

PHYLOGENY OF SEED PLANTS BASED ON EVIDENCE FROM EIGHT GENES¹

DOUGLAS E. SOLTIS,^{2,5} PAMELA S. SOLTIS,³ AND MICHAEL J. ZANIS⁴

²Department of Botany and the Genetics Institute, University of Florida, Gainesville, Florida 32611-5826 USA;

³Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida 32611-7800 USA; and

⁴School of Biological Sciences, Washington State University, Pullman, Washington 99164-4236 USA

Relationships among the five groups of extant seed plants (cycads, *Ginkgo*, conifers, Gnetales, and angiosperms) remain uncertain. To explore relationships among groups of extant seed plants further and to attempt to explain the conflict among molecular data sets, we assembled a data set of four plastid (cpDNA) genes (*rbcL*, *atpB*, *psaA*, and *psbB*), three mitochondrial (mtDNA) genes (*mtSSU*, *coxI*, and *atpA*), and one nuclear gene (18S rDNA) for 19 exemplars representing the five groups of living seed plants. Analyses of the combined eight-gene data set (15 772 base pairs/taxon) with maximum parsimony (MP), maximum likelihood (ML), and Bayesian approaches reveal a gymnosperm clade that is sister to angiosperms. Within the gymnosperms, a conifer clade includes Gnetales as sister to Pinaceae. Cycads and *Ginkgo* are either successive sisters to this conifer clade (including Gnetales) or a clade that is sister to conifers and Gnetales. All analyses of the mtDNA partition and ML analyses of the nuclear partition yield very similar topologies. However, MP analyses of the combined cpDNA genes place Gnetales as sister to all other seed plants with strong bootstrap support, whereas ML and Bayesian analyses of the cpDNA data set place Gnetales as sister to Pinaceae. Maximum parsimony and ML analyses of first and second codon positions of the cpDNA partition also place Gnetales as sister to Pinaceae. In contrast, MP analyses of third codon positions place Gnetales as sister to other seed plants, although ML analyses of third codon positions place Gnetales with Pinaceae. Thus, most of the discrepancies in seed plant topologies involve third codon positions of cpDNA genes. The likelihood ratio (LR) and Shimodaira-Hasegawa (SH) tests were applied to the cpDNA data. The preferred topology based on the LR test is that Gnetales are sister to *Pseudotsuga*. The SH test based on first and second codon and all three codon positions indicated that there is no significant difference between the best topology (Gnetales sister to *Pseudotsuga*) and Gnetales sister to a conifer clade. However, there is a significant difference between the best topology and topologies in which Gnetales are sister to the rest of the seed plants or Gnetales sister to angiosperms.

Key words: Bayesian inference; hypothesis tests; phylogeny; seed plants.

Despite repeated efforts to resolve phylogenetic relationships among extant seed plants—angiosperms and four lineages of gymnosperms, cycads, *Ginkgo*, conifers, and Gnetales—using both morphological and molecular data sets (reviewed in Donoghue and Doyle, 2000; Sanderson et al., 2000), uncertainty remains. Analyses of morphological data generally concur in suggesting that angiosperms and Gnetales are sister groups (the “anthophyte” hypothesis), with extant gymnosperms paraphyletic (Crane, 1985; Doyle and Donoghue, 1986, 1992; Loconte and Stevenson, 1990; Rothwell and Serbet, 1994; Doyle, 1996). However, the sister-group relationship of Gnetales and angiosperms has not been supported by most molecular analyses. In fact, early molecular studies based on single genes provided conflicting results regarding the relationship of Gnetales and angiosperms (see Doyle, 1998, for review). For example, some analyses of *rbcL* alone and analyses of partial 18S and 26S rRNA sequences placed Gnetales as sister to all other seed plants; angiosperms were sister to a clade of cycads, *Ginkgo*, and conifers (e.g., Hamby and Zimmer, 1992; Albert et al., 1994). In contrast, a maximum likelihood (ML) analysis of *rbcL* placed angiosperms as the sister to a gymnosperm clade (Hasebe et al., 1992).

Single-gene, multigene, and multigenome analyses provide conflicting reconstructions of seed plant phylogeny. Analyses

of 18S rDNA alone suggested that the gymnosperms constitute a clade (Chaw et al., 1997; Soltis et al., 1999), as have plastid (Goremykin et al., 1996) and mitochondrial (Malek et al., 1996) sequences, although measures of support are generally weak, where provided. Recent multigene analyses have found strong support for the monophyly of extant gymnosperms (80% bootstrap value in Chaw et al., 2000; >90% bootstrap values in Bowe, Coat, and dePamphilis, 2000). In addition, both Bowe, Coat, and dePamphilis (2000) and Chaw et al. (2000) obtained strong support for cycads followed by *Ginkgo* as successive sisters to a conifer clade that includes Gnetales.

Analyses of both a four-gene data set (*rbcL*, 18S rDNA, and two slowly evolving mitochondrial genes, *atpA* and *coxI*; Bowe, Coat, and dePamphilis, 2000) and a three-gene data set (mt SSU rDNA, 18S rDNA, and *rbcL*; Chaw et al., 2000) provided strong support for the sister-group relationship of Pinaceae and Gnetales (the “gne-pine” hypothesis of Chaw et al., 2000). The possible derivation of Gnetales from within conifers is a result supported largely by mitochondrial genes. However, the first and second codon positions of *rbcL* (Chaw et al., 2000) and the chloroplast photosystem genes *psaA* and *psbB* (Sanderson et al., 2000) further support the Gnetales plus Pinaceae relationship, although analyses of all positions of *rbcL* (Chaw et al., 2000) and third positions of *psaA* and *psbB* (Sanderson et al., 2000) placed Gnetales as sister to all other seed plants. Thus, different data partitions support alternative topologies, and relationships among extant seed plants remain unresolved.

The monophyly of extant gymnosperms and the close relationship of Gnetales and conifers are supported by other mo-

¹ Manuscript received 22 January 2002; revision accepted 7 May 2002.

This research was supported in part by a joint Fulbright Distinguished Professorship to DES and PSS and by NSF grant DEB-0090283. We thank C. dePamphilis and C. Parkinson for providing mtDNA data sets and information regarding RNA editing sites. K. Pryer kindly provided *atpB* sequences.

⁵ Author for reprint requests (e-mail: dsoltis@botany.ufl.edu).

lecular data, although the taxon sampling in these studies was sparse (e.g., Hansen et al., 1999; Winter et al., 1999; Frohlich and Parker, 2000). For example, Hansen et al. (1999) analyzed sequence data for a 9.5-kilobase (kb) portion of the chloroplast genome, but included only *Pinus*, *Gnetum*, and three angiosperms and used *Marchantia* as the outgroup. Winter et al. (1999) analyzed MADS box genes and found similar results for five genes, but only *Gnetum* was used to represent Gnetales, and cycads and *Ginkgo* were not included. Analyses of *Floricaula/LEAFY (LFY)* sequences also suggested gymnosperms are monophyletic and that Gnetales are sister to pines (Frohlich and Parker, 2000), but only five gymnosperms were analyzed. Additional analyses, aimed at inferring relationships within tracheophytes as a whole (e.g., Pryer et al., 2001) or within angiosperms (e.g., Qiu et al., 1999, 2000; Soltis, Soltis, and Chase, 1999; Graham and Olmstead, 2000), also concur in reconstructing the monophyly of the gymnosperms and/or the sister-group relationship of Gnetales and Pinaceae. Because of small sample sizes (at least for gymnosperms) and low internal support, many of these analyses should be considered equivocal regarding seed plant relationships in general and the phylogenetic position of Gnetales in particular (see also Doyle, 1998; Donoghue and Doyle, 2000).

To examine the problem of seed plant relationships further, we constructed a data set containing sequences from four cpDNA genes (*rbcL*, *atpB*, *psaA*, and *psbB*), three mtDNA genes (*mtSSU*, *coxI*, and *atpA*), and one nuclear gene (18S rDNA) for 16 seed plants and three outgroups. All major groups of seed plants are represented, and the total aligned sequence per taxon is 15 772 base pairs (bp). We addressed the following questions: (1) Do the genome partitions (i.e., plastid, mitochondrial, and nuclear) yield consistent phylogenetic results? (2) For protein-coding genes, do separate partitions of first and second vs. third codon positions contain the same phylogenetic information? (3) Does method of analysis affect the relationships reconstructed for seed plants? Although some of these questions have been addressed previously (e.g., Bowe, Coat, and dePamphilis, 2000; Chaw et al., 2000; Sanderson et al., 2000), none of these studies addressed all three of these questions using a data set with more than two genes from each of the organellar genomes.

MATERIALS AND METHODS

Sampling and sources of sequences—All five lineages of extant seed plants were represented in this analysis: cycads, *Ginkgo*, conifers, Gnetales, and angiosperms. Trees were rooted with the lycophyte *Lycopodium* and the ferns *Angiopteris* and *Asplenium*. The sequences analyzed were published previously (Table 1). The 18S rDNA sequences are from Chaw et al. (1997), Soltis et al. (1997), and Soltis et al. (1999), and the alignment follows that of Soltis et al. (1999). As reported previously, a few small regions were difficult to align and were excluded from phylogenetic analysis. Three genes from the mitochondrial genome were used: mitochondrial small subunit (*mtSSU*) sequences, *atpA* (= *atp1*), and *coxI*. The *mtSSU* sequences and alignment are from Parkinson, Adams, and Palmer (1999) and Chaw et al. (2000). All but one of the *atpA* (= *atp1*) sequences are from Bowe, Coat, and dePamphilis (2000); the *atpA* sequence for *Welwitschia* is from Qiu et al. (1999). The *coxI* sequences are from Bowe, Coat, and dePamphilis (2000). The alignments for *atpA* and *coxI* are those of Bowe, Coat, and dePamphilis (2000). Known or presumed RNA editing sites were excluded (Bowe, Coat, and dePamphilis, 2000; C. dePamphilis, Penn State University, personal communication). Sequences of four chloroplast genes were used: *rbcL*, *psaA*, *psbB*, and *atpB*. Most of the *rbcL* sequences are from Chase et al. (1993) and Qiu et al. (1993); others are from Hasebe et al. (1992, 1995). The *psaA* and *psbB*

sequences are from Sanderson et al. (2000). The *atpB* sequences are from Yoshinaga et al. (1992), Wolf (1997), Wolf et al. (1998), Qiu et al. (1999), Savolainen et al. (2000), and Pryer et al. (2001) (Table 1). Alignment of sequences was straightforward and accomplished by eye (see Chase et al., 1993; Sanderson et al., 2000; Savolainen et al., 2000; Pryer et al., 2001).

The lengths of the aligned sequences were: 18S rDNA = 1738 bp; *mtSSU* = 7189 bp; *atpA* = 1584 bp; *coxI* = 1553 bp; *rbcL* = 1321 bp; *psaA* = 2265 bp; *psbB* = 1530 bp; *atpB* = 1504 bp. The chloroplast partition was 6620 bp, the mitochondrial partition was 7189 bp, and the total aligned sequences were 15 772 bp.

To construct this data set, different congeneric species or closely related genera were occasionally used for different genes as placeholders for a lineage (Table 1). For example, *Cryptomeria* represents the conifer family Taxodiaceae for three genes (18S rDNA, *rbcL*, and *mtSSU*), whereas *Metasequoia* was used for *atpA*, *coxI*, and *atpB* and *Sequoia* was used for *psaA* and *psbB*. For Pinaceae, *Pseudotsuga* was used for 18S rDNA, *rbcL*, and *mtSSU*, and *Pinus* was used for *atpA*, *coxI*, *psaA*, *psbB*, and *atpB* (Table 1). In trees derived from multigene analyses, the name of the taxon included for 18S rDNA is provided.

Sequences of 18S rDNA, *rbcL*, and *mtSSU* were available for all 19 taxa. However, sequences of the remaining genes were not available for all taxa. *Angiopteris* lacks *coxI*, and *Araucaria* is missing *atpB*. The largest numbers of missing sequences are for *psbB* (*Podocarpus*, *Nymphaea*, *Illicium*, *Gnetum*, *Cryptomeria*), *atpA* (*Angiopteris*, *Asplenium*, *Taxus*, *Ephedra*), and *psaA* (*Podocarpus*, *Nymphaea*, *Illicium*, *Gnetum*). To examine the impact of missing data, a second eight-gene data set was constructed for 13 taxa (11 seed plants, two outgroups), with only a single taxon missing any gene. For this smaller data set, phylogenetic analyses of the 18S rDNA, combined mtDNA, combined cpDNA, and combined eight genes were conducted. Although the results of these analyses are not shown, the topologies and conclusions are identical to those obtained using the 19-taxon data set and will not be discussed further.

Phylogenetic analyses: maximum parsimony—Maximum parsimony (MP) analyses were conducted on the genomic partitions separately and on the entire data set. The cpDNA genes were also analyzed by codon position. Heuristic searches were conducted using PAUP* 4.02 (Swofford, 1999), with tree-bisection-reconnection (TBR) branch swapping and 1000 random taxon addition replicates, saving all most parsimonious trees. For analyses of individual plastid genes, searches that did not complete were terminated after a minimum of 20 000 shortest trees were obtained. Internal support for relationships was assessed using fast bootstrap analyses with 1000 replicates. Fast bootstrapping provides equivalent values to bootstrapping with branch swapping when support is high and slightly lower values when support is less (Mort et al., 2000).

Phylogenetic analyses: maximum likelihood—Maximum likelihood (ML) analyses were conducted using PAUP 4.0b2 (Swofford, 1999) and an HKY85 + Γ model of nucleotide substitution (Hasegawa, Kishino, and Yano, 1985; Yang, 1994). The MP tree (if more than one) with the greatest $-\ln$ score was used to estimate the model parameters and base frequencies. We used one of the most parsimonious trees and ten equiprobable random trees as starting trees and swapped on all trees, saving all optimal trees. The cpDNA data were analyzed in three ways: (1) all sites; (2) first and second codon positions; and (3) third codon positions.

Phylogenetic analyses: Bayesian approach—The Bayesian phylogenetic analyses were conducted using MrBayes version 1.10 (Huelsenbeck, 2000). We used uniform prior probabilities, the general time-reversible + Γ model of molecular evolution, and a random starting tree. The Markov chain was run for 41 000 generations and sampled every 100 generations. We ran four chains of the Markov Chain Monte Carlo, sampling every 100 generations for 41 000 generations, starting with a random tree. Stationarity was reached at approximately generation 5500; thus, the first 55 trees were the “burn in” of the chain, and phylogenetic inferences are based on those trees sampled after generation 5500.

TABLE 1. Sources of DNA sequences used. NO = no sequence available. GenBank numbers are given in original references.

| Species | Reference |
|--------------------------------|-------------------------------------|
| 18S rDNA | |
| <i>Lycopodium inundata</i> | Soltis et al. (1999) |
| <i>Angiopteris lygodifolia</i> | Chaw et al. (1997) |
| <i>Asplenium nidus</i> | Chaw et al. (1997) |
| <i>Zamia pumila</i> | Chaw et al. (1997) |
| <i>Cycas taitungensis</i> | Chaw et al. (1997) |
| <i>Taxus mairei</i> | Chaw et al. (1997) |
| <i>Cryptomeria japonica</i> | Chaw et al. (1997) |
| <i>Podocarpus costalis</i> | Chaw et al. (1997) |
| <i>Araucaria excelsa</i> | Chaw et al. (1997) |
| <i>Pseudotsuga menziesii</i> | Chaw et al. (1997) |
| <i>Ginkgo biloba</i> | Chaw et al. (1997) |
| <i>Ephedra sinica</i> | Chaw et al. (1997) |
| <i>Welwitschia mirabilis</i> | Soltis et al. (1999) |
| <i>Gnetum gnemon</i> | Soltis et al. (1997) |
| <i>Nymphaea</i> sp. | Soltis et al. (2000) |
| <i>Illicium parviflorum</i> | Soltis et al. (2000) |
| <i>Magnolia tripetala</i> | Soltis et al. (2000) |
| <i>Glycine max</i> | Chaw et al. (1997) |
| <i>Nicotiana tabacum</i> | Soltis et al. (2000) |
| rbcL | |
| <i>Lycopodium digitatum</i> | Hasebe et al. (1995) |
| <i>Angiopteris lygodifolia</i> | Yoshinaga et al. (1992) |
| <i>Asplenium nidus</i> | Hasebe et al. (1995) |
| <i>Zamia pumila</i> | Chase et al. (1993) |
| <i>Cycas circinalis</i> | Chase et al. (1993) |
| <i>Taxus mairei</i> | Chase et al. (1993) |
| <i>Cryptomeria japonica</i> | Chase et al. (1993) |
| <i>Podocarpus costalis</i> | Chase et al. (1993) |
| <i>Araucaria excelsa</i> | Chase et al. (1993) |
| <i>Pseudotsuga menziesii</i> | Chase et al. (1993) |
| <i>Ginkgo biloba</i> | Hasebe et al. (1992) |
| <i>Ephedra tweediana</i> | Chase et al. (1993) |
| <i>Welwitschia mirabilis</i> | Chase et al. (1993) |
| <i>Gnetum gnemon</i> | Chase et al. (1993) |
| <i>Nymphaea variegata</i> | Chase et al. (1993) |
| <i>Illicium parviflorum</i> | Qiu et al. (1993) |
| <i>Magnolia tripetala</i> | Qiu et al. (1993) |
| <i>Pisum sativum</i> | Chase et al. (1993) |
| <i>Nicotiana tabacum</i> | Chase et al. (1993) |
| mtSSU | |
| <i>Lycopodium digitatum</i> | Chaw et al. (2000) |
| <i>Angiopteris lygodifolia</i> | Chaw et al. (2000) |
| <i>Asplenium nidus</i> | Chaw et al. (2000) |
| <i>Zamia floridana</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Cycas revoluta</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Taxus mairei</i> | Chaw et al. (2000) |
| <i>Cryptomeria japonica</i> | Chaw et al. (2000) |
| <i>Podocarpus costalis</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Araucaria excelsa</i> | Chaw et al. (2000) |
| <i>Abies homolepis</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Ginkgo biloba</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Ephedra tweediana</i> | Chaw et al. (2000) |
| <i>Welwitschia mirabilis</i> | Chaw et al. (2000) |
| <i>Gnetum gnemon</i> | Chaw et al. (2000) |
| <i>Nymphaea variegata</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Schisandra spenanthera</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Magnolia grandiflora</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Pisum sativum</i> | Parkinson, Adams, and Palmer (1999) |
| <i>Nicotiana tabacum</i> | Parkinson, Adams, and Palmer (1999) |
| AtpA | |
| <i>Lycopodium inundata</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Angiopteris lygodifolia</i> | NO |
| <i>Asplenium nidus</i> | NO |
| <i>Zamia pumila</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Cycas taitungensis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Taxus mairei</i> | NO |

TABLE 1. Continued.

| Species | Reference |
|-------------------------------------|-------------------------------------|
| <i>Metasequoia glyptostroboides</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Podocarpus costalis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Araucaria excelsa</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Pinus luchuensis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Ginkgo biloba</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Ephedra sinica</i> | NO |
| <i>Welwitschia mirabilis</i> | Qiu et al. (1999) |
| <i>Gnetum gnemon</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Nymphaea</i> sp. | Bowe, Coat, and de Pamphilis (2000) |
| <i>Illicium parviflorum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Magnolia tripetala</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Pisum sativum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Lycopersicon esculentum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>coxI</i> | |
| <i>Lycopodium</i> sp. | Bowe, Coat, and de Pamphilis (2000) |
| <i>Angiopteris lygodifolia</i> | NO |
| <i>Polystichum</i> sp. | Bowe, Coat, and de Pamphilis (2000) |
| <i>Zamia pumila</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Cycas taitungensis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Taxus</i> sp. | Bowe, Coat, and de Pamphilis (2000) |
| <i>Metasequoia glyptostroboides</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Podocarpus macrophyllus</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Araucaria</i> sp. | Bowe, Coat, and de Pamphilis (2000) |
| <i>Pinus strobus</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Ginkgo biloba</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Ephedra viridis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Welwitschia mirabilis</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Gnetum leyboldii</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Nymphaea odorata</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Illicium lanceolatum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Magnolia grandiflora</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Pisum sativum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>Lycopersicon esculentum</i> | Bowe, Coat, and de Pamphilis (2000) |
| <i>psaA</i> | |
| <i>Huperzia squarrosom</i> | Sanderson et al. (2000) |
| <i>Angiopteris evecta</i> | Sanderson et al. (2000) |
| <i>Adiantum capillus-veneris</i> | Sanderson et al. (2000) |
| <i>Encephalartos lebobombensis</i> | Sanderson et al. (2000) |
| <i>Cycas taiwaniana</i> | Sanderson et al. (2000) |
| <i>Torreya californica</i> | Sanderson et al. (2000) |
| <i>Sequoia sempervirens</i> | Sanderson et al. (2000) |
| <i>Podocarpus costalis</i> | NO |
| <i>Araucaria araucana</i> | Sanderson et al. (2000) |
| <i>Pinus thunbergii</i> | Sanderson et al. (2000) |
| <i>Ginkgo biloba</i> | Sanderson et al. (2000) |
| <i>Ephedra tweediana</i> | Sanderson et al. (2000) |
| <i>Welwitschia mirabilis</i> | Sanderson et al. (2000) |
| <i>Gnetum gnemon</i> | NO |
| <i>Nymphaea</i> sp. | NO |
| <i>Illicium parviflorum</i> | NO |
| <i>Drimys winteri</i> | Sanderson et al. (2000) |
| <i>Pisum sativum</i> | Sanderson et al. (2000) |
| <i>Nicotiana tabacum</i> | Sanderson et al. (2000) |
| <i>psbB</i> | |
| <i>Huperzia squarrosom</i> | Sanderson et al. (2000) |
| <i>Angiopteris evecta</i> | Sanderson et al. (2000) |
| <i>Adiantum capillus-veneris</i> | Sanderson et al. (2000) |
| <i>Encephalartos lebobombensis</i> | Sanderson et al. (2000) |
| <i>Cycas taiwaniana</i> | Sanderson et al. (2000) |
| <i>Torreya californica</i> | Sanderson et al. (2000) |
| <i>Sequoia sempervirens</i> | Sanderson et al. (2000) |
| <i>Podocarpus costalis</i> | NO |
| <i>Araucaria araucana</i> | Sanderson et al. (2000) |
| <i>Pinus thunbergii</i> | Sanderson et al. (2000) |
| <i>Ginkgo biloba</i> | Sanderson et al. (2000) |
| <i>Ephedra tweediana</i> | Sanderson et al. (2000) |
| <i>Welwitschia mirabilis</i> | Sanderson et al. (2000) |
| <i>Gnetum gnemon</i> | NO |

TABLE 1. Continued.

| Species | Reference |
|-------------------------------------|-------------------------------------|
| <i>Nymphaea</i> sp. | NO |
| <i>Illicium parviflorum</i> | NO |
| <i>Drimys winteri</i> | Sanderson et al. (2000) |
| <i>Pisum sativum</i> | Sanderson et al. (2000) |
| <i>Nicotiana tabacum</i> | Sanderson et al. (2000) |
| <i>atpB</i> | |
| <i>Huperzia lucidula</i> | Wolf (1997) |
| <i>Angiopteris lygodifolia</i> | Yoshinaga et al. (1992) |
| <i>Adiantum capillus-veneris</i> | Wolf et al. (1998) |
| <i>Zamia pumila</i> | Qiu et al. (1999) |
| <i>Cycas revoluta</i> | Pryer et al. (2001) |
| <i>Taxus baccata</i> | Savolainen et al. (2000) |
| <i>Metasequoia glyptostroboides</i> | Qiu et al. (1999) |
| <i>Podocarpus costalis</i> | Qiu et al. (1999) |
| <i>Araucaria excelsa</i> | NO |
| <i>Pinus thunbergiana</i> | Savolainen et al. (2000) |
| <i>Ginkgo biloba</i> | Savolainen et al. (2000) |
| <i>Ephedra tweediana</i> | Savolainen et al. (2000) |
| <i>Welwitschia mirabilis</i> | Savolainen et al. (2000) |
| <i>Gnetum gnemon</i> | Savolainen et al. (2000) |
| <i>Nymphaea odoratum</i> | Savolainen et al. (2000) |
| <i>Illicium parviflorum</i> | Hoot et al. (1997) |
| <i>Magnolia tripetala</i> | Savolainen et al. (2000) |
| <i>Pisum sativum</i> | Savolainen et al. (2000) |
| <i>Nicotiana tabacum</i> | Deno, Shinozaki, and Sugiura (1983) |

Likelihood ratio and Shimodaira-Hasegawa tests of alternative topologies—We used two methods to investigate the degree to which alternative phylogenetic hypotheses were supported by the cpDNA data, the likelihood ratio (LR) test using the parametric bootstrap (e.g., Swofford et al., 1996; Goldman, Anderson, and Rodrigo, 2000) and the Shimodaira-Hasegawa test (SH test; Shimodaira and Hasegawa, 1999).

Likelihood ratio test—Our MP analysis of first and second codon positions of cpDNA genes place Gnetales within the conifers sister to *Pseudotsuga*, in agreement with several other analyses (see Results). We tested the hypothesis that for first and second codon positions there is no significant difference between the placement of Gnetales as sister to all conifers (conifers monophyletic) and a topology in which Gnetales are sister to *Pseudotsuga* (conifers nonmonophyletic). We first approximated the null distribution using parametric bootstrapping (Huelsenbeck and Hillis, 1996). To conduct the parametric bootstrap we constructed a constraint tree with Gnetales as sister to a monophyletic conifer clade. We then performed MP analysis, enforcing the constraints imposed by the placement of Gnetales sister to a conifer clade. The constrained analysis found a single most parsimonious tree, from which we estimated, using maximum likelihood, parameter values for the HKY85 + Γ model of molecular evolution and base frequencies (Hasegawa, Kishino, and Yano, 1985; Yang, 1994). We then used the most parsimonious tree found in the constrained analysis (Gnetales as sister to monophyletic conifers) to simulate 500 data sets using seq-gen (Rambaut and Grassly, 1997), with the size of the data set identical to the 19-taxon data set (i.e., 19 taxa, 15 772 bp), for each of the two hypotheses under consideration: Gnetales sister to *Pseudotsuga* (conifers nonmonophyletic) vs. Gnetales sister to all conifers (conifers monophyletic).

Each of the simulated data sets was analyzed via MP using methods similar to those described above: TBR branch swapping, 100 random taxon addition replicates, and saving all most parsimonious trees. We enforced a topological constraint of Gnetales sister to *Pseudotsuga* and saved all tree scores and trees. These analyses of the simulated data allowed us to calculate the null distribution of the likelihood-ratio test statistic, $\Delta = (\log L_1 - \log L_0)$ where $\log L_1$ is the $-\ln$ tree score of the unconstrained tree using data simulated under the assumption that Gnetales are sister to a monophyletic conifer clade and $\log L_0$ is the $-\ln$ tree score of the tree(s) obtained from the con-

strained analysis, with Gnetales sister to *Pseudotsuga* and the data simulated assuming that Gnetales are sister to monophyletic conifers. We then tested whether Δ calculated from analysis of the real 19-taxon data set fell within this null distribution of Δ .

Shimodaira-Hasegawa test—Using the Shimodaira-Hasegawa test (SH test; Shimodaira and Hasegawa, 1999) as implemented in PAUP 4.02 we tested three topologies using both the full cpDNA data set (all codon positions) and the partition containing the first and second codon positions of the cpDNA data set. We compared the optimal tree obtained (Gnetales sister to *Pseudotsuga*; conifers nonmonophyletic) to: (1) Gnetales sister to all other seed plants (the topology obtained in MP analyses of third codon positions—see Results), (2) Gnetales sister to a conifer clade (conifers monophyletic), and (3) Gnetales sister to angiosperms. In ML analyses of first and second codon positions, as well as in ML analyses of all codon positions, we found Gnetales sister to *Pseudotsuga* (see Results). For the other hypotheses of relationships, we performed ML analyses in which we constrained Gnetales to be sister to: (1) all other seed plants, (2) monophyletic conifers, and (3) angiosperms. For the SH test, we obtained the test distribution using the re-estimated log likelihoods (RELL) approximation with 1000 nonparametric bootstrap replicates.

RESULTS

Nuclear 18S rDNA—Maximum parsimony analysis of the 18S rDNA data set yielded a single tree (length = 833 steps; consistency index [CI] = 0.591; retention index [RI] = 0.672), in which the gymnosperms are not monophyletic. The cycads (*Cycas* and *Zamia*) and *Ginkgo* are subsequent sisters to the remaining seed plants. However, these deep-level relationships are not well supported (Fig. 1). Gnetales are sister to the conifers (81% bootstrap value), and this Gnetales and conifers clade is the sister group of the angiosperms. The monophyly of Gnetales, conifers, and angiosperms are each well supported (Fig. 1). Both MP and ML analyses of larger 18S rDNA data sets with greater taxon sampling found the gymnosperms to be monophyletic and Gnetales as sister to

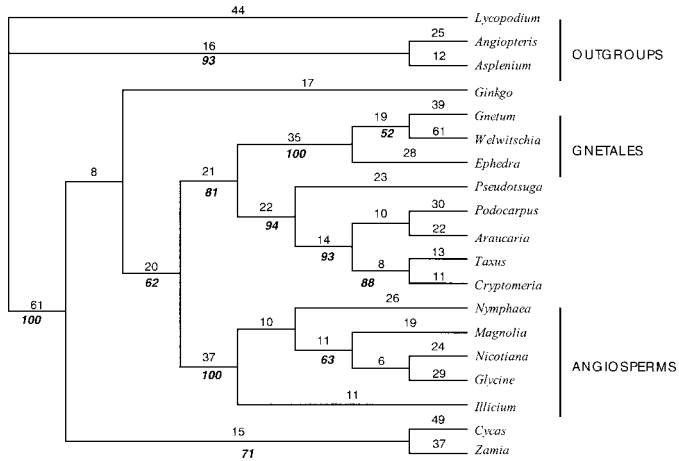


Fig. 1. The single shortest tree obtained in a parsimony analysis of 18S rDNA sequence data. Numbers above branches are branch lengths; those below are bootstrap values. Tree length = 883; CI = 0.610; RI = 0.633.

conifers (Chaw et al., 1997, 2000; Soltis et al., 1999). The difference between our MP tree and the topologies from these earlier studies likely reflects the smaller taxon sampling we used; the sequences included here are taken from the previous studies.

Mitochondrial DNA sequences—Parsimony analysis of the combined data set of three mtDNA genes yielded three shortest trees (length = 2721 steps; CI = 0.770; RI = 0.654). The topology (Fig. 2) is identical to those obtained in previous analyses of mtDNA sequence data using both MP and ML methods (Bowe, Coat, and dePamphilis, 2000; Chaw et al., 2000). Furthermore, also as in earlier analyses, bootstrap support is very high. The mtDNA tree provides strong support (93%) for the monophyly of the gymnosperms, which appear as the sister group to a well-supported angiosperm clade (100%). Within the gymnosperms, there is also strong support for cycads (95%) followed by *Ginkgo* (96%) as successive sisters to the remaining gymnosperms and for a conifer clade that includes Gnetales (96%). The sister-group relationship of Gnetales and *Pseudotsuga* of the Pinaceae (the “gne-pine” hypothesis of Chaw et al., 2000) is also strongly supported.

Chloroplast DNA sequences—For each of the four cpDNA genes we analyzed all sites, first and second codon positions, and third codon positions with both MP and ML. Relevant aspects of the topologies reconstructed for the cpDNA positions are summarized in Table 2. For *atpB*, MP analyses of all codon positions, as well as first and second positions only, and third positions alone, did not resolve the relationships of Gnetales. Even the majority-rule consensus of shortest trees placed

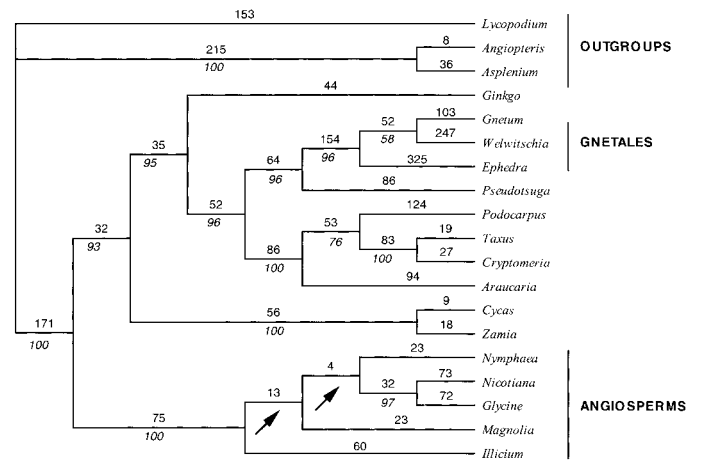


Fig. 2. One of three shortest trees obtained in a parsimony analysis of a combined mtDNA data set of three genes (*coxI*, *19S*, *atpA*). The arrows indicate the nodes that collapse in the strict consensus tree. Numbers above branches are branch lengths; those below are bootstrap values. Tree length = 2721; CI = 0.770; RI = 0.654.

Gnetales in a polytomy with other seed plant lineages. In contrast, ML analysis of all three codon positions retrieved a gymnosperm clade with Gnetales derived from within conifers and sister to a clade of *Podocarpus* and (*Taxus* and *Cryptomeria*). ML analysis of first and second positions for *atpB* resulted in a gymnosperm clade with Gnetales derived from within conifers and sister to a clade of *Taxus* and *Cryptomeria*. In the ML analysis of third codon positions of *atpB*, Gnetales appeared as sister to the angiosperms.

Maximum parsimony analyses of *rbcL* sequences suggested that Gnetales are sister to all other seed plants (Table 2). In contrast, MP analyses of first and second positions only and third positions alone did not resolve the relationship of Gnetales. Maximum likelihood analyses of all codon positions placed Gnetales sister to *Pseudotsuga*; the same topology was obtained in ML analyses of third positions. However, ML analyses of first and second positions placed Gnetales sister to all other gymnosperms.

Analyses of *psaA* and *psbB* sequences provided nearly identical topologies in all analyses, except one. The MP analyses of all three codon positions for *psaA*, as well as *psbB*, did not resolve the position of Gnetales. However, the MP analyses of first and second positions of *psaA*, and of *psbB*, indicated that Gnetales are sister to *Pseudotsuga*, whereas MP analyses of third codon positions for *psaA*, as well as *psbB*, suggested that Gnetales are sister to all other seed plants. Using ML and both *psaA* and *psbB*, analyses of all codon positions and first and second codon positions found Gnetales sister to *Pseudotsuga*. In ML analysis of third codon positions for *psaA*, Gnetales

TABLE 2. Results of phylogenetic analyses of cpDNA data partitions describing the position of Gnetales, listing the taxa that appear as sister to Gnetales under maximum parsimony (MP) and maximum likelihood (ML).

| | All positions | | First and second | | Third | |
|-------------|-----------------------|--------------------|--------------------|-----------------------------------|-----------------------|-----------------------|
| | MP | ML | MP | ML | MP | ML |
| <i>atpB</i> | All other seed plants | <i>Pseudotsuga</i> | Unresolved | <i>Taxus</i> + <i>Cryptomeria</i> | Unresolved | Angiosperms |
| <i>rbcL</i> | All other seed plants | <i>Pseudotsuga</i> | Unresolved | All other gymnosperms | Unresolved | <i>Pseudotsuga</i> |
| <i>psaA</i> | All other seed plants | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | All other seed plants | <i>Pseudotsuga</i> |
| <i>psbB</i> | All other seed plants | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | All other seed plants | All other seed plants |
| All cpDNA | All other seed plants | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | <i>Pseudotsuga</i> | All other seed plants | <i>Pseudotsuga</i> |

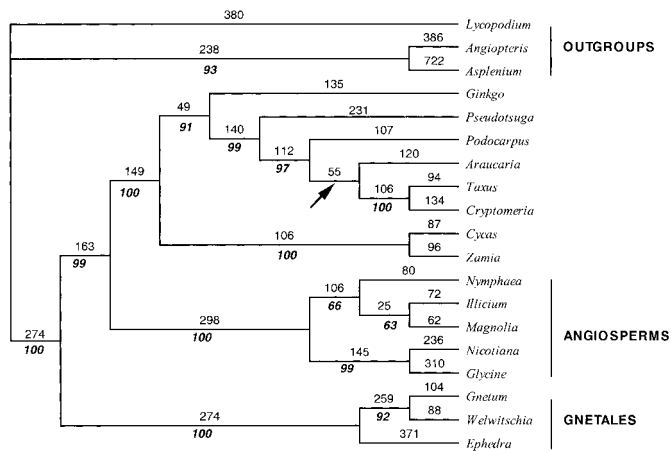


Fig. 3. One of two shortest trees obtained in a parsimony analysis of a combined plastid data set of four genes (*rbcl*, *atpB*, *psaA*, *psbB*). The arrow indicates the node that collapses in the strict consensus of two shortest trees. Numbers above branches are branch lengths; those below are bootstrap values. Tree length = 6314; CI = 0.570; RI = 0.476.

again appeared as sister to *Pseudotsuga*. However, ML analysis of third codon positions of *psbB* placed Gnetales sister to all other seed plants.

Maximum parsimony searches of the combined data set of four cpDNA genes yielded two shortest trees (length = 6314; CI = 0.570; RI = 0.476). The two MP trees differed only in the placement of *Araucaria*, which appeared as sister to either *Podocarpus* or *Taxus* and *Cryptomeria*. The relationships in the MP trees are strongly supported but differ from those obtained with either 18S rDNA or mtDNA sequence data. Gnetales are sister to all other extant seed plants, the monophyly of which receives very strong support (99%). The angiosperms (100% support) are sister to a well-supported (100%) clade of cycads, *Ginkgo*, and conifers. Within this clade, the cycads are sister to *Ginkgo* and conifers, the latter clade receiving bootstrap support of 91% (Fig. 3). Thus, MP analyses of the combined cpDNA data appear to be in strong conflict with mtDNA data and 18S rDNA data, both of which suggest that the gymnosperms are monophyletic and that Gnetales are related to conifers (see also Chaw et al., 1997, 2000; Soltis et al., 1999; Bower, Coat, and dePamphilis, 2000).

For the ML analysis of the cpDNA partition, model parameter values estimated from a single MP tree ($-\ln = 36582.83402$) are as follows: (1) base frequencies, A: 0.274863, C: 0.199218, G: 0.205064, T: 0.320856; (2) ti/tv = 3.515509; and (3) the gamma shape parameter $\alpha = 0.256015$. The ML tree ($-\ln = 36519.52472$) is essentially identical to that obtained with mtDNA gene sequence data. The gymnosperms are monophyletic and sister to the angiosperms. Within the gymnosperms, cycads are sister to *Ginkgo* and the remaining gymnosperms. Gnetales are embedded within the conifers and appear as sister to *Pseudotsuga* (Fig. 4).

MP analysis of first and second codon positions yielded two most parsimonious trees (length = 2102; CI = 0.623; RI = 0.524) that differed only in the placement of *Araucaria*, which appeared as sister to either *Podocarpus* or *Taxus* and *Cryptomeria*. Both trees placed Gnetales as sister to *Pseudotsuga*. In the ML analysis of first and second codon positions, model parameter values estimated using a single most parsimonious tree with a score of $-\ln 16704.27212$ were: (1) base fre-

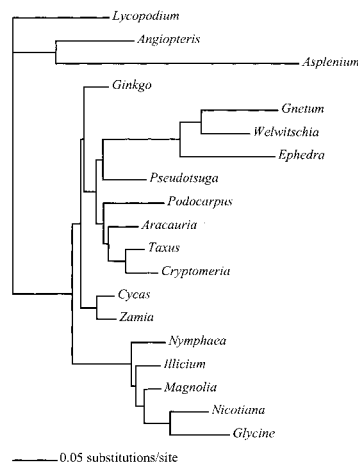


Fig. 4. Tree obtained in maximum likelihood analysis of combined cpDNA data set ($-\ln = 36519.52472$).

quencies, A: 0.257348, C: 0.218534, G: 0.253843, T: 0.270275, (2) ti/tv: 2.436609, and (3) the gamma shape parameter $\alpha = 0.150712$. The ML tree was the same as the MP tree from which the parameters were estimated. Gnetales again appeared as sister to *Pseudotsuga*.

Maximum parsimony analysis of the cpDNA third positions produced a single most parsimonious tree (length = 4189; CI = 0.547; RI = 0.460) in which Gnetales appeared as sister to the rest of the seed plants. Maximum likelihood parameter values estimated from the single most parsimonious tree were: (1) base frequencies, A: 0.295411, C: 0.171523, G: 0.151231, T: 0.381834; (2) ti/tv: 4.915934; and (3) gamma shape parameter $\alpha = 0.810807$. In contrast to the MP topology, the ML tree ($-\ln 18794.58221$) placed Gnetales as sister to *Pseudotsuga* and cycads as sister to the angiosperms.

The analysis of the combined cpDNA data set using a Bayesian approach yielded a tree (Fig. 5) comparable to that obtained with ML. Most clades, including the angiosperms (1.0), all gymnosperms (1.0), Gnetales and Pinaceae (0.987), and the successive sister-group relationships of cycads fol-

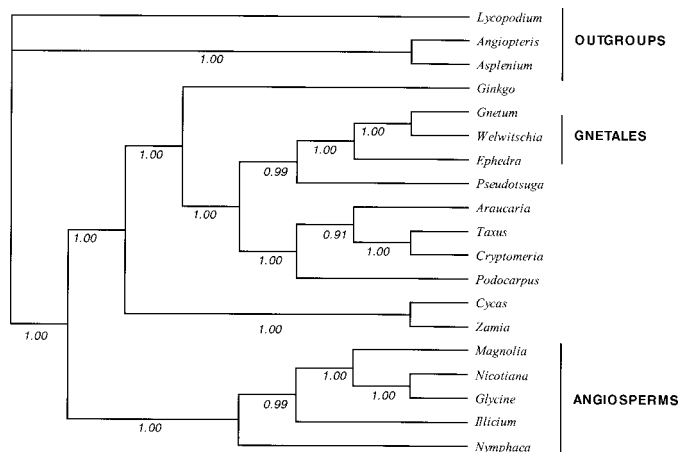


Fig. 5. One of the 691 best trees (also equivalent to the majority rule tree) obtained in a MrBayes analysis (Huelsenbeck, 2000) of a combined plastid data set of four genes (*rbcl*, *atpB*, *psaA*, *psbB*). Numbers below branches are the frequency of recovery of each clade.

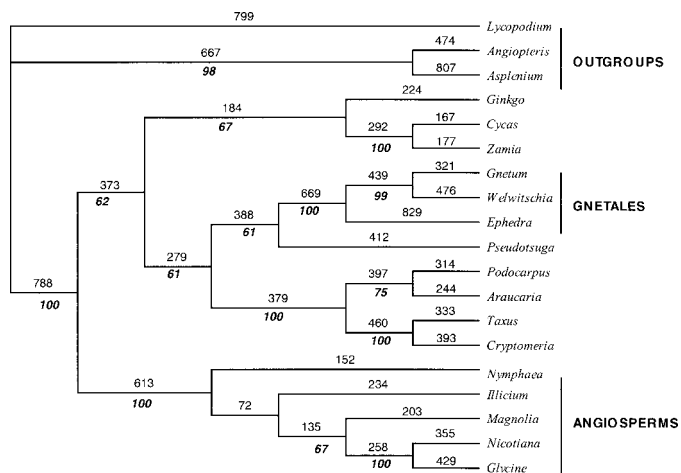


Fig. 6. The single shortest tree obtained in a parsimony analysis of a combined plastid, mtDNA, 18S rDNA data set of eight genes (*rbcL*, *atpB*, *psaA*, *psbB*, *coxI*, *19S rDNA*, *atpA*, *18S rDNA*). Numbers above branches are branch lengths; those below are bootstrap values. Tree length = 13736; CI = 0.649; RI = 0.511.

lowed by *Ginkgo* were supported by posterior probabilities of 1.0 or nearly 1.0.

Combined 18S rDNA, mtDNA, and cpDNA sequences—Parsimony analysis of all eight genes representing all three genomes yielded a single shortest tree (length = 13736; CI = 0.649; RI = 0.511). The topology recovered in this analysis (Fig. 6) is nearly identical to that retrieved with mtDNA sequences and in the ML analysis of cpDNA genes. However, bootstrap support for relationships in the eight-gene tree is in some cases much lower than obtained in the mtDNA analyses. The angiosperms are again strongly supported (100%) and are sister to the gymnosperms, but the monophyly of the gymnosperms is only weakly (62%) supported. Within the gymnosperms, there is weak support for a clade of cycads plus *Ginkgo*, rather than a grade of cycads, *Ginkgo*, and the remaining extant gymnosperms. Gnetales are derived from within the conifers, but the monophyly of this conifer/Gnetales clade is only weakly supported (61%). Gnetales appear as sister to *Pseudotsuga*, but with low support (61%). The lower support for relationships in this eight-gene tree reflects the conflict in parsimony analyses between the cpDNA and mtDNA partitions, each of which yields strongly supported, yet differing, topologies (compare Figs. 2–5).

Analysis of the combined data set using ML resulted in a tree (not shown) similar to that retrieved using MP. Gymnosperms are monophyletic, with Gnetales sister to Pinaceae. Cycads and *Ginkgo* appear as successive sisters to the conifer and Gnetales clade. The analysis of the combined data set of all genes using a Bayesian approach yielded a tree (Fig. 7) in which a sister group of Gnetales and Pinaceae was strongly supported (probability of 1.0). As in the eight-gene MP tree, cycads and *Ginkgo* form a clade with weak support (0.72 probability), rather than appearing as successive sisters to conifers and Gnetales. The gymnosperms form a clade, but their monophyly is also only weakly supported (0.72). The single difference between the Bayesian tree and that obtained with MP involves the placement of *Magnolia*, which is sister to other angiosperms in the Bayesian analysis, but is sister to *Nicotiana* and *Glycine* in the MP tree (compare Figs. 6 and 7).

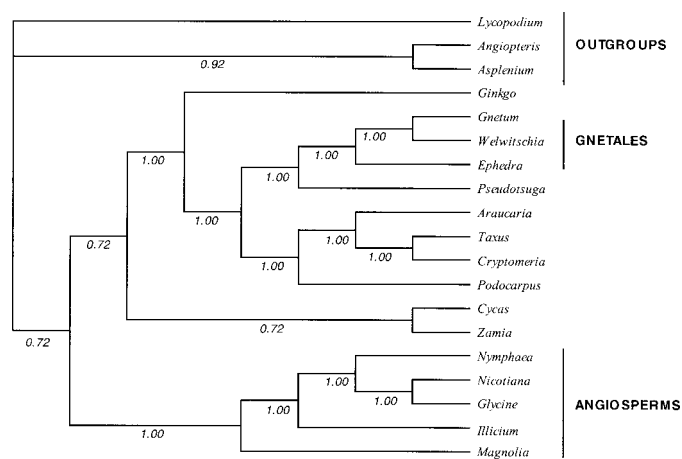


Fig. 7. One of the 505 best trees (also equivalent to the majority rule tree) obtained in a MrBayes analysis of a combined plastid, mtDNA, 18S rDNA data set of eight genes (*rbcL*, *atpB*, *psaA*, *psbB*, *coxI*, *19S rDNA*, *atpA*, *18S rDNA*). Numbers below branches are the frequency of recovery of each clade.

Likelihood ratio and Shimodaira-Hasegawa tests—The likelihood ratio (LR) test with the parametric bootstrap indicates a significant difference in the MP tree length between the two topologies examined, Gnetales as sister to all conifers (conifers monophyletic), and a topology in which Gnetales are sister to *Pseudotsuga* (conifers non-monophyletic). The preferred topology based on this test is that Gnetales are sister to *Pseudotsuga*; conifers are therefore nonmonophyletic (Fig. 8).

The SH tests of first and second codon positions and all codon positions provided similar results. The tree topologies we examined are: Gnetales sister to all other seed plants, Gnetales sister to a conifer clade (conifers monophyletic), Gnetales sister to *Pseudotsuga* (conifers nonmonophyletic), and Gnetales sister to the angiosperms. Analyses of first and second codon positions indicated that there is a significant difference in $-\ln$ score between the best topology (Gnetales sister to *Pseudotsuga*) and two of these alternative topologies (see Table 3). The *P* value for Gnetales sister to a conifer clade is 0.257 and clearly not significant (Table 3); however, the *P* value for Gnetales sister to all other seed plants (0.018) is significant (critical value is 0.05), as is the *P* value for Gnetales sister to angiosperms (0.006). Analyses of the full cpDNA data set (all codon positions) similarly indicated that there is no significant difference in $-\ln$ score between the optimal tree and a topology in which Gnetales are sister to conifers. As with first and second codon positions, however, there is a significant difference in $-\ln$ score between the best topology and the topology in which Gnetales are sister to the rest of the seed plants and the topology in which Gnetales are sister to angiosperms (Table 3).

DISCUSSION

Plastid genes—Although analyses of mtDNA genes have provided strong support for a monophyletic gymnosperm clade with Gnetales sister to Pinaceae (Bowe, Coat, and dePamphilis, 2000; Chaw et al., 2000), phylogenetic studies of seed plants employing cpDNA genes have yielded varying topologies depending on the sampling of taxa, the method of analysis, and the codon positions analyzed (e.g., Hasebe et al., 1992; Albert et al., 1994; Goremykin et al., 1996; Chaw et al.,

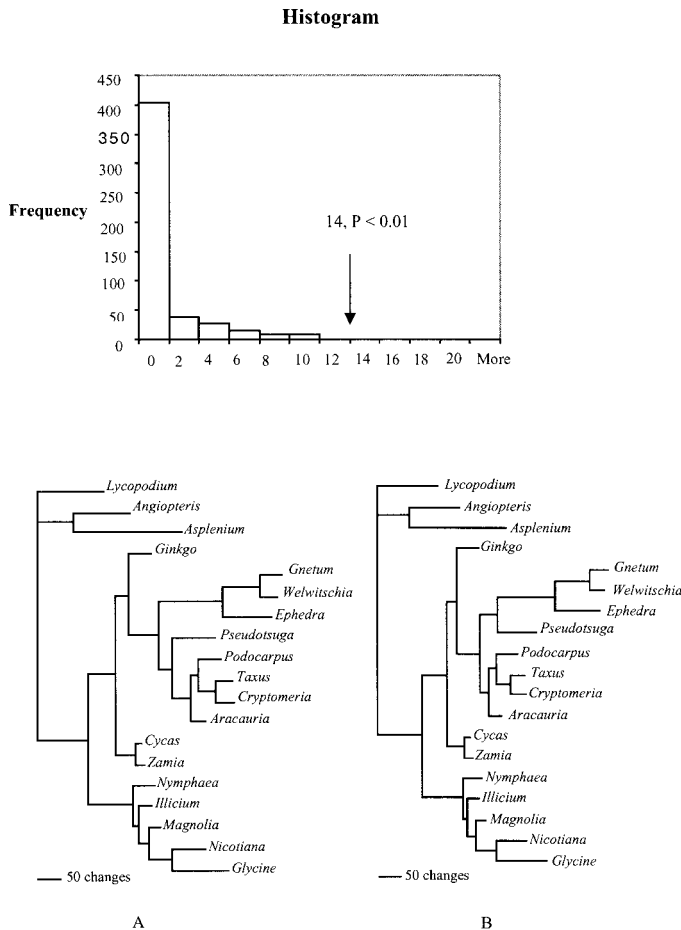


Fig. 8. Graph of the likelihood ratio test results approximating the distribution using parametric bootstrapping. This test compared the two trees below for maximum parsimony for first and second codon positions: Tree A is the suboptimal tree placing Gnetales sister to a conifer clade; B is the optimal or best tree with Gnetales sister to *Pseudotsuga*.

2000; Sanderson et al., 2000). The difference in signal we observed between first and second vs. third positions for cpDNA genes is a pattern observed for seed plants in other studies. Sanderson et al. (2000) found strongly supported, but conflicting, topologies with *psaA* and *psbB* gene sequences when the data were partitioned into first and second vs. third codon positions. In our MP analyses of first and second positions, both genes supported the monophyly of the gymno-

sperms, with Gnetales related to Pinaceae; with third positions, Gnetales were sister to all other seed plants. The ML analyses of first and second as well as third positions of *psaA* yielded a topology with Gnetales sister to Pinaceae, but ML analyses of *psbB* third positions placed Gnetales as sister to all other seed plants.

A difference in phylogenetic signal between first and second vs. third positions is also apparent for *rbcL*. Chaw et al. (2000) obtained different topologies for seed plants depending on the treatment of third positions. With all positions equally weighted, ML analyses of *rbcL* placed Gnetales as sister to the other seed plants, whereas MP analyses placed Gnetales as sister to the angiosperms. However, when third position transitions were excluded, both ML and MP placed Gnetales as sister to Pinaceae.

Our analyses of cpDNA gene sequences likewise indicate that the signal of third codon positions differs from that of first and second codon positions (see also Chaw et al., 2000; Sanderson et al., 2000). Sanderson et al. (2000) concluded that a factor potentially contributing to bias and inconsistency in their analyses of *psaA* and *psbB* is the combination of extremely short branches at the base of the seed plants as well as extremely high rates of molecular evolution in Gnetales and the outgroups. Sanderson et al. (2000) suggested that the difference in topology obtained with MP for first and second codon positions vs. third codon positions of these two cpDNA genes might be the result of saturation of third position sites at this deep phylogenetic level (i.e., all seed plants). Furthermore, analyses of the cpDNA genes *rbcL*, *atpB*, and *rps4* across all tracheophytes indicate heterogeneous rates of evolution in first, second, and third codon positions (Soltis et al., 2002) that could translate into conflicting phylogenetic signal.

Our results, coupled with those of Chaw et al. (2000) and Sanderson et al. (2000), provide an intriguing paradox in that recent studies have pointed to the value of third codon positions of cpDNA genes in phylogeny reconstruction (e.g., Lewis, Mishler, and Vilgalys, 1997; Källersjö et al., 1998; Källersjö, Albert, and Farris, 1999; Olmstead, Reeves, and Yen, 1998). Even at very deep phylogenetic levels, such as all photosynthetic life, third positions in *rbcL* sequences provide most of the phylogenetic signal (Källersjö et al., 1998; Källersjö, Albert, and Farris, 1999). However, our results suggest that third codon positions of cpDNA genes may be misleading in phylogenetic analyses of seed plants.

The conflict between first and second vs. third codon positions in plastid genes that we observed has similarly been reported in other investigations of seed plant relationships (San-

TABLE 3. Results of Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999).

| Tree | -ln L | Diff - ln L | P |
|--|---------------|-------------|--------|
| First and second position SH test | | | |
| Gnetales + <i>Pseudotsuga</i> | 16 704.272 25 | Best | |
| Gnetales sister to all other seed plants | 16 743.052 74 | 38.780 62 | 0.018* |
| Gnetales sister to conifer clade | 16 716.749 59 | 12.477 47 | 0.257 |
| Gnetales sister to angiosperms | 16 752.911 68 | 48.639 56 | 0.006* |
| Full data set SH test | | | |
| Gnetales + <i>Pseudotsuga</i> | 36 518.896 81 | Best | |
| Gnetales sister to all other seed plants | 36 555.701 50 | 36.804 69 | 0.010* |
| Gnetales sister to conifer clade | 36 530.728 06 | 11.831 26 | 0.233 |
| Gnetales sister to angiosperms | 36 572.061 47 | 53.164 66 | 0.001* |

* Significant at the 0.05 level.

derson et al., 2000; Rydin, Källersjö, and Friis, 2002). The nature of this conflict appears to be complex. Sanderson et al. (2000) reported conflict between first and second vs. third codon positions in the plastid genes *psaA* and *psbB*. Rydin, Källersjö, and Friis (2002) analyzed the plastid genes *rbcL* and *atpB* across seed plants and found that the conflict in codon position is more pronounced in *rbcL* than in *atpB*, but the conflict in these genes is less than that observed for *psaA* and *psbB*. Although third codon positions of plastid genes generally have most of the phylogenetic signal (e.g., Källersjö et al., 1998; Olmstead, Reeves, and Yen, 1998), it may be that third positions are saturated in some, but not all, instances. Chaw et al. (2000) found that third positions of *rbcL* were saturated based on transition/transversion ratios across seed plants. In contrast, however, Sanderson et al. (2000) found that for *psaA* the third codon positions were not sufficiently saturated to generate a misleading phylogenetic result. Adding to the complexity of the conflict between first, second, and third positions is the fact that transitions within each codon position conflict with transversions (Rydin, Källersjö, and Friis, 2002).

The preponderance of evidence—gene trees based on (1) first and second positions of four cpDNA genes, as well as third positions with ML and all positions with ML and Bayesian approaches, (2) nuclear 18S rDNA, and (3) three mtDNA genes—supports a relationship between Gnetales and conifers, although 18S rDNA suggests that Gnetales are sister to all conifers rather than to Pinaceae (Soltis et al., 1999; Chaw et al., 2000).

General conclusions—An overall conclusion of our analyses is that some aspects of relationship among extant seed plants appear to be largely resolved. We conclude that extant gymnosperms are monophyletic and that cycads and *Ginkgo* are sisters to the remaining gymnosperms, with Gnetales sister to Pinaceae (if not embedded within Pinaceae) within the conifers. However, the relationship between cycads and *Ginkgo* differs in our analyses; in some analyses, cycads and *Ginkgo* are successive sisters to a clade of conifers and Gnetales, whereas in others, *Ginkgo* and cycads form a clade that is sister to the other gymnosperms.

Most of the conflict observed in our own analyses, as well as in previous analyses, involves the difference between first and second vs. third codon positions of cpDNA genes. When ML analyses of cpDNA genes are conducted, or third positions are excluded in MP analyses, a topology highly similar to those obtained with mtDNA genes emerges: gymnosperms are strongly supported as monophyletic and are sister to the angiosperms. Within the gymnosperms, cycads and *Ginkgo* are successive sisters to the remainder of the clade; Gnetales are closely related to conifers. However, the precise relationship of Gnetales to conifers is unclear. Some analyses of both plastid and mitochondrial genes support a placement of Gnetales within conifers as sister to Pinaceae (the “gne-pine” hypothesis; Bowe, Coat, and dePamphilis, 2000; Chaw et al., 2000). However, with 18S rDNA sequences (Chaw et al., 1997; Soltis et al., 1999) and with multiple genes that include a broad representation of conifers (Rydin, Källersjö, and Friis, 2002), Gnetales are sister to all conifers. Hence, the gne-pine hypothesis may be an artifact of inadequate taxon sampling in some analyses (Rydin, Källersjö, and Friis, 2002).

Hypothesis tests using the LR and SH tests and the cpDNA data provide conflicting results. The LR test based on first and second positions indicates that there is a significant difference

in the MP tree length between the two topologies examined, Gnetales as sister to all conifers (conifers monophyletic) and a topology in which Gnetales are sister to *Pseudotsuga* (conifers non-monophyletic). The preferred topology based on the LR test is that Gnetales are sister to *Pseudotsuga*. The SH test based on first and second codon positions and all codon positions indicated that there is no significant difference in $-\ln$ score between the best topology (Gnetales sister to *Pseudotsuga*) and Gnetales sister to a conifer clade. However, there is a significant difference in $-\ln$ score between the best topology and the topology in which Gnetales are sister to the rest of the seed plants and the topology in which Gnetales are sister to the angiosperms. Thus, we found that the alternative placement of Gnetales as sister to conifers (rather than as sister to Pinaceae) could not be rejected consistently, in agreement with Rydin, Källersjö, and Friis (2002), who, in some analyses, recovered Gnetales as sister to a monophyletic conifer clade. Lastly, a structural mutation in the chloroplast genome (Raubeson and Jansen, 1992) also suggests that a Gnetales-conifer sister-group relationship may be a more parsimonious explanation of the data.

Future considerations—With 15 772 bp of sequence data representing all three plant genomes, relationships among the five lineages of extant seed plants are not yet fully resolved. The value of additional molecular studies of a restricted number of taxa therefore needs to be carefully considered. The precise relationship of cycads and *Ginkgo*—are they sister taxa or subsequent branches to the remaining gymnosperms?—has yet to be answered. Furthermore, although most molecular data suggest a close relationship between conifers and Gnetales, the exact nature of this relationship is unclear: Are Gnetales derived from within conifers or are they the sister group to conifers? To determine the placement of Gnetales relative to conifers we suggest that additional conifers be added to existing gene sequence data sets, rather than simply adding sequences of more genes.

Although molecular data can provide some additional insights into the relationships among extant seed plants, efforts at achieving a better understanding of relationships among extant seed plants would be better placed by integrating fossil taxa into comprehensive analyses of seed plant phylogeny. For example, DNA data strongly support the monophyly of extant gymnosperms, but is this a valid conclusion when fossil taxa are included? If no single extant lineage of gymnosperms is the sister group of angiosperms, can a fossil lineage be identified that fills this role? More attention must be paid to fossils and morphological features, with future efforts directed at the integration of fossils into phylogenetic analyses based on both morphological and molecular data (see also Doyle, 1998; Donoghue and Doyle, 2000). Such analyses will require a careful reconsideration of the morphological characters included in previous studies of seed plant phylogeny. Clearly, the homology of those characters that linked angiosperms and Gnetales must be reevaluated. Analyses that integrate fossils and morphology with molecular characters may be more fruitful than further analyses of molecular data alone.

LITERATURE CITED

- ALBERT, V. A., A. BACKLUND, K. BREMER, M. W. CHASE, J. R. MANHART, B. D. MISHLER, AND K. C. NIXON. 1994. Functional constraints and *rbcL* evidence for land plant phylogeny. *Annals of the Missouri Botanical Garden* 81: 534–567.

- BOWE, L. M., G. COAT, AND C. W. DEPAMPHILIS. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4092–4097.
- CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.
- CHAW, S.-M., C. L. PARKINSON, Y. CHENG, T. M. VINCENT, AND J. D. PALMER. 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4086–4091.
- 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. *Molecular Biology and Evolution* 14: 56–58.
- CRANE, P. R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Annals of the Missouri Botanical Garden* 72: 716–793.
- DENO, H., K. SHINOZAKI, AND M. SUGIURA. 1983. Nucleotide sequence of tobacco chloroplast gene for the alpha subunit of proton-translocating ATPase. *Nucleic Acids Research* 11: 2185–2191.
- DONOGHUE, M. J., AND J. A. DOYLE. 2000. Seed plant phylogeny: demise of the anthophyte hypothesis? *Current Biology* 10: R106–R109.
- DOYLE, J. A. 1996. Seed plant phylogeny and the relationships of Gnetales. *International Journal of Plant Sciences* 157: S3–S39.
- DOYLE, J. A. 1998. Molecules, morphology, fossils, and the relationships of angiosperms and Gnetales. *Molecular Phylogenetics and Evolution* 9: 448–462.
- DOYLE, J. A., AND M. J. DONOGHUE. 1986. Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach. *Botanical Review* 52: 321–431.
- DOYLE, J. A., AND M. J. DONOGHUE. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* 44: 89–104.
- FROHLICH, M. W., AND D. S. PARKER. 2000. The mostly male theory of flower evolutionary origins: from genes to fossils. *Systematic Botany* 25: 155–170.
- GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652–670.
- GOREMYKIN, V., V. BOBROVA, J. PAHNKE, A. TROITSKY, A. ANTONOV, AND W. MARTIN. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support Gnetalean affinities of angiosperms. *Molecular Biology and Evolution* 13: 383–396.
- GRAHAM, S. W., AND R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.
- HAMBY, R. K., AND E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], *Molecular systematics of plants*, 50–91. Chapman and Hall, New York, New York, USA.
- HANSEN, A., S. HANSMANN, T. SAMIGULLIN, A. ANTONOV, AND W. MARTIN. 1999. *Gnetum* and the angiosperms: molecular evidence that their shared morphological characters are convergent, rather than homologous. *Molecular Biology and Evolution* 16: 1006–1009.
- HASEBE, M., ET AL. 1995. Fern phylogeny based on *rbcL* nucleotide sequences. *American Fern Journal* 85: 134–181.
- HASEBE, M., R. KOFUJI, M. ITO, M. KATO, K. IWATSUKI, AND K. UEDA. 1992. Phylogeny of gymnosperms inferred from *rbcL* gene sequences. *Botanical Magazine of Tokyo* 105: 673–679.
- HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 32–38.
- HOOT, S. B., J. W. KADEREIT, F. R. BLATTNER, K. B. JORK, A. E. SCHWARZBACH, AND P. R. CRANE. 1997. Data congruence and phylogeny of the Papaveraceae s. l. based on four data sets: *atpB* and *rbcL* sequences, *trnK* restriction sites, and morphological characters. *Systematic Botany* 22: 575–590.
- HUELSENBECK, J. P. 2000. MrBayes. Distributed by the author. Department of Biology, University of Rochester, Rochester, New York, USA.
- HUELSENBECK, J. P., AND D. M. HILLIS. 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. In J. D. Ferrarier and S. R. Palumbi [eds.], *Molecular zoology: advances, strategies, and protocols*, 19–45. Wiley-Liss, New York, New York, USA.
- KÄLLERSJÖ, M., V. ALBERT, AND J. S. FARRIS. 1999. Homoplasy increases phylogenetic structure. *Cladistics* 15: 91–93.
- KÄLLERSJÖ, M., J. S. FARRIS, M. W. CHASE, B. BREMER, M. F. FAY, C. J. HUMPHRIES, G. PETERSEN, O. SEBERG, AND K. BREMER. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants, and flowering plants. *Plant Systematics and Evolution* 213: 259–287.
- LEWIS, L. A., B. D. MISHLER, AND R. VILGALYS. 1997. Phylogenetic relationships of the liverworts (Hepatitae), a basal Embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Molecular Phylogenetics and Evolution* 7: 377–393.
- LOCOYTE, H., AND D. W. STEVENSON. 1990. Cladistics of the Spermophyta. *Brittonia* 42: 197–211.
- MALEK, O., K. LATTIG, R. HIESEL, A. BRENNICKE, AND V. KNOOP. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO* 14: 1403–1411.
- MORT, M. E., P. S. SOLTIS, D. E. SOLTIS, AND M. L. MABRY. 2000. A comparison of three methods for estimating internal support on phylogenetic trees. *Systematic Biology* 49: 102–113.
- OLMSTEAD, R. G., P. A. REEVES, AND A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. In D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], *Molecular systematics of plants II*, 164–187. Kluwer, New York, New York, USA.
- PARKINSON, C. L., K. L. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488.
- PRYER, K. M., H. SCHNEIDER, A. R. SMITH, R. CRANFILL, P. WOLF, J. S. HUNT, AND S. D. SIPES. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* 409: 618–622.
- QIU, Y.-L., M. W. CHASE, D. H. LES, AND C. R. PARKS. 1993. Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 587–606.
- QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.
- QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 2000. Phylogenetic analyses of basal angiosperms based on five genes from all three genomes. *International Journal of Plant Sciences* 161: S3–S27.
- RAMBAUT, A., AND N. C. GRASSLY. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* 13: 235–238.
- RAUBESON, L. A., AND R. K. JANSEN. 1992. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochemical Systematics and Ecology* 20: 17–24.
- ROTHWELL, G. W., AND R. SERBET. 1994. Lignophyte phylogeny and the evolution of spermatophytes: a numerical cladistic analysis. *Systematic Botany* 19: 443–482.
- RYDIN, C., M. KÄLLERSJÖ, AND E. M. FRIIS. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: conflicting data, rooting problems, and the monophyly of conifers. *International Journal of Plant Sciences* 163: 197–214.
- SANDERSON, M. J., M. F. WOJCIECHOWSKI, J.-M. HU, T. S. KHAN, AND S. G. BRADY. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* 17: 782–797.
- SAVOLAINEN, V., M. W. CHASE, C. M. MORTON, S. B. HOOT, D. E. SOLTIS, C. BAYER, M. F. FAY, A. DE BRUIJN, S. SULLIVAN, AND Y.-L. QIU. 2000. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* 49: 306–362.
- SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.
- SOLTIS, D. E., ET AL. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.
- SOLTIS, D. E., ET AL. 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.
- SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny

- inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.
- SOLTIS, P. S., D. E. SOLTIS, V. SAVOLAINEN, P. R. CRANE, AND T. G. BAR-
RACLOUGH. 2002. Integration of molecular and fossil data and evidence
for molecular living fossils. *Proceedings of the National Academy of
Sciences, USA* 99: 4430–4435.
- SOLTIS, P. S., D. E. SOLTIS, P. G. WOLF, D. L. NICKRENT, S.-M. CHAW, AND
R. L. CHAPMAN. 1999. The phylogeny of land plants inferred from 18S
rDNA sequences: pushing the limits of rDNA signal? *Molecular Biology
and Evolution* 16: 1774–1784.
- SWOFFORD, D. L. 1999. PAUP*: phylogenetic analysis using parsimony (*and
other methods). Version 4.0b4a. Sinauer, Sunderland, Massachusetts,
USA.
- SWOFFORD, D. L., G. J. OLSEN, P. WADDELL, AND D. M. HILLIS. 1996. Phy-
logenetic inference. In D. M. Hillis, C. Moritz, and B. K. Mable [eds.],
Molecular systematics, 407–425. Sinauer, Sunderland, Massachusetts,
USA.
- WINTER, K.-U., A. BECKER, T. MUNSTER, J. T. KIM, H. SAEDLER, AND G.
THIESSEN. 1999. The MADS-box genes reveal that gnetophytes are more
closely related to conifers than to flowering plants. *Proceedings of the
National Academy of Sciences, USA* 96: 7342–7347.
- WOLF, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic
studies of ferns and other pteridophytes. *American Journal of Botany* 84:
1429–1440.
- WOLF, P. G., K. M. PRYER, A. R. SMITH, AND M. HASEBE. 1998. Phyloge-
netic studies of extant pteridophytes. In D. E. Soltis, P. S. Soltis, and J.
J. Doyle [eds.], Molecular systematics of plants II, 541–556. Kluwer,
Boston, Massachusetts, USA.
- YANG, Z. 1994. Phylogenetic analysis using parsimony and likelihood meth-
ods. *Journal of Molecular Evolution* 39: 294–307.
- YOSHINAGA, K., Y. KUBOTA, T. ISHII, AND K. WADA. 1992. Nucleotide se-
quence of *atpB*, *rbcL*, *trnR*, *dedB* and *psaI* chloroplast genes from a fern
Angiopteris lygodiiifolia: a possible emergence of Spermatophyta lineage
before the separation of Bryophyta and Pteridophyta. *Plant Molecular
Biology* 18: 79–82.