Syst. Biol. 49(2):306–362, 2000

Phylogenetics of Flowering Plants Based on Combined Analysis of Plastid *atpB* and *rbcL* Gene Sequences

VINCENT SAVOLAINEN,[1,2,8] MARK W. CHASE,[1] SARA B. HOOT,[3] CYNTHIA M. MORTON,[1,4,9] DOUGLAS E. SOLTIS,[5] CLEMENS BAYER,[6] MICHAEL F. FAY,[1] ANETTE Y. DE BRUIJN,[1] STUART SULLIVAN,[1] AND YIN-LONG QIU[7,10]

[1]*Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK; E-mail: v.savolainen@rbgkew.org.uk (V.S.), m.chase@rbgkew.org.uk (M.W.C.), m.fay@rbgkew.org.uk (M.F.F.)*

[2]*Conservatoire et Jardin Botaniques, Geneva 1293, and Institute of Ecology, University of Lausanne, Lausanne 1015, Switzerland*

[3]*Department of Biological Sciences, University of Wisconsin-Milwaukee, Wisconsin 53201, USA; E-mail: hoot@uwm.edu*

[4]*Department of Botany, University of Reading, Reading RG6 2AS, UK*

[5]*Department of Botany, Washington State University, Pullman, Washington 99164–4238, USA; E-mail: dsoltis@mail.wsu.edu*

[6]*Institut für Allgemeine Botanik, Ohnhorststrasse 18, Hamburg 22609, Germany; E-mail: c.bayer@botanik.botanik.uni-hamburg.de*

[7]*Department of Biology, Indiana University, Bloomington, Indiana 47405, USA*

Abstract.—Following (1) the large-scale molecular phylogeny of seed plants based on plastid *rbcL* gene sequences (published in 1993 by Chase et al., Ann. Missouri Bot. Gard. 80:528–580) and (2) the 18S nuclear phylogeny of flowering plants (published in 1997 by Soltis et al., Ann. Missouri Bot. Gard. 84:1–49), we present a phylogenetic analysis of flowering plants based on a second plastid gene, *atpB*, analyzed separately and in combination with *rbcL* sequences for 357 taxa. Despite some discrepancies, the *atpB*-based phylogenetic trees were highly congruent with those derived from the analysis of *rbcL* and 18S rDNA, and the combination of *atpB* and *rbcL* DNA sequences (comprising ~3000 base pairs) produced increased bootstrap support for many major sets of taxa. The angiosperms are divided into two major groups: noneudicots with inaperturate or uniaperturate pollen (monocots plus Laurales, Magnoliales, Piperales, Ceratophyllales, and Amborellaceae–Nymphaeaceae–Illiciaceae) and the eudicots with triaperturate pollen (particularly asterids and rosids). Based on *rbcL* alone and *atpB/rbcL* combined, the noneudicots (excluding *Ceratophyllum*) are monophyletic, whereas in the *atpB* trees they form a grade. *Ceratophyllum* is sister to the rest of angiosperms with *rbcL* alone and in the combined *atpB/rbcL* analysis, whereas with *atpB* alone, Amborellaceae, Nymphaeaceae, and Illiciaceae/Schisandraceae form a grade at the base of the angiosperms. The phylogenetic information at each codon position and the different types of substitutions (observed transitions and transversions in the trees vs. pairwise comparisons) were examined; taking into account their respective consistency and retention indices, we demonstrate that third-codon positions and transitions are the most useful characters in these phylogenetic reconstructions. This study further demonstrates that phylogenetic analysis of large matrices is feasible. [Angiosperm; *atpB;* complex phylogenies; large molecular data sets; *rbcL.*]

As we near the end of the 20th century, systematics had gained a new perspective because of the extensive and intensive use of molecular data in evolutionary studies. In 1994 Donoghue showed that at least one new phylogenetic hypothesis was being published every day, most of these being based on DNA data. Like other groups of organisms (and perhaps even more so at suprafamilial levels), flowering plants have been the subject of many phylogenetic studies. Despite this emphasis, many aspects of higher-level angiosperm phylogeny remain uncertain, including the initial branching patterns, relationships of the monocots to the dicots, and identification of the major groups of eudicots and their interrelationships (see Chase et al., 1993; Soltis et al., 1997b; Nandi et al., 1998).

[8]Address correspondence to this author, at Jodrell Laboratory.

[9]Present address: Department of Botany and Microbiology, Auburn University, Auburn, Alabama 36849–5407, USA; E-mail: mortocy@mail.auburn.edu

[10]Present address: Department of Biology, University of Massachusetts, Amherst, MA 01003-5810, USA; Email: yqiu@bio.umass.edu

In 1993, 42 plant systematists published an analysis of nucleotide sequences of the plastid gene *rbcL*, including 499 taxa representing what was hoped would be all major lineages of seed plants (Chase et al., 1993). These authors discussed many problems concerning the classification and evolution of angiosperms as well as some aspects of conducting phylogenetic analyses of large data sets. Some workers accepted most of the relationships observed in these cladograms, whereas others argued that the accuracy of these trees was unproven and that many of the relationships were too preposterous to be seriously entertained (e.g., Takhtajan, 1997). Others maintained that the analysis itself was flawed; the trees illustrated were only a few of the thousands that existed at that tree length, and still shorter trees were likely (Rice et al., 1998). Reanalyzing the 499 taxa matrix using several years of computer time, Rice et al. (1998) found shorter trees, although the major clades depicted in the 1993 paper remained unchanged.

Concomitantly, researchers using 18S rDNA expanded the coverage of angiosperms from that of earlier studies (Hamby and Zimmer, 1992; Nickrent and Soltis, 1995). This research resulted in the analysis of Soltis et al. (1997b), which included 223 taxa covering all major lineages of flowering plants. The 18S rDNA and *rbcL* topologies are highly concordant; virtually all of the same major clades and subclades are retrieved by both genes, although the branching order of major clades sometimes differs. Neither the 18S rDNA nor the *rbcL* analysis provided internal support (as estimated by the bootstrap/jackknife) for the spine of the tree. Indeed, the 1993 *rbcL* paper contained no estimates of internal support for the large analysis; it did, however, contain a smaller analysis of the eudicots, for which "decay" values (e.g., Bremer support; Bremer, 1988) were produced, and from these it was clear that little support existed for major clades within the eudicots (see Chase and Albert, 1998, for a bootstrap analysis). Chase et al. (1995) demonstrated a similar pattern of low support for major clades and higher-level relationships within the monocots.

Taxonomically equivalent data sets of 18S rDNA and *rbcL* (Soltis et al., 1997a, 1999) as well as for 18S rDNA, *rbcL*, and *atpB* (Chase and Cox, 1998; Soltis et al., 1998; Hoot et al., 1995, 1997, 1999; Hoot & Douglas, 1998) have been constructed and compared across the angiosperms and for certain large subgroups of flowering plants. Significantly, combining these data sets has resulted in a higher number of strongly supported clades and greater resolution than have analyses of the individual data sets.

Chase and Cox (1998) examined starting-tree lengths relative to the shortest trees ultimately found in parsimony searches for seven matrices each of 141 taxa for *rbcL*, *atpB*, and 18S rDNA (each used alone, in pairs, and all three combined). They found that the differences between the length of starting trees and the shortest trees ultimately obtained was greatly decreased in all combined-gene compared with single-gene matrices. This was also one of the factors responsible for the decreased analysis time for the combined data sets compared with that for the individual data sets. Parsimony searches involving combined matrices actually ran to completion, whereas none of the searches using individual matrices or the pairwise combinations was ever completed (Chase and Cox, 1998; Soltis et al., 1998).

Several studies indicated that a second plastid gene, *atpB*, would be a good candidate for comparison and combination with *rbcL* (Ritland and Clegg, 1987; Hoot et al., 1995). The rate of *atpB* evolution appeared to be similar to that of *rbcL*, and *atpB* was described as being easy to amplify and sequence with universal PCR primers (Hoot et al., 1995). Moreover, the *atpB* gene has been used successfully in phylogenetic studies at family and higher levels (Hoot et al., 1997, 1999; Hoot and Douglas, 1998; Bayer et al., 1999; Chase et al., 1999).

The genes coding for *atpB* and *rbcL* are both located in the large single-copy region of the plastid genome; their coding sequences are on opposite strands separated by an intergenic spacer of ~600–800 base pairs (bp; Savolainen et al., 1997). Different evolutionary constraints are likely to be involved because the two genes code for distinct enzymatic functions: *rbcL* codes for the

large subunit of the ribulose-1,5-bisphosphate-carboxylase/oxygenase, a free enzyme in the stroma, whereas *atpB* codes for the beta subunit of the ATP synthase, which is bound to the thylakoid-membrane (Zurawski et al., 1982). Consequently, *atpB* and *rbcL* data represent independent data sets for which comparative analyses should be suitable. Because both are part of the same nonrecombining piece of DNA, they should have the same history.

We present here the results of phylogenetic analyses using matrices for these two plastid genes, *atpB* and *rbcL* (~3,000 bp total), analyzed separately and in combination, for 357 taxa (~250 families) representing all major lineages of angiosperms indicated by the large *rbcL* and 18S rDNA studies (Chase et al., 1993; Soltis et al., 1997b). We also explore more thoroughly the substitution patterns and quality of the phylogenetic signal present in both *rbcL* and *atpB*.

Materials and Methods

Sampling

The taxa used, as well as voucher information, references, and DNA databank accession numbers, are provided in the Appendix. We tried to sequence *atpB* from the same DNA sample used previously to sequence *rbcL*. This was not always feasible, however, and in several instances we had to reextract DNAs. Whenever possible, we used the same species or another species from the same genus, but in some cases we used another genus from the same family (Appendix 1; see Kellogg and Linder, 1995, for some discussion on pitfalls of combining taxa). Our final sampling set comprises 357 species (714 gene sequences) representing 261 families (sensu Watson and Dallwitz, 1991; version 1997 is available on the web at http://www.keil.ukans.edu/delta) or 250 families according to a recent reclassification of the angiosperms that is based largely on the results obtained from molecular phylogenetic studies (Angiosperm Phylogeny Group [APG], 1998). We followed the APG treatment for familial and ordinal circumpscriptions and names.

DNA Sequencing

A standard procedure used to extract DNA and sequence *rbcL* and *atpB* for many species is described below; not all sequences were prepared in this manner (e.g., some of the *rbcL* sequences had been generated before widespread use of polymerase chain reaction [PCR]). Total DNAs were extracted from 0.2–1.0 g leaf tissue (fresh, silica gel–dried, or herbarium specimens) by using the 2 × CTAB method of Doyle and Doyle (1987) and then purified on 1.55 g ml^{-1} cesium chloride gradients. The *rbcL* gene was amplified (Gene-Amp PCR system, Perkin-Elmer 9600: 35 cycles, 1 min of denaturation at 95°C, 30 sec of annealing at 50°C, 1 min of extension at 72°C, and 7 min for final extension) by using primers 1F (5′-ATGTCACCACAAACAGAAAC-3′) and 1460R (5′-TCCTTTTAGTAAAAGATTGGG CCGAG-3′; Olmstead et al., 1992). The *atpB* gene was amplified by the same protocol as above but using primers 2F (5′-TATGAGA ATCAATCCTACTACTTCT-3′) and 1494R (5′-TCAGTACACAAAGATTTAAGGTCAT -3′; Hoot et al., 1995). Bovine serum albumin (0.40% w/v) was added to the PCR mix because it is useful for recalcitrant DNA samples, particularly those obtained from herbarium specimens (Savolainen et al., 1995). Amplification products were purified by using Magic minicolumns (Promega, Inc.) according to the manufacturer's protocols. Dideoxy cycle sequencing (26 cycles: 10 sec of denaturation at 96°C, 5 sec of annealing at 50°C, 4 min of extension at 60°C) with dye terminators was performed in 5-μl volumes directly on the cleaned PCR products. These reactions were then purified by simple precipitation. The resuspended sample was run on an Applied Biosystems Inc. 373A or 377 automated sequencer according to the manufacturer's protocols. Both strands were sequenced by using the amplification primers and the following additional internal primers: 636F (5′-GCGTTG GAGAGATCGTTTCT-3′) and 724R (5′-TCGCATGTACCTGCAGTAGC-3′ for all dicots and 5′-TCGCATGTACCYGCAGT TGC-3′ for monocots) for *rbcL*, and 611F (5′-AACGTACTCGTGAAGGAAATGATCT-3′) and 766R (5′-TAACATCTCGGAAATATTC CGCCAT-3′) for *atpB* (Hoot et al., 1995). These sequencing primers provided two 80–90% overlapping and complementary pairs of sequences. Additional *atpB* primers 40F (5′-TCCTCTTGTTCTTG GGGTTTCC-3′), 73F (5′-CAAATCATTGGYCCRGTACTGG

ATG-3'), 385R (5'-GCGCAGATCTATGAA TAGGAGACGT-3', Hoot et al., 1995) and 1186R (5'-TGTCCTGAAGTTCTTTGTAAC GTTG-3', Hoot et al., 1995) were used to sequence *atpB* from gymnosperms, which were used as the outgroups.

Phylogenetic Analysis

Sequence data were analyzed by using various versions of the PAUP* package (PAUP 4.0 d50–d63, by special arrangement with David Swofford). Most-parsimonious trees were obtained through use of the following strategies: (1) For each gene separately, 100 replicates of RANDOM taxon additions were performed, using equal weights and tree bisection-reconnection (TBR) branch swapping, with only five trees held at each step (NCHUCK = 5). The trees collectively found in these 100 replicates were then used as starting trees for new searches utilizing nearest-neighbor-interchange (NNI) swapping until 3,000 trees at this length were found (MAXTREES = 3,000). (2) When both genes were analyzed together, the procedure described above was applied, but with 350 replicates of RANDOM taxon additions and holding 10 trees at each step (NCHUCK = 10); the trees collected in these replicates were then swapped on again by using TBR with only 10 trees held at each step (NCHUCK = 10) until 200 best trees were found (MAXTREES = 200). These latter trees were then finally used as starting trees in a round of NNI swapping until completion. Each of these searches typically required ~1,000 hrs of CPU time for an Ultrasparc Enterprise 3000 Sun Microsystems (5 × 250 MHz, 1 GB of RAM). Taking substantially more time, similar results were obtained with a Power Macintosh with 32 MB of RAM; searches on this machine lasted more than a month, even for the combined matrix of both genes. This basic search strategy has been commonly used with large data sets and has been shown to be reasonably efficient (see Chase et al., 1993; Soltis et al., 1997b; Chase and Cox, 1998).

Internal support was evaluated by using bootstrap resampling (Felsenstein, 1985). For each matrix (singly or in combination), 1,000 bootstrap replicates were performed, using the NNI swapping algorithm with simple addition of taxa and only 10 trees held at each step (random deletion with replacement as implemented in bootstrapping eliminates the need to perform randomization of taxon entry order; the deletion/substitution of characters alters the distance calculations, thus randomly altering the taxon-addition patterns). Each bootstrap set (1,000 replicates) typically required 6 days using a Power Macintosh (7300/166, 32 MB of RAM).

MacClade 3.04 (Maddison and Maddison, 1992) was used to evaluate codon frequencies and to calculate various statistics (i.e., consistency index [CI] and retention index [RI] for each codon position, excluding autapomorphies for the former), and to map the morphological and chemical characters of Nandi et al. (1998) onto the combined trees. MEGA 1.0.1 (Kumar et al., 1993) was used to compare the molecular evolution of *atpB* and *rbcL*. MEGA provides widely used calculations (e.g., transitions, transversions, and so forth) that we wished to compare with tree-based values. For this purpose, 40 taxa were randomly chosen (the 40 first taxa in alphabetical order) and analyzed by pairwise comparisons without correction for multiple substitutions. LI93 also was used to calculate unbiased rates of synonymous, nonsynonymous, and overall substitutions, based on the same pairwise comparisons as above (Li, 1993). For comparison with distance-based calculations, we also calculated the number of inferred transitions and transversions (as well as their CIs and RIs), using a step matrix in PAUP* in which transversions were weighted "1" and transitions "0." From the number of transversions and their collective CI and RI, we calculated the number of transitions and their CI and RI.

We calculated the CI and RI for transversions, transitions, and each codon position to evaluate the hypothesis that frequency is a valid basis on which to implement relative weighting (i.e., that rarer events would be more reliable in phylogeny reconstruction than more frequent ones). We used RI as the criterion to discuss relative information content of different character types (Farris, 1989). CI measures overall homoplasy, whereas RI takes into account the maximum number of changes that could have occurred on an unresolved bush and

thus measures amount of structure (or phylogenetic signal) retained on optimal trees (Farris, 1989).

To compare the distribution of starting tree lengths with the length of optimal trees, we followed the same procedure as Chase and Cox (1998): The lengths of 5,000 starting trees (stepwise addition with random taxa order, no swapping, as implemented in PAUP*) were recorded for each data set (*atpB* and *rbcL* alone and in combination) and plotted as the percentage by which they were longer than most-parsimonious trees (i.e., those we found in each of the full heuristic searches after extensive swapping).

RESULTS

Molecular Evolution and Patterns of Change

Sequences of both genes were aligned easily by eye; the combined matrix contained 1,408 characters for *rbcL* (1,428 bp minus the length of the 1F primer) and 1,447 for *atpB* (1,497 bp minus the length of the 2F and 1494R primers). There were no insertions or deletions (except for some rare exceptions in *atpB;* these insertions were simply omitted from the matrix because they occurred only in single taxa at the 3′ end of the gene). The number of variable and potentially parsimony-informative characters found in each data set, the number and length of the most-parsimonious trees obtained, and the CIs and RIs for each of the analyses performed are presented in Table 1. Data sets are available at the Systematic Biology Web site (www.utexas.edu/ftp/depts/systbiol/).

Branch lengths, CIs, and RIs are shown for each partition (genes alone or in combination) by codon position (Fig. 1); most of the variation (75% and 71% for *atpB* and *rbcL*, respectively) is at third positions as previously reported (Chase et al., 1995; Hoot et al., 1995). Second positions experienced the fewest steps (10% in both genes), and first positions were intermediate (15% and 19% for *atpB* and *rbcL*, respectively). Whereas CIs were greater at the second positions (0.28 and 0.29, respectively) and less at third positions (0.11 for both), RIs were similar for the first (0.48 and 0.46) and second (0.46 and 0.44) positions, with the highest values at the third positions (0.75 and 0.59, respectively). In contrast, Naylor and Brown (1997, 1998) found RI was lowest for third-position sites in animal mitochondrial genes. Using computer-generated data sets, Hauser and Boyajian (1997) demonstrated that RI was primarily influenced by the percentage of characters that change per node but was only weakly correlated with the number of taxa. Here, in the *atpB/rbcL* tree, the average rate of change (number of steps/number of variable sites) was 7.74 for first, 5.35 for second, and 20.09 for third-codon positions. We can thus infer that the only factor responsible for the higher RI of the third-codon positions is their distribution within the tree.

Table 2 compares the rates of synonymous, nonsynonymous, and overall substitution for *atpB* and *rbcL*. As previously reported by Hoot et al. (1995), the overall rate of substitution is slightly higher for *rbcL* than *atpB*. This contrasts with the tree length, which is shorter for *rbcL* than for *atpB* (12,772 vs. 12,979, Table 1; for contradictory results, see Hoot et al., 1995, 1999). Tree-based estimates of sequence change are therefore at odds with those calculated a priori, in the absence of a topology, by using pairwise comparisons. Synonymous versus nonsynonymous sites do not differ significantly between *rbcL* and *atpB*.

Transition/transversion ratios (ts/tv) for *atpB* and *rbcL* also differ. We calculated the

TABLE 1. Comparison of indices for the various trees illustrated in Figures 5 and 6 (tree length includes uninformative characters; consistency index [CI] excludes uninformative characters).

Data set	Number of variable characters	Number of informative characters	Number of trees	Length	CI	RI
atpB	1,023	787	>>3,000	12,979	0.15	0.56
rbcL	1,051	734	>>3,000	12,772	0.14	0.56
atpB/rbcL	2,074	1,521	8,600	25,936	0.14	0.56

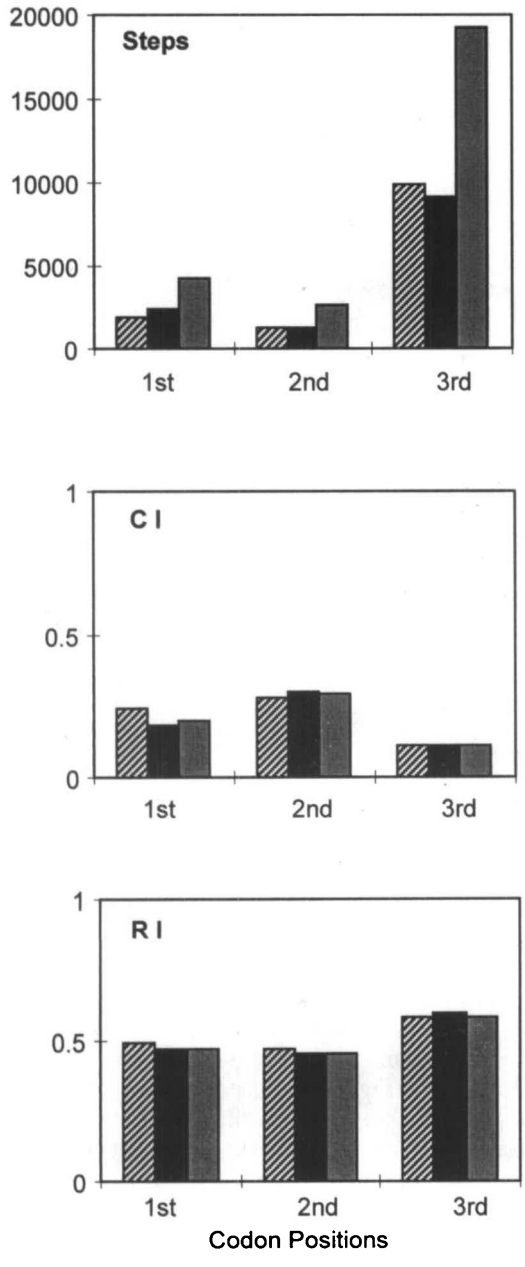

FIGURE 1. Number of steps, consistency index (CI), and retention index (RI) for each codon position for the trees inferred from the analysis of *atpB* alone (hatched), *rbcL* alone (solid), and *atpB/rbcL* combined (shaded). Note that the RI for third-codon positions is higher than that for first and second positions.

ts/tv ratio on one of the trees obtained from the analysis of the combined *rbcL* and *atpB* data set: for *atpB* ts/tv is 2.09, whereas for *rbcL* it is 1.65 (Table 3); similar ts/tv ratios were found by Hoot et al. (1995). Using instead the shortest trees from each of the

TABLE 2. Comparison of the rates of synonymous (K_S), nonsynonymous (K_A), and overall rates of substitutions (K_0) between *atpB* and *rbcL* (see Materials and Methods).

	atpB	*rbcL*
K_S	0.261 ± 0.088	0.270 ± 0.080
K_A	0.019 ± 0.007	0.025 ± 0.007
K_0	0.092 ± 0.029	0.098 ± 0.027

separate analyses affects only the second decimal place. Figure 2 shows the percentage of divergence versus the number of transitions and transversions in pairwise comparisons (independently from the recovered trees). The *atpB* gene exhibits the greater number of transitions, which agrees with the ts/tv ratio we calculated based on the trees (see Table 3).

CIs and RIs for transitions and transversions (Table 3) exhibit a pattern similar to that for codon-based change. Transitions had slightly lower CIs than transversions (ts *atpB* = 0.12 and ts *rbcL* = 0.13 vs. tv *atpB* = 0.17 and tv *rbcL* = 0.15), but the RIs for transitions were higher than those for transversions (ts *atpB* = 0.64 and ts *rbcL* = 0.62 vs. tv *atpB* = 0.47 and tv *rbcL* = 0.49). Thus, the much more numerous transitions performed better (had higher RIs) than the rarer transversions.

Phylogenetic Patterns and Support

Neither of the two genes individually has any striking differences in the number of groups receiving bootstrap support (Figs. 3–6), but the combined matrix clearly pro-

TABLE 3. Number of steps (length), consistency index (CI), and retention index (RI) for inferred transversions (tv) and transitions (ts) in one of the trees obtained from the analysis of the combined *rbcL* and *atpB* data set (see Results).

	tv/ts	tv	ts	tv/ts
atpB:				
Length	13,089	4230	8859	2.09
CI	0.15	0.17	0.12	
RI	0.56	0.48	0.64	
rbcL:				
Length	12,847	4852	7995	1.65
CI	0.14	0.15	0.13	
RI	0.56	0.49	0.62	

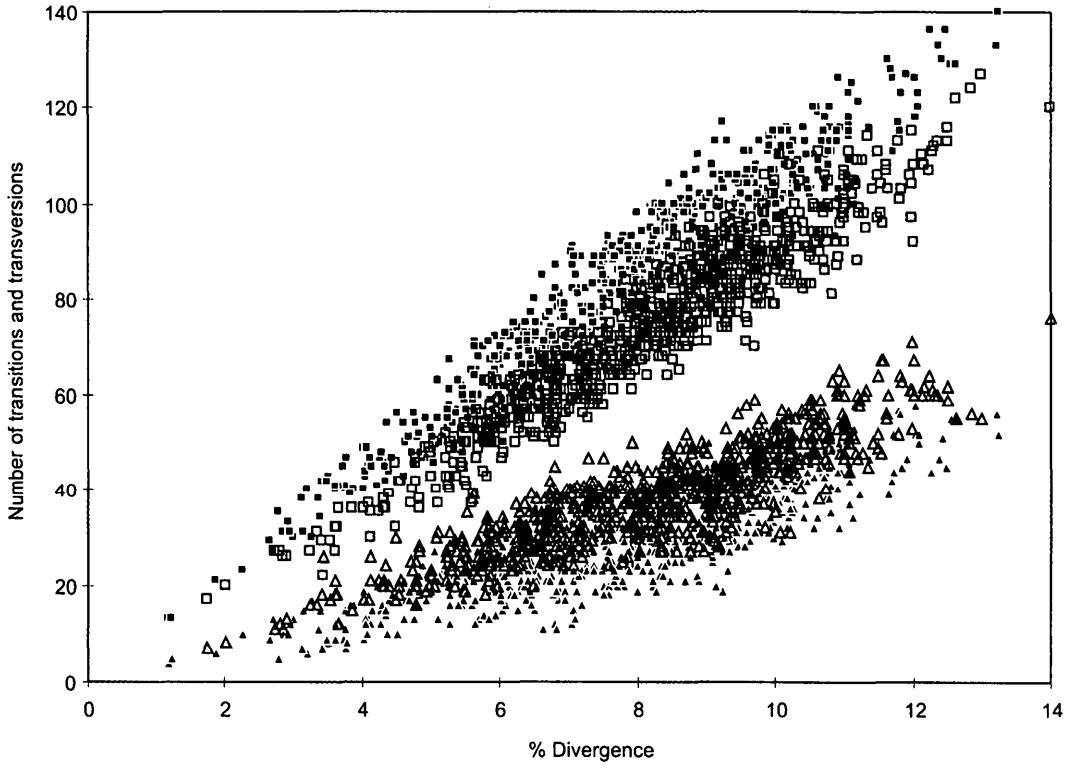

FIGURE 2. Percent divergence plotted against the number of transitions (squares) and transversions (triangles) for pairwise comparisons of *atpB* (solid) and *rbcL* (open) sequences. Note that although both genes are similar for both kinds of substitutions and no saturation is observed, *atpB* is slightly biased towards transitions, whereas *rbcL* is biased toward transversions.

vides more resolution and robust relationships than the individual matrices. Figures 3 and 4 present topologies found when both genes were analyzed separately and in combination, respectively (for composition of the named clades, see Figs. 5 and 6; the names of families and ordinal groups follow those recommended by the APG, 1998). Figures 5a–p present one of the 3,000 shortest trees found for *atpB* and *rbcL* analyzed separately (see Table 1; because of memory limitations, only 3,000 trees were kept; however, more trees at this length exist). Figures 6a–h present one of the 8,600 shortest trees found in the analysis of combined data (see Table 1 for tree statistics). In all figures, arrows indicate the branches that are not found in the strict consensus of the shortest trees.

Figure 7 shows the distribution of starting tree lengths in comparison with the length of the optimal trees for *atpB* and *rbcL* alone and in combination. The shape of the curves is not smooth but rather jagged: This could indicate that suboptima occur locally (e.g., starting *atpB* trees at length 13,143 or 13,146 are more likely to be built than at 13,144 steps; the former found 77 and 76 trees at each, respectively, vs. only 62 for 13,144; see Fig. 7). The combined data set produced the greatest number of starting trees nearest to the shortest tree length (the *atpB/rbcL* curve is left-skewed compared with *atpB* or *rbcL* alone), which explains why combined matrices ran to completion and required less swapping time to reach optima than did the analyses involving single genes (Chase and Cox, 1998; Soltis et al., 1998). Because we set a tree limit in each case (combined vs. single-genes), the faster analysis time exhibited for the combined matrix is strictly a factor of the tree lengths (starting vs. final), rather than the number of starting trees.

Apart from the above considerations of molecular evolution, *atpB* and *rbcL* also differ slightly in the phylogenetic patterns inferred. We consider the trees produced by

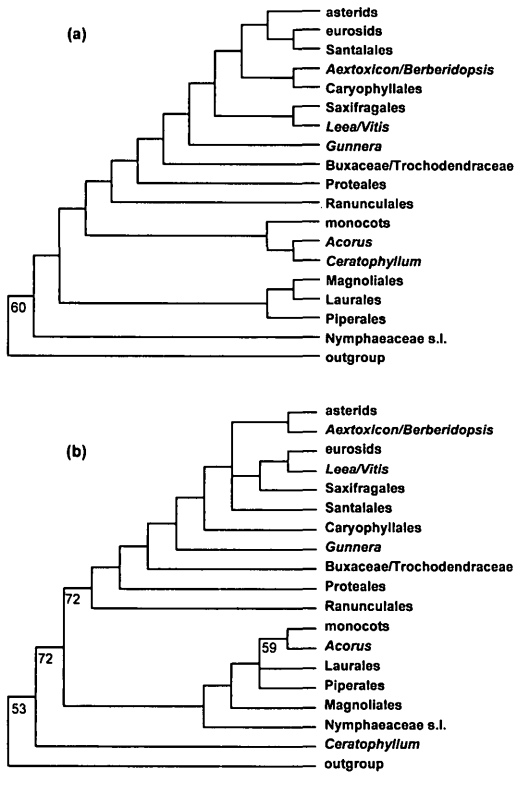

FIGURE 3. Summary of the phylogenetic trees representing only the major groupings inferred from the separate analysis of the *atpB* (a) and *rbcL* (b) coding sequences (see Fig. 5a–p for detailed topology). Bootstrap values >50% are indicated.

the combined analysis to be more accurate than trees obtained from either gene individually, given the higher levels of bootstrap support (see earlier examples presented in Hoot and Crane, 1995; Chase and Cox, 1998; Soltis et al., 1998; Hoot et al., 1999). We are not suggesting that bootstrap support is infallible—spurious groupings with high bootstrap values are well documented (see Lecointre et al., 1993)—but for Lecointre et al. this occurred concomitantly with long branch attraction and sparse taxon sampling. We will not discuss in detail the differences between the three trees (*atpB* and *rbcL* alone plus *atpB/rbcL* combined); we stress, however, that no strongly supported and incongruent patterns were obtained. The *atpB* tree compares well with the *rbcL* trees of Chase et al. (1993); only the topologies with weak bootstrap support (<50%) vary (Qiu et al., 1993; Chase and Cox, 1998; Nandi et al., 1998).

For descriptive purposes, we have divided the flowering plants into two major groups: (1) noneudicots with inaperturate or uniaperturate pollen (monocotyledons plus Laurales, Magnoliales, Piperales, Ceratophyllales, and the Amborellaceae–Nymphaeaceae–Illiciaceae group; see Figs. 5a–d, 6a,b) and (2) eudicots with triaperturate pollen (as defined by Chase et al., 1993; see Figs. 5e–p, 6c–h). The most obvious exceptions to the monosulcate pollen characteristic of the former group are Illiciaceae and Schisandraceae, which have tricolpate pollen, but the pollen of these families is known to have been derived in a nonhomologous manner (Huynh, 1976).

NONEUDICOTS

The noneudicots comprise woody (e.g., Magnoliales and Laurales), herbaceous (e.g., most Chloranthaceae and Piperales), and aquatic (e.g., Ceratophyllaceae and Nymphaeaceae) taxa. Although many studies (Soltis et al., 1997b; Nandi et al., 1998), as well as the *atpB* tree presented here, have shown the noneudicots to be paraphyletic, this informal name is useful for describing a largely monophyletic group (if *Ceratophyllum* is excluded) recognized here in the *rbcL* and combined trees. The monocots are embedded within the noneudicots in all trees (Figs. 5a,b, 6a). In this study, the monocots have not been sampled extensively but are represented by a taxonomically diverse spectrum (Figs. 5c,d, 6b). We refer readers to Duvall et al. (1993), Soreng and Davis (1998), and Chase et al. (1995, 2000) for broader analyses of the monocots. However, both *atpB* and *rbcL* provide highly similar topologies, with *Acorus* being sister to the remaining monocots (Figs. 5c,d, 6b).

A major difference between *atpB* and *rbcL* concerns the monophyly of the noneudicots (excluding *Ceratophyllum*). In the *rbcL* tree (Fig. 5b; Chase et al., 1993; Qiu et al., 1993), the noneudicots are monophyletic, as they are in the combined tree (Fig. 6b). However, with *atpB* (Fig. 5a), 18S rDNA (Soltis et al., 1997b), and 18S rDNA/*rbcL* (Soltis et al., 1997a), the noneudicots form a grade, with the root attached between *Amborella* and all other angiosperms. Because the bootstrap does not provide support >50% for either view, we argue that neither of these results

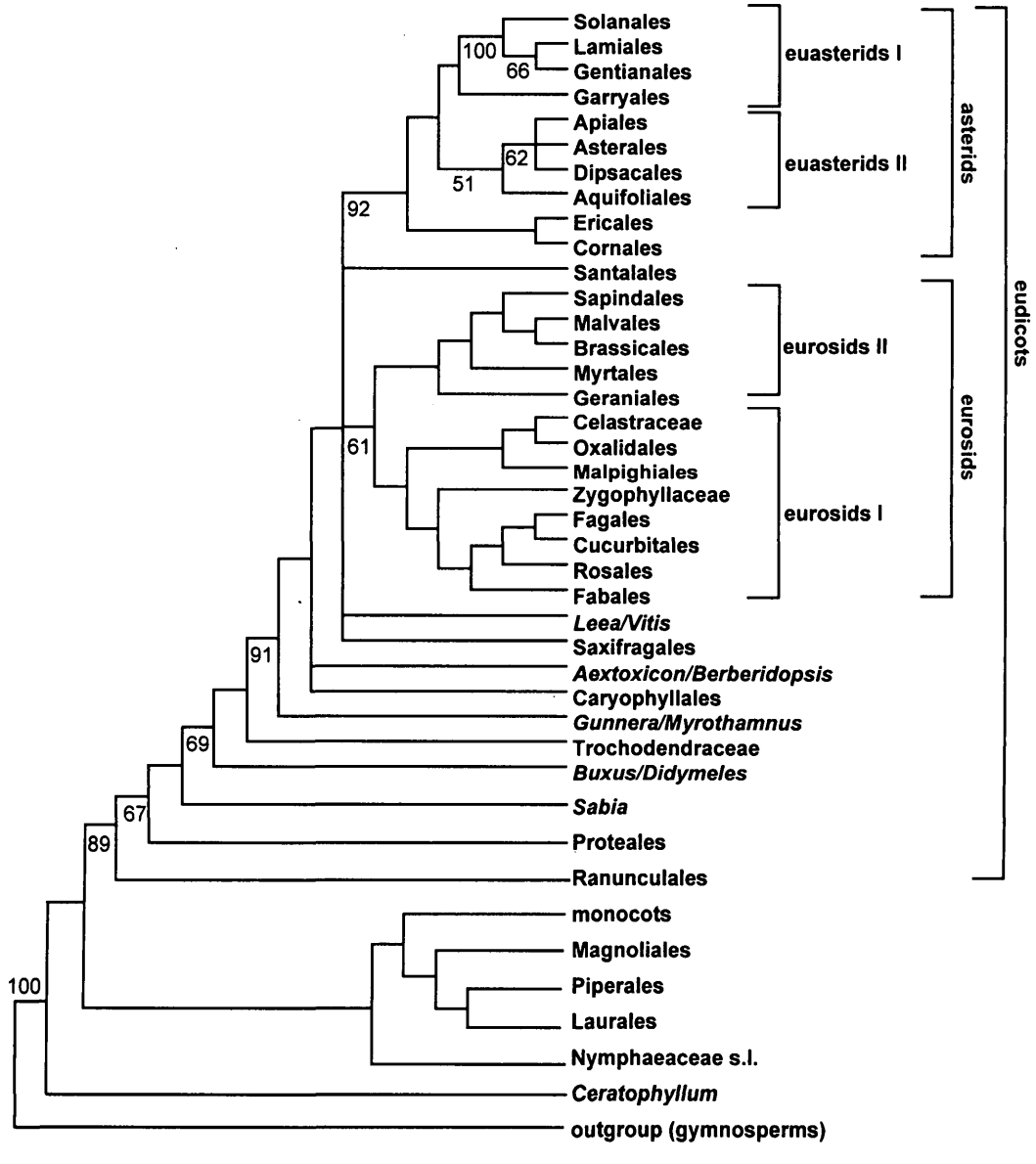

FIGURE 4. Summary of the phylogenetic trees representing only the major groupings inferred from the combined analysis of the *atpB/rbcL* coding sequences (see Fig. 6a–h for detailed topology). Bootstrap values >50% are indicated.

nor any of the previously published studies is reliable, and we await further data for a robust resolution of the problem.

Relationships within the noneudicots are consistent in all three trees, but patterns of bootstrap support are variable. Two inconsistent patterns are those for Aristolochiaceae and *Acorus*. Aristolochiaceae are paraphyletic to the rest of Piperales with *atpB* and are monophyletic and sister to the rest of Piperales with *rbcL* (Figs. 5a,b). *Acorus* (Figs. 5c,d) is sister to *Ceratophyllum* with *atpB* and is alone as the sister to the monocots with *rbcL*. In both cases, the combined tree favors the *rbcL* pattern, although there is <50% bootstrap support for the monophyly of Aristolochiaceae (Fig. 6a); support is greater for the position of *Acorus* alone as sister to the rest of the monocots: 86% for the monophyly of the monocots excluding *Ceratophyllum*, and 83% for the monophyly of the monocots minus *Acorus* (Fig. 6b).

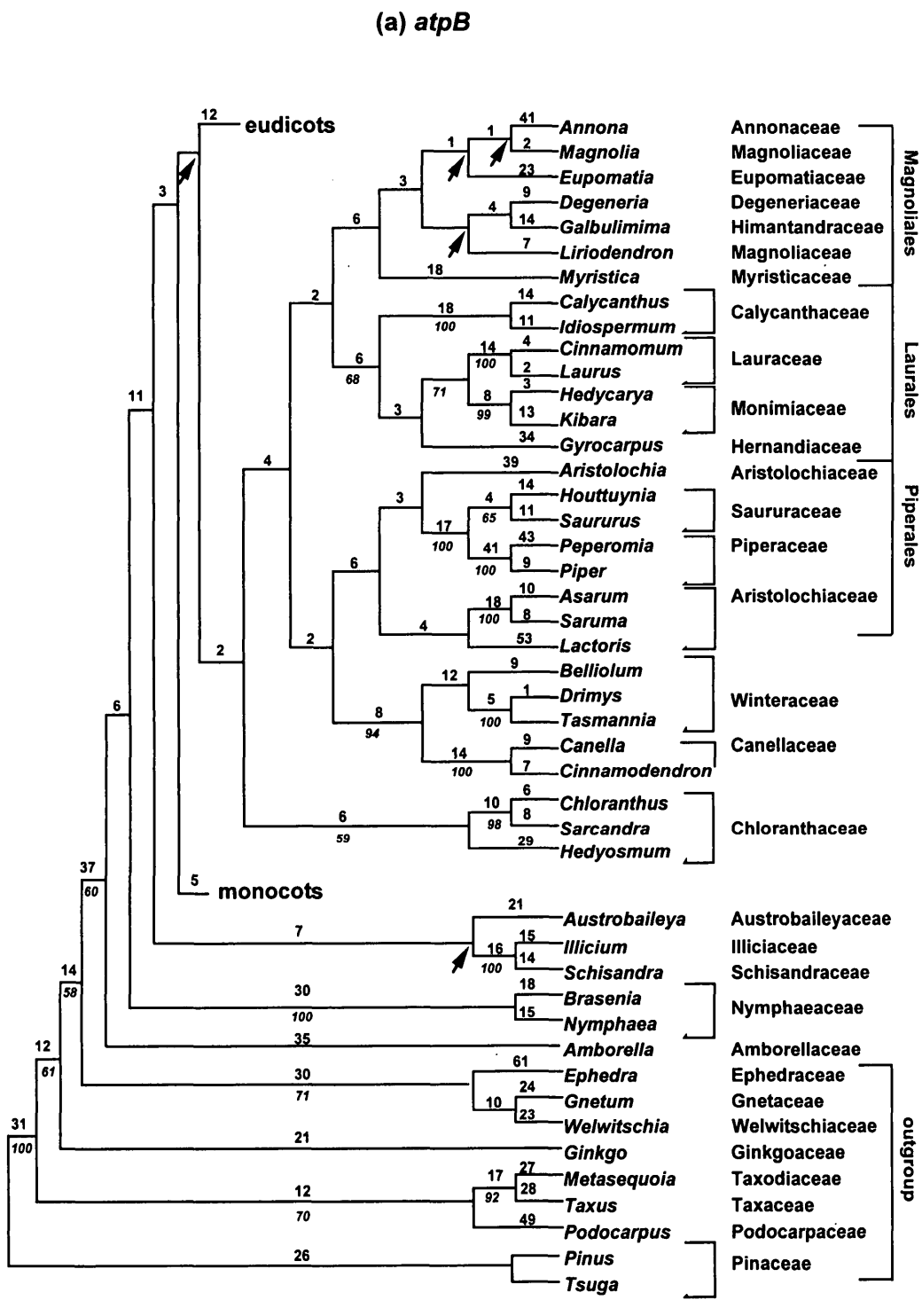

FIGURE 5. One of 3,000 best trees resulting from the exploratory phylogenetic analysis of *atpB* alone and *rbcL* alone for 357 taxa. For *rbcL*, tree length is 12,772 steps, CI = 0.14, and RI = 0.56; for *atpB*, tree length is 12,979 steps, CI = 0.15, and RI = 0.56. Arrows indicate branches collapsing in the strict consensus tree of all shortest trees. The numbers of steps are indicated above the branches, and bootstrap values >50% are indicated below the branches.

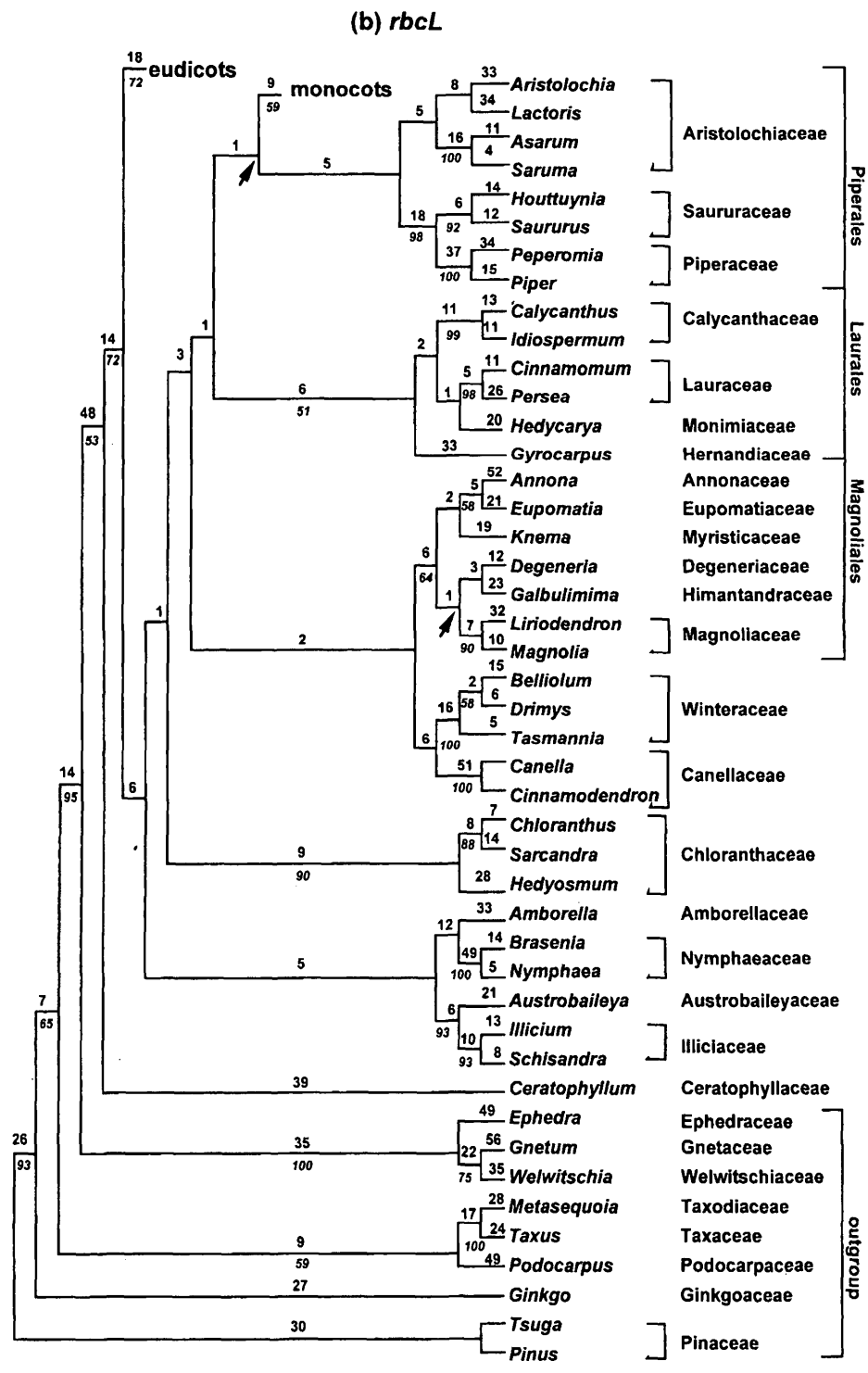

FIGURE 5. (Continued) Because of their size, the trees have been broken into eight parts each. Each *atpB* and *rbcL* tree is presented on opposite pages to facilitate comparisons: (a, b) noneudicots, (c, d) monocots, (e, f) eudicots, (g, h) caryophyllids, (i, j) eurosids I, (k, l) eurosids II, (m, n) asterids, (o, p) euasterids. Suprageneric nomenclature follows that published by the Angiosperm Phylogeny Group (1998).

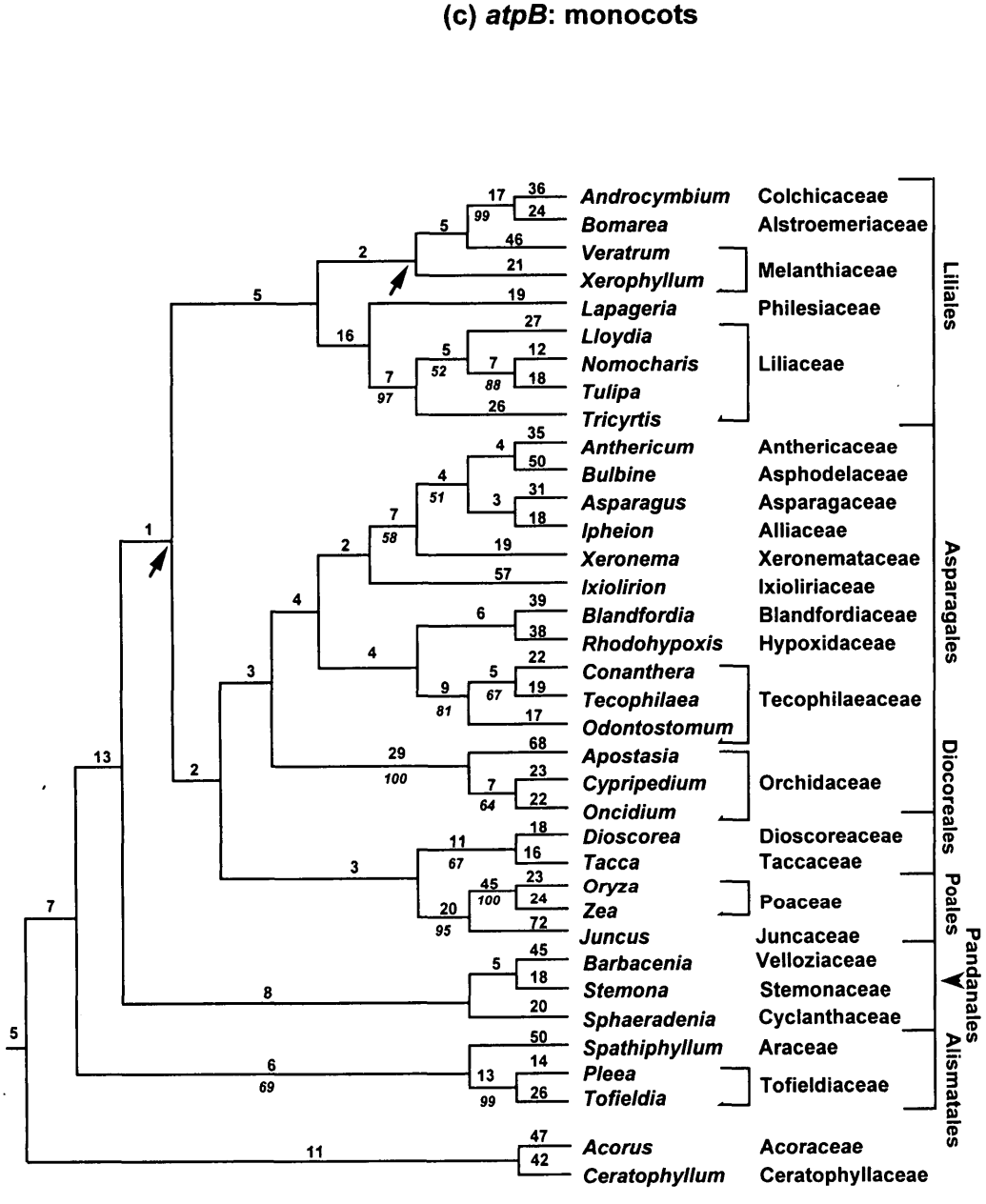

FIGURE 5. (Continued)

Support for some clades is greater with *atpB* than for *rbcL* (e.g., bootstrap support for Canellaceae/Winteraceae is 94% with *atpB* and <50% with *rbcL;* support for the monophyly of Laurales is 68% with *atpB* vs. 51% with *rbcL*). In other groups, in contrast, support is greater for *rbcL* than for *atpB* (e.g., support for the monophyly of Magnoliales is 64% with *rbcL* vs. <50% with *atpB;* support for the monophyly of Chloranthaceae is 90% with *rbcL* vs. 59% with *atpB*); in all cases, groups supported by one gene are not contradicted by different groups with strong support from the other gene.

Eudicots

The eudicots consist of (1) a basal (asymmetric) grade, comprising putatively "ancient" lineages with relatively few species

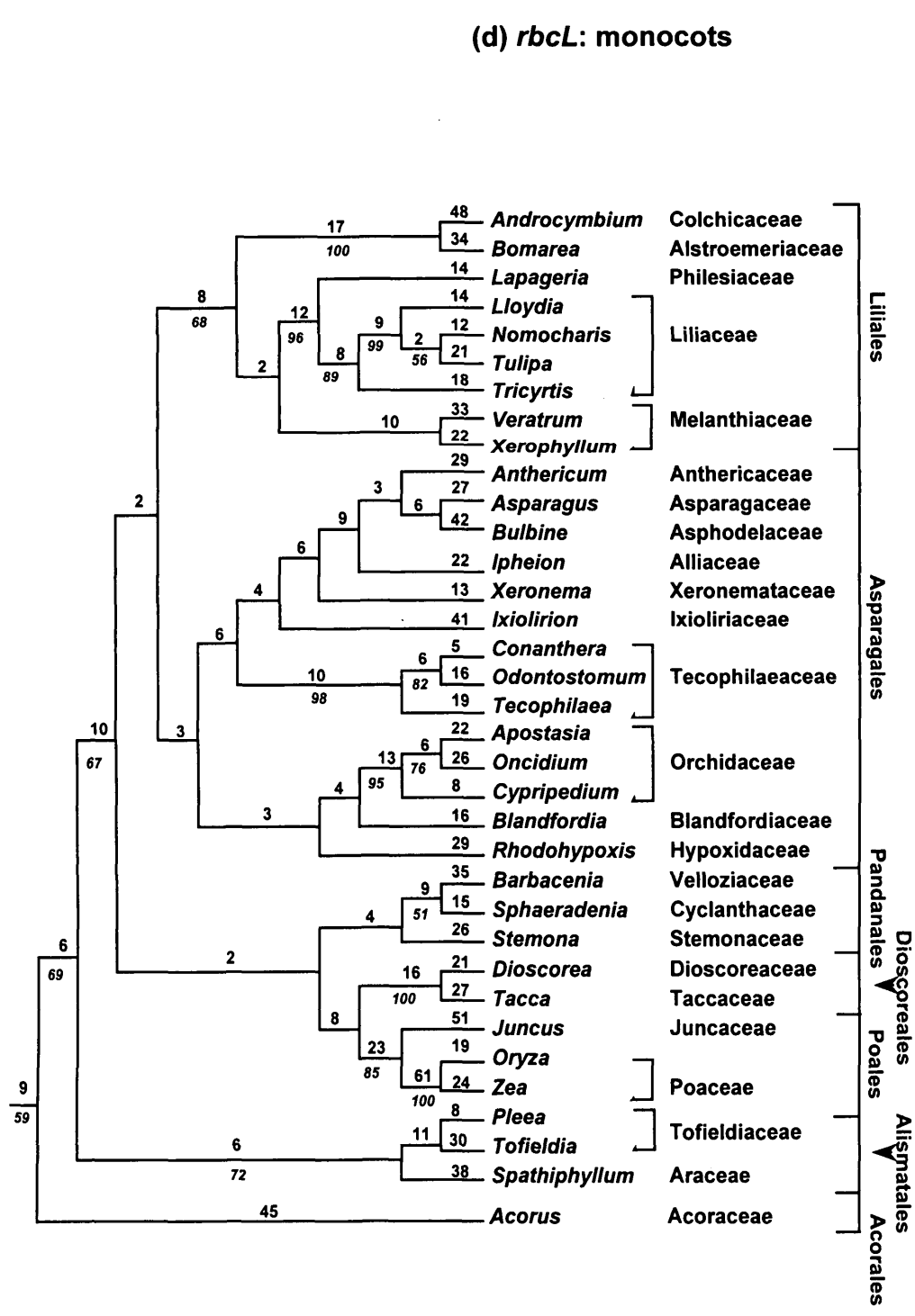

FIGURE 5. (Continued)

each, i.e., Ranunculales, Proteales, Buxaceae/Didymelaceae, Trochodendraceae, Sabiaceae; and (2) a large symmetric core clade consisting of Gunneraceae/Myrothamnaceae, Dilleniaceae, Vitaceae, Santalales, Caryophyllales, Saxifragales, and Berberidopsidaceae/Aextoxicaceae plus two major subclades, eurosids and asterids (Figs. 5i–p,

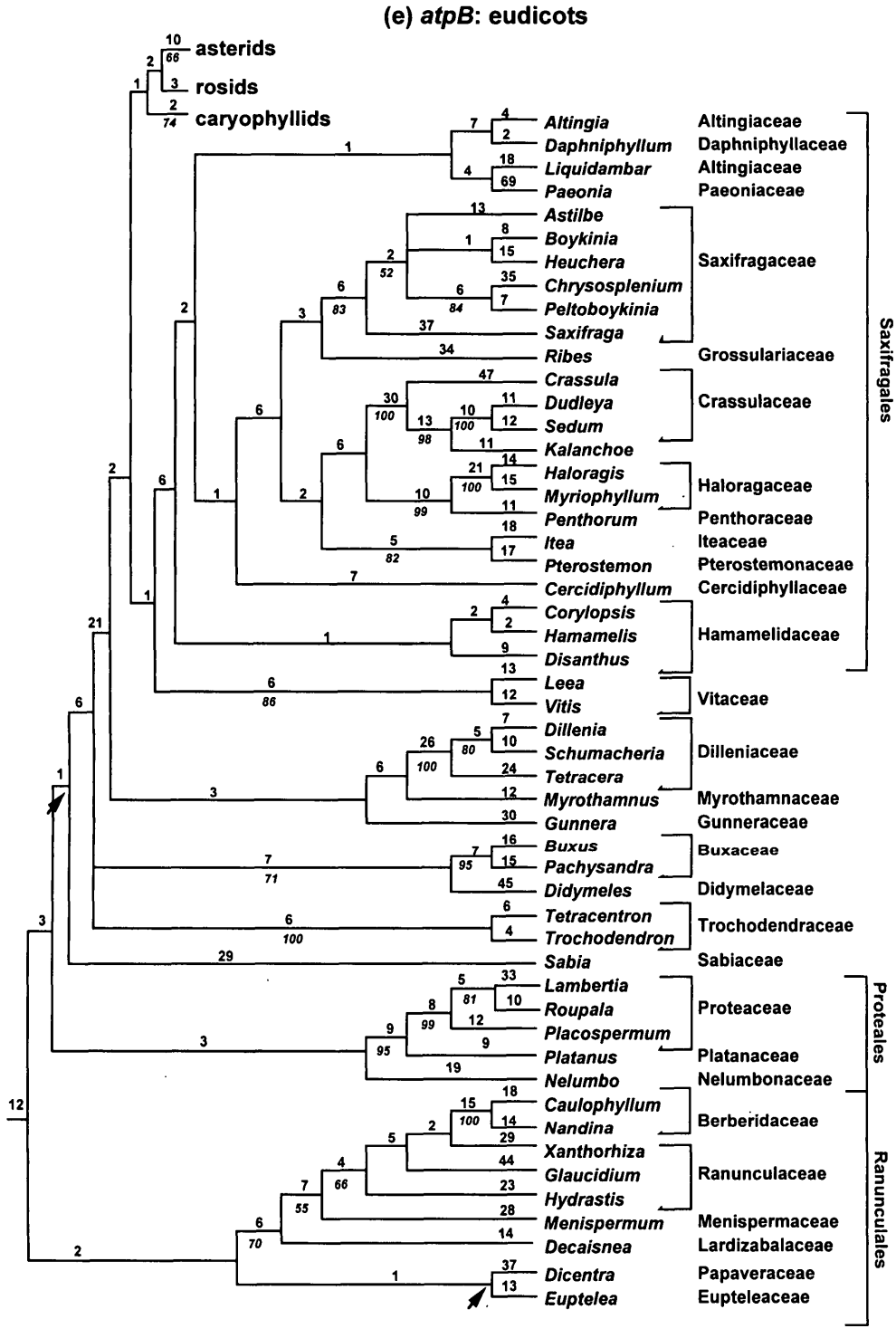

FIGURE 5. (Continued)

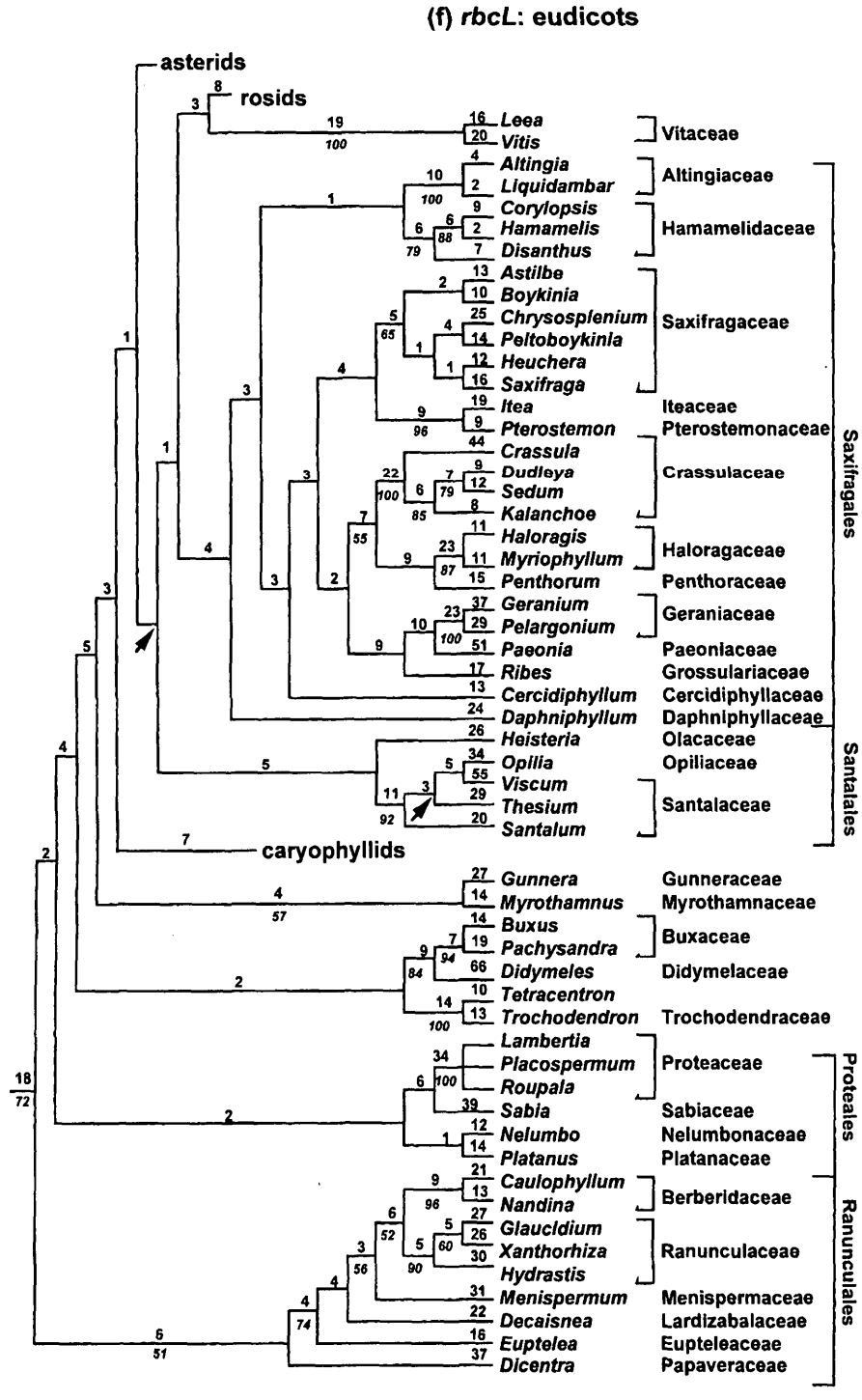

FIGURE 5. (Continued)

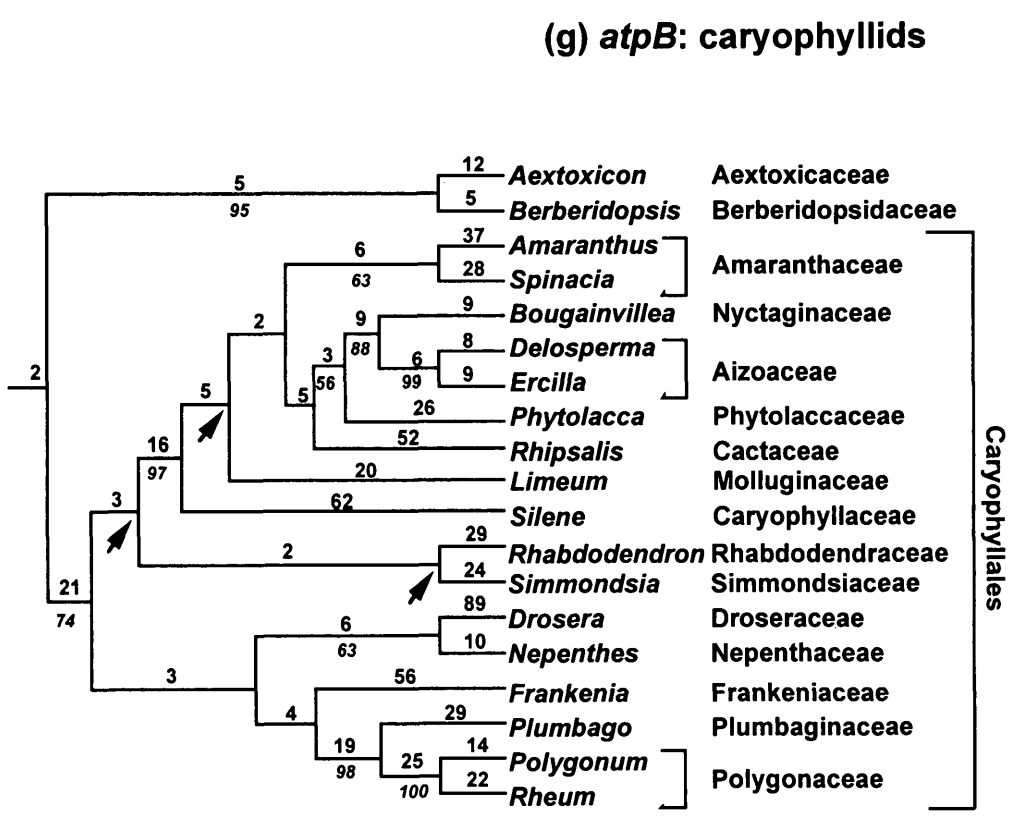

FIGURE 5. (Continued)

6e–h). The same basal grade was found in previous analyses of the "lower" eudicots based on three genes—these two plastid genes plus 18S rDNA (Soltis et al., 1998; Hoot et al., 1999).

In the separate *atpB* and *rbcL* trees, the spine of the lower eudicot portion of the tree receives no support >50%, whereas in the combined analysis three large clades are supported (Fig. 6c): (1) the dichotomy that separates Ranunculales (including Euptelea) from all other eudicots (67% for eudicots excluding Ranunculales, 94% for the monophyly of Ranunculales); (2) the eudicots excluding Ranunculales, Proteales, and Sabiaceae (69%); and (3) the core eudicots (91%, just above Trochodendraceae). Ranunculales and core eudicots are more highly supported (97% and 100%, respectively) in a study of basal eudicots with more extensive sampling for Ranunculales (Hoot et al., 1999). Other notable relationships that receive increased bootstrap support in our combined *atpB/rbcL* analysis are the monophyly of Nelumbonaceae/[Proteaceae/Platanaceae] (<50% in both *atpB* and *rbcL* alone, 60% in the combined trees), Gunneraceae/Myrothamnaceae (<50% in *atpB*, 57% in *rbcL*, 80% in the combined), and Buxaceae/Didymelaceae (71% in *atpB*, 84% in *rbcL*, and 100% in the combined).

Two groupings actually lose support in the combined tree over that in one of the individual trees: the clade consisting of Platanaceae/Proteaceae (95% in *atpB*, <50% in *rbcL*, 84% in the combined), and the monophyly of Ranunculaceae (90% in *rbcL*, <50% in *atpB*, 78% in the combined). Alternative patterns found in the individual analyses are not supported by the bootstrap, but apparently these destabilize the combined analysis, resulting in slightly decreased bootstrap percentages. Even in this situation it would be difficult to argue that this decrease in support is attributable to "hard" incongruence (Seelanan et al., 1997);

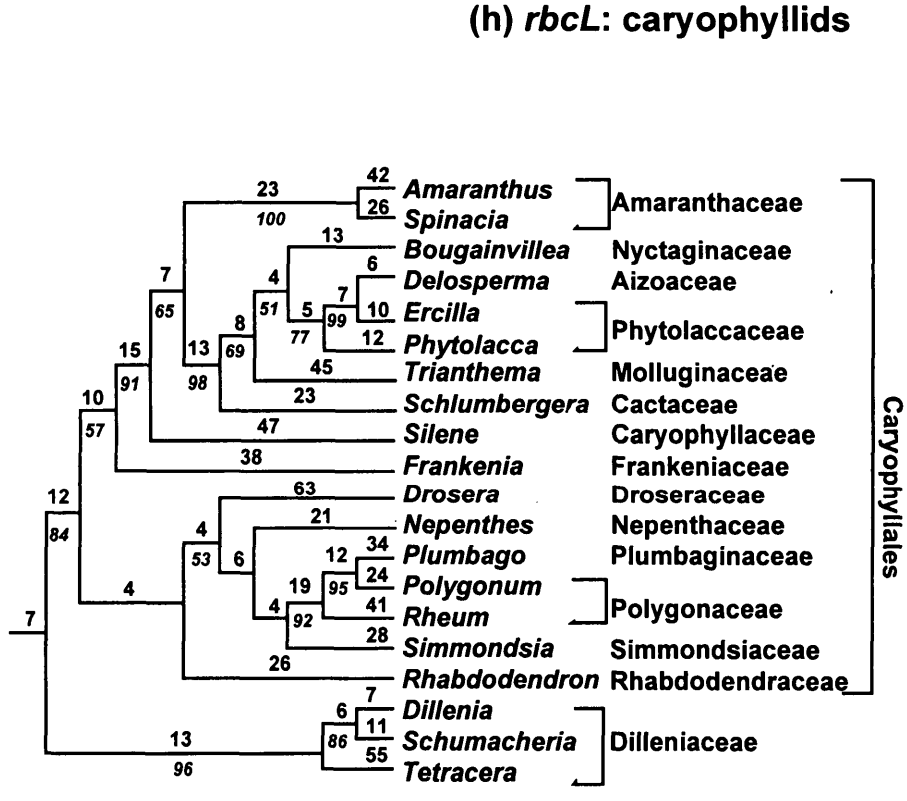

FIGURE 5. (Continued)

rather, we suggest that it is due to the simple addition of an unclear pattern in one gene to a clear one in the other, resulting in less support in the combined analysis. With three genes combined (*atpB/rbcL*/18S rDNA), these two clades, Platanaceae/Proteaceae and Ranunculaceae, are each strongly supported (Soltis et al., 1998; Hoot et al., 1999).

All three data sets provide evidence for a monophyletic Saxifragales (Figs. 5e,f, 6c), but in the *rbcL* tree, Geraniaceae are embedded within this order near Paeoniaceae. This is most likely a result of the sampling used here; in the Chase et al. (1993) *rbcL* tree with greater sampling, Geraniaceae appeared with the group here named Geraniales (Fig. 6f). The position of Paeoniaceae within Saxifragales is highly unstable and has little support for any particular placement (but see Soltis et al., 1997b, 1999; Soltis and Soltis, 1998, for relationships in Saxifragales). The relationship of Saxifragales to the other major clades is also unstable, mostly because of the short branches along the spine of the tree (Fig. 6c). As in Hoot et al. (1999), Soltis and Soltis (1998), and Soltis et al. (1999), the various members of the Hamamelidaceae represented here (*Corylopsis, Hamamelis, Disanthus*) are found in a clade consisting of Saxifragales and other assorted rosids.

Similarly, all three data sets support an expanded Caryophyllales (74% in *atpB*, 84% in *rbcL*, and 97% in the combined; Figs. 5g,h, 6d), but their placement relative to the rosids (including Saxifragales and Vitaceae) and asterids has bootstrap values <50% in all three analyses. The expanded Caryophyllales includes a core clade consisting of Amaranthaceae, Aizoaceae, Cactaceae, Caryophyllaceae, Molluginaceae, Nyctaginaceae, and Phytolaccaceae. A clade of Droseraceae, Plumbaginaceae, Polygonaceae, and Nepenthaceae is also found in all three trees but exhibits only low support in the combined analysis (67%; Fig. 6d). Many of the relationships within both core Caryophyllales and this second clade are also well supported, but additional taxa not included here are also found to be related to Caryophyllales s.l. (e.g., Physenaceae and

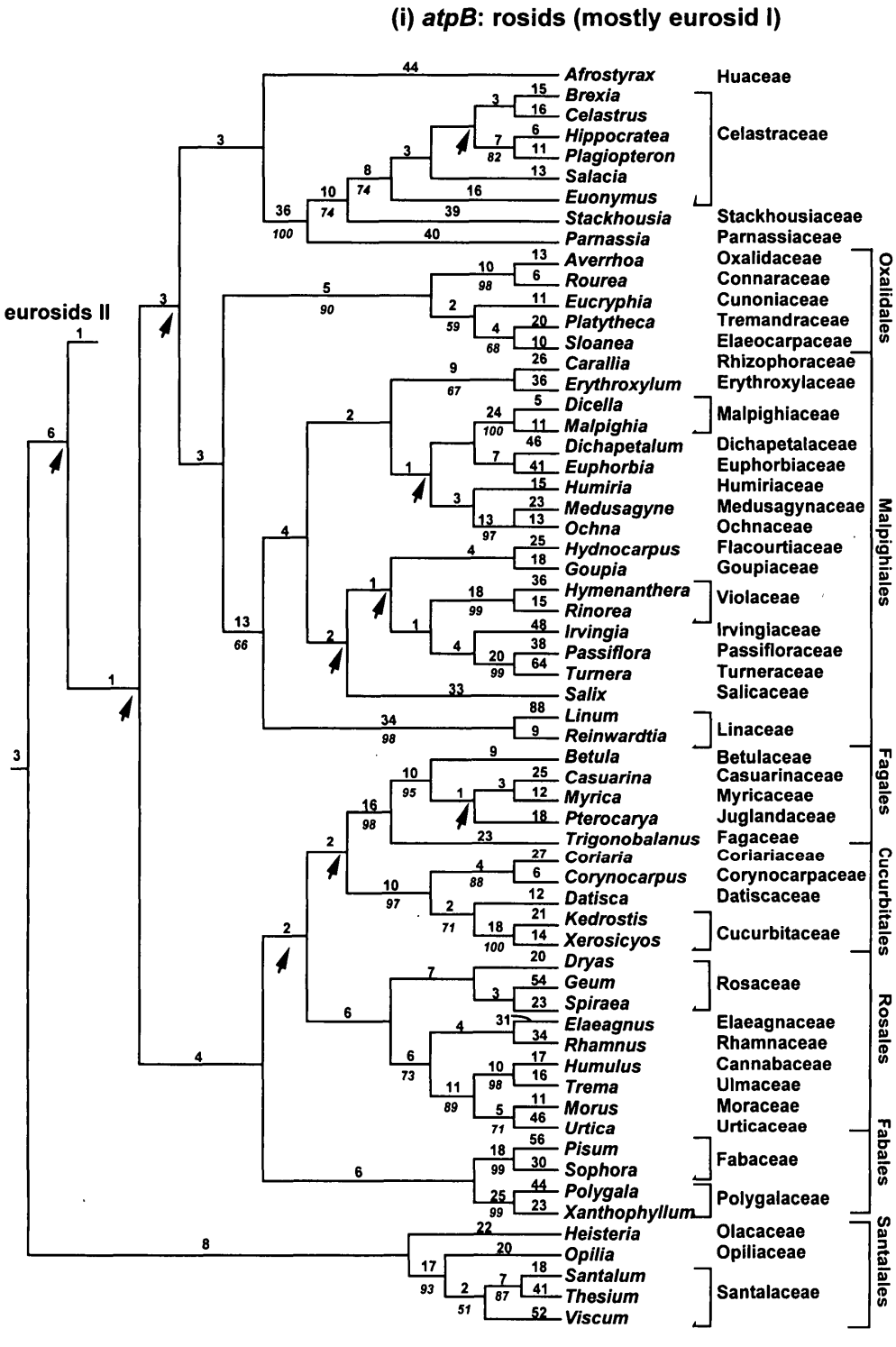

FIGURE 5. (Continued)

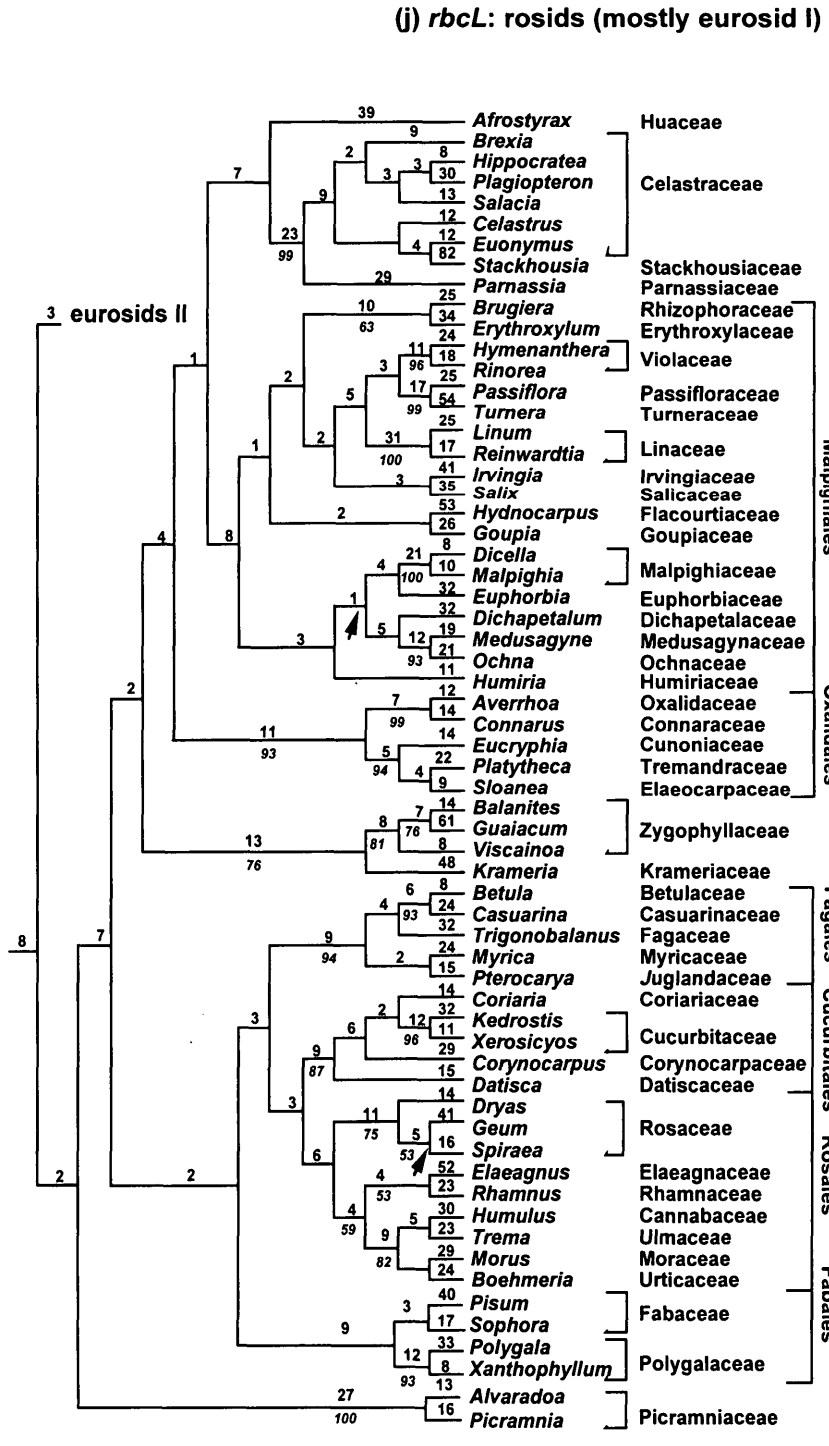

FIGURE 5. (Continued)

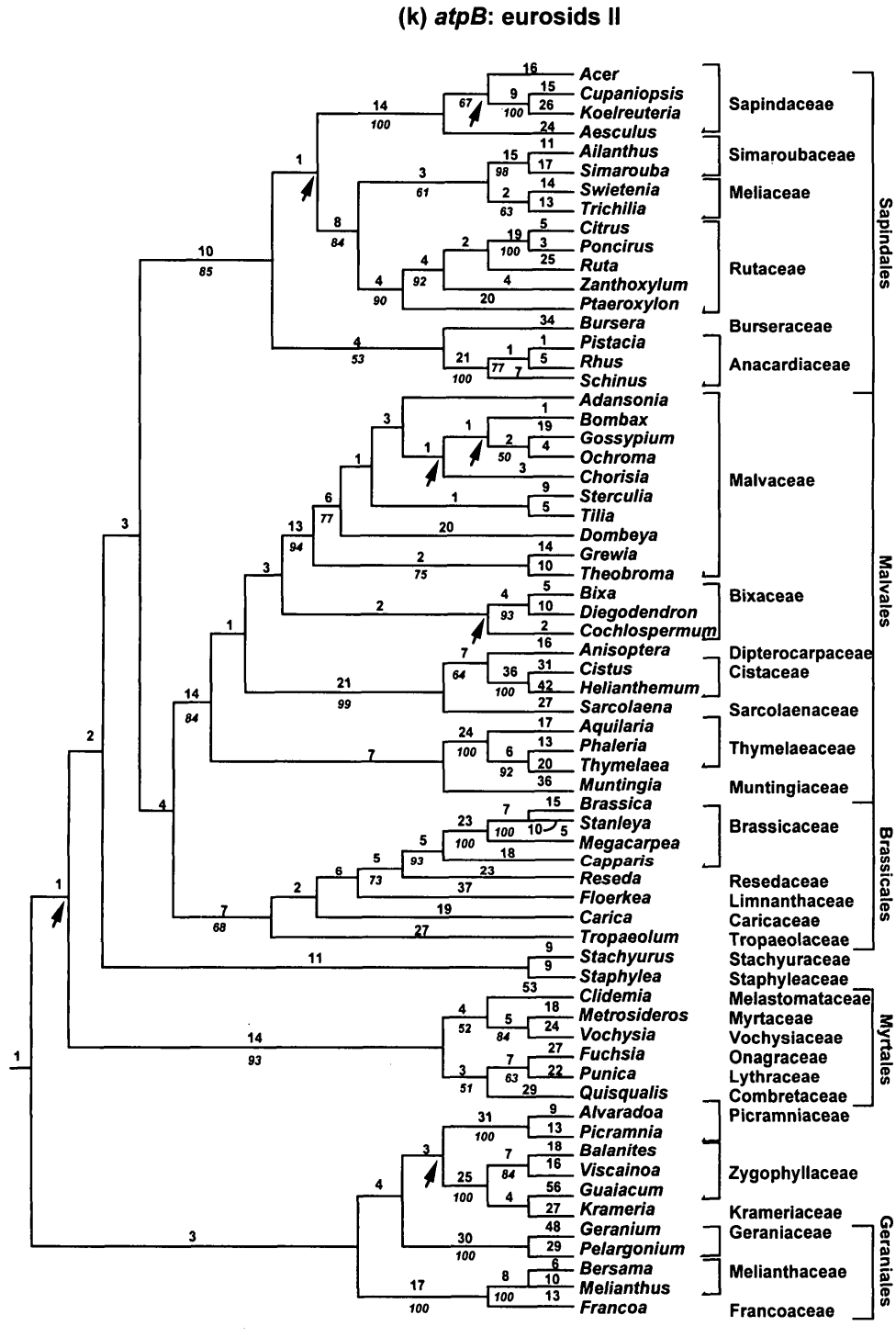

FIGURE 5. (Continued)

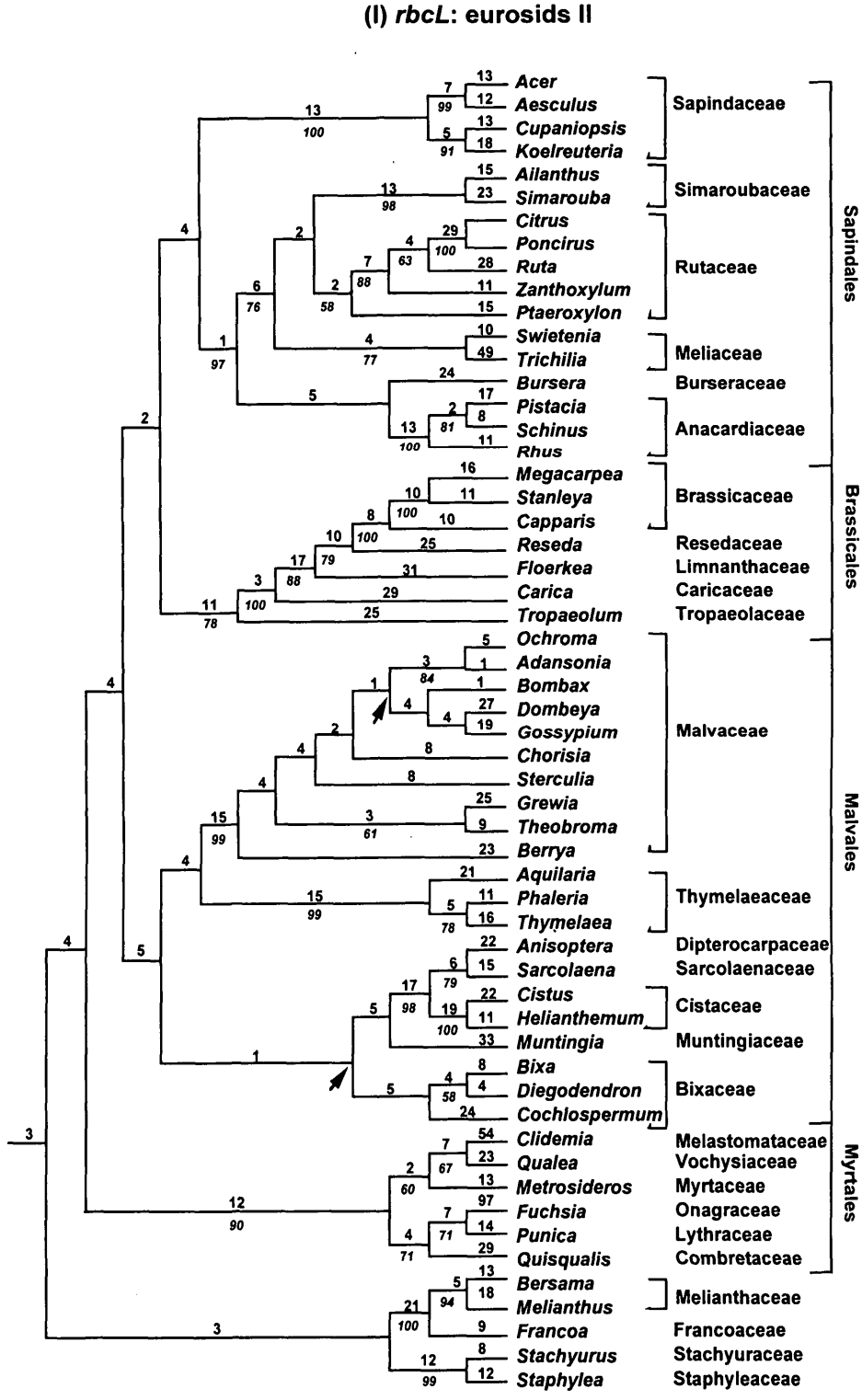

FIGURE 5. (Continued)

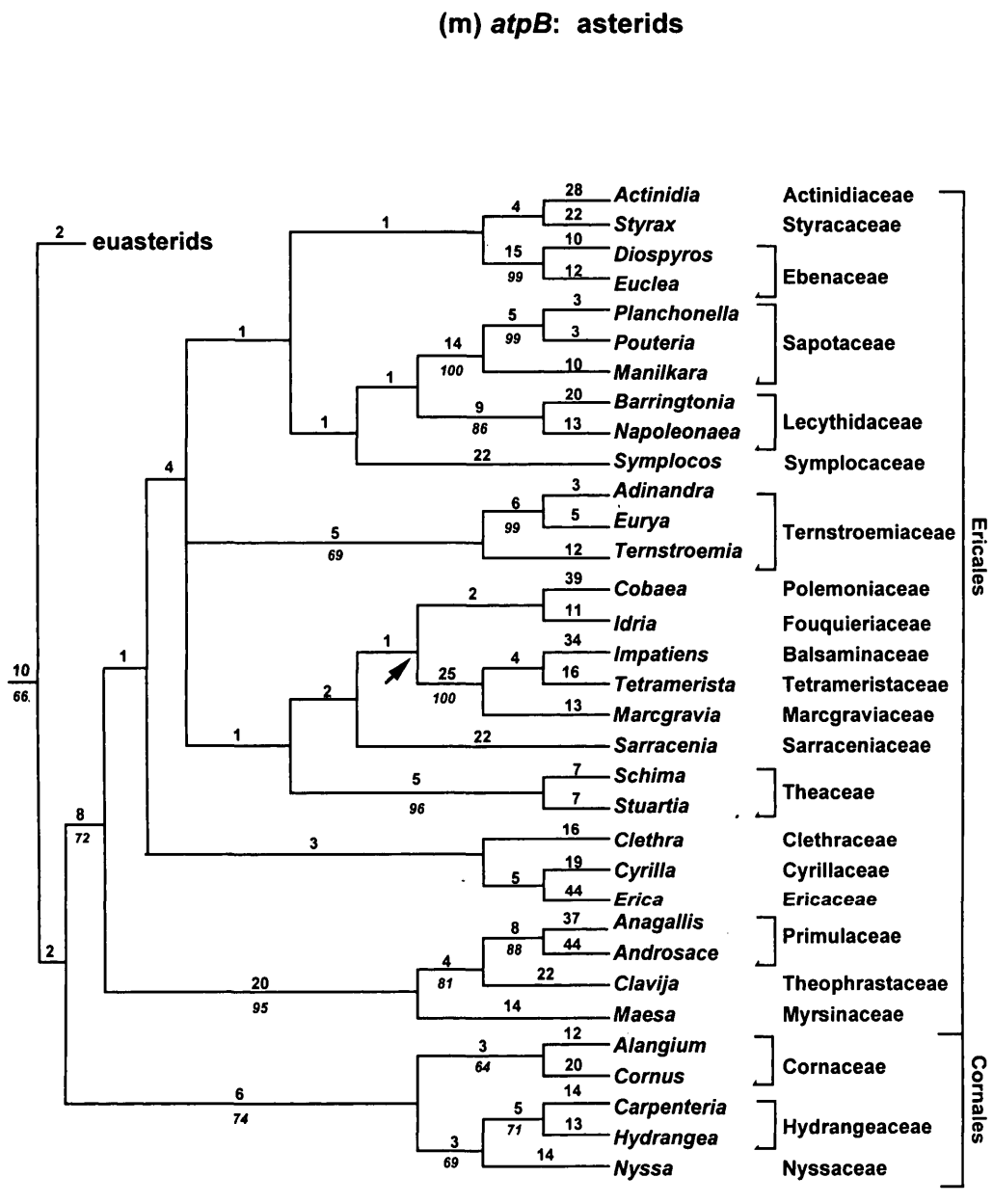

FIGURE 5. (Continued)

Asteropeiaceae in Morton et al., 1997; Tamaricaceae, Ancistrocladaceae, and Dioncophyllaceae in Fay et al., 1997, and Lledó et al., 1998) and require evaluation in using this combined-gene matrix.

Rosids

Within the rosids (61% *atpB/rbcL* bootstrap support), two major clades are identified in all shortest trees (neither of which, however, receives support >50%): eurosid I, composed of Celastraceae (including Hippocrateaceae plus Huaceae, Parnassiaceae, and Stackhousiaceae), Cucurbitales, Fabales, Fagales, Malpighiales, Oxalidales, and Rosales (Figs. 5i,j, 6e); and eurosid II, including Brassicales, Malvales, Myrtales, and Sapindales (Figs. 5k,l, 6f). There are also several clades that cannot be clearly as-

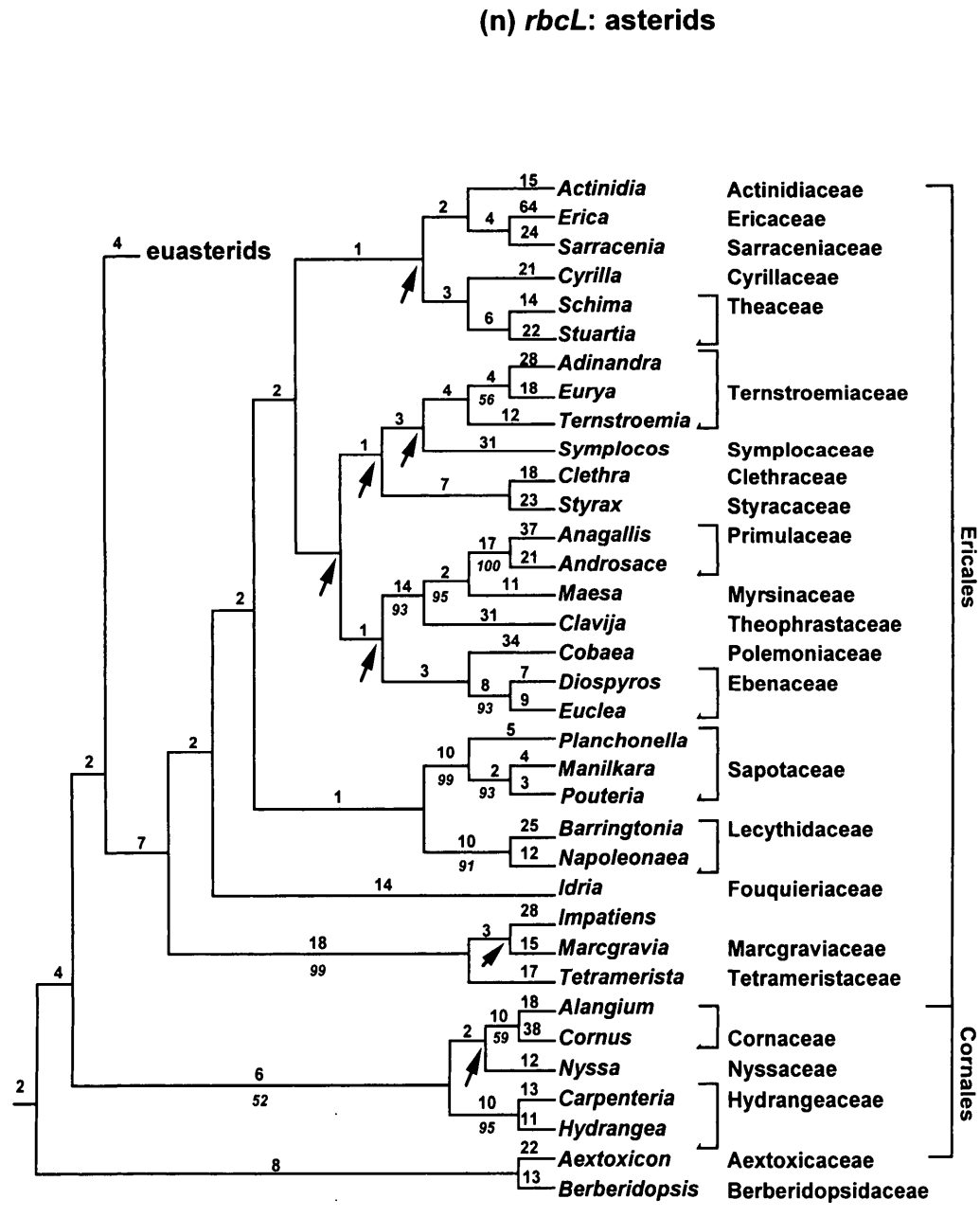

FIGURE 5. (Continued)

signed to either of these major rosid groupings: Geraniales (based on the 1993 *rbcL* tree and unpublished results: Francoaceae, Geraniaceae, Melianthaceae, Staphyleaceae, and Stachyuraceae and perhaps Crossosomataceae, Geissolomataceae, Greyiaceae, and Vivianiaceae), Zygophyllaceae/Krameriaceae (monophyly supported by bootstrap of 86% in the combined trees), and Picramniaceae. The placement of Zygophyllaceae/Krameriaceae and Picramniaceae as members of eurosid I receives <50% bootstrap support in the combined analysis. Each of the other orders of the eurosid I clade is supported in the combined analysis: Celastraceae/Huaceae/Parnassiaceae/Stackhousiaceae (<50%, <50%, 60% for *atpB*, *rbcL*, and combined, respectively), Cu-

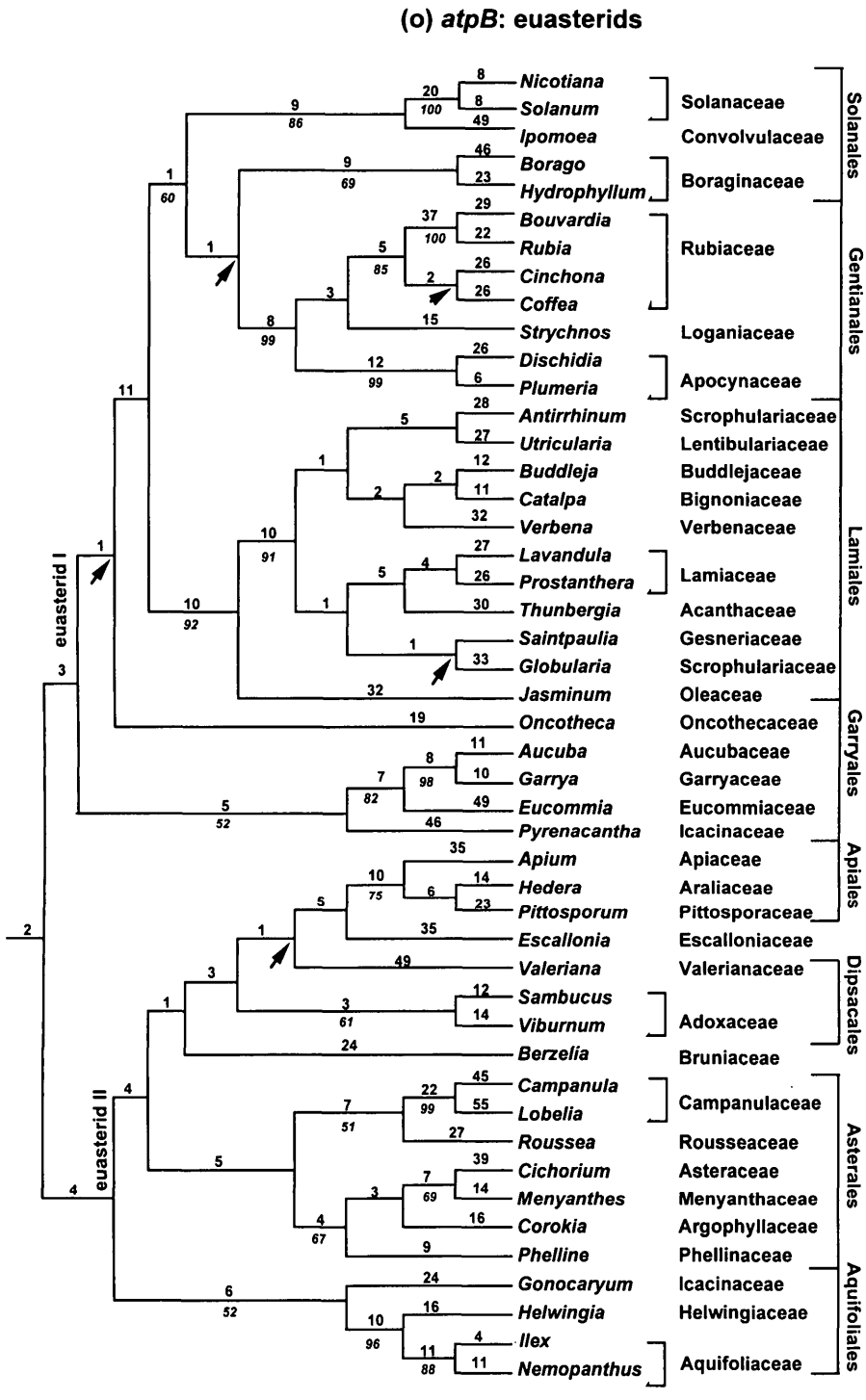

FIGURE 5. (Continued)

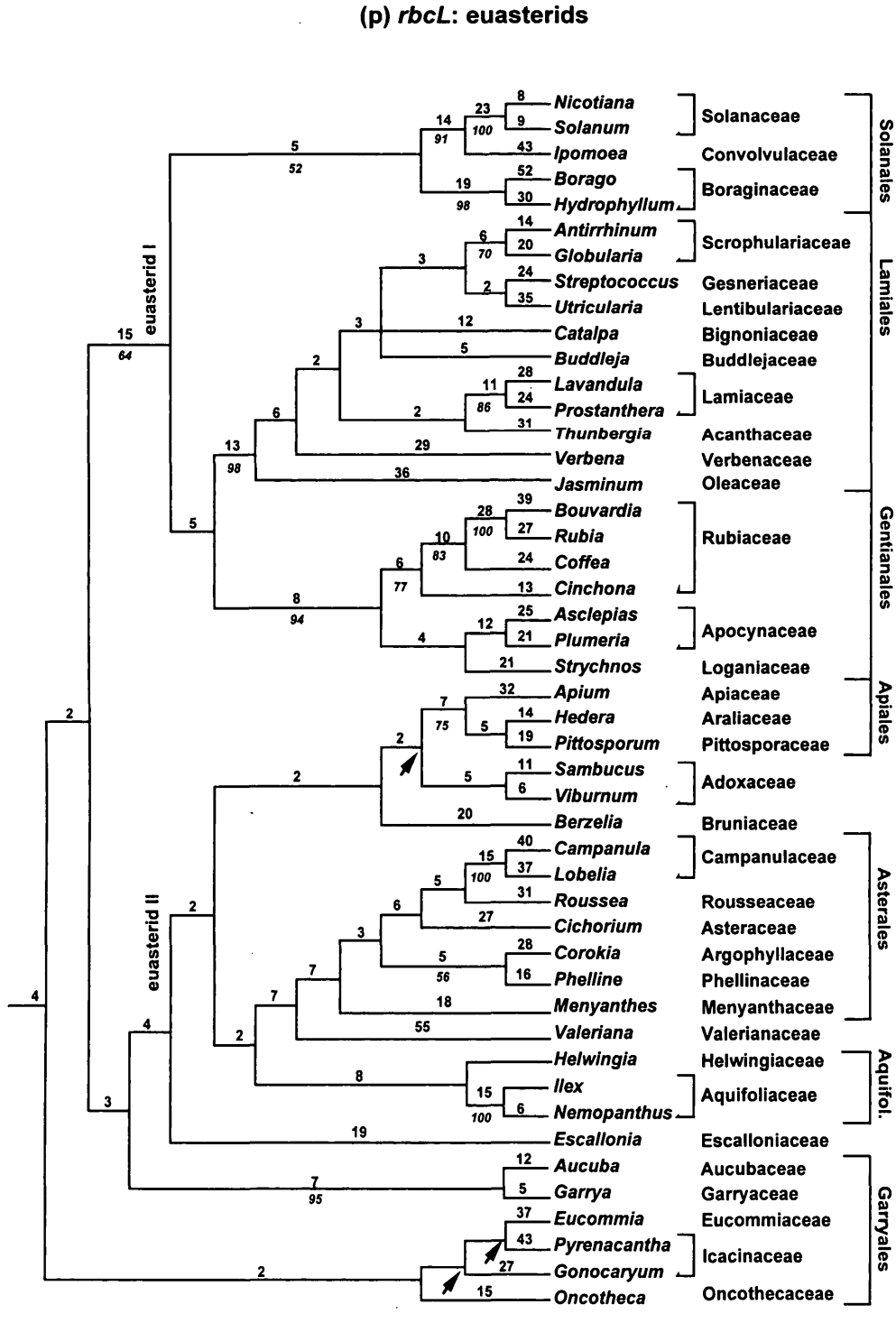

FIGURE 5. (Continued)

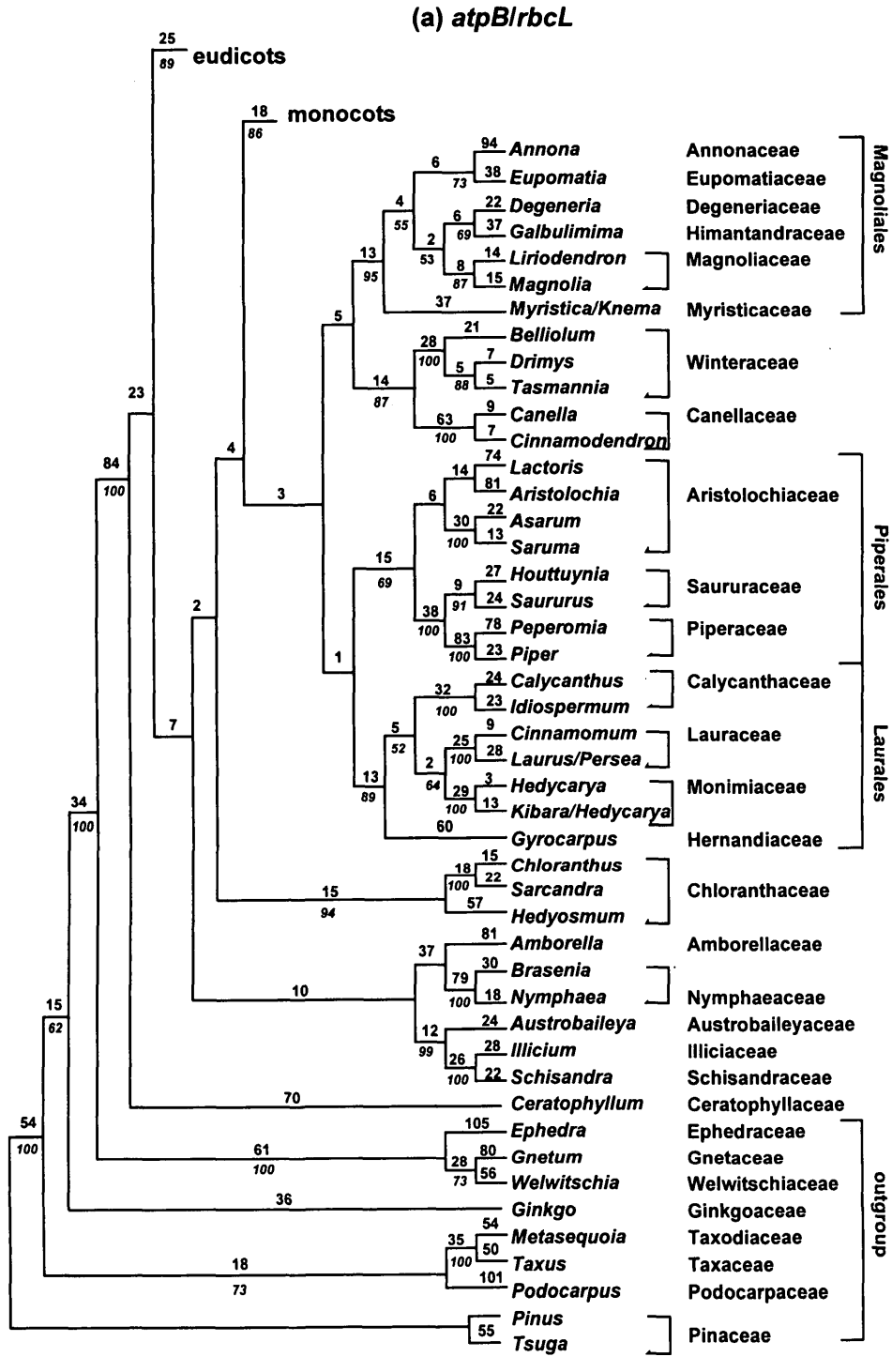

FIGURE 6. One of 8,600 best trees resulting from the exploratory phylogenetic analysis of *atpB/rbcL* combined for 357 taxa. The tree length is 25,936 steps, CI = 0.14, and RI = 0.56. Arrows indicate branches collapsing in the strict consensus tree of 2,000 shortest trees. The numbers of steps are indicated above the branches, and bootstrap values >50% are indicated below the branches. Because of its size, the tree has been broken into eight parts: (a) noneudicots, (b) monocots, (c) eudicots, (d) caryophyllids, (e) eurosids I, (f) eurosids II, (g) asterids, (h) euasterids. Suprageneric nomenclature follows that published by the Angiosperm Phylogeny Group (1998).

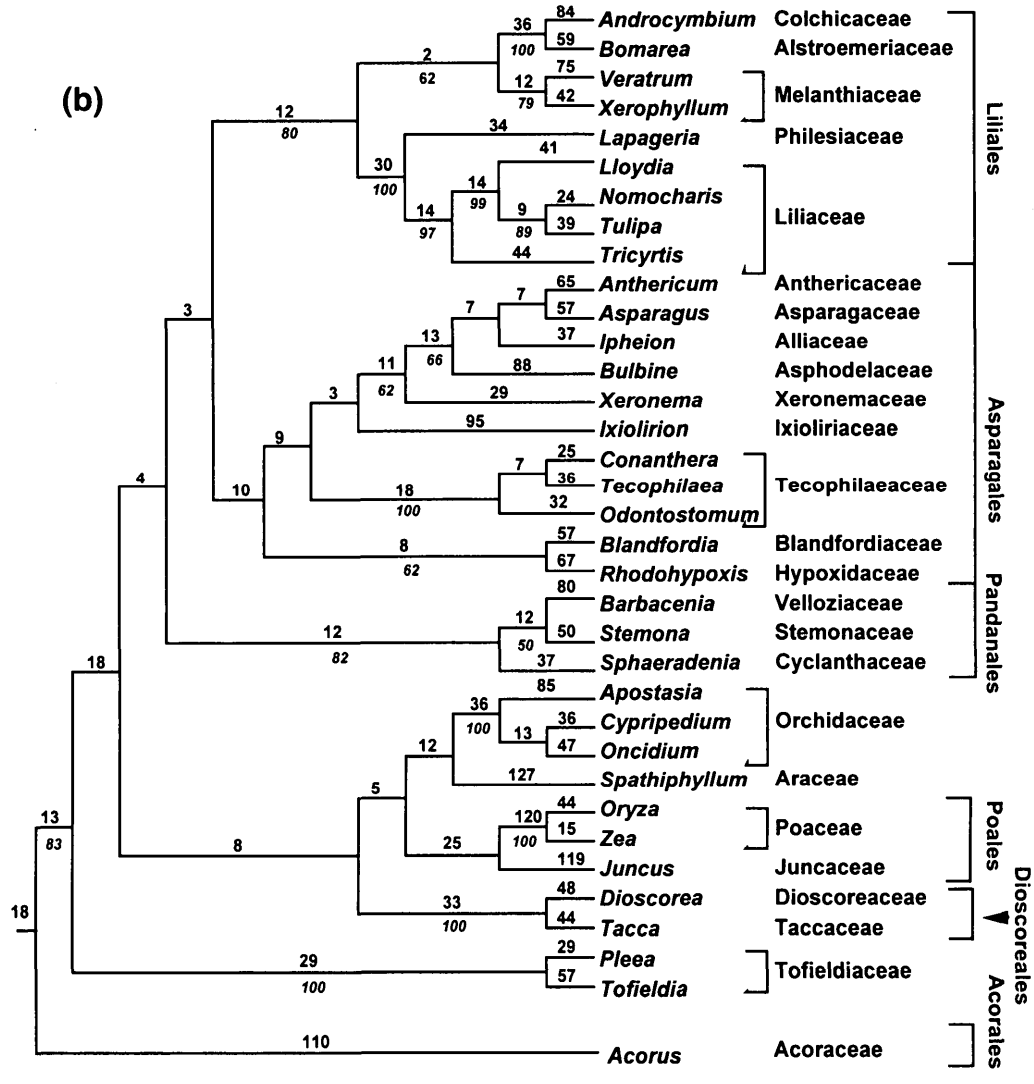

FIGURE 6. (Continued)

curbitales (97%, 87%, 100%), Fabales (<50%, <50%, 89%), Fagales (98%, 94%, 100%), Malphigiales (66%, <50%, 92%), Oxalidales (90%, 93%, 100%), and Rosales (<50%, <50%, 73%). Geraniales are placed as sister to the eurosid II clade but without bootstrap support >50% (Fig. 6f). The other orders of eurosid II also receive support from these analyses (Figs. 5k,l, 6f): Brassicales (68%, 78%, 99%), Malvales (84%, <50%, 97%), Myrtales (93%, 90%, 100%), and Sapindales (85%, <50%, 100%). Many other relationships within the eurosid I and II clades also are well supported (for example, nearly all dichotomies in Brassicales, Cucurbitales, Fabales, Fagales, Geraniales, Myrtales, Oxalidales, Rosales, and Sapindales receive bootstrap support >50%).

Asterids

An expanded asterid clade (Asteridae sensu Olmstead et al., 1992) is recovered from analysis of all three matrices. Broad analysis of 18S rDNA similarly revealed an expanded Asteridae, although some analyses showed Caryophyllales embedded in Ericales, within Asteridae s.l. (Soltis et al., 1997b). In the combined trees, this expanded asterid clade is strongly supported (92%; Fig. 6g). Cornales (52%, 74%, 96%) and Ericales (<50%, 72%, 97%) together

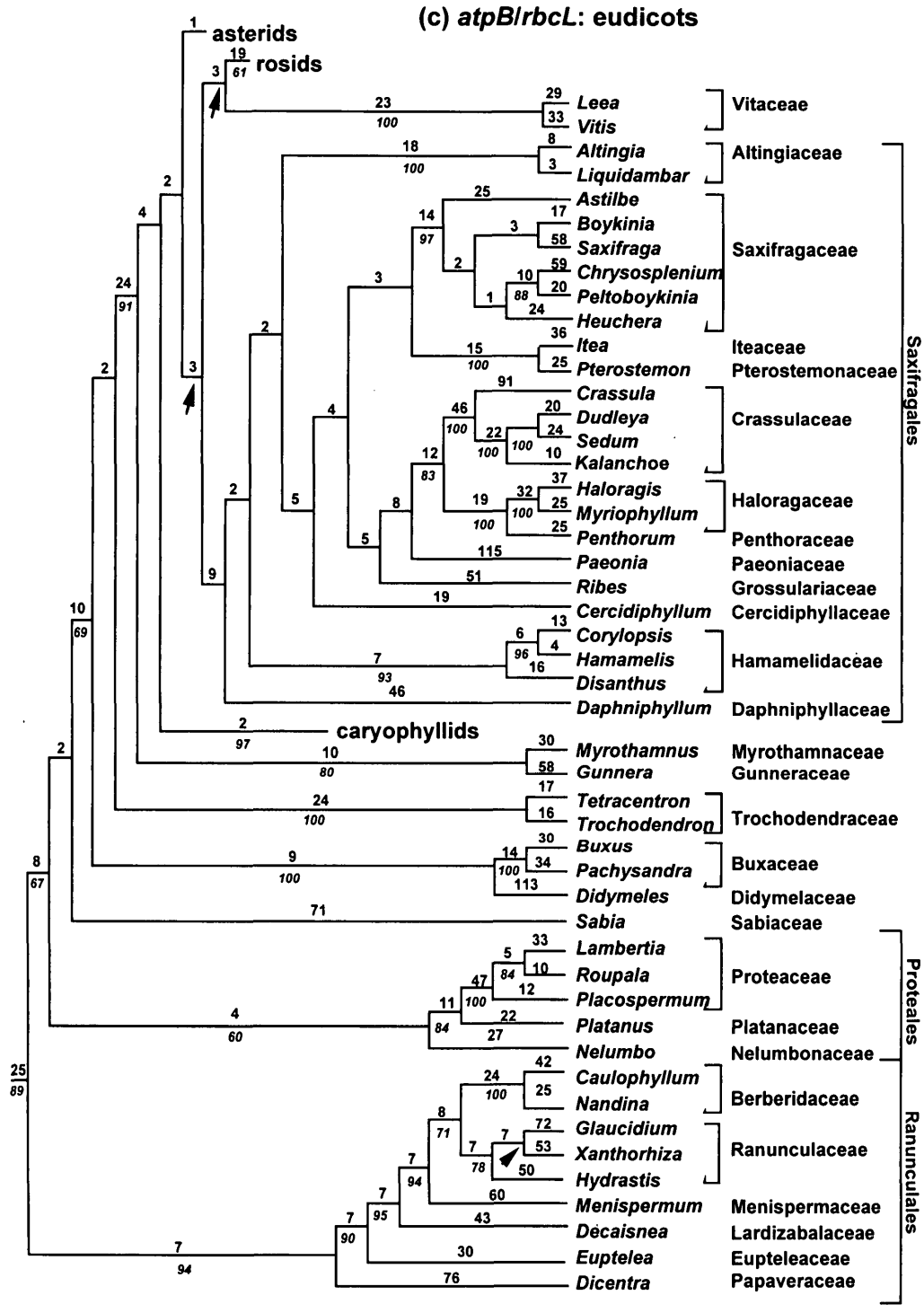

FIGURE 6. (Continued)

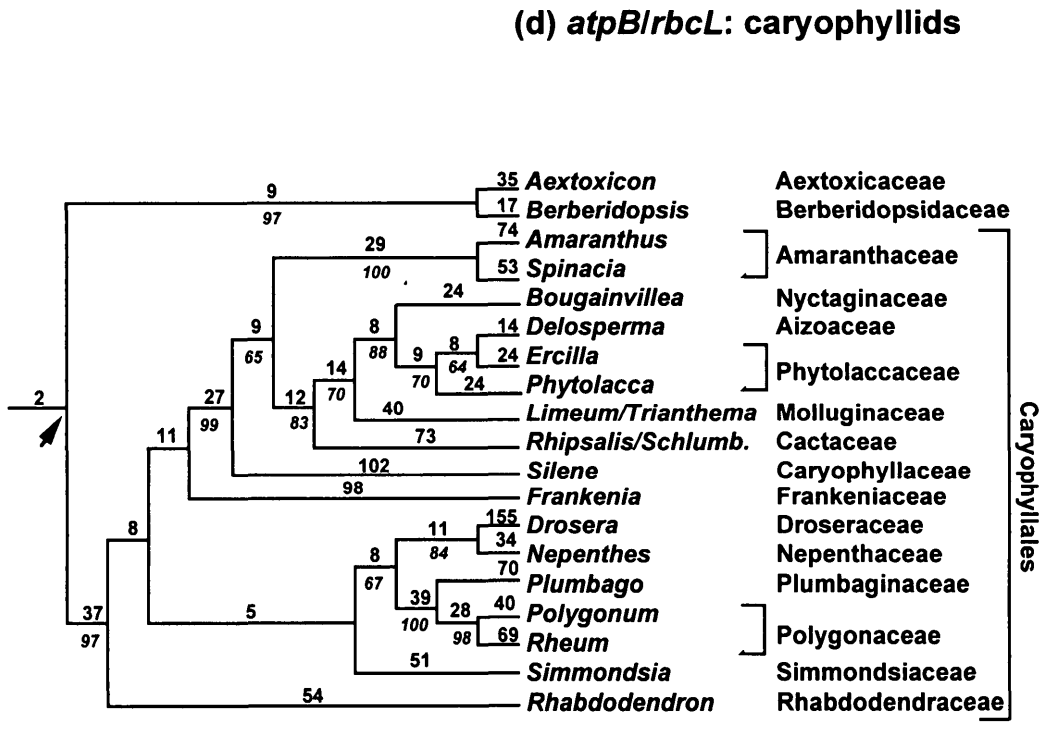

FIGURE 6. (Continued)

form a clade (<50%) in the combined analysis (Figs. 5m,n, 6g), and this clade is sister to the clade composed of euasterid I (Solanales, Lamiales, Gentianales, and Garryales) and euasterid II (Apiales, Asterales, Dipsacales, and Aquifoliales; Figs. 5o,p, 6h) clades. Although patterns within Cornales are fairly well supported, those within Ericales are not. Cornales should also include Grubbiaceae, Loasaceae, and probably Hydrostachydaceae (Xiang et al., 1993; Hempel et al., 1995; Morton et al., 1996), but no additional families have been found to be members of Ericales since Morton et al. (1996).

Within the euasterids (Figs. 5o,p, 6h), there are two orders of uncertain relationships: Garryales, in which *Oncotheca* may not be a member (the rest have weak support as a clade: 52%, not found, 55%) and Aquifoliales (52%, not found, 55%). The monophyly of the latter and euasterid II has low bootstrap support (51% in the combined analysis). Support for euasterid II (62% in the combined trees) and the relationships therein are almost all weak (Fig. 6h); only Apiales (75%, 75%, 99%) and Asterales (<50%, <50%, 71%) receive bootstrap support of 50% or more. In contrast, within euasterid I (<50%, 64%, 100%), many relationships are well supported: Gentianales (99%, 94%, 100%), Lamiales (92%, 98%, 100%), and Solanales (not present, 52%, 60%). Relationships within Gentianales and Solanales appear to be well resolved, whereas those within Lamiales are generally unclear. Within Lamiales, only the monophyly of all families (excluding Oleaceae) receive strong bootstrap support (91%, <50%, 91%).

DISCUSSION

The trees derived from analysis of a combined *rbcL/atpB* data set are a marked improvement in terms of support for the terminal groups identified, mostly defined as orders here and elsewhere (APG, 1998). This represents one of the first phylogenetic analyses that has dealt with combined gene matrices for such wide taxonomic sampling. Overall, the phylogenetic relationships observed with *atpB* compare extremely well with those of *rbcL;* they are

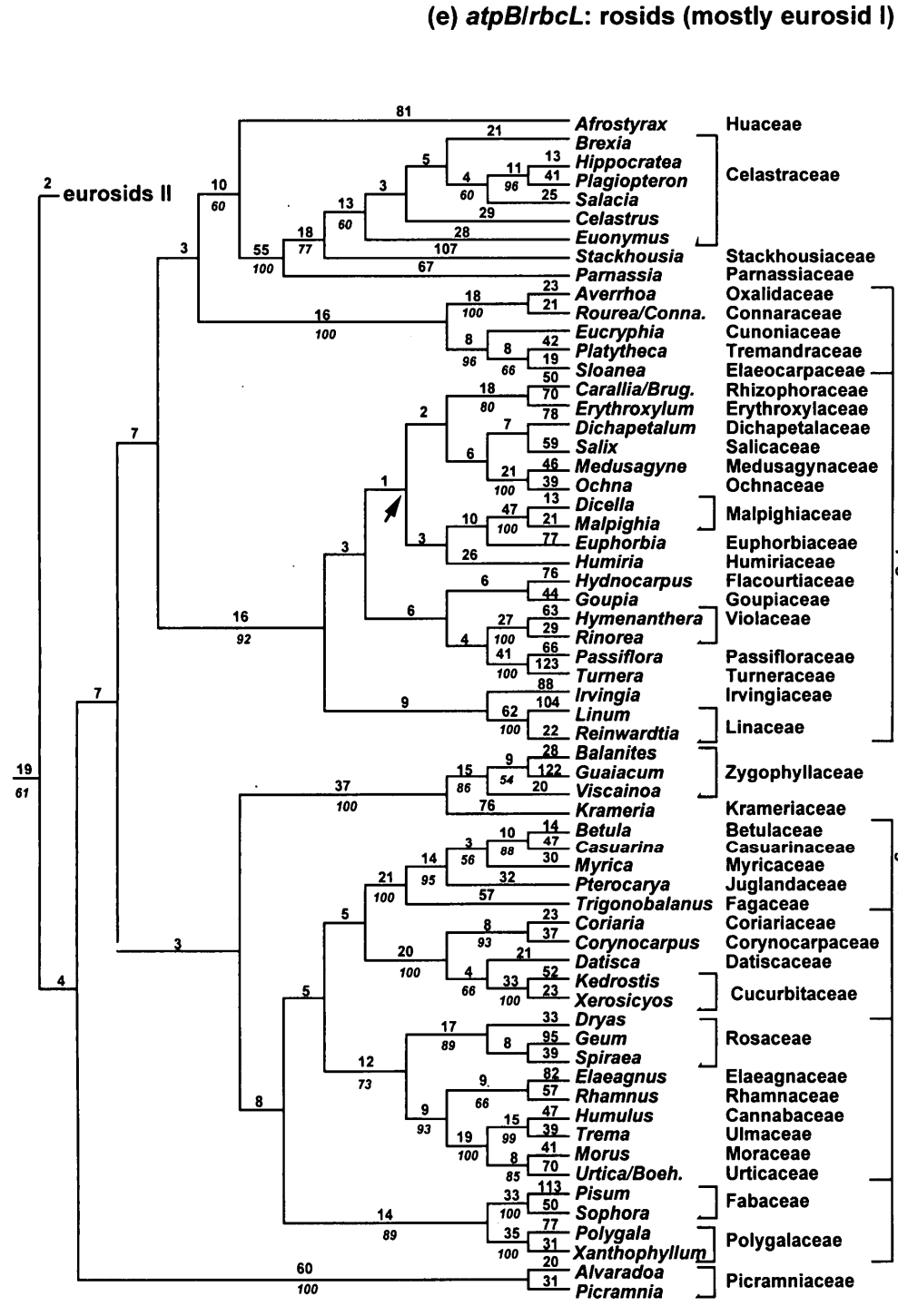

FIGURE 6. (Continued)

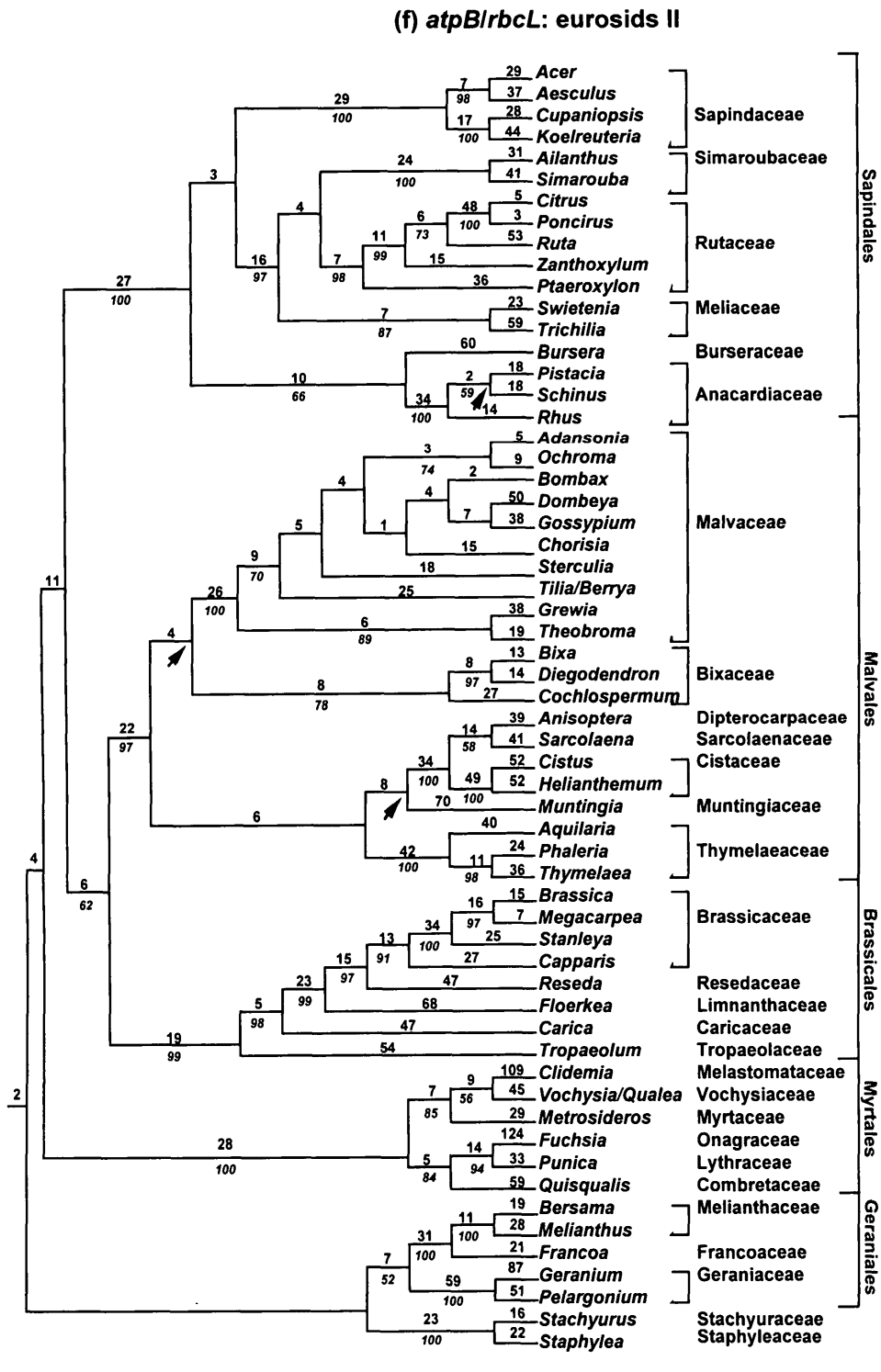

FIGURE 6. (Continued)

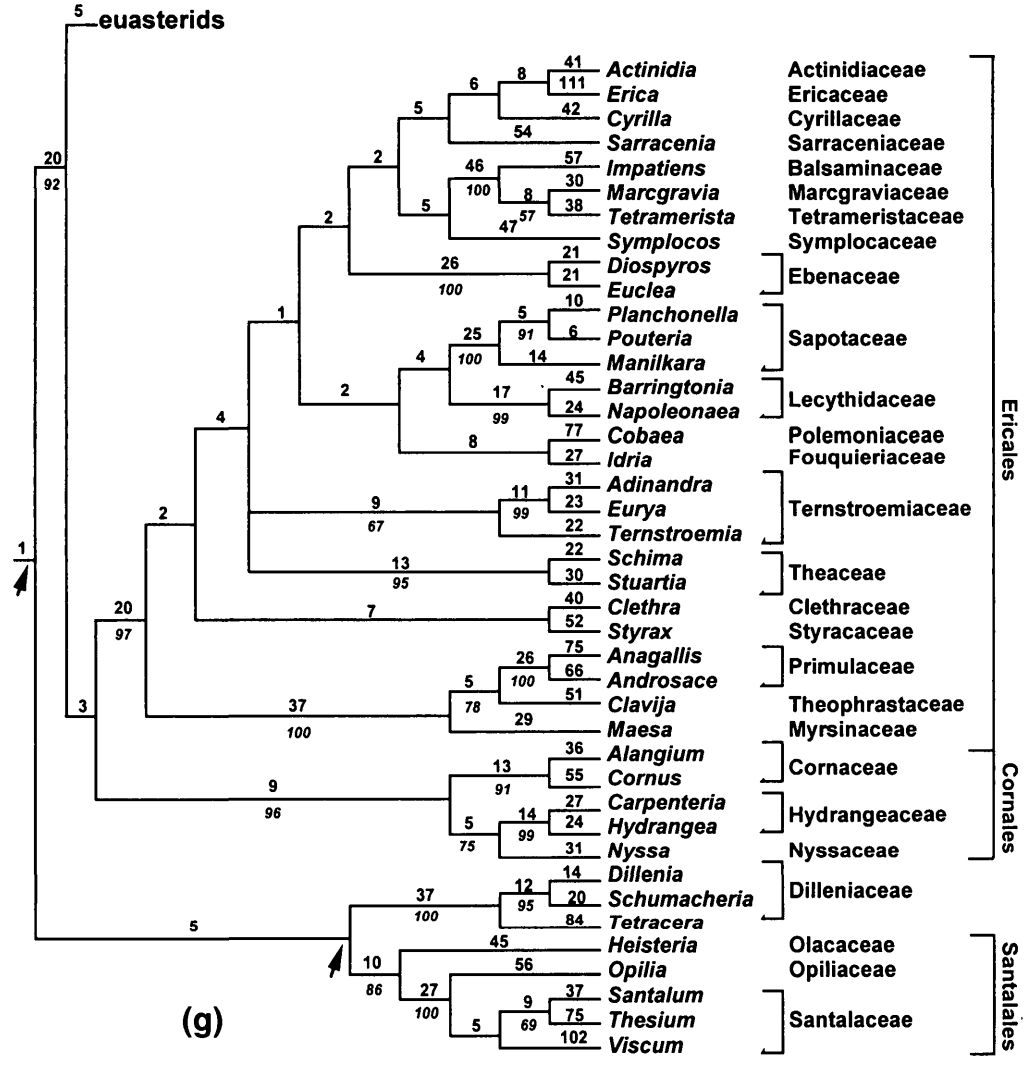

FIGURE 6. (Continued)

also in general agreement with those from 18S rDNA (Soltis et al., 1997b).

Debates on phylogenetic methods and use of molecular characters in large data sets have been numerous (e.g., Graur et al., 1991; Patterson et al., 1993; Hillis et al., 1994; Mishler, 1994; Hillis, 1995, 1996, 1998; D'Erchi et al., 1996; Graybeal, 1998; Kim, 1998). The analyses presented here contradict several widely held ideas concerning molecular phylogenetics and thus have broad implications beyond angiosperm relationships.

Rates of Divergence versus Phylogenetic Signal

When considering potential loci to be sequenced, systematists are often interested in studies that have demonstrated similar amounts of variability (e.g., in a previously published molecular study at the same taxonomic level). A great deal of attention has been paid to the rate of molecular evolution as a feature of prime importance for resolution at different taxonomic levels (e.g., Taberlet et al., 1991; Graybeal, 1993; Meyer,

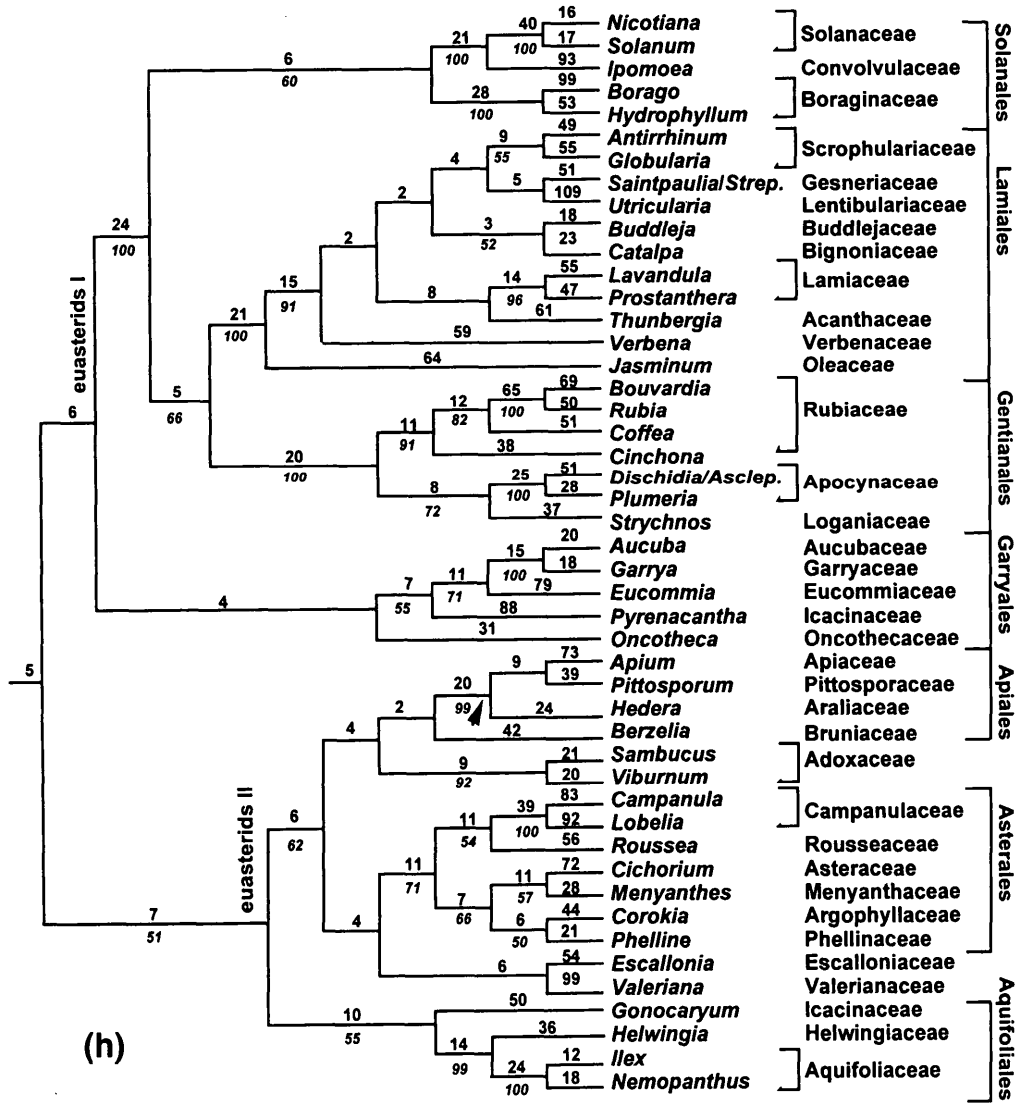

FIGURE 6. (Continued)

1994; Soltis and Soltis, 1998). This topic is not clear-cut, however, for several reasons.

As noted, pairwise comparisons indicate that *rbcL* has a slightly, but statistically insignificant, faster rate of change than *atpB*. The former also has more numerous variable sites (which are more important in pairwise comparisons than in tree-based methods for estimating substitution rates), but the *atpB* trees are longer than those from *rbcL* (see Hoot et al., 1995, 1999, for differing results). Similarly, Lledó et al. (1998) found that although the noncoding *trnL-F* regions had approximately twice as many variable sites as *rbcL* (in Plumbaginaceae), the *rbcL* tree was actually longer than that for *trnL-F*. Thus, if rates estimated from pairwise comparisons are emphasized, a different answer will be obtained from that based on looking at trees. Counterintuitively, the number of steps in a tree can be greater when using a gene with lower estimated rates from pairwise comparisons. These examples illustrate that emphasizing only a generalized rate of evolution for a given gene can be misleading (Olmstead et al., 1998). Obviously, an increased number of steps reflects more detected homoplasies, but in a parsimony framework "homoplasy is considered as

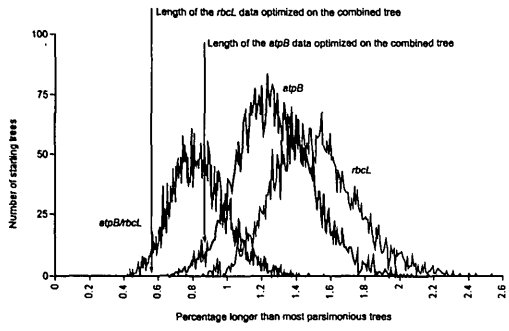

FIGURE 7. Distribution of starting tree length for both *atpB* and *rbcL* data sets alone and in combination. The percentage longer than the most-parsimonious trees was calculated by computing the length of 5,000 replicates of random taxa addition without swapping (i.e., starting trees) in comparison with the length of trees found in each full heuristic search (i.e., optima reached after extensive swapping). Median values are 1.49% longer for *rbcL*, 1.25% longer for *atpB*, and 0.82% longer for *atpB/rbcL* combined. Arrows indicate the length of the trees when the single genes are optimized onto the combined topology (best topology); swapping below these limits leads to ever-greater underestimates (trees that are shorter than the optimum for the combined data set).

deserving of explanation as is homology" (Siddall and Kluge, 1997:317). Homoplasy is evidence, and the more evidence that is available, the more accurate is the resulting tree. Increased homoplasy (i.e., lower CI) does not necessarily mean a weaker signal and a lower RI.

It is not rate that should be of interest, but rather how "decisive" a data set is (Goloboff, 1996; Davis et al., 1998). Similar ideas have emerged in other studies: Analyzing mitochondrial genes in vertebrates, Zardoya and Meyer (1996:939) stated that "performance of genes in recovering the expected . . . trees seems to be not strongly dependent on their rate of evolution and concomitant saturation processes"; they stressed that density of lineage-initiation events in time and completeness of taxon representation are more important factors than the overall rate of gene evolution. To detect ancient phyletic radiations, our data indicate that using the more rapidly evolving genes would be more appropriate than using the more slowly evolving regions, which are unlikely to contain much evidence of a rapid radiation.

Not all genes or sequence regions contain the same strength of signal for the same monophyletic groups. This is obvious here; for some clades it is *rbcL* that contains the strongest signal (e.g., bootstrap of 64% for the monophyly of Magnoliales, 90% for Chloranthaceae, 59% for the monocots), whereas for others it is *atpB* (e.g., bootstrap of 94% for the monophyly of Canellaceae/Winteraceae, 68% for Laurales); Soltis et al. (1998) and Bayer et al. (1999) give a detailed analysis of the variation in signal for these genes for different clades. Thus, each separate matrix has an heterogeneous pattern of support, and no overall measure of matrix signal or rate is adequate to predict whether a specific group will receive high bootstrap support.

Information Content in Codon Position and Substitution Type

If the use of equally weighted analyses does not appear to be efficient at detecting a clear set of relationships (Huelsenbeck and Hillis, 1993; Hillis et al., 1994), many authors have used some form of relative weighting in the hope of improving resolution (see Albert et al., 1993; Manhart, 1994; Allard and Carpenter, 1996; Nandi et al., 1998). Our results indicate that simple down-weighting or omission of third positions or transitions (as is usually performed in weighted parsimony) is an oversimplification and may lead to loss of resolution and support. Thus, downweighting or eliminating third positions as a class is unwarranted; in fact, the reverse—giving greater weight to third positions—appears to be more appropriate. For example, even though their variation is much more frequent (75% and 71% of the variable positions in *atpB* and *rbcL*, respectively, are third positions), the RI of third positions was higher than those for the first and second positions. Yang (1996) reached this same conclusion, using mitochondrial coding sequences for hominoids, and added that the notion of saturation appeared to depend on analytical method. Similarly, Lewis et al. (1997:377), using *rbcL* sequences in basal embryophytes, found that "the presence of signal in third codon positions . . . means that definitions of saturation based on pairwise comparisons of sequences inadequately assess phylogenetic signal." With respect to codon positions on the *rbcL* and

atpB trees, some of each position (including third) were invariant, whereas others changed many times: first positions up to 43 and 102 times, second positions up to 63 and 67 times, and third positions up to 78 and 114 times for *atpB* and *rbcL*, respectively. As stated by Olmstead et al. (1998), most studies have estimated means and, unfortunately, have ignored variance when comparing rates of genes and coding positions.

Ts/tv weighting is probably the most common form of weighting because it can be used with coding and noncoding regions alike; transitions are often down-weighted because of their higher frequency and thus greater attendant homoplasy (see Allard and Carpenter, 1996). For *atpB* and *rbcL*, we determined both the frequencies as well as CI and RI for each type of substitution. Transitions (calculated on the trees) were more frequent (1.65 times for *rbcL* and 2.09 times for *atpB;* see Table 3), but their CIs were close to that of transversions (i.e., homoplasy for both types of substitution was similar), whereas the RIs for the more frequent transitions were actually higher, the phylogenetic signal of transitions was stronger. We are convinced that the results obtained here are unbiased by the use of Fitch parsimony (Fitch, 1971), and several workers have found that parsimonious trees derived from transversion weighting did not differ from those obtained with the data equally weighted (e.g., Allard and Carpenter, 1996; Hoot and Douglas, 1998). With respect to our data, the much higher internal support produced by the combined matrices in which nearly three-fourths of the variability could be attributed to third positions and two-thirds to transitions demonstrates that these most frequent categories of change can produce robust estimates of evolutionary relationships.

Root of the Angiosperms

In all phylogenetic analyses of morphological data, the noneudicots formed a grade, not a clade (Donoghue and Doyle, 1989; Loconte and Stevenson, 1991; Nandi et al., 1998; Hoot et al., 1999), but here with *rbcL* alone (as in Chase et al., 1993) and with *rbcL/atpB* combined, they (excluding *Ceratophyllum*) form a clade that is sister to the eudicots. With *atpB* alone and in the 18S rDNA/*rbcL* and 18S rDNA trees (Soltis et al., 1997a, 1997b), as well as in trees based on three genes (Soltis et al., 1998; Hoot et al., 1999), the noneudicots also formed a grade. In general, the traits of the noneudicots have been assumed to be plesiomorphic for the angiosperms (largely because we have all learned that these are the "archaic" or "primitive" angiosperms), and so they have been coded in such a way that monophyly was precluded. For example, in the matrix of Nandi et al. (1998), no outgroups for the angiosperms were used, and so it would have been impossible to view binary data as forming mutual synapomorphies; one character has to be plesiomorphic by default if the other is derived. However, if all or some portion of the noneudicots and the eudicots are sister taxa, as in the combined *atpB/rbcL* trees, then some proportion of eudicot traits could equally well be viewed as plesiomorphic for the angiosperms.

There are two major categories of noneudicots, which were referred to as magnoliid I and II in Nandi et al. (1998). Magnoliid I (= eumagnoliids) comprises monocotyledons plus Laurales, Magnoliales, and Piperales and perhaps Chloranthaceae and Canellaceae/Winteraceae. Non-DNA characters that are frequent in eumagnoliid families (and perhaps are synapomorphies) are the presence of asarone, gibacin, licarin A, veraguensin, liriodenine, and indole alkaloids; rod- or tube-shaped epicuticular waxes; trimery in the calyx (and in the androecium); successive microsporogenesis; perisperm (or nucellar-derived storage tissue), expanded stamens, and endotestal crystals (see details in Nandi et al., 1998). Many of these characters are polymorphic within the taxa in which they occur, and others are poorly sampled, but the list of possible synapomorphies is longer and more diverse than for many clades of eudicots. These features are all either absent or rare characters in eudicots, and it is likely that at least some of these are synapomorphies for the eumagnoliids. Magnoliid II contains only Amborellaceae, Austrobaileyaceae, Cabombaceae, Illiciaceae, Nymphaeaceae, and Schisandraceae. These Magnoliid II families are the focus of controversy because in various phylogenetic studies they

occupy different positions, including being positioned at the root within the angiosperms (making them paraphyletic). Magnoliid II are monophyletic for *rbcL* (Chase et al., 1993; Qiu et al., 1993; this paper), *rbcL*/morphology (Nandi et al., 1998), and *rbcL/atpB* combined (this paper); they are paraphyletic with *atpB* (this paper) and according to combined analyses based on three genes (Hoot et al., 1999) and 18S rDNA (Soltis et al., 1997b). They also share para- or tetracytic stomata and expanded stamen connectives with the magnoliid I noneudicots. Magnoliid II have unfused carpels filled with mucilage as potential synapomorphy, which is perhaps shared with Chloranthaceae but is not found in Illiciaceae (Endress and Igersheim, 1997). This could be interpreted as primitive within the angiosperms (Endress and Igersheim, 1997), but there is no way to determine whether filling an open cavity with mucilage was inherited from a common ancestor or was an innovation, given that all other clades of angiosperms have eliminated the cavity by postgenital fusion. Mapped onto the 18S rDNA (Soltis et al., 1997a, 1997b, 1998, 1999) and our *atpB* trees, this character would be viewed as plesiomorphic, but on the *rbcL* and *atpB/rbcL* trees it would be an apomorphy. Many families of magnoliid II are problematic, and more data are required to resolve the rooting of the angiosperms. Not only are more gene sequence data needed, but particularly critical would be characters that can be polarized by reference to outgroups. Unfortunately, many traits such as the presence of mucilage in an unfused carpel, microsporogenesis, and chemistry are unlikely to be available from the fossil record. Most recently, Qiu et al. (1999) identified *Amborella* at the root of the angiosperms based on a combined analysis of five genes from the mitochondrial, plastid and nuclear genomes.

Nonmolecular Characters and Relationships in the Eudicots

The larger clades found in our analysis do not correspond to the distribution of many of the morphological characters used in several previous taxonomic schemes (e.g., centrifugal stamen initiation, parietal placentation). In contrast, the two major categories of angiosperms revealed in the combined *atpB/rbcL* analysis—noneudicots and eudicots (with their respective sets of familial and ordinal relationships)—have never been recognized previously. The sole character that delimits these groups is the nature of pollen development, which in the former results in uniaperturate (mostly monosulcate) pollen and in the latter triaperturate pollen.

The eudicots, in addition to triaperturate pollen, have a secretory anther tapetum, simultaneous microsporogenesis, filaments frequently much longer than anthers (an especially pronounced feature of the higher eudicots), and two leaf traces (one in many Ericales and Myrtales). Leaf venation in which there is a single primary vein with the lateral veins terminating at the margin (often in a tooth; craspedodromous) is also typical of the lower eudicots plus Dilleniaceae, Saxifragales, and Vitaceae, thus leaving this condition as a synapomorphy that is further modified in Caryophyllales, Santalales, asterids, and rosids (not in Rosales). The unanticipated *Nelumbo/Platanus*/ Proteaceae clade is difficult to characterize, given the diverse habits of its member taxa, but these plants are marked by epicuticular waxes that are rod- or tube-shaped, large seeds with scanty or no endosperm (the latter condition found only in one genus of Proteaceae), and alternate vessel pitting.

The higher or core eudicots (here including *Gunnera/Myrothamnus*) exhibit calyx and corolla differentiation (although this is not well developed in some Saxifragales, Rosales, and members of Malpighiales such as Flacourtiaceae), calyx and corolla organs in fives (but not in *Gunnera/Myrothamnus*), a floral disk (but not in Caryophyllales, many Ericales, and Geraniales), nondecurrent stigmas, and antesepalous/antepetalous carpels that are partially to wholly fused (but not in some Saxifragales such as Paeoniaceae and some Crassulaceae). The general arrangement of floral organs in nearly all higher eudicots is thus quite stereotyped, but deviation (apparent reversals) from this syndrome marks subclades within several orders.

The lower eudicots (Buxaceae, Didymelaceae, Proteales, Ranunculales, and Trochodendraceae) deviate substantially from these characteristics, and in many respects

have features otherwise typical of the non-eudicots (e.g., see Nandi et al., 1998). Sabiaceae are an exception to this syndrome: They are like other lower eudicots in their craspedodromous venation and lack of ellagic acid, but their floral characters are much like those of the higher eudicots, from which, on the basis of the *atpB/rbcL* trees, they are excluded.

Caryophyllales as defined here are a remarkably well-supported group that no previous classification had identified. Their characters include the presence of pinitol and ancistrocladine, a spinulose sexine, nuclear endosperm development (also typical of the rosids), lack of hypostase, starchy endosperm (although not in core Caryophyllales), alternate vessel pitting, and simple perforations of vessel end-walls.

Although Caryophyllales form a well-defined clade, their relationships to asterids, rosids, and Saxifragales are not clear. Caryophyllales represent highly divergent members of the higher eudicots and lack several of the synapomorphies that otherwise characterize the asterids and rosids. Like most of the large and reasonably diverse clades identified here, members of Caryophyllales are highly specialized and have converged on floral and vegetative traits that are typical of other lineages; they lack, however, mucilage cavities/cells typical of the rosid/asterid clade, a hypostase in their seeds, and tricolporate pollen, features found in nearly all of the other more advanced groups. The distribution of anomalous secondary growth in Caryophyllales (e.g., Aizoaceae, Cactaceae, Caryophyllaceae, Chenopodiaceae, Didiereaceae, Dioncophyllaceae, Droseraceae, Nepenthaceae, Nyctaginaceae, Phytolaccaceae, Plumbaginaceae, Polygonaceae, Portulacaceae, Rhabdodendraceae, and Simmondsiaceae) indicates that this condition is ancestral within the order. Likewise, anomalous floral development has been reported in several families (e.g., Aizoaceae, Caryophyllaceae, Phytolaccacaeae, and Portulacaceae [Ronse Decraene et al., 1998]) such that the apparently well-organized flowers of these plants, which are diplostemonous like the rosids, have arisen from a polymerous developmental pattern and should perhaps be best described as "pseudodiplostemonous." Although Caryophyllales appear in many respects to be typical advanced eudicots in terms of habit and floral structure, they clearly have independently developed these traits, which is consistent with their isolation from the asterid/rosid clade in the *atpB/rbcL* trees.

Groups of Uncertain Position

For several groups the affinities are still not clear. The position of Zygophyllaceae is not yet well supported, but the occurrence of anthroquinones (also found in Gentianales, Myrtales, and Lamiales) indicates that their position in the combined *atpB/rbcL* tree as sister to the nitrogen-fixing clade of eurosid I may be reasonable. The sister group relationship of Zygophyllaceae with Krameriaceae is well supported, although these two families are extremely divergent morphologically (Sheahan and Chase, 1996). A reticulate sexine might indicate a closer relationship with other eurosid I or II families in which this feature predominates. However, because the trait is presumably plesiomorphic, the psilate/granulate sexine found in the nitrogen-fixing clade does not exclude a sister group relationship for Zygophyllaceae/Krameriaceae.

Vitaceae likewise fit the rosid pattern; they have the nuclear endosperm development typical of eurosid I and II and a hypostase, which is present in all rosid/asterid families except those of Ericales. Dilleniaceae share nuclear endosperm development and a reticulate sexine with the eurosids, but a hypostase is absent, as in Santalales and Caryophyllales. Aextoxicaceae, Berberidopsidaceae, and Picramniaceae are poorly studied, which precludes any hypothesis about their relationships, although the last are clearly rosids. Berberidopsidaceae (two genera: *Berberidopsis* from temperate South America and *Streptothamnus* from eastern Australia), tentatively placed near the caryophyllids, are interesting because of their primitive wood and flowers, lacking clear differentiation of calyx and corolla (Miller, 1975). That all of these represent higher eudicots is clear on the basis of general floral organization (Berberidopsidaceae being the obvious ex-

ception). However, characters exhibited by these taxa are mixtures of the traits otherwise characterizing the four major clades: asterids, Caryophyllales, rosids, and Saxifragales. As mentioned above, the lack of isomery in many of these enigmatic taxa places them outside the rosids and asterids. Santalales are isomerous, but they vary from one to more than five whorls, perhaps indicating that they too are outside the core lineages. Vitaceae are obhaplostemonous, which clearly distinguishes them from both rosids and asterids, although this was the character emphasized by some (e.g., Cronquist, 1981) to link them to Rhamnaceae, which is a member of Rosales (eurosid I). Before an overall synthesis of floral evolution in eudicots can be developed, these problematic groups must be accurately placed into the general phylogenetic scheme for angiosperms.

Overall, the types of characters that mark clades at the interordinal levels (sensu APG, 1998) within the angiosperms are not those of gross morphology that have been the mainstay of nearly all previous classifications; stipules and stamen organization are obvious exceptions. The results of DNA sequence analyses provide evidence that many underused characters are of great systematic importance: phytochemistry, development, and anatomy. Seed anatomy (presence of hypostase) in particular appears to be a rich source of phylogenetically important information and should be more extensively studied. Considerably more chemotaxonomic work using the DNA phylogenies to focus attention on particular taxa and compounds may also prove useful.

There is little doubt that the intuitive classifications of the past with their emphasis on weighting of selected characters are not useful in either a phylogenetic or predictive context. As compendia of characters, treatments such as those of Cronquist (1981) and Takhtajan (1997) are useful (particularly if they also contain extensive literature citations, as in Takhtajan, 1997). Fortunately, the incorporation of large numbers of DNA sequences with the extensive literature from chemotaxonomic, developmental, and anatomical studies into a phylogenetic framework offers for the first time a robust alternative to evolutionary classifications.

Comparison with the Large 18S rDNA Phylogeny

Chase et al. (1993) stated that no specific sampling plan guided their study. These authors tried to compile all available *rbcL* sequences for this first broad-scale phylogenetic analysis of angiosperms. Consequently, some plant groups were oversampled, whereas others were poorly or not represented at all. Since then, many additional sequences have been collected (especially for some rare and geographically restricted taxa), and a huge literature has been produced, which guided the sampling for this analysis. Hence, we chose 357 taxa to represent all major lineages, but this presented us with the limitations inherent in analysis of such large matrices. There are $\sim 7.5 \times 10^{863}$ possible rooted trees and 10^{861} unrooted trees. Obviously, therefore, we cannot guarantee that we have found the shortest trees.

When we started to analyze the *atpB* sequences, we were amazed at how the new results matched the *rbcL* phylogeny. Because two plastid genes provide evidence for the same groups of families, despite the computational problems related to the sizes of the matrices, this may be taken as evidence that a clear historical pattern is being detected. Such consistency is unlikely to be due to chance alone. Conversely, the 18S rDNA trees (Soltis et al., 1997b) at least present only a degree of "soft" incongruence (Seelanan et al., 1997) with the two plastid trees. Application of the random partition test (Farris et al., 1995) indicated that incongruence between 18S rDNA and either *atpB* or *rbcL* is actually lower than between the two plastid genes (Soltis et al., 1997a, 1998).

The placement of *Ceratophyllum* remains problematic. That *Ceratophyllum* should be either a monocot or sister to the monocots (with *Acorus*), as in the *atpB* tree, seems less plausible than its position as sister to the rest of angiosperms, as in the *rbcL* and the combined *atpB/rbcL* trees. Ceratophyllaceae are a cosmopolitan family (comprising a single genus and ~2–30 species) of highly

specialized and reduced aquatics, lacking roots, cuticle, stomata, perianth, and woody tissues. Their affinities have been uncertain, but a placement near the base of the angiosperms compares favorably with their inaperturate pollen as well as with their floral features (Endress, 1994) and fossil record (Les, 1988).

Using 18S rDNA sequences, Soltis et al. (1997b) also presented as problematic the placement of the paleoherbs, Chloranthaceae, Aristolochiaceae, and Lactoridaceae, plus Winteraceae. We also found some discrepancies between the *atpB* and *rbcL* trees for these same groupings. Based on *atpB*, Winteraceae are not close to Magnoliales but instead go within the paleoherb group mentioned above. Based on *rbcL* or *atpB/rbcL*, Winteraceae are placed as sister to Magnoliales, and Chloranthaceae form an isolated lineage at the base of the remaining magnoliids. However, there is no bootstrap support >50% for any of these relationships in our trees.

General patterns among eudicots are highly congruent among all molecular studies published so far. Based either on the nuclear 18S rDNA or the plastid genes, Ranunculales, Saxifragales, and Caryophyllales are well defined (corresponding, respectively, to the ranunculids, saxifragoids, and Caryophyllidae s.l. of Soltis et al., 1997b) as are the two largest clades, the rosids and asterids. Hamamelidae and Dilleniidae sensu Cronquist (1981) or Takhtajan (1997) are highly polyphyletic in the 18S rDNA trees and in the trees presented here, and these concepts should no longer be maintained (we subsume Hamamelidaceae in Saxifragales, and Dilleniaceae are still unplaced as to order). All trees contain the glucosinolate clade (Brassicales), the nitrogen-fixing clade (Cucurbitales, Fabales, Fagales, and Rosales), and the subclades of euasterids, for which the "early" versus "late sympetaly" of Erbar and Leins (1996) fits well.

Soltis et al. (1997b) stated that "perhaps the most unusual consistent feature of the 18S rDNA trees involved the placement of Caryophyllidae s. l. within Asteridae s. l." Because the spines of all these separate trees, *rbcL*, *atpB*, and 18S, are without clear patterns, mostly because of their very short branches, accurate placement of Caryophyllales, Saxifragales, asterids, and rosids is not yet possible. In all the analyses here, Caryophyllales appear at the base of asterids plus rosids, but with simple branch removal and replacement experiments in MacClade (Maddison and Maddison, 1992), only a few additional steps (~1–5) are required to put Caryophyllales within any clade of asterids.

The basic problem of the 18S rDNA result is that by itself it provides little clear pattern. The 18S rDNA trees are highly unstable and have relatively low numbers of supported groups compared with those based on *rbcL* and *atpB* (Soltis et al., 1998; Chase and Cox, 1998; Hoot et al., 1995, 1999); the trees deviate largely because 18S rDNA has fewer variable positions experiencing most of the change (Chase and Cox, 1998). Nevertheless, 18S rDNA data make a valuable contribution to angiosperm phylogenetics. The sequences represent another genome, and in studies combining 18S rDNA, *atpB*, and *rbcL*, the three genes produce substantially stronger evidence of relationships than any of the separate analyses (Hoot et al., 1995, 1999; Soltis et al., 1997a, 1998; Chase and Cox, 1998). In spite of the differences in topology, the patterns of variation in 18S rDNA appear to be the same as those in *rbcL* and *atpB*.

Conflict with Evolutionary Classifications

In a recently proposed system of angiosperm classification, Takhtajan (1997:3) entirely dismissed DNA studies, stating that "these [DNA studies] often point to relationships that are clearly not compatible with other data and sometimes even quite outside the realm of possibility. . . . Besides the random noise in DNA sequences, molecular characters are subject to evolutionary convergence, parallelism, and reversal; therefore molecular methods are not a panacea. Molecular evidence should be used with, not in place of, morphological evidence". The published molecular results are not incompatible with morphological evidence itself but rather are incongruent with some interpretations of morphology. When analyzed phylogenetically (Nandi et al., 1998), non-DNA data produced patterns similar to those found with DNA sequences. When Takhtajan (1997:3) stated that DNA

studies produced results that were "sometimes even quite outside the realm of possibility", he clearly implied that these DNA-based patterns were erroneous simply because they were in conflict with other data that he felt were more accurate. However, several morphological studies have been published that support DNA trees (e.g., see Patterson, 1988; Atcheley and Fitch, 1991; Novacek, 1992; Patterson et al., 1993; Hoot et al., 1995, 1997; Erbar and Leins, 1996; Spichiger and Savolainen, 1997; Hoot and Douglas, 1998; Rudall et al., 1998).

We can imagine that certain DNA results upset those who hold traditional perspectives. For example, it is easy to see that *Nelumbo* is not morphologically similar to *Platanus* and Proteaceae, although molecular data indicate they form a clade; but then, what alternative is more robust? Takhtajan classified *Nelumbo* in a subclass of its own, Nelumbonideae; this is not a refutation of the DNA trees but rather an admission that he has no evidence of what its affinities might be. A high frequency of small suprafamilial taxa is an admission of a lack of alternative hypotheses. Takhtajan's new system (1997) recognized two monofamilial and one bifamilial subclasses, 18 monofamilial and 11 bifamilial superorders, and 135 monofamilial and 26 bifamilial orders. The number of monogeneric families among the 591 he recognized is the highest of any published system. Such high numbers of small taxa reduce the information content of his classification and leave the impression that little is known about their higher-level relationships.

Gene sequences can create strongly supported patterns without being influenced by the multiple morphological convergences, reversals, and extreme morphological specializations that have occurred during plant evolution. For example, *Aextoxicon* was found with molecular data to be strongly supported in all three analyses as sister to *Berberidopsis* (Figs. 5g,n, 6d), but no obvious morphological features link these two genera. They probably represent specialized relics from ancient floras for which the intermediate linking taxa have simply disappeared. *Nelumbo* and its relationships to *Platanus* and Proteaceae are another example of how molecular information provides new and radically different hypotheses that might resolve long-standing impasses.

In the trees presented here, the results are basically of two types, which set the stage for future research: (1) groups that are well supported and sometimes drastically different from classical views, which merit further study because some evidence of this pattern is likely to have been retained during evolution (e.g., micromorphology, biochemistry, or palynology); and (2) groups that are weakly supported for which additional data are required to define their position accurately.

An example of the first category is the clade of families here named Malpighiales; no one had previously suggested a group composed of these families, and there is no obvious suite of characters that unites all these families. Similarly, Saxifragales appear as a strongly supported clade, but their circumscription based on DNA analysis differs from that suggested by any previous author. A parallel example can be seen with the families of Brassicales: No previous taxonomist ever included all these families in the same group except Dahlgren (1975). The single clue to the common ancestry of Brassicales is mustard oils.

Groups that fall into the second category noted above include Zygophyllaceae/Krameriaceae (Sheahan and Chase, 1996), which appear in different places in every analysis conducted, including those presented here, but never with high bootstrap support. They most often appear as the sister of the nitrogen-fixing families, as in the combined analysis (Fig. 6e). The combination of being fairly sequence-divergent (37 hypothesized substitutions) and low divergence near the rosid I and rosid II split (3, 4, or 7 hypothesized substitutions; Fig. 6e) makes a robust placement unlikely until more sequence data are available. Dilleniaceae, Vitaceae, Picramniaceae, and Santalales also fall into this category.

Prospects for Finding an Accurate Angiosperm Tree

The flowering plants represent one of the largest groups of organisms, comprising >250,000 species in ~13,000 genera and 500 families. The analysis of large data sets containing hundreds of taxa is the only way to

address the phylogeny of such a large diverse group. If obtaining only a reasonably optimal tree is unsatisfactory, then we are at an impasse. Further progress is precluded until new methods of analysis are developed or substantially greater amounts of data are available. However, an inability to recover the shortest tree is definitely less of a problem than has previously been maintained.

If we can assume that the accuracy of a tree derived from combined data sets is improved over those of single genes (e.g., because the strength of the phylogenetic signal has been enhanced when multiple genes are used; see Chase and Cox, 1998; Soltis et al., 1998; Hoot et al., 1999), then the shortest trees based on the single genes are actually underestimates of the length of the true phylogeny. Our mapping experiments support this hypothesis: The individual *atpB* and *rbcL* trees are each roughly 0.6% and 0.9% (respectively) too short relative to the combined tree. If we map the *atpB* characters onto the topology produced by the combined *atpB/rbcL* matrix, the length is 110 steps longer than that from the analysis of *atpB* alone (13,089 vs. 12,979). Similarly, if we map the *rbcL* characters onto the combined topology (Fig. 7), 75 extra steps are found compared with the *rbcL*-alone analysis (12,847 vs. 12,772 steps). In neither case are there important differences in CI and RI. Although this may not seem like a substantial underestimate, it is in the range required to move *Ceratophyllum* from sister to *Acorus* to sister to the rest of the angiosperms, or vice versa. Figure 7 presents the starting tree lengths of each of 5,000 replicates (genes alone or in combination) as the percentage longer than the length of the most-parsimonious tree. In addition, we have indicated how much longer are the trees depicted from each individual gene when optimized onto the combined topology. These percentages represent the lower limits of the starting trees obtained for *atpB* or *rbcL* trees without swapping: Beyond these limits, further swapping leads to trees that are too short, and a more accurate topology (as found with the combined matrix) can never be recovered by continued search for yet shorter trees (see Fig. 7).

Given that our goal is to find the true tree for angiosperms, there is no point spending extensive computing time in the attempt to find shorter trees based on any of the single genes (contrary to the admonitions of Rice et al., 1998). Rather, this analysis of a combined *atpB/rbcL* data set, as well as other recent analyses of *atpB* /*rbcL* /18S rDNA (Chase and Cox, 1998; Soltis et al., 1998) indicate clearly that our efforts are better placed in sequencing more taxa and genes. Finding yet shorter trees for individual genes will never recover new groups with high levels of internal support; all such well-supported groups are present in the starting trees before swapping begins. Groups that require extensive swapping to be found have a high probability of being spurious because single-gene trees inevitably are underestimates. The only relationships that we can be confident about are those that have high internal support, and performing a bootstrap analysis does not first require swapping to find the shortest tree.

Because we can demonstrate that a data set with weak phylogenetic patterns leads to underestimates of levels of homoplasy, we are also suspicious that all optimality criteria are unreliable if the patterns in the data are weak because they will also inevitably underestimate tree length (no matter which algorithm is used). A distance-based algorithm can correct the distances, but it cannot correct the relationships any better than we can when we know that the *atpB* tree produced by parsimony analysis is an underestimate. We know that homoplasy has been underestimated, but we cannot know where the missed steps should be added to make a more accurate tree. Similarly, although we know roughly that the more accurate tree(s) fall near the shorter limit of the starting trees, we can suggest no method of winnowing out this tree from the undoubtedly thousands of trees at this length. Without evidence, either from internal support or congruence (these are highly correlated, e.g., in Soltis et al., 1998), we are in a quagmire from which a rigorous analysis cannot extricate us.

If we had spent another year of computing time on analyses of each of the two individual genes, we almost certainly would have found shorter trees, but these would have added nothing to our knowledge of angiosperm relationships. Patterns for which

there is bootstrap support <50% (or congruence with other analyses) are unreliable even if the shortest tree(s) can be found. This is not an argument against the application of parsimony or other optimality criteria, but rather an acknowledgment of the limitations of data sets that lack clear patterns (i.e., those with too few or conflicting characters).

Conclusion

Our purpose in this paper was not only to produce a plastid tree for the angiosperms but also to bring into focus a series of issues pertaining to the phylogeny of flowering plants and the analysis of large data sets. Here, by adding another plastid gene (*atpB*) for an extended sampling of flowering plants, we performed phylogenetic analyses in which numerous clades are (1) highly supported (>80% bootstrap) on the combined analysis of two plastid genes, (2) identified in the most-parsimonious trees recovered when analyzing each matrix independently, and (3) congruent with many of the clades identified by 18S rDNA. This provides convincing evidence that such analyses can provide the framework for a new classification of flowering plants that is based largely on the DNA patterns but has a great deal of corroboration from other lines of evidence (APG, 1998). These DNA-based trees are in close agreement with many morphological, anatomical, and chemical characters that were underused in earlier classifications (e.g., Cronquist, Takhtajan, etc.).

Furthermore, we fully agree that large data sets are not as tractable as we would wish, but the increasing availability of computer power will certainly permit further analytical improvements; further, however, we have demonstrated that not obtaining the shortest possible trees is also less important than previously recognized. In spite of limitations, the large analyses thus far published demonstrate a level of congruence that is inconsistent with the notion that simply because they are large they must be highly flawed. We expect the next 5 years of molecular systematics to usher in even faster and more drastic change. As a result of this exciting period of change, evolutionary and ecological processes will be addressed in ways previously considered speculative (e.g., see Sanderson and Donoghue, 1996; Savolainen and Goudet, 1998). In short, analyses of large data sets are not only feasible but have proven to be robust. Combining both *atpB* and *rbcL*, which together represent only ~3,000 bp (with about half of these variable in two or more taxa, i.e., 1,521 sites), was useful because a clearer phylogenetic signal was detected. Yet, more characters are needed to infer robustly the branching pattern of ancient lineages such as Ceratophyllaceae and the magnoliids. Some of what might be phyletic radiations (for example, the polytomies composed of asterids, Berberidopsidaceae/Aextoxicaceae, Caryophyllales, Dilleniaceae, Santalales, Saxifragales, and Vitaceae) also require more data to be adequately estimated. The addition of 18S rDNA sequences undoubtedly greatly improved the results (e.g., Soltis et al., 1998; Hoot et al., 1999; also see Soltis et al. [1999] for a recent phylogenetic analysis of the angiosperms based on *rbcL*, *atpB* and 18S rDNA combined), but other molecular information can be useful, for example, patterns of genome organization (Qiu et al., 1998).

Now that we can see that the future holds the possibility for producing a robust phylogeny by the direct combination of several genes, we can also expect to improve the models of molecular evolution by optimizing various features on these well-supported phylogenetic relationships; however, the use of models to develop accurate phylogenetic estimates is itself circular and can prevent the recognition of patterns that contradict models. If individual matrices are inadequate to produce robust estimates of relationships, then estimating probabilities from the data themselves can mislead. If clear patterns of relationships can be estimated without resorting to lengthy or assumption-laden methods of analysis, particularly for large matrices such as these, why then would anyone favor particular results or have greater confidence in them just because they are based on application of a specific optimality criterion? We prefer robustness and congruence as the best measures of accuracy, and simple tree-building algorithms are more than adequate for these purposes. We need to know in more

detail how the commonly sequenced genes, 18S rDNA, *atpB*, and, *rbcL* evolve so that specific models of their evolution can be designed, but this should come after we understand relationships of the taxa, not before (e.g., Soltis and Soltis, 1998; Chase and Albert, 1998; Olmstead et al., 1998).

Sampling still needs further improvement, and representatives of several lineages are still missing here. Addition of more taxa will not solve the problems at the base of the tree (simply because we have already sampled surviving lineages of these families thoroughly); rather, we need more accurate ideas of character and life-history evolution. For these purposes, as well as for the production of a comprehensive classification of the angiosperms, we must have all lineages and their positions identified.

Ultimately, we envisage a future in which we worry primarily about managing the vast wealth of systematic data available to us (including nonmolecular as well as DNA sequences, and in particular the electronic files of the raw data produced by the automated sequencers that have made this upsurge of sequences possible). We can expect not only to produce large data sets but also to be able to analyze them in a robust manner. Clearly our biggest problems are not methodological or theoretical. They are instead mundane and unexciting and consist of how we document, manage, and communicate the vast amounts of systematic information that we are set up to produce and need to integrate. Compared with these problems, phylogenetic analysis of even larger matrices is becoming a relatively simple and straightforward task.

ACKNOWLEDGMENTS

We thank David Swofford for the use of test versions of PAUP 4.0; Greg L'Hoste and Jérôme Goudet for their help with some calculations; Ken Wolfe for the LI93 program; and Jim Doyle, Toby Kellogg, Manuel Ruedi, an anonymous reviewer, and the Editor, Dick Olmstead, for many useful comments on the manuscript. We also thank all persons and institutes who helped us acquire plant samples, and the Royal Botanic Gardens, Kew, the Royal Society (UK), the Swiss National Science Foundation (31–52378.97 to V.S.), the National Environment Research Council (UK; to C.M.), the Academic Society of Geneva, and the National Science Foundation (USA; DEB-9596011 to S.B.H., DEB-9707868 to Pamela Soltis and D.S.), which funded this study.

REFERENCES

ALBERT, V. A., M. W. CHASE, AND B. D. MISHLER. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. Ann. Mo. Bot. Gard. 80:752–766.

ALBERT, V. A., S. E. WILLIAMS, AND M. W. CHASE. 1992. Carnivorous plants: Phylogeny and structural evolution. Science 257:1491–1495.

ALLARD, M. W., AND J. M. CARPENTER. 1996. On weighting and congruence. Cladistics 12:183–198.

ALVERSON, W. S., K. G. KAROL, D. A. BAUM, M. W. CHASE, S. M. SWENSON, R. MCCOURT, AND K. J. SYTSMA. 1998. Circumscription of the Malvales and relationships to other Rosidae: Evidence from *rbcL* sequence data. Am. J. Bot. 85:876–887.

ANGIOSPERM PHYLOGENY GROUP. 1998. An ordinal classification of the families of flowering plants. Ann. Mo. Bot. Gard. 85:531–553.

ATCHEY, W. R., AND W. M. FITCH. 1991. Gene trees and the origins of inbred strains of mice. Science 254:554–558.

BAKKER, F. T., D. D. VASSILIADES, C. M. MORTON, AND V. SAVOLAINEN. 1998. Phylogenetic relationships of *Biebersteinia* Stephan (Geraniaceae) inferred from *rbcL* and *atpB* sequence comparisons. Bot. J. Linn. Soc. 127:149–158.

BAYER, C., M. F. FAY, A. Y. DE BRUIJN , V. SAVOLAINEN, C. M. MORTON, K. KUBITZKI, W. S. ALVERSON, AND M. W. CHASE. 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: A combined analysis of plastid *atpB* and *rbcL* sequences. Bot. J. Linn. Soc. 129:267–303.

BOUSQUET, J., S. H. STRAUS, A. H. DOERKSEN, AND R. A. PRICE. 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proc. Natl. Acad. Sci. USA 89:7844–7848.

BREMER, B., K. ANDREASEN, AND D. OLSSON. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbcL* sequence data. Ann. Mo. Bot. Gard. 82:383–397.

BREMER, K. 1988. The limits of amino acid sequence in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

CHASE, M. W., AND V. A. ALBERT. 1998. A perspective on the contribution of plastid *rbcL* DNA sequences to angiosperm phylogenetics. Pages 488–507 *in* Molecular systematics of plants II (D. E. Soltis, P. S. Soltis, and J. A. Doyle, eds). Kluwer, London.

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

CHASE, M. W., C. M. MORTON, AND J. KALLUNKI. 1999. Molecular phylogenetics of Rutaceae: Evidence from combined plastid *atpB* and *rbcL*. Am. J. Bot. 86:1191–1199.

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.-J. KIM, C. F. WIMPEE, J. F. SMITH, G. R. FURNIER, 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. BARRETT, S. DAYANANDAN, AND V. A. ALBERT. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Mo. Bot. Gard. 80:528–580.

CHASE, M. W., D. E. SOLTIS, P. S. SOLTIS, P. J. RUDALL, M. F. FAY, W. H. HAHN, S. SULLIVAN, J. JOSEPH, T. J. GIVINISH, K. J. SYTSMA, C. PIRES, AND J. S. FARRIS. 2000. Higher-level systematics of the monocotyledons: An assessment of current knowledge and a new classification. Proc. Second Int. Symp. on the Comparative Biology of the Monocotyledons. CSIRO, Sydney, Australia (in press).

CHASE, M. W., D. W. STEVENSON, P. WILKIN, AND P. J. RUDALL. 1995. Monocotyledon systematics: A combined analysis. Pages 685–730 *in* Monocotyledons: Systematics and evolution (P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries, eds). Royal Botanic Gardens, Kew, U.K.

CONTI, E., A. LITT, AND K. J. SYTSMA. 1996. Circumscription of Myrtales and their relationships to other rosids: Evidence from *rbcL* sequence data. Am. J. Bot. 83:221–233.

CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York.

DAHLGREN, R. M. T. 1975. The distribution of characters within an angiosperm system. Bot. Not. 128:181–197.

DAVIS, J. I., P. SIMMONS, D. W. STEVENSON, AND J. F. WENDEL. 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: An example from the monocotyledons using mitochondrial atpA sequences. Syst. Biol. 47:282–310.

DENO, H., K. SHINOZAKI, AND M. SUGIURA. 1983. Nucleotide sequence of tobacco chloroplast gene for the alpha subunit of proton-translocating ATPase. Nucleic Acids Res. 11:2185–2191.

D'ERCHI, A. M., C. GISSI, G. PESOLE, C. SACCONE, AND U. ARNASSON. 1996. The guinea-pig is not a rodent. Nature 381:597–599.

DONOGHUE, M. J. 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Mo. Bot. Gard. 81:405–418.

DONOGHUE, M. J., AND J. A. DOYLE. 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. Pages 17–45 *in* Evolution, systematics and fossil history of the Hamamelidae (P. R. Crane and S. Blackmore, eds). Clarendon Press, Oxford, U.K.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbcL* sequences. Ann. Mo. Bot. Gard. 79:333–345.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

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, AND G. H. LEARN. 1993. Phylogenetic hypothesis for the monocotyledons constructed from *rbcL* sequence data. Ann. Mo. Bot. Gard. 80:607–619.

ENDRESS, P. K. 1994. Evolutionary aspects of the floral structure in *Ceratophyllum*. Pl. Syst. Evol. S8: 175–183.

ENDRESS, P. K., AND A. IGERSHEIM. 1997. Gynoecium diversity and systematics of the Laurales. Bot. J. Linn. Soc. 125:93–168.

ERBAR, C., AND P. LEINS. 1996. Distribution of the character state "early sympetaly" and "late sympetaly" within the "sympetalae tetracyclicae" and presumably allied groups. Bot. Acta 109:427–440.

FARRIS, J. S. 1989. The retention index and the rescaled consistency index. Cladistics 5:417–419.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Testing significance of incongruence. Cladistics 10:315–319.

FAY, M. F., C. BAYER, W. S. ALVERSON, A. Y. DE BRUIJN, AND M. W. CHASE. 1998. Plastid *rbcL* sequence data indicate a close affinity between *Diegodendron* and *Bixa*. Taxon 47:43–50.

FAY, M. F., K. M. CAMERON, G. T. PRANCE, M. D. LLEDÓ, AND M. W. CHASE. 1997. Familial relationships of *Rhabdodendron* (Rhabdodendraceae): Plastid *rbcL* sequences indicate a caryophyllid placement. Kew Bull. 52:923–932.

FELSENSTEIN, J. 1985. Confidence limits on phylogenetics: An approach using the bootstrap. Evolution 39:783–791.

FERNANDO, E. S., P. A. GADEK, D. M. CRAYN, AND C. J. QUINN. 1995. Simaroubaceae, an artificial construct: Evidence from *rbcL* sequence variation. Am. J. Bot. 82:344–350.

FITCH, W. M. 1971. Toward defining the course of evolution: Minimal change for a specific tree topology. Syst. Zool. 20:406–416.

GADEK, P. A., E. S. FERNANDO, C. J. QUINN, S. B. HOOT, T. TERRAZAS, M. C. SHEANAN, AND M. W. CHASE. 1996. Sapindales: Molecular delimitation and infraordinal groups. Am. J. Bot. 83:802–811.

GIANNASI, D. E., G. ZURAWSKI, G. LEARN, AND M. T. CLEGG. 1992. Evolutionary relationships of the Caryophyllales based on comparative *rbcL* sequences. Syst. Bot. 17:1–15.

GOLENBERG, E. M., D. E. GIANNASI, M. T. CLEGG, C. J. SMILEY, M. DURBIN, D. HENDERSON, AND G. ZURAWSKI. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. Nature 344:656–658.

GOLOBOFF, P. A. 1996. Methods for faster parsimony analysis. Cladistics 12:199–220.

GRAUR, D., W. A. HIDE, AND W.-H. LI. 1991. Is the guinea-pig a rodent? Nature 351:649–651.

GRAYBEAL, A. 1993. The phylogenetic utility of cytochrome *b*: Lessons from bufonid frogs. Mol. Phylogenet. Evol. 3:256–269.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.

HAMBY, K. R., AND E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Pages 50–91 *in* Molecular systematics of plants (P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds). Chapman and Hall, New York.

HAUSER, D. L., AND G. BOYAJIAN. 1997. Proportional change and patterns of homoplasy: Sanderson and Donoghue revisited. Cladistics 13:97–100.

HEMPEL, A. L., P. A. REEVES, R. G. OLMSTEAD, AND R. K. JANSEN. 1995. Implication of *rbcL* sequence data for higher order relationships of the Loasaceae and the anomalous aquatic plant *Hydrostachys* (Hydrostachyaceae). Plant Syst. Evol. 194:25–37.

HILLIS, D. M. 1995. Approaches for assessing phylogenetic accuracy. Syst. Biol. 44:3–16.

HILLIS, D. M. 1996. Inferring complex phylogenies. Nature 383:130–131.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47:3–8.

HILLIS, D. M., J. P. HUELSENBECK, AND C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. Science 264:671–677.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C. R. SUN, B. Y. MENG, Y. Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, AND M. SUGIURA. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Mol. Gen. Genet. 217:185–194.

HOOT, S. B., AND P. R. CRANE. 1995. Interfamilial relationships in the Ranunculidae based on molecular systematics. Plant Syst. Evol. S9:119–131.

HOOT, S. B., A. CULHAM, AND P. R. CRANE. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: Comparison within *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. Ann. Mo. Bot. Gard. 82:194–207.

HOOT, S. B., AND A. DOUGLAS. 1998. Phylogeny of the Proteaceae based on *atpB* and *atpB/rbcL* intergenic spacer region sequences. Aust. J. Bot. 46:301–320.

HOOT, S. B., J. W. KADEREIT, F. R. BLATTNER, K. B. JORK, A. E. SCHWARZBACH, AND P. R. CRANE. 1997. Data congruency and phylogeny of the Papaveraceae s.l. based on four data sets: *atpB* and *rbcL* sequences, *trnK* restriction sites, and morphological characters. Syst. Bot. 22:575–590.

HOOT, S. B., S. MEGALLON-PUEBLA, AND P. R. CRANE. 1999. Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequences. Ann. Mo. Bot. Gard. 86:1–32.

HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247–264.

HUYNH, K.-L. 1976. L'arrangement du pollen du genre *Schisandra* (Schisandraceae) et sa signification phylogénétique chez les angiospermes. Beitr. Biol. Pflanz. 52:227–253.

JOHNSON, L. A., AND D. E. SOLTIS. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Syst. Bot. 19:143–156.

KÄSS, E., AND M. WINK. 1995. Molecular phylogeny of the Papilionoideae (family Leguminosae): *rbcL* sequences versus chemical taxonomy. Bot. Acta 108: 149–162.

KAUFMANN, M., AND M. WINK. 1994. Molecular systematics of the Nepetoideae (family Labiatae): Phylogenetic implications from *rbcL* gene sequences. Biosci. Rep. 49:635–645.

KELLOGG, E. A., AND H. P. LINDER. 1995. Phylogeny of Poales. Pages 511–542 *in* Monocotyledons: Systematics and evolution (P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries, eds). Royal Botanic Gardens, Kew, U.K.

KIM, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. Syst. Biol. 47:43–60.

KIM, K.-J., R. K. JANSEN, R. S. WALLACE, H. J. MICHAELS, AND J. D. PALMER. 1992. Phylogenetic implications of *rbcL* sequence variation in Asteraceae. Ann. Mo. Bot. Gard. 79:428–445.

KRON, K. A., AND M. W. CHASE. 1993. Systematics of the Ericaceae, Empetraceae, Epacridaceae and related taxa based upon *rbcL* sequence data. Ann. Mo. Bot. Gard. 80:735–741.

KUMAR, S., K. TAMURA, AND M. NEI. 1993. MEGA: Molecular evolutionary genetics analysis, version 1.0.1. Pennsylvania State Univ., University Park.

LECOINTRE, G., H. PHILIPPE, H. L. VÂN LÊ, AND H. LE GUYADER. 1993. Species sampling has a major impact on phylogenetic inference. Mol. Phylogenet. Evol. 2:205–224.

LES, D. H. 1988. The origin and affinities of Ceratophyllaceae. Taxon 37:326–345.

LES, D. H., D. K. GARVIN, AND C. F. WIMPEE. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proceedings of the National Academy of Sciences USA:10119–10123.

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

LI, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36:96–99.

LIN, C. M., Z. Q. LIU, AND S. D. KUNG. 1986. *Nicotiana* chloroplast genome: X. Correlation between the DNA sequences and the isoelectric focusing patterns of the LS of *Rubisco*. Plant Mol. Biol. 6:81–87.

LLEDÓ, M. D., M. B. CRESPO, K. M. CAMERON, M. F. FAY, AND M. W. CHASE. 1998. Systematics of Plumbaginaceae based upon cladistic analysis of *rbcL* sequence data. Syst. Bot. 23:21–29.

LOCONTE, H., AND D. W. STEVENSON. 1991. Cladistics of the Magnoliidae. Cladistics 7:267–296.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution. Version 3.04. Sinauer, Sunderland, Massachusetts.

MAIER, R. M., K. NECKERMANN, G. L. IGLOI, AND H. KOESSEL. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Mol. Biol. 251:614–628.

MANEN, J.-F., AND A. NATALI. 1995. Comparison of the evolution of ribulose-1,5-bisphosphate carboxylase (*rbcL*) and *atpB-rbcL* noncoding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). J. Mol. Evol. 41:920–927.

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

MEYER, A. 1994. Shortcomings of the cytochrome *b* gene as a molecular marker. Trends Ecol. Evol. 9: 278–280.

MICHALOWSKI, C. B., H. J. BOHNERT, D. F. KLESSIG, AND J. O. BERRY. 1990. Nucleotide sequence of *rbcL* from *Amaranthus hypochondriacus*. Nucleic Acids Res. 18:2187–2187.

MILLER, R. B. 1975. Systematic anatomy of the xylem and comments on the relationships of the Flacourtiaceae. J. Arnold Arbor. Harv. Univ. 56:20–102.

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

MORGAN, D. R., AND D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbcL* sequence data. Ann. Mo. Bot. Gard. 80:1152–1154.

MORTON, C. M., M. W. CHASE, K. A. KRON, AND S. A. SWENSEN. 1996. A molecular evaluation of the monophyly of the order Ebenales based upon *rbcL* sequence data. Syst. Bot. 21:567–586.

MORTON, C. M., K. G. KAROL, AND M. W. CHASE. 1997. Taxonomic affinities of *Physena* (Physenaceae) and *Asteropeia* (Theaceae). Bot. Rev. 63:231–239.

NANDI, O., M. W. CHASE, AND P. K. ENDRESS. 1998. A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. Ann. Mo. Bot. Gard. 85:137–212.

NAYLOR, G. J. P., AND W. M. BROWN. 1997. Structural biology and phylogenetic estimation. Nature 388:527–528.

NAYLOR, G. J. P., AND W. M. BROWN. 1998. *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47:61–76.

NICKRENT, K. A., AND D. E. SOLTIS. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcL* sequences. Ann. Mo. Bot. Gard. 81:484–533.

NOVACEK, M. J. 1992. Mammalian phylogeny: Shaking the tree. Nature 356:121–125.

OLMSTEAD, R. G., B. BREMER, K. M. SCOTT, AND J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbcL* sequences. Ann. Mo. Bot. Gard. 80:700–722.

OLMSTEAD, R. G., H. J. MICHAELS, K. M. SCOTT, AND J. D. PALMER. 1992. Monophyly of the Asteridae sensu lato and identification of their major lineages inferred from DNA sequences of *rbcL*. Ann. Mo. Bot. Gard. 79:249–265.

OLMSTEAD, R. G., P. A. REEVES, AND A. C. YEN. 1998. Pattern of sequence evolution and implication for parsimony analysis of chloroplast DNA. Pages 164–187 *in* Molecular systematics of plants II (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds). Kluwer, London.

PATTERSON, C. 1988. Homology in classical and molecular biology. Mol. Biol. Evol. 5:603–625.

PATTERSON, C., D. M. WILLIAMS, AND C. J. HUMPHRIES. 1993. Congruence between molecular and morphological phylogenies. Annu. Rev. Ecol. Syst. 24: 153–188.

PRICE, R. A., AND J. D. PALMER. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from *rbcL* sequences comparison. Ann. Mo. Bot. Gard. 80:661–671.

QIU, Y.-L., M. W. CHASE, D. H. LES, AND C. R. PARKS. 1993. Molecular phylogenetics of the Magnoliidae: Cladistic analysis of nucleotide sequences of the plastid gene *rbcL*. Ann. Mo. Bot. Gard. 80:587–606.

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

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 1999. The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407.

RETTIG, J. H., H. D. WILSON, AND J. M. MANHART. 1992. Phylogeny of the Caryophyllales—Gene sequence data. Taxon 41:201–209.

RICE, K. A., M. J. DONOGHUE, AND R. G. OLMSTEAD. 1998. Analyzing large data sets: *rbcL* 500 revisited. Syst. Biol. 46:554–563.

RITLAND, K., AND M. T. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. Am. Nat. 130:S74–S100.

RODMAN, J., R. PRICE, K. KAROL, E. CONTI, K. SYTSMA, AND J. D. PALMER. 1993. Nucleotide sequences of the *rbcL* gene indicate monophyly of mustard oil plants. Ann. Mo. Bot. Gard. 80:686–699.

RONSE DECRAENE, L. P., E. F. SMETS, AND P. VANVINCKENROYE. 1998. Pseudodiplostemony, and its implications for the evolution of the androecium in the Caryophyllaceae. J. Plant Res. 111:25–43.

RUDALL, P. J., M. W. CHASE, D. F. CUTLER, J. RUSBY, AND A. Y. DE BRUIJN. 1998. Anatomical and molecular systematics of Asteliaceae and Hypoxidaceae. Bot. J. Linn. Soc. 127:1–42.

SANDERSON, M. J., AND M. J. DONOGHUE. 1996. Reconstructing shifts in diversification rates on phylogenetic trees. Trends Evol. Ecol. 11:15–20.

SAVOLAINEN, V., P. CUENOUD, R. SPICHIGER, M. D. P. MARTINEZ, M. CREVECOEUR, AND J.-F. MANEN. 1995. The use of herbarium specimens in DNA phylogenetics: Evaluation and improvement. Plant Syst. Evol. 197:87–98.

SAVOLAINEN, V., AND J. GOUDET. 1998. Rate of gene sequence evolution and species diversification in flowering plants: A re-evaluation. Proc. R. Soc. Lond. B 265:603–607.

SAVOLAINEN, V., R. SPICHIGER, AND J.-F. MANEN. 1997. Polyphyletism of Celastrales deduced from a chloroplast noncoding DNA region. Mol. Phylogenet. Evol. 7:145–157.

SEELANAN, T., A. SCHNABEL, AND J. WENDEL. 1997. Congruence and consensus in the cotton tribe (Malvaceae). Syst. Bot. 22:259–290.

SHEAHAN, M. C., AND M. W. CHASE. 1996. A phylogenetic analysis of Zygophyllaceae R. Br. based on morphological, anatomical, and *rbcL* DNA sequence data. Bot. J. Linn. Soc. 122:279–300.

SHINWARI, Z. K., H. KATO, R. TERAUCHI, AND S. KAWANO. 1994. Phylogenetic relationships among genera in the Liliaceae–Asparagoideae–Polygonatae s.l. inferred from *rbcL* gene sequence data. Plant Syst. Evol. 192:263–277.

SIDDALL, M. E., AND A. G. KLUGE. 1997. Probabilism and phylogenetic inference. Cladistics 13:313–336.

SOLTIS, D. E., C. HIBSCH-JETTER, P. S. SOLTIS, M. W. CHASE, AND J. S. FARRIS. 1997a. Molecular phylogenetic relationships among angiosperms: An overview based on *rbcL* and 18S rDNA sequences. Pages 157–178 *in* Evolution and diversification of land plants (K. Iwatsuki and P. H. Raven, eds). Springer-Verlag, Tokyo.

SOLTIS, D. E., D. R. MORGAN, A. GRABLE, P. S. SOLTIS, AND R. KUZOFF. 1993. Molecular systematics of Saxifragaceae sensu stricto. Am. J. Bot. 80:1056–1081.

SOLTIS, D. E., M. E. MORT, P. S. SOLTIS, C. HIBSH-JETTER, E. A. ZIMMER, AND D. MORGAN. 1999. Phylogenetic relationships of the enigmatic angiosperm family

Podostemaceae inferred from 18S rDNA and *rbcL* sequence data. Mol. Phylogenet. Evol. 11:261–272.

SOLTIS, D. E., AND P. S. SOLTIS. 1998. Phylogenetic analysis of Saxifragaceae sensu lato: A comparison of topologies based in 18S rDNA and *rbcL* sequences. Am. J. Bot. 84:504–522.

SOLTIS, D. E., P. S. SOLTIS, M. T. CLEGG, AND M. DURBIN. 1990. *rbcL* sequence divergence and phylogenetic relationships in Saxifragaceae sensu lato. Proc. Natl. Acad. Sci. USA 87:4640–4644.

SOLTIS, D. E., P. S. SOLTIS, D. R. MORGAN, S. M. SWENSEN, B. C. MULLIN, J. M. DOWD, AND P. G. MARTIN. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc. Natl. Acad. Sci. USA 92:2647–2651.

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

SOLTIS, D. E., P. S. SOLTIS, D. L. NICKRENT, L. A. JOHNSON, W. J. HAHN, S. B. HOOT, J. A. SWEERE, R. K. KUZOFF, K. A. KRON, M. W. CHASE, S. M. SWENSEN, E. A. ZIMMER, S.-M. CHAW, L. J. GILLESPIE, W. J. KRESS, AND K. J. SYTSMA. 1997b. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann. Mo. Bot. Gard 84:1–49.

SOLTIS, P. S., D. E. SOLTIS, AND M. W. CHASE. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404.

SORENG, R. J., AND J. I. DAVIS. 1998. Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of morphological and chloroplast DNA restriction site character sets. Bot. Rev. 64:1–85.

SPICHIGER, R., AND V. SAVOLAINEN. 1997. Present state of Angiospermae phylogeny. Candollea 52:435–455.

SWENSEN, S. M. 1996. The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. Am. J. Bot. 83:1503–1512.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding region of chloroplast DNA. Plant Mol. Biol. 17:1105–1109.

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

WAKASUGI, T., J. TSUDZUKI, S. ITO, K. NAKASHIMA, T. TSUDZUKI, AND M. SUGIURA. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA 91:9794–9798.

WATSON, L., AND M. J. DALLWITZ. 1991. The families of angiosperms: Automated descriptions, with interactive identification and information retrieval. Aust. Syst. Bot. 4:681–695.

XIANG, Q.-Y., D. E. SOLTIS, D. R. MORGAN, AND P. S. SOLTIS. 1993. Phylogenetic relationships of *Cornus* L. sensu lato and putative relatives inferred from *rbcL* sequence data. Ann. Mo. Bot. Gard. 80:723–734.

YANG, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42:587–596.

ZARDOYA, R. AND A. MEYER. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13:933–942.

ZURAWSKI, G., W. BOTTOMLEY, AND P. R. WHITFELD. 1982. Structure of the genes for the beta and epsilon subunits of the spinach chloroplast ATPase indicates a dicistronic mRNA and an overlapping translation stop/start signal. Proc. Natl. Acad. Sci. USA 79:6260–6264.

ZURAWSKI, G., P. R. WHITFELD, AND W. BOTTOMLEY. 1986. Sequence of the genes for the beta and epsilon subunits of ATP synthase from pea chloroplasts. Nucleic Acids Res. 14:3974.

ZURAWSKI, G., B. PERROT, W. BOTTOMLEY, AND P. R. WHITFELD. 1981. The structure of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res. 9:3251–3270.

Received 10 November, 1998; accepted 9 August, 1999
Associate Editor: R. Olmstead

APPENDIX

List of taxa used in this study, with vouchers, citation information, and EMBL/GenBank accession numbers.

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Acanthaceae	*Thunbergia coccinea* Wall.	Chase 2539 K	this paper	AJ235625	*Thunbergia usambarica* Lindau	Chase et al., 1993	L12956
Acoraceae	*Acorus calamus* L.	Chase 2758 K	this paper	AJ235381	same species	Duvall et al., 1993	M901625
Actinidiaceae	*Actinidia chinensis* Planch.	Kron 2117 NCU	this paper	AJ235382	same DNA	Albert et al., 1992	L01882
Adoxaceae	*Sambucus nigra* L.	Chase 2509 K	this paper	AJ235591	*Sambucus racemosa* L.	Donoghue et al., 1992	L14066
Adoxaceae	*Viburnum opulus* L.	Chase 2519 K	this paper	AJ235640	*Viburnum acerifolia* L.	Olmstead et al., 1992	L01959
Aextoxicaceae	*Aextoxicon punctatum* Ruiz and Pav.	Chase 959 K	this paper	AJ235384	same DNA	Alverson et al., 1998	X83986
Aizoaceae	*Delosperma echinatum* Schwantes	Chase 2539 K	this paper	AJ235452	same DNA	this paper	AJ235778
Alliaceae	*Ipheion dialystemon* Guaglianone	Chase 744 K	this paper	AJ235504	same DNA	Chase et al., 1995	Z77253
Alstroemeriaceae	*Bomarea hirtella* Herb.	Chase 520 K	this paper	AJ235413	same DNA	Chase et al., 1995	Z77255
Altingiaceae	*Altingia excelsa* Noronha	Hoot 9225 UWM	Hoot et al., 1999	AF092103	unknown	Chase et al., 1993	AJ131769
Altingiaceae	*Liquidambar styraciflua* L.	Kron 162 NCU	Hoot et al., 1999	AF092104	*Liquidambar formosana* Hance	Chase et al., 1993	AJ131772
Amaranthaceae	*Amaranthus hypochondriacus* L.	unknown	this paper	AJ235388	same species	Michalowski et al., 1990	X51964
Amaranthaceae	*Spinacia oleracea* L.	unknown	Zurawski et al., 1982	U23082	same species	Zurawski et al., 1981	J01443
Amborellaceae	*Amborella trichopoda* Baill.	Thien 500 NCU	this paper	AJ235389	same DNA	Qiu et al., 1993	L12628
Anacardiaceae	*Pistacia vera* L.	Terrazas sn CHAPA	Bayer et al., 1999	AJ132282	same DNA	this paper	AJ235786
Anacardiaceae	*Rhus vernix* L.	Terrazas sn CHAPA	Bakker et al., 1998	AF035912	same DNA	Gadek et al., 1996	U00440
Anacardiaceae	*Schinus molle* L.	Anderson 13601 MICH	Bakker et al., 1998	AF035914	same DNA	Gadek et al., 1996	U39270
Annonaceae	*Annona muricata* L.	Qiu 90031 NCU	this paper	AJ235393	same DNA	Qiu et al., 1993	L12629
Anthericaceae	*Anthericum liliago* L.	Chase 515 K	this paper	AJ235394	same DNA	Chase et al., 1995	Z69225
Apiaceae	*Apium graveolens* L.	Chase 2523 K	this paper	AJ235396	same species	Albert et al., 1992	L01885
Apocynaceae	*Dischidia lanceolata* Decne.	Chase 734 K	this paper	AJ235458	*Asclepias exaltata* L.	Olmstead et al., 1993	L14390
Apocynaceae	*Plumeria obtusa* Bert.	Chase 724 K	this paper	AJ235566	*Plumeria inodora* Jacq.	Sennblad and Bremer, unpubl.	X91767
Aquifoliaceae	*Ilex crenata* Thunb.	Chase 119 NCU	this paper	AJ235502	same DNA	Albert et al., 1992	L01928
Aquifoliaceae	*Nemopanthus mucronatus* Druce	Savolainen nmu1 G	this paper	AJ235541	same DNA	this paper	X69747
Araceae	*Spathiphyllum wallisii* Hort.	Chase 201 NCU	this paper	AJ235606	same DNA	Chase et al., 1993	AJ235807
Araliaceae	*Hedera helix* L.	Chase 2743 K	this paper	AJ235488	same species	Xiang et al., 1993	L01924
Argophyllaceae	*Corokia cotoneaster* Raoul	Chase 2752 K	this paper	AJ235445	same species	Xiang et al., 1993	L11221
Aristolochiaceae	*Aristolochia macrophylla* Lam.	Qiu 91019 NCU	this paper	AJ235399	same DNA	Qiu et al., 1993	L12630
Aristolochiaceae	*Asarum canadense* L.	Hoot 923 UWM	Hoot et al., 1999	U86383	same species	Chase et al., 1993	L14290
Aristolochiaceae	*Lactoris fernandeziana* Phil.	Stuessy 11335 OS	this paper	AJ235515	same species	Chase et al., 1993	L08763
Aristolochiaceae	*Saruma henryi* Oliv.	Chase 3077 K	this paper	AJ235595	same species	Qiu et al., 1993	L12664
Asparagaceae	*Asparagus officinalis* L.	Chase 513 K	this paper	AJ235400	same species	Duvall et al., 1993	L05028
Asphodelaceae	*Bulbine succulenta* Compton	UCI Arb. 7174	this paper	AJ235421	same DNA	Chase et al., 1995	AJ131947
Asteraceae	*Cichorium intybus* L.	Chase 2511 K	this paper	AJ235433	same species	Kim et al., 1992	L13640

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Aucubaceae	*Aucuba japonica* Thunb.	Chase 1095 K	this paper	AJ235402	same species	Xiang et al., 1993	L11210
Austrobaileyaceae	*Austrobaileya scandens* C.T.White	Qiu 90030 NCU	this paper	AJ235403	same DNA	Qiu et al., 1993	L12632
Balsaminaceae	*Impatiens repens* Moon	Chase 901 K	this paper	AJ235503	*Impatiens capensis* Meerb.	Chase et al., 1993	Z83142
Berberidaceae	*Caulophyllum thalictroides* (L.) Michx.	Hoot 925 UWM	Hoot et al., 1999	AF092108	same species	Chase et al., 1993	L08760
Berberidaceae	*Nandina domestica* Thunb.	Hoot 922 UWM	Hoot et al., 1999	L37930	same species	Hoot et al., 1995	L37920
Berberidopsidaceae	*Berberidopsis corallina* Hook.	Chase 555 K	this paper	AJ235409	same DNA	this paper	AJ235773
Betulaceae	*Betula pendula* L.	Chase 2539 K	this paper	AJ235411	*Betula nigra* L.	Albert et al., 1992	L01889
Bignoniaceae	*Catalpa bignonioides* Walt.	Chase 2539 K	this paper	AJ235428	same species	Olmstead et al., 1992	L11679
Bixaceae	*Bixa orellana* L.	Chase 243 NCU	Bakker et al., 1998	AF035897	same DNA	Fay et al., 1998	Y15139
Bixaceae	*Cochlospermum intermedium* Mildbr.	Chase 2434 K	Bayer et al., 1999	AJ233060	same DNA	Fay et al., 1998	Y15143
Bixaceae	*Diegodendron humbertii* Capuron	Capuron 23034 K	Bayer et al., 1999	AJ233061	same DNA	Fay et al., 1998	Y15138
Blandfordiaceae	*Blandfordia punicea* Sweet	Chase 519 K	this paper	AJ235412	same DNA	Chase et al., 1995	Z73694
Boraginaceae	*Borago officinalis* L.	Chase 2746 K	this paper	AJ235414	same species	Olmstead et al., 1992	L11680
Boraginaceae	*Hydrophyllum canadense* L.	Chase 2548 K	this paper	AJ235498	*Hydrophyllum virginianum* L.	Olmstead et al., 1992	L01927
Brassicaceae	*Brassica balearica* Pers.	Chase 1534 K	this paper	AJ132281	*Brassica oleracea* L	Rodman et al., 1993	M88342
Brassicaceae	*Capparis spinosa* L.	Chase 2751 K	Bakker et al., 1998	AF035900	*Capparis hastata* Jacq.	Rodman et al., 1993	M95755
Brassicaceae	*Megacarpaea polyandra* Benth.	Chase 565 K	this paper	AJ235531	*Brassica oleracea* L	Rodman et al., 1993	M88342
Brassicaceae	*Stanleya pinnata* Britton	Chase 2748 K	Bayer et al., 1999	AJ132284	same species	Chase et al., 1993	AJ235809
Bruniaceae	*Berzelia lanuginosa* Brongn.	Kirstenbosch 7589	Hoot et al., 1999	AF095731	same species	Olmstead et al., 1993	L14391
Buddlejaceae	*Buddleja auriculata* Benth.	Chase 2467 K	this paper	AJ235420	*Buddleja davidii* Franch	Olmstead et al., 1993	L14392
Burseraceae	*Bursera inaguensis* Britton	Fairchild Trop Garden 64–269 D	Bakker et al., 1998	AF035899	same DNA	Albert et al., 1992	L01890
Buxaceae	*Buxus sempervirens* L.	Hoot 921 UWM	Hoot et al., 1999	AF092110	same species	Hoot et al., 1999	AF093717
Buxaceae	*Pachysandra procumbens* Michx.	Hoot 917 UWM	Hoot et al., 1999	AF092111	same species	Chase et al., 1993	AJ235815
Cactaceae	*Rhipsalis teres* Steud.	Chase 2545 K	this paper	AJ235581	*Schlumbergera truncata* Moran	Manhart et al., unpubl.	M83543
Calycanthaceae	*Calycanthus floridus* L.	Qiu 94155 NCU	this paper	AJ235422	same DNA	Chase et al., 1993	L14291
Calycanthaceae	*Idiospermum australiense* Blake	Qiu 91042 NCU	this paper	AJ235500	same DNA	Qiu et al., 1993	L12651
Campanulaceae	*Campanula trachelium* Brot.	Chase 2546 K	this paper	AJ235423	*Campanula ramulosa* Wall.	Olmstead et al., 1992	L13861
Campanulaceae	*Lobelia angulata* Forst.	Chase 2540 K	this paper	AJ235524	*Lobelia erinus* L.	Albert et al., 1992	L01931
Canellaceae	*Canella winterana* Gaertn.	Qiu 90017 NCU	this paper	AJ235424	same DNA	Qiu et al., 1993	AJ131928
Canellaceae	*Cinnamodendron ekmanii* Sleumer	Qiu 47067 NCU	this paper	AJ235435	same DNA	this paper	AJ235776
Cannabaceae	*Humulus lupulus* L.	Chase 2749 K	this paper	AJ235495	same species	Chase et al., 1993	U02729
Caricaceae	*Carica papaya* L.	Chase 2508 K	Bakker et al., 1998	AF035901	same species	Rodman et al., 1993	M95671
Caryophyllaceae	*Silene nutans* L.	Chase 2292 K	this paper	AJ235601	*Silene gallica* L.	Manhart et al., unpubl.	M83544
Casuarinaceae	*Casuarina litorea* L.	Chase 215 NCU	this paper	AJ235427	same DNA	Albert et al., 1992	L01893
Celastraceae	*Brexia madagascariensis* Thouars	Schwerdtfeger 25471 B	this paper	AJ235419	same DNA	Morgan and Soltis, 1993	L11176
Celastraceae	*Celastrus orbiculatus* Humb. and Bonpl.	Chase 2274 K	this paper	AJ235429	same DNA	this paper	AJ235775
Celastraceae	*Euonymus alatus* Siebold	Chase 137 NCU	this paper	AJ235471	same DNA	Chase et al., 1993	L13184
Celastraceae	*Hippocratea barbata* Muell.	Chase 2971 K	this paper	AJ235493	*Hippocratea richardiana* Cambess.	Savolainen et al., 1997	X69740
Celastraceae	*Plagiopteron suaveolens* Griff.	Chase 1335 K	this paper	AJ235562	same DNA	this paper	AJ235787

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Celastraceae	*Salacia pallescens* Oliv.	Van Der Laan 373 WAG	this paper	AJ235589	same DNA	this paper	X69754
Ceratophyllaceae	*Ceratophyllum demersum* L.	Qiu 91027 NCU	this paper	AJ235430	same species	Les et al., 1991	M77030
Cercidiphyllaceae	*Cercidiphyllum japonicum* Siebold and Zucc.	Olmstead 90–016 COLO	Hoot et al., 1999	AF092112	same species	Olmstead et al., 1992	L11673
Chloranthaceae	*Chloranthus japonicus* Siebold	Chase 204 NCU	this paper	AJ235431	same DNA	Qiu et al., 1993	L12640
Chloranthaceae	*Hedyosmum arborescens* Sw.	Chase 338 NCU	this paper	AJ235490	same DNA	Qiu et al., 1993	L12649
Chloranthaceae	*Sarcandra grandiflora* Subr. and Henry	Qiu 92002 NCU	this paper	AJ235593	same DNA	Qiu et al., 1993	L12663
Cistaceae	*Cistus revolii* Coste and Soulie	Chase 525 K	Bakker et al., 1998	AF035902	same DNA	Fay et al., 1998	Y15140
Cistaceae	*Helianthemum grandiflorum* DC.	Chase 524 K	Bakker et al., 1998	AF035907	same DNA	Fay et al., 1998	Y15141
Clethraceae	*Clethra arborea* Vent.	Chase 902 K	this paper	AJ235438	*Clethra alnifolia* L.	Kron and Chase 1993	L12609
Colchicaceae	*Androcymbium ciliolatum* Schltr.	Chase 272 NCU	this paper	AJ235391	same DNA	Chase et al., 1995	Z72265
Combretaceae	*Quisqualis indica* L.	Chase 128 NCU	this paper	AJ235576	same DNA	Albert et al., 1992	L01948
Connaraceae	*Rourea minor* Leenk.	Chase 1221 K	this paper	AJ235585	*Connarus conchocarpus* F.Muell.	Morgan and Soltis, 1993	U06798
Convolvulaceae	*Ipomoea mauritiana* Jacq.	Chase 2525 K	this paper	AJ235505	*Ipomoea coccinea*	Olmstead et al., 1993	L14400
Coriariaceae	*Coriaria myrtifolia* L.	Chase 245 NCU	this paper	AJ235443	same DNA	Albert et al., 1992	L01897
Cornaceae	*Alangium* sp.	Chase 2541 K	this paper	AJ235386	*Alangium chinense* Harms	Xiang et al., 1993	L11209
Cornaceae	*Cornus mas* L.	Chase 2520 K	this paper	AJ235444	same species	Xiang et al., 1993	L11216
Cornaceae	*Nyssa sylvatica* Marsh.	Chase 2530 K	this paper	AJ235545	*Nyssa ogeche* Marsh.	Xiang et al., 1993	L11228
Corynocarpaceae	*Corynocarpus laevigatus* Forst.	Chase 236 NCU	this paper	AJ235446	same DNA	Savolainen et al., 1997	X69731
Crassulaceae	*Crassula marnierana* Huber and Jacobsen	Morgan 2152 WS	this paper	AJ235447	same DNA	Albert et al., 1992	L01899
Crassulaceae	*Dudleya viscida* Moran	Huntington 62801 BG	this paper	AJ235461	same DNA	Morgan and Soltis, 1993	L11182
Crassulaceae	*Kalanchoe daigremontana* Hamet and Perrier	Morgan 2151 WS	this paper	AJ235510	same DNA	Morgan and Soltis, 1993	L11189
Crassulaceae	*Sedum nudum* Aiton	Chase 2459 K	this paper	AJ235600	*Sedum rubrotinctum* Clausen	Albert et al., 1992	L01956
Cucurbitaceae	*Kedrostis nana* Cogn.	Chase 274 NCU	this paper	AJ235511	same DNA	this paper	AJ235782
Cucurbitaceae	*Xerosicyos danguyi* Humb.	Chase 321 NCU	this paper	AJ235648	*Xerosicyos decaryi* Guillaumin	Swensen, 1996	U59824
Cunoniaceae	*Eucryphia milliganii* Hook.	Chase 2528 K	this paper	AJ235470	*Eucryphia lucida* Druce	Albert et al., 1992	L01918
Cyclanthaceae	*Sphaeradenia pendula* Hammel	Chase 222 NCU	this paper	AJ235607	same DNA	Chase et al., 1993	AJ235808
Cyrillaceae	*Cyrilla racemiflora* L.	Chase 2531 K	this paper	AJ235449	same species	Albert et al., 1992	L01900
Daphniphyllaceae	*Daphniphyllum* sp.	Wagner et al. 6599 HAST	Hoot et al., 1999	AF092118	same species	Albert et al., 1992	L01901
Datiscaceae	*Datisca cannabina* L.	Chase 2745 K	this paper	AJ235450	same species	Chase et al., 1993	L21939
Degeneriaceae	*Degeneria* sp.	Qiu 1202–55 NCU	this paper	AJ235451	same DNA	Qiu et al., 1993	L12643
Dichapetalaceae	*Dichapetalum brownii* Baill.	Fison s.n. 10/8/93 K	this paper	AJ235455	*Dichapetalum crassifolium* Chod.	Savolainen et al., 1997	X69733
Didymeleaceae	*Didymeles perrieri* Leandri	Andrianantonnina 387 MO	Hoot et al., 1999	AF092119	same DNA	Hoot et al., 1999	AF094541
Dilleniaceae	*Dillenia retusa* Thunb.	Chase 2103 K	this paper	AF095732	*Dillenia indica* L.	Albert et al., 1992	L01903
Dilleniaceae	*Schumacheria* sp.	Chase 308 NCU	Hoot et al., 1999	AF092121	same DNA	Hoot et al., 1999	AF095734
Dilleniaceae	*Tetracera asiatica* Hoogl.	Chase 1238 K	this paper	AJ235622	same DNA	this paper	AJ235796
Dioscoreaceae	*Dioscorea polygonoides* Plum. and Bompl.	Chase 197 NCU	this paper	AJ235456	same DNA	Chase et al., 1993	AJ235803

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Dipterocarpaceae	*Anisoptera marginata* Korth.	Chase 2486 K	Bakker et al., 1998	AF035918	same DNA	Fay et al., 1998	Y15144
Droseraceae	*Drosera communis* A.St.-Hil.	Chase 2582 K	this paper	AJ235459	*Drosera spathulata* Labill.	Chase et al., 1993	L13168
Ebenaceae	*Diospyros kaki* Thunb.	Chase 920 K	this paper	AJ235457	*Diospyros virginiana* L.	Kron and Chase, 1993	L12613
Ebenaceae	*Euclea natalensis* DC.	Goldblatt 9275 MO	this paper	AJ235468	same DNA	Morton et al., 1997	Z80186
Elaeagnaceae	*Elaeagnus* sp.	Chase 2414 K	this paper	AJ235462	*Elaeagnus angustifolia* L.	Soltis et al., 1995	U17038
Elaeocarpaceae	*Sloanea berteriana* Choisy	Chase 343 NCU	this paper	AJ235603	*Sloanea latifolia* Schumann	Alverson et al., 1998	AF022131
Ephedraceae	*Ephedra tweediana* C.A.Mey.	none	this paper	AJ235463	same DNA	Chase et al., 1993	L12677
Ericaceae	*Erica sicula* Guss.	Chase 892 K	this paper	AJ235465	*Erica australis* L.	Kron and Chase, 1993	L12617
Erythroxylaceae	*Erythroxylum confusum* Britton	Fairchild Trop Garden 63–251E	this paper	AJ235466	same DNA	Chase et al., 1993	L13183
Escalloniaceae	*Escallonia* sp.	Chase 2499 K	this paper	AJ235467	*Escallonia coquimbensis* Remy	Morgan and Soltis, 1993	L11183
Eucommiaceae	*Eucommia ulmoides* Oliv.	Chase 2755 K	this paper	AJ235469	same species	Albert et al., 1992	L01917
Euphorbiaceae	*Euphorbia polychroma* Kerner	Chase 102 NCU	this paper	AJ235472	same DNA	Chase et al., 1993	L13183
Eupomatiaceae	*Eupomatia bennettii* Muell.	Qiu 90022 NCU	this paper	AJ235473	same DNA	Qiu et al., 1993	L12644
Eupteleaceae	*Euptelea polyandra* Siebold et Zucc.	Qiu 90026 NCU	Hoot et al., 1997	U86384	same DNA	Chase et al., 1993	L12645
Fabaceae	*Pisum sativum* L.	unknown	Zurawski et al., 1986	X03852	same species	Zurawski et al., 1986	X03853
Fabaceae	*Sophora toromiro* Skottsb.	Chase s.n. K	this paper	AJ235605	*Sophora davidii* Kom.	Käss and Wink, 1995	Z70138
Fagaceae	*Trigonobalanus verticellatus* Forman	Chase 595 K	this paper	AJ235631	same species	Chase et al., 1993	AJ235812
Flacourtiaceae	*Hydnocarpus heterophylla* Blume	Chase 1278 K	this paper	AJ235496	same DNA	this paper	AJ235781
Fouquieriaceae	*Idria columnaria* Kellogg	U. California, Irvine Arb.	this paper	AJ235501	same DNA	Morton et al., 1996	Z80210
Francoaceae	*Francoa appendiculata* A.Juss.	Chase 2502 K	Bakker et al., 1998	AF035905	*Francoa sonchifolia* Cav.	Soltis et al., 1990	L11184
Frankeniaceae	*Frankenia pulverulenta* L.	Collenette 6/93 K	this paper	AJ235475	same DNA	Fay et al., 1997	Z97638
Garryaceae	*Garrya elliptica* Douglas	Chase 1098 K	this paper	AJ235478	same species	Albert et al., 1992	L01919
Geraniaceae	*Geranium sanguineum* L.	Chase 125 NCU	Bakker et al., 1998	AF035906	*Geranium cinereum* Cav.	Price and Palmer, 1993	L14695
Geraniaceae	*Pelargonium cotyledonis* L'Hér.	Chase 2477 K	Bakker et al., 1998	AF035911	same species	Price and Palmer, 1993	L14704
Gesneriaceae	*Saintpaulia ionantha* Wendl.	Chase 700 K	this paper	AJ235588	*Streptocarpus holstii* Engl.	Olmstead et al., 1993	L14409
Ginkgoaceae	*Ginkgo biloba* L.	Chase 2489 K	this paper	AJ235480	same species	Chase et al., 1993	AJ235804
Gnetaceae	*Gnetum gnemon* L.	Chase 208 NCU	this paper	AJ235482	same DNA	Chase et al., 1993	L12680
Goupiaceae	*Goupia glabra* Aubl.	Prevost 3031 CAY	this paper	AJ235484	same DNA	this paper	AJ235780
Grossulariaceae	*Ribes speciosum* Pursh.	Chase 2542 K	this paper	AJ235583	*Ribes aureum* Pursh	Morgan and Soltis, 1993	L11204
Gunneraceae	*Gunnera hamiltonii* Kirk. ex W.S. Ham.	Chase 562, K	Hoot et al., 1999	AF093374	*Gunnera manicata* Linden	Morgan and Soltis, 1993	L11186
Haloragaceae	*Haloragis aspera* Lindl.	Chase 453 K	this paper	AJ235487	*Haloragis serra* Brongn.	Conti et al., 1996	U26325
Haloragaceae	*Myriophyllum exalbescens* Fernald	Broch 30 Aug 1991 WS	this paper	AJ235538	same DNA	Morgan and Soltis, 1993	L11195
Hamamelidaceae	*Corylopsis pauciflora* Siebold . and Zucc	Reznicek 9239 MICH	Hoot et al., 1999	AF093377	same DNA	Hoot et al., 1999	AF094548
Hamamelidaceae	*Disanthus cercidifolia* Maxim.	Hoot 9221 UWM	Hoot et al., 1999	AF093378	same DNA	Hoot et al., 1999	AF094549
Hamamelidaceae	*Hamamelis virginiana* L.	Hoot 910 UWM	Hoot et al., 1999	AF093380	*Hamamelis mollis* Oliv.	Albert et al., 1992	L01922
Helwingiaceae	*Helwingia japonica* Dietr.	Chase 1104 K	this paper	AJ235492	same species	Xiang et al., 1993	L11226
Hernandiaceae	*Gyrocarpus americana* Jacq.	Chase 317 NCU	this paper	AJ235486	same DNA	Qiu et al., 1993	L12647
Himantandraceae	*Galbulimima belgraveana* Sprague	Qiu 90034 NCU	this paper	AJ235477	same DNA	Qiu et al., 1993	L12646
Huaceae	*Afrostyrax* sp.	Cheek 5007 K	this paper	AJ235385	same DNA	this paper	AJ235771

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Humiriaceae	*Humiria balsaminifera* Aubl.	Anderson 13654 MICH	this paper	AJ235494	same DNA	Albert et al., 1992	L01926
Hydrangeaceae	*Carpenteria californica* Torr.	Chase 2497 K	this paper	AJ235426	same species	Soltis et al., 1990	L11177
Hydrangeaceae	*Hydrangea macrophylla* Torr.	Chase 2537 K	this paper	AJ235497	same species	Morgan and Soltis, 1993	L11187
Hypoxidaceae	*Rhodohypoxis milloides* Hilliard and B.L.Burtt	Chase 479 K	this paper	AJ235582	same DNA	Chase et al., 1995	Z77280
Icacinaceae	*Gonocaryum litorale* Sleum.	Chase 1294 K	this paper	AJ235483	same DNA	this paper	AJ235779
Icacinaceae	*Pyrenacantha malvifolia* Engl.	Chase 683 K	this paper	AJ235575	same DNA	this paper	AJ235791
Illiciaceae	*Illicium parviflorum* Michx.	Naczi 2784 MICH	Hoot et al., 1997	U86385	same species	Qiu et al., 1993	L12652
Iteaceae	*Itea virginica* L.	Ware 9401 WS	Hoot et al., 1999	AF093383	same species	Soltis et al., 1990	L11188
Ixioliriaceae	*Ixiolirion tataricum* Herb.	Chase 489 K	this paper	AJ235507	same DNA	Chase et al., 1995	Z73704
Juglandaceae	*Pterocarya fraxinifolia* Spach	Chase 860 K	this paper	AJ235572	same DNA	this paper	AJ235790
Juncaceae	*Juncus effusus* L.	Chase 200 NCU	this paper	AJ235509	same DNA	Chase et al., 1993	L12681
Krameriaceae	*Krameria ixine* L.	Litt 1 NY	this paper	AJ235514	*Krameria lanceolata* Torr.	Chase et al., 1993	Y15032
Lamiaceae	*Lavandula bipinnata* Kuntze	Upson sn RNG	this paper	AJ235519	*Lavandula angustifolia* Moench	Kaufman and Wink, 1994	Z37404
Lamiaceae	*Prostanthera ovalifolia* R.Br.	Chase 2522 K	this paper	AJ235571	*Prostanthera rotundifolia* R.Br.	Olmstead et al., 1993	L14408
Lardizabalaceae	*Decaisnea fargesii* Franch.	Reznicek 9236 MICH	Hoot et al., 1995	L37926	same DNA	Hoot et al., 1995	L37916
Lauraceae	*Cinnamomum camphora* Nees and Eberm.	Qiu 102 NCU	this paper	AJ235436	same DNA	Qiu et al., 1993	L12641
Lauraceae	*Laurus nobilis* Cav.	Qiu 94209 NCU	this paper	AJ235518	*Persea americana* Mill.	Golenberg et al., 1990	X54347
Lecythidaceae	*Barringtonia asiatica* Kurz	Chase 328 K	this paper	AJ235407	same DNA	Morton et al., 1996	Z80174
Lecythidaceae	*Napoleonaea vogelii* Hook. and Planch.	Chase 329 K	this paper	AJ235540	same DNA	Morton et al., 1996	Z80173
Lentibulariaceae	*Utricularia biflora* Roxb.	Chase 143 NCU	this paper	AJ235636	same DNA	Albert et al., 1992	L13190
Liliaceae	*Lloydia serotina* Sweet	Jones sn K	this paper	AJ235523	same DNA	Chase et al., 1995	Z77294
Liliaceae	*Nomocharis pardanthina* Planch.	Chase 934 K	this paper	AJ235543	same DNA	Chase et al., 1995	Z77295
Liliaceae	*Tricyrtis latifolia* Maxim.	Chase 548 K	this paper	AJ235630	*Tricyrtis affinis* Makino	Shinwari et al., 1994	D17382
Liliaceae	*Tulipa kolpakowskiana* Baker	Chase 438 K	this paper	AJ235633	same DNA	Chase et al., 1995	Z77292
Limnanthaceae	*Floerkea proserpinicoides* Willd.	Reznicek 8609 MICH	Bakker et al., 1998	AF035904	same DNA	Chase et al., 1993	L12679
Linaceae	*Linum perenne* Guss.	Chase 111 NCU	this paper	AJ235521	same DNA	Fay et al., 1997	Z75681
Linaceae	*Reinwardtia indica* Dumort.	Chase 230 NCU	this paper	AJ235577	same DNA	Fay et al., 1997	Z13188
Loganiaceae	*Strychnos nux-vomica* L.	Chase 2538 K	this paper	AJ235613	same species	Olmstead et al., 1993	L14410
Lythraceae	*Punica protopunica* Balf.	Chase 1905 K	this paper	AJ235574	*Punica granatum* L.	Conti et al., 1993	L10223
Magnoliaceae	*Liriodendron tulpifera* L.	Qiu 94126 NCU	this paper	AJ235522	same species	Golenberg et al., 1990	X54346
Magnoliaceae	*Magnolia tripetala* L.	Qiu 3 NCU	this paper	AJ235526	same DNA	Qiu et al., 1993	AJ131927
Malpighiaceae	*Dicella nucifera* Chodat	Anderson 13607 MICH	this paper	AJ235453	same DNA	Chase et al., 1993	AJ235802
Malpighiaceae	*Malpighia coccigera* L.	Mathaei BG 20626 MICH	this paper	AJ235527	same DNA	this paper	AJ235784
Malvaceae	*Adansonia rubrostipa* Jum. and Perrier	Chase 3043 K	Bayer et al., 1999	AJ233050	same DNA	Bayer et al., 1999	AJ233115
Malvaceae	*Bombax ceiba* L.	Chase 3049 K	Bayer et al., 1999	AJ233051	*Bombax buonopozense* P.Beauv.	Chase et al., 1993	AF022118
Malvaceae	*Chorisia speciosa* A..St.-Hil.	Chase 3188 K	Bayer et al., 1999	AJ233052	same DNA	Bayer et al., 1999	AJ233116
Malvaceae	*Dombeya tiliacea* Planch.	Chase 273 UCI	Bayer et al., 1999	AJ233075	same DNA	Bayer et al., 1999	AJ233125
Malvaceae	*Gossypium hirsutum* L.	Chase 3014 K	Bayer et al., 1999	AJ233052	*Gossypium robinsoni* F.Muell.	Chase et al., 1993	L13186

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Malvaceae	*Grewia occidentalis* L.	Chase 3042 K	Bayer et al., 1999	AJ233105	same DNA	Bayer et al., 1999	AJ233152
Malvaceae	*Ochroma pyramidale* Urb.	Chase 244 NCU	Bayer et al., 1999	AJ233055	same DNA	Bayer et al., 1999	AJ233118
Malvaceae	*Sterculia apetala* Karsten	Chase 352 K	Bayer et al., 1999	AJ233089	*Sterculia tragacantha* Lindl.	Alverson et al., 1998	AF022126
Malvaceae	*Theobroma cacao* L.	Chase 3016 K	Bayer et al., 1999	AJ233090	same species	Chase et al., 1993	AF022125
Malvaceae	*Tilia platyphyllos* Scop.	Chase 3018 K	Bayer et al., 1999	AJ233113	*Tilia americana* L.	Chase et al., 1993	AF022127
Marcgraviaceae	*Marcgravia rectiflora* Triana and Planch.	Chase 331 NCU	this paper	AJ235529	same DNA	Morton et al., 1996	Z83148
Medusagynaceae	*Medusagyne oppositifolia* Baker	Chase 670 K	this paper	AJ235530	same DNA	Fay et al., 1997	Z75670
Melanthiaceae	*Veratrum viride* Aiton	Chase 551 K	this paper	AJ235638	*Veratrum parviflorum* Michx.	Chase et al., 1993	AJ235813
Melanthiaceae	*Xerophyllum tenax* Nutt.	Chase 527 K	this paper	AJ132285	same DNA	Chase et al., 1995	AJ131949
Melastomataceae	*Clidemia petiolaris* Triana	Chase 2534 K	this paper	AJ235439	same DNA	this paper	AJ235777
Meliaceae	*Swietenia macrophylla* King	Chase 250 K	this paper	AJ235616	same DNA	Gadek et al., 1996	U39080
Meliaceae	*Trichilia emetica* Vahl	Chase 552 K	this paper	AJ235629	same DNA	Gadek et al., 1996	U39082
Melianthaceae	*Bersama lucens* Szyszyl.	Kirstenboch 385/83	this paper	AJ235410	same DNA	this paper	AJ235774
Melianthaceae	*Melianthus major* L.	U California, Irvine Arb	this paper	AJ235532	same species	Gadek et al., 1996	pers. comm.
Menispermaceae	*Menispermum canadense* L.	Naczi 2837 (MICH)	Hoot et al., 1999	AF093384	same species	Hoot et al., 1999	AF093726
Menyanthaceae	*Menyanthes trifoliata* L.	Chase 3501 K	this paper	AJ235533	same species	Olmstead et al., 1993	L14006
Molluginaceae	*Limeum* sp.	Hoot 983 UWM	Hoot et al., 1999	AF093385	*Trianthema portulacastrum* L.	Manhart et al., unpubl.	M62572
Monimiaceae	*Hedycarya arborea* Forst.	Qiu 90028 NCU	this paper	AJ235489	same DNA	Qiu et al., 1993	L12648
Monimiaceae	*Kibara* sp.	Coode 7879 K	this paper	AJ235512	*Hedycarya arborea* Forst.	Qiu et al., 1993	L12648
Moraceae	*Morus nigra* L.	Chase 2512 K	this paper	AJ235536	*Morus alba* L.	Albert et al., 1992	L01933
Muntingiaceae	*Muntingia calabura* L.	Chase 346 NCU	this paper	AF035908	same DNA	Fay et al., 1998	Y15146
Myricaceae	*Myrica cerifera* L.	Chase 2500 K	Bayer et al., 1999	AJ233068	same species	Albert et al., 1992	L01934
Myristicaceae	*Myristica fragans* Houtt.	Qiu 92014 NCU	this paper	AJ235539	*Knema latericia* Elmer	Qiu et al., 1998	L12653
Myrothamnaceae	*Myrothamnus flabellifolius* Welw.	Winter 72 RAV	Hoot et al., 1999	AF093386	same DNA	Qiu et al., 1999	AF060707
Myrsinaceae	*Maesa myrsinoides* Lév.	Chase 309 K	this paper	AJ235525	same DNA	Morton et al., 1996	Z80203
Myrtaceae	*Metrosideros nervulosa* Moore and Muell.	Chase 2451 K	this paper	AJ235535	same DNA	this paper	AJ235785
Nelumbonaceae	*Nelumbo lutea* Pers.	Hoot 9212, UWM	Hoot et al., 1999	AF093387	same species	Les et al., 1991	M77032
Nepenthaceae	*Nepenthes alata* Blanco	Chase 145, NCU	this paper	AJ235542	same DNA	Albert et al., 1992	L01935
Nyctaginaceae	*Bougainvillea glabra* Choisy	Chase 2485 K	this paper	AJ235415	same species	Manhart et al., unpubl.	M88340
Nymphaeaceae	*Brasenia schreberi* Gmelin	Qiu 91031 NCU	this paper	AJ235418	same species	Les et al., 1991	M77028
Nymphaeaceae	*Nymphaea odorata* Aiton	Qui 91029 NCU	this paper	AJ235544	same species	Les et al., 1991	M77031
Ochnaceae	*Ochna multiflora* DC.	Chase 229 NCU	this paper	AJ235546	*Ochna serrulata* Walp.	Chase et al., 1993	Z75273
Olacaceae	*Heisteria parvifolia* Sm.	Cheek 5985 K	this paper	AJ235491	same DNA	this paper	AJ131771
Oleaceae	*Jasminum polyanthum* Franch.	Chase 2474 K	this paper	AJ235508	*Jasminum suavissimum* Lindl.	Albert et al., 1992	L01929
Onagraceae	*Fuchsia procumbens* Cunn.	Chase 2530 K	this paper	AJ235476	*Fuchsia cyrtandroides* Moore	Conti et al., 1993	L10220
Oncothecaceae	*Oncotheca balansae* Baill.	Jaffre 3238 NOU	this paper	AJ235549	same DNA	this paper	AJ131950
Opiliaceae	*Opilia* sp.	Chase 1902 K	this paper	AJ235550	same DNA	this paper	AJ131773
Orchidaceae	*Apostasia stylidoides* Rchb.	Clements 4843 CANB	this paper	AJ235397	same DNA	Chase et al., 1995	Z73705
Orchidaceae	*Cypripedium calceolus* Walt.	Chase O-714 K	this paper	AJ235448	*Cypripedium irapeanum* Lex.	Chase et al., 1993	Z73706
Orchidaceae	*Oncidium excavatum* Lindl.	Chase O-86 K	this paper	AJ235548	same DNA	Olmstead et al., 1992	AF074201

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Oxalidaceae	*Averrhoa carambola* L.	Chase 214 NCU	this paper	AJ235404	same species	Price and Palmer, 1993	L14692
Paeoniaceae	*Paeonia mlokosewitschi* Lomakin	Chase 505 K	this paper	AJ235551	*Paeonia tenuifolia* L.	Chase et al., 1993	L13687
Papaveraceae	*Dicentra chrysantha* Walp.	Chase 534 K	this paper	AJ235454	*Dicentra spectabilis* Lem.	Chase et al., 1993	L08761
Parnassiaceae	*Parnassia palustris* L.	Fay sn K	this paper	AJ235552	*Parnassia fimbriata* Banks	Soltis et al., 1990	L01939
Passifloraceae	*Passiflora coccinea* Aubl.	Chase 2475 K	this paper	AJ235553	*Passiflora quadrangalis* L.	Albert et al., 1992	L01940
Penthoraceae	*Penthorum sedoides* L.	Hayden 2232 WS	this paper	AJ235555	same DNA	Soltis et al., 1990	L11197
Phellinaceae	*Phelline comosa* Labill.	Savolainen pco1 G	this paper	AJ235557	same DNA	this paper	X69748
Philesiaceae	*Lapageria rosea* Ruiz and Pav.	Chase 181 NCU	this paper	AJ235517	same DNA	Chase et al., 1995	Z77301
Phytolaccaceae	*Ercilla volubilis* A.Juss.	Chase 2526 K	this paper	AJ235464	same DNA	this paper	AJ235800
Phytolaccaceae	*Phytolacca dioica* L.	Chase 2535 K	this paper	AJ235558	*Phytolacca americana* L.	Rettig et al., 1992	M62567
Picramniaceae	*Alvaradoa amorphoides* Liebm.	21621 UNSW	this paper	AJ235387	same DNA	Fernando et al., 1995	AF123277
Picramniaceae	*Picramnia pentandra* SW.	21620 UNSW	this paper	AJ235559	same DNA	Fernando et al., 1995	pers. comm.
Pinaceae	*Pinus thunbergiana* Franco	unknown	Wakasugi et al., 1994	D17510	same species	Wakasugi et al., 1994	D17510
Pinaceae	*Tsuga canadensis* Carr.	Chase 2514 K	this paper	AJ235632	*Tsuga heterophylla* Sarg.	Chase et al., 1993	X63659
Piperaceae	*Peperomia obtusifolia* Miq.	Qiu 94135 NCU	this paper	AJ235556	*Peperomia obtusifolia* Miq.	Qiu et al., 1993	L12661
Piperaceae	*Piper betle* L.	Qiu 91048 NCU	this paper	AJ235560	same DNA	Qiu et al., 1993	L12660
Pittosporaceae	*Pittosporum fairchildii* Cheeseman	Chase 2468 K	this paper	AJ235561	*Pittosporum japonicum* Hort.	Morgan and Soltis, 1993	L11202
Platanaceae	*Platanus occidentalis* L.	Qiu P90005 NCU	Hoot et al., 1997	U86386	same DNA	Albert et al., 1992	L01943
Plumbaginaceae	*Plumbago zeylanica* L.	Chase 994 K	this paper	AJ235565	*Plumbago capensis* Thunb.	Giannasi et al., 1992	M77701
Poaceae	*Oryza sativa* L.	unknown	Hiratsuka et al., 1989	X15901	same species	Hiratsuka et al., 1989	X15901
Poaceae	*Zea mays* L.	unknown	Maier et al., 1995	X86563	same species	Maier et al., 1995	X86563
Podocarpaceae	*Podocarpus milanjianus* Rendl.	Chase 2482 K	this paper	AJ235567	*Podocarpus gracilior* Pilg.	Bousquet et al., 1992	X58135
Polemoniaceae	*Cobaea scandens* Cav.	Chase 961 K	this paper	AJ235440	same DNA	Morton et al., 1996	Z83143
Polygalaceae	*Polygala cruciata* L.	Chase 155 NCU	this paper	AJ235568	same DNA	Albert et al., 1992	L01945
Polygalaceae	*Xanthophyllum* sp.	Coode 7760 K	this paper	AJ235646	same DNA	this paper	AJ235799
Polygonaceae	*Polygonum sachalinense* Schmidt	Chase 896 K	this paper	AJ235569	same DNA	this paper	AJ235789
Polygonaceae	*Rheum pinchonii* Pierre	Chase 926 K	this paper	AJ235580	*Rheum X cultorum*	Giannasi et al., 1992	M77701
Primulaceae	*Anagallis tenella* L.	Chase 1910 K	this paper	AJ235390	*Anagallis arvensis* L.	Chase et al., 1993	M88343
Primulaceae	*Androsace spinulifera* Knuth	Chase 954 K	this paper	AJ235392	same DNA	this paper	AJ235772
Proteaceae	*Lambertia inermis* R.Br.	Natl. Trop. BG Hawai	this paper	AJ235516	same species	Morgan and Soltis, 1993	L11190
Proteaceae	*Placospermum coriaceum* White and Francis	Douglas 110 MEL	Hoot and Douglas, 1998	AF060391	same DNA	Hoot et al., 1999	AF093729
Proteaceae	*Roupala macrophylla* Pohl	Douglas 131 MEL	Hoot and Douglas, 1998	AF060416	same DNA	Hoot et al., 1999	AF093728
Pterostemonaceae	*Pterostemon rotundifolius* Ramírez	Jordan s.n. HO	this paper	AJ235573	same DNA	Morgan and Soltis 1993	L11203
Ranunculaceae	*Glaucidium palmatum* Siebold and Zucc.	Hoot 924 UWM	Hoot et al., 1999	AF093375	same DNA	Hoot & Crane, 1995	L75848
Ranunculaceae	*Hydrastis canadense* L.	Naczi 2883, MICH	Hoot et al., 1999	AF093382	same DNA	Hoot & Crane, 1995	L75849
Ranunculaceae	*Xanthorhiza simplicissima* Marshall	Qiu 91030 NCU	Hoot et al., 1999	AF093394	same DNA	Chase et al., 1993	L12669
Resedaceae	*Reseda alba* L.	Chase 3017 K	Bayer et al., 1999	AJ132283	same species	Rodman et al., 1993	L11359

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Rhabdodendraceae	*Rhabdodendron amazonicum* Huber	Ribeiro 1187 K	this paper	AJ235578	same DNA	Fay et al., 1997	Z97649
Rhamnaceae	*Rhamnus cathartica* L.	Chase 100 NCU	this paper	AJ235579	same DNA	Chase et al., 1993	L13189
Rhizophoraceae	*Carallia brachiata* Merill.	Chase 2151 K	this paper	AJ235425	*Bruguiera gymnorhiza* Savigny	Conti et al., 1996	U26320
Rosaceae	*Dryas drummondi* Richardsons	Chase 917 K	this paper	AJ235460	same DNA	Swensen, 1996	U59818
Rosaceae	*Geum* sp.	Chase 2507 K	this paper	AJ235479	*Geum chiloense* Balb.	Albert et al., 1992	L01921
Rosaceae	*Spiraea betulifolia* Pall.	Chase 2503 K	this paper	AJ235608	*Spirea vanhouttei* Zabel	Morgan and Soltis, 1993	L11206
Rousseaceae	*Roussea simplex* Sm.	Mauritius Sugar Res. Inst.	this paper	AJ235586	same DNA	this paper	AJ235792
Rubiaceae	*Bouvardia glaberrina* Engelm.	Natali and Manen 3 G	this paper	AJ235416	same DNA	Manen and Natali, 1995	X81093
Rubiaceae	*Cinchona pubescens* Vahl.	McDowell 4613 DUKE	this paper	AJ235434	same DNA	Bremer et al., 1995	X83630
Rubiaceae	*Coffea arabica* L.	Natali and Manen 5 G	this paper	AJ235441	same DNA	Manen and Natali, 1995	X81095
Rubiaceae	*Rubia tinctorum* L.	916690 G	this paper	AJ235587	same DNA	Manen and Natali, 1995	X81104
Rutaceae	*Citrus paradisi* Macfad.	Chase 2473 K	Chase et al., 1999	AJ238408	same DNA	Chase et al., 1999	AJ238407
Rutaceae	*Poncirus trifoliata* Raf.	Chase 117 NCU	Chase et al., 1999	AJ238409	same DNA	Chase et al., 1993	AJ235806
Rutaceae	*Ptaeroxylon obliquum* Radlk.	Baker s.n. cult KIRST	Chase et al., 1999	AF066848	same DNA	Gadek et al., 1996	pers. comm.
Rutaceae	*Ruta graveolens* L.	Chase 510 K	Bakker et al., 1998	AF035913	same DNA	Gadek et al., 1996	U39281
Rutaceae	*Zanthoxylum monophyllum* P.Wilson	Chase 332 K	Bakker et al., 1998	AF035919	same DNA	Gadek et al., 1996	U39282
Sabiaceae	*Sabia swinhoei* Hemsl.	Wagner 6518 HAST	Hoot et al., 1999	AF093395	*Sabia* sp.	Chase et al., 1993	L12662
Salicaceae	*Salix reticulata* L.	Chase 840 K	this paper	AJ235590	same DNA	this paper	AJ235793
Santalaceae	*Santalum album* L.	Chase 1349 K	this paper	AJ235592	same species	Nickrent and Soltis, 1995	L26077
Santalaceae	*Thesium humile* Vahl.	M. M. A. Ghanik s.n. K	this paper	AJ235624	same DNA	this paper	AJ235797
Sapindaceae	*Acer saccharum* L.	Chase 106 NCU	Bakker et al., 1998	AF035893	same DNA	Chase et al., 1993	L01881
Sapindaceae	*Aesculus pavia* Castigl.	Chase 503 K	Bakker et al., 1998	AF035894	same DNA	Gadek et al., 1996	U39277
Sapindaceae	*Cupaniopsis anacardioides* Radl.	Chase 217 NCU	Bakker et al., 1998	AF035903	same DNA	Chase et al., 1993	L13182
Sapindaceae	*Koelreuteria paniculata* Laxm.	Chase 115 NCU	this paper	AJ235513	same DNA	Gadek et al., 1996	U39283
Sapotaceae	*Manilkara zapota* Royen	Chase 342 NCU	this paper	AJ235528	same DNA	Albert et al., 1992	L01932
Sapotaceae	*Planchonella pohlmanniana* Burkill	Chase 3184 K	this paper	AJ235398	same DNA	this paper	AJ235788
Sapotaceae	*Pouteria macrantha* Baehni	Chase 1370 K	this paper	AJ235570	*Pouteria eerwah* Baehni	Morton et al., 1996	Z80188
Sarcolaenaeaceae	*Sarcolaena* sp.	Chase 903 K	Bayer et al., 1999	AJ233070	same DNA	Fay et al., 1998	Y15147
Sarraceniaceae	*Sarracenia flava* L.	Chase 144 NCBG	this paper	AJ235594	same DNA	Albert et al., 1992	L01952
Saururaceae	*Houttuynia cordata* Thunb.	Reznicek 9238 MICH	Hoot et al., 1999	AF093397	same species	Chase et al., 1993	L08762
Saururaceae	*Saururus chinensis* Hort. ex Lond.	Qiu 91023 NCU	this paper	AJ235596	*Saururus cernuus* L.	Chase et al., 1993	L14924
Saxifragaceae	*Astilbe taquetii* Koizd.	Soltis and Soltis 2477 WS	this paper	AJ235401	same DNA	Soltis et al., 1990	L11173
Saxifragaceae	*Boykinia rotundifolia* Parry	Gornall 101 UBG	this paper	AJ235417	same DNA	Morgan and Soltis, 1993	L11175
Saxifragaceae	*Chrysosplenium iowense* Rydb.	Wendel s.n. ISC	this paper	AJ235432	same DNA	Johnson and Soltis, 1994	L19935
Saxifragaceae	*Heuchera sanguinea* Engelm.	Hoot 932 UWM	Hoot et al., 1999	AF093399	*Heuchera micrantha* Douglas	Soltis et al., 1990	L01925
Saxifragaceae	*Peltoboykinia tellimoides* Hara	Nikko Bot. Gard. Japan	this paper	AJ235554	same DNA	Soltis et al., 1993	U06213
Saxifragaceae	*Saxifraga retusa* Gouan.	Chase 778 K	this paper	AJ235597	*Saxifraga integrifolia* Hook	Morgan and Soltis, 1993	L01953
Schisandraceae	*Schisandra sphenanthera* Rehder and Wilson	Qiu 94165 NCU	this paper	AJ235599	same DNA	Qiu et al., 1993	L12665

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Scrophulariaceae	*Antirrhinum majus* L.	Chase 2570 K	this paper	AJ235395	same DNA	Olmstead et al., 1992	L11688
Scrophulariaceae	*Globularia salicina* Lam.	Chase 2547 K	this paper	AJ235481	*Globularia cordifolia* L.	Oxelman, unpubl.	AJ001764
Simaroubaceae	*Ailanthus altissima* L.	Chase 126 NCU	this paper	AF035895	same DNA	Chase et al., 1993	L12566
Simaroubaceae	*Irvingia malayana* Oliv.	Simpson 2638 K	this paper	AJ235506	same DNA	Fernando et al., 1995	AF123278
Simaroubaceae	*Simarouba glauca* DC.	Tomlinson 21623 UNSW	this paper	AJ235602	same DNA	Fernando et al., 1995	U38927
Simmondsiaceae	*Simmondsia chinensis* C.K.Schneid.	Boyd et al. 3355 F	Hoot et al., 1999	AF093401	same DNA	Fay et al., 1997	AF093732
Solanaceae	*Nicotiana tabacum* L.	unknown	Deno et al., 1983	V00162	same species	Lin et al., 1986	Z00044
Solanaceae	*Solanum nodiflorum* Desv. ex Dun.	Chase 2372 K	this paper	AJ235604	*Lycopersicon esculentum* Mill.	Olmstead et al., 1993	L14403
Stachyuraceae	*Stachyurus praecox* Siebold and Zucc.	Chase 900 K	this paper	AJ235609	same DNA	this paper	AJ235794
Stackhousiaceae	*Stackhousia minima* Hook.f.	Molloy s.n. CHR	this paper	AJ235610	same DNA	this paper	AJ235795
Staphyleaceae	*Staphylea trifoliata* Marsh	Chase 116 NCU	this paper	AJ235611	same DNA	Gadek et al., 1996	AJ238406
Stemonaceae	*Stemona japonica* Franch. and Sav.	Chase 258 NCU	this paper	AJ235612	same DNA	Chase et al., 1995	AJ131948
Styracaceae	*Styrax japonica* Siebold and Zucc.	Chase 960 K	this paper	AJ235615	*Styrax americana* Lam.	Kron and Chase, 1993	L12623
Symplocaceae	*Symplocus costata* Choisy	Chase 1374 K	this paper	AJ235617	*Symplocos paniculata* Miq.	Kron and Chase, 1993	L12624
Taccaceae	*Tacca chantrieri* André	Chase 175 NCU	this paper	AJ235618	same DNA	Chase et al., 1993	AJ235810
Taxaceae	*Taxus baccata* L.	Chase 2527 K	this paper	AJ235619	*Taxus media* Rehder	Chase et al., 1993	AJ235811
Taxodiaceae	*Metasequoia glyptostroboides* Hu and Cheng	Chase 2516 K	this paper	AJ235534	same species	Chase et al., 1993	AJ235805
Tecophilaeaceae	*Conanthera campanulata* Lindl.	Chase 523 K	this paper	AJ235442	same DNA	Chase et al., 1995	Z77311
Tecophilaeaceae	*Odontostomum hartwegii* Torc.	Chase 491 K	this paper	AJ235547	same DNA	Chase et al., 1995	Z77314
Tecophilaeaceae	*Tecophilaea cyanocrocus* Leyb.	Chase 447 K	this paper	AJ235620	same DNA	Chase et al., 1995	Z73709
Ternstroemiaceae	*Adinandra dumosa* Jack	Chase 1379 K	this paper	AJ235383	same DNA	Morton et al., 1996	Z83149
Ternstroemiaceae	*Eurya japonica* Thunb.	Chase 1448 K	this paper	AJ235474	same DNA	Morton et al., 1996	Z80207
Ternstroemiaceae	*Ternstroemia stahlii* Krug and Urb.	Axelrod 4538 UPR	this paper	AJ235621	same DNA	Morton et al., 1996	Z80211
Tetrameristaceae	*Tetramerista* sp.	Coode 7925 K	this paper	AJ235623	same DNA	Morton et al., 1996	Z80199
Theaceae	*Schima superba* Gardn. and Champ.	Chase 261 NCU	this paper	AJ235598	same DNA	Morton et al., 1996	Z80208
Theaceae	*Stuartia pseudocamellia* Maxim.	Chase 964 K	this paper	AJ235614	same DNA	Morton et al., 1996	Z80209
Theophrastaceae	*Clavija eggersiana* Mez.	Chase 216 K	this paper	AJ235437	same DNA	Kron and Chase, 1993	L12608
Thymelaeaceae	*Aquilaria beccariana* Tiegh.	Chase 1380 K	Bayer et al., 1999	AJ233079	same DNA	Fay et al., 1998	Y15149
Thymelaeaceae	*Phaleria capitata* Jack	Chase 1383 K	Bayer et al., 1999	AJ233096	*Phaleria chermsideana* C.T.White	Conti et al., 1996	U26332
Thymelaeaceae	*Thymelaea hirsuta* Endl.	Chase 1882 K	this paper	AJ235626	same DNA	Fay et al., 1998	Y152151
Tofieldiaceae	*Pleea tenuifolia* Michx.	Chase 152 NCU	this paper	AJ235564	same DNA	Chase et al., 1993	AJ131774
Tofieldiaceae	*Tofieldia calyculata* Wahlenb.	Fay sn K	this paper	AJ235627	same DNA	this paper	AJ235798
Tremandraceae	*Platytheca verticellata* Baill.	Chase 179 NCU	this paper	AJ235563	same DNA	Chase et al., 1993	L01944
Trochodendraceae	*Tetracentron sinensis* Oliv.	Qiu 90009 NCU	Hoot et al., 1999	AF093422	same species	Chase et al., 1993	L12668
Trochodendraceae	*Trochodendron aralioides* Siebold	Qiu 90026 NCU	Hoot et al., 1999	AF093423	same species	Albert et al., 1992	L01958
Tropaeolaceae	*Tropaeolum tricolor* Lindl.	Chase 2518 K	Bakker et al., 1998	AF035917	*Tropaeolum majus* L.	Price and Palmer, 1993	L14706
Turneraceae	*Turnera ulmifolia* L.	Chase 220 NCU	this paper	AJ235634	same DNA	Fay et al., 1997	Z75691
Ulmaceae	*Trema micrantha* Blume	Chase 335 NCU	this paper	AJ235628	same DNA	Chase et al., 1993	U03844
Urticaceae	*Urtica dioica* L.	Chase 2754 K	this paper	AJ235635	*Boehmeria nivea* Gaudich.	Chase et al., 1993	AJ235801
Valerianaceae	*Valeriana officinalis* L.	Chase 2524 K	this paper	AJ235637	same species	Olmstead et al., 1992	L13934

APPENDIX (CONTINUED)

	atpB				*rbcL*		
Family	Species	Voucher	Citation	EMBL/ Genbank	Species	Citation	EMBL/ Genbank
Velloziaceae	*Barbacenia elegans* Pax	Chase 253 K	this paper	AJ235406	same DNA	Chase et al., 1995	AJ131946
Verbenaceae	*Verbena scabrido-glandulosa* Turrill.	Chase 2460 K	this paper	AJ235639	*Verbena bonariensis* L.	Olmstead et al., 1993	L14412
Violaceae	*Hymenanthera alpina* Oliv.	Chase 501 K	this paper	AJ235499	same DNA	Fay et al., 1997	Z75692
Violaceae	*Rinorea bengalensis* Kuntze	Chase 2148 K	this paper	AJ235584	*Rinorea crenata* Blake	Alverson et al., 1998	AJ237591
Viscaceae	*Viscum album* L.	Sheahan sn K	this paper	AJ235642	same species	Nickrent and Soltis, 1995	L26078
Vitaceae	*Leea guineensis* G.Don.	Chase 712 K	this paper	AJ235520	same DNA	this paper	AJ235783
Vitaceae	*Vitis aestivalis* Michx.	Chase 226 NCU	this paper	AJ235643	same DNA	Albert et al., 1992	L01960
Vochysiaceae	*Vochysia rufescens* W.A.Rodrigues	Litt 14 NY	this paper	AJ235644	*Qualea* sp.	Olmstead et al., 1992	U02730
Welwitschiaceae	*Welwitschia mirabilis* Hook.	Geneva Bot Gard	this paper	AJ235645	same species	Chase et al., 1993	AJ235814
Winteraceae	*Belliolum* sp.	Qiu 90025 NCU	this paper	AJ235408	same DNA	Qiu et al., 1993	L12633
Winteraceae	*Drimys winteri* Forster and Forster	Nickrent 3013 SIU	Hoot et al., 1999	AF093425	same DNA	Albert et al., 1992	L01905
Winteraceae	*Tasmannia lanceolata* Smith	Raleigh 109 MEL	Hoot et al., 1999	AF093424	*Tasmannia insipida* DC.	Albert et al., 1992	L01957
Xeronemaceae	*Xeronema callistemon* Oliv.	Chase 653 K	this paper	AJ235647	same DNA	Chase et al., 1995	Z69235
Zygophyllaceae	*Balanites maughamii* Sprague	Sheahan sn K	this paper	AJ235405	same DNA	Sheahan and Chase, 1996	Y15016
Zygophyllaceae	*Guaiacum sanctum* L.	Chase 133 NCU	this paper	AJ235485	same DNA	Chase et al., 1993	AJ131770
Zygophyllaceae	*Viscainoa geniculata* Greene	Sheahan sn K	this paper	AJ235641	same DNA	Sheahan and Chase, 1996	Y15029