Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers

L. Michelle Bowe*, Gwénaële Coat†, and Claude W. dePamphilis†‡

*Department of Biology, Frostburg State University, Frostburg, MD 21532; and †Department of Biology, Institute of Molecular Evolutionary Genetics, and Life Sciences Consortium, Pennsylvania State University, University Park, PA 16802

Edited by Peter H. Raven, Missouri Botanical Garden, St. Louis, MO, and approved December 28, 1999 (received for review July 15, 1999)

Efforts to resolve Darwin's ''abominable mystery''—the origin of angiosperms—have led to the conclusion that Gnetales and various fossil groups are sister to angiosperms, forming the ''anthophytes.'' Morphological homologies, however, are difficult to interpret, and molecular data have not provided clear resolution of relationships among major groups of seed plants. We introduce two sequence data sets from slowly evolving mitochondrial genes, *cox*1 and *atp*A, which unambiguously reject the anthophyte hypothesis, favoring instead a close relationship between Gnetales and conifers. Parsimony- and likelihood-based analyses of plastid *rbc*L and nuclear 18S rDNA alone and with *cox*1 and *atp*A also strongly support a gnetophyte–conifer grouping. Surprisingly, three of four genes (all but nuclear rDNA) and combined three-genome analyses also suggest or strongly support Gnetales as derived conifers, sister to Pinaceae. Analyses with outgroups screened to avoid long branches consistently identify all gymnosperms as a monophyletic sister group to angiosperms. Combined three- and four-gene rooted analyses resolve the branching order for the remaining major groups—cycads separate from other gymnosperms first, followed by *Ginkgo* and then (Gnetales + Pinaceae) sister to a monophyletic group with all other conifer families. The molecular phylogeny strongly conflicts with current interpretations of seed plant morphology, and implies that many similarities between gnetophytes and angiosperms, such as ''flower-like'' reproductive structures and double fertilization, were independently derived, whereas other characters could emerge as synapomorphies for an expanded conifer group including Gnetales. An initial angiosperm–gymnosperm split implies a long stem lineage preceding the explosive Mesozoic radiation of flowering plants and suggests that angiosperm origins and homologies should be sought among extinct seed plant groups.

The origin of angiosperms has long been considered a fundamental mystery of plant evolution (1–4), and until recently, the main data available for addressing this question came from morphological and anatomical analysis of living and fossil species, with subsequent cladistic analysis. Morphological homologies, however, are notoriously difficult to ascertain, and many of the relevant characters have been interpreted and reinterpreted many times (3–13). Despite this challenge, a consensus has emerged among morphological cladistic analyses that Gnetales—three bizarre and enigmatic seed plant genera (*Welwitschia*, *Gnetum*, and *Ephedra*)—form a clade with angiosperms and various fossil groups (2–14). This ''anthophyte'' clade (2) is ostensibly the only well-supported relationship among the five main extant seed plant groups (13) and has become the subject of a wide range of evolutionary and molecular developmental studies (14–20), several of which have not affirmed a Gnetales–angiosperm relationship (19, 20).

Molecular phylogenies have been largely unable to reach strong conclusions about questions of seed plant evolution other than confirming the monophyly of three of the main groups: Gnetales, cycads, and angiosperms (13). Published chloroplast and nuclear gene phylogenies, some unrooted, have either potentially supported (21–23) or, more often, conflicted (23–26) with the anthophyte hypothesis, and there were indications that *rbc*L at least (25) was saturated at the depth needed to resolve basal seed plant relationships. Several papers using chloroplast *rbc*L and nuclear 18S reached varied conclusions when different taxa, sequence samples, outgroups, or analysis methods were used (13, 23–26), suggesting that the issue was far from settled.

We reasoned that optimal resolution of this question may be obtained by using molecular sequences with slower underlying rates of nucleotide substitution than previously utilized, such as sequences from the mitochondrial genome of plants (27–30). Here we have sampled two mitochondrial protein genes, *cox*1 (cytochrome oxidase I) and *atp*A (= *atp*1, ATPase I), from all extant seed plant lineages, including all widely recognized gymnosperm families. We compare phylogenies for these genes to ones based on plastid and nuclear genes for closely matched taxa, developing (along with Chaw *et al.*, ref. 39) a comprehensive molecular phylogeny of seed plants on the basis of all three plant genomic compartments. Our results strongly conflict with the anthophyte hypothesis, suggesting instead that Gnetales' closest relatives are conifers, and that the extant sister group to angiosperms is all other seed plants.

Methods

Twenty-eight new gene sequences were obtained for mitochondrial *cox1* and 15 for *atp*A (29). Approximately 1,416 bp of *cox*1 and 1,239 bp of *atp*A were PCR amplified and sequenced. Known or presumed RNA editing sites (with nonsynonymous C-T transitions at otherwise conserved amino acid residues) were excluded, as described (30–32). Plastid *rbc*L and nuclear small subunit (18S rDNA) sequences were sampled from the database. Latin names and voucher data, primers, protocols, additional sequence information, as well as taxon substitutions to allow combined multigene analyses, are available at http://depcla4.bio.psu.edu/Seedpl. Aligned sequences and internal primer sequences are available from C.W.D.

Phylogenetic Analyses. Maximum likelihood (ML) analyses of individual data sets were performed by using PAUP* (Ver. 4.0b2) (33), with empirical estimates of base composition. Starting parameters for transition–transversion ratio (ti/tv) and invariant (inv) sites were obtained from initial ML or parsimony (MP) trees and then estimated in formal ML analyses. Gamma, a parameter to account for among-site rate variation, was also

Abbreviations: inv, invariant; ML, maximum likelihood; MP, parsimony; NJ, neighbor-joining; BS, bootstrap analysis; ti/tv, transition/transversion ratio; KH*P*, Kishino—Hasegawa *P*-value.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. *cox*1, AF020556–AF020585; *atp*A, AF209099–AF209113).

‡To whom reprint requests should be addressed. E-mail: cwd3@psu.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.

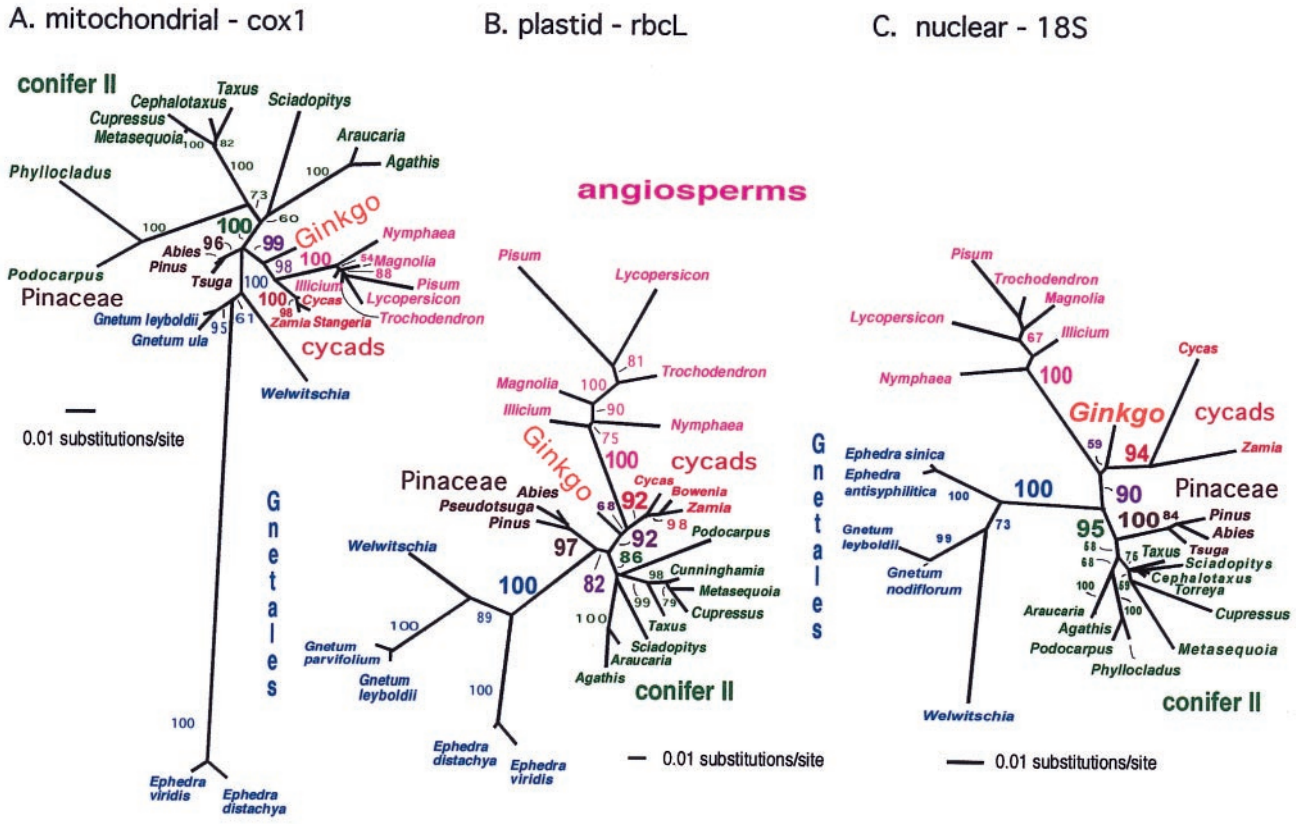

Fig. 1. Molecular phylogenies of seed plants are inconsistent with the anthophyte hypothesis (= gnetophytes + angiosperms) and instead support a close Gnetales–conifer relationship. Unrooted phylograms of seed plant DNA sequences found with PAUP* (heuristic search algorithm, ML) by using Macintosh G3 computers. ML bootstrap percentages on the basis of 100 replicates are also shown on key branches with data and tree characteristics as follows: (*A*) mitochondrial *cox*1: 27 taxa, 1,324 (aligned total) and 325 (parsimony informative) bases; -ln likelihood = 6,196.7314; ti/tv = 1.165; inv sites = 0.261; gamma = 0.698. (*B*) plastid *rbc*L: 26 taxa, 1,329 and 365 bases; -ln likelihood = 8292.7942; ti/tv = 3.03; inv sites = 0.52; gamma = 1.3. (*C*) nuclear 18S rDNA, 27 taxa, 1,715 and 224 bases; -ln likelihood = 6,694.0414; ti/tv = 2.26, inv sites = 0.6199, gamma = 0.624.

estimated in the ML models with inv sites and/or ti/tv held constant. None of the genes exhibited significant heterogeneity in base composition among taxa χ^2 test by using PAUP* 4.0. For each gene, at least 10 random input orders were used, and both the Kishino–Hasegawa–Yano (K–H–Y) model for unequal base frequencies and the Felsenstein model were tested; starting parameters and method for ancestral-state reconstruction were also varied; ML topologies were generally insensitive to these choices. For MP analyses, 250 replicates of random step-wise addition with tree bisection reconnection (TBR) branch-swapping and no weighting were indicated in a heuristic search. Neighbor joining (NJ) was performed with several distance models and parameter values, and the Kimura-2-parameter model with gamma = 0.5 and ML estimates of inv sites were used for analyses shown here unless indicated otherwise. Support for each node was tested with standard bootstrap analysis (BS); 100, 250, and 250 replicates were used for ML, MP, and NJ, respectively. Sequences were also analyzed with and without indel regions, and with several different alignments—no significant differences among trees were found. In the final analyses, five small regions of uncertain alignment were excluded from the 18S analyses, and *rbc*L analyses were restricted to positions 31 to 1,359 from the ATG start codon to include only those portions that were available for Gnetales and other critical taxa.

Detection of Long Branches and Outgroup Selection. Data sets were examined for long branches by using RELATIVE APPARENT SYNAPOMORPHY ANALYSIS (RASA) 2.3.7; refs. 34–36). Initial unrooted analyses were performed with and without putative long-branch taxa. Because rooting of seed plant phylogenies necessarily involves distantly related (nonseed plant) taxa that might easily cause artifactual effects in phylogenetic analyses (13, 24, 34–37), analyses were also performed to screen outgroups for suitability in rooted analyses. Potential outgroups were considered "safe for use" if: (*i*) their addition did not disrupt ingroup topology when compared with unrooted analyses; (*ii*) the assigned outgroup(s) did not cause a decline in tRASA (phylogenetic signal; ref. 34); and (*iii*) the assigned outgroup(s) did not result in a significantly long branch according to RASA taxon-variance plot (36).

Results

Unrooted Analyses. Phylogenetic analysis of mitochondrial *cox*1 sequences (Figs. 1*A* and 2) identifies strongly supported monophyletic groups corresponding to major lineages of seed plants: angiosperms (BS 100%; values pertain to ML trees unless otherwise stated), Gnetales (BS 100%), cycads (BS 100%), Pinaceae (a conifer family; BS 100%), and a group we term Conifer II (BS 100%), containing all conifer families other than Pinaceae. The anthophyte hypothesis (angiosperms + Gnetales), is strongly rejected by the *cox*1 analysis: two well-supported nodes (BS 98% and 99%) separate them on the *cox*1 tree, which groups Gnetales weakly with Pinaceae within the conifers. ML Kishino–Hasegawa (K-H; ref. 38) tests confirm that *cox*1 unrooted and rooted trees have a significantly higher likelihood than trees constrained to include an anthophyte clade

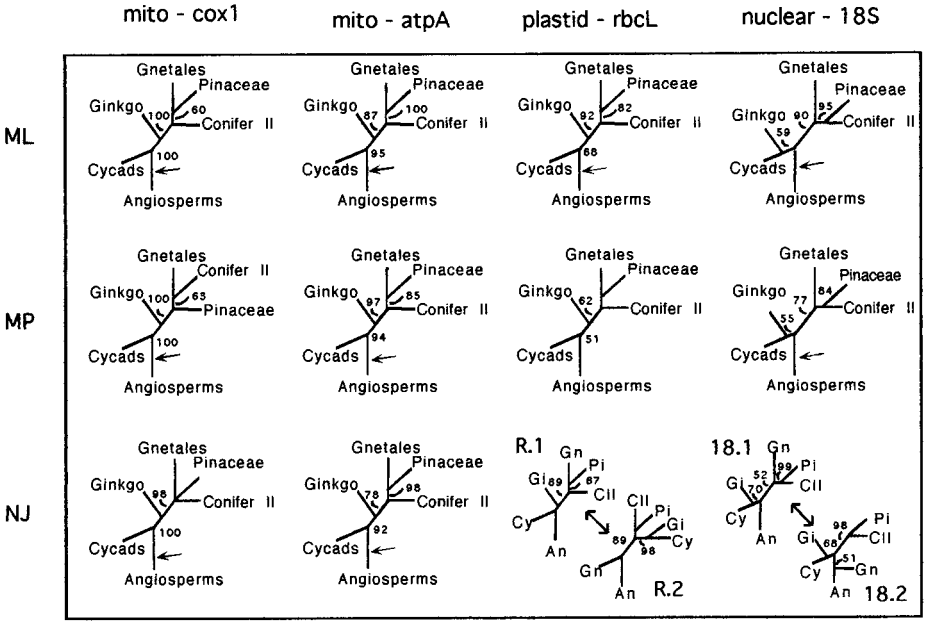

Fig. 2. Alternative phylogenetic analyses of four genes by using ML, MP, and NJ. Analyses are unrooted, with arrows indicating position of root in additional rooted analyses (see text for details). BS above 50% for each analysis are given on internal branches. Taxa and tree characteristics are as Fig. 1, except long-branch taxa (*Ephedra, Podocarpus, Phyllocladus)* are excluded for *cox1*. Additional tree characteristics for each analysis: for *cox*1: 23 taxa, ML: -ln likelihood = 4827.7315 (ti/tv = 1.464, inv sites = 0.335, gamma = 0.758); MP: 1 tree at 502 steps, CI = 0.707, RI = 0.850; NJ: min. evol. score = 0.5217; for *atp*A: 16 taxa, 1,283 and 228 bases; ML: -ln likelihood = 5,118.0457 (ti/tv = 1.783, inv sites = 0.542, gamma = 3.911); MP = 1 tree at 639 steps; CI = 0.655, RI = 0.767; NJ: min. evol. score = 0.8169; for *rbc*L: 26 taxa, MP: 3 trees at 1,329 steps; CI = 0.438, RI = 0.637; NJ: min. evol. score = see text; for 18S: 27 taxa, MP: 2 trees at 780 steps, CI = 0.499, RI = 0.717; NJ: min. evol. = see text.

[K-H *P*-value (KH*P*) = 0.0005 and 0.0016, respectively]. Two ephedras appear to have a greatly accelerated rate of evolution and are highly diverged (≥15% total; ≥30% third position) from other *cox*1 sequences.

A second mitochondrial gene (*atp*A, Fig. 2; details in Fig. 3*B*) also unambiguously rejects the anthophyte hypothesis in favor of a Gnetales–conifer relationship (ML KH*P* = 0.005 for unrooted trees). Although the *atp*A data set is smaller than *cox*1 (16 vs. 27 taxa; 1,283 vs. 1,324 bp), a virtually fully resolved tree of seed plant relationships is obtained for *atp*A with a topology (Fig. 3*B*) identical to *cox*1 analyses. Surprisingly, *atp*A strongly supports (BS 100%) a specific (*Pinus* + *Gnetum*) relationship within conifers.

Plastid *rbc*L and nuclear 18S rDNA identify the same six groups (angiosperms, gnetophytes, cycads, Pinaceae, Conifers II, and *Ginkgo*) with high bootstrap support (Fig. 1). As with *atp*A and *cox*1, *rbc*L associates Gnetales with Pinaceae, implying that conifers are not monophyletic. rDNA trees disagree with *atp*A and *rbc*L on this point (Fig. 1): with 18S, Pinaceae and Conifer II form a strongly supported group (BS 95%) sister to the Gnetales. Also in contrast with the other three genes, 18S weakly places *Ginkgo* and cycads together (BS 59%).

When the sensitivity of these findings to method of phylogenetic reconstruction was examined (Fig. 2), all unrooted ML and MP trees and NJ analyses of *cox*1 and *atp*A provide strong support for a Gnetales–conifer relationship. *cox*1, *rbc*L, 18S, and *atp*A unrooted MP trees require a minimum of 16, 3, 3, and 20 additional steps, respectively, when constrained to have Gnetales and angiosperms together. Alternative analyses with *cox*1, however, provide weak support for Gnetales + Pinaceae (ML), Gnetales + conifer II (MP), or an unresolved trichotomy of the three groups (NJ). NJ analyses for *rbc*L and 18S were found to be sensitive to parameter choice. For example, if no parameter is included to account for site-to-site variation, then trees are obtained (Fig. 2, R.2 and 18.2) that do identify an anthophyte clade, trivially supported with 18S (51% BS, 18.2) but stronger with *rbc*L (89% BS, R.2). Some NJ models that do account for site-to-site variation recover trees identical to those found with ML and MP analysis, e.g., tree R.1 (inv sites = 0.5, gamma = 0.25, 89% BS for Gnetales + Pinaceae + Conifer II) and tree 18.1 (inv sites = 0.5, gamma = 1.0, 52% BS for Gnetales + Conifer). When larger data sets with 20 or 60 phylogenetically balanced angiosperm species representing all major angiosperm lineages were tested, major group patterns (assessed with MP and NJ) were unaltered (not shown).

RASA. ML trees (Fig. 1) suggested large differences in rates of evolution among taxa and across genes for the same taxa. For example, *Ephedra* is on a very long branch for *cox*1, whereas the *Gnetum* branch is much shorter, as is *Ephedra* for the other two genes. For this study, we sought to determine whether rate heterogeneities were large enough to mislead phylogenetic analyses because of potential long-branch attractions (34–37). RASA regression analysis of the *cox*1 data set identified both ephedras, *Phyllocladus* and *Podocarpus,* as long branches potentially disruptive to phylogenetic analysis. When these were removed from the data set, the tRASA test statistic improved from 6.35 (with all species included) to 14.77, and no other significant long branches were detected.

Similar RASA analyses were performed for each of the other data sets. Gnetales and angiosperms were identified as potentially long branches in both *rbc*L and 18S in RASA regressions, but taxon variance ratio tests (details not shown) were not significant, suggesting that these taxa could safely be included in phylogenetic analyses. No significantly long branches were detected in the *atp*A data set or the combined three-gene and four-gene data sets (below). Finally, we performed optimal outgroup analysis (36) using the five potential outgroup taxa

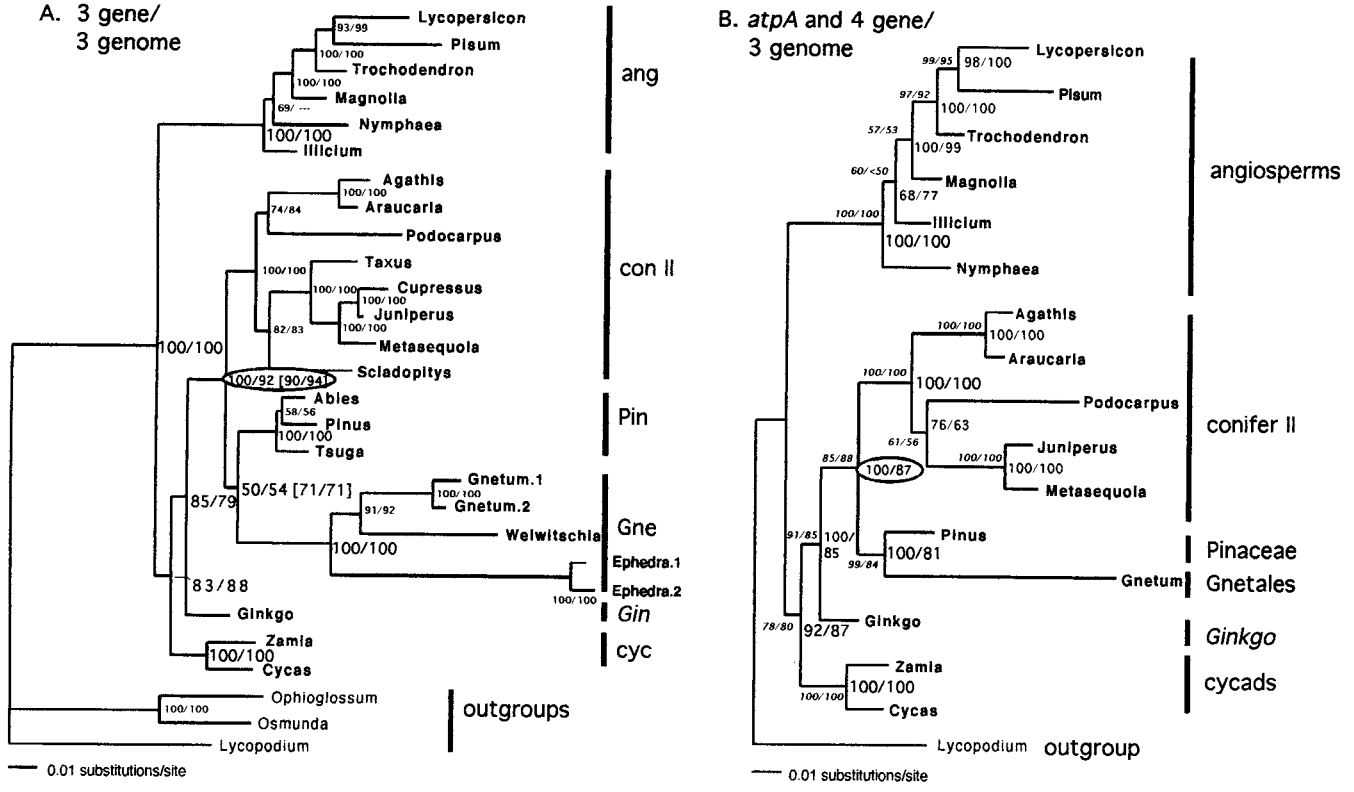

Fig. 3. Rooted seed-plant phylogenies by using three-genome combined data from (*A*) *cox1, rbc*L, and 18S rDNA, and (*B*) including *atp*A. Identical topologies were obtained by using ML, MP (except arrangement of angiosperms, where *Nymphaea* is basal in three-gene MP), and NJ analyses, rooted or unrooted, with or without long-branch *Ephedra*. (*A*) ML/MP bootstraps are shown. Twenty-eight taxa; 4,367 and 1,025 bases; ML: –ln likelihood = 24,645.044 (ti/tv = 2.137, inv sites = 0.427, gamma = 0.685); MP: 3 trees at 3,583 steps (CI = 0.49; RI = 0.67). ML/MP bootstraps also given for two key nodes (in brackets) after deletion of long-branch *Ephedra*. (*B*) ML/MP bootstraps shown for *atp*A alone (*Left*) and all four genes (*Right*). For *atp*A (17 taxa, 1,283 and 244 bases): ML: -ln likelihood = 5,678.420 (ti/tv = 1.774, inv sites = 0.431, gamma = 1.836; MP: 1 tree at 777 steps (CI = 0.621; RI = 0.723). For all 4 genes (5,650 and 920 bases): ML: -ln likelihood = 23,323.1563 (ti/tv = 2.224, inv sites = 0.508, gamma = 0.938); MP: 1 tree at 3,044 steps, CI = 0.549, RI = 0.641.

available for *cox*1: *Marchantia* (a liverwort), *Lycopodium* (a lycophyte), and *Ophioglossum*, *Osmunda,* and *Polystichum* (ferns). *Polystichum* was a significantly long branch for *cox*1, and *Marchantia* was significantly long for *rbc*L and *atp*A. Therefore, we used *Marchantia*, *Lycopodium*, *Ophioglossum*, and *Osmunda* to root *cox*1 trees (Fig. 2), but only the last three for rooting *rbc*L, 18S, and the three-gene combined trees, and *Lycopodium* for *atp*A and four-gene trees (Figs. 2 and 3).

Rooted analyses with individual genes (Fig. 2) and with "optimal" outgroups selected as above consistently placed the root between angiosperms and all gymnosperms in all ML and most MP analyses, implying monophyly of extant gymnosperms. *cox*1 and *atp*A trees also rooted to the same spot in all NJ analyses. Rooted *rbc*L MP analyses placed Gnetales at the base of the seed plants (as in ref. 24), and trees constrained to the anthophyte hypothesis were three steps longer, whereas trees constrained to include Gnetales + all conifers were four steps longer. Rooted *rbc*L ML analyses did not alter ingroup topologies, and this tree was significantly better than one constrained to include a Gnetales–angiosperm group (KH*P* = 0.0276). NJ analyses for *rbc*L and 18S were sensitive to rooting, with Gnetales attracting readily to the outgroup in most cases. Attempts to root *rbc*L and 18S MP and NJ trees with a much broader collection of nonseed plants showed the same sensitivities (not shown). Because of the significant disruption of ingroup topologies in the face of rooting with (necessarily) distant nonseed outgroups, we considered rooting to be unreliable for NJ and MP analyses with *rbc*L and for NJ analyses with 18S, so no root is presented for these specific analyses (Fig. 2).

With the possible exception of the arrangement of Gnetales and conifers in 18S vs. the other genes, the points of difference among the trees are not strongly supported. Partition homogeneity tests in PAUP* (33) indicate that the data sets are not significantly heterogeneous—*P* = 0.17 and 0.10 for three and four gene rooted analyses, and *P* = 0.22 and 0.10 for unrooted analyses.

Analyses of combined *cox*1*, rbc*L, and 18S rRNA data rooted with outgroups *Ophioglossum*, *Lycopodium,* and *Osmunda* show the same major group patterns as *cox*1 ML trees (Fig. 3): well-supported monophyletic angiosperm, gymnosperm, cycad, and Gnetales clades are resolved, and the ingroup topology remains the same as in unrooted trees. Gnetales and conifers form a highly supported clade, and cycads are at the base of the gymnosperms, with *Ginkgo* branching next. Three gene-combined data trees constrained to include the anthophyte clade are 20 steps longer (3,583 vs. 3,603 steps) and have significantly worse likelihood (KH*P* = 0.0035). Relationships among Gnetales and conifers are essentially unresolved (Fig. 1), but when long-branch *Ephedra* is excluded, a Gnetales + Pinaceae clade is found (BS 70%). With all four genes combined and rooted with *Lycopodium*, the same major groups appear, and ML and MP trees are significantly better than trees constrained to the anthophyte hypothesis (3,039 vs. 3,090 steps; KH*P* = 0.0001). *Pinus* + *Gnetum* is resolved with moderate to high support (ML BS 100%; MP BS 79%) in the four-gene rooted analysis.

Discussion

Although earlier studies offered little hope that a molecular consensus would be forthcoming (13), it now appears we are rapidly converging on a broadly based well-supported molecular phylogeny of extant seed plants. Our study indicates several important conclusions that differ from those of recent morphological cladistic studies (13): (*i*) Gnetales are closely related to conifers and, more specifically, they may be derived from within conifers, sister to Pinaceae; (*ii*) rooted phylogenetic trees separate angiosperms from all gymnosperms, implying that extant gymnosperms are monophyletic; and (*iii*) Gnetales are unambiguously monophyletic (unlike refs. 11 and 12). Strong support is also provided for the monophyly of other major seed plant groups: cycads, angiosperms, two principal lineages of conifers (Pinaceae and "conifer II," with all other conifer families), and cycads are resolved as the basal gymnosperm lineage, sister to a clade with *Ginkgo*, conifers, and Gnetales. These same basic findings are also reported by Chaw *et al.* (39), despite differences in taxon sampling, genic region, and methods of analysis.

Early research (21, 23) and one recent data set (nuclear large subunit rDNA; ref. 22) suggested that molecular data might weakly support an anthophyte grouping, but current molecular evidence clearly does not. This includes recent analyses of nuclear 18S data (Figs. 1 and 2; refs. 26 and 39) and analyses of *rbc*L with various methods to account for site-to-site variation and associated saturation of third base positions (Figs. 1 and 2; refs. 25 and 26). In addition, at least four mitochondrial loci [*cox*1 and *atp*A (here), mtSSU rDNA (39), *cox*3 (28), several plastid loci (cpITS, ref. 25, *rpo*C1, ref. 40)], studies of multiple plastid genes (ref. 41; R. Olmstead, personal communication; M. Sanderson, personal communication), and several nuclear genes [leguminins (42) and LEAFY (refs. 19 and 20, and M. Frohlich, personal communication) yield phylogenies inconsistent with an anthophyte clade, with varying taxa and support. Combined evidence from five genes, including mitochondrial *mat*R and *atp*1 (*atp*A), strongly rejects the anthophytes in unrooted trees (43). Finally, an unusual derived gene order in the reduced plastid inverted repeat of conifers (44) is also found in Gnetales (L. Raubeson, personal communication), a pattern in agreement with a conifer–Gnetales relationship suggested by most of the above studies.

Our analyses showed only minor sensitivity to choice of gene or method of phylogenetic analysis, but these may explain some of the differing results of previous studies by using *rbc*L (13, 21, 24, 25) and nuclear 18S sequences (23, 26). Apparently both loci have been subject to a greater frequency of multiple substitutions (homoplasy) during seed plant evolution than have the generally slowly evolving mitochondrial *cox*1 and *atp*A genes. This is illustrated by lower consistency index (CI) values in MP analyses [Fig. 2 and analyses with identical species number per gene (not shown)]. Given the relatively long branches for both gnetophytes and angiosperms in most of the gene trees (Figs. 1 and 3), it is probable that these analyses can be affected by long-branch attraction between these groups (34–37). Addition of distantly related outgroups that act as long-branch attractors (Fig. 2 and refs. 13 and 24) clearly exacerbate the challenge of obtaining a true tree. Here, unrooted analyses, screens for long branches by using RASA, and likelihood analysis with models to account for site-to-site rate variation were used to try to minimize the long-branch attractions (34–37). RASA provided an objective means to identify and justify the exclusion of specific long-branch taxa before phylogenetic analysis, and some taxa excluded by RASA did disrupt topologies. For example, the very long-branch *Ephedra* in the *cox*1 data set apparently misled relationships among the three Gnetales genera (Fig. 1). If the distant outgroup *Marchantia* is added to *rbc*L analyses, then Gnetales are drawn to the base (not shown). In other cases, long-branch taxa did not disrupt topology but did alter bootstrap support values (e.g., Fig. 3*A*). This suggests that studies with distant outgroups must be viewed with caution, and that reanalysis or expansion of earlier seed plant data sets may be fruitful (40–42). Finally, multigene multigenome evidence was used to minimize potential gene-specific and compartment-specific effects on phylogeny. Combined analyses with as many as four genes (5,650 bp per taxon) or ML analyses of individual genes (also see ref. 40) tended to be less sensitive to problems of rooting or differences in taxon choice.

Morphological cladistic analyses have consistently supported a clade with angiosperms and gnetophytes to the exclusion of other extant seed plants, but the molecular hypothesis shown here strongly disagrees. Hypotheses that associate the Gnetales with conifers instead of angiosperms are indicated by all four genes studied, despite marked differences in gene function and genomic location, evolutionary rate (27), RNA editing (32), and mode of inheritance (45). Unless an unaccounted bias is present in these data or analyses, so that the gene trees do not reflect organismal history, then morphological and other characters that associate gnetophytes with angiosperms in one or more cladistic studies [e.g., double integuments, vessels, lignin biochemistry, tunica formation in the apical meristem, pollen wall structure, and lack of archegonia (2–13)] probably were derived independently in these two groups. Double fertilization in angiosperms and some Gnetales, of great interest as possibly homologous in the two groups (4, 10, 15–18), was probably also separately derived. Alternatively, some of the similarities between Gnetales and angiosperms might be ancestral seed plant traits that have been lost several times during seed plant evolution (16).

Many earlier (noncladistic) workers actually considered Gnetales to be close relatives of conifers and convergent with angiosperms (3, 6, 10, 46–52); this view was widely accepted for decades (3). They noted similarities between Gnetales and conifers, including xylem anatomy, cone structure, and the presence of simple linear leaves, inconsistent with the anthophyte hypothesis. Some of these "conifer convergences" may support a Gnetales + Conifer or a Gnetales + Conifer + *Ginkgo* relationship in combined analyses of morphological and molecular data sets.

An unexpected finding is evidence favoring a sister group relationship between Pinaceae and Gnetales, especially in *rbc*L, *atp*A, and four-gene trees. Molecular phylogenies strongly favoring a "gnepines" hypothesis are presented as well by Chaw *et al.* (39) and unrooted five-gene analysis (43). If correct, these phylogenies would imply that Gnetales are derived conifers that have diverged markedly from their common ancestor with Pinaceae. To our knowledge, this surprising hypothesis has not been suggested by any published morphological cladistic study. Molecular evidence for a Gnetales–Pinaceae clade is decidedly mixed: some studies found a Gnetales–Pinaceae group (25, 28, 40), but these used *Pinus* as the sole conifer or lacked other major gymnosperm groups. However, the 18S rDNA data (Figs. 1 and 2; refs. 25 and 39) clearly favor an alternate hypothesis, with Gnetales sister to all extant conifers. In fact, the two resolutions of Gnetales (sister to Pinaceae in most analyses or sister to all conifers in 18S) represent the only strong disagreement seen in these multigene comparisons. The loss of the large inverted repeat from the plastid genome of *Pinus* and other conifers is not shared by Gnetales (44); this would require (under the gnepines hypothesis) either independent losses in Pinaceae and Conifer II or the reacquisition of a large inverted repeat in Gnetales. Although the multigene phylogenies give clear evidence that Gnetales and conifers are close relatives and presently support a Gnetales + Pinaceae relationship, these mixed signals suggest that additional data and analyses are needed to evaluate the alternate hypotheses.

Conclusions and Darwin's "Mystery"

One result of this work is to deepen Darwin's "abominable mystery" of the origin of the flower (1) by separating angiosperms from all extant gymnosperm lineages. Darwin puzzled over the sudden appearance of diverse flowering plants in the fossil record (1), where the earliest definitive angiosperm fossils occur only about 130 million years ago (mya), although much earlier fossils with angiosperm-like features are known (4). Instead of helping to focus our understanding of the origin and diversification of angiosperms, as a Gnetales + angiosperm hypothesis so obviously has (2–20), the molecular phylogeny implies that angiosperms originated some time along a very long stem lineage. This would reach back possibly to the Carboniferous, about 300 mya or more, before joining with the lineage containing all other living seed plants. It also implies (14) that all gymnosperms (or perhaps just the slowly evolving cycads and Pinaceae) are a more appropriate living outgroup for angiosperms than the frequently used (and rapidly evolving) Gnetales. However, our results in no way exclude the possibility that an extinct gymnosperm or seed fern group such as Bennettitales or Caytoniales may be sister to angiosperms as suggested (3, 10, 12, 13, 46–53). This broadens the challenging search for evidence of angiosperm origins and places even greater significance on clear resolution of relationships among the most basal angiosperms (Fig. 3*B* and unpublished data; refs. 43, 54, and 55) and on new fossil evidence (53). Independent origins of Gnetales and angiosperms should prompt reassessment of assumed homologies and might lead to a clearer understanding of how homoplasy and long-branch attractions can challenge phylogenetic analysis of both morphological and molecular data sets.

We dedicate this paper to the fond memory and inspiration of Warren H. (Herb) Wagner. We thank L. Raubeson for unpublished chloroplast DNA structural data; J. Palmer, L. Raubeson, R. Jansen, R. Olmstead, M. Frohlich, D. Soltis, and M. Sanderson for prepublication information; N. Young, G. Chenery, W. Elisens, A. Wolfe, A. G. Moore, T. Barkman, and J. McNeal for several *cox*1 and *atp*A sequences; R. Price for DNAs and *rbc*L sequences; D. Soltis for 18S sequences and initial alignments; and D. Nickrent for two DNAs. N. Friedman, S. Mathews, M. Donoghue, M. Frohlich, O. Pellmyr, B. Swalla, J. Leebens-Mack, N. Young, D. McCauley, T. Barkman, J. Lyons-Weiler, J. McNeal, and especially J. A. Doyle are thanked for helpful comments. This work was supported by grants from the American Society of Plant Taxonomists (to L.M.B.) and the National Science Foundation (to C.W.D.).

1. Darwin, C. (1903) in *More Letters of Charles Darwin: A Record of His Work in a Series of Hitherto Unpublished Letters, Vol. 2,* eds. Darwin, F. & Seward, A. C. (John Murray, London).
2. Donoghue, M. J. (1994) *Ann. Mo. Bot. Gard.* **81,** 405–418.
3. Doyle, J. A. (1978) *Ann. Rev. Ecol. Syst.* **9,** 365–392.
4. Crane, P. R., Friis, E. M. & Pedersen, K. R. (1995) *Nature (London)* **374,** 27–33.
5. Loconte, H. & Stevenson, D. W. (1990) *Brittonia* **42,** 197–211.
6. Doyle, J. A. & Donoghue, M. J. (1986) *Bot. Rev.* **52,** 321–431.
7. Rothwell, G. W. & Serbet, R. (1994) *Syst. Bot.* **19,** 443–482.
8. Crane, P. R. (1985) *Ann. Mo. Bot. Gard.* **72,** 716–793.
9. Doyle, J. A. & Donoghue, M. J. (1992) *Brittonia* **44,** 89–106.
10. Doyle, J. A. (1996) *Int. J. Plant Sci.* **157,** S3–S39.
11. Nixon, K. C., Crepet, W. L., Stevenson, D. & Friis, E. M. (1994) *Ann. Mo. Bot. Gard.* **81,** 484–533.
12. Hickey, L. J. & Taylor, D. W. (1996) in *Flowering Plant Origin, Evolution and Phylogeny*, eds. Taylor, D. W. & Hickey, L. J. (Chapman & Hall, New York), pp. 176–231.
13. Doyle, J. A. (1998) *Ann. Rev. Ecol. Syst.* **29,** 567–599.
14. Doyle, J. A. (1998) *Mol. Phyl. Evol.* **9,** 448–462.
15. Friedman, W. E. (1992) *Science* **255,** 336–339.
16. Friedman, W. E. (1992) *Int. Rev. Cytology* **140,** 319–355.
17. Carmichael, J. S. & Friedman, W. E. (1996) *Am. J. Bot.* **83,** 767–780.
18. Friedman, W. E. & Carmichael, J. S. (1998) *Evolution* **52,** 1016–1030.
19. Frohlich, M. W. & Meyerowitz, E. M. (1997) *Int. J. Plant Sci.* **158,** S131–142.
20. Winter, K.-U., Becker, A., Münster, T., Kim, J. T., Saedler, H. & Theissen, G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 7342–7347.
21. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, D. B., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., *et al.* (1993) *Ann. Mo. Bot. Gard.* **80,** 528–580.
22. Stefanovic, S., Jager, M., Deutsch, J., Broutin, J. & Masselot, M. (1998) *Am. J. Bot.* **85,** 688–697.
23. 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.
24. Albert, V. A., Backlund, A., Bremer, K., M. W. Chase, Manhart, J. R., Mishler, B. D. & Nixon, K. C. (1994) *Ann. Mo. Bot. Gard.* **81,** 534–568.
25. Goremykin, V., Bobrova, V., Pahnke, J., Troitsky, A., Antonov, A. & Martin, W. (1996) *Mol. Biol. Evol.* **13,** 383–396.
26. Chaw, S.-M., Zharkikh, A., Sung, H.-M., Lau, T.-C. & Li, W.-H. (1997) *Mol. Biol. Evol.* **14,** 56–68.
27. Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987) *Proc. Natl Acad. Sci. USA* **84,** 9054–9058.
28. Malek, O., Lättig, K., Hiesel, R., Brennicke, A. & Knoop, V. (1996) *EMBO J.* **15,** 1403–1411.
29. Davis, J. I., Simmons, M. P., Stevenson, D. W. & Wendel, J. F. (1998) *Syst. Biol.* **47,** 282–310.
30. Bowe, L. M. (1997) Ph.D. thesis (Vanderbilt University, Nashville, TN).
31. Laroche, J., Li, P. & Bousquet, J. (1995) *Mol. Biol. Evol.* **12,** 1151–1156.
32. Bowe, L. M. & dePamphilis, C. W. (1996) *Mol. Biol. Evol.* **13,** 1159–1166.
33. Swofford, D. L. (1999) PAUP*: *Phylogenetic Analysis Using Parsimony and Other Methods* (Sinauer, Sunderland, MD).
34. Lyons-Weiler, J., Hoelzer, G. A. & Tausch, R. J. (1996) *Mol. Biol. Evol.* **13,** 749–757.
35. Lyons-Weiler, J., Hoelzer, G. A. & Tausch, R. J. (1998) *Biol. J. Linn. Soc.* **64,** 493–511.
36. Lyons-Weiler, J. & Hoelzer, G. A. (1997) *Mol. Phyl. Evol.* **8,** 375–384.
37. Felsenstein, J. (1978) *Syst. Zool.* **27,** 401–410.
38. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29,** 170–179.
39. Chaw, S.-M., Parkinson, C. L., Cheng, Y., Vincent, T. M. & Palmer, J. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4086–4091.
40. Samgiullin, T. K., Martin, W. F., Troitsky, A. V. & Antonov, A. S. (1999) *J. Mol. Evol.* **49,** 310–315.
41. Hansen, A., Hansmann, S., Samigullin, T., Antonov, A. & Martin, W. (1999) *Mol. Biol. Evol.* **16,** 1006–1009.
42. Shutov, A. D., Braun, G., Chesnokov, Y. V., Horstmann, C., Kakhovskaya, I. A. & Bäumlein, H. *J. Mol. Evol.* **47,** 486–492.
43. Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, D. E., Soltis, P. S., Soltis, M., Zanis, E. A., Zimmer, Z., Chen, Z., Savolainen, V. & Chase, M. W. (1999) *Nature (London)* **402,** 404–407.
44. Raubeson, L. A. & Jansen, R. K. (1992) *Biochem. Syst. Ecol.* **20,** 17–24.
45. Mogensen, H. L. (1996) *Am. J. Bot.* **83,** 383–404.
46. Bailey, I. W. (1944) *Am. J. Bot.* **31,** 421–428.
47. Eames, A. J. (1952) *Phytomorphology* **2,** 79–100.
48. Cronquist, A. (1960) *Bot. Rev.* **26,** 425–482.
49. Bierhorst, D. W. (1971) *Morphology of Vascular Plants*. (Macmillan, New York).
50. Carlquist, S. (1996) *Int. J. Plant Sci.* **157,** S58–S76.
51. Gifford, E. M. & Foster, A. S. (1989) *Morphology and Evolution of Vascular Plants* (Freeman, New York).
52. Stebbins, G. L. (1974) *Flowering Plants: Evolution Above the Species Level* (Harvard Univ. Press, Cambridge, MA).
53. Sun, G., Dilcher, D. L., Zheng, S. & Zhou, Z. (1998) *Science* **282,** 1692–1695.
54. Soltis, P. S., Soltis, D. E. & Chase, M. W. (1999) *Nature (London)* **402,** 402–404.
55. Mathews, S. & Donoghue, M. J. (1999) *Science* **286,** 947–950.