unambiguously and were excluded from the analyses. Parsimony, as implemented in PAUP*, version 4.0 (Swofford 1998), was used to infer phylogenies based on nucleotide substitutions in aligned sequences.

Unweighted parsimony analyses were performed by heuristic search with tree bisection-reconnection (TBR) branch swapping, the MULPARS option, ACCTRAN optimization, and 1,000 random-addition replicates for the 4CL data set, or by branch-and-bound search with the options of Multree and farthest sequence addition for the *matK, nad5,* and combined data sets. Bootstrap analyses (Felsenstein 1985) were carried out with 1,000 replications of heuristic search with simple taxon addition, while all trees were saved. *Cycas* was chosen as the outgroup for phylogenetic analyses of the *matK* and *nad5* sequences because sequence divergence of the *rbcL* gene is lower between Pinaceae and *Cycas* than between Pinaceae and Podocarpaceae or Araucariaceae (Wang, Han, and Hong 1998*a*). However, we were unable to amplify the 4CL gene from *Cycas. Arabidopsis* was used as the outgroup when only exon sequences of the 4CL gene were analyzed. In the resulting parsimony tree, *Cedrus* formed the sister group to the remaining genera with 78% bootstrap support. The same basal relationship of *Cedrus* was obtained from both *matK* and *nad5* phylogenies when *Cycas* was used as the outgroup (see *Results*). Thus, *Cedrus* was chosen as the functional outgroup for further parsimony analysis of both exon and intron regions of the 4CL gene.

Congruence among the three data sets was examined with the partition-homogeneity test (Farris et al. 1995), implemented in PAUP*, version 4.0. For the purposes of this test, data sets of the three genes were reduced so that they shared the same set of 13 taxa. In the reduced data sets, each genus was represented by a single species, except for *Pinus* and *Tsuga,* of which subgenera were also represented. In the reduction of the 4CL data set, a single clone was chosen randomly to represent a species with multiple distinct 4CL sequences. *Cedrus* was used as the functional outgroup, while *Cycas* was excluded from the *matK* and *nad5* data sets to maintain consistency with the 4CL data set. The tests were performed with 100 replications of heuristic search with TBR branch swapping. Topological congruence between the gene trees was evaluated with the Templeton (1983) test, implemented in PAUP*, version 4.0.

Maximum-likelihood analyses were performed using PAUP*, version 4.0. The program Modeltest, version 2.1 (Posada and Crandall 1998), was utilized to find the model of sequence evolution that best fit each data set by the hierarchical likelihood ratio (LR) test ($\alpha = 0.05$). When the models of sequence evolution are nested, the LR test statistic is distributed as χ^2 with degrees of freedom equal to the number of free parameters between the two models (Goldman 1993). Once the best sequence evolution model was determined (table 2), maximum-likelihood tree searches were performed for each data set. The molecular-clock hypothesis was tested with the LR test by calculating the log likelihood score of the chosen model with the molecular clock enforced and comparing it with the log likelihood score without the molecular clock enforced (Muse and Weir 1992; Baldwin and Sanderson 1998). The number of degrees of freedom is equivalent to the number of terminals minus two (Sorhannus and Van Bell 1999).

Table 2
Sequence Evolution Models Best Fit to Each Data Set as Determined by Hierarchical Likelihood Ratio Tests

Data Set	Models
matK	K3Puf+Γ
nad5	K3P+I+Γ
4CL	K2P+Γ
Three-gene combined	K3Puf+Γ

NOTE.—The models of DNA substitution are the Kimura (1986) two-parameter model (K2P); the Kimura (1981) three-parameter model (K3P), K3P with unequal base frequencies (K3Puf, Kimura 1981). Λ = shape parameter of the gamma distribution (estimated via maximum-likelihood); I = proportion of invariable sites (estimated via maximum-likelihood).

Results

The aligned *matK* sequences were 1,551 bp in length, of which 545 nucleotide sites were variable and 210 were parsimony-informative. Parsimony analysis generated a single most-parsimonious tree with a tree length of 778, a consistency index (CI) of 0.80, and a retention index (RI) of 0.76 (fig. 1*A*). The aligned sequences of the *nad5* gene included 285 bp of exon and 1,042 bp of intron, of which 141 nucleotide sites were variable and 54 were parsimony-informative. Parsimony analysis yielded three equally most parsimonious trees (tree length = 184, CI = 0.80, RI = 0.70). The parsimonious tree that is topologically identical to the maximum-likelihood (ML) tree is shown in figure 1*B*. Although the basal position of *Cedrus* collapsed on the strict consensus of the three parsimonious trees, *Cedrus* is the sister group of the remaining genera of the family on the *nad5* ML tree. This result supports the utility of *Cedrus* as a functional outgroup in analyses of the 4CL and combined data sets.

After screening 10–30 4CL clones for each species, only one type of clone was found for *Cathaya argyrophylla, Cedrus atlantica, T. mertensiana, Keteleeria evelyniana, Picea smithiana,* and *N. longibracteata.* Two types of clones were identified for each of the following species: *Abies holophylla, Larix gmelini, Pinus banksiana, Pseudolarix amabilis,* and *Tsuga canadensis.* Three types of clones were isolated for each of the following species: *Abies firma, Abies beshanzuensis, Pinus armandi, Pseudotsuga menziesii,* and *Pseudotsuga sinensis.* The 4CL data set contained 827 bp of exon and 126 bp of alignable intron sequences, of which 360 sites were variable and 264 were parsimony-informative. Parsimony analysis resulted in six equally most parsimonious trees (tree length = 649, CI = 0.71, RI = 0.86). The parsimonious tree that is topologically identical to the ML tree is shown in figure 1*C*.

When the three data sets were reduced to 13 taxa for the homogeneity tests, the *matK, nad5,* and 4CL data sets contained 131, 47, and 153 parsimony-informative sites, respectively. The partition-homogeneity tests indicated that all pairs of the three data sets are congruent ($P = 0.17$ for *matK-nad5*; $P = 0.20$ for *matK*-4CL; $P = 0.17$ for *nad5*-4CL). Therefore, the three data sets were combined for further phylogenetic analysis. Three equally most parsimonious trees (tree length = 1,110,
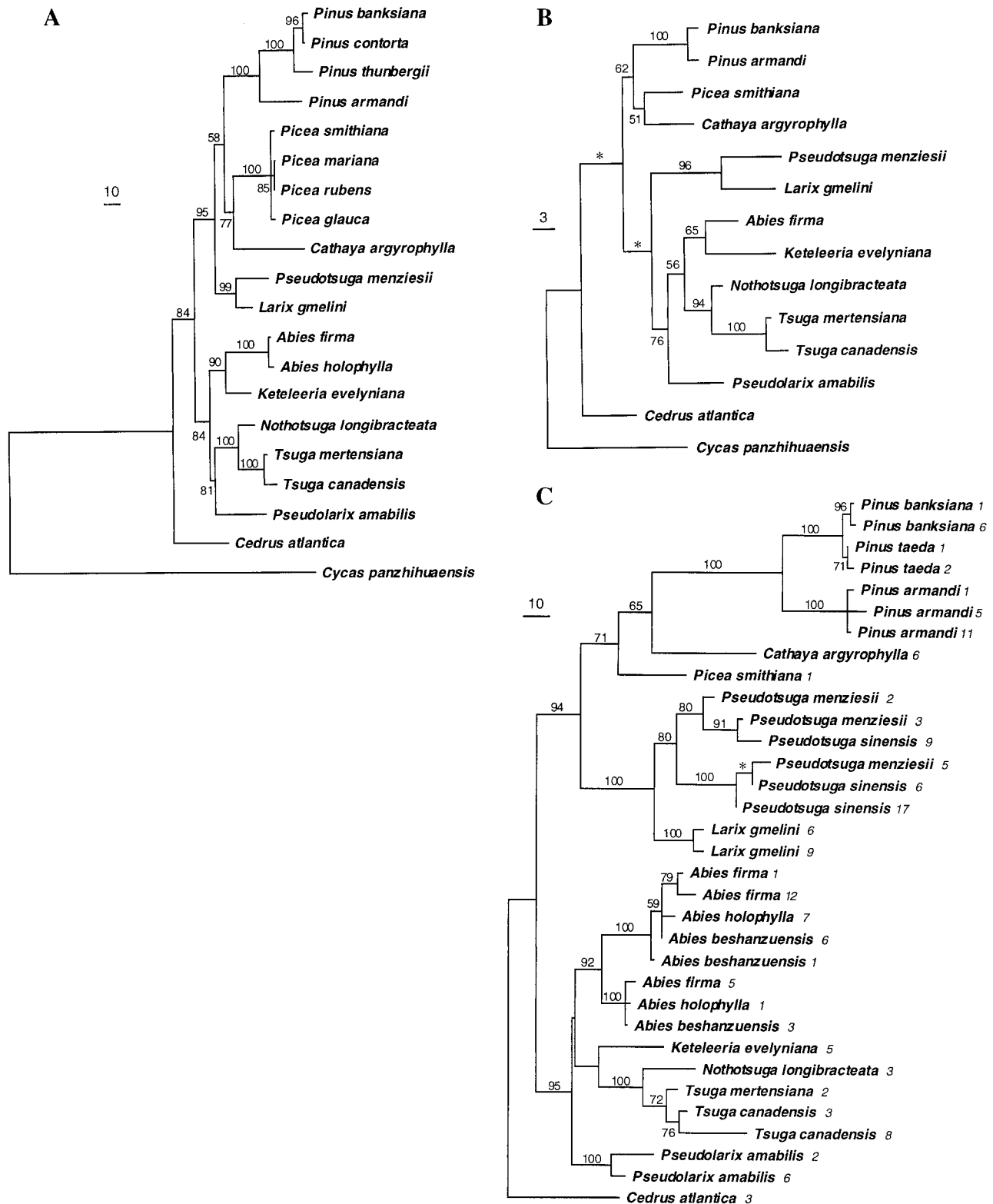
A

```
        96 ┌ Pinus banksiana
    100 ┌──┴─ Pinus contorta
 100 ┤      └── Pinus thunbergii
     │       └── Pinus armandi
  58 ┤    ┌ Picea smithiana
     │ 100 ┤ Picea mariana
     │   85│├ Picea rubens
  95 ┤    77│└ Picea glauca
     │      └── Cathaya argyrophylla
     │   99 ┌── Pseudotsuga menziesii
  84 ┤     └── Larix gmelini
     │    100 ┌ Abies firma
     │   90 ┤└ Abies holophylla
     │     │ └── Keteleeria evelyniana
  84 ┤ 100 ┌── Nothotsuga longibracteata
     │  100│┌ Tsuga mertensiana
  81 ┤    └── Tsuga canadensis
     │   └── Pseudolarix amabilis
     └── Cedrus atlantica
          └────────── Cycas panzhihuaensis
```

10

B

```
       100 ┌ Pinus banksiana
    62 ┤  └ Pinus armandi
   * ┤  ┤ Picea smithiana
   │51 └ Cathaya argyrophylla
   │    96 ┌── Pseudotsuga menziesii
   │      └── Larix gmelini
   * ┤  65 ┌── Abies firma
   │ 56┤ └── Keteleeria evelyniana
   │ 94┤ ┌ Nothotsuga longibracteata
   │76 100┤┌ Tsuga mertensiana
   │    └ Tsuga canadensis
   │    └── Pseudolarix amabilis
   └── Cedrus atlantica
        └──────── Cycas panzhihuaensis
```

3

C

```
                         96 ┌ Pinus banksiana 1
                    100 ┤└ Pinus banksiana 6
                        │┌ Pinus taeda 1
               100 ┤  71└ Pinus taeda 2
                   │     100 ┌ Pinus armandi 1
           65 ┤       ┤─ Pinus armandi 5
        71 ┤  │       └ Pinus armandi 11
           │  └── Cathaya argyrophylla 6
           │    └ Picea smithiana 1
  94 ┤   80 ┌ Pseudotsuga menziesii 2
     │ 80 ┤91┤ Pseudotsuga menziesii 3
     │   │  └── Pseudotsuga sinensis 9
     │   │  * ┌ Pseudotsuga menziesii 5
     │ 100┤100┤ Pseudotsuga sinensis 6
     │   │   └ Pseudotsuga sinensis 17
     │ 100 ┌ Larix gmelini 6
     │    └ Larix gmelini 9
     │   79 ┌ Abies firma 1
     │  59 ┤└ Abies firma 12
     │ 100 ┤ Abies holophylla 7
     │92 ┤ Abies beshanzuensis 6
     │   │└ Abies beshanzuensis 1
     │   100┌ Abies firma 5
     │    ┤ Abies holophylla 1
     │    └ Abies beshanzuensis 3
  95 ┤  ┌── Keteleeria evelyniana 5
     │ 100┤ ┌ Nothotsuga longibracteata 3
     │   72┤ Tsuga mertensiana 2
     │    ┤ Tsuga canadensis 3
     │  76└ Tsuga canadensis 8
     │ 100 ┌ Pseudolarix amabilis 2
     │    └ Pseudolarix amabilis 6
     └── Cedrus atlantica 3
```

10

FIG. 1.—Phylogenies of chloroplast *matK,* mitochondrial *nad5,* and nuclear 4CL genes of Pinaceae. *A, matK* gene phylogeny. The single most parsimonious tree (tree length = 778, consistency index [CI] = 0.80, retention index [RI] = 0.76). *B, nad5* gene phylogeny. One of three equally most parsimonious trees (tree length = 184, CI = 0.80, RI = 0.70). *C,* 4CL gene phylogeny. One of six equally most parsimonious trees (tree length = 649, CI = 0.71, RI = 0.86). Small numbers following a species name indicate clone numbers. Numbers associated with branches are bootstrap percentages greater than 50%. When multiple parsimonious trees are found from the *nad5* and 4CL data sets, the one with the same topology as the maximum-likelihood tree is shown. * = branch collapses on the strict consensus. Branch lengths are proportional to the numbers of nucleotide substitutions and are measured by scale bars.
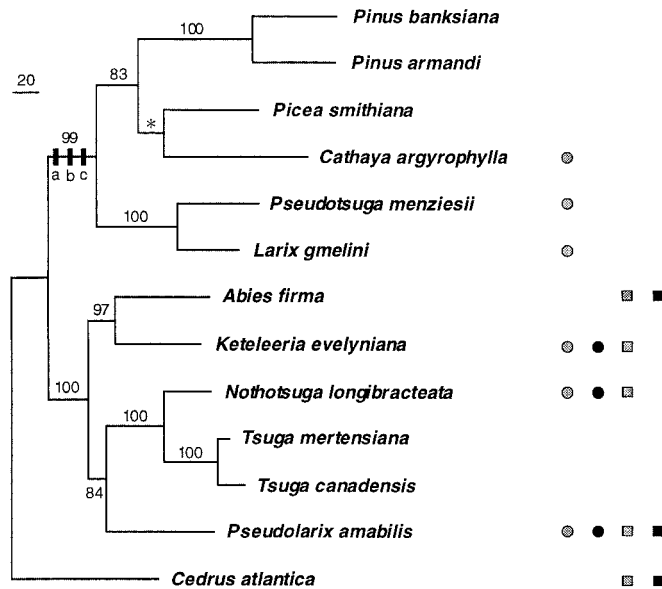
Fig. 2.—Phylogeny of Pinaceae based on combined sequences of three genes, *matK, nad5,* and *4CL*; one of three equally most parsimonious trees (tree length = 1,110, CI = 0.78, RI = 0.68) which is topologically identical to the maximum-likelihood tree. * = branch collapses on the strict consensus. Numbers associated with branches are bootstrap percentages greater than 50%. Branch lengths are proportional to the numbers of nucleotide substitutions and are measured by the scale bar. Synapomorphies supporting a branch are indicated by black bars: (*a*) absence of resin vesicles in seed coat, (*b*) absence of narrowed, pedicellate base of seed scales, and (*c*) presence of two resin canals in vascular cylinder of young taproot. Four morphological characters that may have undergone parallel changes are labeled next to the species names: gray circle, cones on leaved peduncles; black circle, male strobili in clusters from a single bud; gray square, erect position of mature cones; black square, seed scale abscission.

CI = 0.78, RI = 0.68) were obtained from the combined data set. The parsimonious tree that is topologically identical to the ML tree is shown in figure 2. For each gene, the average sequence divergence among these 13 taxa was estimated with the Jukes-Cantor model (Jukes and Cantor 1969) as 10.93% for the 4CL exons, 5.62% for the *matK* gene, and 2.21% and 2.46% for the exon and intron of the *nad5* gene, respectively.

The *matK* phylogeny is topologically congruent with the tree resulting from the combined analysis (figs. 1*A* and 2). Topological incongruence between the *nad5* and the combined tree, which is supported by bootstrap values higher than 50% on both trees, involves only the position of *Pseudolarix* (figs. 1*B* and 2). The Templeton test was performed on the *nad5* data set, while the topology of the combined tree was used as a constraint. The analysis with the constraint did not lead to an increase in tree length, and thus the topological incongruence was not significant. While topological incongruence between the 4CL and the combined trees involves the positions of *Cathaya, Keteleeria,* and *Pseudolarix* (figs. 1*C* and 2), bootstrap support is found only for *Cathaya* on the 4CL tree. Using the topology of the combined tree as a constraint (fig. 2), the Templeton test indicated that the incongruence was not significant (Ts = 35.0, N = 9, P = 0.10).

Results of the LR test of the molecular-clock hypothesis for the reduced data sets (each containing 13 taxa) of the three genes are as follows: *matK,* -2 ln LR = 24.64, df = 11, 0.025 > P >0.01; *nad5,* -2 ln LR = 22.56, df = 11, 0.025 > P > 0.01; and 4CL, -2 ln LR = 39.50, df = 11, P < 0.001. Because the molecular

clock of the *matK* and *nad5* genes cannot be rejected at the significance level of P = 0.01, sequence divergence of these two genes may be useful in estimating divergence times. However, when the molecular clock was enforced, ML analyses of the *matK* and *nad5* data sets yielded trees (not shown) with topologies different from the parsimonious trees. On the *matK* ML tree with molecular clock (the ML-MC tree), *Cedrus* formed a sister group with the clade containing *Abies, Keteleeria, Nothotsuga, Tsuga,* and *Pseudolarix.* On the *nad5* ML-MC tree, *Cedrus* formed a sister group with the clade containing *Pinus, Picea, Cathaya, Pseudotsuga,* and *Larix,* and the clade of *Larix* and *Pseudotsuga* became the sister group of the remaining genera of the family.

By excluding *Cedrus* from the *matK* data set and rooting the tree between the next two major basal clades of the three-gene phylogeny (fig. 2), the resulting ML-MC tree (fig. 3) has the same topology as the *matK* (fig. 1*A*) and three-gene phylogenies. The molecular clock for the remaining sequences could not be rejected at P = 0.025 (-2 ln LR = 19.78, df = 10, 0.05 > P > 0.025). Because *Cedrus* has the shortest branch length on the *matK* gene tree (fig. 1*A*), the slow divergence rate of the *matK* sequences of *Cedrus* may have contributed in part to the rate heterogeneity of the *matK* data set. For the *nad5* data set, although the molecular clock could not be rejected after excluding *Cedrus,* the clade of *Larix* and *Pseudotsuga* remained as the sister group of the rest of the genera (tree not shown). Therefore, only *matK* sequences were used to estimate divergence times of all genera except *Cedrus.* Branch lengths were estimated by ML with the molecular clock enforced (fig. 3).
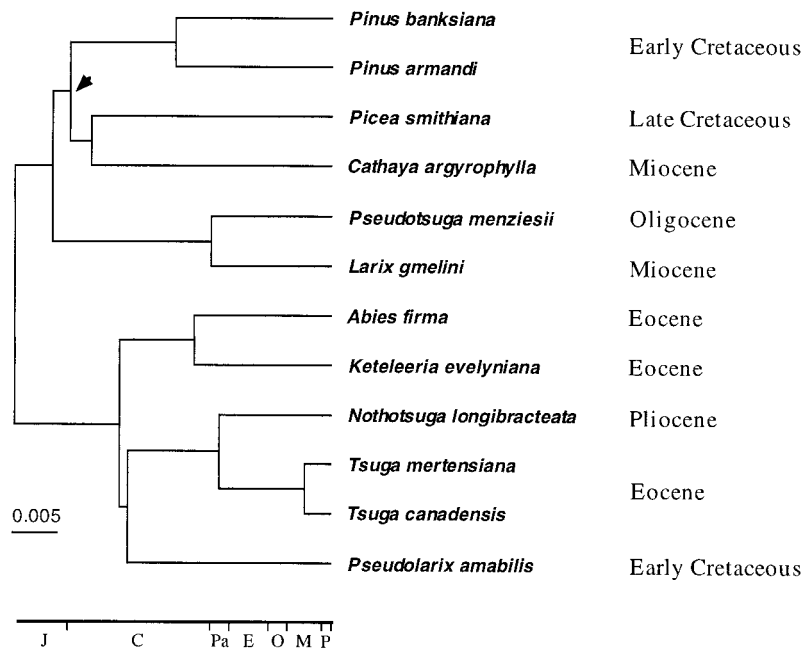
Fig. 3.—Maximum-likelihood tree of Pinaceae based on *matK* sequences with a molecular clock enforced. The earliest fossil record of each genus (not necessarily the species sampled in this study) is indicated. Branch lengths are proportional to sequence divergence estimated by maximum-likelihood and are measured by the scale bar. The geological timescale was calculated from the branch lengths according to the molecular clock: J, Jurassic; C, Cretaceous; Pa, Paleocene; E, Eocene; O, Oligocene; M, Miocene; P, Pliocene. The arrow indicates the point at which the molecular clock is calibrated (140 MYA).

Discussion
Evolution of 4CL Gene

A better understanding of the dynamics of gene duplication/deletion will provide insights into evolution of the nuclear genome, as well as the phylogenetic utility of low-copy nuclear genes (Morton, Gaut, and Clegg 1996; Clegg, Cummings, and Durbin 1997). Two 4CL loci were previously found in *P. taeda* (Zhang and Chiang 1997), and their sequences formed a monophyletic group on the 4CL phylogeny (fig. 1*C*). This study identified as many as three distinct clones from individuals of some species. Observed sequence divergence between clones isolated from the same individual ranged from 3 bp (between *P. armandi* 1 and *P. armandi* 11) to 57 bp (between *P. sinensis* 9 and *P. sinensis* 17). Distinct clones isolated from an individual plant may represent different loci or allelic variation. Given that the two 4CL loci isolated previously from *P. taeda* differ by only 2 bp in the partial sequence analyzed here, different sequences cloned from a species in this study, which differ by at least 3 bp, could also represent different 4CL loci. Although only one type of 4CL sequence has been found in a number of species, it is still possible that additional loci in some of the species remain unidentified due to PCR selection (Wagner et al. 1994). Apparently, the 4CL gene of Pinaceae may exist as a single copy in some species, but it constitutes a small gene family with two or three members in others.

It is remarkable that all 4CL clones from each genus form a strongly supported monophyletic group (fig. 1*C*). Sequences cloned from a species, however, do not necessarily form a monophyletic group. Notably, two or three types of sequences were cloned from each of the three *Abies* species. They group into two strongly supported clades, which may represent a gene duplication prior to the diversification of the three *Abies* species. A similar pattern was also found for the two *Pseudotsuga* species. These results indicate that the 4CL gene has a tempo of duplication/deletion cycles such that paralogous loci are maintained between species but not between genera. Therefore, this gene can serve as an efficient phylogenetic marker for studying relationships at or above the intergeneric level. However, caution must be exercised in distinguishing paralogy and orthology when the 4CL gene is used for phylogenetic studies at the interspecific level.

Of the three genes, the nuclear 4CL gene evolved most rapidly. The average sequence divergence of the exon region of the 4CL gene is approximately twice as high as that of the chloroplast *matK* gene and five times as high as that of the mitochondrial *nad5* gene. Because the nuclear ribosomal DNA internal transcribed spacers exhibit a high level of length variation and exist in multiple diverged copies in Pinaceae (Liston et al. 1996; Gernandt and Liston 1999), low-copy nuclear genes will provide useful alternative markers for phylogenetic reconstructions at the intergeneric and interspecific levels. The *matK* gene diverged about twice as fast as the *rbcL* gene (Wang, Han, and Hong 1998*a*) in Pinaceae, which is similar to the rate differences between these two chloroplast genes in angiosperms (Johnson and Soltis 1994, 1995; Steele and Vilgalys 1994). The higher evolutionary rate of the *matK* gene may be responsible for the better resolution and support in the *matK* phylogeny

than in the *rbcL* phylogeny (Wang, Han, and Hong 1998*a*) of Pinaceae. Sequences of the mitochondrial *nad5* gene, including a large intron, have diverged most slowly, concordant with previous observations of relative sequence divergence rates among the three plant genomes (Palmer 1992). Nevertheless, the *nad5* gene tree has offered reasonable resolution among genera of Pinaceae (fig. 1*B*).

Phylogeny of Pinaceae and Evolution of Morphological Characters

Despite the different inheritance pathways of the three genomes, the *matK, nad5,* and 4CL data sets are congruent based on the homogeneity test. Although the three gene trees are not identical in topology, incongruence among them is not supported by high bootstrap values, and the three gene trees are topologically congruent with the combined tree based on the Templeton test. Therefore, the tree resulting from the combined analysis is very likely to represent the true intergeneric relationships of Pinaceae. Furthermore, the strongly supported monophyly of *N. longibracteata, T. mertensiana,* and *T. canadensis* on all three gene phylogenies provides evidence against previous hypotheses of the hybrid origins of *N. longibracteata* (between *Tsuga* and *Keteleeria*) and *T. mertensiana* (between *Tsuga* and *Picea*) (Van Campo-Duplan and Gaussen 1948; Gaussen 1966). If these were intergeneric hybrids, they would likely have significantly incongruent positions between the chloroplast (paternal) and mitochondrial (maternal) gene phylogenies (Sang, Crawford, and Stuessy 1997; Wendel and Doyle 1998).

The three-genome phylogeny is similar to the phenogram generated from immunological data (Price, Olsen-Stojkovich, and Lowenstein 1987) except for the position of *Cedrus.* The immunological phenogram, which did not sample *Cathaya* or *Nothotsuga,* placed *Cedrus* and *Abies* as sister genera. In contrast, a sister group relationship between *Cedrus* and the rest of the family is revealed here by the *matK* phylogeny (with 84% bootstrap support), the 4CL exon sequences (with 78% bootstrap support with *Arabidopsis* as the outgroup), and the ML tree of the *nad5* gene.

In comparison with the commonly accepted classification systems of Pinaceae, the three-genome phylogeny largely agrees with the classification based on the number and position of resin canals in the central vascular cylinder of the young taproot, which divided the family into two major groups: Cédrées, containing *Abies, Cedrus, Keteleeria, Pseudolarix,* and *Tsuga,* and Pinées, containing *Larix, Picea, Pinus,* and *Pseudotsuga* (Van Tieghem 1891). The three-genome phylogeny, however, differs markedly from the conventional classification of Pinaceae, which recognizes three subfamilies: Pinoideae (*Pinus*), Laricoideae (*Cedrus, Larix,* and *Pseudolarix*), and Abietoideae, consisting of the remaining genera (Melchior and Werdermann 1954). Our results support the previous speculation that shoot and foliage morphology, on which the classification is based,

has undergone considerable convergent evolution within Pinaceae (Frankis 1988).

The three-genome phylogeny agrees to a certain extent with the phylogenetic hypothesis based on combined evidence from morphology of both vegetative and reproductive organs, wood and root anatomy, and immunological data (Farjon 1990). By labeling the characters that Farjon (1990) used to define major groups on the three-genome phylogeny, both synapomorphies and parallelisms are illustrated (fig. 2). Synapomorphies, which support the clade of *Cathaya, Larix, Picea, Pinus,* and *Pseudotsuga,* include absence of resin vesicles in the seed coat; absence of a narrowed, pedicellate base of seed scales; and presence of two resin canals in the vascular cylinder of the young taproot. In contrast, assuming homology of the morphological feature "cones on leaved peduncles" leads to grouping *Cathaya* with *Larix* and *Pseudotsuga.* This character, together with "male strobili in clusters from a single bud," grouped *Keteleeria, Pseudolarix,* and *Nothotsuga* together. Two characters, "seed scale abscission" and "erect position of mature cones," were mainly responsible for grouping *Abies* and *Cedrus.* These results suggest that the morphology of reproductive organs may have also undergone convergent evolution.

Divergence Times

When the molecular clock was used to estimate divergence times in Pinaceae, *Cedrus* was excluded because its *matK* gene appeared to have diverged at a slower rate. Even when *Cedrus* was excluded, there still existed a certain degree of rate heterogeneity among the remaining *matK* sequences ($0.05 > P > 0.025$). When *Keteleeria,* which has the second shortest branch on the *matK* gene tree (fig. 1*A*), was also excluded, the LR test could not reject the molecular clock ($-2 \ln LR = 15.69$, df = 9, $P > 0.05$). Exclusion of *Keteleeria,* however, had little impact on the estimated divergence times of the remaining genera and did not alter the estimated divergence times at the broad geological timescale where the molecular clock and fossil record were compared. Therefore, the data set that still contains *Keteleeria* was used for estimating divergence times and for further comparison with the fossil record.

Pinaceae has one of the most extensive fossil records of extant plant families. Among genera of Pinaceae, *Pinus* has the the best fossil record, dating back to the early Cretaceous (Miller 1977; Florin 1963). Thus, we calibrated the molecular clock by using 140 MYA as the time when *Pinus* diverged from the other genera (Savard et al. 1994). The geological timescale is estimated accordingly along the branch length of the tree (fig. 3). The earliest fossil records for the genera (Miller 1977, 1998; Florin 1963; Farjon 1990; LePage and Basinger 1995*a,* 1995*b*) are labeled on the *matK* ML-MC tree (fig. 3).

Remarkably, divergence times estimated from the molecular clock correspond well with the fossil record for the majority of the genera. Of four genera, *Pinus, Picea, Cathaya,* and *Pseudolarix,* which became estab-

lished in the early and middle Cretaceous according to the ML-MC tree, three have fossil records from the Cretaceous. Only *Cathaya,* currently endemic to China, has a much more recent fossil record, first documented in the Miocene. Although the divergence time of *Cedrus* could not be estimated directly, its basal position in Pinaceae revealed by the three genes is concordant with its fossil record from the early Cretaceous (Arnold 1953). The ML-MC tree suggests that the next period of major diversification within the pine family was around the Paleocene. This corresponds well with the earliest fossil records of *Abies, Keteleeria, Larix,* and *Tsuga* from the Eocene and that of *Pseudotsuga* from the Oligocene. *Nothotsuga,* however, has a rather recent fossil record, dating back only to the Pliocene. The lack of early fossil records of the monotypic genera *Nothotsuga* and *Cathaya* may be due to their limited historical distributions and/or less extensive studies of fossils at these sites.

Acknowledgments

LITERATURE CITED

ARNOLD, C. A. 1953. Silicified plant remains from the Mesozoic and Tertiary of North America. II. Some fossils from northern Alaska. Mich. Acad. Sci. Lett. **38**:9–12.

BALDWIN, B. G., and M. J. SANDERSON. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae) Proc. Natl. Acad. Sci. USA **95**:9402–9406.

BROMHAM, L., A. RAMBAUT, R. FORTEY, A. COOPER, and D. PENNY. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. Proc. Natl. Acad. Sci. USA **95**:12386–12389.

CHAW, S.-M., A. ZHARKIKH, H.-M. SUNG, T.-C. LAU, and W.-H. LI. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol. Biol. Evol. **14**:56–68.

CLEGG, M. T., M. P. CUMMINGS, and M. L. DURBIN. 1997. The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA **94**:7791–7798.

DOYLE, J. J. 1997. Trees within trees: genes and species, molecules and morphology. Syst. Biol. **46**:537–553.

DOYLE, J. J., and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19**:11–15.

DOYLE, J. J., V. KANAZIN, and R. C. SHOEMAKER. 1996. Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine*: Leguminosae). Mol. Phylogenet. Evol. **6**:438–447.

FARJON, A. 1990. Pinaceae. Koleltz Scientific Books, Konigstein, Germany.

———. 1998. World checklist and bibliography of conifers. Royal Botanic Gardens, Kew, England.

FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. Cladistics **10**:315–319.

FELSENSTEIN, J. 1985. Confidence limits on phylogenetics: an approach using the bootstrap. Evolution **39**:783–791.

FLORIN, R. 1963. The distribution of conifer and taxad genera in time and space. Acta Hort. Berg. **20**:121–312.

FRANKIS, M. P. 1988. Generic inter-relationships in Pinaceae. Notes RBG Edinb. **45**:527–548.

GAUSSEN, H. 1966. Les Gymnosperms actuelles et fossils. Trav. Lab. For. Toulouse Tome **2**:481–715.

GERNANDT, D. S., and A. LISTON. 1999. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). Am. J. Bot. **86**:711–723.

GILLHAM, N. W. 1994. Organelle genes and genomes: transmission and compatibility of organelle genomes. Oxford University Press, New York.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:182–198.

GOTTLIEB, L. D., and V. S. FORD. 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC*. Syst. Bot. **21**:45–62.

HART, J. A. 1987. A cladistic analysis of conifers: preliminary results. J. Arn. Arb. **68**:269–307.

HIESEL, R., A. V. HAESELER, and A. BRENNICKE. 1994. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. Proc. Natl. Acad. Sci. USA **91**:634–638.

HIPKINS, V. D., K. V. KRUTOVSKII, and S. H. STRAUSS. 1994. Organelle genomes in conifers: structure, evolution, and diversity. For. Genet. **1**:179–189.

JOHNSON, L. A., and D. E. SOLTIS. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae sensu stricto. Syst. Bot. **19**:143–156.

———. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. Ann. Mo. Bot. Gard. **82**:149–175.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

———. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78**:454–458.

KINLAW, C. S., and D. B. NEALE. 1997. Complex gene families in pine genomes. Trends Plant Sci. **2**:356–359.

LAROCHE, J., P. LI, L. MAGGIA, and J. BOUSQUET. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. Proc. Natl. Acad. Sci. USA **94**:5722–5727.

LEE, D., M. ELLARD, L. A. WANNER, K. R. DAVIS, and C. J. DOUGLAS. 1995. The *Arabidopsis thaliana* 4-coumarate : CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. Plant Mol. Biol. **28**:871–884.

LEPAGE, B. A., and J. F. BASINGER. 1995*a*. The evolutionary history of the genus *Larix* (Pinaceae). USDA For. Serv. Int. Res. Sta. GTR-INT **319**:19–29.

———. 1995*b*. Evolutionary history of the genus *Pseudolarix* Gordon (Pinaceae). Int. J. Plant Sci. **156**:910–950.

LIDHOLM, J., and P. GUSTAFSSON. 1991. A three-step model for the rearrangement of the chloroplast *trnK-psbA* region of the gymnosperm *Pinus contorta*. Nucleic Acids Res. **19**:2881–2887.

LIN, J.-X., Y.-S. HU, and F. H. WANG. 1995. Wood and bark anatomy of *Nothotsuga* (Pinaceae). Ann. Mo. Bot. Gard. **82**:603–609.

LISTON, A., W. A. ROBINSON, J. M. OLIPHANT, and E. R. ALVAREZ-BUYLLA. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of nonflowering seed plants. Syst. Bot. **21**:109–121.

MADDISON, W. P. 1997. Gene trees in species trees. Syst. Biol. **46**:523–536.

MARTIN, W., A. GIERL, and H. SAEDLER. 1989. Molecular evidence for pre-Cretaceous angiosperm origins. Nature **339**: 46–48.

MASON-GAMER, R. J., C. F. WEIL, and E. A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. **15**:1658–1673.

MELCHIOR, H., and E. WERDERMANN. 1954. Engler, Syllabus der Pflanzenfamilien. 12th edition. Berlin.

MILLER, C. N. 1977. Mesozoic conifers. Bot. Rev. **43**:217–281.

———. 1988. The origin of modern conifer families. Pp.448–486 *in* C. B. BECK, ed. Origin and evolution of gymnosperms. Columbia University Press, New York.

MOGENSEN, H. L. 1996. The hows and ways of cytoplasmic inheritance in seed plants. Am. J. Bot. **83**:383–404.

MORTON, B. R., B. S. GAUT, and M. T. CLEGG. 1996. Evolution of alcohol dehydrogenase genes in the Palm and Grass families. Proc. Natl. Acad. Sci. USA **93**:11735–11739.

MURRAY, B. G. 1998. Nuclear DNA amounts in gymnosperms. Ann. Bot. **82**(Suppl. A):3–15.

MUSE, S. V., and B. S. WEIR. 1992. Testing for equality of evolutionary rates. Genetics **132**:269–276.

PAGE, C. N. 1988. New and maintained genera in the conifer families Podocarpaceae and Pinaceae. Notes RBG Edinb. **45**:377–395.

PALMER, J. D. 1992. Mitochondrial DNA in plant systematics: applications and limitations. Pp. 36–49 *in* P. S. SOLTIS, D. E. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants. Chapman Hall, New York.

PERRY, D. J., and G. R. FURNIER. 1996. *Pinus banksiana* has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc. Natl. Acad. Sci. USA **93**:13020–13023.

POSADA, D., and K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

PRICE, R. A., J. OLSEN-STOJKOVICH, and J. M. LOWENSTEIN. 1987. Relationships among the genera of Pinaceae: an immunological comparison. Syst. Bot. **12**:91–97.

QIU, Y.-L., and J. D. PALMER. 1999. Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci. **4**:26–30.

SANG, T., D. J. CRAWFORD, and T. F. STUESSY. 1997. Chloroplast phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. **84**:1120–1136.

SANG, T., M. J. DONOGHUE, and D. ZHANG. 1997. Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. Mol. Biol. Evol. **14**:994–1007.

SAVARD, L., P. LI, S. H. STRAUSS, M. W. CHASE, M. MICHAUD, and J. BOUSQUET. 1994. Chloroplast and nuclear gene sequences indicated late Pennsylvanian time for the last common ancestor of extant seed plants. Proc. Natl. Acad. Sci. USA **91**:5163–5167.

SMALL, R. L., J. A. RYBURN, R. C. CRONN, T. SEELANAN, and J. F. WENDEL. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. **85**:1301–1315.

SORHANNUS, U., and C. VAN BELL. 1999. Testing for equality of molecular evolutionary rates: a comparison between a relative-rate test and a likelihood ratio test. Mol. Biol. Evol. **16**:848–855.

STEELE, K. P., and R. VILGALYS. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. Syst. Bot. **19**:126–142.

STEFANOVIC, S., M. JAGER, J. DEUTSCH, J. BROUTIN, and M. MASSELOT. 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. Am. J. Bot. **85**:688–697.

SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endoclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution **37**:221–244.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

TSUDZUKI, J., K. NAKASHIMA, T. TSUDZUKI, J. HIRATSUKA, M. SHIBATA, T. WAKASUGI, and M. SUGIURA. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ, trnK, psbA, trnI* and *trnH* and the absence of *rps16*. Mol. Gen. Genet. **232**:206–214.

TSUMURA, Y., K. YOSHIMURA, N. TOMARU, and K. OHBA. 1995. Molecular phylogeny of conifers using RFLP analysis of PCR-amplified specific chloroplast genes. Theor. Appl. Genet. **91**:1222–1236.

VAN CAMPO-DUPLAN, M., and H. GAUSSEN. 1948. Sur quatre hybrides de genres chez les Abietinees. Trav. Lab. For. Toulouse Tome **24**:1–14.

VAN TIEGHEM, P. 1891. Structure et affinites des Abies et des genres les plus voisins. Bull. Soc. Bot. Fr. **38**:406–415.

WAGNER A., N. BLACKSTONE, P. CARTWRIGHT, M. DICK, B. MISOF, P. SNOW, G. P. WAGNER, J. BARTELS, M. MURTHA, and J. PENDLETON. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Syst. Biol. **43**:250–261.

WANG, X.-Q., Y. HAN, and D. Y. HONG. 1998*a*. A molecular systematic study of *Cathaya,* a relic genus of the Pinaceae in China. Plant Syst. Evol. **213**:165–172.

———. 1998*b*. PCR-RFLP analysis of the chloroplast gene *trnK* in the Pinaceae, with special reference to the systematic position of *Cathaya*. Isr. J. Plant Sci. **46**:265–271.

WENDEL, J. F., and J. J. DOYLE. 1998. Phylogenetic incongruence: window into genomes history and molecular evolution. Pp. 265–296 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants, II: DNA sequencing. Kluwer, Boston.

WOLFE, K. H., M. GOUY, Y-W. YANG, P. M. SHARP, and W.-H. LI. 1989. Date of the monocot dicot divergence estimated from chloroplast DNA-sequence data. Proc. Natl. Acad. Sci. USA **86**:6201–6205.

ZHANG, X.-H., and V. L. CHIANG. 1997. Molecular cloning of 4-coumarate : coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. **113**:65–74.