



## Angiosperm Phylogeny Inferred from 18S Ribosomal DNA Sequences

Douglas E. Soltis; Pamela S. Soltis; Daniel L. Nickrent; Leigh A. Johnson; William J. Hahn; Sara B. Hoot; Jennifer A. Sweere; Robert K. Kuzoff; Kathleen A. Kron; Mark W. Chase; Susan M. Swensen; Elizabeth A. Zimmer; Shu-Miaw Chaw; Lynn J. Gillespie; W. John Kress; Kenneth J. Sytsma

*Annals of the Missouri Botanical Garden*, Volume 84, Issue 1 (1997), 1-49.

Stable URL:

<http://links.jstor.org/sici?sici=0026-6493%281997%2984%3A1%3C1%3AAPIF1R%3E2.0.CO%3B2-9>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Annals of the Missouri Botanical Garden* is published by Missouri Botanical Garden Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mobot.html>.

---

*Annals of the Missouri Botanical Garden*  
©1997 Missouri Botanical Garden Press

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

---

Volume 84  
Number 1  
1997

Annals  
of the  
Missouri  
Botanical  
Garden



---

ANGIOSPERM PHYLOGENY  
INFERRED FROM 18S  
RIBOSOMAL DNA  
SEQUENCES<sup>1</sup>

Douglas E. Soltis,<sup>2</sup> Pamela S. Soltis,<sup>2</sup>  
Daniel L. Nickrent,<sup>3</sup> Leigh A. Johnson,<sup>2</sup>  
William J. Hahn,<sup>4</sup> Sara B. Hoot,<sup>5</sup>  
Jennifer A. Sweere,<sup>4</sup> Robert K. Kuzoff,<sup>2</sup>  
Kathleen A. Kron,<sup>6</sup> Mark W. Chase,<sup>7</sup>  
Susan M. Swensen,<sup>8</sup> Elizabeth A.  
Zimmer,<sup>4</sup> Shu-Miaw Chaw,<sup>9</sup> Lynn J.  
Gillespie,<sup>10</sup> W. John Kress,<sup>11</sup> and  
Kenneth J. Sytsma<sup>12</sup>

---

ABSTRACT

Parsimony analyses were conducted for 223 species representing all major groups of angiosperms using entire 18S ribosomal DNA (rDNA) sequences. Although no search swapped to completion, the topologies recovered are highly concordant with those retrieved via broad analyses based on the chloroplast gene *rbcl*. The general congruence of 18S rDNA and *rbcl* topologies further clarifies the broad picture of angiosperm phylogeny. In all analyses, the first-branching angiosperms are Amborellaceae, Austrobaileyaceae, Illiciaceae, and Schisandraceae, all woody magnoliids. These taxa are always followed by the paleoherb family Nymphaeaceae. This same general order of early-branching taxa is preserved with several suites of outgroups. In most searches, the remaining early-branching taxa represent Piperales and other orders of subclass Magnoliidae (sensu Cronquist). With the exception of *Acorus*, the monocots are supported as monophyletic and typically have as their sister *Ceratophyllum*. In most analyses, taxa with uniaperturate pollen form a grade at the base of the angiosperms; a large eudicot clade is composed primarily of taxa having triaperturate pollen. Two large subclades are present within the eudicots, one consisting largely of Rosidae and a second corresponding closely to Asteridae sensu lato. Subclasses Dilleniidae and Hamamelidae are highly polyphyletic. These data sets of 18S rDNA sequences also permit an analysis of the patterns of molecular evolution of this gene. Problems deriving from both the prevalence of indels and uncertain alignment of 18S rDNA sequences have been overstated in previous studies. With the exception of a few well-defined regions, insertions and deletions are relatively uncommon in 18S rDNA; sequences are therefore easily aligned by eye across the angiosperms. Indeed, several indels in highly conserved regions appear to be phylogenetically informative. Initial analyses suggest that both stem and loop bases are important sources of phylogenetic information, although stem positions are prone to compensatory substitutions. Of the stem changes analyzed, only 27% destroy a base-pairing couplet; 73% maintain or restore base pairing.

---

<sup>1</sup> This research was supported in part by grants from the National Science Foundation (DEB 9307000 to DES, DEB 9407984 to DLN, DEB 9303266 to WJH, DEB 9407350 to KAK, DEB 9306913 to SMS and Loren H. Rieseberg); the National Science Council, Republic of China (2818F) to SMC, the Mellon Foundation grant (to EAZ, PSS, and DES), the Betty W. Higinbotham Trust (Department of Botany, Washington State University, to LAJ and RKK), and the Scholarly Studies Program of the Smithsonian Institution (to WJK and EAZ). We thank D. Swofford for access to PAUP\* 4.0. and S. Farris for conducting a parsimony jackknife analysis. We also thank B. Alverson, J. Doyle, M. HersHKovitz, D. Olmstead, J. Palmer, Y.-L. Qiu, J. Rodman, and Q.-Y. Xiang for providing plant material and DNAs used in this study. We appreciate the valuable advice of D. Swofford on analyzing large data sets and the help of M. HersHKovitz in using the UNIX version of PAUP\*. We also thank D. Olmstead and M. Sanderson for helpful comments on the manuscript.

Although the angiosperms are almost universally considered to be monophyletic, many basic questions of angiosperm phylogeny remain unanswered, including: (1) what are the first-branching angiosperms? (2) what is the ancestor of the monocots? (3) what are the major groups of angiosperms and the relationships among these groups? Despite intensive study, these questions have been difficult to answer for a variety of reasons. Most notable, perhaps, is the inadequacy of the fossil record alone to answer these questions conclusively. In addition, the apparent rapid radiation of the angiosperms following their origin resulted in few morphological synapomorphies among lineages at the base of the angiosperm tree, hindering attempts to resolve relationships among major groups (Crane et al., 1995). Finally, the angiosperms present relatively few morphological characters for comparison at higher levels. For example, recent cladistic analyses of morphological characters for angiosperms (Donoghue & Doyle, 1989a, b) and all seed plants (Doyle et al., 1994) included only 54 and 82 characters, respectively. As recently demonstrated by Doyle et al. (1994), careful analysis of both morphological and molecular data is required to understand angiosperm phylogeny.

During the past decade, several attempts have been made to reconstruct the phylogeny of the angiosperms. Morphological and molecular analyses usually identify the Gnetales as the extant sister group to the angiosperms, in either the shortest trees or those slightly longer (e.g., Crane, 1985, 1988; Donoghue & Doyle, 1989a, b; Doyle & Donoghue, 1986, 1992; Loconte & Stevenson, 1991; Hamby & Zimmer, 1992; Chase et al., 1993; Doyle et al., 1994; Nixon et al., 1994; but see Goremykin et al., 1996; Chaw et al., 1997). Molecular phylogenetic analyses include those based on *rbcL* sequences (Chase et al., 1993), partial 18S and 26S ribosomal RNA sequences (Hamby & Zimmer, 1992), and *rbcS* amino acid sequences (Martin & Dowd, 1991). These analyses tend to identify many of the same major groups of taxa, but they often

present different views of relationships among these groups.

In the largest phylogenetic analysis of angiosperms, Chase et al. (1993) presented the results of two parsimony analyses of DNA sequences from the chloroplast gene *rbcL* for 475 and 499 species of seed plants. More recently, Rice et al. (1997) have reanalyzed the 499-taxon *rbcL* data matrix to search for shorter trees. The benefits to the systematics community of performing these large phylogenetic analyses of seed plants in general, and angiosperms in particular, have been considerable. These studies provide comprehensive, explicit phylogenetic hypotheses of higher-level relationships in the angiosperms. Furthermore, the need for similar studies of angiosperms based on other character sets has been recognized, and such studies have been encouraged (e.g., Chase et al., 1993). Particularly important is the comparison of chloroplast-based phylogenetic estimates (Chase et al., 1993) with topologies derived from analyses of nuclear genes.

For reasons reviewed elsewhere, phylogenetic analyses based on nuclear DNA have largely involved portions of the rDNA cistron (e.g., Mindell & Honeycutt, 1990; Hillis & Dixon, 1991; Hamby & Zimmer, 1992; Sanderson & Doyle, 1993a; Nickrent & Soltis, 1995). Analyses of 18S rDNA and rRNA sequences have been used for phylogenetic inference at higher taxonomic levels in animals (e.g., Sogin et al., 1986; Field et al., 1988; Wainright et al., 1993; Wada & Satoh, 1994), protozoa (Schlegel et al., 1991), algae (Buchheim et al., 1990; Huss & Sogin, 1990; Kantz et al., 1990; Hendricks et al., 1991; Chapman & Buchheim, 1991; Bakker et al., 1994; Ragan et al., 1994; Olsen et al., 1994), fungi (Forster et al., 1990; Swann & Taylor, 1993; Hinkle et al., 1994), lichens (Gargas et al., 1995), bryophytes (Mishler et al., 1994; Capesius, 1995; Kranz et al., 1995), gymnosperms (e.g., Chaw et al., 1993, 1995, 1997), and even among the deepest branches of life (Wolters & Erd-

<sup>2</sup> Department of Botany, Washington State University, Pullman, Washington 99164, U.S.A.

<sup>3</sup> Department of Plant Biology, Southern Illinois University, Carbondale, Illinois 62901, U.S.A.

<sup>4</sup> Laboratory of Molecular Systematics, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

<sup>5</sup> Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin 53201, U.S.A.

<sup>6</sup> Department of Biology, Wake Forest University, Winston-Salem, North Carolina 27109, U.S.A.

<sup>7</sup> Laboratory of Molecular Systematics, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom.

<sup>8</sup> Department of Biology, Ithaca College, Ithaca, New York 14850, U.S.A.

<sup>9</sup> Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China.

<sup>10</sup> Research Division, Canadian Museum of Nature, Ottawa, Ontario K1P 64P, Canada.

<sup>11</sup> Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

<sup>12</sup> Department of Botany, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

mann, 1986; Olsen, 1987; Woese, 1987; Embley et al., 1994; Bhattacharya & Medlin, 1995).

Despite this wide usage in other major groups of organisms, the 18S rRNA gene has received comparatively little attention in angiosperms. In large part this reflects the enormous interest in, and demonstrated utility of, *rbcL* sequences for inferring phylogeny, particularly at the family level and above (e.g., Chase et al., 1993; Morgan & Soltis, 1993; Olmstead et al., 1993; Kron & Chase, 1993; Qiu et al., 1993; Rodman et al., 1993). In addition, skepticism apparently exists among many angiosperm systematists regarding the utility of 18S rDNA sequences for inferring plant phylogeny. Early analyses of 18S rDNA or rRNA sequences in angiosperms (e.g., Nickrent & Franchina, 1990; Hamby & Zimmer, 1992; Nickrent & Starr, 1994), while in general pointing to the possible phylogenetic utility of these data, raised concerns that 18S rDNA may be too evolutionarily conservative to address phylogenetic questions at the family level and above and that insertion and deletion events (indels) occur frequently in at least some portions of 18S rDNA, making alignment of sequences difficult. In addition, other basic background information regarding the molecular evolution of the 18S rRNA gene is not available for angiosperms. For example, given that 18S rRNA, as well as rRNAs in general, have inherent secondary structure that includes characteristic loop (non-paired) and stem (paired) stretches of RNA, should changes in the encoding stem and loop bases be considered equally informative in phylogenetic analyses? Models of rRNA evolution suggest that paired (stem) bases will undergo compensatory changes to maintain the appropriate base pairing. However, empirical studies of angiosperm rRNA structure are few (e.g., Senecoff & Meagher, 1992), and available data sets have not been used to evaluate patterns of evolution of the 18S rRNA gene in angiosperms.

The history of the use of 18S rRNA and rDNA sequences for phylogeny reconstruction in angiosperms was recently reviewed (Nickrent & Soltis, 1995). To date, the largest studies of 18S sequences are those of Hamby and Zimmer (1992) and Nickrent and Soltis (1995). Zimmer and collaborators conducted phylogenetic analyses using direct sequencing of rRNA from approximately 60 species of vascular plants, of which 29 were dicots and 17 monocots (Zimmer et al., 1989; Hamby & Zimmer, 1992). These investigators sequenced portions of both 18S and 26S rRNA, yielding a total of 1701 base positions per taxon, 1097 base positions from the 18S gene and 604 from the 26S gene. The shortest trees obtained had a number of features in

accord with existing classifications, but sampling of nonmagnoliid taxa was sparse and may explain some of the unusual relationships suggested among more derived angiosperms. Furthermore, many of the nodes were poorly supported. As a result of the unusual relationships suggested for some taxa and the poor resolution obtained in this study, angiosperm systematists remained unsure of the utility of 18S and 26S rRNA (and rDNA) sequences for inferring phylogeny.

More recently, Nickrent and Soltis (1995) compared the rate of evolution and phylogenetic resolution of entire 18S rDNA sequences with those for the chloroplast gene *rbcL* using a taxonomically similar suite of 59 angiosperms. Pairwise comparisons showed that *rbcL* is generally about three times more variable than 18S rDNA. However, because of the longer length of 18S rDNA, the ratio of the number of phylogenetically informative sites per molecule is only about 1.4 times greater for *rbcL* than for 18S rDNA. Exploratory parsimony analyses of angiosperms showed that several clades were strongly supported by both *rbcL* and 18S rDNA data sets. Nickrent and Soltis (1995) concluded that complete 18S rDNA sequences are sufficiently variable to conduct phylogenetic studies at higher levels within the angiosperms.

Here we explore further the higher-level phylogenetic relationships within the angiosperms using entire nuclear 18S rDNA sequences. More specifically, we provide phylogenetic hypotheses for flowering plants based on analyses of four 18S rDNA data sets, differing in both the number of taxa and the inclusion of indels as additional characters. We also compare the phylogenetic estimates based on 18S rDNA sequences with those obtained from phylogenetic analysis of *rbcL* sequences (Chase et al., 1993). Using the phylogenetic estimates, we examine patterns of molecular evolution of 18S rDNA by assessing the frequency of insertions and deletions, the prevalence of compensatory changes, and the relative phylogenetic importance of stem versus loop changes in angiosperm 18S rDNA.

## MATERIALS AND METHODS

### SPECIES SAMPLED AND SOURCES OF PLANT MATERIAL

The species included in this analysis are given in Table 1, along with family membership, general collection information, and GenBank accession numbers for the 18S sequences. In Table 1, and throughout the text, we generally follow the taxonomic circumscriptions of Cronquist (1981) for dicots and Dahlgren et al. (1985) for monocots. This

Table 1. Species analyzed for 18S rDNA sequence variation. Species are arranged alphabetically by families (dicots according to Cronquist; monocots according to Dahlgren). Within families, species are arranged alphabetically by genus. † indicates those angiosperms present in the 228-taxon data sets, but not the 199-taxon data sets.

| Species  | Family            | Voucher/source           | Literature citation     | Genbank | Sequence by     |
|--|-------------------|--------------------------|-------------------------|---------|-----------------|
| <i>Pachystachys lutea</i> Nees                   | Acanthaceae       | Johnson 95-003, WS       |                         | L49290  | Johnson         |
| <i>Acer rubrum</i> L.                            | Aceraceae         | Soltis & Soltis 2515, WS |                         | U42494  | Soltis          |
| <i>Actinidia</i> sp.                             | Actinidiaceae     | Morgan s.n., WS          |                         | U42495  | Soltis          |
| <i>Tetragonia expansa</i> Murr.                  | Aizoaceae         | Herikowitz 111, WS       |                         | U42496  | Soltis          |
| <i>Sagittaria trifoliata</i> L.                  | Alismataceae      | Chaw 1371, HAST          |                         | D29781  | Chaw            |
| † <i>Allium thunbergii</i> G. Don                | Alliaceae         | NA 55049, US             |                         | U42071  | Hahn & Kress    |
| † <i>Eucharis grandiflora</i> Planch & Linden    | Amaryllidaceae    | Hahn 6868, WIS           |                         | U42069  | Hahn            |
| <i>Hippeastrum</i> sp.                           | Amaryllidaceae    | Hahn 6875, WIS           |                         | U42065  | Hahn            |
| <i>Amborella trichopoda</i> Baill.               | Amborellaceae     | Suh 44, US               |                         | U42497  | Soltis & Soltis |
| † <i>Isolona</i> sp.                             | Annonaceae        | Chase 542, K             |                         | L54061  | Chase           |
| <i>Mikania fragrans</i>                          | Annonaceae        | Schultz 3364, WIS        |                         | L54060  | Hoot            |
| <i>Chlorophytum nepalense</i> Baker              | Antheriaceae      | Kress 92-3434, US        |                         | U42066  | Hahn & Kress    |
| <i>Lomatium triternatum</i> (Pursh) Coul. & Rose | Apiaceae          | Soltis 2266, WS          |                         | U42498  | Soltis & Soltis |
| <i>Tabernaemontana divaricata</i> (L.) R. Br.    | Apocynaceae       | Nickrent 2978, SIU       |                         | U42499  | Soltis & Soltis |
| † <i>Acorus calamus</i> L.                       | Araceae           | Nickrent 2941, SIU       | Nickrent & Soltis, 1995 | L24078  | Nickrent        |
| <i>Callia palustris</i> L.                       | Araceae           | Hahn 6959, WIS           |                         | U42073  | Hahn            |
| <i>Veitchia sessilifolia</i> Burret              | Arecaceae         | Hahn, US                 |                         | U42070  | Hahn            |
| <i>Hedera helix</i> L.                           | Araliaceae        | Plunkett 1368, WS        |                         | U42500  | Soltis & Soltis |
| <i>Aristolochia tomentosa</i> Sims.              | Aristolochiaceae  | Nickrent 2922, SIU       | Nickrent & Soltis, 1995 | L24083  | Nickrent        |
| <i>Asarum canadense</i> L.                       | Aristolochiaceae  | Nickrent 2888, SIU       | Nickrent & Starr, 1994  | L24043  | Nickrent        |
| † <i>Saurau henryi</i> Oliver                    | Aristolochiaceae  | Qiu 91018, NEU           | Nickrent & Soltis, 1995 | L24417  | Nickrent        |
| <i>Tagetes</i> sp.                               | Asteraceae        | Nickrent 3061, SIU       |                         | U42501  | Soltis & Soltis |
| <i>Tragopogon dubius</i> Scop.                   | Asteraceae        | Soltis 2472, WS          |                         | U42502  | Soltis & Soltis |
| <i>Austrobaileya scandens</i> C. T. White        | Austrobaileyaceae | Nickrent 2953, SIU       |                         | U42503  | Nickrent        |
| <i>Impatiens wailleana</i> Hook.                 | Balsaminaceae     | Johnson 95-071, WS       |                         | L49285  | Johnson         |
| <i>Batis maritima</i> L.                         | Bataceae          | Ilits 30500, WIS         |                         | U42504  | Soltis & Soltis |

Table 1. Continued.

| Species   | Family           | Voucher/source       | Literature citation        | Genbank | Sequence by     |
|---|------------------|----------------------|----------------------------|---------|-----------------|
| <i>Bauera rubioides</i> Andrews                       | Baueraaceae      | Kew 1977-6377, K     |                            | U42505  | Soltis          |
| <i>Begonia metallica</i> × <i>sanguinea</i>           | Begoniaceae      | Chase 225, NCU       |                            | U42506  | Soltis          |
| <i>Symbegonia</i> sp.                                 | Begoniaceae      | Kew U012, K          |                            | U43613  | Swensen         |
| <i>Caulophyllum thalictroides</i> (L.) Michx.         | Berberidaceae    | Hoot 925, UWM        |                            | L54064  | Hoot            |
| † <i>Podophyllum peltatum</i> L.                      | Berberidaceae    | Nickrent 2891, SIU   |                            | L24413  | Nickrent        |
| <i>Alnus glutinosa</i> L. Gaertn.                     | Betulaceae       | unknown              | Savard & Lalonde, 1991     | X59984  |                 |
| <i>Parmentiera cerifera</i> Seem.                     | Bignoniaceae     | Johnson 95-005, WS   |                            | L49291  | Johnson         |
| <i>Bombax ceiba</i> Burm.                             | Bombacaceae      | Alverson s.n., WS    |                            | U42507  | Kuzoff          |
| <i>Bourreria succulenta</i> Jacq.                     | Boraginaceae     | J. Miller 6421, MO   |                            | U38319  | Nickrent        |
| <i>Arabisopsis thaliana</i> (L.) Heyn.                | Brassicaceae     | unknown              | Unfried et al., 1989       | X16077  |                 |
| <i>Brassica hirta</i> Moench                          | Brassicaceae     | unknown              | Rathgeber & Capesius, 1990 | X17062  |                 |
| <i>Glomeropitcairnia penduliflora</i> (Grisebach) Mez | Bromeliaceae     | Kress 92-3466, US    |                            | U42075  | Hahn & Kress    |
| <i>Berzèia lanuginosa</i> (L.) Brongn.                | Bruniaceae       | Price s.n., IND      |                            | U42508  | Soltis          |
| <i>Buddleia davidii</i> Franch.                       | Buddlejaceae     | Johnson 95-031, WS   |                            | L49275  | Johnson         |
| <i>Buxus sempervirens</i> L.                          | Buxaceae         | Hoot 921, UWM        |                            | L54065  | Hoot            |
| <i>Byblis gigantea</i>                                | Byblidaceae      | Palmengarten B. G.   |                            | U42509  | Soltis & Soltis |
| <i>Calycanthus floridus</i> L.                        | Calycanthaceae   | Nickrent 2893, SIU   |                            | U38318  | Nickrent        |
| <i>Campanula ramulosa</i> Wall.                       | Campanulaceae    | Jansen 984, MICH     |                            | U42510  | Soltis & Soltis |
| <i>Lobelia erina</i> L.                               | Campanulaceae    | Jansen 989, MICH     |                            | U42785  | Soltis & Soltis |
| <i>Canna coccinea</i> Mill.                           | Cannaceae        | Chaw 1371, HAST      |                            | D29784  | Chaw            |
| <i>Cleome hassleriana</i> Chodat                      | Capparaceae      | Al-Shehbaz, s.n., MO |                            | U42511  | Soltis & Soltis |
| <i>Koeberlinia spinosa</i> Zucc.                      | Capparaceae      | Al-Shehbaz, s.n., MO |                            | U42512  | Soltis & Soltis |
| † <i>Lonicera maackii</i> (Rupr.) Maxim.              | Caprifoliaceae   | Nickrent 3060, SIU   |                            | U66701  | Soltis & Soltis |
| <i>Symphoricarpos albus</i> (L.) Blake                | Caprifoliaceae   | Olmstead s.n., COLO  |                            | U42513  | Soltis & Soltis |
| <i>Carica papaya</i> L.                               | Caricaceae       | Missouri B. G., MO   |                            | U42514  | Soltis & Soltis |
| <i>Casuarina equisetifolia</i> L.                     | Casuarinaceae    | Nickrent 2977, SIU   |                            | U42515  | Soltis          |
| <i>Euonymus alatus</i> (Thunb.) Siebold               | Celastraceae     | Nickrent 2894, SIU   | Nickrent & Franchina, 1990 | X16600  | Nickrent        |
| <i>Cephalotus foliolaris</i> Labill.                  | Cephalotaceae    | Chase 147, NCU       |                            | U42516  | Soltis          |
| <i>Ceratophyllum demersum</i> L.                      | Ceratophyllaceae | Quiu 91027, NCU      |                            | U42517  | Soltis          |

Table 1. Continued.

| Species   | Family            | Voucher/source               | Literature citation     | Genbank | Sequence by     |
|---|-------------------|------------------------------|-------------------------|---------|-----------------|
| <i>Cercidiphyllum japonicum</i> Siebold & Zucc. | Cercidiphyllaceae | Soltis 2540, WS              |                         | U42518  | Soltis          |
| <i>Spinacia oleracea</i> L.                     | Chenopodiaceae    | Nickrent 2896, SIU           | Nickrent & Soltis, 1995 | L24420  | Nickrent        |
| <i>Hedysmum arborescens</i> Sw.                 | Chloranthaceae    | Nickrent 3022, SIU           |                         | U38536  | Nickrent        |
| <i>Chrysobalanus icaco</i> L.                   | Chrysobalanaceae  | Fairchild Trop. G 76-311     |                         | U42519  | Soltis          |
| <i>Licania tomentosa</i> Fritsch                | Chrysobalanaceae  | Fairchild Trop. B. G. 64-734 |                         | U42520  | Soltis          |
| <i>Clethra alnifolia</i> L.                     | Clethraceae       | Kron 1884s, NCU              |                         | U42521  | Soltis          |
| <i>Colchicum autumnale</i> L.                   | Colchicaceae      | Hahn 6864, WIS               |                         | U42072  | Hahn            |
| <i>Elatis</i> sp.                               | Commelinaceae     | Evans s.n., WIS              |                         | U42076  | Hahn            |
| <i>Ipomoea hederacea</i> Jacq.                  | Convolvulaceae    | Cobwell s.n., MO             |                         | U38310  | Nickrent        |
| <i>Aucuba japonica</i> Thunb.                   | Cornaceae         | U.S. Natl. Arb.              |                         | U42522  | Soltis          |
| <i>Corokia cotoneaster</i> Raoul                | Cornaceae         | Strybing Arb. 74211          |                         | U42523  | Soltis & Soltis |
| <i>Helwingia japonica</i> (Thunb.) F. Dietr.    | Cornaceae         | Arnold Arb. 912              |                         | U42524  | Soltis & Soltis |
| <i>Costus barbatus</i> Suess.                   | Costaceae         | Kress 94-3710, US,           |                         | U42080  | Hahn & Kress    |
| <i>Crassula marnierana</i> Huber & Jacobsen     | Crassulaceae      | Morgan 2152, WS              |                         | U42525  | Soltis          |
| <i>Dudleya viscidula</i> (S. Watson) Moran      | Crassulaceae      | Huntington B. G. 62801       |                         | U42526  | Soltis          |
| <i>Kalanchoe diademantana</i> Hamet & Perrier   | Crassulaceae      | Morgan 2151, WS              |                         | U42527  | Soltis          |
| <i>Sedum rubrotinctum</i> Clausen               | Crassulaceae      | Morgan 2153, WS              |                         | U42528  | Soltis          |
| <i>Crossosoma californicum</i> Nutt.            | Crossosomataceae  | Rancho Santa Ana B. G.       |                         | U42529  | Soltis          |
| <i>Abobra tenuiflora</i> Naudin.                | Cucurbitaceae     | Chase 915, K                 |                         | U41501  | Swensen         |
| <i>Ceraopectalum gummiferum</i> Small           | Cunoniaceae       | Keller 2135, CAS             |                         | U42530  | Soltis          |
| <i>Cuscuta gronovi</i> Willd.                   | Cuscutaceae       | Nickrent 3015, SIU           | Nickrent & Soltis, 1995 | L24747  | Nickrent        |
| <i>Cyperus albobristatus</i> Schrader           | Cyperaceae        | Kress 92-3463, US            |                         | U42077  | Hahn & Kress    |
| <i>Cyrilla racemifolia</i> L.                   | Cyrtaceae         | Kron s.n., NCU               | Kron, 1996              | U43294  | Kron            |
| <i>Daphniphyllum</i> sp.                        | Daphniphyllaceae  | Qiu 91026, NCU               |                         | U42531  | Soltis          |
| <i>Datisca glomerata</i> (Presl) Baill.         | Datisceae         | Rancho Santa Ana B. G.       |                         | U42426  | Swensen         |

Table 1. Continued.

| Species                                     | Family         | Voucher/source   | Literature citation    | Genbank | Sequence by     |
|---|----------------|--|------------------------|---------|-----------------|
| <i>Terameles nudiflora</i> R. Br.           | Datiaceae      | Central B. G., Sara Bori, Thailand; Philbrick & Wongprasert 2272 |                        | U41502  | Swensen         |
| <i>Diapensia lapponica</i> L.               | Diapensiaceae  | Hills 89018, NCU   |                        | L49278  | Johnson         |
| <i>Galax urceolata</i> L.                   | Diapensiaceae  | Kron 163, NCU  |                        | L49281  | Johnson         |
| † <i>Dillenia alata</i> (DC) Mart.          | Drilleniaceae  | Nickrent 2956, SIU   |                        | U38314  | Nickrent        |
| <i>Dipsacus</i> sp.                         | Dipsacaceae    | Jansen 931, MICH   |                        | U43150  | Soltis & Soltis |
| <i>Diospyros virginiana</i> L.              | Droseraceae    | Palmengarten B. G.   |                        | U42532  | Soltis          |
| † <i>Elaeagnus umbellata</i> Thunb.         | Ebenaceae      | Kron 3004, NCU   |                        | U43295  | Kron            |
| <i>Sloanea cf. latifolia</i> K. Sch.        | Elaeagnaceae   | Nickrent 2898, SIU   |                        | L24090  | Nickrent        |
| <i>Ephedra sinica</i> Stapf                 | Elaeocarpaceae | Alberson s.n., WIS   |                        | U42826  | Kuzoff          |
| <i>Ephedra torreyana</i> S. Wats.           | Ephedraceae    | TI-9297, TI  |                        | D38242  | Chaw            |
| <i>Arctostaphylos uva-ursi</i> (L.) Spreng. | Ephedraceae    | Gillespie 4236, US   |                        | U42414  | Sweere          |
| <i>Vaccinium macrocarpon</i> Ait.           | Ericaceae      | Johnson 94-085, WS   |                        | L49272  | Johnson         |
| † <i>Eucommia almoides</i> Oliv.            | Ericaceae      | Kron 2937, NCU   |                        | L49297  | Johnson         |
| <i>Eucryphia lucida</i> Druce               | Eucommiaceae   | Qiu 91024, NCU   |                        | L54066  | Hoot            |
| <i>Drypetes roxburghii</i> (Wall.) Hu-rus.  | Eucryphiaceae  | Strybing Arb. 86-0250  |                        | U42533  | Soltis          |
| <i>Euphorbia pulcherrima</i>                | Euphorbiaceae  | FTC-83463A, K  |                        | U42534  | Soltis          |
| <i>Euptelea polyandra</i> Siebold & Zucc.   | Euphorbiaceae  | Soltis & Soltis 2541, WS   |                        | U42535  | Soltis          |
| <i>Albizia julibrissin</i> Durazz.          | Euphorbiaceae  | Qiu 9001, NCU  |                        | L75831  | Hoot            |
| <i>Bauhinia</i> sp.                         | Fabaceae       | Doyle 1526, BH   |                        | U42536  | Soltis          |
| † <i>Glycine max.</i> (L.) Merrill          | Fabaceae       | Doyle s.n., MSU  |                        | U42537  | Soltis          |
|   | Fabaceae       | unknown  | Eckernode et al., 1985 | X02623  |                 |
| <i>Pisum sativum</i> L.                     | Fabaceae       | Carolina Biological Laboratory, Burlington N.C.                  |                        | U43011  | Sweere          |
| * <i>Muntingia calabura</i> Griseb.         | Flacourtiaceae | Alberson 2172, WIS   |                        | U42539  | Kuzoff          |
| <i>Dicentra exima</i> Torrey                | Fumariaceae    | Reznicek 9756, MICH  |                        | L37908  | Hoot            |
| <i>Fouquieria splendens</i> Engelm.         | Fouquieriaceae | Missouri B. G. 860162, MO  |                        | L49280  | Johnson         |
| <i>Garrya elliptica</i> Douglas ex Lind.    | Garryaceae     | Rancho Santa Ana B. G. 13280                                     |                        | U42540  | Soltis          |
| <i>Geranium cinereum</i> Cav.               | Geraniaceae    | Price s.n., IND  |                        | U42541  | Soltis          |

\*A recent detailed phylogenetic analysis of Flacourtiaceae using *rbcL* sequences indicates that *Muntingia* is only distantly related to the "true" flacourts (Chase, unpublished).



Table 1. Continued.

| Species   | Family          | Voucher/source                             | Literature citation | Genbank | Sequence by     |
|---|-----------------|--|---------------------|---------|-----------------|
| <i>Aeschynanthus radicans</i> Jack.                 | Gesneriaceae    | Nickrent 2979, SIU                         |                     | U42542  | Soltis & Soltis |
| <i>Gnetum gnetum</i> L.                             | Gnetaceae       | Gillespie 4212, US                         |                     | U42416  | Sweere          |
| <i>Gnetum nodiflorum</i> Brongn.                    | Gnetaceae       | Gillespie 4246, US                         |                     | U42415  | Sweere          |
| <i>Gnetum urens</i> (Aubl.) Blume                   | Gnetaceae       | Kress et al. 91-3271, US                   |                     | U42417  | Sweere          |
| <i>Greigia radlkoferi</i> Szysyl.                   | Greyniaceae     | Strybing Arb. 640406                       |                     | U43151  | Soltis          |
| <i>Brexia madagascariensis</i> Thouars ex Ger Gawl. | Grossulariaceae | Kew 1977-14901, K                          |                     | U42543  | Soltis          |
| <i>Escallonia coquimbensis</i> Reamy                | Grossulariaceae | U. Calif. B. G. 52-1333                    |                     | U42544  | Soltis & Soltis |
| <i>Itea virginica</i> L.                            | Grossulariaceae | Ware 9401, WS                              |                     | U42545  | Soltis          |
| <i>Phyllonoma laucuspis</i> (Turcz.) Engl.          | Grossulariaceae | Morgan 2124, WS                            |                     | U42546  | Soltis          |
| <i>Pterostemon rotundifolius</i> Ramirez            | Grossulariaceae | Sanchez 259, TEX                           |                     | U42547  | Soltis          |
| <i>Ribes aureum</i> Pursh                           | Grossulariaceae | Soltis & Soltis 2220, WS                   |                     | L28143  | Soltis          |
| <i>Roussea simplex</i> J. E. Smith                  | Grossulariaceae | Herbarium, Mauritius Sugar Ind. Res. Inst. |                     | U42548  | Soltis          |
| <i>Tetracarpaea tasmanica</i> Hook f.               | Grossulariaceae | Jordan s.n., HO                            |                     | U42549  | Soltis          |
| <i>Gunnera manicata</i> Linden                      | Gunneraceae     | Kruckeberg s.n., WTU                       |                     | U43787  | Soltis          |
| <i>Haloragis erecta</i> (Banks ex Murr.) Eichler    | Haloragaceae    | Chase 453, K                               |                     | U42550  | Soltis          |
| <i>Alungia</i> sp.                                  | Hamamelidaceae  | RBC, Edinburgh, Qiu 93006                  |                     | U42552  | Soltis          |
| <i>Liquidambar styraciflua</i> L.                   | Hamamelidaceae  | Soltis & Soltis 2516, WS                   |                     | U42553  | Soltis          |
| <i>Heliconia indica</i> Lamark                      | Heliconiaceae   | Kress 80-1118, US                          |                     | U42082  | Hahn & Kress    |
| <i>Bowiea volubilis</i> Harvey ex Hook f.           | Hyacinthaceae   | Hahn 6882, WIS                             |                     | U42067  | Hahn            |
| <i>Hydrangea macrophylla</i> Torr.                  | Hydrangeaceae   | Morgan 2150, WS                            |                     | U42781  | Soltis          |
| <i>Philadelphus lewisii</i> Pursh.                  | Hydrangeaceae   | Soltis & Soltis 2411, WS                   |                     | U42782  | Soltis          |
| <i>Phacelia bicolor</i> Torr. ex Wats.              | Hydrophyllaceae | Johnson 92-005, WS                         |                     | L49292  | Johnson         |
| <i>Illicium parviflorum</i> Michx. ex Vent.         | Illiciaceae     | Nazzi 2784, MICH                           |                     | L75832  | Hoot            |
| <i>Gladiolus buckenildii</i> (L. Bol.) Goldblatt    | Iridaceae       | Goldblatt & Manning 9504, MO               |                     | L54062  | Hoot            |
| † <i>Isophysis tasmanica</i> (Hook.) T. Moore       | Iridaceae       | J. Bruhl s.n., TAS                         |                     | L64063  | Hoot            |
| <i>Lactoris fernandeziana</i> Phil.                 | Lactoridaceae   | Steussy et al. 11,784, OSU                 |                     | U42783  | Soltis          |

Table 1. Continued.

| Species                                 | Family          | Voucher/source           | Literature citation     | Genbank | Sequence by     |
|---|-----------------|--------------------------|-------------------------|---------|-----------------|
| <i>Lamium amplexicaule</i> L.           | Lamiaceae       | Johnson 95-001, WS       |                         | U49287  | Johnson         |
| <i>Akebia quinata</i> (Houtt.) Decne.   | Lardizabalaceae | Nickrent 2945, SIU       |                         | L31795  | Nickrent        |
| <i>Sassafras albidum</i> (Nutt.) Nees   | Lauraceae       | Soltis & Soltis 2518, WS |                         |         |                 |
| † <i>Lilium formosanum</i> Wallace      | Liliaceae       | H0962, HAST              |                         | D29775  | Chaw            |
| <i>Floerkea proserpinicoides</i> Willd. | Limnanthaceae   | Reznicek 8609, MICH      |                         | U42784  | Soltis & Soltis |
| <i>Linum perenne</i> L.                 | Linaceae        | Nickrent 2900, SIU       |                         | L24401  | Nickrent        |
| <i>Malpighia coccigera</i> L.           | Malpighiaceae   | Nickrent 2905, SIU       |                         | U42046  | Nickrent        |
| <i>Gossypium hirsutum</i> L.            | Malvaceae       | Abrerson s.n., WS        |                         | U42827  | Kuzoff          |
| <i>Maranta bicolor</i> Vell.            | Marantaceae     | Kress 94-3724, US        |                         | U42079  | Hahn & Kress    |
| <i>Menispermum canadensis</i> L.        | Menispermaceae  | Nazi 2837, MICH          |                         | L75834  | Hoot            |
| † <i>Tinospora caffra</i> Miets         | Menispermaceae  | Jaarsveld 2131, NBG      |                         | L37914  | Hoot            |
| <i>Mollugo verticillata</i> L.          | Molluginaceae   | Hershkovitz 37, WS       |                         | U42828  | Kuzoff          |
| † <i>Monotropa uniflora</i> L.          | Monotropaceae   | Nickrent 3018, SIU       | Nickrent & Soltis, 1995 | L25680  | Nickrent        |
| <i>Morus alba</i> L.                    | Moraceae        | Nickrent 2924, SIU       |                         | L24398  | Nickrent        |
| <i>Moringa oleifera</i> Lam.            | Moringaceae     | Ilitis 30501, WS         |                         | U42786  | Soltis & Soltis |
| <i>Musa acuminata</i> Colla             | Musaceae        | U.S. Bot. Gard. s.n., US |                         | U42083  | Hahn & Kress    |
| <i>Nelumbo lutea</i> (Willd.) Pers.     | Nelumbonaceae   | Hoot 9212, 1713          |                         | L75835  | Hoot            |
| <i>Nepenthes</i> sp.                    | Nepenthaceae    | Nickrent 3056, SIU       |                         | U42787  | Soltis          |
| <i>Nymphaea tuberosa</i> Paine          | Nymphaeaceae    | Nickrent 2906, SIU       | Nickrent & Soltis, 1995 | L24404  | Nickrent        |
| <i>Mirabilis jalapa</i> L.              | Nyctaginaceae   | Hershkovitz 60, WS       |                         | U42788  | Soltis          |
| <i>Campithecra acuminata</i> Decne.     | Nyssaceae       | Strybing Arb. 74-180     |                         | U42789  | Soltis & Soltis |
| <i>Olea europaea</i> L.                 | Oleaceae        | Johnson 95-004, WS       |                         | L49289  | Johnson         |
| † <i>Clarkia xaniana</i> A. Gray        | Onagraceae      | Gottlieb 7436, DAV       |                         | U67930  | Sytma           |
| <i>Opilia amentacea</i> Roxb.           | Opiliaceae      | Nickrent 2816, SIU       |                         | U42790  | Soltis          |
| <i>Oncidium excavatum</i> Lindl.        | Orchidaceae     | Chase 83427, K           |                         | U42791  | Soltis          |
| <i>Paeonia suffruticosa</i>             | Paeoniaceae     | Chase 486, K             |                         | U42792  | Soltis          |
| <i>Hypocoum imberba</i> Sm.             | Papaveraceae    | Chase 528, K             |                         | L75836  | Hoot            |
| <i>Helmholtzia</i> sp.                  | Phyllodraceae   | Kress 92-3505, US        |                         | U42074  | Hahn & Kress    |
| <i>Phytolacca americana</i> L.          | Phytolaccaceae  | Hershkovitz 38, WS       |                         | U42793  | Soltis          |
| <i>Peperomia serpens</i> (Swartz) Loud. | Piperaceae      | Nickrent 2907, SIU       | Nickrent & Soltis, 1995 | L24411  | Nickrent        |

Table 1. Continued.

| Species   | Family         | Voucher/source   | Literature citation     | Genbank | Sequence by              |
|---|----------------|--|-------------------------|---------|--------------------------|
| <i>Pittosporum japonicum</i> Hort. ex C. Presl.             | Pittosporaceae | Rieseberg s.n., RSA                                      | Nickrent & Soltis, 1995 | L28142  | Soltis & Kuzoff          |
| <i>Platanus occidentalis</i> L.                             | Platanaceae    | Soltis & Soltis 2514, WS                                 |                         | U42794  | Sweere, Zimmer, & Soltis |
| <i>Plumbago auriculata</i> Lam.                             | Plumbaginaceae | Nickrent s.n., SIU                                       |                         | U42795  | Kuzoff                   |
| <i>Oryza sativa</i> L.                                      | Poaceae        | unknown  | Takaiwa et al., 1984    | X00755  | Takaiwa et al., 1984     |
| <i>Zea mays</i> L.  | Poaceae        | (Genomic DNA) Clontech Laboratories, Inc., Palo Alto, CA |                         | U42796  | Sweere, Zimmer, & Soltis |
| <i>Acanthogilia gloriosa</i> (Brandegee) A. G. Day & Moran  | Polemoniaceae  | Johnson & Mori, 95-070, WS                               |                         | L49271  | Johnson                  |
| <i>Cobaea scandens</i> Cav.                                 | Polemoniaceae  | Patterson s.n., WS                                       |                         | L49277  | Johnson                  |
| <i>Gilia capitata</i> Sims.                                 | Polemoniaceae  | Johnson 92-15, WS  | Nickrent & Soltis, 1995 | L24143  | Soltis & Kuzoff          |
| <i>Polygala pauciflora</i> Willd.                           | Polygalaceae   | Doyle 1567, BH   |                         | U42797  | Soltis                   |
| <i>Coccoloba uvifera</i> L.                                 | Polygonaceae   | Nickrent 2927, SIU                                       |                         | U42798  | Sweere, Zimmer, & Soltis |
| <i>Primula</i> sp.  | Primulaceae    | Johnson 95-006, WS                                       |                         | L49295  | Johnson                  |
| <i>Knightsia excelsa</i> R. Br.                             | Proteaceae     | Univ. of California, Santa Clara, B. G.                  | Nickrent & Soltis, 1995 | L24155  | Nickrent                 |
| † <i>Placospermum coriaceum</i> C. T. White & W. D. Francis | Proteaceae     | Douglas 110, MEL   |                         | L75837  | Hoot                     |
| † <i>Punica granatum</i> L.                                 | Punicaceae     | Nickrent 2931, SIU                                       |                         | U38311  | Nickrent                 |
| † <i>Pyrola picta</i> Sm.                                   | Pyrolaceae     | Cobwell, s.n., MD  |                         |         | Colwell & Nickrent       |
| † <i>Coptis trifolia</i> (L.) Salisb.                       | Ranunculaceae  | Voss & Howard, MICH                                      |                         | L75838  | Hoot                     |
| <i>Ranunculus sardous</i> Crantz                            | Ranunculaceae  | Nickrent 2932, SIU                                       | Nickrent & Soltis, 1995 | L24092  | Nickrent                 |
| <i>Xanthorhiza simplicissima</i> Marshall                   | Ranunculaceae  | Qui 91030, NCU   |                         | L75839  | Hoot                     |
| <i>Ceanothus sanguineus</i> Pursh                           | Rhamnaceae     | Morgan 2155, WS  |                         | U42799  | Soltis                   |

Table 1. Continued.

| Species   | Family           | Voucher/source              | Literature citation     | Genbank | Sequence by             |
|---|------------------|-----------------------------|-------------------------|---------|-------------------------|
| <i>Photinia Fraseri</i> Dress                       | Rosaceae         | Morgan 2131, WS             |                         | U42800  | Soltis                  |
| <i>Prunus persica</i> (L.) Batsch                   | Rosaceae         | E. E. Dickson s.n., BH      |                         | L28749  | Soltis                  |
| <i>Spiraea vanhouttei</i> (Briot) Zubei             | Rosaceae         | Morgan 2130, WS             |                         | U42801  | Soltis                  |
| <i>Michelia repens</i> L.                           | Rubiaceae        | Xiang s.n., OSU             |                         | U42802  | Soltis & Soltis         |
| † <i>Cirsium aurantiacum</i> L.                     | Rutaceae         | Nickrent 2977, SIU          |                         | U38312  | Nickrent                |
| <i>Sabia swinhoei</i>                               | Sabiaceae        | Wagner 6518, HAST           |                         | L75840  | Hoot                    |
| <i>Ostrya lanceolata</i> Hochst. & Steud.           | Santalaceae      | Nickrent 2731, SIU          |                         | U42803  | Soltis                  |
| † <i>Koeleruteria paniculata</i> Laxm.              | Sapindaceae      | Nickrent 2915, SIU          |                         | U38313  | Nickrent                |
| <i>Manilkara zapota</i> (L.) Royen                  | Sapotaceae       | Chase 129, NCU              | Kron, 1996              | U43080  | Kron                    |
| † <i>Sargentodoxa cuneata</i> (Oliv.) Rehder & Wil. | Sargentodoxaceae | Qiu s.n., PE                |                         | L75841  | Hoot                    |
| <i>Sarracenia purpurea</i> L.                       | Sarraceniaceae   | Morgan s.n., WS             |                         | U42804  | Soltis                  |
| <i>Houttuynia cordata</i> Thunb.                    | Saururaceae      | Nickrent 2940, SIU          | Nickrent & Soltis, 1995 | L24147  | Nickrent                |
| <i>Saururus cernuus</i> L.                          | Saururaceae      | Suh 128, US                 |                         | U42805  | Sweere, Zimmer & Soltis |
| <i>Boykinia intermedia</i> (Piper) Jones            | Saxifragaceae    | Grable 11638, WS            |                         | U42806  | Soltis                  |
| <i>Eremosyne pectinata</i> Endl.                    | Saxifragaceae    | Annelis & Hearn 4795, PERTH |                         | U42807  | Soltis                  |
| <i>Francoa sonchifolia</i> Cav.                     | Saxifragaceae    | Soltis & Soltis 2479, WS    | Nickrent & Soltis, 1995 | L28137  | Soltis                  |
| <i>Heuchera micrantha</i> Douglas                   | Saxifragaceae    | Soltis & Soltis 1949, WS    |                         | X28139  | Soltis                  |
| <i>Lepuropetalon spathulatum</i> (Muhl.) Elliott    | Saxifragaceae    | Thomas s.n., NLU            |                         | L28141  | Soltis                  |
| <i>Montinia caryophyllacea</i> Thunb.               | Saxifragaceae    | Williams 2833, MO           |                         | U42808  | Soltis & Soltis         |
| <i>Parnassia fimbriata</i> Banks                    | Saxifragaceae    | Soltis & Soltis s.n., WS    |                         | U42809  | Soltis                  |
| <i>Penthorum sedoides</i> L.                        | Saxifragaceae    | Hayden 2232, WS             |                         | U25660  | Soltis                  |
| <i>Saxifraga integrifolia</i> Hook.                 | Saxifragaceae    | Soltis & Soltis 2253, WS    |                         | U42810  | Soltis                  |
| <i>Vahlia capensis</i> Thunb.                       | Saxifragaceae    | Van Wyk 10-579, PRU         |                         | U42813  | Soltis & Soltis         |
| <i>Schizandra chinensis</i> (Turcz.) Baill.         | Schisandraceae   | Reznicek s.n., MICH         |                         | L75842  | Hoot                    |
| † <i>Linaria vulgaris</i> P. Mill                   | Scrophulariaceae | Colwell MO CA1, SIU         |                         | U38315  | Colwell & Nickrent      |

Table 1. Continued.

| Species   | Family           | Voucher/source                      | Literature citation     | Genbank | Sequence by              |
|---|------------------|-------------------------------------|-------------------------|---------|--------------------------|
| † <i>Orthocarpus erianthus</i> Benth.                       | Scrophulariaceae | <i>Colwell</i> CA SVD10, SIU        |                         | U38316  | Colwell & Nickrent       |
| † <i>Pedicularis lanceolata</i> Michx. Schlechtend.) Benth. | Scrophulariaceae | <i>Colwell</i> MO SFPL, SIU         |                         | U38317  | Colwell & Nickrent       |
| <i>Brunfelsia pauciflora</i> (Cham. & Schlechtend.) Benth.  | Solanaceae       | <i>Johnson</i> 95-002, WS           |                         | L49274  | Johnson                  |
| † <i>Sparganium eurycarpum</i> Engelm.                      | Sparganiaceae    | <i>Nickrent</i> 2943, SIU           | Nickrent & Soltis, 1995 | L24419  | Nickrent                 |
| † <i>Styrax americana</i> Lam.                              | Styracaceae      | <i>Kron</i> 3002, NCU               | Kron, 1996              | U43296  | Kron                     |
| <i>Symplocos paniculata</i> Miq.                            | Symplocaceae     | <i>Kron</i> 3005, NCU               | Kron, 1996              | U43297  | Kron                     |
| <i>Tacca</i> sp.  | Taccaceae        | Missouri Bot. Garden 894904, US     |                         | U42063  | Hahn                     |
| <i>Cyanella capensis</i> L.                                 | Tecophilaceae    | <i>Hahn</i> 6966, WIS               |                         | U42078  | Hahn                     |
| <i>Tetracentron sinensis</i> Oliv.                          | Tetracentraceae  | <i>Qiu</i> 90009, NCU               |                         | U42814  | Soltis                   |
| <i>Camelia japonica</i> L.                                  | Theaceae         | <i>Nickrent</i> 2929, SIU           |                         | U42815  | Soltis                   |
| <i>Luehea seemannii</i> Cuatr.                              | Tiliaceae        | <i>Alverson</i> 2184, WIS           |                         | U42829  | Kuzoff                   |
| <i>Trochodendron aralioides</i> Siebold & Zucc.             | Trochodendraceae | <i>Qiu</i> 90026, NCU               |                         | U42816  | Soltis                   |
| <i>Tropaeolum majus</i> L.                                  | Tropaeolaceae    | <i>Chase</i> 113, NCU               | Nickrent & Soltis, 1995 | L31796  | Soltis                   |
| <i>Turnera ulmifolia</i>                                    | Turneraceae      | <i>Chase</i> 220, NCU               |                         | U42817  | Soltis                   |
| <i>Celtis yunnanensis</i> C. K. Schneid.                    | Ulmaceae         | <i>Qiu</i> P90002, NCU              |                         | U42818  | Soltis                   |
| <i>Zeakoya serrata</i>                                      | Ulmaceae         | <i>Soltis &amp; Soltis</i> 2517, WS |                         | U42819  | Soltis                   |
| <i>Pilea cadieri</i> Gagnep. & Guillem.                     | Urticaceae       | <i>Nickrent</i> 2972, SIU           |                         | U42820  | Soltis                   |
| <i>Viscum album</i> L.                                      | Viscaceae        | <i>Nickrent</i> 2253, SIU           |                         | U42821  | Soltis                   |
| <i>Drimys aromatica</i>                                     | Winteraceae      | <i>Suh</i> 9, US                    |                         | U42822  | Soltis & Soltis          |
| <i>Drimys winteri</i> J. R. & G. Forster                    | Winteraceae      | <i>Suh</i> 47, US                   |                         | U42823  | Sweere, Zimmer, & Soltis |
| <i>Xanthorrhoea</i> sp.                                     | Xanthorrhoeaceae | <i>Kress</i> 92-3422, US            |                         | U42064  | Hahn & Kress             |
| <i>Zamia pumila</i> L.                                      | Zamiaceae        | unknown                             | Naim & Ferl, 1988       | M20017  | Naim & Ferl              |
| <i>Zingiber gramineum</i> Norton                            | Zingiberaceae    | <i>Kress</i> 91-3266, US            |                         | U42081  | Hahn & Kress             |
| <i>Guaiacum sanctum</i> L.                                  | Zygophyllaceae   | <i>Anderson</i> s.n., MICH          |                         | U42824  | Soltis                   |

approach parallels that in Chase et al. (1993) and should facilitate comparison.

As with the broad analyses of *rbcL* sequences, close examination of the genera included in this study will reveal an uneven taxonomic distribution. Some groups are relatively well represented (e.g., Saxifragaceae sensu stricto [Saxifragaceae s. str.] and allies, ranunculids, Asteridae sensu lato [Asteridae s.l.]), whereas others are not as thoroughly sampled (e.g., portions of Dilleniidae and Magnoliidae). However, our selection of taxa was not random. We attempted to include samples from all major angiosperm orders and subclasses sensu Cronquist (1981). Furthermore, in selecting genera for sequencing, we tried to sample representatives from each of the major clades recovered in the analyses of Chase et al. (1993) (e.g., rosoid I, asterid I, asterid III, etc.), as well as from the various subclades present within those major clades. We also used, when available, the same DNAs used previously for the sequencing of *rbcL* (Chase et al., 1993). If a given DNA was no longer available, we attempted to obtain leaf material of the same species, and if that failed, from a congeneric species.

Another factor that influenced our choice of taxa was sequence quality. As discussed in detail below, one outcome of this study was the discovery that many available sequences are erroneous, some highly so. We therefore attempted to eliminate any dubious sequences from our data sets. In addition, some available sequences were not included because they were incomplete or contained numerous ambiguities or extensive gaps.

Several laboratories were involved in this project; hence, several different protocols were used to generate the sequences analyzed. Although both automated and manual sequencing strategies were employed, 70% of the sequences analyzed were generated via automated sequencing. The general methods used for PCR amplification and subsequent manual sequencing of 18S rDNA are provided in Nickrent (1994), Nickrent and Starr (1994), and Bult et al. (1992). General methods for the automated sequencing approach for 18S rDNA are given in D. Soltis and Soltis (1997). The base composition of the oligonucleotide primers used for PCR and sequencing are provided in Nickrent and Starr (1994) and Bult et al. (1992).

#### ALIGNMENT OF THE 18S rDNA SEQUENCES

With the exception of a few, small, well-defined regions, alignment of 18S rDNA sequences was easily accomplished by eye across all taxa. This general ease of alignment is due not only to the

highly conserved nature of the 18S rRNA gene, but also to the fact that most length mutations involve insertions or deletions of a single base pair. Straightforward alignment of sequences was further facilitated by the fact that most indels in 18S rDNA are confined to a few specific regions that are particularly prone to variation in primary sequence and length, such as the termini of helices E10-1, 17, E23-1, and 43 (see also Nickrent & Soltis, 1995). Because they were difficult to align over a broad taxonomic scale, no attempt was made to align four small regions of 18S rDNA over all of the taxa analyzed: positions 230-237; 496-501; 666-672; 1363-1369 (see Appendix). These base positions correspond to the sequence of *Glycine max* (Eckenrode et al., 1985), which provides a convenient reference sequence because of the availability of a proposed ribosomal RNA secondary structure model (Nickrent & Soltis, 1995). These four regions of ambiguous alignment were subsequently eliminated from the phylogenetic analyses, following Swofford and Olsen (1990). In addition, the extreme 5' and 3' ends of the sequences were not included in the analyses. Positions 1-20 were excluded because they correspond to the forward PCR primer. Because most of the sequences were clearly readable at, or just before, base position 41, we began analysis of our data set at position 42. At the 3' end of the 18S sequences, base positions 1751-1808 (on *Glycine*) were often difficult to read and also were eliminated from the analysis. Some sequences are incomplete at the 3' end and are approximately 1700 base pairs in length. The total length of the aligned 18S rDNA sequences was 1850 base pairs.

Two indels, each of a single base pair, were detected in highly conserved regions of the 18S rRNA gene not prone to insertion-deletion (Table 2). One indel (A), an apparent deletion based on outgroup comparison, is present in all higher dicots (i.e., the large clade consisting of Rosidae and Asteridae s.l.). A second indel (B), an apparent insertion, occurs in all members of the saxifragoid clade (also referred to as Saxifragales; D. Soltis & Soltis, 1997). These two indels were included as characters in two of the phylogenetic analyses, as described below.

#### PHYLOGENETIC ANALYSIS

We constructed four data sets for phylogenetic analysis: (1) a data set of 223 angiosperms plus five members of Gnetales as outgroups, without the two indels noted above; (2) taxon sampling as in (1), but with indels A and B (see also Table 2) included;

Table 2. Potentially phylogenetically informative indels located in conserved regions of 18S rDNA. Indel A is one-bp deletion that characterizes all higher eudicots (i.e., the Rosidae and Asteridae s.l. clades). Indel B is a one-bp insertion that unites all members of the saxifragoid clade. Base positions correspond to the last position given in the sequence of *Glycine max*.

|                      |                   |      |
|----------------------|-------------------|------|
| Indel A              |                   | 1529 |
| <i>Glycine</i>       | CCGGGTAATCTTTG -  |      |
| <i>Trochodendron</i> | CCGGGTAATCTTTGA   |      |
| Indel B              |                   | 1406 |
| <i>Glycine</i>       | TATGGCCGCTTA -GGC |      |
| <i>Heuchera</i>      | TATGGCCGATTTAAGGC |      |

(3) a data set of 194 angiosperms, plus five Gnetales as outgroups, without the two indels; (4) taxon sampling as in (3) above, with indels A and B included. For data sets 2 and 4, the indels A and B were added to the data matrix, and the position scored as either present (1) or absent (0).

Four data sets were used for several reasons. First, the approach used permitted an assessment of the phylogenetic informativeness of the two indels. Second, our goal in constructing the two smaller data sets was to improve the phylogenetic analysis by removing incomplete and/or possibly erroneous sequences and by reducing the size of the matrix to make the problem more tractable. The two smaller data sets (3 and 4) differ from the larger data sets (1 and 2) in the removal of several taxa having long branch lengths (e.g., *Dillenia*, *Acorus*) and several taxa for which the sequences were incomplete (e.g., several of the ranunculids). In addition, representatives from some of the larger clades (e.g., monocots, Asteridae s.l.) and from some families for which multiple sequences were available (e.g., Annonaceae, Aristolochiaceae, Proteaceae) were removed to construct data sets 3 and 4.

The outgroups were five members of Gnetales: *Ephedra sinica*, *E. torreyana*, *Gnetum nodiflorum*, *G. gnemon*, and *G. urens*. Gnetales were the logical choice of outgroup because they appear as the extant sister to the angiosperms in most recent phylogenetic analyses (e.g., Crane, 1985, 1988; Doyle & Donoghue, 1986, 1992; Donoghue & Doyle, 1989a, b; Loconte & Stevenson, 1991; Chase et al., 1993; Doyle et al., 1994; Nixon et al., 1994). In addition, to ascertain the topological impact of other outgroups, particularly with regard to the first-branching angiosperms, we conducted several other searches. Using the smaller 18S data sets (194 an-

giosperms), both with and without indels, we used as outgroups: (1) the five Gnetales and *Zamia pumila*; (2) the five Gnetales, *Zamia pumila* and *Cycas revoluta*. Similarly, for the large data sets (223 angiosperms) we used as outgroups: (1) the five Gnetales and *Zamia pumila*; (2) the five Gnetales and a recently acquired sequence of *Welwitschia mirabilis*.

Because of the large number of taxa involved, we used two basic search strategies. The first method was a heuristic search performed using PAUP\* 4.0 (Swofford, pers. comm.) and to a lesser extent PAUP 3.1.1 (Swofford, 1993) with MULPARS, RANDOM taxon addition, and TBR branch-swapping. These searches were permitted to run for a week or more using either a Macintosh Quadra 650 or Sun Sparc Server 600P. These searches did not produce trees as short as those produced by the method below and will not be discussed further.

The primary search strategy was inspired by Maddison et al. (1992) and suggested by D. Swofford (pers. comm.). For each of the four data sets, we used 50–100 consecutive searches without MULPARS using RANDOM taxon addition and NNI branch-swapping. We then performed multiple searches (300–500 replicates; a Sun Sparc Server 600P typically required 19–25 hours to complete five replicates) using RANDOM addition, MULPARS, and TBR branch swapping, where only two trees (NCHUCK = 2) of a specified length (CHUCKLEN) or longer were saved per replicate. The length of the shortest trees obtained in the NNI searches was used as the initial CHUCKLEN value. As shorter trees were found, additional searches were conducted with lower CHUCKLEN values. This approach prevented the searches from being overwhelmed with trees.

The final portion of this search strategy involved use of the shortest trees obtained above as starting points for subsequent searches, again with MULPARS and TBR branch swapping. These searches were permitted to run for weeks or months using Macintosh Quadra/Centris 650 or PowerMac 6100 or 7100 computers. Typically no more than 2000–5000 trees were saved in any of these searches. We used starting trees of several different lengths when implementing this final portion of the search strategy to explore tree space from multiple perspectives and to prevent the analysis from stalling while swapping on suboptimal trees (P. Soltis & Soltis, 1997). For data set 1, 94 starting trees of lengths 3929, 3930, 3934, 3936, 3937, 3938, 3939, 3940, and 3941 were used (shortest tree ultimately obtained had a length of 3923 steps). For data set 2, 78 starting trees of lengths 3938, 3939, 3940,

3941, 3942, 3944, 3946, and 3947 were used (shortest tree ultimately obtained had a length of 3930 steps). For data set 3, 192 starting trees of lengths 3506, 3508, 3509, 3512, 3514, 3515, and 3517 were used (shortest tree ultimately obtained had a length of 3501 steps). For data set 4, 96 starting trees of lengths 3513, 3514, 3515, 3516, 3517, 3520, and 3521 (shortest tree ultimately obtained had a length of 3507 steps). Many of these searches resulted in trees one to several steps longer than the shortest trees ultimately obtained; these longer trees were also examined to help ascertain the general support for some branches. Implementing the above search strategy for the data sets described ultimately entailed well over two years of computer time.

Implementing decay analyses (Bremer, 1988; Donoghue et al., 1992) is impractical with data sets of this size. To obtain an estimate of support for the 18S rDNA topologies, we applied the parsimony jackknife approach (Farris et al., 1997) to data set 1 (this analysis was kindly conducted by S. Farris). The jackknife is a resampling approach, similar to the bootstrap, in which the characters of a data set are resampled to generate replicate data sets. Each replicate is analyzed, and the proportion of replicates supporting a given conclusion (in this case a clade) is considered a measure of support. Jackknife percentages can therefore be interpreted similarly to bootstrap percentages. In this analysis, 1000 replicates were conducted, and a minimum jackknife value of 50 (CUT = 50) was used (i.e., only clades supported by jackknife values of 50% or greater were retained). If a node is supported by one uncontradicted character, the jackknife value is 63% (Farris et al., 1997). Thus, clades with values of 63% or more are strongly supported; nodes with values of 51–62% are less well supported, and those with values of 50% or less receive no support.

## RESULTS AND DISCUSSION

### 1. THE EVOLUTION OF THE 18S rRNA GENE

The accumulation of a large data set of entire 18S rDNA sequences has permitted a more thorough assessment of the general evolution of the 18S rRNA gene. Unlike protein-coding genes, such as the widely sequenced *rbcL*, *matK*, and *ndhF*, there is no clear frame of reference for aligning sequences or revealing errors. For example, with protein-coding genes, translation of a sequence to amino acids will potentially reveal some errors such as frameshifts and internal stop codons. No such internal check is available, however, for rDNA. As a result, general features concerning the evolution of

18S rDNA have, in large part, been greatly misunderstood. In particular, insertion and deletion events have long been considered common in 18S rDNA; concomitantly, alignment was considered highly problematic. Until now, the existing database of angiosperm 18S rDNA sequences was insufficient to assess these views.

### INSERTION-DELETION AND ALIGNMENT

This study reveals that indels are not widespread in the 18S rDNA sequences of angiosperms, but instead are confined to a few, small regions of the gene. Furthermore, with the exception of these same small regions, alignment of 18S rDNA sequences is straightforward. Several possibilities may explain the misconception that the 18S rRNA gene is highly prone to insertion and deletion. First, the literature contains a number of erroneous 18S rDNA sequences. We have resequenced the 18S rDNA of over 20 taxa, and have found that some published sequences are in error by as many as 33 bases, which corresponds to 1.8% of the total gene. In several instances, we discovered numerous errors in the 18S rDNA sequence for a taxon using the same DNA originally used to produce the reported sequence. These errors in previously generated sequences are not confined to base composition, but also involve the presence of what we refer to here as “false” insertions and deletions. For example, we resequenced 18S rDNA from *Zea mays* and discovered that the original sequence included a large number of “false” insertions relative to all other angiosperms. Integrating our new sequence for *Zea* into the angiosperm data matrix and removal of the previously published sequence resulted in the elimination of 14 indels from our 18S rDNA data set, four of which were alignment gaps that had to be added to all other angiosperms. We were able to remove additional alignment gaps through the resequencing of other taxa for which 18S rDNA sequences were reported previously. As a point of comparison, the total length of the aligned sequences in the data matrix of Nickrent and Soltis (1995) for 64 taxa was 1853 bp. In contrast, the length of the aligned sequences in our 228-taxon data matrix is only 1850 bp. The resequencing of additional taxa for which published sequences cause numerous alignment gaps would likely decrease further the total length of the aligned sequences.

The numerous erroneous 18S rDNA sequences in the literature perhaps result from inherent difficulties in sequencing rDNA. Secondary structure in the rRNA for which this gene codes is also pres-



Table 3. Area initially thought to be prone to insertion and deletion; however apparent gaps result from sequencing difficulties. The underlined portion of the *Hydrangea* sequence shows the actual base composition that we have determined to be present for many taxa for this area.

|                      | 215<br>*                |
|----------------------|-------------------------|
| <i>Hydrangea</i>     | AAAGGTTGACGCGGGCTTTGCCC |
| <i>Glycine</i>       | AAAGGTCAACACAGGCTCTGCCT |
| <i>Heuchera</i>      | AAAGGCCAAC----GCTTTGCCC |
| <i>Lepuropetalon</i> | AAAGGTCAACGCTTGCTTCGGCT |
| <i>Prunus</i>        | AAAGGCCAAC----GCTCTGCCC |
| <i>Francoa</i>       | AAAGGTCGAC----GCTTTGCCC |
| <i>Podophyllum</i>   | AAAGGTCAACG---GCTTTGCCC |
| <i>Austrobaileya</i> | AAAGGCCGAT--CGGCTCTGCCC |
| <i>Caulophyllum</i>  | AAAGGTCAAC?????CTGCCC   |
| <i>Buxus</i>         | AAAGGTCGA-----TGCC-     |

ent in the gene itself and may lead to compressions and associated sequencing problems. More than one molecular systematist with substantial experience in the sequencing of chloroplast genes such as *rbcL* has referred to the sequencing of 18S rDNA as "tricky." We have found that preparation of samples via cycle sequencing followed by automated sequencing yields reliable 18S rDNA sequences that appear more accurate than most manually generated sequences. The critical procedural step is likely the cycle sequencing reactions, in which secondary structure is reduced or eliminated by high temperature. Several specific regions of 18S rDNA are particularly difficult to sequence. These include base positions 215–230, 1355–1365, and 1705–1715 (all positions mentioned in this paper correspond to those of *Glycine max*; Eckenrode et al., 1985), as well as several of those small regions noted earlier that are prone to insertion and deletion (positions 230–237; 496–501; 666–672; 1363–1369) (see Appendix).

We will use the first of these regions (base positions 215–230) to illustrate the errors that can result in 18S rDNA sequencing. Based on manual sequences (generated by D. Soltis & R. Kuzoff), the base composition of this region in Saxifragaceae and several other rosoid families initially appeared to involve a large deletion relative to some other available sequences (see Table 3). Similarly, the 18S rDNA sequences generated manually by other investigators, representing a diversity of angiosperms, typically were lacking one or more base pairs in this region. Alternatively, researchers scored this region as uncertain, using either "?" or "N." Thus, sequences available prior to this study suggested that this region was highly prone to in-

sertion and deletion. As a result, alignment of this region was initially difficult. Alignment problems were exacerbated by the apparent occurrence of base substitution in the region. Further compounding the difficulty of alignment is the fact that the region just 3' of this area actually is prone to insertion-deletion, as well as to considerable variation in primary sequence. Base positions 230–237 correspond to one of the variable helix termini noted above. Cycle and automated sequencing of over 100 taxa, however, revealed no indels in the area of positions 215–230. In fact, after resequencing this region in many taxa for which manual sequences initially suggested the presence of numerous indels, we have concluded that indels in this region are either extremely rare or nonexistent. This region is G-C rich; as a result, sequencing "stops" often occurred, yielding only a portion of the base pairs actually present in the region. Alignment of these incomplete sequences suggested numerous indels in this region, leading to the misconception that indels were frequent. Similar sequencing problems were encountered in other portions of the 18S rRNA gene. Taken together, these regions had contributed to the view that insertion and deletion are common in 18S rDNA.

Although the frequency of indels has been overstated for the 18S gene, several regions of 18S rDNA are, in fact, prone to variation in primary sequence and length. However, these regions are small, easily located, and, as noted by Nickrent and Soltis (1995), typically confined to the termini of helices on the proposed secondary structure model for 18S rRNA (e.g., Nickrent & Soltis, 1995). Four such regions, represented by base positions 230–237, 496–501, 666–672, and 1363–1369, correspond to the termini of helices E10-1, 17, E23-1, and 43, respectively (see Appendix). These regions are difficult to align over a broad taxonomic scale, such as all angiosperms, and were therefore not used in our phylogenetic analyses (see Materials and Methods above). On a lower taxonomic scale (e.g., closely related families), however, even these highly variable regions are easily aligned, permitting the use of these regions in more focused studies (e.g., Polemoniaceae and related Asteridae s.l., Johnson et al., unpublished; portions of Saxifragaceae s.l., D. Soltis & Soltis, 1997; Orchidaceae, Cameron & Chase, unpublished).

Because indels in 18S rDNA are neither as prevalent nor as problematic as previously thought, alignment of clean 18S rDNA sequences is straightforward. With the exclusion of the few small regions noted above, alignment of over 200 angiosperm sequences was straightforward and easily accom-

plished by eye. This is also true for the alignment of 18S rDNA sequences on a broader scale across vascular plants (P. Soltis et al., unpublished). As noted above, the resequencing of some taxa greatly facilitated the alignment process in angiosperms.

#### PHYLOGENETICALLY INFORMATIVE INDELS

Not only do indels in angiosperm 18S rDNA sequences not cause alignment problems, but some indels may be phylogenetically informative at the level of our investigation. Here we do not consider those indels located within the variable regions noted above, but only those indels located in highly conserved regions not normally prone to changes in length. Two such indels in particular (Table 2) appear to be phylogenetically informative across the angiosperms.

One indel involves an apparent deletion of one base pair that unites all higher eudicots (Table 2, indel A). This base pair is present in Gnetales, monocots, paleoherbs, Magnoliales, Laurales, ranunculids, Trochodendraceae, Tetracentraceae, Proteaceae, Sabiaceae, Platanaceae, and Nelumbonaceae and is absent from all higher dicots (i.e., the large Rosidae clade and Asteridae s.l.). Thus, the distribution of this indel agrees with the results of phylogenetic analyses based only on base substitutions (Figs. 1, 2, 4; all figures, plus Appendix, follow Lit. Cit.). In addition, the distribution of this indel also agrees with topologies based on analyses of *rbcL* sequences. It appears, however, that this base pair may also have been lost independently in two of the monocots analyzed here (i.e., *Calla* and *Chlorophytum*).

The second phylogenetically informative indel involves an apparent insertion (Table 2, indel B) that unites all members of the saxifragoid clade (Saxifragaceae s. str. and close allies; this is the Saxifragales of D. Soltis & Soltis, 1997). The saxifragoid clade also is united by base substitutions and represents one of the most strongly supported clades resulting from the phylogenetic analyses.

Additional examples of potentially informative indels can be found at lower taxonomic levels. For example, in Zingiberales, two insertions, each of a single base pair, are found at positions 117 and 260 in all members of the Zingiberaceae (Kress et al., 1995). None of the other approximately 70 monocots for which 18S rDNA has been sequenced exhibits these insertions. Similarly, an insertion of one base at position 655 appears to unite members of Viscaceae.

Undoubtedly, additional 18S rDNA sequencing will reveal more examples of phylogenetically in-

formative indels. However, the discovery of such indels depends on the availability of a large database of accurate sequences. Previously published sequences containing errors and ambiguities, "false" indels, and incompletely sequenced regions have made assessment of the phylogenetic potential of indels in 18S rDNA impossible.

#### STEM VERSUS LOOP CHANGES/COMPENSATORY CHANGES

The secondary structure of the 18S rRNA transcript may have significant implications for phylogeny reconstruction using rRNA or rDNA sequences (e.g., Mishler et al., 1988; Dixon & Hillis, 1993). The question remains as to whether both loop bases (non-pairing bases) and stem bases (pairing bases) should be used in phylogeny reconstruction and, if so, whether bases from stems and loops should be considered equally informative and independent. Assuming near-perfect compensatory mutation (substitutions that maintain or restore stem base complementarity—e.g., Noller, 1984; Curtiss & Vournakis, 1984; Wheeler & Honeycutt, 1988) in stem regions to maintain secondary structure of the 18S or 26S (28S) rRNA, Wheeler and Honeycutt (1988) recommended that stem bases either be eliminated from phylogenetic analysis or weighted by one-half relative to loop bases. However, in their analyses of 28S rRNA sequences from vertebrates, Dixon and Hillis (1993) found that characters from both stems and loops contain phylogenetic information. They also found that the number of compensatory mutations in stem bases was less than 40% of that expected under a hypothesis of perfect compensation to maintain secondary structure. Dixon and Hillis therefore suggested that the weighting of stem characters be reduced by no more than 20% relative to loop characters in phylogenetic analyses.

The large database of 18S rDNA sequences reported here affords the opportunity to address these and other issues regarding the impact of the secondary structure of the 18S rRNA transcript on phylogeny reconstruction in angiosperms. Although it is not our goal to examine such issues in detail here, we will provide some initial observations regarding the relative importance of both stem and loop mutations and the prevalence of compensatory mutations.

We followed the definitions of stem and loop bases used elsewhere (e.g., Dixon & Hillis, 1993): stem bases are those that participate in base-pairing interactions; loop bases do not engage in base pairing in the mature rRNA. Mapping base substitutions on the proposed 18S rRNA secondary struc-

ture for *Glycine max* (see Appendix), we examined 120 positions (in 60 taxa) at which phylogenetically informative base substitutions had occurred (based on the results of the searches; discussed below). Emphasis was placed on those substitutions that provided synapomorphies for those clades that appear in all shortest trees and that represent several different taxonomic levels (e.g., Asteridae s.l., Caryophyllidae s.l., monocots, glucosinolates, santaloids, Caryophyllales, saxifragoids, celastroids, *Parnassia-Lepuropetalon*). Of these 120 positions, 70 (58%) were stem bases, and 50 (42%) were loop bases. Although this initial survey considers only a subset of synapomorphous base substitutions, it suggests that both stem and loop regions appear to contain phylogenetic information, with perhaps a somewhat greater proportion of informative sites found in stem, rather than loop, regions. This topic certainly requires a more rigorous examination. The relative information content of stem versus loop bases may, in fact, vary at different taxonomic levels. For example, some of the more variable loop regions (several of which were removed from these phylogenetic analyses because the sequences were difficult to align) may hold relatively more information at lower levels (among and within closely related families) than at higher taxonomic levels (ordinal and above), at which the sequences become too divergent for confident alignment.

The frequency of compensatory changes was examined in 21 stem regions located throughout the 18S rRNA gene. Following the general approach of others (e.g., Dixon & Hillis, 1993), we considered two classes of substitutions within stem regions. The first class involves substitutions that change one pair of complementary bases to another pair of complementary bases. This includes "double compensatory" substitutions in which one pair of complementary bases is converted to another (e.g., C-G to A-U). This class also includes changes that require only a single substitution event. That is, because uracil can pair with guanine as well as with adenine, it is possible to have a single change from one base-pairing couplet to another (e.g., U-G to C-G; U-A to U-G). The latter represent one type of "single compensatory" substitutions (sensu Dixon & Hillis, 1993). The second class of stem substitutions involves those that change one pair of complementary bases to a pair of noncomplementary bases, or vice versa (e.g., C-G to G-G; or C-C to C-G). For example, a change of C-G to G-G destroys a base-pairing couplet. Conversely, a change from C-C to C-G creates a base-pairing couplet and represents another example of a "single compensatory" substitution. Of the 216 stem changes we

analyzed, 19% were "double compensatory"; 46% were single base substitutions involving uracil that changed one base-pairing couplet to another ("single compensatory"); 8% changed a pair of noncomplementary bases to complementary (the second type of "single compensatory" change); 27% destroyed a base-pairing couplet. Of these changes that result in mispairing of nucleotides, over one-third are adjacent to loop regions. Hence, the loop regions may simply be expanded in these instances. Nearly three quarters (73%) of the stem mutations we analyzed maintain or restore base pairing and would be considered compensatory. In their comparable analysis of 28 rRNA sequences, Dixon and Hillis (1993) observed that only 47% of the mutations maintained or restored base pairing. Our results are more similar to observations for 5S rRNA (Curtiss & Vourmakis, 1984), where approximately 88% of the base substitutions analyzed from stem regions were compensatory.

A similar pattern of molecular evolution is seen for 18S rDNA within a single family, Polemoniaceae, and its closest relatives (Johnson et al., unpublished), where 228 variable nucleotide positions were examined. Although most of these 228 positions are located on stems (133 compared to 95 loop characters), the average number of substitutions per site over the potentially informative characters is greater for the loop characters (5.0) than for the stem characters (3.1). Using one of the most parsimonious trees as a framework, Johnson et al. (unpublished) also considered in more detail 67 substitutions that either unite or appear within Polemoniaceae. Of these substitutions, 36 (53.7%) occur in loops. Considering just the 31 stem substitutions, 23 (74.2%) either maintain or restore base-pairing. The remaining eight stem substitutions (25.8%) result in mispairing of nucleotides, with four sites located adjacent to loops.

These initial studies of the relative information content of stem and loop regions and the frequency of compensatory changes have implications for the use and relative weighting of stem and loop bases in phylogeny reconstruction. These data reinforce the findings of others (e.g., Dixon & Hillis, 1993; Smith, 1989) that both stem and loop regions provide important information for phylogeny reconstruction. In addition, the high frequency of observed compensatory change in angiosperm 18S rDNA suggests that perhaps stem characters should receive less weight than loop characters in future analyses. However, weighting of stem versus loop characters is more complex than it might seem initially. Recent work with 18S rDNA sequences in Polemoniaceae (Johnson et al., unpublished) dem-

onstrates that loop regions evolve more rapidly than do stem regions. Thus, in more focused studies in which it is possible to align and use the entire 18S rDNA sequence, stem and loop regions should perhaps be given equal weight. In broader studies in which the rapidly evolving loops are removed due to alignment difficulties and only the more conserved loop regions are included in the analysis, stems should be downweighted; however, more detailed analyses are required to estimate appropriate weights.

## II. PHYLOGENETIC RELATIONSHIPS

Each broad phylogenetic analysis yielded thousands of most parsimonious trees; it is likely that shorter trees exist for all four data sets and that additional classes of most-parsimonious trees were not recovered. Nonetheless, we feel that it is significant that analyses of three of the four data sets suggest the same general topology. The shortest trees obtained from searches of data sets 1 and 2 are essentially identical, and differences between the shortest trees from analysis of these two data sets and data set 4 are minor and weakly supported. Although phylogenetic analysis of data set 3 revealed many of the same major clades recovered by searches of the other data sets, relationships among some of these clades differ; most notable are the weakly supported, unusual positions of monocots and saxifragoids (see below).

All searches revealed the same major clades (e.g., Rosidae, Asteridae s.l., Caryophyllidae s.l., monocots, saxifragoids), as well as the same suite of taxa as sister to all remaining angiosperms. In general, the trees obtained in these exploratory, broad analyses of 18S rDNA sequences depict relationships very similar to those obtained in broad analyses of *rbcL* sequences (Chase et al., 1993). The general features observed in the shortest trees obtained from the four searches are discussed below. The several unusual relationships among major clades suggested by analyses of data set 3 are discussed in more detail below under "Differences Among the Shortest Trees."

### FIRST-BRANCHING FAMILIES

Phylogenetic analyses of three of four data sets (data sets 1, 2, and 4) suggest that those taxa having uniaperturate pollen (monosulcate and monosulcate-derived) and ethereal oils appear as early-branching angiosperms, forming a grade (labeled monosulcate grade in Figs. 1, 2, and 4). Those plants having triaperturate pollen (tricolpate and tricolpate-derived), with a few exceptions (see be-

low), and tannins and alkaloids as secondary compounds form a clade (labeled eudicot clade). The latter group has been referred to as the eudicots (Donoghue & Doyle, 1989b; Doyle & Hotton, 1991; Chase et al., 1993). Although the term eudicot has been variously defined, we will use the Chase et al. (1993) definition to facilitate comparison between the two studies. A eudicot clade was also recovered in analyses of *rbcL* sequences (Chase et al., 1993), but, instead of forming a grade, as they do here, those plants with uniaperturate pollen form a weakly supported clade in the *rbcL* trees.

There are two major exceptions to the general correspondence between the eudicot clade and the distribution of the triaperturate pollen types (other than the obvious departures observed in the trees derived from analysis of data set 3). First, the Winteraceae and several families of paleoherebs (Chloranthaceae, Lactoridaceae, Aristolochiaceae) all possess uniaperturate pollen, yet appear within the eudicot (triaperturate) clade in the shortest trees obtained in analyses of data sets 1 and 2 (Figs. 1, 2). These exceptions may reflect low taxon density and/or the low resolving power of 18S rDNA sequence data (see below); these taxa seem to be unstable in position in the various searches. In broad analyses of *rbcL* sequences, in contrast, these taxa are clearly members of the uniaperturate clade.

The second exception involves Illiciaceae and Schisandraceae. Unlike the examples above, however, which we suspect represent spurious phylogenetic placements, Illiciaceae and Schisandraceae appear to be true early-branching angiosperms (see below), yet possess triaperturate pollen. These families similarly appear as early-branching angiosperms in analyses based on *rbcL* sequences (Chase et al., 1993; Qiu et al., 1993). As reviewed by Doyle et al. (1990), however, the tricolpate condition in Illiciaceae and Schisandraceae is different from that which characterizes eudicots. Hence, the 18S rDNA analyses further support the *rbcL*-based inferences of Qiu et al. (1993) that Illiciaceae and Schisandraceae represent an independent evolution of tricolpate pollen.

Four families of woody Magnoliidae consistently appear as sister taxa to all remaining angiosperms analyzed: Amborellaceae and a clade of Austrobaileyaceae, Illiciaceae, and Schisandraceae. The latter three families form one of the most strongly supported clades in this study (jackknife value of 94%). In searches of data sets 1 and 2, a clade of Austrobaileyaceae, Illiciaceae, and Schisandraceae is the sister group to all other angiosperms, followed subsequently by Amborellaceae; in analyses of data sets 3 and 4, the positions of these two

lineages are reversed. Surprisingly, given the close relationship suggested between Illiciaceae and Schisandraceae by others (e.g., Cronquist, 1981; Qiu et al., 1993), *Illicium* is sister to *Austrobaileya-Schisandra* in all four analyses. These four genera, with Nymphaeales, form a clade in the *rbcL* analyses of Chase et al. (1993) and Qiu et al. (1993).

In the shortest trees resulting from searches of all four data sets, from one to several families of paleoherbs (sensu Donoghue & Doyle, 1989a) subsequently follow Austrobaileyaceae, Illiciaceae, Schisandraceae, and Amborellaceae. Nymphaeaceae immediately follow these four families in all shortest trees. In searches of data sets 1 and 2, Nymphaeaceae form a clade with Piperaceae and Saururaceae (represented by *Peperomia* and *Houttuynia* and *Saururus*, respectively; jackknife value of 85%), whereas in searches of data set 4, these same three families form a grade with Nymphaeaceae as sister to all remaining angiosperms, followed by a clade of Piperaceae and Saururaceae. In searches of data set 3, Nymphaeaceae also follow Austrobaileyaceae, Illiciaceae, Schisandraceae, and Amborellaceae, but Nymphaeaceae are then followed by saxifragoids, an unusual placement discussed in more detail below.

Amborellaceae, followed by (1) a clade of Austrobaileyaceae, Illiciaceae, and Schisandraceae, (2) Nymphaeaceae, (3) a clade or grade of Piperaceae, Saururaceae, Aristolochiaceae, and Lactoridaceae (similar to Fig. 4), appear as the first-branching angiosperms when *Zamia* and *Cycas* are used as additional outgroups (see Materials and Methods). In preliminary analyses of a larger data set of 271 angiosperms using species of *Welwitschia*, *Gnetum*, and *Ephedra* as outgroups, Amborellaceae, a clade of Austrobaileyaceae–Illiciaceae–Schisandraceae, and Nymphaeaceae again appear as successive sisters to all remaining angiosperms.

The position of woody magnoliids as first-branching taxa in these 18S rDNA trees is in general agreement with traditional views of angiosperm relationships (e.g., Cronquist, 1968, 1981; Stebbins, 1974; Takhtajan, 1969, 1980) that suggest that woody Magnoliidae are the most primitive extant angiosperms. The morphological analyses of Donoghue and Doyle (1989a) and Loconte and Stevenson (1991) also support the woody Magnoliidae as the most ancestral living group of angiosperms. Other data also point to the antiquity of at least some of these genera. For example, Endress and Honegger (1980) determined that the pollen of *Austrobaileya* resembles *Clavatipollenites*, one of the oldest probable angiosperm fossils, and concluded that *Austrobaileya* may be “especially archaic

among the angiosperms.” If the 18S rDNA inference is correct in suggesting that Amborellaceae, a family lacking vessel elements, are among the first-branching angiosperms, this analysis may support the hypothesis that the angiosperms were primitively vesselless (Bailey, 1957; Cronquist, 1981; Young, 1981).

In contrast to this study, analyses based on partial 18S and 26S rRNA sequences suggested that a group of paleoherbs (Aristolochiales, Piperales, Nymphaeales) is the sister taxon to all other flowering plants (Hamby & Zimmer, 1992). However, of the four woody families of Magnoliidae appearing as first-branching taxa in our 18S rDNA trees (Amborellaceae, Schisandraceae, Illiciaceae, and Austrobaileyaceae), only Illiciaceae were sampled by Hamby and Zimmer (1992). Other phylogenetic analyses similarly support the position of some paleoherbs as first-branching taxa among the angiosperms (e.g., Doyle et al., 1994; Nixon et al., 1994). Paleoherbs are sister to other angiosperms in trees based on a combination of morphology and rRNA sequence data and in those derived independently from morphological and rRNA data (Doyle et al., 1994). However, this topology is only weakly supported by morphological data, with trees rooted next to Magnoliales only one step longer. Furthermore, the rRNA data set employed by Doyle et al. (1994) is that of Hamby and Zimmer (1992), which, as noted above, lacked several critical woody magnoliids.

Broad phylogenetic analyses of *rbcL* sequences (Chase et al., 1993; Rice et al., 1997) place the aquatic genus *Ceratophyllum* as sister to all remaining angiosperms. This placement of *Ceratophyllum* also has been suggested on morphological grounds (Les, 1988; Les et al., 1991; Nixon et al., 1994), although alternative trees in the latter study place the paleoherb family Chloranthaceae as sister to the remaining flowering plants. A number of floral features of *Ceratophyllum* also conform to the view that the genus represents a primitive angiosperm (Endress, 1994). However, *Ceratophyllum* does not appear as first-branching in any of our phylogenetic analyses. Searches involving three of the four data sets (1, 2, and 4) place *Ceratophyllum* as sister to the monocots, a finding in general agreement with earlier rRNA sequence analyses (Hamby & Zimmer, 1992).

Subsequent to the Amborellaceae, Austrobaileyaceae, Illiciaceae, Schisandraceae, Nymphaeaceae, and Piperales, in analyses of three of four data sets (1, 2, and 4) are additional families and orders of Magnoliidae: Annonaceae, Calycanthaceae, and

Lauraceae, all woody families traditionally considered among the most primitive extant angiosperms.

With the exception of the shortest trees resulting from analysis of data set 3, the monocots also appear as an early lineage of angiosperms. The monocots are monophyletic, with the exception of *Acorus*, which does not appear closely related to the other member of Araceae included (*Calla*). In analyses of the two data sets (1 and 2) that included *Acorus*, *Acorus* follows Nymphaeaceae–Piperales as the subsequent sister to all remaining angiosperms. In analyses of *rbcL* sequences, *Acorus* was considered “phylogenetically isolated” as sister to the remaining monocots (Duvall et al., 1993). Phylogenetic results based on 18S rDNA sequences also suggest that *Acorus* is anomalous among monocots. Given its long branch length and unexpected position, the 18S rDNA of *Acorus* should be resequenced, and additional monocots should be added to the data set before the affinities of this enigmatic genus are addressed further.

Because our sampling of monocots was limited, to permit more thorough treatment elsewhere, relationships within the monocots will not be discussed here in any detail. Nonetheless, several traditionally recognized groups of monocots appear to be monophyletic, including Zingiberales, Liliales, and higher commelinoids. Furthermore, the bromeliads are grouped with the grasses and allies, as expected (Duvall et al., 1993). The two best supported clades within the monocots are Zingiberales (*Maranta*, *Zingiber*, *Costus*, *Canna*, *Heliconia*, and *Musa*; jackknife value of 58%) and a clade composed of Sparganiaceae, Cyperaceae, Poaceae, and Bromeliaceae (*Sparganium*, *Cyperus*, *Oryza*, *Zea*, and *Glomeropitcairnia*; jackknife value of 59%). Surprising results, given *rbcL* topologies (Duvall et al., 1993) and morphological features, include the placement of Orchidaceae (*Oncidium*) as the sister, or one of the sisters, to the remaining monocots and the placement of Arecaceae (*Veitchia*) and Alismataceae (*Sagittaria*) within the Asparagales (Figs. 1, 2). These unusual placements should not be considered seriously, however, due to the low representation of the monocots.

#### EUDICOT CLADE

Analyses of three of four data sets (1, 2, and 4) clearly reveal a eudicot (or triaperturate) clade (Figs. 1, 2, 4), with the following successive sister groups at its base (Figs. 1, 2): Proteaceae, Nelumbonaceae, Platanaceae, a clade of ranunculids, Trochodendraceae/Tetracentraceae, and a clade composed of Winteraceae (*Drimys*), Aristolochi-

aceae (*Aristolochia*, *Asarum*, *Saruma*), Lactoridaceae, Sabiaceae, and Chloranthaceae (*Hedyosmum*). The latter clade is an unexpected grouping (see below) of paleoherbs (Aristolochiaceae, Lactoridaceae, Chloranthaceae), woody Magnoliales (Winteraceae), and eudicots (Sabiaceae). With the exception of Aristolochiaceae, Lactoridaceae, Chloranthaceae, and Winteraceae, the presence of the remaining taxa on branches at the base of the eudicots closely parallels results retrieved from the phylogenetic analyses of *rbcL* sequences (Chase et al., 1993).

In the shortest trees obtained in analyses of data set 4, the distinction between the monosulcate grade and the eudicot clade is less clear than in the shortest trees obtained from the analyses of data sets 1 and 2 (see “Lower Eudicots/Monosulcates,” Fig. 4). Platanaceae, Trochodendraceae/Tetracentraceae, ranunculids (which are paraphyletic), Proteaceae (*Knightsia*), Buxaceae, Sabiaceae, and a clade of Chloranthaceae (*Hedyosmum*)/Winteraceae (*Drimys*) again appear as sister groups to the remainder of the eudicot clade. Also in this same lower eudicot/monosulcate grade, however, are Calycanthaceae, Annonaceae (*Mkuluu*), and Lauraceae (*Sassafras*), uniaiperturate families that appear in a clade with some ranunculids (Fig. 4).

In the shortest trees obtained in analyses of all data sets, the remainder of the eudicot clade is essentially composed of two large subclades, one consisting largely of Rosidae plus some Dilleniidae and the other corresponding to the Asteridae s.l. (labeled Rosidae and Asteridae s.l., respectively, in Figs. 1, 2, 4). With a few exceptions, most notably the placement of the monocots within the Rosidae clade, these two large clades also are present in the trees derived from searches of data set 3. The Rosidae and Asteridae s.l. clades were also recovered in broad analyses of *rbcL* sequences (Chase et al., 1993; Olmstead et al., 1992, 1993; Rice et al., submitted), although the placement of Caryophyllidae s.l. is very different in the 18S rDNA and *rbcL* topologies (see below). These two large clades, Rosidae and Asteridae s.l., reflect the basic division of higher dicots into two major groups (Young & Watson, 1970), with (1) polypetalous corollas and non-tenuinucellate ovules and (2) sympetalous corollas and tenuinucellate ovules, respectively. Below we discuss in more detail the major clades of eudicots based on phylogenetic analyses of 18S rDNA sequences. We use informal names in most instances to refer to strongly supported clades (e.g., celastroids, saxifragoids, ranunculids), some of which differ dramatically from traditional views of rela-

tionship, but formal taxonomic change may be warranted for many of these.

**Ranunculids.** The searches of the larger data sets (Figs. 1, 2) recovered a clade (labeled "Ranunculids") containing Lardizabalaceae, Berberidaceae, Ranunculaceae, Menispermaceae, Eupteleaceae, Fumariaceae, Sargentodoxaceae, and Papaveraceae. This same clade was found in the broad analyses of *rbcL* sequences (Chase et al., 1993); it represents the core of the Ranunculales (sensu Cronquist, 1981) and corresponds closely to the Berberidales of Thorne (1992) and the Ranunculiflorae of Dahlgren (1980). Analyses of not only *rbcL* and 18S rDNA sequences, but also *atpB* sequences, place Eupteleaceae (Hamamelidae) within this clade (Hoot & Crane, 1995). Also part of this clade in the 18S rDNA analyses is Sargentodoxaceae, a family typically placed in Ranunculales and allied with Lardizabalaceae (e.g., Cronquist, 1981). In contrast, analyses of *rbcL* sequences placed Sargentodoxaceae with Fabaceae (Chase et al., 1993). This result is due to the misidentification of leaf material in the *rbcL* analysis (Qiu, pers. comm.). Reanalysis of *Sargentodoxa* for *rbcL* places it as sister to the Lardizabalaceae (Hoot & Crane, 1995; Hoot et al., 1995).

Searches involving the two small data sets (3 and 4) employed fewer representatives of Ranunculales. In trees resulting from searches of data set 4, these taxa form a grade as some of the early-branching eudicots. In trees from searches of data set 3, in contrast, the ranunculids appear polyphyletic. The placements of the ranunculids in analyses of data sets 3 and 4 may well reflect their decreased representation (lower taxon density) in these searches. In preliminary analyses of a 271-taxon 18S rDNA data set including more ranunculids, the ranunculids again form a monophyletic group.

**Saxifragoids.** All analyses of 18S rDNA sequences (Figs. 1–4) reveal a clade composed of *Heuchera*, *Boykinia*, *Saxifraga* (Saxifragaceae s. str.), *Crassula*, *Sedum*, *Dudleya*, and *Kalanchoe* (Crassulaceae), *Pterostemon*, *Tetracarpaea*, *Ribes*, and *Itea* (Grossulariaceae), *Penthorum* (placed in Saxifragaceae by Cronquist, 1981), *Altingia* and *Liquidambar* (Hamamelidaceae), Haloragaceae, Cercidiphyllaceae, Daphniphyllaceae, and Paeoniaceae; this clade is referred to here as saxifragoids. This clade is one of the most strongly supported findings of this investigation (jackknife value of 68%). The same saxifragoid clade (also referred to as Saxifragales) was identified in an analysis of 130 18S rDNA sequences aimed at elucidating the affinities of the morphologically diverse members of

Saxifragaceae s.l. (D. Soltis & Soltis, 1997). The monophyly of this clade is supported not only by base substitutions, but also by the presence of an insertion (see Table 2) located in a portion of the 18S rRNA gene that is highly conserved in length. An identical clade (referred to as rosoid III) is revealed in the 499-taxon analysis of *rbcL* sequences (Chase et al., 1993) and is also retrieved in preliminary analyses of a 271-taxon 18S rDNA data set including more Hamamelidaceae, as well as in phylogenetic analyses involving *matK* sequences (Hibsch-Jetter & Soltis, unpublished) and preliminary work with *atpB* sequences (Hoot, unpublished). As reviewed in more detail by D. Soltis and Soltis (1997), this small clade is noteworthy in that it contains taxa traditionally placed in three subclasses: Paeoniaceae (Dilleniidae); Hamamelidaceae, Daphniphyllaceae, Cercidiphyllaceae (Hamamelidae); the remaining taxa are all members of Rosidae.

Although this saxifragoid clade is recovered by analyses of both 18S rDNA and *rbcL* sequences, this same group of taxa has never been recognized in any classification. Whereas Saxifragaceae s. str., *Ribes*, *Itea*, *Tetracarpaea*, *Pterostemon*, *Penthorum*, and Crassulaceae are considered closely related members of Rosidae in virtually all recent treatments (e.g., Cronquist, 1981; Thorne, 1992; Takhtajan, 1987; Dahlgren, 1980, 1983), the affinities of the rosoid family Haloragaceae and the dilleniid family Paeoniaceae have been considered enigmatic (e.g., Cronquist, 1981). The hamamelid families found in this clade (Hamamelidaceae, Cercidiphyllaceae, and Daphniphyllaceae) typically have not been considered close relatives of Saxifragaceae s. str. and allied rosids. The relationships of these more anomalous members of this clade are discussed in more detail by D. Soltis and Soltis (1997).

**Glucosinolate clade.** Another clade revealed by all analyses (Figs. 1–4) comprises glucosinolate-producing taxa. The families that compose this clade in Figures 1–4 are 7 of the 15 families known to produce glucosinolates (mustard oil glucosides): Limnanthaceae, Brassicaceae, Capparaceae, Moringaceae, Caricaceae, Bataceae, and Tropaeolaceae. Whereas Brassicaceae and Capparaceae have long been recognized as closely related, the remaining families included in this study (Limnanthaceae, Moringaceae, Caricaceae, Bataceae, and Tropaeolaceae) are morphologically diverse and have been placed in distinct orders (e.g., Cronquist, 1981; see review by Rodman et al., 1993). The genus *Drypetes* (Euphorbiaceae) also produces glucosinolates, but it does not appear to be closely

related to the glucosinolate clade in any of the four searches. Phylogenetic analyses of 18S rDNA sequences involving additional glucosinolate taxa further demonstrate the monophyly of the glucosinolate-producers, with the exception of *Drypetes*, and also clarify relationships among the members of this clade (Rodman et al., submitted). These results closely parallel findings based on the phylogenetic analysis of *rbcL* sequences (Rodman et al., 1993; Chase et al., 1993) and morphology (Rodman, 1991). Thus, both *rbcL* and 18S rDNA sequence data indicate that there were two independent origins of the mustard oil syndrome (see Rodman et al., 1993, submitted).

**Nitrogen-fixing clade.** Species of only 10 families of angiosperms are known to form symbiotic associations with nitrogen-fixing bacteria in root nodules (Fabaceae, Betulaceae, Casuarinaceae, Coriariaceae, Datisceae, Elaeagnaceae, Myricaceae, Rhamnaceae, Rosaceae, and Ulmaceae). These families are distributed among four of Cronquist's (1981) six subclasses of dicotyledons, implying that many of these families are only distantly related. Recent phylogenetic analyses of *rbcL* sequences reveal, however, that representatives of all ten of these families occur together in a single clade ("nitrogen-fixing clade"; Soltis et al., 1995). In addition to these ten families, this clade also contains several families not known to form associations with nitrogen-fixing bacteria, including Moraceae, Cannabaceae, Urticaceae, Polygalaceae, Fagaceae, Begoniaceae, and Cucurbitaceae.

Analyses of three of four 18S rDNA data sets (Figs. 1, 2, 4) suggest an alliance of taxa similar to that revealed by *rbcL* sequences. This clade in large part represents a subset of the taxa present in the *rbcL*-based nitrogen-fixing clade. The families in the 18S rDNA-based nitrogen-fixing clade include Betulaceae, Casuarinaceae, Datisceae, Elaeagnaceae, Rhamnaceae, and Ulmaceae, all families that form symbiotic associations with nitrogen-fixing bacteria. Other families known to form such associations (i.e., Coriariaceae and Myricaceae) and that appeared in the *rbcL*-based nitrogen-fixing clade were not analyzed for 18S rDNA sequence variation. Also part of the nitrogen-fixing clade retrieved here are Begoniaceae, Moraceae, Urticaceae, and Cucurbitaceae, families also found to be part of this alliance based on analyses of *rbcL* sequences. However, neither Rosaceae nor Fabaceae, two families involved in nitrogen-fixing symbioses, are included within the 18S rDNA nitrogen-fixing clade, although both families are part of this alliance in the *rbcL*-based trees (Soltis et al., 1995).

Searches involving the two larger data sets (1 and 2) also place three families of Malvales (Malvaceae, Bombacaceae, and Tiliaceae) within the nitrogen-fixing clade; these taxa were not part of the nitrogen-fixing clade in the *rbcL*-based trees. In analyses of data set 4, however, these three families of Malvales are not part of the nitrogen-fixing clade (Fig. 4). No clear nitrogen-fixing clade emerged in analyses of data set 3; instead, these taxa are part of a grade that represents the first branches of a primarily rosoid-dillenioid clade (Fig. 3).

**Asteridae sensu lato.** Analyses of all four 18S rDNA data sets also reveal an expanded Asteridae clade (Asteridae s.l.) that agrees closely with that recovered by analyses of *rbcL* sequences (Olmstead et al., 1992, 1993; Chase et al., 1993). In addition to the conventionally circumscribed Asteridae, this clade also includes a number of families placed in Dilleniidae, such as Ericaceae, Clethraceae, Pyrolaceae, Styracaceae, Ebenaceae, Actinidiaceae, Sarracenaceae, Fouquieriaceae, Theaceae, and Primulaceae. Also present in Asteridae s.l. are Nysaceae, Pittosporaceae, Apiaceae, Araliaceae, and Hydrangeaceae, all members of Rosidae. In addition, Eucommiaceae, a member of Hamamelidae, and *Byblis*, a genus of carnivorous plants usually placed in Rosidae, also appear within Asteridae s.l. All analyses also place an expanded Caryophyllidae (Caryophyllidae s.l.) within the Asteridae s.l. clade, an unexpected result that is discussed in more detail below.

Within Asteridae s.l., several subclades or grades can be identified that agree, in large part, with some of the groups identified in analyses of *rbcL* sequences (Chase et al., 1993; Olmstead et al., 1993). Perhaps most noteworthy of these is the ericalean grade (the asterid III clade of Chase et al., 1993) observed in all of the shortest 18S rDNA trees (Figs. 1–4). Other clades of Olmstead et al. are also observed to be monophyletic, including Dipsacales, Boraginales, Gentianales, Asterales s.l., and Lamiidae. Additional asterid taxa should be sequenced for 18S rDNA to assess more rigorously the monophyly of these groups and their interrelationships.

**Caryophyllidae sensu lato.** All analyses of 18S rDNA sequences reveal a clade composed of Nysagaceae (*Mirabilis*), Chenopodiaceae (*Spinacia*), Phytolaccaceae (*Phytolacca*), Aizoaceae (*Tetragonia*), and Molluginaceae (*Mollugo*). These five families represent Caryophyllales (e.g., Cronquist, 1981), the monophyly of which is supported in this study by a jackknife value of 58%, as well as by numerous lines of morphological and molecular



data (e.g., Rodman et al., 1984; Rettig et al., 1992). Sister to this clade of Caryophyllales is another strongly supported clade comprising Plumbaginaceae and Polygonaceae (jackknife value of 77%); this group collectively represents Caryophyllidae (sensu Cronquist, 1981). The monophyly of Caryophyllidae is only weakly supported by cladistic analysis of morphological, chemical, anatomical, and palynological features (Rodman et al., 1984). Analyses of 18S rDNA sequences also suggest that two families of carnivorous plants, Droseraceae and Nepenthaceae, are sister to Caryophyllidae, and we refer to this entire assemblage as Caryophyllidae s.l. (Figs. 1–4).

Phylogenetic analyses of *rbcL* sequences similarly recovered a Caryophyllidae s.l. clade composed of Caryophyllales, Polygonaceae, Plumbaginaceae, Droseraceae, and Nepenthaceae (Chase et al., 1993). One of the broad analyses of *rbcL* sequences (search A, Chase et al., 1993) placed Vitaceae and Dilleniaceae with this expanded Caryophyllidae clade. In the analyses of 18S rDNA sequences, Vitaceae were not sampled, and Dilleniaceae are well removed from Caryophyllidae s.l. The anomalous placement of Dilleniaceae near the monocots (Figs. 1, 2) is discussed below.

**Santaloids.** Analyses of all four data sets reveal a monophyletic santaloid clade or Santalales, which are represented here by only three families (Opiliaceae, Santalaceae, and Viscaceae). However, in preliminary analyses in which Santalales are represented by seven families (Opiliaceae, Santalaceae, Viscaceae, Eremolepidaceae, Misodendraceae, Loranthaceae, and Olacaceae), santaloids again form a clade. These seven families are widely considered to form a natural group based on morphology (e.g., Cronquist, 1981) and have been shown to form a clade in previous, smaller analyses of 18S rDNA sequences (Nickrent & Franchina, 1990; Nickrent & Soltis, 1995).

Although santaloids appear monophyletic, the position of this clade varies among the analyses. In analyses of data sets 1 and 2, santaloids are sister to *Polygala* and closely related to the legumes. Analysis of data set 4 again places santaloids with *Polygala* and a legume (*Pisum*), as well as with *Gunnera*. Analysis of data set 3 results in an unusual placement of santaloids with several paleoherbs. These findings parallel those of Chase et al. (1993) based on *rbcL* sequences in which the position of santaloids differed greatly between the 476- and 499-taxon searches. In the former, santaloids and *Gunnera* form the asterid V clade; in the latter, santaloids are sister to the caryophyllids,

but again appear near *Gunnera*. Thus, whereas both *rbcL* and 18S rDNA searches occasionally place santaloids near *Gunnera*, analyses of three of the four 18S rDNA data sets place santaloids close to Fabaceae and Polygalaceae.

**Celastroids.** Another small clade revealed in all analyses consists of *Lepuropetalon* and *Parnassia* (Saxifragaceae s.l.), *Brexia* (Grossulariaceae), and *Euonymus* (Celastraceae). This clade, labeled celastroids (Figs. 1–4), was also recovered in analyses of *rbcL* sequences (Chase et al., 1993; Morgan & Soltis, 1993). Although this initially appears to be an eclectic assemblage (*Brexia* is a genus of small trees; *Lepuropetalon spathulatum* is the smallest terrestrial angiosperm), embryological and morphological data also unite these taxa (reviewed in Morgan & Soltis, 1993). The celastroid clade consists of two pairs of genera, each of which is strongly supported: *Lepuropetalon*–*Parnassia* (jackknife = 100%) and *Brexia*–*Euonymus* (jackknife = 67%). These same two pairs of genera also were revealed in analyses of *rbcL* sequences (Chase et al., 1993; Morgan & Soltis, 1993).

**Cunonioids.** *Bauera* and *Ceratopetalum* (Cunoniaceae) and *Eucryphia* (Eucryphiaceae) form a clade with a jackknife value of 53%. A close relationship among these genera also was revealed by a cladistic analysis of morphological features (Hufford & Dickison, 1992). *Bauera*, *Ceratopetalum*, and *Eucryphia* constitute the core of a very well supported clade (jackknife value of 89%) labeled cunonioids (Figs. 1–4) that also contains Cephalotaceae, a family of carnivorous plants, and *Sloanea* (Elaeocarpaceae). A close relationship of Cephalotaceae to these same representatives of Cunoniaceae and Eucryphiaceae also is suggested by analyses of *rbcL* sequences (Chase et al., 1993; Morgan & Soltis, 1993). *Sloanea* was not represented in the broad analyses of *rbcL* sequences. Other taxa that appear closely allied with Cunoniaceae, Eucryphiaceae, and Cephalotaceae in *rbcL* analyses include Tremandraceae and Oxalidaceae; these families were not included, however, in the 18S rDNA analyses.

**Other noteworthy relationships.** As recently reviewed (Qiu et al., 1993), the placement of Lactoridaceae has been controversial, with relationships to Magnoliales, Laurales, and Piperales all proposed. Analyses of *rbcL* sequences suggested a close relationship of Lactoridaceae to Aristolochiaceae (Chase et al., 1993), and analyses of 18S rDNA sequences similarly suggest that these two

families are sisters (Figs. 1–4), an inference strongly supported by a jackknife value of 82%.

Additional, small monophyletic groups also merit brief discussion. Bombacaceae, Tiliaceae, and Malvaceae (represented by *Bombax*, *Luhea*, and *Gossypium*, respectively) form a strongly supported clade (jackknife value of 78%) in all 18S rDNA analyses, in agreement with both traditional treatments (all are members of Malvales) and topologies based on *rbcL* sequences. However, *Sloanea* (Elaeocarpaceae–Malvales) does not appear with Bombacaceae–Tiliaceae–Malvaceae in any of the 18S rDNA trees (Figs. 1–4). As noted above, this malvoid clade sometimes is embedded within the nitrogen-fixing clade (Figs. 1, 2), a placement at odds with analyses based on *rbcL* sequences. This unusual placement could be the result of insufficient taxon density in that many of the closest putative relatives of Malvales were not included here (e.g., Anacardiaceae, Simaroubaceae, Leitneriaceae, Sterculiaceae, Dipterocarpaceae).

On a broader scale, all 18S rDNA topologies suggest that Hamamelidae comprise a number of phylogenetically distinct lineages. For example, Trochodendraceae, Tetracentraceae, and Platanaceae appear near the base of the eudicots in trees derived from searches of data sets 1, 2, and 4 (Figs. 1, 2, 4). Eupteleaceae also appear near the base of the eudicots, but as part of the ranunculid clade. Three traditional families of Hamamelidae, Hamamelidaceae, Cercidiphyllaceae, and Daphniphyllaceae, are part of a well supported saxifragoid clade (Figs. 1–4). Still other families of Hamamelidae (i.e., Betulaceae, Urticaceae, Moraceae, and Ulmaceae) are part of the nitrogen-fixing clade, and Eucommiaceae are nested within the Asteridae s.l. The pronounced polyphyly of Hamamelidae was similarly revealed by analyses of *rbcL* sequences. Both 18S rDNA and *rbcL* sequence data suggest similar placements for representatives of this subclass.

Topologies based on 18S rDNA sequences also reveal the polyphyly of subclass Dilleniidae. Taxa attributed to Dilleniidae appear in several phylogenetically well separated clades. Paeoniaceae appear in the saxifragoid clade, Nepenthaceae and Droseraceae appear in Caryophyllidae s.l., Capparales, Batales, and Violales appear in the glucosinolate clade, and several orders (e.g., Violales, Ebenales, Ericales, Diapensiales, Primulales, and Theales) appear in Asteridae s.l. Other representatives of Dilleniidae (e.g., Turneraceae, Elaeocarpaceae) are scattered throughout the large Rosidae clade.

#### DIFFERENCES AMONG THE SHORTEST TREES

The shortest trees resulting from analyses of data sets 1 and 2 are essentially identical (Figs. 1, 2) and in turn are very similar to those derived from searches of data set 4 (Fig. 4). The most unusual topology results from searches of data set 3 (Fig. 3). For example, the distinction between the monosulcate grade and the eudicot clade does not occur in the shortest trees from this analysis, with the monocots part of a predominantly rosoid assemblage and saxifragoids appearing as one of the early-branching lineages of angiosperms. The ranunculids are not monophyletic in trees from searches of data set 3, with two genera (*Hypecoum* and *Dicentra*) appearing as sister to the monocots and the remaining ranunculids appearing as part of a clade that occupies the unusual position of sister to Asteridae s.l. (see Asteridae s.l. Plus, Fig. 3). However, some of our numerous searches of data set 3 found trees only one step longer than the shortest trees that have a topology essentially identical to that resulting from analysis of the other small data set (4).

Although searches of data sets 1, 2, and 4 yielded similar topologies, several weakly supported differences also exist among the shortest trees found. For example, in trees derived from analyses of data sets 1 and 2, one group of paleoherbs (Aristolochiaceae, Lactoridaceae) appears within the eudicot clade, rather than within the monosulcate grade, as would be expected. In contrast, in trees derived from the smaller data sets (3 and 4), Aristolochiaceae and Lactoridaceae appear within the monosulcate grade, close to other families of paleoherbs (e.g., Piperaceae, Saururaceae). In addition, the shortest trees obtained from analyses of data sets 1 and 2 show a more well defined break between the monosulcate grade and lower eudicots than do trees from data set 4 (compare Figs. 1, 2, and 4). For example, trees resulting from analyses of data set 4 place the monosulcate families Calycanthaceae, Annonaceae, and Lauraceae with Proteaceae (see Fig. 4, Lower eudicots/monosulcates). These and other differences may be the result of insufficient taxon density in certain portions of the tree, incomplete analysis, or lack of signal (see Caveats below).

#### ANOMALOUS PLACEMENTS

Perhaps the most unusual consistent feature of the 18S rDNA trees involves the placement of Caryophyllidae s.l. within Asteridae s.l. Although Caryophyllidae s.l. form a well supported clade, the position of this clade within Asteridae s.l. is not strongly supported. Some of the many searches con-

ducted retrieved trees only two steps longer than the shortest trees obtained in which Caryophyllidae s.l. are not part of Asteridae s.l., but appear instead within the Rosidae clade. Furthermore, analyses of 130 dicot 18S rDNA sequences aimed at resolving the relationships of Saxifragaceae s.l. did not place the Caryophyllidae s.l. within Asteridae s.l., but instead showed the caryophyllids to be embedded within a rosid clade (D. Soltis & Soltis, 1997). Although the placement of Caryophyllidae s.l. varies in the broad analyses of *rbcL* sequences (Chase et al., 1993), this clade does not appear closely related to the asterids in any of the shortest trees obtained. The 476-taxon analysis places Caryophyllidae s.l. within a clade of rosids, whereas the 499-taxon analysis places them near the split between the clades of higher eudicots (i.e., Rosidae and Asteridae s.l.).

Other anomalous placements include the position in some analyses (Figs. 1, 2) of one group of paleoherbs (Chloranthaceae, Aristolochiaceae, Lactoridaceae) plus Winteraceae of Magnoliales near the base of the eudicot clade. These taxa often are considered to represent early-branching or primitive angiosperms in both analyses of *rbcL* sequences (Chase et al., 1993; Qiu et al., 1993) and recent classification schemes (e.g., Cronquist, 1981; Thorne, 1992; Takhtajan, 1987). Based on phylogenetic analyses of *rbcL* sequences (e.g., Chase et al., 1993; Qiu et al., 1993), for example, Chloranthaceae, Aristolochiaceae, and Lactoridaceae are part of the monosulcate clade. As noted above, however, in some of our searches (see Figs. 3 and 4), Chloranthaceae, Aristolochiaceae, and Lactoridaceae do appear closer to the base of the angiosperms with other monosulcate taxa.

Analyses of data set 4 recovered a clade consisting of *Sagittaria* (Alismataceae) and *Cuscuta* (Cuscutaceae), placed in the Rosidae clade. In trees resulting from analyses of all other data sets, these genera appear with the monocots and Asteridae s.l., respectively, in agreement with traditional views and with trees based on *rbcL* sequences (Chase et al., 1993). The unusual relationship suggested by searches of data set 4 likely results from the more limited taxon sampling of this data set (fewer monocots are included, for example, compared to data sets 1 and 2) and long-branch attraction. *Sagittaria* and *Cuscuta* have very long branches (e.g., 39 and 65 steps, respectively, in Fig. 2) in all of the shortest trees obtained. In analyses of data sets 1 and 2, the long branch of *Cuscuta* also seems to affect the placement of *Ipomoea* (Convolvulaceae), with both appearing in Lamiales instead of Solanales.

The three subfamilies of Fabaceae, Papilionoideae, Mimosoideae, and Caesalpinioideae (represented by *Pisum* and *Glycine*; *Albizia*; *Bauhinia*, respectively), although present in the same small clade with several other families in Figures 1 and 2, do not form a monophyletic group in any of our searches. Rather than representing a true case of discordance between 18S rDNA and *rbcL* trees, this likely represents either the lower limits of resolution of 18S rDNA sequences (see below) or retrieval of only a small sample of all equally most parsimonious trees (i.e., the strict consensus of all shortest trees, had they been found, would have led to the collapse of this part of the tree). In support of the former conclusion is the observation that more focused phylogenetic studies of 18S sequences representing only Rosidae, some of which swapped to completion, similarly suggest a polyphyletic Fabaceae; bootstrap analyses indicate that these relationships are poorly supported, however (D. Soltis & Soltis, 1997, unpublished).

The position of the monocot genus *Acorus* (Araaceae) (Figs. 1, 2) also is unusual. Rather than appearing with the monocots, *Acorus* appears as an early-branching angiosperm, as it did in a previous analysis of 64 18S rDNA and rRNA sequences (Nickrent & Soltis, 1995). Other anomalous placements include the position of Dilleniaceae near the monocots (Figs. 1 and 2) and the unexpected position of *Oncidium* (Orchidaceae) as a first-branching monocot.

Several taxa are noteworthy not only because their phylogenetic positions are unusual, but also because their phylogenetic position varies from search to search. For example, the close relationship of *Gunnera* to the monocots (Figs. 1 and 2) is unexpected, but it is not seen in the trees resulting from analysis of data sets 3 and 4 where *Gunnera* appears in a clade with Santalales, Polygalaceae, and *Pisum* (Fabaceae). The relationship of *Gunnera* also is uncertain in *rbcL* topologies, in which its placement varies from being embedded within Asteridae s.l. (the 476-taxon search) to sister group of the higher dicots (the 499-taxon search).

#### COMPARISON WITH HAMBY AND ZIMMER (1992)

Hamby and Zimmer (1992) used partial 18S and 26S rRNA sequences to examine relationships among land plants. Because their analyses involved only 46 angiosperms, taxon sampling clearly differs between that and the present study. Nonetheless, brief comparison of the topologies resulting from the two studies is instructive.

In most of the shortest trees obtained here (Figs.

1, 2, 4), as well as in the study of Hamby and Zimmer (1992), *Ceratophyllum* is allied with the monocots. Both studies also concur in suggesting that Nymphaeaceae appear near the base of the angiosperm radiation. Nymphaeaceae are the sister group to all other angiosperms in Hamby and Zimmer's (1992) shortest trees; however, Amborellaceae, Austrobaileyaceae, and Schisandraceae were not included in that study. In all of our shortest trees, Nymphaeaceae follow the latter three families and Illiciaceae as the sister group to all remaining flowering plants.

Another similarity between the shortest trees in both studies is the placement of *Drimys* (Winteraceae). *Drimys* occupies an unusual phylogenetic position in trees presented by both Hamby and Zimmer (1992) and Nickrent and Soltis (1995), appearing as sister to *Glycine* and *Pisum* (Fabaceae), rather than as an early-branching angiosperm. *Drimys* occupies an unusual position in trees derived from the current analyses as well, appearing among the lower eudicots. In trees resulting from the analysis of data set 3, *Drimys* again appears with *Pisum*. The 18S rDNA sequence of *Drimys* exhibits a number of substitutions not found in other magnoliids. In an attempt to ascertain the relationships of Winteraceae, we sequenced two species of *Drimys*, *D. winteri* and *D. aromatica*, and they have identical sequences. More recently, another member of Winteraceae (*Pseudowintera*) has been sequenced for 18S rDNA (Hoot, unpublished); this sequence is nearly identical to the sequences for *Drimys*. Adding *Pseudowintera* to the analysis does not alter the unusual position of Winteraceae (trees not shown).

The unusual phylogenetic relationships that exist among the eudicots in the shortest trees of Hamby and Zimmer (1992) probably derive from insufficient sampling in that study. The present analysis with its greater representation of eudicots reveals relationships much more in accord with recent classifications (e.g., Cronquist, 1981; Takhtajan, 1987) and/or the *rbcL* topologies of Chase et al. (1993). Thus, the present study suggests that many of the highly unusual relationships seen in Hamby and Zimmer are likely to reflect low taxon density rather than an inherent inability of 18S rDNA sequences to resolve relationships.

#### CAVEATS

A number of limitations are inherent in any large phylogenetic study such as this. Several factors may contribute to the anomalous positions of certain taxa, including uncertainty regarding maximum parsimony, insufficient taxon sampling and/or den-

sity, the presence of "older," erroneous 18S rDNA sequences in the data matrix, and the overall lower rate of evolution of 18S rDNA compared to *rbcL*. We discuss these potential factors in more detail below.

An analysis of this magnitude cannot be expected to achieve maximum parsimony in a reasonable amount of time. It is likely that we did not find all classes of most-parsimonious trees, despite a search strategy (cf. Maddison et al., 1992) designed to identify multiple islands (Maddison, 1991) of shortest trees, and that even shorter trees exist that were not recovered. Furthermore, although our search strategy involved well over two years of computer time, no search swapped to completion; there is no assurance, therefore, that these trees represent even a local parsimony optimum. Although it is, of course, impossible to know how far from completion any search is when it is truncated, the search design used here offers an insightful basis for comparison. Data sets 1 and 2, and 3 and 4 are identical except for the inclusion of two gap characters (indels) in data sets 2 and 4, each of which apparently accounts for only four steps on the shortest trees obtained. Thus, the fact that the shortest trees obtained in searches of data set 2 are seven steps longer than those obtained in searches of data set 1 indicates that the shortest trees obtained in our searches of data set 2 are three steps less parsimonious than trees derived from searches of data set 1. A similar comparison of the searches of data sets 3 and 4 reveals that the shortest trees from searches of data set 4 are two steps less parsimonious than those obtained from data set 3.

We also sampled among the large set of equally parsimonious trees following Sanderson and Doyle (1993b). Using trees obtained in searches of data set 1, we examined the number of distinct components (clades) as a function of the size of the sample of trees (number of trees). We wanted to determine whether increasing the set of trees uncovers new components that bear on the relationships of particular taxa or, in contrast, includes different subsets of the components that are essentially variations on the same theme (Sanderson & Doyle, 1993b). We found that a plot of the number of distinct clades versus the number of trees sampled reaches an asymptote for a small number of trees, suggesting that most of the clade diversity has been found, despite the fact that all most parsimonious trees have not been retrieved. The development of improved methods of phylogenetic analysis of large data sets will ultimately be one of the central issues of phylogeny reconstruction during the next several years (see discussions in Chase et al., 1993; Doyle

et al., 1994; Mishler, 1994; P. Soltis & Soltis, 1997).

Although the anomalous relationships described for some taxa may be unsettling, extremely short branches characterize most of the major clades in the 18S rDNA trees. The internal support for many branches is very low, as indicated by the parsimony jackknife analysis (Farris et al., 1997). Although the monophyly of the angiosperms is well supported (jackknife value of 100%), few major clades within the angiosperms have high jackknife values. For example, large clades such as eudicots and Rosidae do not have jackknife values above 50%; the saxifragoids represent the largest clade having a high jackknife value (jackknife value of 68%). The other monophyletic groups with high jackknife values are relatively small, such as eunonioids, Zingiberales, Malvales, Caryophyllales, Lactoridaceae–Aristolochiaceae, and Schisandraceae–Illiciaceae–Austrobaileyaceae. Significantly, a number of major clades seen in all shortest trees, as well as in trees many steps longer than the most parsimonious trees, do not have jackknife values above 50%, including monocots, glucosinolates, Caryophyllidae s.l., and Asteridae s.l. The majority of high jackknife values correspond to pairs of sister taxa representing terminal nodes (e.g., *Calycanthus*–*Sassafras*, *Brexia*–*Euonymus*, *Lepuropetalon*–*Parnassia*, *Plumbago*–*Cocoloba*, *Helwingia*–*Phyllonoma*, *Tragopogon*–*Tagetes*, *Francoa*–*Greyia*, *Trochodendron*–*Tetracentron*, *Menispermum*–*Tinospora*).

Examination of trees obtained from searches that found trees one or a few steps longer than the shortest trees also suggests low internal support for some branches. The phylogenetic position of the monocots appears weakly supported. In some searches of data set 2, for example, trees only one step longer than the shortest trees place the monocots within the eudicots, as part of Rosidae, a position also observed in the shortest trees obtained from searches of data set 3 (Fig. 3). Although all of the starting trees and shortest trees showed Amborellaceae, Illiciaceae, Schisandraceae, and Austrobaileyaceae to be at the base of the angiosperms, one search of data set 2 resulted in trees two steps longer than the shortest trees and placed these four families near the monocots, with *Acorus* and *Oncidium* as the first-branching angiosperms. Trees two steps longer than the shortest trees show the Asteridae s.l. embedded within Rosidae, rather than sister to this large clade. In trees two steps longer than the shortest trees found for data set 3, Caryophyllidae s.l. are not part of Asteridae s.l. but instead are part of the large Rosidae clade.

These few examples illustrate well the uncertain-

ty that surrounds some angiosperm relationships inferred from analyses of 18S rDNA sequences. Furthermore, because relatively few character-state changes occur on many of the branches, a small amount of homoplasy or error in the data set may be sufficient to distort some relationships.

Additionally, some of the anomalous placements could reflect insufficient and/or uneven taxon sampling. The somewhat uneven taxonomic distribution of the sequences presently available means that some groups, such as Asteridae, and much of Rosidae and Hamamelidae, are relatively well represented here, whereas Magnoliidae, the monocots, Dilleniidae, Caryophyllidae, and several orders of Rosidae are under-represented.

The importance of sufficient taxon density is revealed here by some of the differences in topology observed between trees resulting from analyses of the smaller and larger data sets. Many of the taxa not present in the two smaller data sets (3 and 4) represent monosulcates and lower eudicots. It is this portion of the overall topology that shows the most spurious relationships in trees derived from analyses of these two small data sets (the distinction between the monosulcate grade and eudicots largely breaks down in Fig. 3, for example). In contrast, the much more thoroughly represented Asteridae s.l. and Rosidae clades are little affected by slightly decreased representation in data sets 3 and 4. These findings lend further support to the importance of sufficient and equal taxon density in attempts to infer angiosperm phylogeny (e.g., Sytsma & Baum, 1996).

One of the major lessons of this study is that the 18S rRNA gene is difficult to sequence, apparently due in large part to the secondary structure inherent in the rRNA. As a result, many published sequences are erroneous, some highly so, and the extent of insertion and deletion events has been greatly overestimated. We reiterate that whereas the total length of the aligned 18S rDNA data matrix of 64 taxa used by Nickrent and Soltis (1995) was 1853 bp, the length of our 228-taxon data matrix actually is shorter, 1850 bp. After resequencing over 20 dubious 18S rDNA sequences, we were able to remove numerous "false" indels and reduce the length of the aligned sequences. The great majority (70%) of the 18S rDNA sequences used here were generated via cycle sequencing followed by automated sequencing, an approach that provides more reliable rDNA sequences. Additional "older" 18S rDNA sequences should be replaced with sequences generated via this approach.

The overall slower rate of evolution of 18S rDNA compared to *rbcL* (see Nickrent & Soltis, 1995)

contributed, in part, to the widespread belief that 18S rDNA sequences would not contribute greatly to phylogenetic inference in angiosperms. Although this study and other recent papers employing entire 18S rDNA sequences (e.g., Nickrent & Soltis, 1995; Kron, 1996; D. Soltis & Soltis, 1997; Rodman et al., submitted; Johnson et al., unpublished) have dispelled this notion, 18S rDNA sequences will, in most cases, not elucidate relationships to the degree possible with the more rapidly evolving *rbcL*. In some groups such as Orchidaceae, however, 18S rDNA has been found to evolve faster than *rbcL* (Cameron and Chase, unpublished).

#### FUTURE CONSIDERATIONS

These exploratory analyses clearly illustrate the phylogenetic potential of 18S rDNA sequences for elucidating angiosperm relationships at higher taxonomic levels. Future attempts to conduct broad phylogenetic analyses of 18S rDNA sequences should not only add more taxa, but should also involve the resequencing of the 18S rRNA gene for some of those taxa for which erroneous sequences are suspected.

This study suggests that a broad, nuclear-based phylogenetic hypothesis for the angiosperms is achievable via sequence analysis of the 18S rRNA gene. One of the strengths of 18S sequence data appears to be the ability to recognize a suite of groups that appear in all shortest trees (e.g., glucosinolate clade, saxifragoids, Caryophyllidae s.l., Asteridae s.l., celastroids). This may reflect substitutions that occurred in highly conserved portions of the 18S rRNA gene during the early diversification of a lineage, resulting in a well-supported clade. Such substitutions are rare, however, and the result is limited resolution in some areas of the 18S rDNA topologies. Thus, our results also clearly reveal that 18S rDNA topologies will, in most cases, not exhibit the degree of resolution and internal support possible with *rbcL* sequences. Increased sampling of angiosperms for 18S rDNA sequence analysis is desirable. However, to achieve a nuclear-based estimate of angiosperm phylogeny comparable to that realized with *rbcL*, it probably will be necessary to include all, or portions of, the 26S rRNA gene as well. The utility of portions of the 26S gene for inferring family-level relationships has been demonstrated for angiosperms (Hamby & Zimmer, 1992), as well as for other groups of organisms (e.g., Buchheim & Chapman, 1991; Chapman & Buchheim, 1991; Chapela et al., 1994; Waters et al., 1992).

#### CONCLUSIONS

This study provides general insights into the structure and evolution of the 18S rRNA gene in angiosperms and dispels certain "myths" about its evolution. Indels are neither as common nor as problematic for alignment as previously believed. Instead, they are largely confined to a few, small, specific regions that correspond to the termini of certain helices present in the proposed secondary structure for 18S rRNA. When these few, short areas are eliminated from consideration, alignment of 18S rDNA sequences is straightforward and easily accomplished by eye across all angiosperms. Conversely, indels are rare throughout most of the 18S rRNA gene; when present, they typically involve a single base pair. Furthermore, indels present in highly conserved regions of the gene may, in fact, be phylogenetically informative, such as the insertion that unites saxifragoids and the deletion that unites higher eudicots.

Initial attempts to evaluate the impact of secondary structure of the 18S rRNA transcript on phylogeny reconstruction in angiosperms suggest that both stem and loop regions appear to be sources of phylogenetic information, with a slightly greater proportion (58% vs. 42%) of informative sites found in stem rather than loop regions. Of the stem changes we analyzed, only 27% destroyed a base-pairing couplet; 73% restored or maintained stem base pairing and hence are considered compensatory. The most frequent type of stem change observed involved single base substitutions that changed one base-pairing couplet to another (e.g., U-G to C-G; U-A to U-G). The high frequency of compensatory change indicates that some downweighting of stem characters relative to loop bases may be warranted in future broad analyses of 18S rDNA sequences.

The phylogenetic trees obtained in these exploratory, broad analyses of 18S rDNA sequences are largely concordant with those resulting from analyses of *rbcL* sequences. Areas of general concordance include the presence of a tricolpate or eudicot clade, which in turn includes two large clades corresponding mostly to Rosidae and Asteridae s.l., respectively. However, the latter clade also includes Caryophyllidae s.l. in 18S rDNA trees, but not in trees retrieved from analyses of *rbcL* sequences. In addition, the monocotyledons are monophyletic (with the possible exception of *Acorus*) and generally appear with other taxa having monosulcate pollen. One of the most noteworthy differences between this study and that of Chase et al. (1993) concerns the first-branching angiosperms. The

woody magnoliids Amborellaceae, Illiciaceae, Schisandraceae, and Austrobaileyaceae consistently appear as first-branching angiosperms and are always followed by the paleoherb Nymphaeaceae. *Ceratophyllum* is closely allied with the monocots and does not appear as sister to all other angiosperms, as in analyses of *rbcL* sequences (Chase et al., 1993). Monophyletic groups apparent in all analyses include Caryophyllidae s.l., Asteridae s.l., saxifragoids, glucosinolate-producing taxa, santaloids, and Cunonioids. Other clades apparent in most analyses include ranunculids and nitrogen-fixing taxa. Thus, this analysis identifies major clades of angiosperms that are largely consistent with those inferred from *rbcL* analyses.

This study further demonstrates that 18S rDNA sequences contain sufficient information to conduct phylogenetic studies at higher taxonomic levels in the angiosperms. Additional phylogenetic analyses of angiosperms should be conducted using a larger 18S rDNA data set that improves taxon sampling for Magnoliidae and Dilleniidae in particular. In constructing this larger data set, some taxa for which published sequences are available should first be resequenced.

Although comparative sequencing of the entire 18S rRNA gene holds great promise for retrieving phylogeny at the family level and above in the angiosperms, this nuclear gene will rarely elucidate familial and generic relationships to the extent possible with *rbcL* (see also Nickrent & Soltis, 1995). Due to the slower rate of evolution of 18S rDNA compared to *rbcL*, it likely will be necessary to sequence the 26S rDNA as well to obtain a nuclear-based estimate of phylogeny comparable to that achieved with *rbcL*. Lastly, because of the general congruence of 18S rDNA and *rbcL* topologies for angiosperms, this study concomitantly suggests that 18S rDNA and *rbcL* sequences should be combined to provide a more accurate estimate of angiosperm phylogeny. One can anticipate that other sequences (e.g., *atpB* and 26S rDNA) will ultimately also be combined with *rbcL* and 18S rDNA sequences to provide a larger data set from which to infer a more complete picture of angiosperm phylogeny.

#### Literature Cited

- Bailey, I. W. 1957. The potentialities and limitations of wood anatomy in the study of the phylogeny and classification of angiosperms. *J. Arnold Arbor.* 38: 243–254.
- Bakker, F. T., J. L. Olsen, W. T. Stam & C. Van den Hoek. 1994. The *Cladophora* complex (Chlorophyta): New views based on 18S rRNA gene sequences. *Molec. Phylogenetics and Evol.* 3: 365–382.
- Bhattacharya, D. & 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. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42: 795–803.
- Buchheim, M. A. & R. L. Chapman. 1991. Phylogeny of the colonial green flagellates: A study of 18S and 26S rRNA sequence data. *BioSystems* 25: 85–100.
- , M. Turmel, E. A. Zimmer & R. L. Chapman. 1990. Phylogeny of *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S rRNA sequence data. *J. Phycol.* 26: 689–699.
- Bult, C., M. Kallersjö & Y. Suh. 1992. Amplification and sequencing of 16/18S rDNA from gel-purified total plant DNA. *Pl. Molec. Biol. Reporter* 10: 273–284.
- Capesius, I. 1995. A molecular phylogeny of bryophytes based on the nuclear encoded 18S rRNA genes. *J. Pl. Physiol.* 146: 59–63.
- Chapela, I. H., S. A. Rehner, T. R. Schultz & U. G. Mueller. 1994. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* 266: 1691–1694.
- Chapman, R. L. & M. Buchheim. 1991. Ribosomal RNA gene sequences: Analysis and significance in the phylogeny and taxonomy of green algae. *C. R. C. Crit. Rev. Pl. Sci.* 10: 343–368.
- Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-I. Kim, C. F. Wimpee, J. F. Smith, G. R. Fournier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 80: 528–580.
- Chaw, S.-M., H. Long, B.-S. Wang, A. Zharkikh & W.-H. Li. 1993. The phylogenetic position of Taxaceae based on 18S rRNA sequences. *J. Molec. Evol.* 37: 624–630.
- , H.-M. Sung, H. Long, A. Zharkikh & W.-H. Li. 1995. The phylogenetic positions of the conifer genera *Amentotaxus*, *Phyllocladus*, and *Nageia* inferred from 18S rRNA sequences. *J. Molec. Evol.* 41: 224–230.
- , A. Zharkikh, H.-M. Sung, T.-C. Lau & W.-H. Li. 1997. Molecular phylogeny of gymnosperms and seed plant evolution: Analysis of 18S rRNA sequences. *J. Molec. Evol.* in press.
- Crane, P. R. 1985. Phylogenetic analysis of seed plants and the origin of the angiosperms. *Ann. Missouri Bot. Gard.* 72: 716–793.
- , 1988. Major clades and relationships in the “higher” gymnosperms. Pp. 218–272 in C. B. Beck (ed.), *Origin and Evolution of Gymnosperms*. Columbia Univ. Press, New York.
- , E. M. Friis & K. R. Pedersen. 1995. The origin and early diversification of angiosperms. *Nature* 374: 27–33.
- Cronquist, A. 1968. *The Evolution and Classification of Higher Plants*. Houghton, Mifflin, Boston.
- , 1981. *An Integrated System of Classification of Flowering Plants*. Columbia Univ. Press, New York.
- Curtiss, W. C. & J. N. Vournakis. 1984. Quantitation of

- base substitutions in eukaryotic 5S rRNA: Selection for the maintenance of RNA secondary structure. *J. Molec. Evol.* 20: 351–361.
- Dahlgren, R. T. 1980. A revised system of classification of the angiosperms. *Bot. J. Linn. Soc.* 80: 91–124.
- . 1983. General aspects of angiosperm evolution and macrosystematics. *Nordic J. Bot.* 3: 119–149.
- Dahlgren, R. T., H. T. Clifford & P. F. Yeo. 1985. *The Families of the Monocotyledons: Structure, Evolution, and Taxonomy*. Springer-Verlag, New York.
- Dixon, M. T. & D. M. Hillis. 1993. Ribosomal RNA secondary structure—Compensatory mutations and implications for phylogenetic analysis. *Molec. Biol. Evol.* 10: 256–267.
- Donoghue, M. J. & J. A. Doyle. 1989a. Phylogenetic analysis of angiosperms and the relationships of Hamamelidaceae. Pp. 17–45 in P. R. Crane & S. Blackmore (editors), *Evolution, Systematics and Fossil History of the Hamamelidaceae*. Clarendon Press, Oxford.
- & ———. 1989b. Phylogenetic studies of seed plants and angiosperms based on morphological characters. Pp. 181–193 in B. Fernholm, K. Bremer & H. Jornvall (editors), *The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis*. Elsevier Science Publishers, Amsterdam.
- , R. G. Olmstead, J. F. Smith & J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbcL* sequences. *Ann. Missouri Bot. Gard.* 79: 333–345.
- Doyle, J. A. & M. J. Donoghue. 1986. Seed plant phylogeny and the origin of the angiosperms: An experimental cladistic approach. *Bot. Rev.* 52: 321–431.
- & ———. 1992. Fossils and seed plant phylogeny reanalyzed. *Brittonia* 44: 89–106.
- & C. L. Hotton. 1991. Diversification of early angiosperm pollen in a cladistic context. Pp. 169–195 in S. Blackmore & S. H. Barnes (editors), *Pollen and Spores*. Clarendon Press, Oxford.
- , M. J. Donoghue & E. A. Zimmer. 1994. Integration of morphological and ribosomal RNA data on the origin of angiosperms. *Ann. Missouri Bot. Gard.* 81: 419–450.
- , C. L. Hotton & J. V. Ward. 1990. Early Cretaceous tetrads, zonaulculate pollen, and Winteraceae. II. Cladistic analysis and implications. *Amer. J. Bot.* 77: 1558–1568.
- Duvall, M. R., M. T. Clegg, M. W. Chase, W. D. Clark, W. J. Kress, H. G. Hills, L. E. Eguiarte, J. F. Smith, B. S. Gaut, E. A. Zimmer & G. H. Learn, Jr. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. *Ann. Missouri Bot. Gard.* 80: 607–619.
- Eckenrode, V. K., J. Arnold & R. B. Meagher. 1985. Comparison of the nucleotide sequence of soybean 18S rDNA with the sequences of other small-subunit rRNAs. *J. Molec. Evol.* 21: 259–269.
- Embley, M. T., R. P. Hirt & D. M. Williams. 1994. Biodiversity at the molecular level: The domains, kingdoms and phyla of life. *Philos. Trans., Ser. B* 345: 21–31.
- Endress, P. K. 1994. Evolutionary aspects of the floral structure in *Ceratophyllum*. *Pl. Syst. Evol.* 8: 175–183.
- & R. Honneger. 1980. The pollen of the Austrobaileyaaceae and its phylogenetic significance. *Grana* 19: 177–182.
- Farris, J. S., V. A. Albert, M. Kallersjö, D. Lipscomb & A. G. Kluge. 1997. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 11: in press.
- Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace & R. A. Raff. 1988. Molecular phylogeny of the animal kingdom. *Science* 239: 748–753.
- Forster, H., M. D. Coffey, H. Elwood & M. L. Sogin. 1990. Sequence analysis of the small subunit ribosomal RNAs of three zoospore fungi and implications for fungal evolution. *Mycologia* 82: 306–312.
- Gargas, A., P. T. De Priest & J. W. Taylor. 1995. Positions of multiple insertions in SSU rDNA of lichen-forming fungi. *Molec. Biol. Evol.* 12: 208–218.
- Goremykin, V., V. Bobrova, J. Pahnke, A. Troitsky, A. Antonov & W. Martin. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support Gnetalean affinities of angiosperms. *Molec. Biol. Evol.* 13: 383–396.
- Hamby, K. R. & E. A. Zimmer. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Pp. 50–91 in P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), *Molecular Systematics of Plants*. Chapman and Hall, New York.
- Hendricks, L., R. De Baere, Y. Van de Peer, J. Neefs, A. Goris & R. De Wachter. 1991. The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. *J. Molec. Evol.* 32: 167–177.
- Hillis, D. M. & M. T. Dixon. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quart. Rev. Biol.* 66: 411–453.
- Hinkle, G., J. K. Wetterer, T. R. Schultz & M. L. Sogin. 1994. Phylogeny of the attine ant fungi based on analysis of small subunit ribosomal RNA gene sequences. *Science* 266: 1695–1697.
- Hoot, S. B. & P. R. Crane. 1995. Inter-familial relationships in the Ranunculidae based on molecular systematics. *Pl. Syst. Evol.* 9: 119–131.
- , A. Culham & P. R. Crane. 1995. Phylogenetic relationships of the Lardizabalaceae and Sargentodoxaceae: Chloroplast and nuclear DNA sequence evidence. *Pl. Syst. Evol.* 9: 195–199.
- Hufford, L. & W. C. Dickson. 1992. A phylogenetic analysis of Cunoniaceae. *Syst. Bot.* 17: 181–200.
- Huss, V. A. R. & M. L. Sogin. 1990. Phylogenetic position of some *Chlorella* species within the Chlorococcales based upon complete small-subunit ribosomal RNA sequences. *J. Molec. Evol.* 31: 432–442.
- Kant, T. S., E. C. Theriot, E. A. Zimmer & R. L. Chapman. 1990. The Pleurostrophyceae and Micromonadophyceae: Cladistic analysis of nuclear rRNA sequence data. *J. Phycol.* 26: 711–721.
- Kranz, H. D., D. Miks, M.-L. Siegler, I. Capesius, Ch. W. Sensen & 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. Molec. Evol.* 41: 74–84.
- Kress, W. J., W. J. Hahn, T. M. Evans & E. A. Zimmer. 1995. Unraveling the evolutionary radiation of the families of the Zingiberaceae using morphological, molecular, and fossil characters. *Amer. J. Bot.* 82 (suppl.): 142.
- Kron, K. A. 1996. Phylogenetic relationships of Empetraceae, Epacridaceae, and Ericaceae: Evidence from nuclear ribosomal 18S sequence data. *Ann. Bot. (London)* 77: 293–303.
- & M. W. Chase. 1993. Systematics of Ericaceae, Empetraceae, Epacridaceae and related taxa based



- upon *rbcL* sequence data. *Ann. Missouri Bot. Gard.* 80: 735–741.
- Les, D. H. 1988. The origins and affinities of the Cera-  
tophyllaceae. *Taxon* 37: 326–345.
- , D. K. Garvin & C. F. Wimpee. 1991. Molecular  
evolutionary history of ancient aquatic angiosperms.  
*Proc. Natl. Acad. U.S.A.* 88: 10119–10123.
- Loconte, H. & D. W. Stevenson. 1991. Cladistics of the  
Magnoliidae. *Cladistics* 7: 267–296.
- Maddison, D. R. 1991. Discovery and importance of mul-  
tiple islands of most-parsimonious trees. *Syst. Zool.* 40:  
315–328.
- , M. Ruvolo & D. L. Swofford. 1992. Geographic  
origins of human mitochondrial DNA: Phylogenetic evi-  
dence from control region sequences. *Syst. Biol.* 41:  
111–124.
- Martin, P. G. & J. M. Dowd. 1991. Studies of angiosperm  
phylogeny using protein sequences. *Ann. Missouri Bot.  
Gard.* 78: 296–337.
- Mindell, D. P. & R. L. Honeycutt. 1990. Ribosomal RNA  
in vertebrates: Evolution and phylogenetic applications.  
*Ann. Rev. Ecol. Syst.* 21: 541–566.
- Mishler, B. D. 1994. Cladistic analysis of molecular and  
morphological data. *Amer. J. Phys. Anthropol.* 94: 143–  
156.
- , K. Bremer, C. J. Humphries & S. P. Churchill.  
1988. The use of nucleic acid sequence data in phylo-  
genetic reconstruction. *Taxon* 37: 391–395.
- , L. A. Lewis, M. A. Buchheim, K. S. Renzaglia,  
D. J. Garbary, C. F. Delwiche, F. W. Zechman, T. S.  
Kantz & R. L. Chapman. 1994. Phylogenetic relation-  
ships of the “green algae” and “bryophytes.” *Ann. Mis-  
souri Bot. Gard.* 81: 451–483.
- Morgan, D. R. & D. E. Soltis. 1993. Phylogenetic rela-  
tionships among members of Saxifragaceae sensu lato  
based on *rbcL* sequence data. *Ann. Missouri Bot. Gard.*  
80: 631–660.
- Nairn, C. J. & R. J. Ferl. 1988. The complete nucleotide  
sequence of the small-subunit ribosomal RNA coding  
region for the cycad *Zamia pumila*: Phylogenetic im-  
plications. *J. Molec. Evol.* 27: 133–141.
- Nickrent, D. L. 1994. From field to film: Rapid sequenc-  
ing methods for field collected plant species. *Bio-  
Techniques* 16: 470–474.
- & C. R. Franchina. 1990. Phylogenetic relation-  
ships of the Santalales and relatives. *J. Molec. Evol.* 31:  
294–301.
- & D. E. Soltis. 1995. A comparison of angio-  
sperm phylogenies from nuclear 18S rDNA and *rbcL*  
sequences. *Ann. Missouri Bot. Gard.* 82: 208–234.
- & E. M. Starr. 1994. High rates of nucleotide  
substitution in nuclear small-subunit (18S) rDNA from  
holoparasitic flowering plants. *J. Molec. Evol.* 39: 62–  
70.
- Nixon, K. C., W. L. Crepet, D. Stevenson & E. M. Friis.  
1994. A reevaluation of seed plant phylogeny. *Ann.  
Missouri Bot. Gard.* 81: 484–533.
- Noller, H. F. 1984. Structure of ribosomal RNA. *Annual  
Rev. Biochem.* 53: 119–162.
- Olmstead, R. G., H. J. Michaels, K. M. Scott & J. D.  
Palmer. 1992. Monophyly of the Asteridae and iden-  
tification of their major lineages inferred from DNA se-  
quences of *rbcL*. *Ann. Missouri Bot. Gard.* 79: 249–  
265.
- , B. Bremer, K. M. Scott & J. D. Palmer. 1993. A  
parsimony analysis of the Asteridae sensu lato based on  
*rbcL* sequences. *Ann. Missouri Bot. Gard.* 80: 700–722.
- Olsen, G. J. 1987. Earliest phylogenetic branchings:  
Comparing rRNA-based evolutionary trees inferred with  
various techniques. *Cold Spring Harbor Symp. Quant.  
Biol.* 52: 825–837.
- Olsen, J. L., W. T. Stam, S. Berger & D. Menzel. 1994.  
18S rDNA and evolution in the Dasycladales (Chlo-  
rophyta): Modern living fossils. *J. Phycol.* 30: 729–744.
- Qiu, Y.-L., M. W. Chase, D. H. Les & C. R. Parks. 1993.  
Molecular phylogenetics of the Magnoliidae: Cladistic  
analyses of nucleotide sequences of the plastid gene  
*rbcL*. *Ann. Missouri Bot. Gard.* 80: 587–606.
- Ragan, M. A., C. J. Bird, E. L. Rice, R. R. Gutell, C. A.  
Murphy & R. K. Singh. 1994. A molecular phylogeny  
of the marine red algae (Rhodophyta) based on nuclear  
small-subunit rRNA gene. *Proc. Natl. Acad. U.S.A.* 91:  
7276–7280.
- Rathgeber, J. & I. Capesius. 1990. Nucleotide sequence  
of the intergenic spacer and the 18S ribosomal RNA  
gene from mustard (*Sinapis alba*). *Nucl. Acids Res.* 18:  
1288.
- Rettig, J. H., H. D. Wilson & J. R. Manhart. 1992. Phy-  
logeny of the Caryophyllales—gene sequence data. *Taxon*  
41: 201–209.
- Rice, K. A., M. J. Donoghue & R. G. Olmstead. 1997. A  
reanalysis of the large *rbcL* data set: Implications for  
future phylogenetic studies. *Syst. Biol.* in press.
- Rodman, J. E. 1991. A taxonomic analysis of glucosi-  
nolate-producing plants, Part 2: Cladistics. *Syst. Bot.*  
16: 619–629.
- , P. S. Soltis, D. E. Soltis & K. J. Sytsma. Dual  
origin of mustard oil biosynthesis inferred from congru-  
ent nuclear and plastid gene phylogenies. *Proc. Natl.  
Acad. Sci. U.S.A.*, submitted.
- , M. K. Oliver, R. R. Nakamura, J. U. McClammer,  
Jr. & A. H. Bledsoe. 1984. A taxonomic analysis and  
revised classification of Centropsermae. *Syst. Bot.* 9:  
297–323.
- , R. A. Price, K. Karol, E. Conti, K. J. Sytsma &  
J. D. Palmer. 1993. Nucleotide sequences of the *rbcL*  
gene indicate the monophyly of mustard oil plants. *Ann.  
Missouri Bot. Gard.* 80: 686–699.
- Sanderson, M. J. & J. J. Doyle. 1993a. Reconstruction of  
organismal gene phylogenies from data on multigene  
families: Concerted evolution, homoplasy, and confi-  
dence. *Syst. Biol.* 42: 4–17.
- & ———. 1993b. Phylogenetic relationships in  
North American *Astragalus* (Fabaceae) based on chlo-  
roplast DNA restriction site variation. *Syst. Bot.* 18:  
395–408.
- & M. J. Donoghue. 1994. Shifts in diversification  
rate with the origin of angiosperms. *Science* 264: 1590–  
1593.
- Savard, L. & M. Lalonde. 1991. Sequence of 18S rDNA  
of actinorhizal *Alnus glutinosa* (Betulaceae). *Pl. Molec.  
Biol.* 16: 725–728.
- Schlegel, M., H. J. Elwood & M. L. Sogin. 1991. Molec-  
ular evolution in hypotrichous ciliates: Sequence of the  
small subunit ribosomal RNA genes from *Onychodro-  
mus quadricornutus* and *Oxytricha granulifera* (Oxytri-  
chidae, Hypotrichida, Ciliophora). *J. Molec. Evol.* 32:  
64–69.
- Senecoff, J. F. & R. B. Meagher. 1992. In vivo analysis  
of plant RNA structure: Soybean 18S ribosomal and  
ribulose-1,5-bisphosphate carboxylase small subunit  
RNAs. *Pl. Molec. Biol.* 18: 219–234.
- Smith, A. B. 1989. RNA sequence data in phylogenetic

- reconstruction: Testing the limits of its resolution. *Cladistics* 5: 321-344.
- Sogin, M. L., H. J. Elwood & J. H. Gunderson. 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc. Natl. Acad. U.S.A.* 83: 1383-1387.
- Soltis, D. E. & P. S. Soltis. 1997. Phylogenetic relationships among Saxifragaceae sensu lato: A comparison of topologies based in 18S rDNA and *rbcL* sequences. *Amer. J. Bot.* in press.
- , ———, D. R. Morgan, S. M. Swensen, B. C. Mullin, J. M. Dowd & P. G. Martin. 1995. Chloroplast gene sequence data suggest a single origin of the pre-disposition for symbiotic nitrogen fixation in angiosperms. *Proc. Natl. Acad. U.S.A.* 92: 2647-2651.
- Soltis, P. S. & D. E. Soltis. 1997. Phylogenetic analysis of large molecular data sets. *Bol. Soc. Bot. México*. In press.
- Stebbins, G. L. 1974. *Flowering Plants: Evolution Above the Species Level*. Harvard Univ. Press, New York.
- Swann, E. C. & J. W. Taylor. 1993. Higher taxa of basidiomycetes: An 18S rRNA gene perspective. *Mycologia* 85: 923-936.
- Swofford, D. L. 1993. PAUP—Phylogenetic Analysis Using Parsimony, version 3.1. Center for Biodiversity, Illinois Natural History Survey, Champaign, Illinois.
- & G. J. Olsen. 1990. Phylogeny reconstruction. Pp. 411-501 in D. Hillis & C. Moritz (editors), *Molecular Systematics*. Sinauer, Sunderland, Massachusetts.
- Sytsma, K. J. & D. A. Baum. 1996. Molecular phylogenies and the diversification of the angiosperms. Pp. 314-340 in D. W. Taylor & L. J. Hickey (editors), *Flowering Plant Origin, Evolution, and Phylogeny*. Chapman and Hall, New York.
- Takaiwa, F., K. Oono & M. Sugiura. 1984. The complete nucleotide sequence of a rice 17S rRNA gene. *Nucl. Acids. Res.* 12: 5441-5448.
- Takhtajan, A. 1969. *Flowering Plants: Origin and Dispersal*. Oliver & Boyd, Edinburgh.
- . 1980. Outline of the classification of flowering plants (Magnoliophyta). *Bot. Rev.* 46: 225-359.
- . 1987. *System of Magnoliophyta*. Academy of Sciences U.S.S.R., Leningrad.
- Thorne, R. F. 1992. An updated classification of the flowering plants. *Aliso* 13: 365-389.
- Unfried, I., U. Stocker & P. Gruendler. 1989. Nucleotide sequence of the 18S rRNA gene from *Arabidopsis thaliana* Col0. *Nucl. Acids. Res.* 17: 7513.
- Wada, H. & N. Satoh. 1994. Phylogenetic relationships among extant classes of echinoderms, as inferred from sequences of 18S rDNA, coincide with relationships deduced from the fossil record. *J. Molec. Evol.* 38: 41-49.
- Wainright, P. O., O. Hinkle, M. L. Sogin & 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 & R. L. Chapman. 1992. Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. *Amer. J. Bot.* 79: 459-466.
- Wheeler, W. C. & R. L. Honeycutt. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Molec. Biol. Evol.* 5: 90-96.
- Woese, C. R. 1987. Bacterial evolution. *Microbiol. Rev.* 51: 221-271.
- Wolters, J. & V. A. Erdmann. 1986. Cladistic analysis of 5S rRNA and 16S rRNA secondary and primary structure—The evolution of eukaryotes and their relation to Archaeobacteria. *J. Molec. Evol.* 24: 152-166.
- Young, D. A. 1981. Are the angiosperms primitively vesselless? *Syst. Bot.* 6: 313-330.
- Young, D. J. & L. Watson. 1970. The classification of dicotyledons: A study of the upper levels of hierarchy. *Austral. J. Bot.* 18: 387-433.
- Zimmer, E. A., R. K. Hamby, M. L. Arnold, D. A. Leblanc & E. C. Theriot. 1989. Ribosomal RNA phylogenies and flowering plant evolution. Pp. 205-214 in B. Fernholm, K. Bremer & J. Jönvall (editors), *The Hierarchy of Life*. Elsevier Science Publishers, Amsterdam.

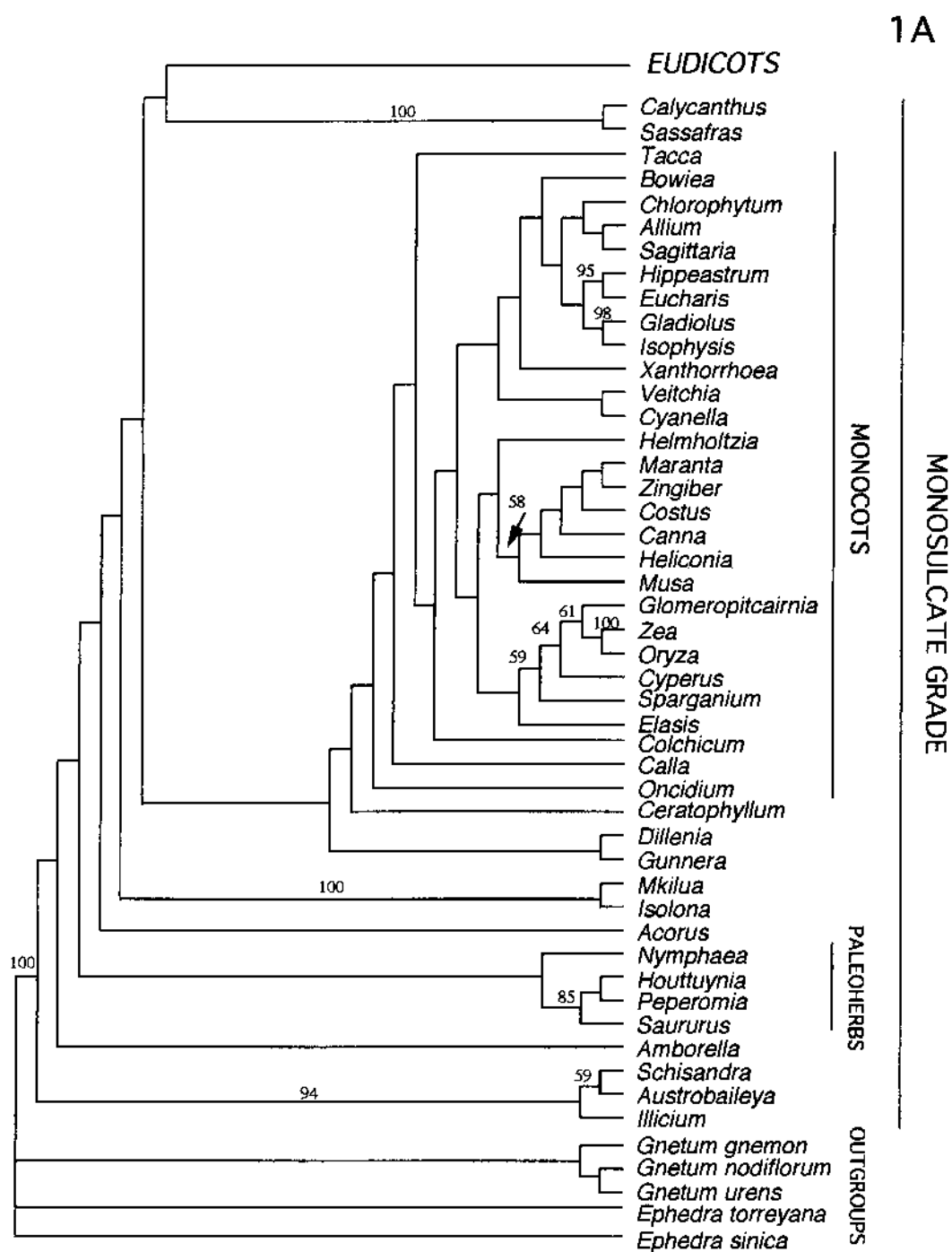


Figure 1. Strict consensus of 5294 shortest trees resulting from the exploratory phylogenetic analysis of 223 species of angiosperms. Each of the shortest trees has a length of 3923 steps, CI = 0.235, and RI = 0.535. Parsimony jackknife values (Farris et al., 1997) of 50 or above (based on 1000 replicates) are given above nodes (run time = 949 sec.). Because of its size, the tree has been broken into four parts (1A, 1B, 1C, and 1D).

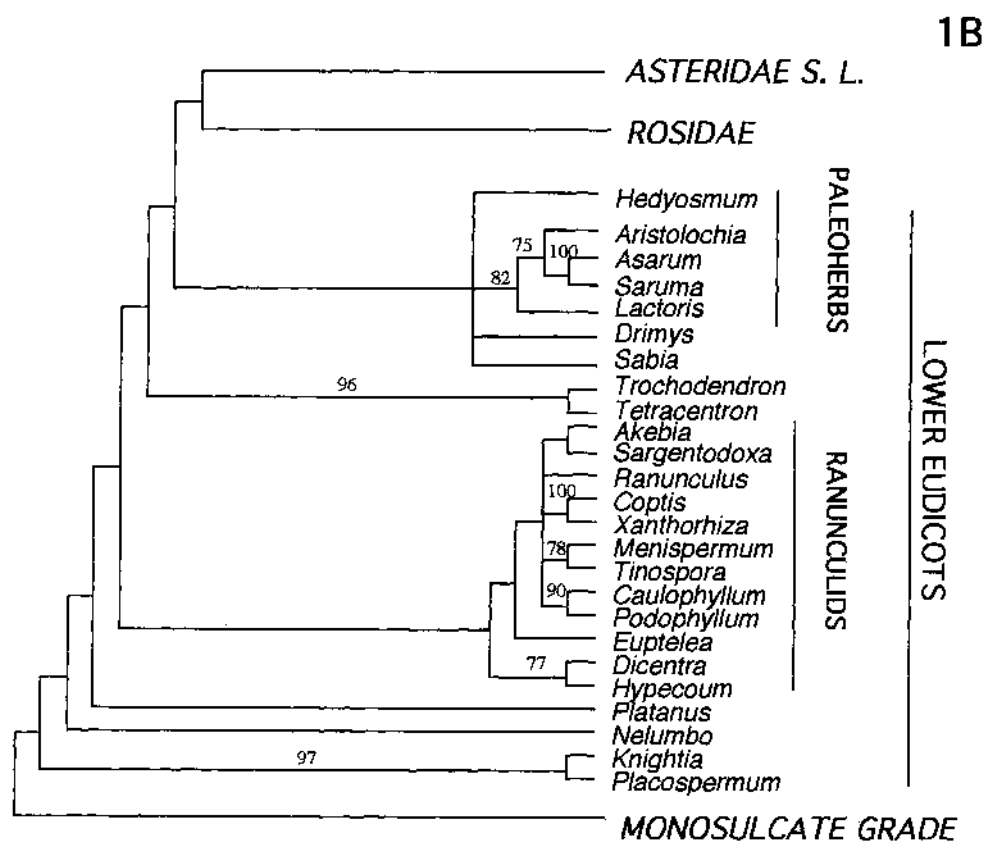


Figure 1B.

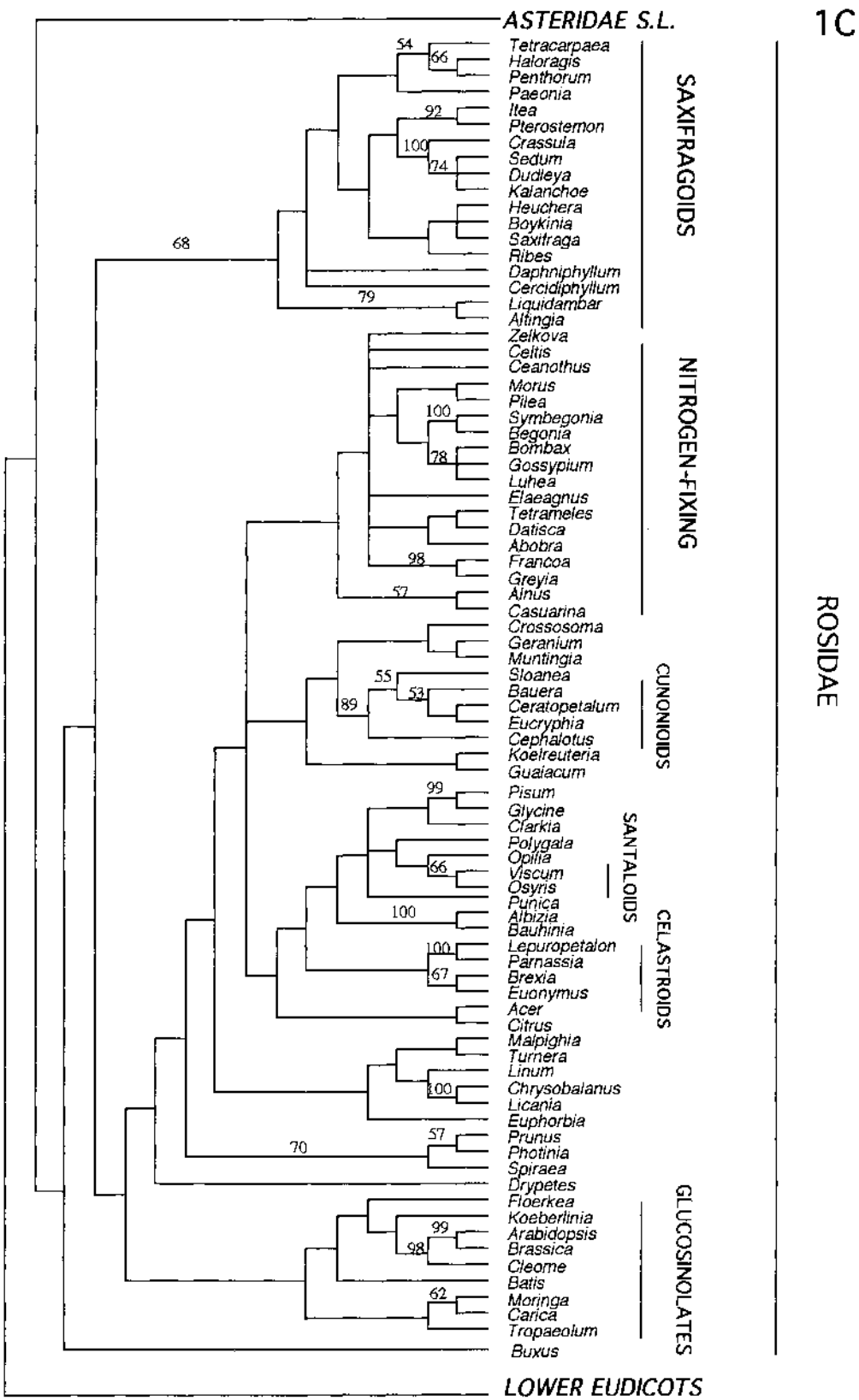


Figure 1C.

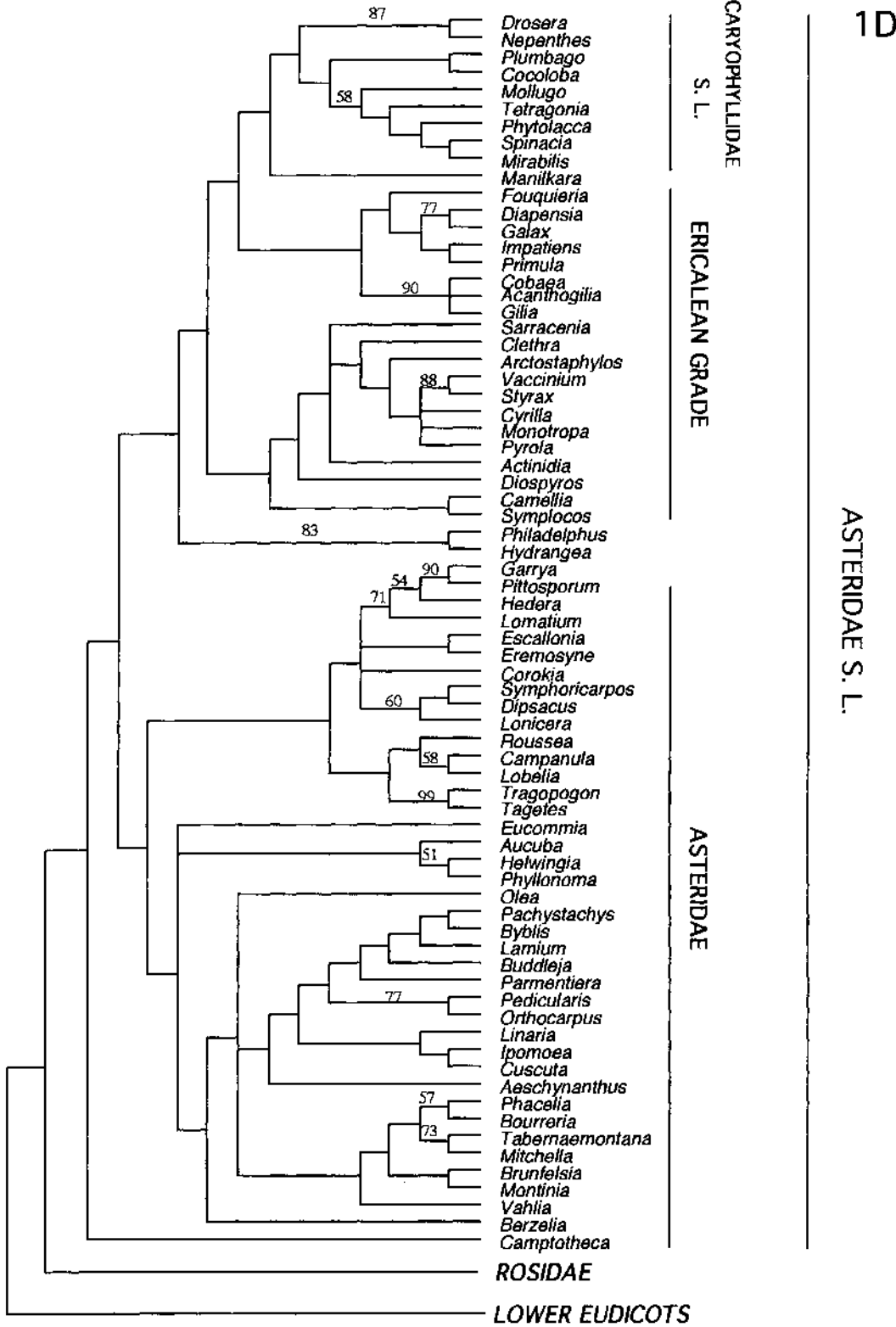


Figure 1D.

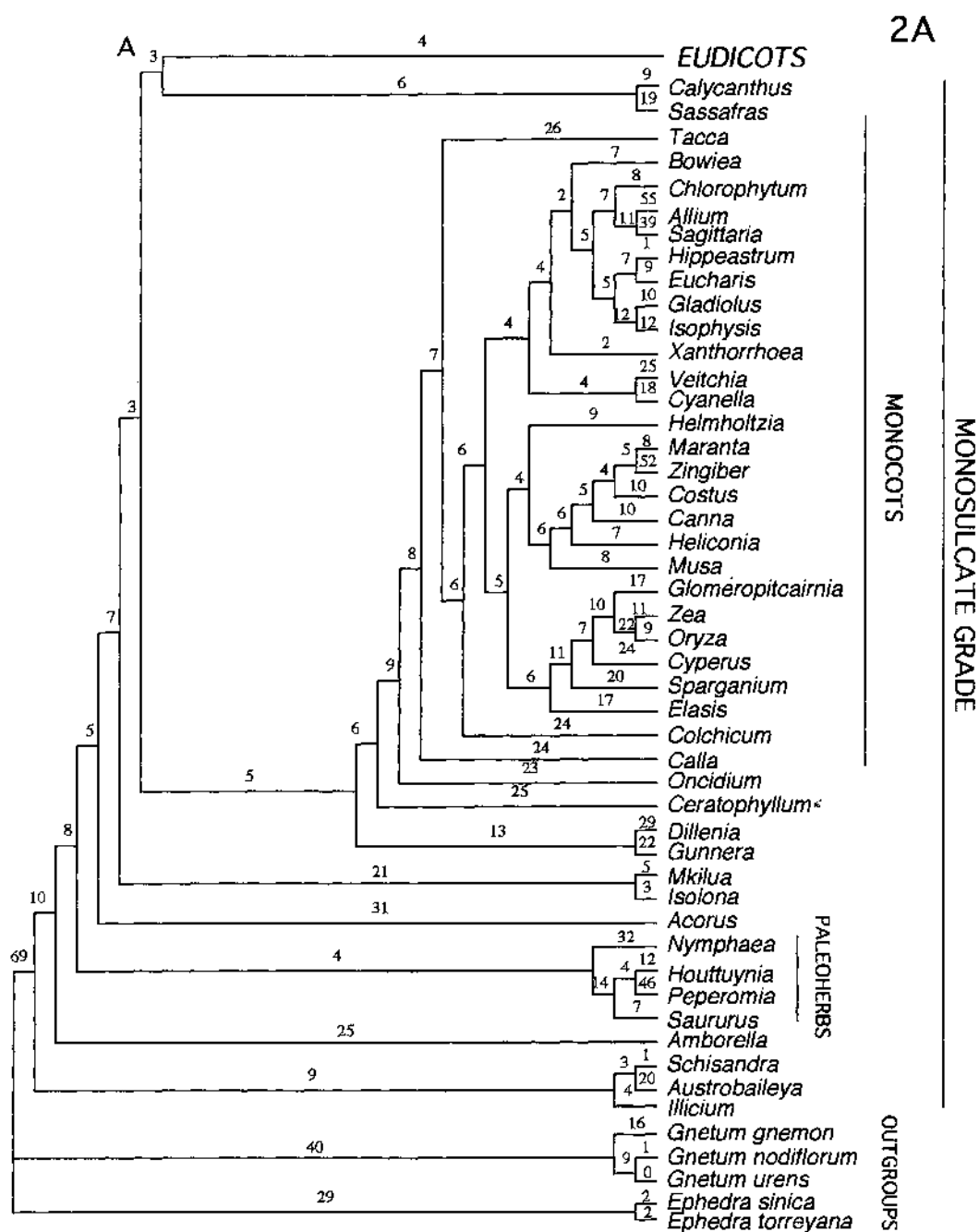


Figure 2. One of 2508 shortest trees resulting from the exploratory phylogenetic analysis of 223 species of angiosperms; two indels were included in the analyses. Each of the shortest trees has a length of 3930 steps, CI = 0.235, and RI = 0.540. Arrows indicate nodes not present in the strict consensus of all shortest trees. The letters A and B indicate the occurrence of the indels described in Table 2. Because of its size, the tree has been broken into four parts (2A, 2B, 2C, and 2D).

2B

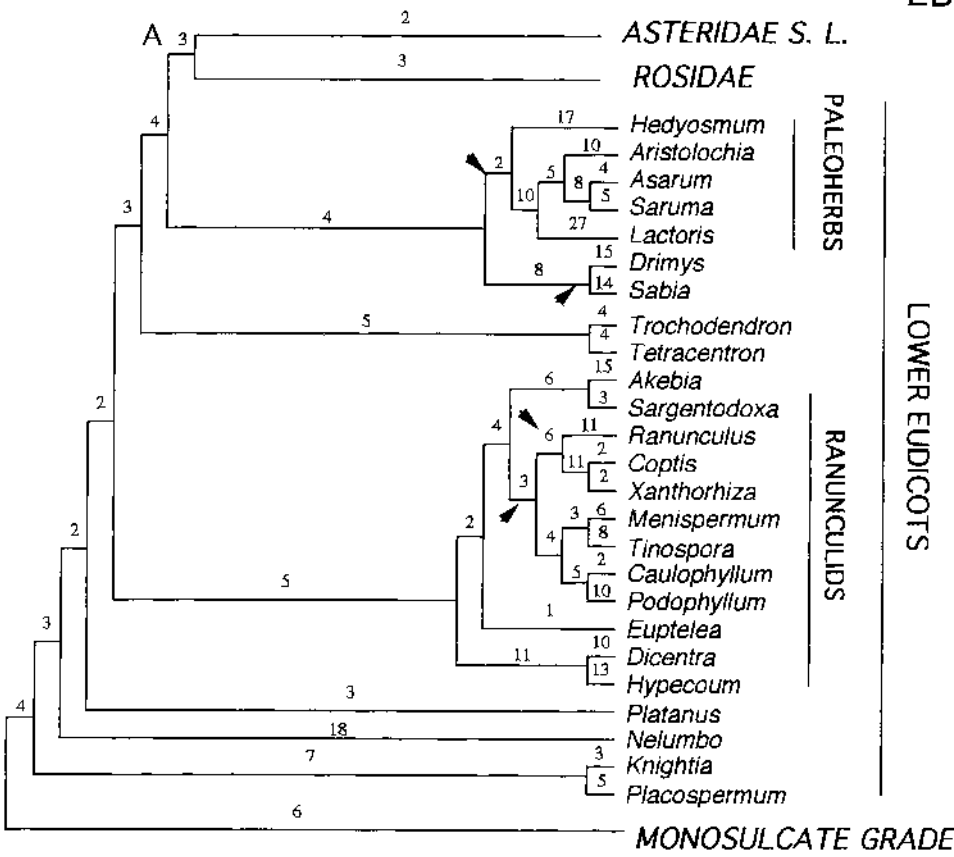


Figure 2B.



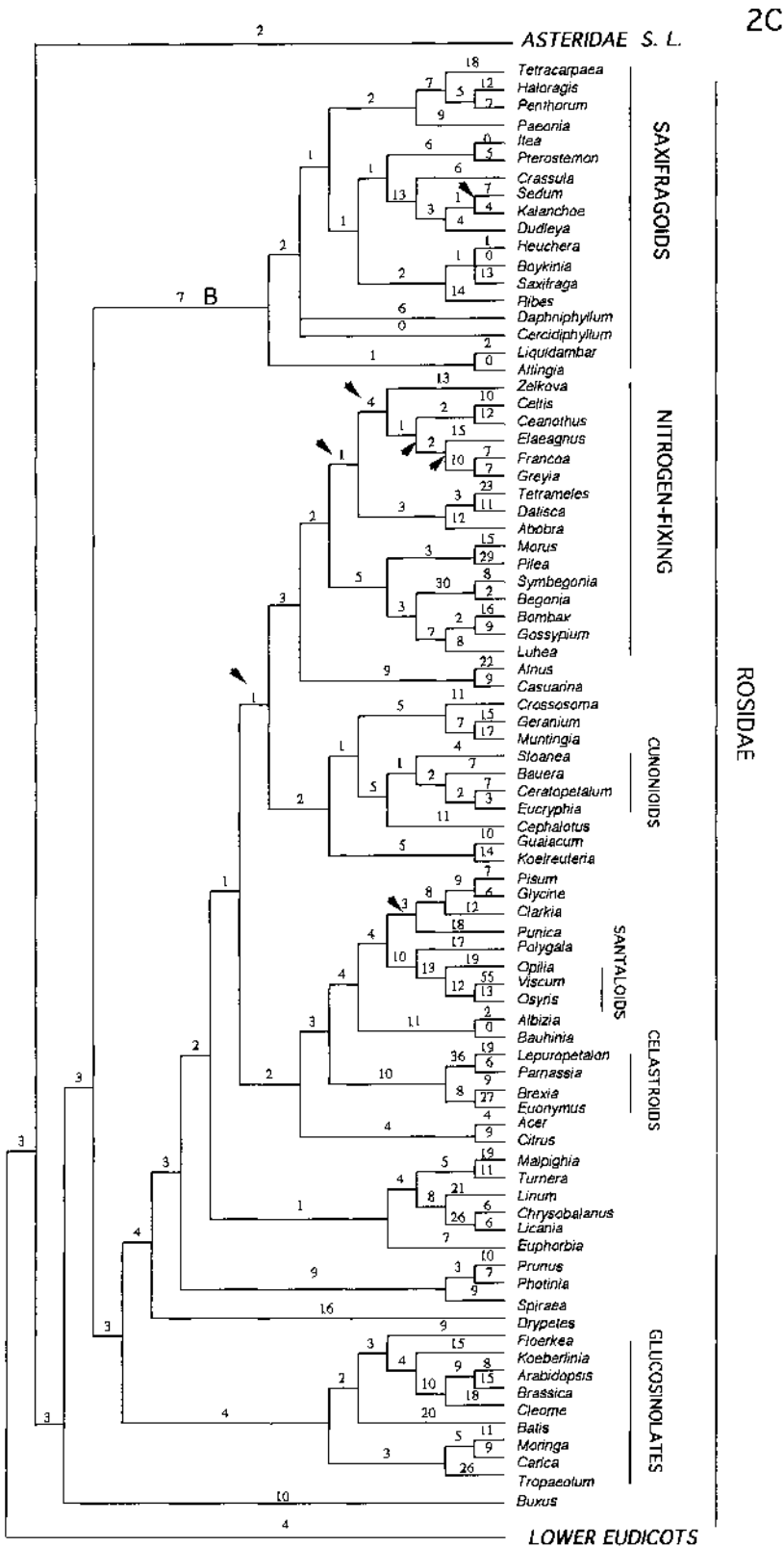


Figure 2C.

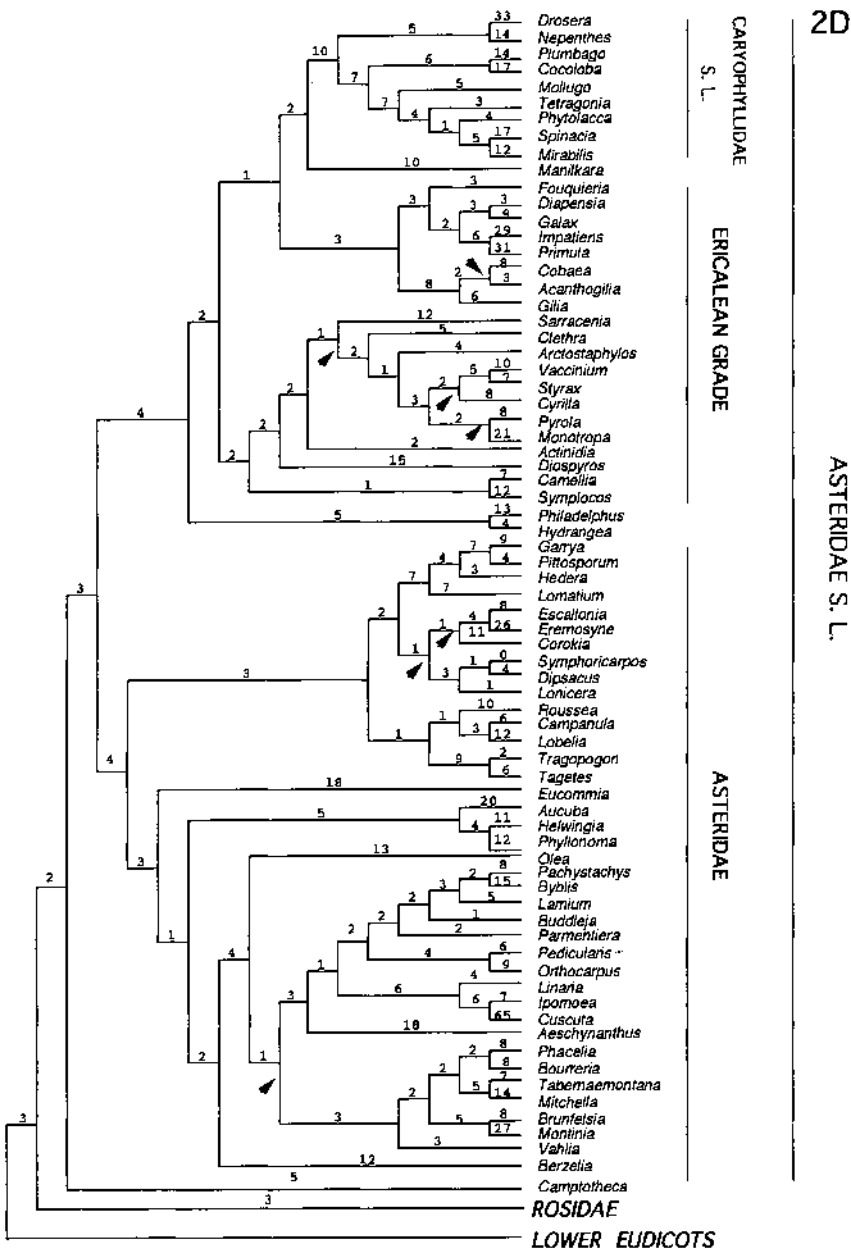


Figure 2D.



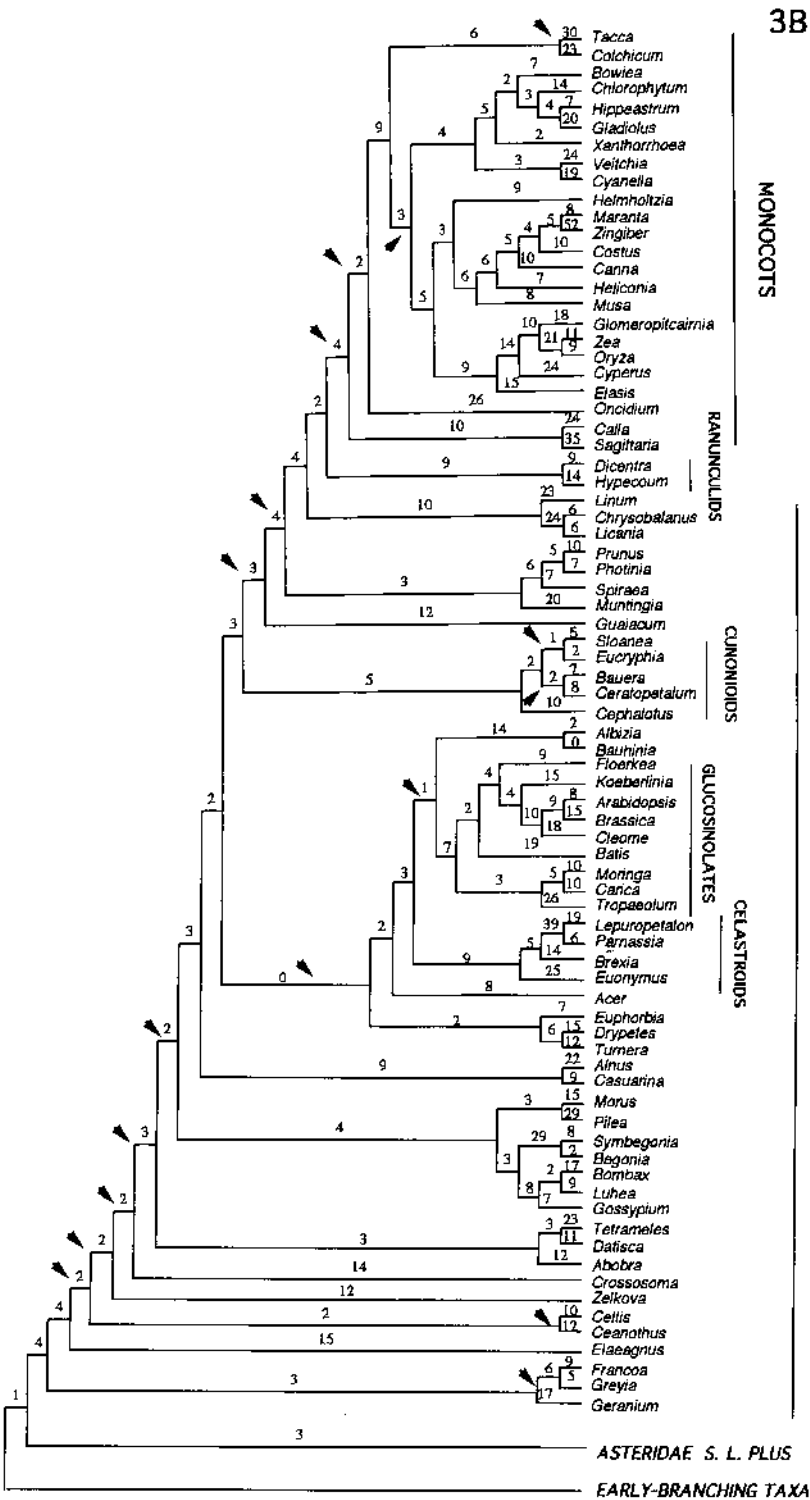


Figure 3B.



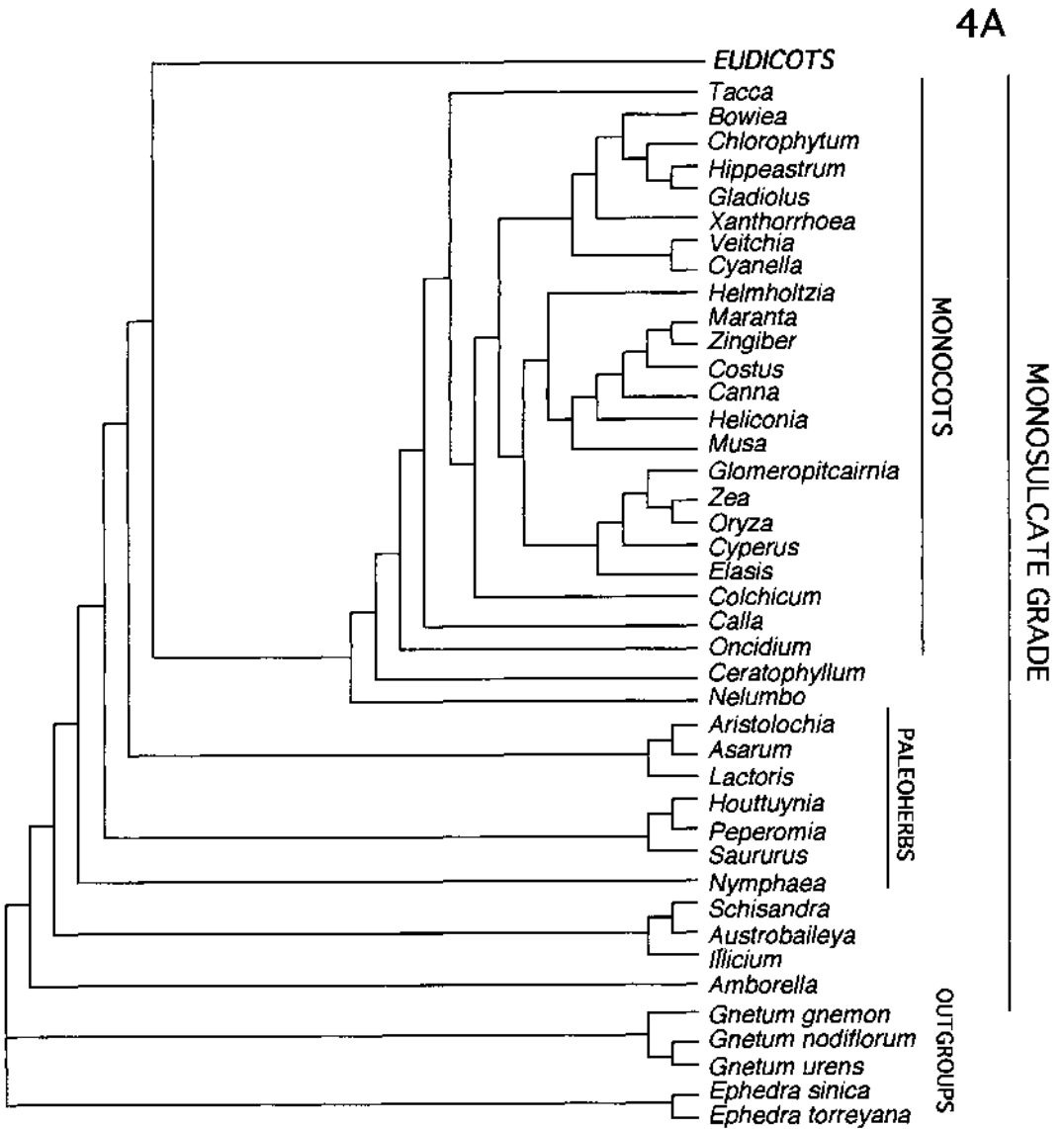


Figure 4. Strict consensus of 2582 shortest trees resulting from the exploratory phylogenetic analysis of 194 species of angiosperms; two indels were included in the analyses. Each of the shortest trees has a length of 3507 steps, CI = 0.249, and RI = 0.536. The letters A and B indicate the occurrence of the indels described in Table 2. Because of its size, the tree has been broken into three parts (4A, 4B, and 4C).

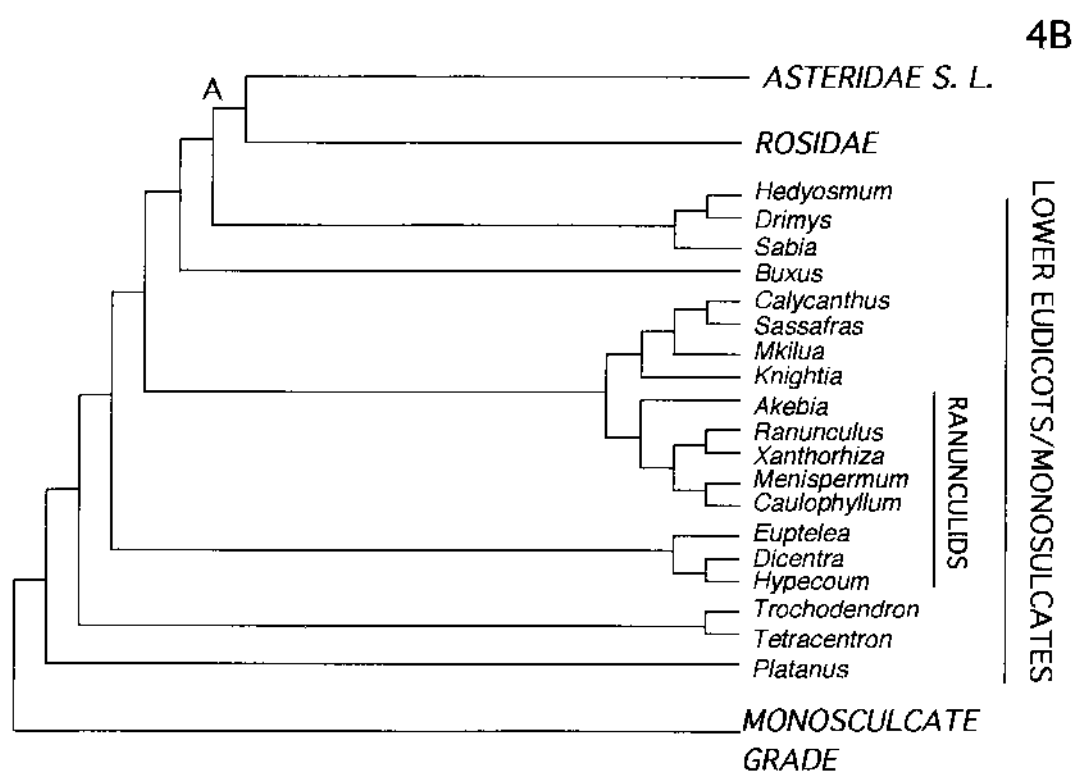


Figure 4B.

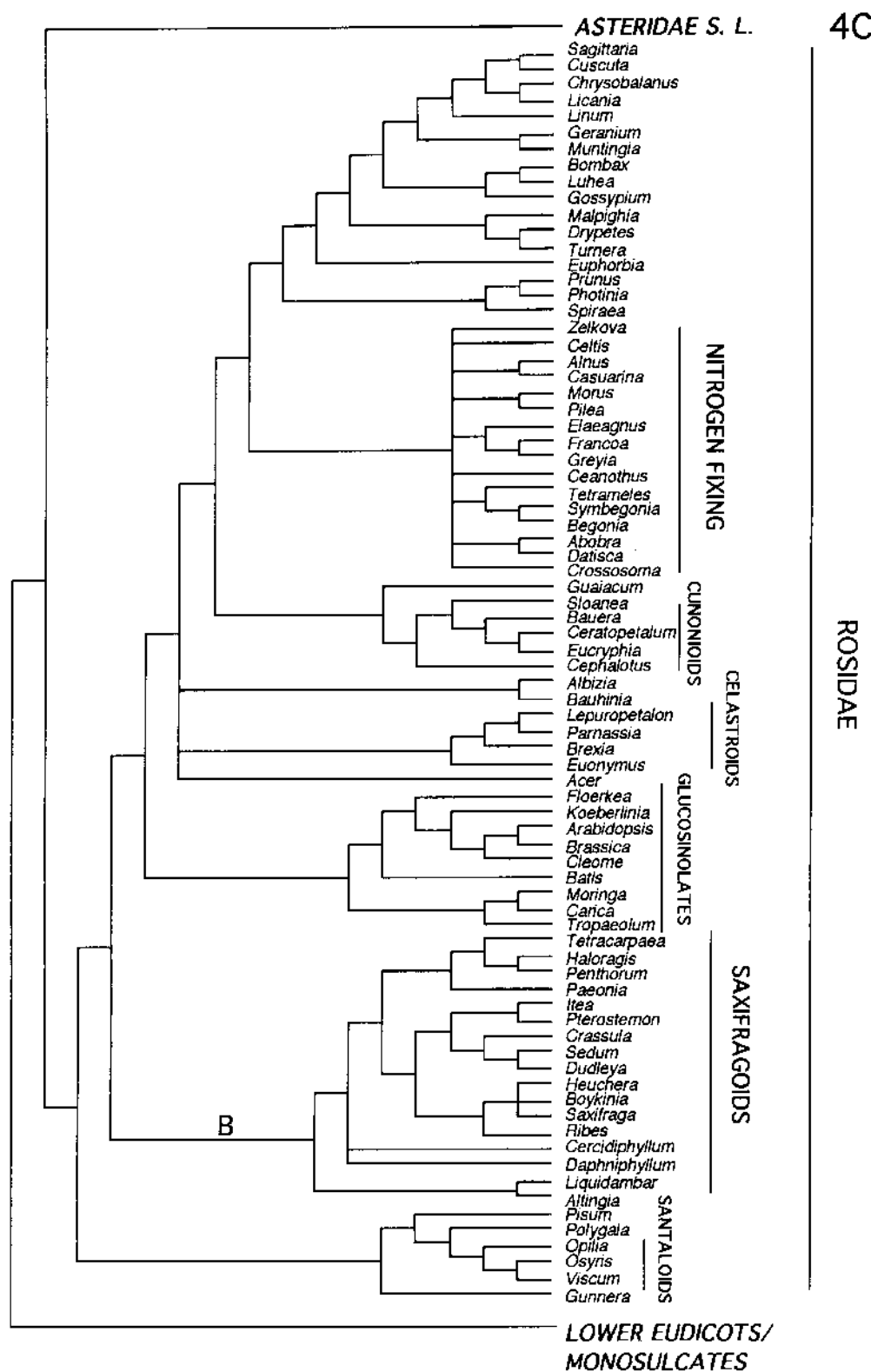


Figure 4C.



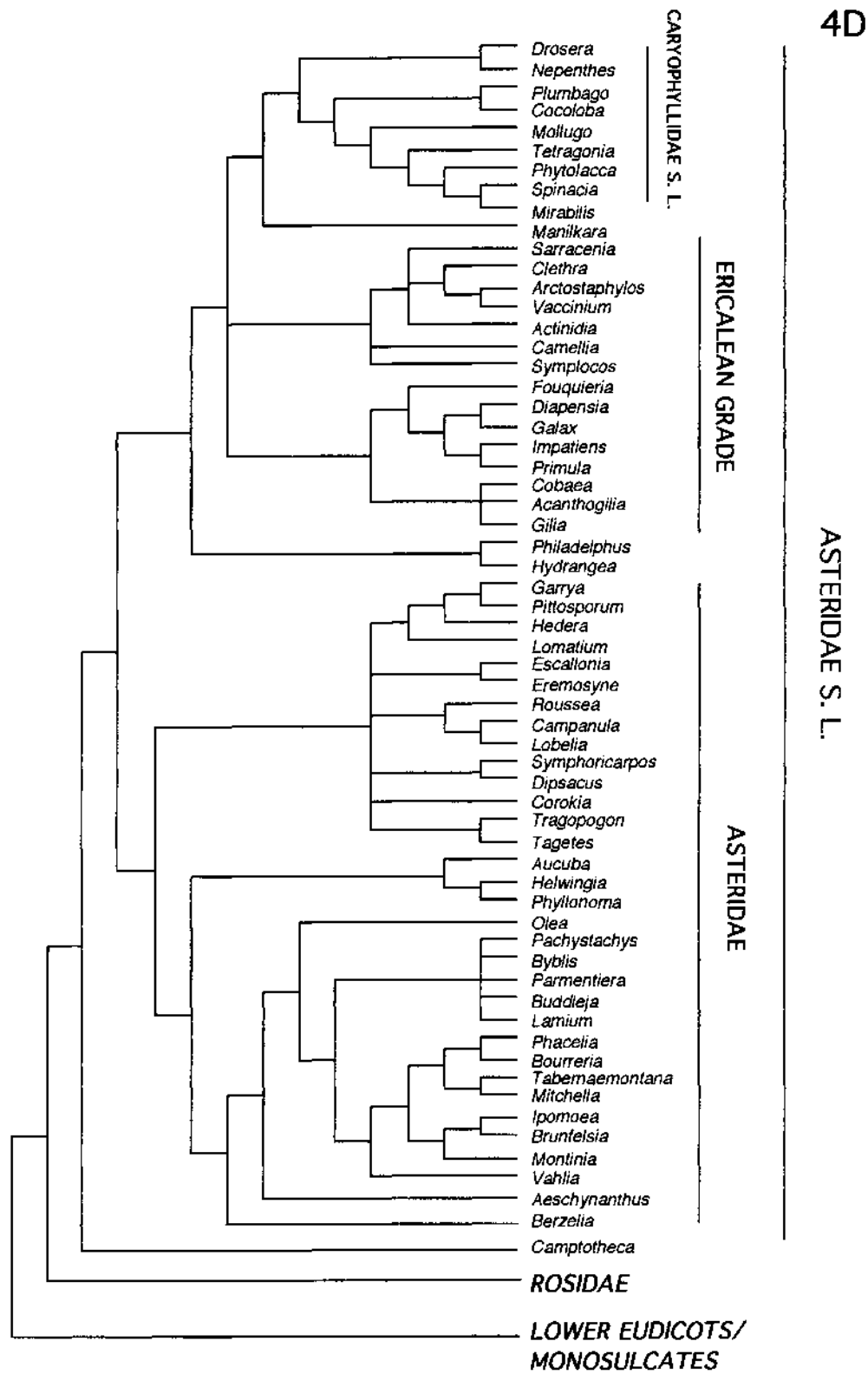


Figure 4D.

APPENDIX. Proposed secondary structure for the 18S ribosomal RNA of *Glycine max* (modified from Nickrent & Soltis, 1995). This structural model is based on the primary sequence of *Glycine* (Eckenrode et al., 1995) and follows the general models proposed for eukaryotes in general. Tertiary interactions are indicated by thick lines. The positions indicated by arrows are those regions particularly prone to variation in primary sequence and length (positions, 230–237, 496–501; 666–672, 1363–1369); these regions are difficult to align over a broad taxonomic scale and were not included in our phylogenetic analyses (see text).

