Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers

Shu-Miaw Chaw*†‡, Christopher L. Parkinson†‡, Yuchang Cheng*, Thomas M. Vincent†§, and Jeffrey D. Palmer†¶

*Institute of Botany, Academia Sinica, Taipei 11529, Taiwan; and †Department of Biology, Indiana University, 1001 East Third Street, Bloomington, IN 47405

Edited by Peter H. Raven, Missouri Botanical Garden, St. Louis, MO, and approved November 11, 1999 (received for review July 15, 1999)

Phylogenetic relationships among the five groups of extant seed plants are presently quite unclear. For example, morphological studies consistently identify the Gnetales as the extant sister group to angiosperms (the so-called ''anthophyte'' hypothesis), whereas a number of molecular studies recover gymnosperm monophyly, and few agree with the morphology-based placement of Gnetales. To better resolve these and other unsettled issues, we have generated a new molecular data set of mitochondrial small subunit rRNA sequences, and have analyzed these data together with comparable data sets for the nuclear small subunit rRNA gene and the chloroplast *rbcL* gene. All nuclear analyses strongly ally Gnetales with a monophyletic conifers, whereas all mitochondrial analyses and those chloroplast analyses that take into account saturation of third-codon position transitions actually place Gnetales *within* conifers, as the sister group to the Pinaceae. Combined analyses of all three genes strongly support this latter relationship, which to our knowledge has never been suggested before. The combined analyses also strongly support monophyly of extant gymnosperms, with cycads identified as the basal-most group of gymnosperms, *Ginkgo* as the next basal, and all conifers except for Pinaceae as sister to the Gnetales + Pinaceae clade. According to these findings, the Gnetales may be viewed as extremely divergent conifers, and the many morphological similarities between angiosperms and Gnetales (e.g., double fertilization and flower-like reproductive structures) arose independently.

Extant seed plants (angiosperms and four groups of gymnosperms: cycads, conifers, *Ginkgo*, and Gnetales) differ from all other living land plants by several characters, the most notable, of course, being reproduction via seeds. Each of the five groups of seed plants is generally thought to be monophyletic; however, relationships among the groups are controversial. A common theme of most morphological studies of seed plant phylogeny is that extant gymnosperms are not monophyletic, with the Gnetales (*Ephedra, Gnetum,* and *Welwitschia*) being the sister group of angiosperms (1–7). Some studies have even concluded that angiosperms arose from within the Gnetales (8, 9). The anthophyte hypothesis, that angiosperms and Gnetales (plus the extinct Bennettitales and *Pentoxylon*) form a monophyletic group, is a major basis for understanding character evolution leading to flowering plants. For example, this hypothesis fits nicely with Friedman's studies (10, 11) showing that Gnetales, like angiosperms, undergo a kind of double fertilization (but without formation of triploid endosperm). Relationships among the three remaining gymnosperm groups vary depending on which morphological characters are used and the sampling of fossil taxa (1–9, 12); the most consistent result is that cycads tend to be the basal-most seed plants.

Molecular studies have generated an even more diverse set of phylogenetic hypotheses for seed plants, especially with respect to the position of the Gnetales. Some molecules and analyses place Gnetales as sister to angiosperms (13, 14), consistent with the anthophyte hypothesis; others place them at the base of seed plants (13, 15), and still others place them within gymnosperms (16–21). These last studies either place Gnetales as sister to conifers within a monophyletic gymnosperms (16–19) or are unable to resolve overall issues of gymnosperm phylogeny because cycads and *Ginkgo* were not sampled (20, 21).

Relationships among the five groups of extant seed plants, including the placement of the Gnetales and the related issue of gymnosperm monophyly, should therefore be regarded as unsettled. More data are evidently needed to better resolve seed plant relationships. To this end, we present analyses of a molecular data set of mitochondrial small subunit rDNA sequences, together with separate and combined analyses of mostly published *rbcL* and nuclear small subunit rDNA data.

Materials and Methods

Total DNA was extracted (22) from 35 plants, whose voucher information and names are at http://nutmeg.bio.indiana.edu/Palmerland/index.html. This site also contains PCR conditions, primers used for PCR and sequencing, and all alignments, with excluded characters and taxon substitutions used to facilitate combined analyses. PCR products were gel isolated and cloned by using either the TA or the TOPO TA cloning kits (Invitrogen). Both strands were sequenced for all 34 mitochondrial small subunit (mtSSU) rDNA genes sequenced. To test for introns, mtSSU cDNAs were generated by reverse transcription–PCR (17). Five new *rbcL* sequences and a single new nuclear small subunit (nuSSU) rDNA sequence were determined by using published protocols (15, 17).

mtSSU rDNA sequences were aligned by using the secondary structure model of *Zea* as a guide, nuSSU rDNA sequences were aligned (17), and *rbcL* sequences were aligned with amino acid translation. Maximum parsimony (MP) analyses employed PAUP* version d56 (23) and used a heuristic search, random addition (100 replicates), and tree bisection with reconnection branch swapping. Maximum likelihood (ML) analyses employed fastDNAml (24) and used the F84 model of Felsenstein (25), with the initial ti/tv ratio estimated with PUZZLE (version 4.02) under the Tamura-Nei model of evolution with the parameter ''estimation'' set to ''approximate'' (26). Ten initial ML trees were inferred by randomizing ''input'' order with jumble, and by using ''global'' swapping across all nodes (equivalent to subtree-pruning-regrafting). The optimal tree (best log-likelihood score) was then input into PAUP* (23) to reoptimize the ti/tv ratio by

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: mtSSU, mitochondrial small subunit; nuSSU, nuclear small subunit; MP, maximum parsimony; ML, maximum likelihood; BS, bootstrap support; KH, Kishino-Hasegawa.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF1610862–AF161096 and AB029351–AB029370).

‡S.-M.C. and C.L.P. contributed equally to this study.

§Present address: Department of Biology, University of California at San Diego, La Jolla, CA 92093.

¶To whom reprint requests should be addressed. E-mail: jpalmer@bio.indiana.edu.

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

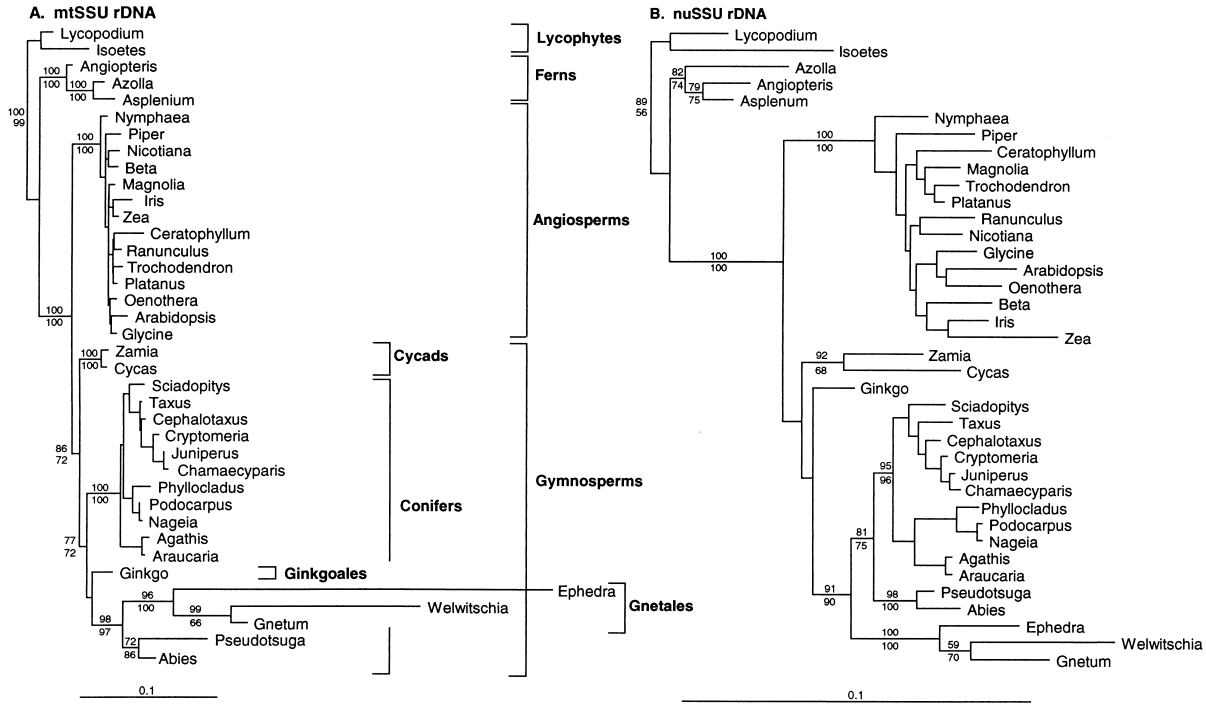

Fig. 1. Phylogenetic analyses of (*A*) mtSSU rDNA and (*B*) nuSSU rDNA. The ML topology (*A*) has a log likelihood of −9,453.548 and was generated using a reoptimized ti/tv ratio of 1.16, whereas the corresponding values for *B* are −10,301.865 and 2.39. BS values >50% are shown above (ML no-rates) and below (MP) all nodes except those within angiosperms and within conifers exclusive of Pinaceae.

using a model that incorporates variability in rates of change. We used the F84 evolutionary model assuming a discrete gamma distribution with four categories of site-to-site rate variability. The resulting ti/tv ratio was used to infer a new tree as above, further optimizing branch lengths. This tree and the optimized ti/tv ratio were then used to estimate evolutionary rates of change for each sequence position by partitioning the sites into 35 "rate" categories with the program DNARATES (http://www.cme.msu.edu/RDP/html/download.html). A new ML tree, incorporating the rate categories and the reoptimized ti/tv ratio, was then inferred. This new optimal tree was then used for a second round of rates estimation and tree inference. This process was iterated until a stable topology was achieved.

Parsimony bootstrapping used PAUP* (23), employing a heuristic search with 100 replications, 10 random additions per replicate, and tree bisection with reconnection swapping. Likelihood bootstrapping used FASTDNAML version 1.06 (24), rewritten in Parallel Virtual Machine language to run in a parallel environment. The no-rates bootstrap data sets were generated with the SEQBOOT program in PHYLIP (25) and then analyzed with FASTDNAML (24), with the input order jumbled for each run and global swapping across all nodes. For ML rates bootstrapping, the individual trees from the 100 ML no-rates pseudoreplicates were each used as a starting tree for two rounds of rates categorization and tree inference with DNARATES as described above, with the final 100 trees used to generate bootstrap values.

Results

mtSSU rDNA Tree Inference. The mtSSU rDNA data set comprises 38 nearly full-length sequences, 34 of which were generated as part of this study. The numerous insertions and deletions present in the mtSSU alignment (one of which was determined to be an intron; unpublished data) were excluded from all phylogenetic analyses. Sequences of the Gnetales, especially *Ephedra*, are particularly divergent in overall length and number of insertions, as well as primary sequence (Fig. 1*A*). Of the 1,595 mtSSU characters used in phylogenetic analyses, 864 were invariant; 731 were variable; and 349 were parsimony informative.

All trees in this study are rooted on two lycophytes, thought to be the basal-most vascular plants (27, 28), and three ferns. Parsimony analysis of the mtSSU data set yielded 490 shortest trees, which differ only within angiosperms. Two kinds of ML analyses were conducted, one with and one without the use of the DNARATES program to account for site-to-site rate variability. The two ML analyses and the MP analysis yielded very similar topologies for the mtSSU rDNA data set. All of the differences are again within angiosperms, except for the placement of *Ginkgo*, which is sister to a Gnetales–Pinaceae clade in both likelihood analyses (Fig. 1*A*), but is a node deeper (the sister to Gnetales and all conifers) in the MP analysis. Given their overall similarities, and to simplify presentation, we have shown only one of the trees (ML no-rates, Fig. 1*A*), but have included on it selected bootstrap support (BS) values for both this and the MP analysis.

Monophyly of seed plants is strongly supported (100% BS) in both ML and MP analyses, as is monophyly of angiosperms (100%), cycads (100%), and Gnetales (≥96%). Refuting the morphology-based anthophyte hypothesis, (extant) gymnosperms are monophyletic (72% and 85% BS; cycads grouped with angiosperms in 27% and 10%, respectively, of the bootstrap trees in which gymnosperms were not monophyletic). Remarkably, the conifers are not monophyletic, and the Gnetales are strongly supported (≥97% BS) as the sister of the two Pinaceae representatives.

This last result is so unexpected, is so at odds with previous analyses of gymnosperm relationships, that we naturally questioned its validity. However, alternative topologies in which the Gnetales were placed at all possible nodes on the tree in Fig. 1*A* were all rejected as significantly worse at the 95% level by using the Kishino-Hasegawa (KH) test (29) except for the Gnetales

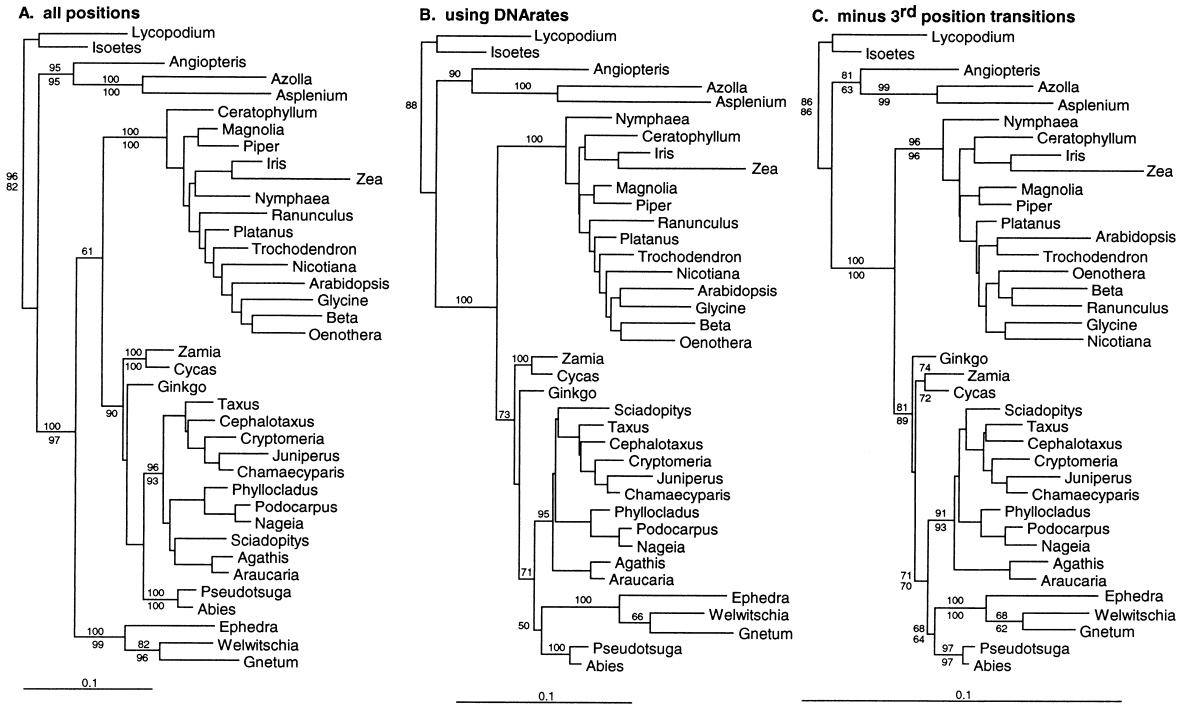

Fig. 2. Maximum likelihood analyses of *rbcL*. (*A*) All positions, no rates correction, ti/tv ratio = 3.14, and log likelihood = −14,561.784. (*B*) All positions, two iterations of DNARATES, ti/tv ratio = 3.14, and log likelihood = −11,910.80. (*C*) Third-position transitions excluded, no rates correction, ti/tv ratio = 0.92, and log likelihood = −7,109.753. BS values >50% are shown above (ML no-rates) and below (MP) all nodes except those within angiosperms and within conifers exclusive of Pinaceae.

placement *within* the Pinaceae, as sister to either *Abies* or *Pseudotsuga*. Noting the exceptionally long-branch lengths leading to both *Ephedra* and *Welwitschia*, we wondered whether the Gnetales placement could reflect a long-branch-attraction artifact. Exclusion of *Ephedra* and *Welwitschia* still yielded, in all three analyses (MP, ML rates, and ML no-rates), a sister-group relationship of Gnetales (i.e., *Gnetum*) and Pinaceae (results not shown). To further evaluate this novel relationship, we performed analyses of largely preexisting data sets, for essentially the same 38 taxa (in a few cases taxonomic substitutes were used), for the nuSSU rDNA and the chloroplast gene *rbcL* gene, as well as for all three genes combined.

nuSSU rDNA Tree Inference. The nuSSU rDNA alignment contained 1,713 positions of unambiguous alignment (out of 1,738 total), of which 1,229 were invariant, 484 were variable, and 290 were parsimony informative. The MP analysis and both ML analyses yielded virtually identical topologies for nuSSU (again excluding numerous differences within angiosperms), and so a single tree is again used to illustrate results (Fig. 1*B*). There are only two differences among these trees involving gymnosperms. First, cycads and *Ginkgo* form the sister group to Gnetales and conifers in the ML rates analysis and in three of the nine shortest MP trees, whereas cycads are the basal-most gymnosperms, with *Ginkgo* the next branch, in the remaining MP trees and in the ML no-rates analysis (Fig. 1*B*). Second, the gymnosperms are monophyletic in the ML no-rates (48% BS; the other bootstrap trees placed cycads and/or *Ginkgo* at the base of seed plants) and rates analyses and in six of nine MP trees, but are paraphyletic (with cycads at the base of seed plants) in the other three MP trees.

The most important difference between the nuclear and mitochondrial rDNA trees concerns conifers and Gnetales. The strongly supported paraphyly of conifers (and sisterhood of Gnetales and Pinaceae) obtained with mtSSU is not recovered with nuSSU, which instead supports monophyly of conifers (75% and 81% BS; but Gnetales and Pinaceae *did* form a clade in all bootstrap trees in which conifers were not monophyletic). Nonetheless, nuSSU rDNA does refute the anthophyte hypothesis by providing relatively strong support (90% and 91% BS) for a sister-group relationship of Gnetales and conifers. Furthermore, with the notable exception of their placement as sister to the Pinaceae (as in Fig. 1*A*), all alternative placements of the Gnetales (including as sister to angiosperms) were rejected at the 95% significance level by the KH test.

***rbcL* Tree Inference.** The *rbcL* alignment contained 1,321 characters, of which 725 were invariant; 596, variable; and 474, parsimony informative. A heuristic MP search found four most parsimonious trees of length 2,503, which differed only within angiosperms. The MP and both ML analyses yielded generally congruent *rbcL* topologies, the most notable difference being the placement of Gnetales. Parsimony placed Gnetales with angiosperms (not shown), ML without rates placed them as the earliest branch within seed plants (Fig. 2*A*; as did the parsimony *rbcL* analyses 15), and ML with rates placed them as sister to Pinaceae (Fig. 2*B*), as in all mtSSU analyses (Fig. 1*A*).

The radical effect of the DNARATES analysis on Gnetales placement suggested that a subset of *rbcL* sites might be evolving very rapidly and perhaps be saturated. To explore this, we examined transitional and transversional divergence at all three codon positions, by plotting uncorrected sequence divergences against Tamura-corrected divergences (30) for all possible pairwise comparisons (Fig. 3). Significant nonlinearity is seen only for third-position transitions, which by this criterion (31) are saturated at the deep taxonomic levels of interest to this study [use of the term "saturation" should not be taken to imply that third-position homoplasy cannot be useful at lower taxonomic levels (e.g., within angiosperms), especially when combined with

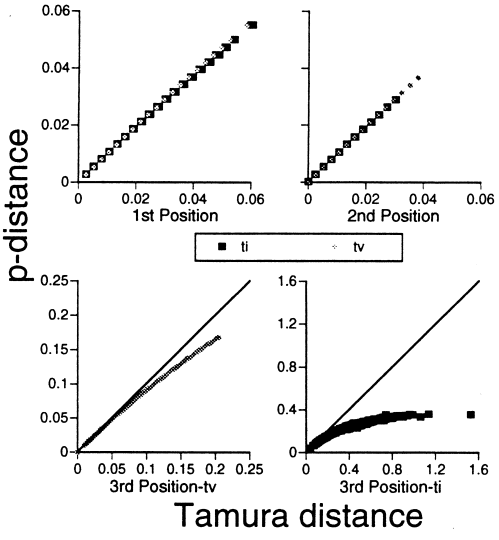

Fig. 3. Plots of uncorrected pairwise sequence divergence (p-distance) versus Tamura-corrected distances for transitions (ti) and transversions (tv) at first-, second-, and third-codon positions. Each plot presents all possible pairwise comparisons (703 data points). Deviation from the $x = y$ line in the bottom two plots is a measure of the degree of saturation for the indicated class of substitution.

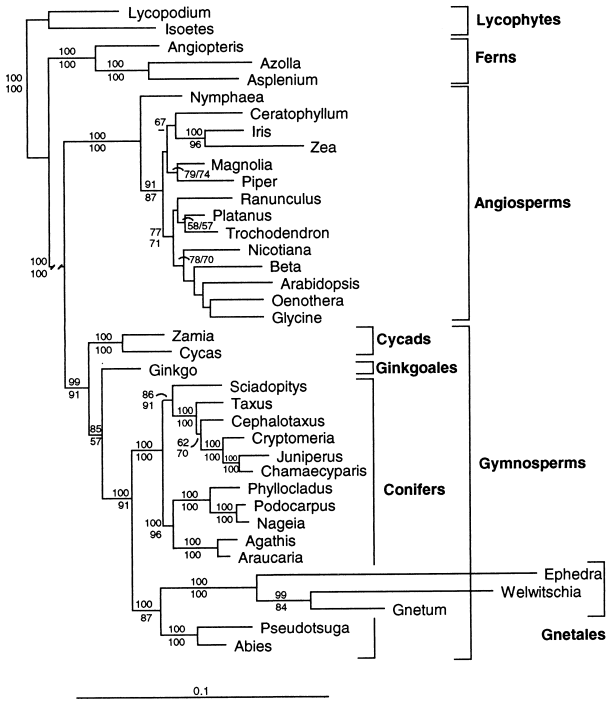

Fig. 4. Analysis of a combined data set of all three genes. The ML tree shown has a log likelihood = −35,774.90 and was estimated by using a ti/tv ratio = 2.1. All BS values >50% are shown above (ML no-rates) and below (MP) nodes.

much denser taxonomic coverage]. These findings extend the analyses of Goremykin *et al.* (16), who concluded that synonymous sites in *rbcL* are effectively saturated at these taxonomic levels, but who did not address the relative contributions of transitions vs. transversions to this saturation.

Accordingly, analyses were performed in which third-position transitions were excluded by recoding all third-position nucleotides as either R (for A or G) or Y (for C or T). In agreement with the *rbcL* rates analysis (Fig. 2*B*) and all mtSSU analyses (Fig. 1*A*), Gnetales emerge as the sister to Pinaceae in the recoded MP and ML analyses (64% and 68% BS, respectively; Fig. 2*C*). Gymnosperms are also monophyletic in the ML rates and recoded *rbcL* analyses (82% and 89% BS; Figs. 2 *B* and *C*). All possible alternative placements of the Gnetales in ML no-rates analyses were examined by using the KH test to see if they were significantly worse than the best trees shown in Figs. 2 *A* and *C*. When all positions were included, numerous alternative placements of the Gnetales, including as sister to angiosperms *or* the Pinaceae, could not be rejected at the 95% level. When third-position transitions were excluded, the anthophyte placement could not be rejected as significantly worse than the best tree, in which the Gnetales were sister to the Pinaceae.

Combined Analyses. The overwhelming similarity of the three sets of individual gene trees indicates that they are fundamentally congruent and therefore can justifiably be combined. The combined data set contained 4,269 unambiguously aligned characters, of which 2,818 were invariant; 1,811, variable; and 1,113, parsimony informative. All three analytical methods gave identical topologies, except for differences within angiosperms. Moreover, virtually all nonangiosperm nodes were strongly supported (Fig. 4). Because the combined tree will serve as the basis for most of the *Discussion*, we will say little about its specifics here. The two most notable points are that (extant) gymnosperms are monophyletic with high support (91% and 99% BS, and a decay value of 14 steps in the MP analyses), and that Gnetales strongly ally with Pinaceae (87% and 100% BS, and a decay of 12 steps). Furthermore, all alternative placements of the Gnetales were rejected at the 99% level by the KH test, and moving the Gnetales to a sister-group relationship with the angiosperms led to a tree fully 24 steps longer than the shortest MP tree (of 5,265 steps).

Molecular phylogenetic analyses of ancient groups are necessarily limited to extant taxa, which is unfortunate for seed plants considering how many important extinct groups they include (1–9). We did, however, examine whether our results, especially the placement of Gnetales, are sensitive to the inclusion of key extant groups. Because our most provocative finding is the sisterhood of Gnetales and Pinaceae, we carried our MP analyses with Pinaceae excluded: Gnetales still grouped, strongly, with the remaining conifers. Exclusion, separately or together, of angiosperms and the outgroups, which are the only long-branched groups other than Gnetales, also did not affect the topology obtained within gymnosperms.

Discussion

The major phylogenetic conclusions from this study are: (*i*) Gnetales are not the sister group to angiosperms among extant seed plants; (*ii*) Gnetales are a monophyletic group; (*iii*) extant gymnosperms are also monophyletic; (*iv*) cycads are the basal group of gymnosperms; (*v*) conifers and Gnetales together comprise a monophyletic group; and (*vi*) Gnetales are nested within a paraphyletic conifers as sister group to the Pinaceae. Except for the basal position of cycads, all of these conclusions receive strong bootstrap support in the three-gene analyses of Fig. 4. Aside from the last, and certainly most provocative conclusion, all of these have been reached in one or more previous studies. However, rarely has such strong support been evident for more than one or two of these conclusions, and never (except for the accompanying paper by Bowe *et al.*; ref. 32) has such a well-resolved framework hypothesis of seed plant phylogeny been produced. Studies relevant to and the evolutionary implications of conclusions *i–iii* are treated in the next section,

followed by discussion of conclusions *v* and *vi*, and lastly, treatment of conclusion *iv* and other implications of our findings.

Monophyly of Extant Gymnosperms and Demise of the Anthophyte Hypothesis. Molecular data are rapidly converging on a clear rejection of the anthophyte hypothesis (that Gnetales are sister to the angiosperms among extant seed plants) in favor of the view that extant gymnosperms, including Gnetales, are a monophyletic lineage to the exclusion of angiosperms. Monophyly of extant gymnosperms is strongly supported in the three-gene/three-genome analyses of Fig. 4. Bowe *et al.* (32) also find considerable support for gymnosperm monophyly with the mitochondrial genes *cox1* and *atpA*, and varying levels of support are found with five other sequences (mitochondrial *cox3*, nuclear legumin and LEAFY, and chloroplast *rpoC1* and rDNA spacer; refs. 16, 19, 33, 34, and M. Frohlich, personal communication). Thus, ten different sequence data sets, distributed across the three plant genomes, provide consistent, sometimes quite strong, support for monophyly of extant gymnosperms. Furthermore, studies that included a number of additional chloroplast (20) or nuclear (21) genes, while unable to address the question of gymnosperm monophyly owing to the absence of cycads and *Ginkgo*, did reject the anthophyte hypothesis by consistently recovering a Gnetales–conifer pairing.

With the demonstrations in this study and in Goremykin *et al.* (16) that *rbcL* does not support the anthophyte hypothesis when saturated sites are excluded, the only molecule for which current analyses still support the anthophyte hypothesis is nuclear large subunit rDNA (14). However, it should be stressed that Stefanovic *et al.* (14) used a relatively short length of this molecule (638 bp), that their analysis found only weak support for an anthophyte clade (BS $<50\%$, decay = 1), and that our reanalysis of their data shows that a variety of alternative placements of the Gnetales, including with other gymnosperms, cannot be rejected by the KH test (data not shown). Overall then, we regard the growing molecular database as providing very strong support for extant gymnosperm monophyly and against the anthophyte hypothesis. Furthermore, the three molecules examined in this study, as well as the five other molecules examined elsewhere (14, 16, 32–34), all firmly reject the morphological cladistic hypotheses of Nixon *et al.* (8) and Hickey and Taylor (9) that Gnetales are paraphyletic (with angiosperms arising from within Gnetales).

The demise of the anthophyte hypothesis, coupled with the well-nested placement of Gnetales within gymnosperms (see next section), means that those morphological and ultrastructural traits that have been regarded by some or most authors (1–9, 12) as uniquely shared by Gnetales and angiosperms were most likely derived separately in the two groups. Such traits include flower-like reproductive structures, lignin chemistry, a tunica in the apical meristem, pollen with granular exine, reduction of the megaspore wall, and vessel-like conducting elements (see refs. 2 and 35 for discussion of the differences between vessels in Gnetales and angiosperms and of the relevance of their absence in such extinct groups as Bennettitales). Most notably, the process of double fertilization in Gnetales (10, 11) almost certainly arose separately from, and thus is not homologous with, the classical double fertilization of angiosperms.

Monophyly of gymnosperms only heightens Darwin's "abominable mystery" concerning the origin of angiosperms. Although extant gymnosperms and angiosperms should now be regarded as sister groups, gymnosperms have a much older fossil record ($\approx$320 million years for the clade including all modern gymnosperms; ref. 6) than angiosperms (at most 130 million years; ref. 4). Are angiosperms substantially older than the current fossil record indicates, or did they arise in the Jurassic, from the Bennettitales, Caytoniales, or some other group of extinct seed plants?

Gnetales as a Sister Group to Conifers and Probably Even Pinaceae. Individually, the three genes analyzed in this study all place the Gnetales as sister to either all conifers (nuSSU, Fig. 1*B*) or specifically the Pinaceae (mtSSU, Fig. 1*A*, and *rbcL*, Figs. 2 *B* and *C*), one of the two fundamental groups of conifers as defined by all molecular studies. Our combined analyses of all three genes strongly support a sister-group relationship of Gnetales and Pinaceae, and the multigene studies of Bowe *et al.* (32) and Qiu *et al.* (36) also find considerable support for this relationship [these three studies include a total of seven different genes (four mitochondrial, two chloroplast, and one nuclear) and over 10,000 bp). The Gnetales/Pinaceae clade was also recovered in ML analyses of all positions and in MP analyses of first and second positions of a combined data set of two other chloroplast genes (M. J. Sanderson, personal communication). Finally, analyses of mitochondrial *cox3* (32), chloroplast *rpoC1* (19), and chloroplast rDNA spacer (16) are all consistent with the sisterhood of Gnetales and Pinaceae, but with the all-important caveat that no other conifers were included in these studies.

A generalized association of Gnetales and coniferopsids (conifers plus *Ginkgo* and the extinct cordaites) has been suggested before based on several lines of evidence, such as xylary pit anatomy, compound strobili, and simple, linear leaves (2, 35). A specific affiliation of Gnetales and conifers is also supported by a derived chloroplast gene order (L. Raubeson, personal communication). However, to our knowledge, no one has ever suggested before that the Gnetales arose from within the conifers. Although this hypothesis, which we hereby name the "gnepines" hypothesis, is strongly supported by our data and those of Bowe *et al.* (32) and Qiu *et al.* (36), it nonetheless requires close scrutiny and testing with more molecular data.

Mixed support for the gnepines hypothesis comes from chloroplast genome architecture. This hypothesis is supported by the correlated loss of all *ndh* genes from chloroplast genomes of Gnetales and Pinaceae (but not other conifers; ref. 37 and M. Ireland, H. Deiderick, and J. D. Palmer, unpublished data), but contradicted by the loss of the large inverted repeat from all conifers (but not Gnetales; ref. 38). If the gnepines hypothesis is correct, then either Pinaceae and other conifers lost the inverted repeat independently (other losses are known in angiosperms; ref. 39) or the repeat was lost in the common ancestor of conifers and Gnetales and then regained in the latter group. Two other structural features that may turn out to support the gnepines clade are a 3-bp deletion in the chloroplast rDNA spacer (16) and a single amino acid insertion in *rpoC1* (19); these are shared by Gnetales and Pinaceae to the exclusion of *Ginkgo*, cycads, and angiosperms, but, critically, are unexamined in other conifers.

If the gnepines hypothesis does stand the test of time, then it demands major reinterpretation of the evolution of conifers and Gnetales. Those traits that currently define the conifers are either not true synapomorphies, i.e., were independently derived in the Pinaceae (after their divergence from Gnetales) and in the common ancestor of all other conifers, or else the Gnetales have undergone such extensive nonmolecular divergence as to have lost most traces of their coniferalean ancestry. Precedent for the latter possibility exists in the form of Taxaceae (yews and relatives). Although unquestionably true conifers by most morphological and all molecular criteria (e.g., refs. 5, 40, and 41, and Figs. 1, 2, and 4), Taxaceae have lost the typical coniferalean cone by reduction to a single-terminal ovule surrounded by a fleshy, berry-like red aril. By analogy, Gnetales may have lost such conifer-defining features as narrowly triangular (one-veined) leaves, resin canals, a tiered proembryo, and flat, woody ovuliferous cone scales (1, 5, 40) but have retained those more generalized coniferopsid features listed two paragraphs above. The specific association of Gnetales with Pinaceae is, however, much more difficult to rationalize morphologically.

Extensive morphological divergence of Gnetales from the

coniferalean ground plan is paralleled, perhaps coincidentally, by generally high rates of molecular evolution in the group. Some or all Gnetales form long branches in most molecular trees, most prominently with mitochondrial genes (Figs. 1, 2 *B* and *C*, and 4; refs. 16, 19, 20, 32, and 36). We cannot rule out the possibility that the surprising placement of Gnetales as sister to Pinaceae is a long-branch artifact. However, we consider this unlikely because (*i*) analyses in which the exceptionally divergent mtSSU sequences of *Ephedra* and *Welwitschia* were excluded still recovered the gnepines group with strong support, and (*ii*) such artifacts are generally manifest as the artifactually deep placement of a long-branch group (often by attraction to a long outgroup branch), whereas in gnepines trees, Gnetales have a relatively nested position and fail to branch with either the long outgroup branch (this is shown foreshortened for space reasons in Fig. 4, but was actually slightly longer than the angiosperm branch) or the long branch leading to angiosperms. Conversely, it is precisely this behavior (the grouping of Gnetales with either angiosperms or the outgroup) that occurs in those *rbcL* analyses in which third-position transitions are not down-weighted (Fig. 2*A* and *Results*) and which we think *is* an artifact of long-branch attraction [similarly, assuming the gnepines hypothesis is correct, we would interpret the monophyly of conifers in nuSSU analyses (Fig. 1*B*) as reflecting attraction of the long Gnetales branch to the relatively long branch (compared with other molecules) between conifers and *Ginkgo*].

Other Relationships Within Seed Plants. Most of our single-gene analyses (Figs. 1 and 2) place cycads as the deepest branch of gymnosperm evolution, with this result enjoying modest support in the combined analyses (Fig. 4). This result is consistent with the multigene analyses of Bowe *et al.* (32) and with many but not all morphological cladistic studies (reviewed in ref. 6), which regard *Ginkgo* (the next deepest branch of gymnosperms according to our results) and conifers as sharing such traits as fertile short shoots, simple leaves, and aspects of wood anatomy. However, it is in conflict with a seemingly unique chloroplast genome rearrangement shared by cycads and *Ginkgo* (42).

Relationships within the main group of conifers (all but Pinaceae) are highly supported at almost all nodes (Fig. 4). Our conifer topology is entirely congruent with that obtained in the nuSSU study of Chaw *et al.* (17), which featured better taxonomic sampling. For these reasons and because of space limitations, we defer to Chaw *et al.* for a discussion of the evolutionary implications of these results.

In contrast to the well-supported phylogeny of gymnosperms obtained in the combined analysis (Fig. 4), relationships within angiosperms are relatively poorly supported (but are nonetheless congruent with those found in more extensive studies; e.g., refs. 36 and 43). This is consistent with the fossil evidence that angiosperms underwent an explosive radiation relatively early in their evolution (4). Of greatest note, the strongly supported basal position of *Nymphaea* (Fig. 4) has now been confirmed and extended by more extensive studies by us (44) and others (36, 45), which place the Nymphaeales as the second earliest branch of angiosperms, after *Amborella*.

We thank Y.-L. Qiu for several DNAs; D. E. Soltis for the *Pseudotsuga* nuSSU sequence; C. dePamphilis, M. W. Frohlich, S. Mathews, Y. L. Qiu, L. A. Raubeson, M. J. Sanderson, and D. E. Soltis for sharing unpublished manuscripts and/or data; W. M. Fischer for Perl scripts and Unix syntax help; W. M. Fischer and S. Turner for lively discussions on phylogenetic analyses; and J. A. Doyle, W. E. Friedman, and R. G. Olmstead for critically reading the manuscript. This work was supported by National Service Center Grants 3519F and 87–2311-B001–075 and an Academia Sinica Grant to S.-M.C.; National Institutes of Health Fellowship GM-19225 to C.L.P.; and National Institutes of Health Grant GM-35087 to J.D.P.

1. Crane, P. R. (1985) *Ann. Mo. Bot. Gard.* **72,** 716–793.
2. Doyle, J. A. & Donoghue, M. J. (1986) *Bot. Rev.* **52,** 321–431.
3. Rothwell, G. W. & Serbet, R. (1994) *Syst. Bot.* **19,** 443–482.
4. Crane, P. R., Friis, E. M. & Pedersen, K. R. (1995) *Nature (London)* **347,** 27–33.
5. Doyle, J. A. (1996) *Int. J. Plant Sci.* **157**, Suppl., S3–S39.
6. Doyle, J. A. (1998) *Mol. Phylogenet. Evol.* **9,** 448–462.
7. Doyle, J. A. (1998) *Annu. Rev. Ecol. Syst.* **29,** 567–599.
8. Nixon, K. C., Crepet, W. L., Stevenson, D. & Friis, E. M. (1994) *Ann. Mo. Bot. Gard.* **81,** 484–533.
9. Hickey, L. J. & Taylor, D. W. (1996) in *Flowering Plant Origin, Evolution & Phylogeny*, eds. Taylor, D. W. & Hickey, L. J. (Chapman & Hall, New York), pp. 176–231.
10. Friedman, W. E. (1990) *Science* **247,** 951–954.
11. Friedman, W. E. (1994) *Am. J. Bot.* **81,** 1468–1486.
12. Loconte, H. & Stevenson, D. W. (1990) *Brittonia* **42,** 197–211.
13. Hamby, R. K. & Zimmer, E. A. (1992) in *Molecular Systematics of Plants*, eds. Soltis, P. S., Soltis, D. E. & Doyle, J. J. (Chapman & Hall, New York), pp. 50–91.
14. Stefanovic, S., Jager, M., Deutsch, J., Broutin, J. & Masselot, M. (1998) *Am. J. Bot.* **85,** 688–697.
15. Albert, V. A., Backlund, A., Bremer, K., Chase, M. W., Manhart, J. R., Mishler, B. D. & Nixon, K. C. (1994) *Ann. Mo. Bot. Gard.* **81,** 534–567.
16. Goremykin, V., Bobrova, V., Pahnke, J., Troitsky, A., Antonov, A. & Martin, W. (1996) *Mol. Biol. Evol.* **13,** 383–396.
17. Chaw, S. M., Zharkikh, A., Sung, H. M., Lau, T. C. & Li, W. H. (1997) *Mol. Biol. Evol.* **14,** 56–68.
18. Soltis, P. S., Soltis, D. E., Wolf, P. G., Nickrent, D. L., Chaw, S.-M. & Chapman, R. L. (1999) *Mol. Biol. Evol*. **16,** 1774–1784.
19. Samgiullin, T. K., Martin, W. F., Troitsky, A. V. & Antonov, A. S. (1999) *J. Mol. Evol.* **49,** 310–315.
20. Hansen, A., Hansmann, S., Samigullin, T., Antonov, A. & Martin, W. (1999) *Mol. Biol. Evol.* **16,** 1006–1009.
21. Winter, K. U., Becker, A., Münster, T., Kim, J. T., Saedler, H. & Theissen, G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7342–7347.
22. Qiu, Y. L., Cho, Y., Cox, J. C. & Palmer, J. D. (1998) *Nature (London)* **394,** 671–674.
23. Swofford, D. L. (1999) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
24. Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994) *Comput. Appl. Biosci.* **10,** 41–48.
25. Felsenstein, J. (1993) *PHYLIP: Phylogeny Inference Package* (University of Washington, Seattle, WA).
26. Strimmer, K. & von Haeseler, A. (1996) *Mol. Biol. Evol.* **13,** 964–969.
27. Kenrick, P. & Crane, P. R. (1997) *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institute, Washington, DC).
28. Raubeson, L. A. & Jansen, R. K. (1992) *Science* **255,** 1697–1699.
29. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29,** 170–179.
30. Tamura, K. (1992) *Mol. Biol. Evol.* **11,** 154–157.
31. Moritz, C., Schneider, C. & Wake, D. B. (1992) *Syst. Biol.* **41,** 273–291.
32. Bowe, L. M., Coat, G. & dePamphilis, C. W. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4092–4097.
33. Malek, O., Lättig, K., Hiesel, R., Brennicke, A. & Knoop, V. (1996) *EMBO J.* **15,** 1403–1411.
34. Shutov, A. D., Braun, H., Chesnokov, Y. V., Horstmann, C., Kakhovskaya, I. A. & Baumlein, H. (1998) *J. Mol. Evol.* **47,** 486–492.
35. Carlquist, S. (1996) *Int. J. Plant Sci.* **157**, Suppl., S58–S76.
36. Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D., Soltis, P., Zanis, M., Zimmer, E., Chen, Z., Savolainen, V. & Chase, M. (1999) *Nature (London),* **402,** 404–407.
37. Tsudzuki, J., Nakashima, K., Tsudzuki, T., Hiratsuka, J., Shibata, M., Wakasugi, T. & Sugiura, M. (1992) *Mol. Gen. Genet.* **232,** 206–214.
38. Raubeson, L. A. & Jansen, R. K. (1992) *Biochem. Syst. Ecol.* **20,** 17–24.
39. Downie, S. R. & Palmer, J. D. (1992) in *Molecular Systematics of Plants*, eds. Soltis, P. S., Soltis, D. E. & Doyle, J. J. (Chapman & Hall, New York), pp. 14–35.
40. Hart, J. (1987) *J. Arnold Arbor. Harv. Univ.* **68,** 269–307.
41. Chaw, S. M., Sung, H. M., Long, H., Zharkikh, A. & Li, W. H. (1995) *J. Mol. Evol.* **41,** 224–230.
42. Raubeson, L. A. & Jansen, R. K. (2000) *Am. J. Bot*., in press.
43. APG (1998) *Ann. Mo. Bot. Gard.* **85,** 531–553.
44. Parkinson, C. L., Adams, K. L. & Palmer, J. D. (1999) *Curr. Biol*. **9,** 1485–1488.
45. Mathews, S. & Donoghue, M. J. (1999*) Science* **286,** 947–950.