The Phylogeny of Land Plants Inferred from 18S rDNA Sequences: Pushing the Limits of rDNA Signal?

Pamela S. Soltis, Douglas E. Soltis,* Paul G. Wolf,† Daniel L. Nickrent,‡ Shu-Miaw Chaw,§ and Russell L. Chapman‖*

*Department of Botany, Washington State University; †Department of Biology, Utah State University; ‡Department of Plant Biology, Southern Illinois University at Carbondale; §Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan; and ‖Department of Biological Sciences, Louisiana State University at Baton Rouge

Previous studies of the phylogeny of land plants based on analysis of 18S ribosomal DNA (rDNA) sequences have generally found weak support for the relationships recovered and at least some obviously spurious relationships, resulting in equivocal inferences of land plant phylogeny. We hypothesized that greater sampling of both characters and taxa would improve inferences of land plant phylogeny based on 18S rDNA sequences. We therefore conducted a phylogenetic analysis of complete (or nearly complete) 18S rDNA sequences for 93 species of land plants and 7 green algal relatives. Parsimony analyses with equal weighting of characters and character state changes and parsimony analyses weighting (1) stem bases half as much as loop bases and (2) transitions half as much as transversions did not produce substantially different topologies. Although the general structure of the shortest trees is consistent with most hypotheses of land plant phylogeny, several relationships, particularly among major groups of land plants, appear spurious. Increased character and taxon sampling did not substantially improve the performance of 18S rDNA in phylogenetic analyses of land plants, nor did analyses designed to accommodate variation in evolutionary rates among sites. The rate and pattern of 18S rDNA evolution across land plants may limit the usefulness of this gene for phylogeny reconstruction at deep levels of plant phylogeny. We conclude that the mosaic structure of 18S rDNA, consisting of highly conserved and highly variable regions, may contain historical signal at two levels. Rapidly evolving regions are informative for relatively recent divergences (e.g., within angiosperms, seed plants, and ferns), but homoplasy at these sites makes it difficult to resolve relationships among these groups. At deeper levels, changes in the highly conserved regions of small-subunit rDNAs provide signal across all of life. Because constraints imposed by the secondary structure of the rRNA may affect the phylogenetic information content of 18S rDNA, we suggest that 18S rDNA sequences be combined with other data and that methods of analysis be employed to accommodate these differences in evolutionary patterns, particularly across deep divergences in the tree of life.

Introduction

Recent analyses of morphological and molecular data have greatly improved our understanding of land plant phylogeny. However, relationships both among and within many of the major lineages of land plants remain unclear. The paucity of clearly homologous morphological characters across the evolutionary span of green algae to angiosperms requires that we look elsewhere for additional characters to refine further the picture of land plant phylogeny.

Ribosomal RNA or DNA sequences (rRNA or rDNA, respectively) have frequently been used to reconstruct deep branches of evolutionary history. Most notable have been those attempts to infer the history of life using small-subunit (SSU; 16S and 16S-like, including 18S) rRNA/rDNA sequences (e.g., Woese and Fox 1977; Olsen 1987; Woese 1987, 1994, 1998; Woese, Kandler, and Wheelis 1990; Olsen and Woese 1996). Within eukaryotes, analyses of 18S rRNA/rDNA sequences have been used to infer the early diversification of eukaryotes (e.g., Embley, Hirt, and Williams 1994; Bhattacharya and Medlin 1995) and have placed fungi as the sister group to animals (e.g., Wainright et al. 1993). In fact, 18S rRNA/rDNA sequences have been used for phylogeny reconstruction within many groups of eukaryotes, including most major groups of plants, e.g., green algae (e.g., Buchheim and Chapman 1991; Chapman and Buchheim 1991; reviewed in Chapman et al. 1998), bryophytes (Waters et al. 1992; Mishler et al. 1994; Capesius 1995; Hedderson, Chapman, and Rootes 1996; Hedderson, Chapman, and Cox 1998), gymnosperms (Chaw et al. 1993, 1995, 1997), and angiosperms (Soltis et al. 1997).

Several studies using 18S rDNA sequences have focused on the origin of land plants and the placement of specific groups in the phylogeny of land plants as a whole (e.g., Mishler et al. 1994; Kranz et al. 1995; Hedderson, Chapman, and Rootes 1996; Kranz and Huss 1996; Huss and Kranz 1997; Hedderson, Chapman, and Cox 1998). These studies portray very different patterns of relationship among major clades of land plants and therefore cannot accurately provide a general framework of land plant phylogeny. These studies have generally suffered from both limited taxon sampling and the use of partial sequences (limiting the number and perhaps biasing the selection of characters included). For example, recent studies emphasizing bryophytes (Capesius 1995; Hedderson, Chapman, and Rootes 1996; Capesius and Bopp 1997) produced trees that fundamentally disagree on the relationships among the major lineages of bryophytes (mosses, hornworts, and liverworts). An analysis focusing on pteridophytes (Kranz and Huss 1996) found *Equisetum* and *Psilotum* to be successive

Key words: land plants, phylogeny, 18S rDNA, molecular evolution.

Address for correspondence and reprints: Pamela S. Soltis, Department of Botany, Washington State University, Pullman, Washington 99164-4238. E-mail: psoltis@wsu.edu.

Mol. Biol. Evol. 16(12):1774–1784. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

sister groups to the seed plants, rather than relatives of the ferns (e.g., Kenrick and Crane 1997*a*, 1997*b*). Previous studies have typically found weak support for those relationships recovered and at least some obviously spurious relationships, resulting in equivocal inferences of land plant phylogeny. We hypothesized that greater sampling of both characters and taxa would improve inferences of land plant phylogeny based on 18S rDNA sequences. In this paper, we (1) describe a phylogenetic analysis of complete (or nearly complete) 18S rDNA sequences for 93 species of land plants and (2) evaluate the usefulness of 18S rDNA sequences for reconstructing deep branches of plant phylogeny.

Materials and Methods

Sources of Plant Material and Sequences

Sequences of 18S rDNA from 93 species of land plants, representing hornworts, liverworts, mosses, ferns, fern allies, all extant lineages of ''gymnosperms,'' and angiosperms, were analyzed. Sources of sequences and GenBank accession numbers for all sequences can be found at http://www.wsu.edu:8080/~soltilab/. Sequences of the charophycean (*sensu* Mattox and Stewart 1984) algal orders Coleochaetales (*Coleochaete orbicularis* and *Coleochaete scutata*) and Charales (*Chara australis, Chara foetida, Chara connivens, Nitella flexilis,* and *Nitella* sp.), the two groups that appear to be the sister groups to land plants (reviewed in Huss and Kranz [1997] and Chapman et al. [1998]), were used as outgroups.

Amplification and sequencing of 18S rDNA followed a variety of procedures. Details are given in Soltis and Soltis (1997) for angiosperms, Chaw et al. (1997) for gymnosperms, Wolf (1995) for ferns and fern allies, and Waters et al. (1992) for bryophytes. Both manual and automated sequencing methods were used.

Sequences were aligned by eye and were generally easy to align across all land plants except the bryophytes; the aligned sequences were 1,869 bp in length. A few areas of ambiguous alignment remained, and these regions and terminal priming sites (totaling 130 bp) were omitted from the phylogenetic analyses. The alignments are available electronically at http://www.wsu.edu:8080/~soltilab/ and on the home page for the Green Plant Phylogeny Research Coordination Group at http://ucjeps.berkeley.edu/bryolab/greenplantpage.html.

Heuristic parsimony analyses were conducted using PAUP* 4.0 (Swofford 1998). The search strategy first involved 200 replicates with RANDOM taxon addition and TBR branch swapping, saving 10 most parsimonious trees per replicate. These initial searches found 10 shortest trees of 2,787 steps; these trees were then used as starting trees in searches that used tree bisection-reconnection (TBR) branch swapping and saved all most-parsimonious trees (MULPARS). In addition, because shorter trees might be reachable from longer starting trees (as in Soltis et al. 1997), all 1,833 trees obtained in the initial 200 replicate searches, ranging in length from 2,787 to 2,798 steps, were used as starting trees to try to find shorter trees than those recovered using only the 10 trees of length 2,787 as starting trees. To assess support for the inferred relationships, we conducted FAST bootstrap analyses with 1,000 replicates. FAST bootstrapping, as implemented in PAUP* 4.0, samples characters with replacement as in typical phylogenetic bootstrapping (Felsenstein 1985), but each replicate search consists only of stepwise addition, with no branch swapping. FAST bootstrapping provides equivalent to slightly lower levels of support than bootstrapping with branch swapping (Mort et al. 2000).

Analyses of patterns of 18S rDNA evolution in angiosperms as a whole and in select, strongly supported clades of angiosperms demonstrated a transition : transversion ratio of approximately 2:1 and differential rates of nucleotide substitution in stem and loop regions of the 18S rRNA molecule (Soltis and Soltis 1998). Although previous analyses found only trivial differences in topology when weights derived from these observations were assigned in phylogenetic analyses of clades of angiosperms (Soltis and Soltis 1998), we implemented weighting schemes in subsequent analyses of the land plants to reflect these patterns of evolution. In this data set for land plants, the transition : transversion ratio is 1.9 (as calculated by MacClade [Maddison and Maddison 1992] and tree 1 of the 12 shortest trees [length 2,787] found in the heuristic search with equal weighting). In the first weighted analysis, we used a step matrix to weight transitions half as much as transversions and conducted the FAST bootstrap parsimony search as before. The second weighted analysis used a model of the secondary structure of rRNA for *Glycine max* (by D. Nickrent in Soltis et al. 1997) to estimate whether specific nucleotides should be considered components of stems or loops. The aligned sequences of 1,869 nt required the insertion of 62 nt in the rDNA sequence of *G. max.* All 1,869 nt were mapped onto the secondary structure for *G. max,* and each base was designated as a stem base or a loop base. Although this designation of bases as either stem or loop is admittedly simplistic and does not accommodate small bulges or single-stranded sequences (e.g., Vawter and Brown 1993), this system accurately classifies almost all nucleotides in plant 18S rDNA. As described in Soltis and Soltis (1998), loops were defined as being four bases or more in length; loops of 4 nt are common in both SSU and LSU rRNAs and make up over 50% of the hairpin loops in 16S rRNA (Woese, Kandler, and Wheelis 1990; Gutell 1993). Therefore, this minimum size of 4 nt largely corresponds to patterns of base-pairing in rRNA. Stem bases were assigned a weight of 1, and loop bases were given a weight of 2. A weighted FAST bootstrap parsimony search was conducted as with the equally weighted characters.

Results

Equally Weighted Parsimony

Of the 1,739 aligned nucleotide positions that were included in the analysis, 448 were parsimony-informative. Heuristic searches using equal weighting of all

characters and all character states found 12 minimal-length trees of 2,787 steps (consistency index [CI] = 0.362; retention index [RI] = 0.762), all on one island (*sensu* Maddison 1991). The same 12 trees were found using the set of 10 starting trees of length 2,787 and the set of 1,833 trees of length 2,787–2,798. An additional 100 replicate searches with random taxon addition, as described above, found 986 trees of length 2,791–2,828. Swapping on all 986 of these trees found 872 trees of length 2,788; no additional trees of length 2,787 were found. The strict consensus of the 12 shortest trees is shown in figure 1.

Using two species of *Coleochaete* and five species of Characeae as outgroups, the Characeae (plus the hornwort *Notothylas*) form a clade with all land plants (bootstrap value = 100%) and appear as the sister group to the land plants in all shortest trees (although bootstrap support for this relationship is less than 50%). The land plants fall into two major clades. One major clade comprises (1) a clade of mosses (bootstrap value < 50%) and (2) a clade of liverworts, *Anthoceros* + *Phaeoceros*, and lycophytes (bootstrap value < 50%). The liverworts are paraphyletic, with a basal clade of *Sphaerocarpus*, *Reboulia*, and *Marchantia* (bootstrap value = 100%), followed by a grade of liverworts, *Anthoceros* + *Phaeoceros*, several mosses, and a lycophyte clade (bootstrap value < 50%); support for Lycopodiaceae is 78%, and support for *Selaginella* + *Isoetes* is 74%.

The second major clade of land plants (bootstrap value < 50%) has *Sphagnum* as sister to all remaining land plants (the tracheophytes; bootstrap value < 50%). Within the tracheophytes is a split between pteridophytes (minus lycophytes, which appeared in the first clade) and seed plants (plus *Salvinia*). A clade of *Ophioglossum* + (*Tmesipteris* + *Psilotum*) (bootstrap value < 50%) is sister to the remaining pteridophytes. *Equisetum* is sister to *Osmunda* (bootstrap value < 50%), and this clade is nested within the ferns.

Salvinia is sister to the seed plants, although bootstrap support for this relationship is less than 50%. The seed plant clade (bootstrap value = 100%) comprises a gymnosperm clade, with bootstrap support less than 50%, and an angiosperm clade (bootstrap value = 100%). Within the monophyletic gymnosperms, the cycads (bootstrap value = 77%) are sister to the remaining genera, although this relationship has bootstrap support less than 50%, followed by *Ginkgo* (bootstrap value < 50%); the conifers (bootstrap value = 77%) and Gnetales (bootstrap value = 100%) are sister groups (bootstrap value = 64%). Relationships within the angiosperms largely follow those reported in broader analyses of angiosperms based on 18S rDNA sequences (e.g., Nickrent and Soltis 1995; Soltis et al. 1997), although branching patterns differ slightly given the limited sampling of the angiosperm clade in the present study. The basal branches are *Austrobaileya, Amborella, Nymphaea,* and *Saururus,* although none of these relationships has bootstrap support greater than 50%. The remaining angiosperms form two clades, neither of which has bootstrap support greater than 50%. The first clade comprises *Phytolacca, Ceratophyllum* + *Sassafras,* and the monocots (bootstrap value < 50%); the second clade contains (*Asarum* + *Saruma*) and the eudicots (bootstrap value < 50%).

Weighted Parsimony

Parsimony analyses that differentially weighted stem versus loop bases or transitions versus transversions did not dramatically alter the overall topology. *Notothylas,* a hornwort, remained the sister to the Characeae, with bootstrap support of 100% in both weighted analyses. Both weighted analyses also recovered the (*Sphaerocarpus* + *Reboulia* + *Marchantia*) liverwort clade (bootstrap value = 100%), although in neither case did *Fossombronia* appear with this clade of liverworts. Both weighted analyses also showed strong support (75% for the stems-vs.-loops analysis and 90% for the transitions-vs.-transversions analysis) for the moss clade, with *Polytrichum* and *Atrichum* as sisters outside this clade. The *Huperzia*/*Lycopodium* clade was also retained in both weighted analyses (75% for the stems-vs.-loops analysis and 86% for the transitions-vs.-transversions analysis); the *Selaginella* + *Isoetes* clade (81% bootstrap in the transitions-vs.-transversions analysis) did not receive bootstrap support greater than 50% in the stems-versus-loops analysis.

The major difference between the bootstrap trees of the two weighted analyses and that of the analysis with equal weighting is that both weighted analyses recovered a large clade of pteridophytes, comprising (*Psilotum* + *Tmesipteris*), *Equisetum,* and all of the ferns included in the analysis, including *Salvinia.* This pteridophyte clade was supported by a bootstrap value of 75% in the stems-versus-loops analysis, but only 51% in the transitions-versus-transversions analysis. Although this clade (minus *Salvinia*) was also recovered in the shortest trees using equal weighting, it did not receive bootstrap support of ≥50%. Relationships at the base of the seed plants were not well resolved in either of the weighted analyses. In the transitions-versus-transversions analysis, a tetrachotomy appears at the base: cycads (84% bootstrap value), *Ginkgo,* conifers (85% bootstrap value) + Gnetales (100% bootstrap value), and angiosperms (100% bootstrap value); the sister status of conifers and Gnetales received bootstrap support of 80%. In the stems-versus-loops analysis, the conifers + Gnetales sister group collapses. In general, differential weighting of stems versus loops or of transitions versus transversions did not substantially alter the results obtained using equal weighting of characters and character states.

Discussion

Comparison of the 18S rDNA Tree with Inferences from Other Data

The relationships inferred from this parsimony analysis are not well supported as measured by the FAST bootstrap analysis. Although the general structure of the shortest trees is consistent with most hypotheses of land plant phylogeny (e.g., Crane 1985; Kenrick and Crane 1997*a*, 1997*b*), several relationships inferred from

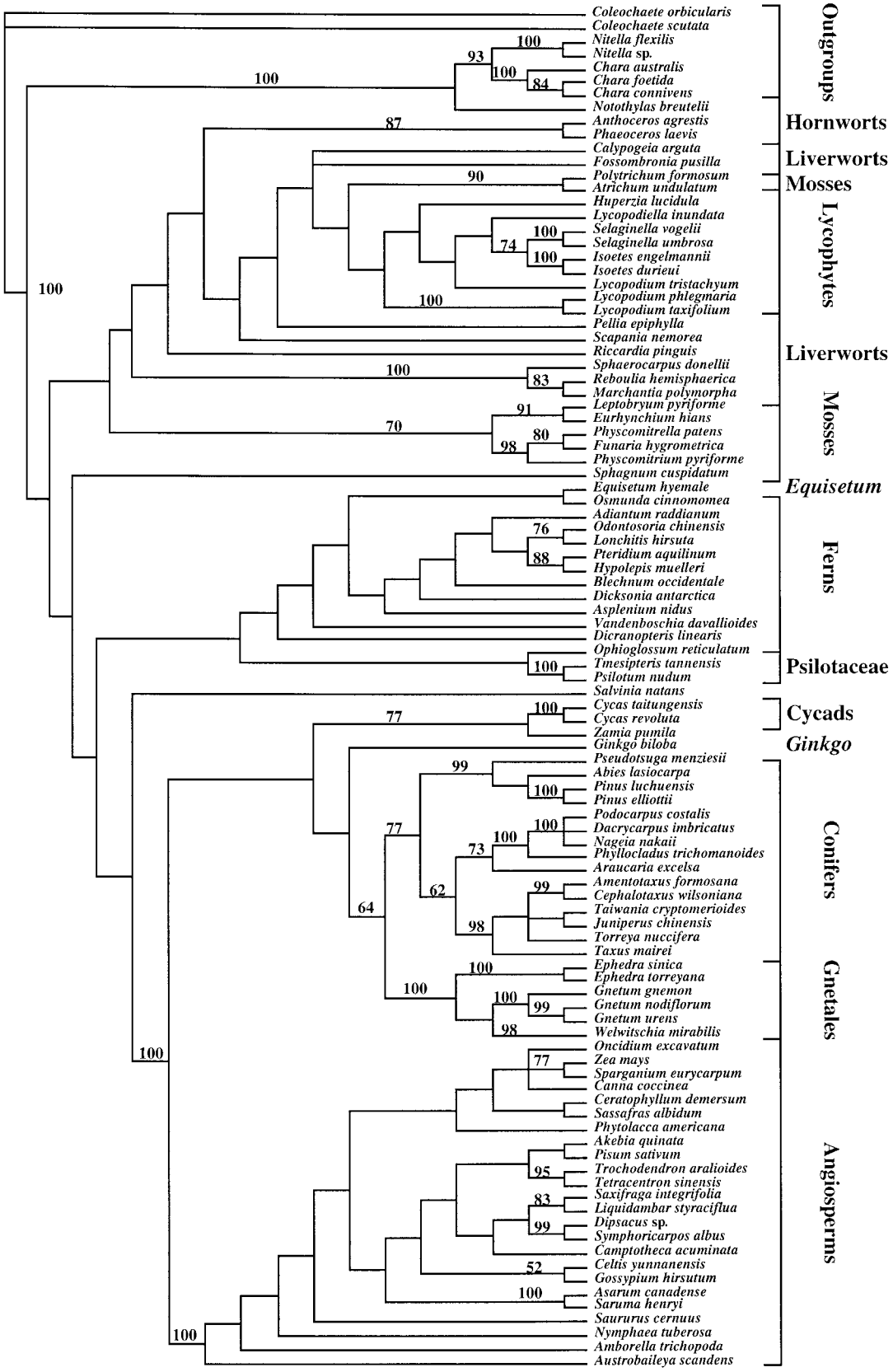

FIG. 1.—Strict consensus of 12 most-parsimonious trees based on parsimony analysis with equal weighting of characters and character state changes. Numbers above branches are bootstrap values calculated from 1,000 replicates of FAST bootstrapping. Length = 2,787 steps, consistency index = 0.362, and retention index = 0.762.

the strict consensus of the 12 shortest trees are clearly spurious. Generic relationships within well-supported clades are not necessarily expected to be accurate in all cases, given the effects of taxon sampling. For example, undersampling of the angiosperm clade, for which 24 species represent at least 260,000 species (Takhtajan 1997), results in the portrayal of some unusual generic relationships that are not evident in more inclusive analyses of the angiosperms based on 18S rDNA (e.g., Soltis et al. 1997). Of greater interest in this study are relationships among major groups of land plants, but the placements of several such groups also appear to be spurious. The hornwort *Notothylas* is sister to Characeae, rather than to *Anthoceros* and *Phaeoceros,* the other hornworts in the analysis. *Anthoceros* + *Phaeoceros,* in turn, are nested within a clade of mosses and liverworts and sister to a clade comprising additional liverworts and the lycophyte clade. The position of hornworts is uncertain in other phylogenetic analyses (e.g., Huss and Kranz 1997), having been placed as sister to all remaining land plants (Hedderson, Chapman, and Rootes 1996; Hedderson, Chapman, and Cox 1998), as sister to the mosses (Waters et al. 1992), as sister to the (mosses + tracheophytes) (Bremer 1985; Mishler and Churchill 1985; Mishler 1986; Bremer et al. 1987; Mishler et al. 1992), and as sister to the liverwort group Jungermanniidae (Capesius 1995). However, the lycophyte clade is clearly misplaced within this bryophyte clade; all other evidence places the lycophytes near or at the base of the tracheophyte clade, and they are now generally regarded as sister to all other tracheophytes (Raubeson and Jansen 1992; Kranz and Huss 1996; Kenrick and Crane 1997*a*, 1997*b*).

Despite more extensive sampling of taxa than in any previous analyses of land plant phylogeny based on 18S rDNA sequences, this analysis does not clearly resolve basal relationships and does not support or refute previous hypotheses. Both the hornworts (e.g., Hedderson, Chapman, and Rootes 1996; Hedderson, Chapman, and Cox 1998; see Garbary and Renzaglia 1998 for review of ultrastructural and developmental data) and the liverworts (e.g., Mishler et al. 1994; Kenrick and Crane 1997*a*; Lewis, Mishler, and Vilgalys 1997; Qiu et al. 1998) have been postulated to be sister to all other land plants; the clearly spurious clade of hornworts (minus *Notothylas*), some liverworts and mosses, and lycophytes recovered in our shortest trees precludes evaluation of these hypotheses. Analyses of mitochondrial 19S rDNA (also small-subunit) sequences were also equivocal with regard to the basal divergence of land plants (Duff and Nickrent 1999).

The position of *Sphagnum* away from other mosses and as sister to the tracheophytes contradicts all classifications but is not without precedent in phylogenetic analyses (e.g., Manhart 1994; Capesius 1995), perhaps because of long-branch attraction (Felsenstein 1978). Within the tracheophytes, the positions of four branches in particular are inconsistent with most current inferences of land plant phylogeny. Three of these branches involve the pteridophytes (*Equisetum, Ophioglossum,* and *Salvinia*), and the fourth involves the placement of the Gnetales (see below). Whereas the positions of *Equisetum* and *Salvinia* are clearly spurious, the placements of *Ophioglossum* and the Gnetales are consistent with the results of at least some other analyses. (1) *Equisetum* is nested within the fern clade, as the sister to *Osmunda,* rather than external to the ferns as inferred from paleobotanical, morphological, and anatomical data (reviewed in Kenrick and Crane 1997*a,* 1997*b*; Doyle 1998*b*) and from other molecular phylogenetic analyses (e.g., Hedderson, Chapman, and Rootes 1996; Manhart 1994). (2) The sister group relationship of Psilotaceae and Ophioglossaceae also contradicts all classifications but was likewise recovered in phylogenetic analyses of land plant sequences of *rbcL* (Manhart 1994; Hasebe et al. 1995), 16S rDNA (Manhart 1995), the internal transcribed spacer of chloroplast 16S rDNA (Pahnke et al. 1996), and the chloroplast gene *atpB* (Wolf 1997; summarized in Wolf et al. 1998). Inclusion of *Botrychium* (also of Ophioglossaceae), in addition to *Ophioglossum,* does not change the sister group relationship of Psilotaceae and Ophioglossaceae (data not shown). In our analysis and others, this Psilotaceae-Ophioglossaceae clade is sister to the leptosporangiate ferns. However, the Psilotaceae-Ophioglossaceae clade was not recovered in analyses of morphological characters (Pryer, Smith, and Skog 1995; Wolf et al. 1998). (3) The heterosporous fern *Salvinia* occurs in an unexpected location in the 18S rDNA tree, as the sister to the seed plants, rather than in the clade with all other ferns. This presumably represents long-branch attraction (Felsenstein 1978) between *Salvinia* and the stem leading to seed plants.

The placement of the Gnetales within a monophyletic gymnosperm clade contradicts evidence for their sister group relationship to the angiosperms, as suggested by analyses of morphological data (e.g., Crane 1985; Doyle and Donoghue 1986; Doyle, Donoghue, and Zimmer 1994; Doyle 1996, 1998*a*; see Crepet [1998] and Doyle [1998*b*] for reviews) and *rbcL* sequences (Manhart [1994] and the depiction of seed-plant relationships in Chase et al. [1993]). This 18S rDNA tree is not the first tree based on molecular characters to suggest that the gymnosperms are monophyletic and that the Gnetales and conifers are sisters. An 18S rDNA analysis focusing on the gymnosperms (Chaw et al. 1997) also showed the gymnosperms to be monophyletic, with the Gnetales as sister to the conifers, as found here (fig. 1) and by Hedderson, Chapman, and Rootes (1996) and Hedderson, Chapman, and Cox (1998). Some analyses of *rbcL* sequences spanning all of photosynthetic life also found a gymnosperm clade as sister to the angiosperms (Källersjö et al. 1998), although this relationship was not found in all of their analyses. Sequences of 19S rDNA from the mitochondrial genome also support the monophyly of the gymnosperms (Duff and Nickrent 1999). Neighbor-joining analyses of internal transcribed spacer regions in the rDNA of the chloroplast genome (cpITS; Goremykin et al. 1996) also found a gymnosperm clade with the Gnetales and conifers as sisters. In a combined cpITS and *rbcL* analysis, the gymnosperms also formed a clade, but the Gnetales

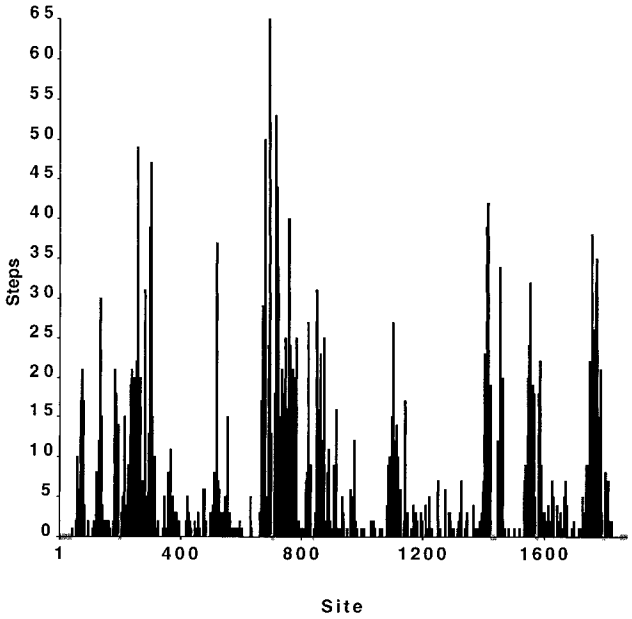

FIG. 2.—Distribution of changes in 18S rDNA sequences across a sample of land plants and green algal outgroups. Positions 1–1869 are plotted along the *x*-axis in intervals of four bases; gray blocks along the *x*-axis represent regions of uncertain alignment that were not included in the analyses. The number of changes (steps) in each block of four characters is plotted on the *y*-axis. This plot was generated using the CHART option of MacClade, version 3.03 (Maddison and Maddison 1992), the data file of 1739 nucleotide characters, and tree 1 of the 12 shortest trees (length = 2,787 steps).

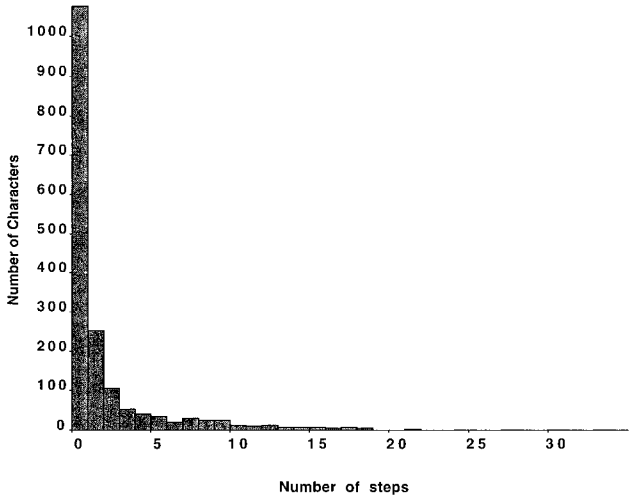

FIG. 3.—Plot showing the amount of variability in 18S rDNA characters across the land plants. The number of characters (*y*-axis) for a given level of variability (number of steps [i.e., changes], *x*-axis) is shown. This plot was generated as described in the legend of figure 2.

appeared as sister to all other lineages of gymnosperms (Goremykin et al. 1996). Most recently, analyses of MADS-box genes have shown that *Gnetum* genes are more closely related to conifer genes than to angiosperm sequences (Winter et al. 1999).

Unusal placements of single taxa or clades may occasionally be caused by one or a few highly divergent taxa that either have no close relatives in the data set or contain errors in their sequences. For example, the hornwort *Notothylas* and the eusporangiate fern *Ophioglossum* occupy unexpected positions in the tree (fig. 1), as sister to the green algae *Chara* and *Nitella* and sister to the Psilotaceae, respectively. Replacing the original *Ophioglossum* sequence with a new sequence did not change its position in the tree. Removing *Notothylas* (cf. Swofford et al. 1996) did not improve the topology near the base of the tree: neither the mosses nor the liverworts formed a clade (tree not shown). Likewise, removal of the seven algal outgroups did not eliminate the problems apparent in figure 1. The lycophytes formed a clade that was sister to the other tracheophytes, but neither the mosses nor the liverworts were monophyletic, and *Sphagnum* and *Notothylas* were successive sister groups to the seed plants.

Limitations of 18S rDNA for Reconstructing Deep Levels of Plant Phylogeny

The rate and pattern of 18S rDNA evolution across land plants may limit the usefulness of this gene for phylogeny reconstruction at deep levels of plant phylogeny. In angiosperms, loop regions of 18S rDNA have higher rates of nucleotide substitution than do stem regions (Soltis and Soltis 1998). Furthermore, some loops, e.g., the termini of helices E10-1, 17, E23-1, and 43, experience considerable length mutation (Soltis et al. 1997; Soltis and Soltis 1998), although most of the 18S rRNA gene is largely conserved in length, even across all land plants. This pattern of conserved stem regions and more variable loops may reflect functional constraints on portions of the gene. If so, then most of the gene may not be free to evolve randomly; consequently, the overall rate of evolution of the gene overestimates sequence evolution in the conserved regions and underestimates it in the more variable regions.

In land plants (and their green algal sister group), as in the angiosperms alone (Nickrent and Soltis 1995; Soltis and Soltis 1998), 18S rDNA is a mosaic of highly conserved and highly variable regions (fig. 2). Considering only the 1,739 aligned positions, most (1,076, calculated by MacClade) were invariant, 253 changed once, 105 changed twice, and the remaining 305 changed 3 or more times, with six characters changing 20 or more times (fig. 3). Of the 663 variable characters (calculated by MacClade), nearly half (292) showed no homoplasy (CI = 1.00), and 413 had CIs of $\geq$0.50 (fig. 4). Of the remaining 250 variable characters, 55 had CIs of 0.330–0.339, 25 had CIs of 0.250–0.229, and 24 had CIs of 0.200–0.209. Thus, of the 1,739 nucleotide positions analyzed, over three fourths changed only once or not at all across this sample of land plant diversity, and the majority of changes occurred at a limited number of sites.

The variable regions of 18S rDNA may evolve too rapidly for the methods used to recover sufficient historical information across the time span of land plant evolution, although they contain signal within the major groups. In fact, 18S rDNA sequences contain less phylogenetic information than do either of the chloroplast genes *rbcL* and *atpB* in analyses of the angiosperms alone (Chase and Cox 1998). Although the reason for

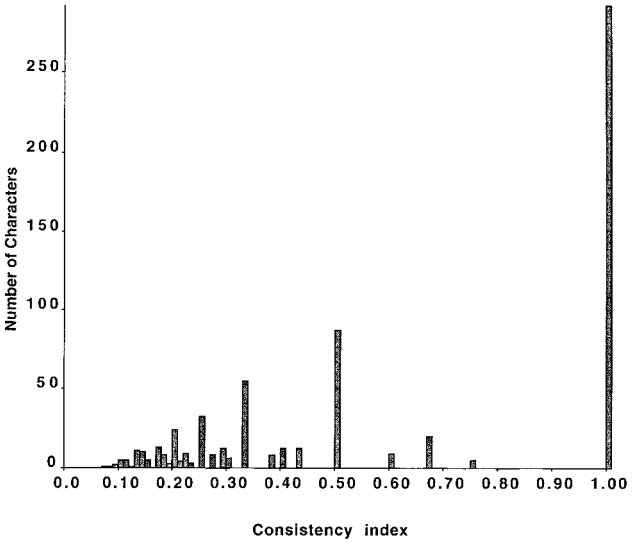

FIG. 4.—Plot of consistency indices (CIs) for variable 18S rDNA characters. A CI value of 1.00 indicates no homoplasy for a character; decreasing CI values reflect increasing amounts of homoplasy. This plot was generated as described in the legend of figure 2.

the lower information content of 18S rDNA relative to *rbcL* and *atpB* is not clear, it may be due to the generally greater functional constraints on much of the 18S rRNA gene and the interspersed evolutionary hot spots (see Soltis and Soltis 1998). The conserved regions may be too invariant to contain sufficient phylogenetic signal to resolve relationships among major groups of land plants. Furthermore, limited variation in the conserved regions may be constrained such that homoplasy is high at the relatively small number of variable sites. In addition, substitution in the highly variable regions may be so high across all land plants that signal is again masked by homoplasy that arises from multiple substitutions at a site.

Alternative methods of analysis may be required to extract the true phylogenetic information from the 18S rDNA data set. For example, approaches that incorporate models of sequence evolution may remove spurious relationships and improve support for clades (e.g., Swofford et al. 1996). However, a neighbor-joining (NJ) analysis using HKY85 distances (Hasegawa, Kishino, and Yano 1985), which allows for unequal base frequencies and for differences in substitution rates between transitions and transversions, did not substantially improve the tree and in many cases was worse than the tree based on parsimony (NJ tree not shown). For example, in the NJ analysis, the lycophytes no longer form a clade: Lycopodiales are sister to a clade of interspersed mosses and liverworts, but *Selaginella* and *Isoetes* form a clade that is sister to *Notothylas* + (*Nitella* + *Chara*). Also in the NJ analysis, *Salvinia* takes a more reasonable position, as sister to the fern clade (including *Psilotum, Tmesipteris,* and *Equisetum*), but *Equisetum* remains as sister to *Osmunda,* nested within this clade, and *Ophioglossum* is no longer the sister of *Psilotum* + *Tmesipteris.* The overall structure of relationships within the seed plants is fairly consistent between the NJ and parsimony analyses. However, in the NJ analysis, *Ginkgo* is sister to cycads, and relationships within the angiosperms are less similar to those inferred from in-depth studies (e.g., Soltis et al. 2000) than are those inferred from parsimony.

Given the mosaic evolution of land plant 18S rDNA, a method that allows for heterogeneity of evolutionary rates among nucleotide positions, such as maximum likelihood with a proportion of sites specified as invariant and the remainder having a distribution of rates that follows a gamma distribution (e.g., Tateno, Takezaki, and Nei 1994; Yang 1994; Sullivan, Holsinger, and Simon 1995; Lockhart et al. 1996), would be useful (an I + gamma model, cf. Gu, Fu, and Li 1995; Waddell and Penny 1996). Unfortunately, maximum-likelihood analysis for a data set of this size is not feasible. However, additional NJ analyses that take into account among-site rate heterogeneity were also conducted on the 18S rDNA data set (NJ trees not shown). The use of HKY85 distances with 60% of sites considered invariant in an NJ analysis did not improve the tree. For example, the lycophytes no longer form a clade, as they do in parsimony analyses or as expected on the basis of all other available data (e.g., Kenrick and Crane 1997*a*). Adding a gamma distribution with $\alpha = 0.5$ to 60% invariant sites had both positive and negative effects on the topology. In this analysis, the mosses are paraphyletic, and the liverworts are polyphyletic, but the lycophytes again form a clade which is sister to all other tracheophytes, consistent with analyses based on morphology (e.g., Kenrick and Crane 1997*a*) and gene order in the chloroplast genome (Raubeson and Jansen 1992). *Equisetum* is still embedded in the ferns, as are both *Ophioglossum* and *Salvinia.* Other differences from the parsimony analyses are that *Psilotum* + *Tmesipteris* are sister to the seed plants and that the gymnosperms are paraphyletic, with (cycads + *Ginkgo*) sister to ((conifers + Gnetales) + angiosperms). Although at least some of the relationships for tracheophytes obtained through these analyses are as plausible as those obtained through parsimony, the incorporation of rate heterogeneity into the distance measure did not improve relationships among the earlier-diverging clades (e.g., liverworts and mosses).

Because differences in base frequencies may result across large divergences, and because these differences may bias phylogenetic analyses, we also performed an NJ analysis with logdet distances (Lockhart et al. 1994) and 60% invariant sites. The results of this analysis are also inconsistent with what we know of land plant phylogeny from other sources. For example, when the tree is rooted with *Coleochaete,* the other green algae (*Chara* and *Nitella*) are nested well within the tree, in a clade with *Selaginella* and *Isoetes* that forms the sister to the seed plants. The remaining lycophytes are far removed from *Selaginella* and *Isoetes.* As before, neither the mosses nor the liverworts form a clade. It therefore appears that the problems with 18S rDNA sequences at this phylogenetic level cannot be accommodated by selecting models of molecular evolution that are thought

to offset the effects of heterogeneity of base frequencies and evolutionary rates.

Another analytical approach that holds promise for large data sets that span large phylogenetic distances is compartmentalization (Mishler 1994). Well-supported and noncontroversial clades are recognized as ''compartments''; compartmentalized analyses then either invoke constraints to reflect these clades or represent compartments by their inferred hypothetical ancestors in phylogeny reconstruction. We performed a compartmentalized analysis of this 18S rDNA data set for land plants (Mishler, Soltis, and Soltis 1998; unpublished data), recognizing as compartments green algae, liverworts, mosses, lycophytes, ferns, conifers, Gnetales, and angiosperms, based partly on bootstrap values obtained in this study and partly on inferences from many other sources of data (e.g., Kenrick and Crane 1997*a*). By constraining these clades, we prevented homoplasy from disrupting these major clades and allowed the analysis to optimize relationships among clades and to place those taxa, such as *Ophioglossum* and the Psilotaceae, that were not assigned to a compartment. The constraints simply identified the clades; branching order among the clades was determined by parsimony analyses. The compartmentalized analysis, with constraints imposed, removed some of the spurious relationships recovered in the unconstrained analysis (tree not shown). For example, the lycophytes are the sister to all other tracheophytes, as indicated by many other lines of evidence (e.g., Raubeson and Jansen 1992; Kenrick and Crane 1997*a*). The branching order of land plants is hornworts, liverworts, mosses, lycophytes, *Equisetum*, ferns, Psilotaceae, and seed plants (still with the gymnosperms monophyletic). In the constrained analysis, after 500 replicate parsimony searches with random taxon addition and nearest neighbor interchanges branch swapping, we found 42 trees of length 2,834 (47 steps longer than the shortest trees in the unconstrained analysis; fig. 1); removing *Ophioglossum* from the fern compartment resulted in 10 trees of length 2,843. These increased lengths indicate that there is considerable homoplasy among the clades; although the constrained trees of length 2,834 are only 1.7% longer than the shortest trees found in the unconstrained analysis, they are significantly longer (Kishino-Hasegawa test [Kishino and Hasegawa 1989]: $t = 3.4112$, $P < 0.0007$).

Some attempts to reconstruct the phylogeny of land plants using *rbcL* sequences (e.g., Manhart 1994) were also plagued by unusual placements of several taxa and/or clades. Possible explanations for those results were (1) insufficient taxon sampling (only 35 species were included) and (2) substantial homoplasy due to rapid evolution of *rbcL* and the consequent large number of multiple substitutions, although use of amino acid sequences did not improve the results (Hasebe et al. 1993). Other authors (e.g., Goremykin et al. 1996) have noted that *rbcL* may not contain sufficient phylogenetic information to resolve land plant relationships, although the analysis of 2,538 *rbcL* sequences spanning all of photosynthetic life successfully recovered the basic clades of land plant phylogeny (Källersjö et al. 1998). *rbcL* evolves more rapidly than does 18S rDNA (two times as rapidly in angiosperms; Nickrent and Soltis 1995) and also experiences fewer functional constraints (e.g., Albert et al. 1994; Chase and Albert 1998) but it appears to show less homoplasy than 18S rDNA (Chase and Cox 1998), and even third positions contain phylogenetic information across broad divergences (e.g., Lewis, Mishler, and Vilgalys 1997; Källersjö et al. 1998). Perhaps *rbcL*, or *rbcL* in conjunction with 18S rDNA, would provide greater signal than 18S rDNA alone for reconstructing the evolutionary history of land plants.

Conclusions

Parsimony analyses of 18S rDNA sequences recovered many relationships that are consistent with results obtained from analyses of other genes and/or morphology. However, much of the strict consensus is at odds with current classifications, the fossil record, the results of other phylogenetic analyses, or all three. Although we originally hypothesized that at least some of the spurious relationships recovered in previous analyses of land plant phylogeny based on 18S rDNA were due to insufficient sampling of characters and/or taxa, it appears that analysis of complete sequences (1,869 aligned/1,739 analyzed positions) for 100 species was not sufficient to correct for apparent conflict in the data. Most of the spine of the tree received low bootstrap support, resulting from either insufficient signal, conflicting signal, or both. We suggest that both limited signal and homoplasy in those characters that are variable are responsible for the generally poor bootstrap support for clades in the 18S rDNA tree. Constraints imposed by the secondary structure of the rRNA allow changes only in certain regions of the molecule; however, these regions likely are prone to homoplasy, perhaps substantial homoplasy, across the timescale spanned by land plants (Ordovician to the present). Thus, analysis of 18S rDNA sequences alone for any nontrivial set of taxa will likely not yield trees that are both well resolved and well supported. However, 18S rDNA sequences, in combination with *rbcL* and/or *atpB* sequences, may improve both the resolution and the internal support for clades of land plants, as they have done for the angiosperms (Soltis et al. 1998). We thus echo Manhart's (1994) suggestion that further analyses of land plant phylogeny follow the ''total evidence'' approach, combining data from multiple genes and morphology.

We conclude that nuclear SSU rDNA may contain historical signal at two levels, perhaps as a consequence of the evolutionary constraints imposed on the sequences. The more rapidly evolving regions are informative regarding recent divergences and relationships within major groups of land plants (e.g., angiosperms, seed plants, ferns). Relationships among these groups cannot be resolved by 18S rDNA sequences alone because of conflicting signal at the variable sites and insufficient signal at the conserved sites. At a deeper level of phylogenetic history, changes in the highly conserved regions may provide signal across all of life (e.g., Woese 1994, 1998; Pace 1997). Additional studies are needed

to examine the distribution of changes along the SSU rDNA sequence across deep levels of the phylogeny of life. Finally, the patterns of molecular evolution reported here and their inherent problems for phylogenetic analysis are likely not unique to ribosomal DNA, although the evolutionary constraints may be clearer for rDNA than for at least some protein-coding genes. Phylogenetic analyses should consider these constraints, and methods of analysis must be developed to accommodate them, in both small and large data sets and across broad evolutionary divergences.

Acknowledgments

This research was supported by a grant from the Mellon Foundation to P.S.S., D.E.S., and E. A. Zimmer and NSF grant DEB-9707868 to P.S.S. and D.E.S. We thank two anonymous reviewers for helpful comments on the manuscript.

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. Ann. Mo. Bot. Gard. **81**:534–567.

BHATTACHARYA, D., and L. MEDLIN. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J. Phycol. **31**:489–498.

BREMER, K. 1985. Summary of green plant phylogeny and classification. Cladistics **1**:369–385.

BREMER, K., C. J. HUMPHRIES, B. D. MISHLER, and S. P. CHURCHILL. 1987. On cladistic relationships in green plants. Taxon **36**:339–349.

BUCHHEIM, M. A., and R. L. CHAPMAN. 1991. Phylogeny of the colonial green flagellates: a study of 18S and 26S rRNA sequence data. BioSystems **25**:85–100.

CAPESIUS, I. 1995. A molecular phylogeny of bryophytes based on the nuclear encoded 18S rRNA genes. J. Plant Physiol. **146**:59–63.

CAPESIUS, I., and M. BOPP. 1997. New classification of liverworts based on molecular and morphological data. Plant Syst. Evol. **207**:87–97.

CHAPMAN, R. L., and M. A. BUCHHEIM. 1991. Ribosomal RNA gene sequences: analysis and significance in the phylogeny and taxonomy of green algae. CRC Crit. Rev. Plant Sci. **10**: 343–368.

CHAPMAN, R. L., M. A. BUCHHEIM, C. F. DELWICHE et al. (11 co-authors). 1998. Molecular systematics of the green algae. Pp. 508–540 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II. Kluwer Academic Publishers, Boston.

CHASE, M. W., and V. A. ALBERT. 1998. A perspective on the contribution of plastid *rbcL* DNA sequences to angiosperm phylogenetics. Pp. 488–507 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II. Kluwer Academic Publishers, Boston.

CHASE, M. W., and A. V. COX. 1998. Gene sequences, collaboration and analysis of large data sets. Aust. Syst. Bot. **11**: 215–229.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD et al. (42 co-authors). 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. **80**:528–580.

CHAW, S.-M., H. LONG, B.-S. WANG, A. ZHARKIKH, and W.-H. LI. 1993. The phylogenetic position of Taxaceae based on 18S rRNA sequences. J. Mol. Evol. **37**:624–630.

CHAW, S.-M., H.-M. SUNG, H. LONG, A. ZHARKIKH, and W.-H. LI. 1995. The phylogenetic positions of the conifer genera *Amentotaxus*, *Phyllocladus*, and *Nageia* inferred from 18S rRNA sequences. J. Mol. Evol. **41**:224–230.

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

CRANE, P. R. 1985. Phylogenetic analysis of seed plants and the origin of the angiosperms. Ann. Mo. Bot. Gard. **72**:716–793.

CREPET, W. L. 1998. The abominable mystery. Science **282**: 1653–1654.

DOYLE, J. A. 1996. Seed plant phylogeny and the relationships of Gnetales. Int. J. Plant Sci. **157**:S3–S39.

———. 1998*a*. Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. Mol. Phylogenet. Evol. **9**:448–462.

———. 1998*b*. Phylogeny of vascular plants. Annu. Rev. Ecol. Syst. **29**:567–569.

DOYLE, J. A., and M. J. DONOGHUE. 1986. Seed plant phylogeny and the origin of the angiosperms: an experimental cladistic approach. Bot. Rev. **52**:321–431.

DOYLE, J. A., M. J. DONOGHUE, and E. A. ZIMMER. 1994. Integration of morphological and ribosomal RNA data on the origin of the angiosperms. Ann. Mo. Bot. Gard. **81**:419–450.

DUFF, R. J., and D. L. NICKRENT. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am. J. Bot. **86**:372–386.

EMBLEY, T. M., R. P. HIRT, and D. M. WILLIAMS. 1994. Biodiversity at the molecular level: the domains, kingdoms and phyla of life. Philos. Trans. Soc. Lond. B Biol. Sci. **345**: 21–31.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**: 401–410.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

GARBARY, D. J., and K. S. RENZAGLIA. 1998. Bryophyte phylogeny and the evolution of land plants: evidence from development and ultrastucture. Pp. 45–63 *in* J. W. BATES, N. W. ASHTON, and J. G. DUCKETT, eds. Bryology for the twenty-first century. Maney and the British Bryological Society, Leeds, England.

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 *rbc*L data do not support Gnetalean affinities of angiosperms. Mol. Biol. Evol. **13**:383–396.

GU, X., Y.-X. FU, and W. H. LI. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. **12**:546–557.

GUTELL, R. R. 1993. Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. Curr. Opin. Struct. Biol. **3**:313–322.

HASEBE, M., M. ITO, R. KOFUJI, K. UEDA, and K. IWATSUKI. 1993. Phylogenetic relationships of ferns deduced from *rbcL* gene sequence. J. Mol. Evol. **37**:476–482.

HASEBE, M., P. G. WOLF, K. M. PRYER et al. (13 co-authors). 1995. Fern phylogeny based on *rbcL* nucleotide sequences. Am. Fern J. **85**:134–181.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:32–38.

HEDDERSON, T. A., R. CHAPMAN, and C. J. COX. 1998. Bryophytes and the origins and diversification of land plants: new evidence from molecules. Pp. 65–77 *in* J. W. BATES, N. W. ASHTON, and J. G. DUCKETT, eds. Bryology for the twenty-first century. Maney and the British Bryological Society, Leeds, England.

HEDDERSON, T. A., R. L. CHAPMAN, and W. L. ROOTES. 1996. Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Plant Syst. Evol. **200**: 213–224.

HUSS, V. A. R., and H. D. KRANZ. 1997. Charophyte evolution and the origin of land plants. Plant Syst. Evol. **11**(Suppl.): 103–114.

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 Syst. Evol. **213**:259–287.

KENRICK, P., and P. R. CRANE. 1997*a*. The origin and early diversification of land plants. Smithsonian Institution Press, Washington, D.C.

———. 1997*b*. The origin and early evolution of plants on land. Nature **389**:33–39.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. J. Mol. Evol. **29**:170–179.

KRANZ, H. D., and V. A. R. HUSS. 1996. Molecular evolution of pteridophytes and their relationship to seed plants: evidence from complete 18S rRNA gene sequences. Plant Syst. Evol. **202**:1–11.

KRANZ, H. D., D. MIKS, M.-L. SIEGLER, I. CAPESIUS, C. W. SENSEN, and V. A. HUSS. 1995. The origin of land plants: phylogenetic relationships between Charophytes, Bryophytes, and vascular plants inferred from complete small subunit ribosomal RNA gene sequences. J. Mol. Evol. **41**: 74–84.

LEWIS, L. A., B. D. MISHLER, and R. VILGALYS. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. Mol. Phylogenet. Evol. **7**:377–393.

LOCKHART, P. J., A. W. D. LARKUM, M. A. STEEL, P. J. WADDELL, and D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA **93**:1930–1934.

LOCKHART, P. J., M. A. STEEL, M. D. HENDY, and D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. **11**:605–612.

MADDISON, D. R. 1991. Discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. **40**:315–328.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade. Version 3.03. Sinauer, Sunderland, Mass.

MANHART, J. R. 1994. Phylogenetic analysis of green plant *rbcL* sequences. Mol. Phylogenet. Evol. **3**:114–127.

———. 1995. Chloroplast 16S rDNA sequences and phylogenetic relationships of ferns and fern allies. Am. Fern J. **85**:182–192.

MATTOX, K. R., and K. D. STEWART. 1984. Classification of the green algae: a concept based on comparative cytology. Pp. 29–72 *in* D. E. G. IRVINE and D. M. JOHN, eds. Systematics of the green algae. Academic Press, London.

MISHLER, B. D. 1986. A Hennigian approach to bryophyte phylogeny. J. Bryology **14**:71–81.

MISHLER, B. D. 1994. Cladistic analysis of molecular and morphological data. Am. J. Phys. Anthropol. **94**:143–156.

MISHLER, B. D., and S. P. CHURCHILL. 1985. Transition to a land flora: phylogenetic relationships of the green algae and bryophytes. Cladistics **1**:305–328.

MISHLER, B. D., L. A. LEWIS, M. A. BUCHHEIM, K. S. RENZAGLIA, D. J. GARBARY, C. F. DELWICHE, F. W. ZECHMAN, T. S. KANTZ, and R. L. CHAPMAN. 1994. Phylogenetic relationships of the ''green algae'' and ''bryophytes''. Ann. Mo. Bot. Gard. **81**:451–483.

MISHLER, B. D., P. S. SOLTIS, and D. E. SOLTIS. 1998. Compartmentalization in phylogeny reconstruction: philosophy and practice. DIMACS Report, Princeton, N.J.

MISHLER, B. D., P. H. THRALL, J. S. HOPPLE JR., E. DE LUNA, and R. VILGALYS. 1992. A molecular approach to the phylogeny of bryophytes: cladistic analysis of chloroplast-encoded 16S and 23S ribosomal RNA genes. Bryologist **95**: 172–180.

MORT, M. E., P. S. SOLTIS, D. E. SOLTIS, and M. L. MABRY. 2000. An empirical test of three methods for estimating internal support on phylogenetic trees. Syst. Biol. (in press).

NICKRENT, D. L., and D. E. SOLTIS. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcL* sequences. Ann. Mo. Bot. Gard. **82**:208–234.

OLSEN, G. J. 1987. Earliest phylogenetic branchings: Comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harb. Symp. Quant. Biol. **52**:825–837.

OLSEN, G. J., and C. R. WOESE. 1996. Lessons from an archael genome: what are we learning from *Methanococcus jannaschii*? Trends Genet. **12**:377–379.

PACE, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science **276**:734–740.

PAHNKE, J., V. GOREMYKIN, V. BOBROVA, A. TROITSKY, A. ANTONOV, and W. MARTIN. 1996. Utility of rDNA internal transcribed spacer sequences from the inverted repeat of chloroplast DNA in pteridophyte molecular phylogenetics. Pp. 217–230 *in* J. M. CAMUS, M. GIBBY, and R. J. JOHNS, eds. Pteridology in perspective. Royal Botanic Garden, Kew, England.

PRYER, K. M., A. R. SMITH, and J. E. SKOG. 1995. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. Am. Fern J. **85**:205–282.

QIU, Y.-L., Y. CHO, J. C. COX, and J. D. PALMER. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature **394**:671–674.

RAGAN, M. A., T. J. PARSONS, T. SAWA, and N. A. STRAUSS. 1994. 18S ribosomal DNA sequences indicate a monophyletic origin of Charophyceae. J. Phycol. **30**:490–500.

RAUBESON, L. A., and R. K. JANSEN. 1992. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science **255**:1697–1699.

SOLTIS, D. E., and P. S. SOLTIS. 1997. Phylogenetic relationships among Saxifragaceae sensu lato: a comparison of topologies based on 18S rDNA and *rbc*L sequences. Am. J. Bot. **84**:504–522.

SOLTIS, D. E., P. S. SOLTIS, M. W. CHASE et al. (13 co-authors). 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. Bot. J. Linn. Soc. (in press).

SOLTIS, D. E., P. S. SOLTIS, M. E. MORT, M. W. CHASE, V. SAVOLAINEN, S. B. HOOT, and C. M. MORTON. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Syst. Biol. **47**:32–42.

SOLTIS, D. E., P. S. SOLTIS, D. L. NICKRENT et al. (16 coauthors). 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann. Mo. Bot. Gard. **84**:1–49.

SOLTIS, P. S., and D. E. SOLTIS. 1998. Molecular evolution of 18S rDNA in angiosperms: implications for character weighting in phylogenetic analysis. Pp. 188–210 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II. Kluwer Academic Publishers, Boston.

SULLIVAN, J., K. E. HOLSINGER, and C. SIMON. 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA in Sigmodontine rodents. Mol. Biol. Evol. **12**:988–1001.

SWOFFORD, D. L. 1998. PAUP*. Version 4.0. Sinauer, Sunderland, Mass.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

TAKHTAJAN, A. 1997. Diversity and classification of flowering plants. Columbia University Press, New York.

TATENO, Y., N. TAKEZAKI, and M. NEI. 1994. Relative efficiencies of the maximum likelihood, neighbor joining, and maximum parsimony methods when substitution rate varies with site. Mol. Biol. Evol. **11**:261–277.

VAWTER, L., and W. M. BROWN. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics **134**:597–608.

WADDELL, P. J., and D. PENNY. 1996. Evolutionary trees of apes and humans from DNA sequences. Pp. 53–73 *in* A. J. LOCK and D. R. PETERS, eds. Handbook of symbolic evolution. Clarendon Press, Oxford, England.

WAINRIGHT, P. O., O. HINKLE, M. L. SOGIN, and S. K. STICKEL. 1993. Monophyletic origins of the Metazoa: an evolutionary link with fungi. Science **260**:340–342.

WATERS, D. A., M. A. BUCHHEIM, R. A. DEWEY, and R. L. CHAPMAN. 1992. Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. Am. J. Bot. **79**:459–466.

WINTER, K.-U., A. BECKER, T. MUNSTER, J. T. KIM, H. SAEDLER, and G. THIESSEN. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Natl. Acad. Sci. USA **96**:7342–7347.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

———. 1994. There must be a prokaryote somewhere: microbiology's search for itself. Microbiol. Rev. **58**:1–9.

———. 1998. Default taxonomy: Ernst Mayr's view of the microbial world. Proc. Natl. Acad. Sci. USA **95**:11043–11046.

WOESE, C. R., and G. E. FOX. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74**:5088–5090.

WOESE, C. R., O. KANDLER, and M. L. WHEELIS. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87**:4576–4579.

WOLF, P. G. 1995. Phylogenetic analyses of *rbcL* and nuclear ribosomal RNA gene sequences in Dennstaedtiaceae. Am. Fern J. **85**:306–327.

———. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. Am. J. Bot. **84**:1429–1440.

WOLF, P. G., K. M. PRYER, A. R. SMITH, and M. HASEBE. 1998. Phylogenetic studies of extant pteridophytes. Pp. 541–556 *in* D. E. SOLTIS, P. S. SOLTIS, and J. J. DOYLE, eds. Molecular systematics of plants II. Kluwer Academic Publishers, Boston.

YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

WILLIAM MARTIN, reviewing editor

Accepted August 19, 1999