American Journal of Botany

American Journal of Botany 98(4): 704–730. 2011.

Angiosperm phylogeny: 17 genes, 640 taxa[1]

Douglas E. Soltis[2,19], Stephen A. Smith[3], Nico Cellinese[4], Kenneth J. Wurdack[5], David C. Tank[6], Samuel F. Brockington[7], Nancy F. Refulio-Rodriguez[8], Jay B. Walker[9], Michael J. Moore[10], Barbara S. Carlsward[11], Charles D. Bell[12], Maribeth Latvis[2,4], Sunny Crawley[13], Chelsea Black[13], Diaga Diouf[13,14], Zhenxiang Xi[15], Catherine A. Rushworth[15], Matthew A. Gitzendanner[2,4], Kenneth J. Sytsma[9], Yin-Long Qiu[16], Khidir W. Hilu[13], Charles C. Davis[15], Michael J. Sanderson[17], Reed S. Beaman[4], Richard G. Olmstead[8], Walter S. Judd[2], Michael J. Donoghue[18], and Pamela S. Soltis[4]

[2]Department of Biology, University of Florida, Gainesville, Florida 32611-8525 USA; [3]Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912 USA; [4]Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611-7800 USA; [5]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20013-7012 USA; [6]Department of Forest Ecology and Biogeosciences & Stillinger Herbarium, University of Idaho, P. O. Box 441133, Moscow, Idaho 83844-1133 USA; [7]Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, UK; [8]Department of Biology, University of Washington, Seattle, Washington 98195-5325 USA; [9]Department of Botany, University of Wisconsin, Madison, Wisconsin 53706 USA; [10]Biology Department, Oberlin College, Oberlin, Ohio 44074-1097 USA; [11]Department of Biological Sciences, Eastern Illinois University, Charleston, Illinois 61920 USA; [12]Department of Biological Sciences, University of New Orleans, New Orleans, Louisiana 70148 USA; [13]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 USA; [14]Département de Biologie Végétale, Université Cheikh AntDeara Diop, Dakar–Fann, BP 5005, Republic of Sénégal; [15]Department of Organismic and Evolutionary Biology, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138 USA; [16]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048 USA; [17]Department of Ecology and Evolutionary Biology, University of Arizona, 1041 East Lowell, Tucson, Arizona 85721-0088 USA; and [18]Department of Ecology and Evolutionary Biology, Yale University, P. O. Box 208106, New Haven, Connecticut 06520-8106 USA

- *Premise of the study*: Recent analyses employing up to five genes have provided numerous insights into angiosperm phylogeny, but many relationships have remained unresolved or poorly supported. In the hope of improving our understanding of angiosperm phylogeny, we expanded sampling of taxa and genes beyond previous analyses.
- *Methods*: We conducted two primary analyses based on 640 species representing 330 families. The first included 25 260 aligned base pairs (bp) from 17 genes (representing all three plant genomes, i.e., nucleus, plastid, and mitochondrion). The second included 19 846 aligned bp from 13 genes (representing only the nucleus and plastid).
- *Key results*: Many important questions of deep-level relationships in the nonmonocot angiosperms have now been resolved with strong support. Amborellaceae, Nymphaeales, and Austrobaileyales are successive sisters to the remaining angiosperms (*Mesangiospermae*), which are resolved into Chloranthales + *Magnoliidae* as sister to *Monocotyledoneae* + [Ceratophyllaceae + *Eudicotyledoneae*]. *Eudicotyledoneae* contains a basal grade subtending *Gunneridae*. Within *Gunneridae*, Gunnerales are sister to the remainder (*Pentapetalae*), which comprises (1) *Superrosidae*, consisting of *Rosidae* (including Vitaceae) and Saxifragales; and (2) *Superasteridae*, comprising Berberidopsidales, Santalales, *Caryophyllales*, *Asteridae*, and, based on this study, Dilleniaceae (although other recent analyses disagree with this placement). Within the major subclades of *Pentapetalae*, most deep-level relationships are resolved with strong support.
- *Conclusions*: Our analyses confirm that with large amounts of sequence data, most deep-level relationships within the angiosperms can be resolved. We anticipate that this well-resolved angiosperm tree will be of broad utility for many areas of biology, including physiology, ecology, paleobiology, and genomics.

Key words: angiosperms; bioinformatics; large data sets; molecular systematics; RAxML; S*uperasteridae*; supermatrix; *Superrosidae*.

[1] Manuscript received 11 October 2010; revision accepted 9 February 2011.

The authors thank the University of Florida High-Performance Computing Center, the UF Genetics Institute, and facilities supported in part by NSF Grant CNS-0821622 for providing computational resources that facilitated the completion of this research. The authors thank E. Mavrodiev for help in preparing some of the figures. This study was carried out as part of the Angiosperm Tree of Life Project (NSF EF-0431266 to D.E.S., P.S.S., W.S.J., and S. R. Manchester, EF-0431242 to C.C.D., DEB-9634231; EF-0431057 to K.H., EF-0431233 to K.J.S., DEB-0431239 to Y.L.Q.) with additional support from NSF grant DEB-0544039 to C.C.D. The authors thank the Royal Botanic Gardens, Kew, and the Missouri Botanical Garden for providing samples used in this study.

[19] Author for correspondence (e-mail: dsoltis@botany.ufl.edu)

doi:10.3732/ajb.1000404

It has been a decade since the 567-taxon, three-gene (*rbcL, atpB*, and 18S rDNA) parsimony-based phylogenetic analysis of angiosperms was published (Soltis et al., 1999, 2000). Since that time, other studies have expanded sampling of additional genes or used alternative methods to evaluate the results of that three-gene study. Hilu et al. (2003) conducted a broad analysis of angiosperms based on *matK* sequences, with results that agreed closely with the three-gene topology. Additional progress was achieved by Davies et al. (2004) by constructing a supertree for angiosperms. Soltis et al. (2007) undertook a Bayesian analysis of the 567-taxon, three-gene data set and obtained a topology nearly identical to that obtained with parsimony. More recently, Bell et al. (2010) analyzed this same three-gene data set using a Bayesian relaxed clock model to simultaneously infer topology and divergence times within angiosperms and found results similar to Soltis et al. (2007). In addition, Burleigh et al. (2009) inferred angiosperm phylogeny using five genes for the same 567 taxa analyzed in Soltis et al. (1999, 2000). Although the five-gene matrix had significantly more missing data (27.5%) than the three-gene matrix (2.9%), the five-gene analysis resulted in higher levels of bootstrap support across the tree.

Recent phylogenomic analyses have shown the value of constructing data sets of many genes to infer deep-level angiosperm phylogeny. Some of these analyses have employed nearly complete plastid genome sequence data (e.g., Leebens-Mack et al., 2005; Jansen et al., 2007; Moore et al., 2007, 2010), but all have been limited in sampling to fewer than 100 taxa. These and other studies based on many genes but focused only on major angiosperm clades (e.g., Schönenberger et al., 2005; Jian et al., 2008; H. Wang et al., 2009; Wurdack and Davis, 2009; Brockington et al., 2010; Tank and Donoghue, 2010) showed that with very large amounts of data (i.e., 13 to 83 genes), many, if not most, deep-level questions of angiosperm phylogeny can be resolved.

While the three- and five-gene analyses of 567 taxa have broad taxonomic coverage, support for many portions of the framework of angiosperm phylogeny is low in these studies. For example, relationships among members of *Mesangiospermae* and of *Pentapetalae* remain unclear. Conversely, studies employing complete plastid genome sequences have deep gene coverage and strong internal support, but taxonomic coverage is often sparse. Hence, it is important to assemble a data set having both broad taxonomic coverage as well as numerous genes. In the hope of further improving our understanding of angiosperm phylogeny, we have sequenced deeply over a broad representation of angiosperms. We constructed a 17-gene data set for 640 species representing 640 genera, 330 families, and 58 of 59 orders (sensu APG III, 2009) using genes that represent all three plant genomic compartments (nucleus, plastid, and mitochondrion).

One goal of the angiosperm Assembling the Tree of Life (AToL) project was to assess morphological synapomorphies for each of the major clades of angiosperms as well as across angiosperms as a whole. We here also provide putative morphological synapomorphies of Saxifragales, although a more detailed morphological exploration of this clade is presented elsewhere (Carlsward et al., unpublished manuscript). This study of Saxifragales represents the first treatment in a series of analyses for large clades of angiosperms that will address morphological synapomorphies and the clades they support (W. S. Judd et al., unpublished data).

MATERIALS AND METHODS

Data sets—DNA samples for most of the species used here were extracted from either fresh or silica-dried material following the general method of Doyle and Doyle (1987) or modifications thereof that employ liquid nitrogen and higher CTAB concentrations (e.g., Soltis et al., 1991; Sytsma, 1994). We attempted to use the same species and DNA samples across all of the genes analyzed here, although multiple species were sometimes used as necessary placeholders to reduce missing data. Many of these DNA samples have been used in earlier analyses (e.g., Chase et al., 1993; Soltis et al., 2000).

We constructed a 17-gene data set for 640 species (for a complete list of taxa, voucher information, and GenBank numbers see Appendix S1, see Supplemental Data online at http://www.amjbot.org/cgi/content/full/ajb.1000404/DC1) using genes from the nuclear, plastid, and mitochondrial genomes. Given the potential problems inherent in phylogeny reconstruction using mtDNA sequences (e.g., RNA editing, Bowe and dePamphilis, 1996; horizontal gene transfer or HGT, Bergthorsson et al., 2003), we also constructed a data set without the mtDNA data, resulting in a matrix of 13 genes. For both data sets, we used the following representatives of *Acrogymnospermae* (extant gymnosperms, sensu Cantino et al., 2007) as the outgroup: *Cycas*, *Ginkgo*, *Gnetum*, *Metasequoia*, *Pinus*, *Podocarpus*, *Welwitschia*, and *Zamia.* Taxon sampling across major clades of angiosperms was not uniform, with poorly resolved clades (e.g., Malpighiales, Saxifragales) targeted for denser taxon sampling to seek improved resolution. We also largely avoided parasitic clades (except Orobanchaceae, Santalales, and *Cuscuta*), which can create analytical problems due to gene loss, accelerated molecular evolution, and horizontal gene transfer (see Davis and Wurdack, 2004; Nickrent et al., 2004; Barkman et al., 2007).

The following 17 genes were sequenced: 18S and 26S rDNA from the nuclear genome; *atpB*, *matK*, *ndhF*, *psbBTNH* (four contiguous genes here treated as one region), *rbcL*, *rpoC2*, *rps16*, and *rps4* from the plastid genome; and *atp1*, *matR*, *nad5*, and *rps3* from the mitochondrial genome. The total length of the aligned 17-gene matrix was 25 260 bp and of the 13-gene matrix was 19 846 bp. The percentage of missing data for the full data set was 41% and for the data set without mtDNA data was 42%.

Alignment and phylogenetic analyses—All sequence data were stored and managed in TOLKIN (Beaman and Cellinese, 2010). TOLKIN is a web application, developed for distance collaboration as part of the Angiosperm Tree of Life project, that allows users to access and share data in real time, as well as automatically generate FASTA files and link to other relevant information (e.g., taxonomy and vouchers) and resources (e.g., GenBank, TreeBASE, and uBIO). The sequences generated here were supplemented with those already available in GenBank to obtain a more complete data set. GenBank sequences were retrieved using the PHLAWD package (Smith et al., 2009; http://code.google.com/p/phlawd), and alignments of combined sequences were generated with the program MAFFT (vers. 6.71; Katoh and Toh, 2008) at the DNA level using the l-ins-i algorithm and default alignment parameters. MAFFT was chosen because of its strong performance over a range of alignment scenarios (Golubchik et al., 2007). Subsequent adjustments were made by eye when there were obvious alignment errors due to particularly divergent or "gappy" sequences. The individual gene regions varied in the amount of missing data per site: 18S rDNA (6%), 26S rDNA (15%), *atpB* (5%), *atp1* (1%), *matK* (13%), *matR* (3%), *nad5* (4%), *ndhF* (20%), *psbBTNH* (19%), *rbcL* (4%), *rpoC2* (21%), *rps16* (26%), *rps3* (10%), and *rps4* (50%). Individual gene regions also varied in the number of taxa with data in the combined analyses: 18S rDNA (78%), 26S rDNA (57%), *atpB* (88%), *atp1* (59%), *matK* (92%), *matR* (76%), *nad5* (59%), *ndhF* (80%), *psbBTNH* (54%), *rbcL* (98%), *rpoC2* (63%), *rps16* (35%), *rps3* (62%), and *rps4* (58%). Sites in the alignment with more than 50% missing data were removed with the program Phyutility (Smith and Dunn, 2008; see discussion in Castresana, 2000) to avoid regions of potentially problematic ambiguous alignment caused by such broad sampling (Talavera and Castresana, 2007).

Phylogenetic analyses using maximum likelihood (Felsenstein, 1973) were conducted in the program RAxML (vers. 7.1; Stamatakis, 2006). For each data set, we searched for the optimal tree, running at least 10 independent maximum likelihood analyses; full analyses also consisted of at least 100 and up to 300 bootstrap replicates (Stamatakis et al., 2008). We conducted analyses on all individual genes, the concatenated 17-gene data set, the concatenated 13-gene data set (no mtDNA data), and genomic compartments (nucleus, plastid, and mitochondrion). The GTRGAMMA substitution model was applied to each gene independently. For analyses of all concatenated data sets, all genes

were partitioned, and unlinked substitution models were applied to each gene. Bootstrapped (BS; Felsenstein, 1985) trees were summarized as majority-rule consensus trees with Phyutility (Smith and Dunn, 2008).

A series of maximum likelihood (ML) analyses was conducted on the combined data sets, ultimately requiring approximately over 1 year of analysis time and approximately 9 years of actual CPU time. Unexpected taxon placements required close examination of the alignment and subsequent reanalysis when problems were detected. For example, initial analyses indicated that the placement of *Polyosma* (Escalloniaceae) differed dramatically when mtDNA data were included vs. when they were omitted for this taxon (see Results). Partitioned analyses suggested that the mtDNA signal for *Polyosma* was providing a spurious placement. In our total evidence analysis (Kluge, 1989; Fig. 1, summary tree, and Fig. 2, complete cladogram; shown as a phylogram in Appendix S1), we therefore omitted the mtDNA data for *Polyosma*. We subsequently sequenced the mitochondrial genes surveyed here for an additional sample of *Polyosma* as well as for a second accession of *Quintinia*, the sister of *Polyosma* in the trees that included mtDNA sequence data.

Maximum parsimony (MP) has not performed well on some recent large angiosperm data sets; long-branch attraction has clearly played a role in previous analyses of deep-level relationships in angiosperms, especially in analyses with limited taxon sampling (reviewed in Leebens-Mack et al., 2005; Soltis et al., 2005, 2007; Burleigh et al., 2009). Given the current availability of ML programs (e.g., RAxML) that can readily handle large data sets of the size employed here, we focused much of our data analyses on this approach. Nonetheless, we also conducted MP searches, using the parsimony ratchet (Nixon, 1999) approach to thorough and rapid tree searching. These analyses were run in the program PAUP* 4.0b10 (Swofford, 2002). Ratchet files were generated with PAUPRat (Sikes and Lewis, 2001) with 50 independent replicates of 500 iterations each. A majority-rule consensus of the best trees from each replicate was generated. Bootstrapping was conducted by generating 500 bootstrap data sets with the SeqBoot module of PHYLIP (Felsenstein, 2005) and running each of these with a PAUPRat-generated ratchet file for a single 500-iteration search. The parsimony searches took just over three CPU years of analysis time. It is noteworthy therefore that, for this data set, a thorough MP analysis took longer than a comparable ML analysis (each total evidence ML analysis required one CPU year of analysis time).

For higher clades, we consistently use PhyloCode names (see Cantino et al., 2007) whenever these are available; these names are always in italics (e.g., *Pentapetalae*, *Mesangiopsermae*, *Rosidae*, *Fabidae*, *Malvidae*). Note that *Rosidae* (sensu Cantino et al., 2007) does include Vitaceae. Our use of family and ordinal names follows APG III (2009) as a formal point of reference; for *Caryophyllales*, we follow Cantino et al. (2007; hence, the use of italics), which matches the APG III circumscription. For additional recent discussion on families and their status, see the Angiosperm Phylogeny Website (Stevens, 2001 onward). We recognize that some broader family circumscriptions favored in APG III are controversial and can obscure underlying diversity (e.g., Passifloraceae s.l.), which would be evident with narrower circumscriptions. All alignments and trees have been deposited in TreeBASE (no. 11267; see http://www.treebase.org/).

After all analyses were complete, the possibility was raised that the *atp1* sequence of *Cardiopteris* was a contaminant. This sequence did not impact the final placement of the genus, but as a result the questionable sequence was not submitted to GenBank.

RESULTS

Each ML analysis of the 17- and 13-gene data sets took 20–32 h on a 32-core (2.93 gHz Xeon ×7350) machine with 128 gb RAM, and analyses of individual genes took 1–11 h.

The best RAxML trees from analyses of the 17-gene and 13-gene matrices are very similar, but with a few noteworthy differences. The 17-gene tree has one major deviation in placement from expected relationships (*Polyosma*; see above). However, other differences are relatively minor, and in most cases the 17-gene tree gives higher BS support than the 13-gene tree. Hence, only the 17-gene tree (total evidence, but no mtDNA data for *Polyosma*; e.g., Lecointre and Deleporte, 2005) is discussed below. This tree (Fig. 1, summary tree) has been divided into separate, interconnected subtrees (Fig. 2a-l).

The 17-gene tree (with mtDNA for all taxa, including *Polyosma*) and the 13-gene tree (without mtDNA) are presented as online supplementary figures (Appendix S2 and Appendix S3, respectively). Additional online supplementary figures are provided for the trees resulting from the analysis of each genomic partition: nuclear rDNA tree (Appendix S4), plastid tree (Appendix S5), and mtDNA tree (Appendix S6), as well as the single-gene trees (Appendices S7–S20; note that *psbBNTH*, representing four genes, are combined in one tree). We also provide a tree resulting from analysis of a reduced, 17-gene, 86-taxon data set (Appendix S21) for comparison with a recent 86-taxon tree based on complete plastid genome sequences (Moore et al., 2010).

Finally, we also provide an MP total evidence topology, with mtDNA for all taxa, except *Polyosma* (Appendix S22). As reviewed below, the MP topology (Appendix S22) is very similar to the ML tree (Figs. 1, 2).

Taxon sampling is so dense, it is decisive for all possible topologies in the sense of Steel and Sanderson (2010; see also Sanderson et al., 2010). This means that there is sufficient taxon coverage to avoid problems of lack of resolution due solely to missing data (though lack of resolution stemming from insufficient sequence information may still be an issue).

DISCUSSION

Overview—Broad phylogenetic analyses involving three to five genes have provided important insights into angiosperm phylogeny, but crucial portions of the backbone of the tree were either not resolved or not well supported (e.g., relationships among major lineages of *Mesangiospermae* and of *Pentapetalae*). Use of large numbers of genes from nearly compete plastid genome sequences established most deep branches of extant angiosperm diversity with strong support, but taxon density was low (Jansen et al., 2007; Moore et al., 2007, 2010). Here, we have employed numerous genes as well as broad taxonomic coverage. The general topology provided here is very similar to recently published trees that include broad taxonomic coverage of the angiosperms, albeit with fewer genes (e.g., Soltis et al., 2000, 2007; Burleigh et al., 2009) or far fewer taxa and nearly complete plastid genome sequences (e.g., Jansen et al., 2007; Moore et al., 2007, 2010). However, the 17-gene tree is a significant improvement in that it provides much higher levels of support for deep-level relationships than obtained with either three or five genes. Our analyses further support those topologies recovered using complete plastid genome analyses (but based on fewer than 100 exemplars).

Within *Acrogymnospermae*, the two Gnetales included here (*Gnetum* and *Welwitschia*) are embedded within the *Coniferae* (conifers) and are sister to Pinaceae (BS = 81%) in any potential rooting other than one in which Gnetales are sister to all other seed plants.

Within the angiosperms, Amborellaceae, Nymphaeales, and Austrobaileyales are subsequent sisters to all other extant flowering plants. These placements all receive BS support (BS) >80%; Fig. 1) and are a result obtained across 17-gene and 13-gene data sets. It is noteworthy, however, of the genome partitions, the mtDNA data have *Amborella* + Nymphaeales (BS = 88%), the cpDNA partition has *Amborella* + Nymphaeales (BS = 78%), and the 18S/26S rDNA partition has Nymphaeales followed by *Amborella* as sister to other angiosperms (BS < 50%; Appendices S4–S6). Our interpretation of these results is that the support

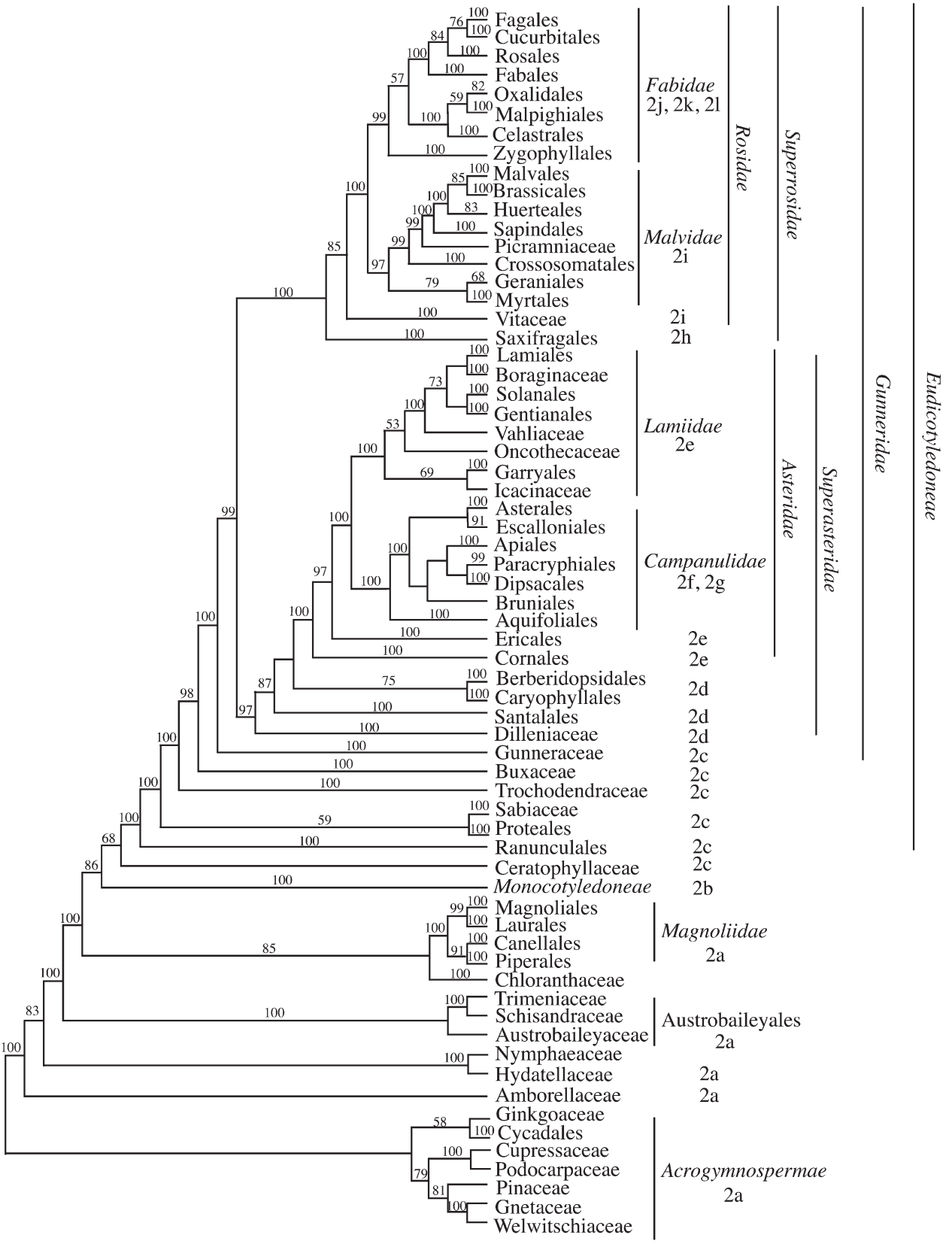

Fig. 1. Summary tree of the maximum likelihood majority-rule consensus from the 17-gene analysis with mtDNA data removed for *Polyosma*. Names of the orders and families follow APG III (2009); other names follow Cantino et al. (2007). Numbers above branches are bootstrap percentages. The number and letter following clade names (i.e., 2a to 2l) refer to clade designations that are used to depict the separate portions of the complete tree in Fig. 2.

from this particular data set for the placement of *Amborella* is not strong. Nonetheless, the placement of *Amborella* as sister to all other extant angiosperms is well supported using complete plastid genome sequences (e.g., Leebens-Mack et al., 2005; Jansen et al., 2007; Moore et al., 2007), as well as in recent analyses of numerous nuclear gene trees (Jiao et al., in press).

Following Amborellaceae, Nymphaeales, and Austrobaileyales, the monophyly of the remaining angiosperms (*Mesangiospermae*) received maximum support. Chloranthales are sister to the magnoliid clade (*Magnoliidae*) with BS = 85%. This clade of *Magnoliidae* + Chloranthales is in turn sister to the remaining angiosperms. *Monocotyledoneae* (monocots), Ceratophyllaceae, and *Eudicotyledoneae* (eudicots) form a well-supported clade (BS = 86%) with *Monocotyledoneae* sister to a weakly supported clade (BS = 68%) of Ceratophyllaceae + *Eudicotyledoneae*. These deep-level relationships agree with those obtained using complete plastid genome sequences (Moore et al., 2007).

Within *Eudicotyledoneae*, there is a grade of basal taxa; in contrast to many previous studies, relationships among the members of this basal grade are well supported. Successively more distant sister groups from *Gunneridae* (= core eudicots) are Buxaceae, Trochodendraceae, Proteales plus Sabiaceae, and Ranunculales.

Within *Gunneridae*, Gunnerales are sister to the remainder of this clade (BS = 99%), which constitute *Pentapetalae. Pentapetalae* comprises (1) a well-supported (BS = 100%) "superrosid" clade, here named *Superrosidae* (see below), consisting of *Rosidae* (including Vitaceae) and Saxifragales; and (2) a well-supported (BS = 87%) "super-asterid" clade, here named *Superasteridae* (see below), consisting of Berberidopsidales, Santalales, *Caryophyllales*, *Asteridae*, and Dilleniaceae, whose position here is at odds with other recent studies (see below).

The MP topology (Appendix S22) is similar to the ML tree (Figs. 1, 2), but there are several noteworthy differences that have also been reported in previous analyses involving MP and ML on smaller data sets (see Soltis et al., 2005, 2007). For example, in the MP tree, Ceratophyllaceae are sister to the *Monocotyledoneae* rather than *Eudicotyledoneae*, as found with ML. MP does recover the *Superasteridae* and *Superrosidae* clades, but there are differences in relationships among members of *Superasteridae* with MP and ML. For example, with MP, Dilleniaceae are sister to *Caryophyllales* (rather than sister to all other *Superasteridae*, as found with ML). We will not discuss the MP topology further.

Basal angiosperms—Our 17-gene analysis (Figs. 1, 2) places Amborellaceae, followed by Nymphaeales, and then Austrobaileyales as well-supported sisters to all other extant angiosperms, in agreement with a series of phylogenetic analyses that have been based on an ever-increasing number of gene sequences (reviewed in Leebens-Mack et al., 2005; Soltis et al., 2005).

Our results also agree with other analyses in providing strong support (BS = 100%) for the placement of Hydatellaceae in Nymphaeales as sister to other members of the clade (Saarela et al., 2007). *Trithuria* has ascidiate carpel development, consistent with placement in Nymphaeales, but as reviewed elsewhere, this placement is important in that it greatly expands the morphological diversity encompassed by Nymphaeales (Rudall et al., 2007).

Following *Amborella* and Nymphaeales, an Austrobaileyales clade (BS = 100%) of Schisandraceae, Austrobaileyaceae, and Trimeniaceae is sister to all remaining angiosperms. Within Austrobaileyales, Austrobaileyaceae are sister to Trimeniaceae + Schisandraceae (including Illiciacae; see APG III, 2009).

Magnoliidae + Chloranthales—Chloranthaceae (the only member of Chloranthales, APG III, 2009) are well supported here as sister to *Magnoliidae* (BS = 85%). This same sister-group relationship also emerged (with BS = 72%) from complete plastid genome sequencing (Moore et al., 2010), but has not been apparent in previous analyses involving three or five genes (Soltis et al., 2000; Burleigh et al., 2009). Within Chloranthaceae, the relationships among the four genera [*Hedyosmum*, (*Ascarina* (*Sarcandra* + *Chloranthus*)] agree with the results of focused analyses of the family (Qiu et al., 1999; Doyle et al., 2003; Zhang and Renner, 2003; Eklund et al., 2004).

In agreement with other recent analyses (e.g., Zanis et al., 2002; Hilu et al., 2003; Qiu et al., 1999, 2005; Soltis et al., 2007; Moore et al., 2010), *Magnoliidae* comprises two well-supported clades, each with BS = 100%: Magnoliales + Laurales and Piperales + Canellales. Relationships within these four clades also agree, for the most part, with previous analyses and are summarized below.

Piperales consist of two well-supported clades: Piperaceae + Saururaceae (BS = 100%) as sister to Aristolochiaceae + Lactoridaceae (BS = 99%). The parasitic Hydnoraceae, also of Piperales (APG III, 2009), were not included here. Canellales comprise Canellaceae + Winteraceae.

In Magnoliales, Magnoliaceae are sister to the remaining Magnoliales: Degeneriaceae + Myristicaceae are sister to Himantandraceae + [Eupomatiaceae + Annonaceae], albeit without BS >50%. These results differ from the three-gene (Soltis et al., 2000, 2007) and five-gene (Burleigh et al., 2009) analyses, as well as focused studies on the clade (Sauquet et al., 2003), all of which have placed Myristicaceae as sister to all other Magnoliales. The problem with the placement of Myristicaceae here appears to be the result of the rDNA data. Both plastid and mtDNA place Myristicaceae as sister to all remaining Magnoliales, following earlier analyses. However, the combined 18S + 26S rDNA tree places a strongly supported clade (BS = 94%) of Myristicaceae plus Degeneriaceae (all with very long branches) as embedded well within Magnoliales, sister to Annonaceae (Appendix S4). This result was not apparent in earlier studies that included 18S rDNA (Soltis et al., 2000). However, in this study, 18S rDNA placed Myristicaceae with Chloranthaceae (albeit without BS support >50%), and the 26S rDNA data placed the family as indicated for the combined data set with strong BS support.

Within Laurales, Calycanthaceae are well supported as sister to the remainder of the clade, as in previous analyses (e.g., Qiu et al., 1999; Soltis et al., 1999, 2000). The current analyses also provide more resolution and support of relationships within the remainder of Laurales than the three- and five-gene studies (Soltis et al., 2000; Burleigh et al., 2009), although reasonable resolution and support were obtained with six genes and morphology (e.g., Renner, 1999). Our trees then divide the remaining Laurales into two subclades: (1) Lauraceae sister to (Hernandiaceae + Monimiaceae), and (2) Siparunaceae sister to (Atherospermataceae + Gomortegaceae). These same two clades were recovered by Renner (1999), although the relationships found here within the two clades are not identical to those recovered in that study.

Monocotyledoneae—The monocots, or *Monocotyledoneae*, are not extensively sampled here, in deference to the ongoing

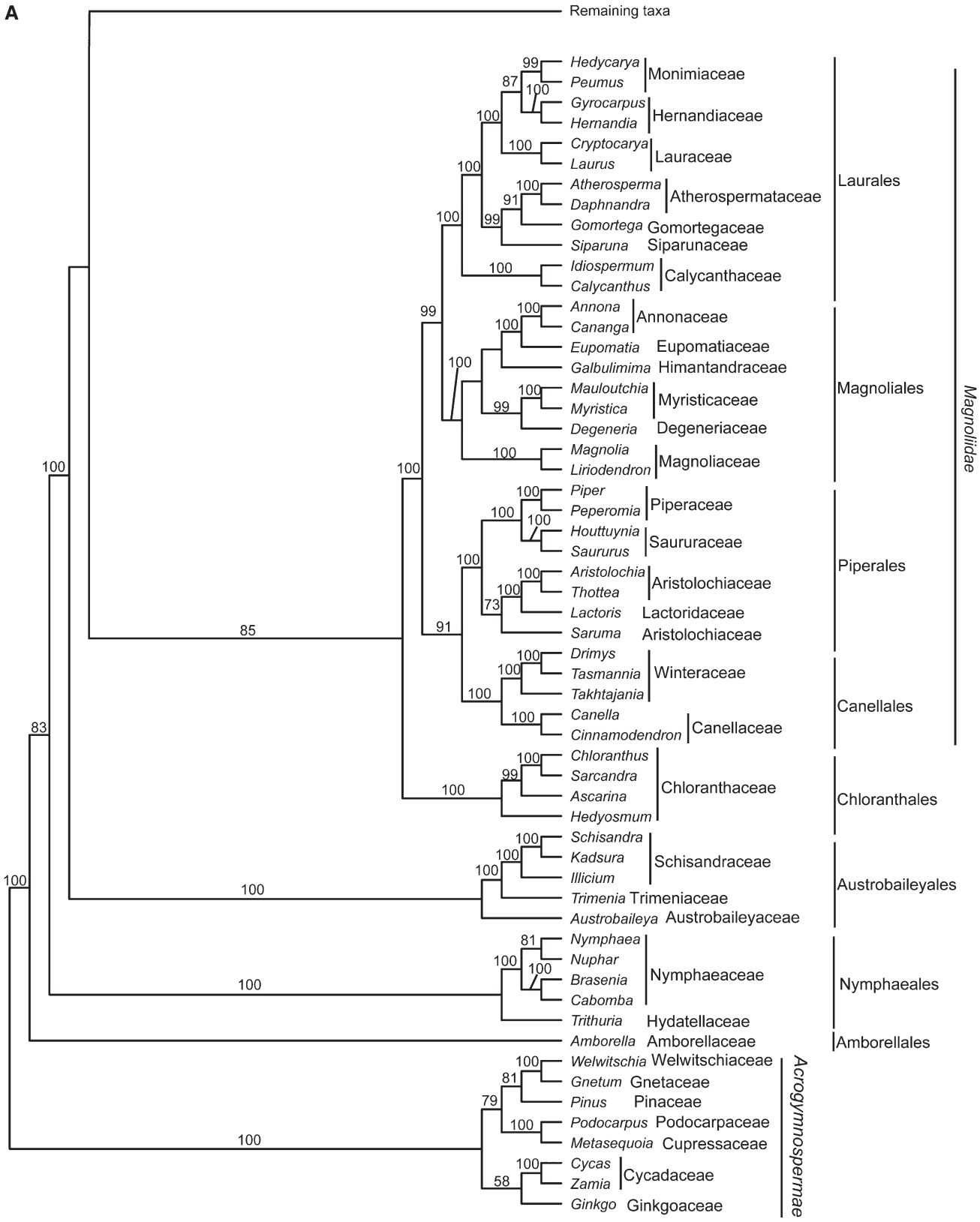

Fig. 2. The maximum likelihood majority-rule consensus from the 17-gene analysis shown as a cladogram with mtDNA data removed for *Polyosma*. This single large tree has been divided into a series of interconnected trees in which subclades are labeled 2a through 2l; these designations match those given in Fig. 1. Names of the orders and families follow APG III (2009); other names follow Cantino et al. (2007). Numbers above branches are bootstrap percentages.

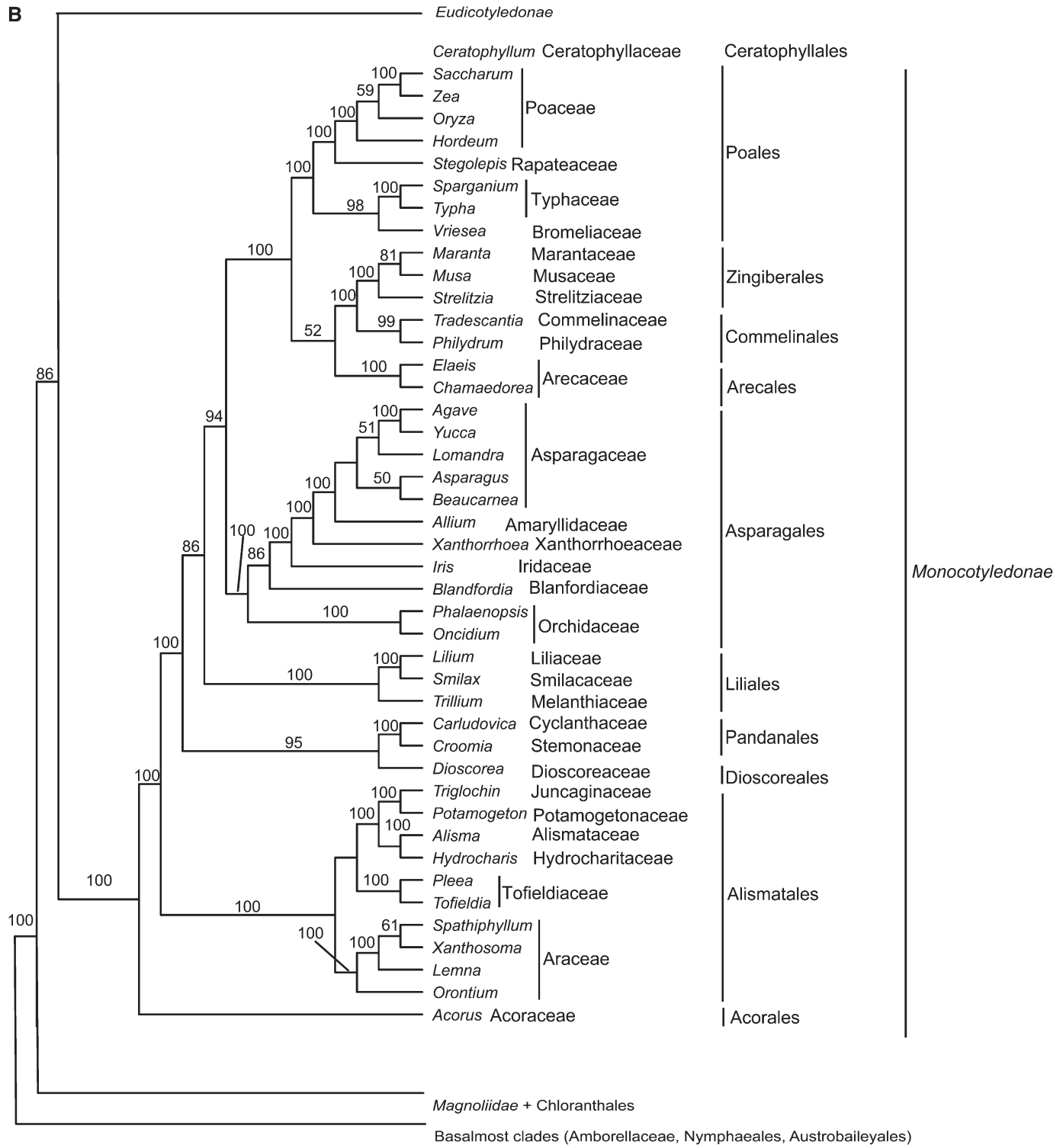

Fig. 2. Continued.

work of the Monocot AToL research group (see Givnish et al., 2010). Nonetheless, the relationships depicted here within the *Monocotyledoneae* mirror those produced in multigene phylogenetic analyses focused on the clade (e.g., Chase et al., 2006; Graham et al., 2006). On a broad scale, *Monocotyledoneae* is well supported (BS = 100%), with Acoraceae (Acorales), followed by Alismatales, each well supported (BS = 100%) as subsequent sisters to all other monocots (Petrosaviales were not sampled). A clade of Dioscoreales + Pandanales (BS = 95%) is sister to the remaining *Monocotyledoneae*. Within this remainder, Liliales followed by Asparagales are sisters to *Commelinidae* (i.e., commelinid clade; BS = 100%; Arecales, Zingiberales,

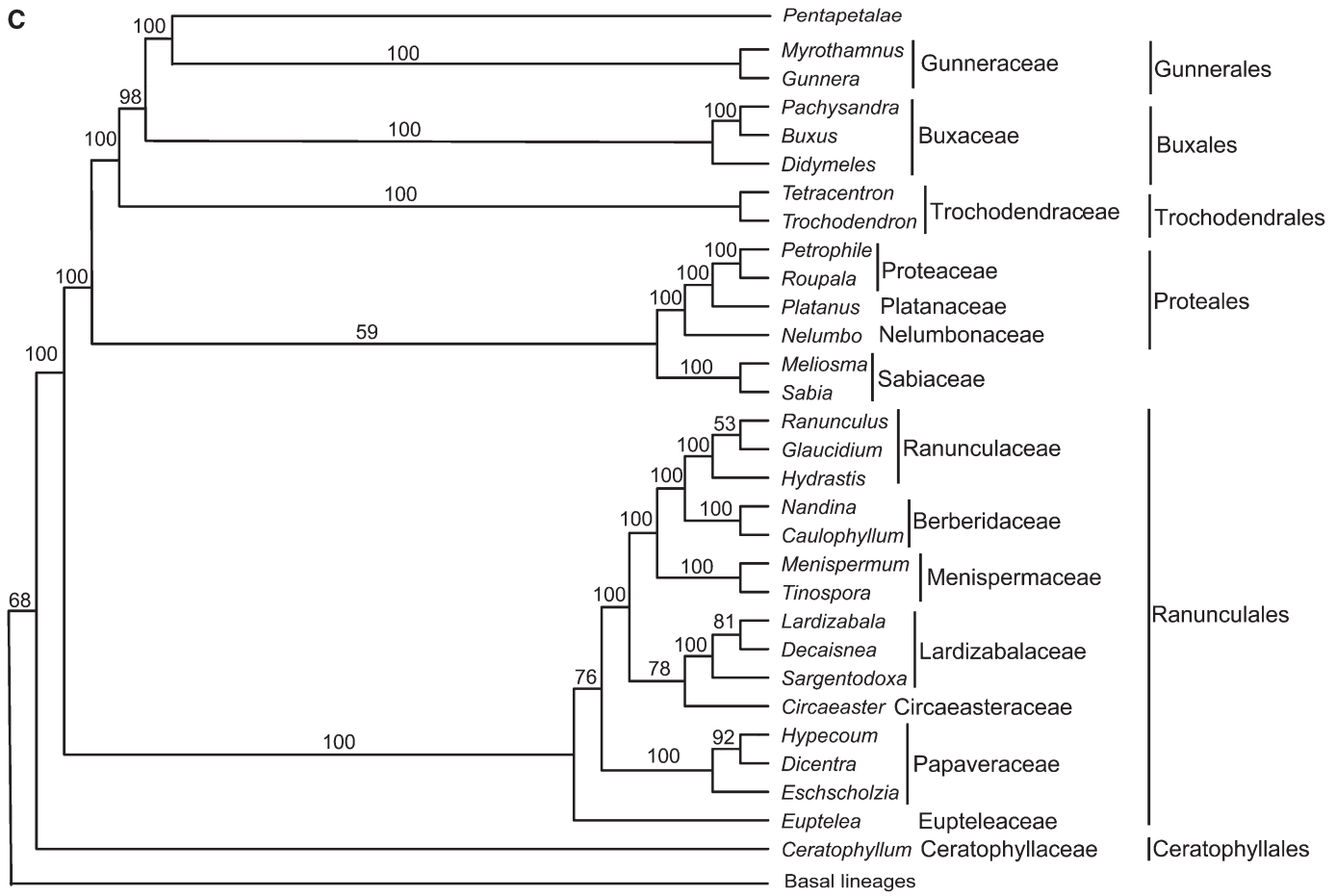

Fig. 2. Continued.

Arecales, Poales, and Dasypogonaceae, a family not included here). Relationships within *Commelinidae* also agree with focused analyses of *Monocotyledoneae*, with Arecales sister to Commelinales + Zingiberales; this clade is in turn sister to Poales. Relationships obtained here within the larger *Monocotyledoneae* clades also largely agree with more focused analyses that have included more taxa (see Chase et al., 2006; Graham et al., 2006; Givnish et al., 2010).

Eudicotyledoneae—*Basal eudicots*—Basal eudicot relationships parallel those in other broad analyses (e.g., Soltis et al., 1999, 2000; Burleigh et al., 2009), albeit with higher BS support here. Ranunculales (BS = 100%) are sister to remaining *Eudicotyledoneae*, which form a clade with strong support (BS = 100%); Ranunculales are then followed by a clade of Sabiaceae + Proteales (BS = 59% in total evidence tree; 65% without mtDNA data); a similar sister-group relationship was recovered with strong support (BS = 80%) with complete plastid genome data (Moore et al., 2010). Following Proteales, both analyses provided strong support for the placements of Trochodendraceae (BS = 100%) followed by Buxaceae (BS = 98%), as subsequent sisters to the core eudicots (*Gunneridae*). The same relationships were recovered by Moore et al. (2010).

Relationships within the basal eudicot clades are generally well resolved and supported. Within Ranunculales, Eupteleaceae are sister to the remaining taxa (BS = 76%), followed by Papaveraceae, which are well supported (BS = 100%) as sister to the rest of the clade. The remaining Ranunculales form two clades: (1) Circaeasteraceae + Lardizabalaceae (BS = 78%), and (2) Menispermaceae sister to (Berberidaceae + Ranunculaceae) (BS = 100%). These results are comparable to most recent analyses (e.g., Kim et al., 2004; W. Wang et al., 2009), although there has been disagreement in terms of the sister group to all other Ranunculales. In some analyses, Papaveraceae have appeared in this position (Soltis et al., 2000), whereas in other studies it has been Eupteleaceae (Kim et al., 2004; W. Wang et al., 2009).

Within Proteales, Nelumbonaceae are sister to Platanaceae + Proteaceae (with BS = 100%) as in other molecular analyses; recent investigations of floral morphology also indicate similarities between Proteaceae and Platanaceae (von Balthazar and Schönenberger, 2009).

Gunneridae—Gunnerales are sister to all other core eudicots (i.e., *Pentapetalae*) with BS = 100%. *Pentapetalae*, in turn, comprises: (1) the *Superrosidae*, consisting of *Rosidae* (including Vitaceae) and Saxifragales (BS = 100%); and (2) the *Superasteridae*, consisting of Berberidopsidales, Santalales, *Caryophyllales*, *Asteridae*, and possibly Dilleniaceae (BS = 97%).

Superrosidae—*Superrosidae* consists of Saxifragales as sister to a well-supported (BS = 85%) *Rosidae*; *Rosidae* comprises Vitaceae as sister to *Malvidae* + *Fabidae* (see

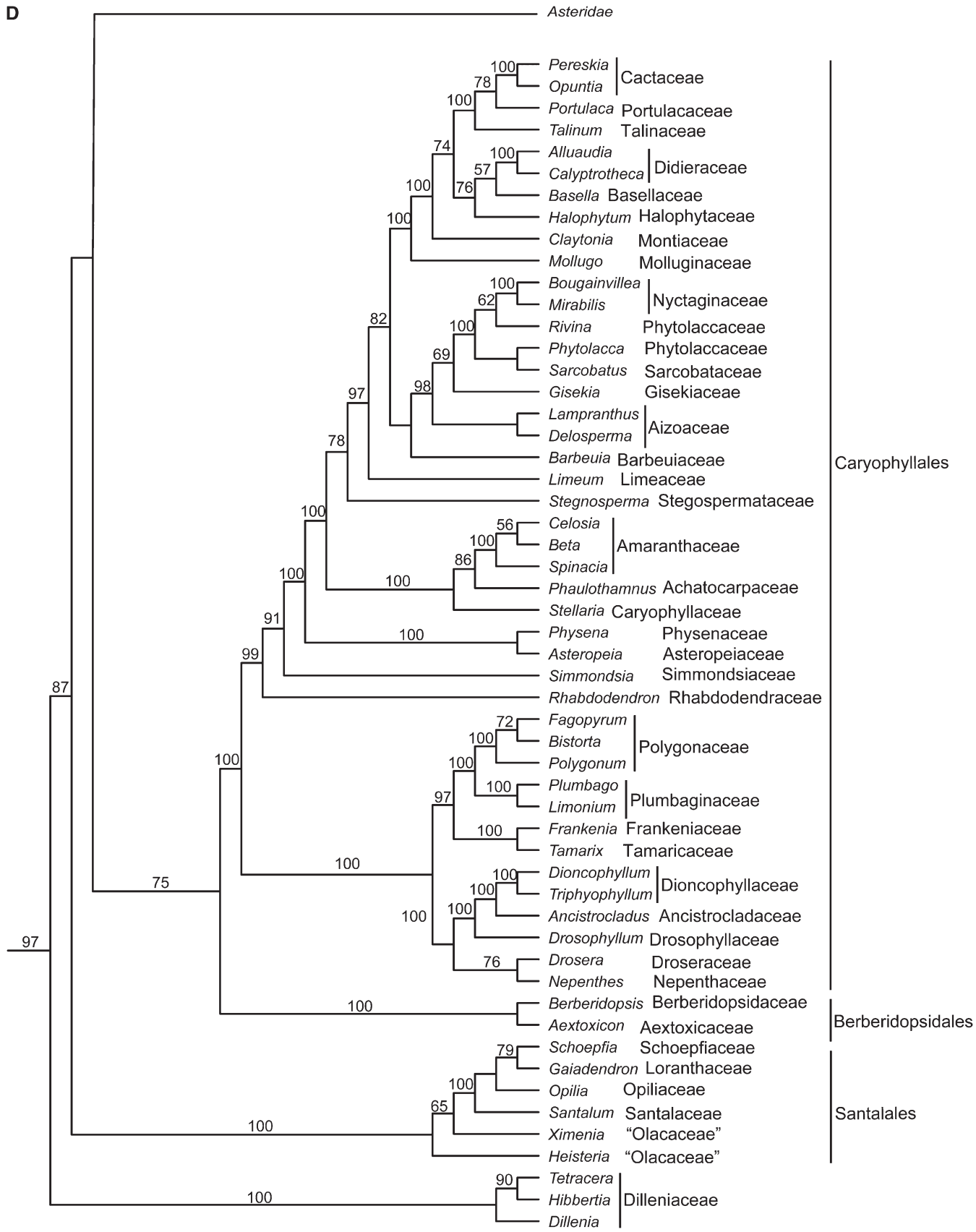

Fig. 2. Continued.

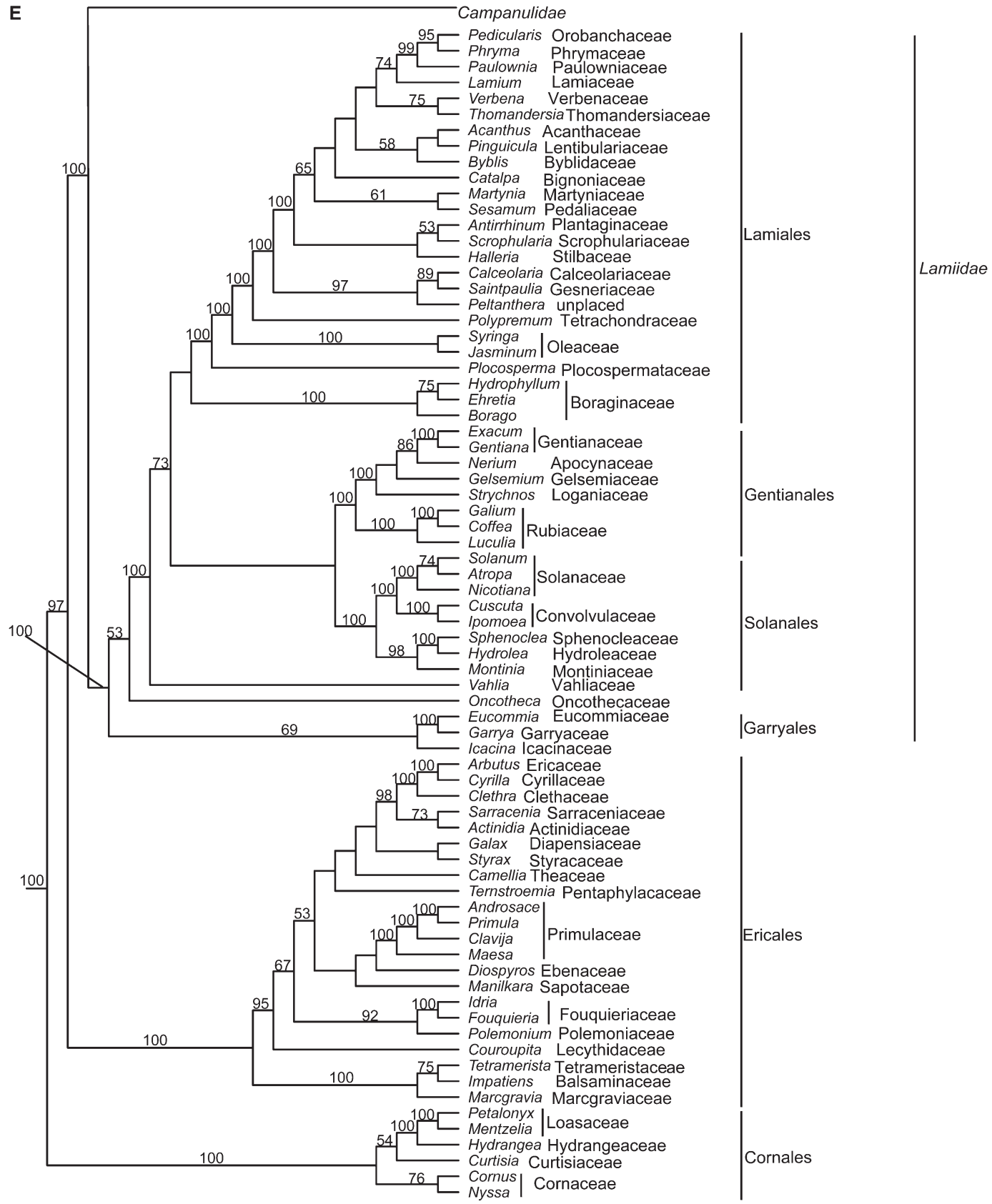

Fig. 2. Continued.

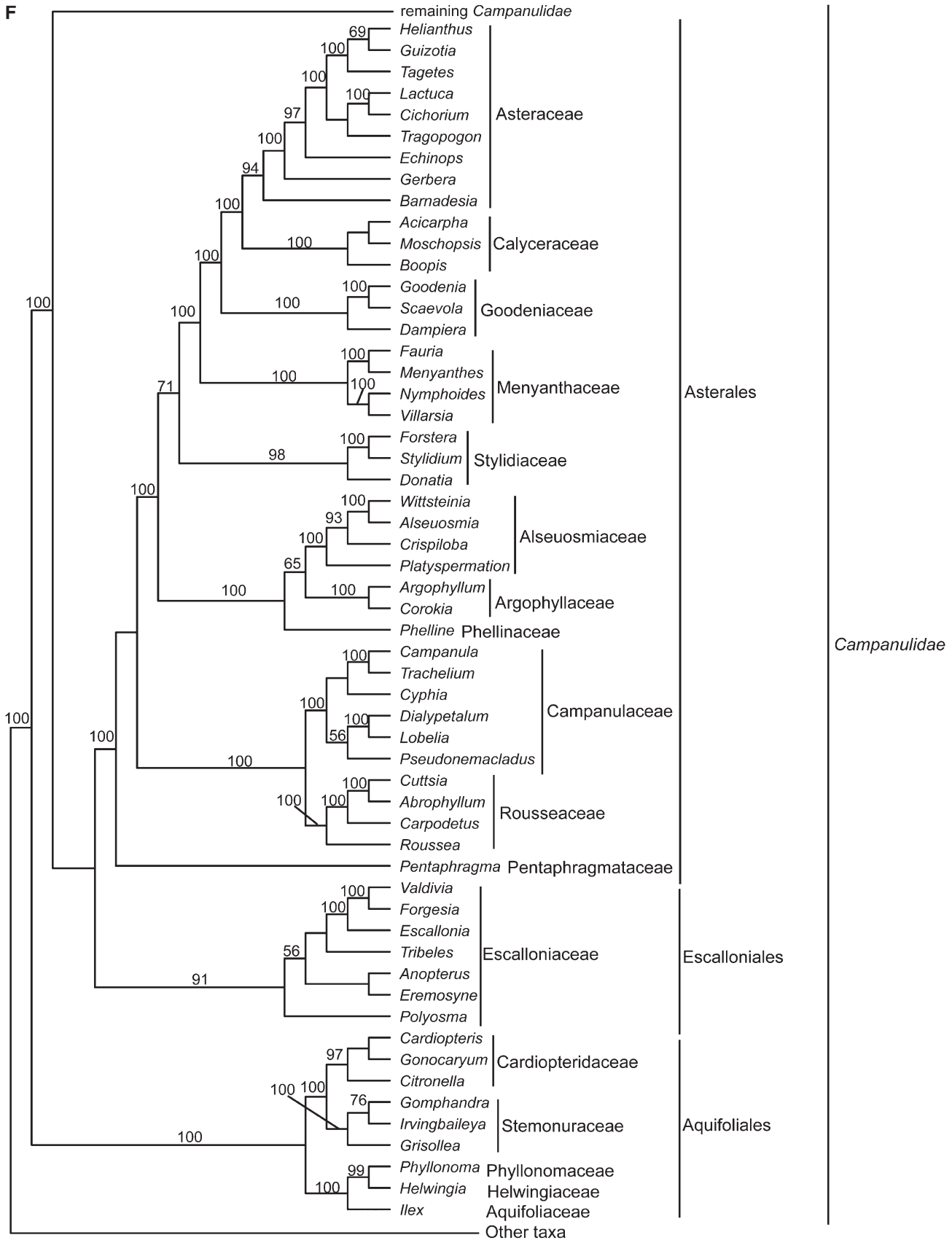

Fig. 2. Continued.

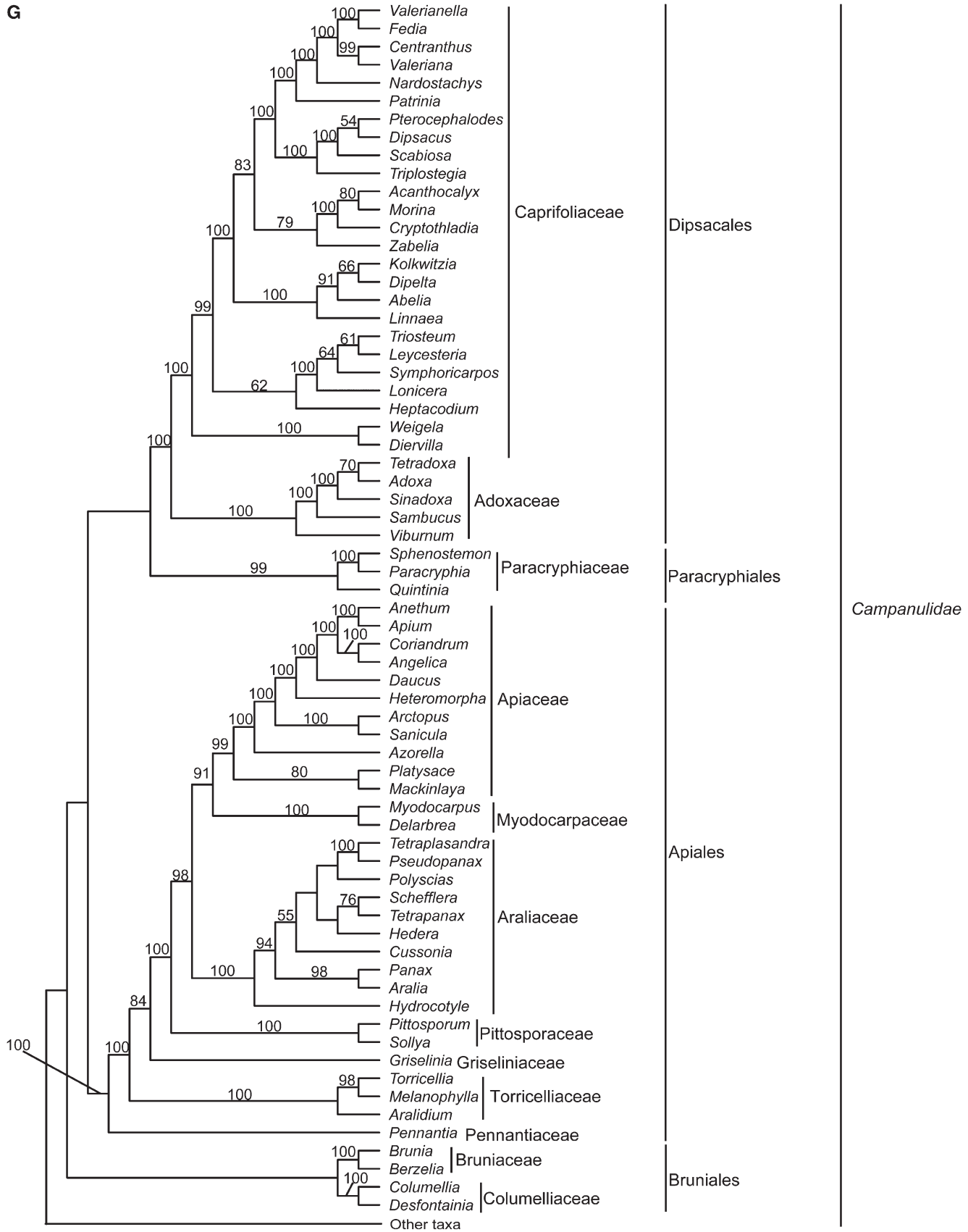

Fig. 2. Continued.

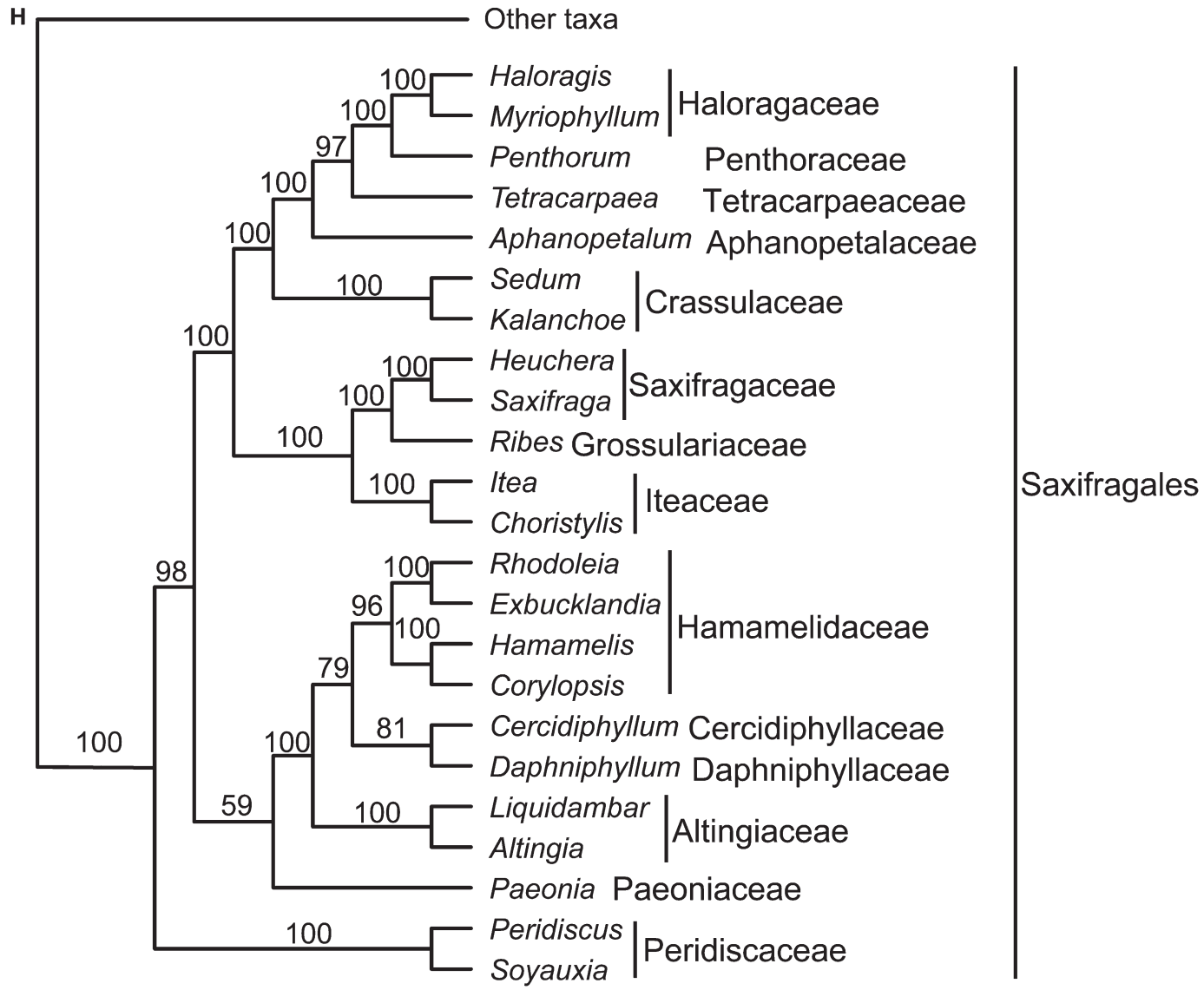

Fig. 2. Continued.

Cantino et al., 2007). Complete plastid genome data (Moore et al., 2010) as well as inverted repeat (IR) sequence data (Moore et al., in press) also recover *Superrosidae* with strong support, but relationships differ among the three constituent clades (Vitaceae, all remaining *Rosidae*, Saxifragales). Complete plastid genome data suggest that *Rosidae* (excluding Vitaceae) are sister to Saxifragales + Vitaceae, whereas IR sequence data recover the topology found here (Vitaceae sister to all remaining *Rosidae* + Saxifragales). Focused studies of *Superrosidae* suggest Saxifragales are sister to Vitaceae + all remaining *Rosidae*, in agreement with the results here (see H. Wang et al., 2009). The divergent results produced by analyses of complete plastid genomes may be due to more limited taxon sampling in that analysis compared to the IR analysis (244 taxa; Moore et al., in press) and the present study.

We provide a formal definition of *Superrosidae* in Appendix 1.

Saxifragales—In the 17-gene tree, Peridiscaceae are sister to all remaining Saxifragales, which are well supported (BS = 98%) and form two major clades: (1) a weakly supported clade (BS = 59%) of Paeoniaceae + the "woody clade" (BS = 100%), comprising Cercidiphyllaceae, Daphniphyllaceae, Altingiaceae, and Hamamelidaceae, and (2) core Saxifragales (BS = 100%), comprising two clades, each with BS = 100%: Crassulaceae + Haloragaceae s.l. (i.e., *Aphanopetalum*, Tetracarpaeaceae, Penthoraceae, and Haloragaceae) and the Saxifragaceae alliance. Within the latter, Saxifragaceae + Grossulariaceae are sister to Iteaceae + Pterostemonaceae. These results agree with the focused analyses of Jian et al. (2008).

Morphological synapomorphies for Saxifragales include basifixed anthers, usually with the filament attached at a basal pit (but reversals to the dorsifixed condition occur in *Ribes*, *Pterostemon*, and Iteaceae), and latrorse dehiscence. All but Peridiscaceae also may be united by follicle fruits, a homoplasious character that is correlated with the ovaries being at least distally distinct. Additionally, the presence of violoid to theoid teeth is possibly synapomorphic, but salicoid teeth occur in *Aphanopetalum*, and nonglandular teeth occur in *Hamamelis*,

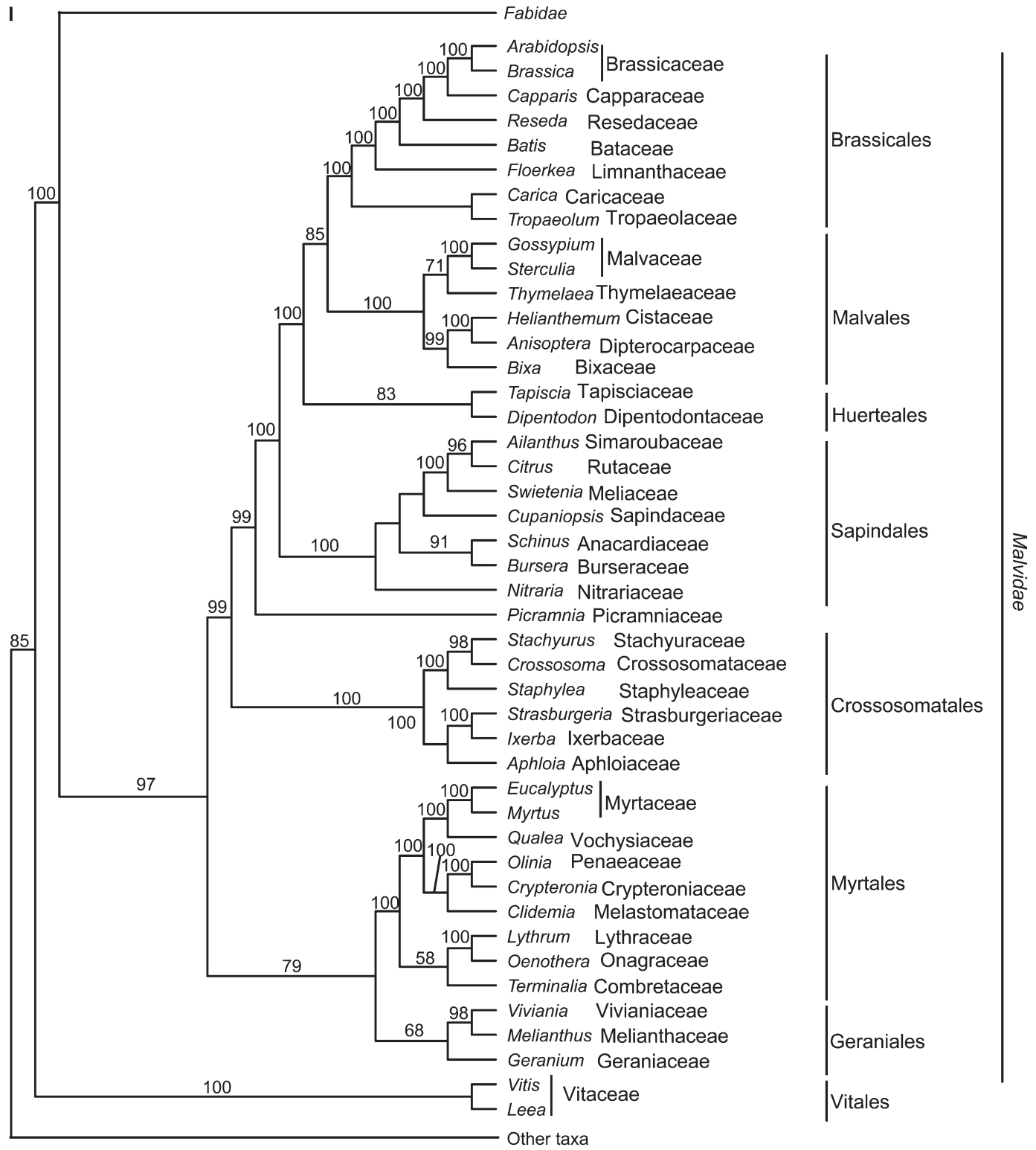

Fig. 2. Continued.

Itea, and *Haloragis*, while the condition cannot be assessed in taxa with entire leaves.

Rosidae—Despite extensive progress in elucidating relationships within the angiosperms, *Rosidae* has long stood out as the largest poorly resolved major clade; deep relationships within *Rosidae* have been particularly problematic. Our analyses agree with recent analyses of this clade based on parsimony and ML analyses of 12-gene (>20 000 bp) and IR (>24 000 bp) data sets for over 100 rosid species (H. Wang et al., 2009). *Rosidae* forms

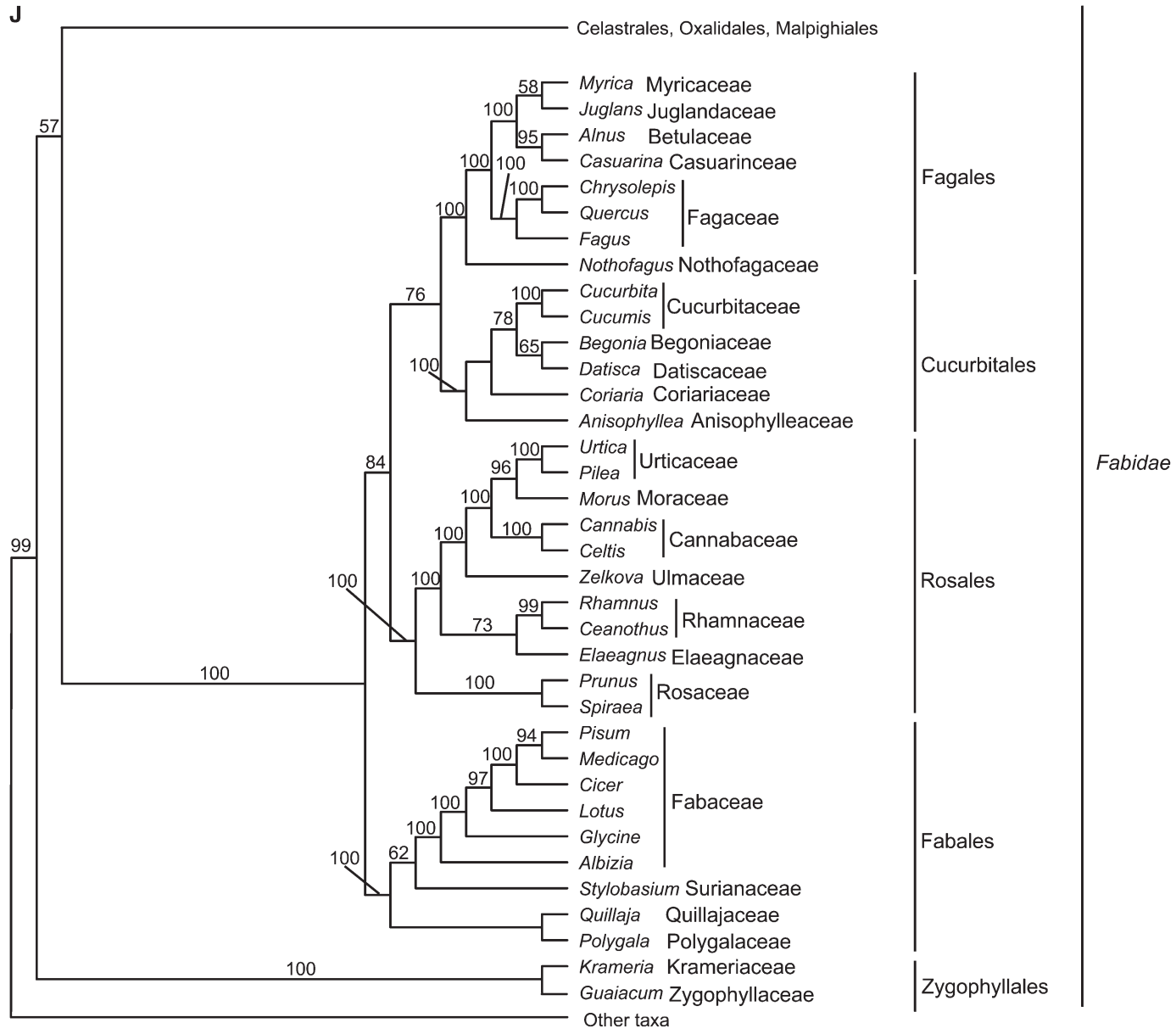

Fig. 2. Continued.

a well-supported clade (BS = 85%), in which Vitaceae are sister to the remainder (BS = 100%), which in turn consists of two large clades: (1) *Fabidae*, which includes the nitrogen-fixing clade, Celastrales, Huaceae, Zygophyllales, Malpighiales, and Oxalidales; and (2) *Malvidae*, which includes Tapisciaceae, Brassicales, Malvales, Sapindales, Geraniales, Myrtales, Crossosomatales, and Picramniaceae.

Malvidae—Within a well-supported *Malvidae* (BS = 97%), a clade (BS = 79%) of Myrtales (BS = 100%) + Geraniales (BS = 68%) is sister to the remaining *Malvidae* (BS = 99%). Within the latter, a well-supported (BS = 100%) Crossosomatales are sister to the rest (BS = 100%), followed by Picramniaceae as sister to a clade (BS = 100%) of Sapindales (BS = 100%) + Huerteales + (Malvales + Brassicales; each with BS = 100%). These results are identical to recent topologies resulting from analyses focused on *Rosidae* (H. Wang et al., 2009).

Over half of the families of Brassicales (Akaniaceae, Emblingiaceae, Gyrostemonaceae, Koeberliniaceae, Moringaceae, Pentadiplandraceae, Salvadoraceae, Setchellanthaceae, and Tovariaceae) were not included in our study. Of those Brassicales sampled, relationships agree closely with other analyses (see Rodman et al., 1996, 1998; Hall et al., 2004). Tropaeolaceae and Caricaceae are sisters to a well-supported Limnanthaceae followed by Bataceae, as subsequent sisters to Resedaceae + (Capparidaceae + Brassicaceae). Malvales are also poorly sampled, with only 5 of 10 families included, but the topology matches other studies (e.g., Bayer et al., 1999; Alverson et al., 1999; H. Wang et al., 2009).

Within Sapindales, Nitrariaceae are sister to two clades (although the sister-group relationship is not very well supported):

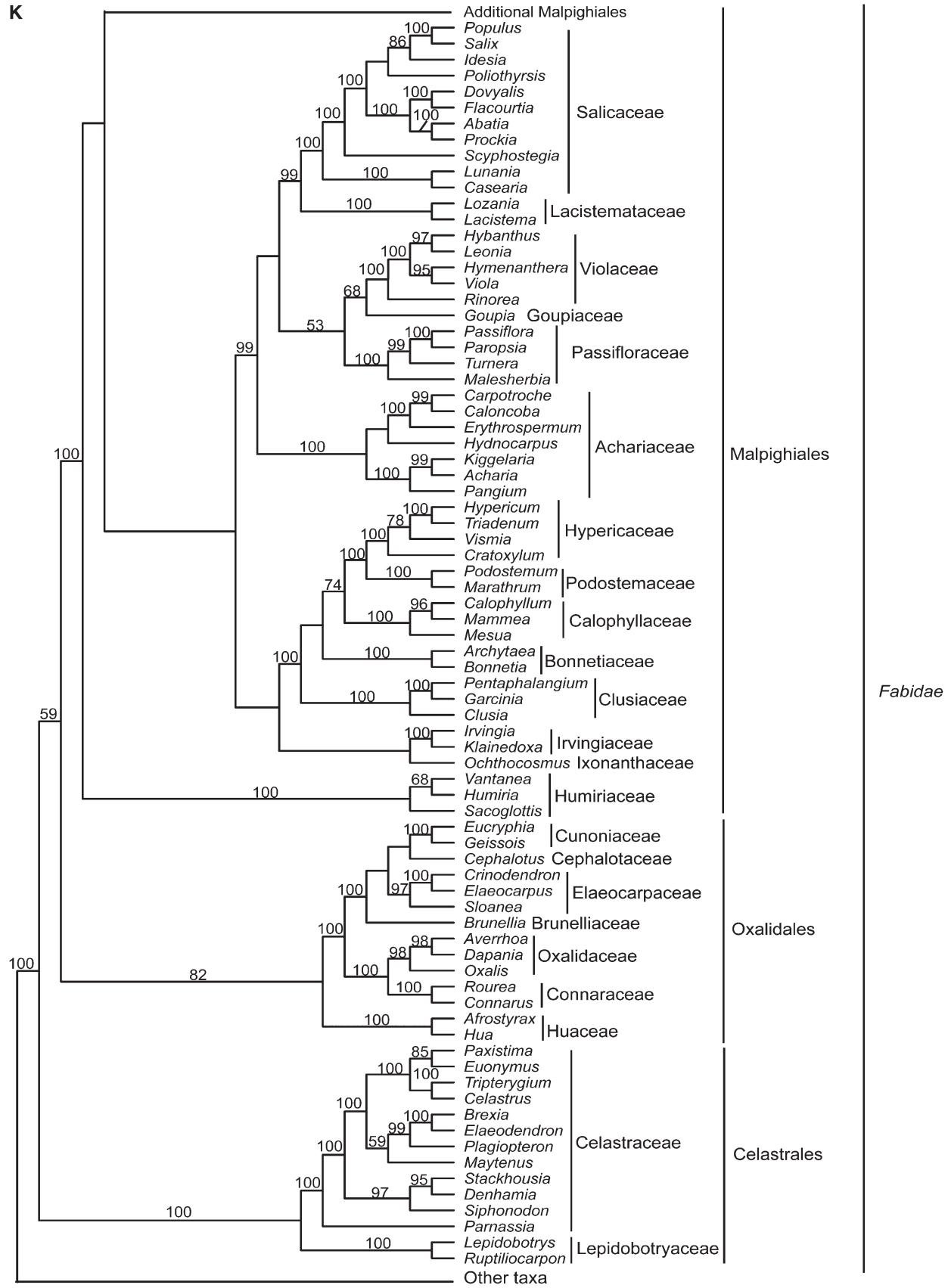

Fig. 2. Continued.

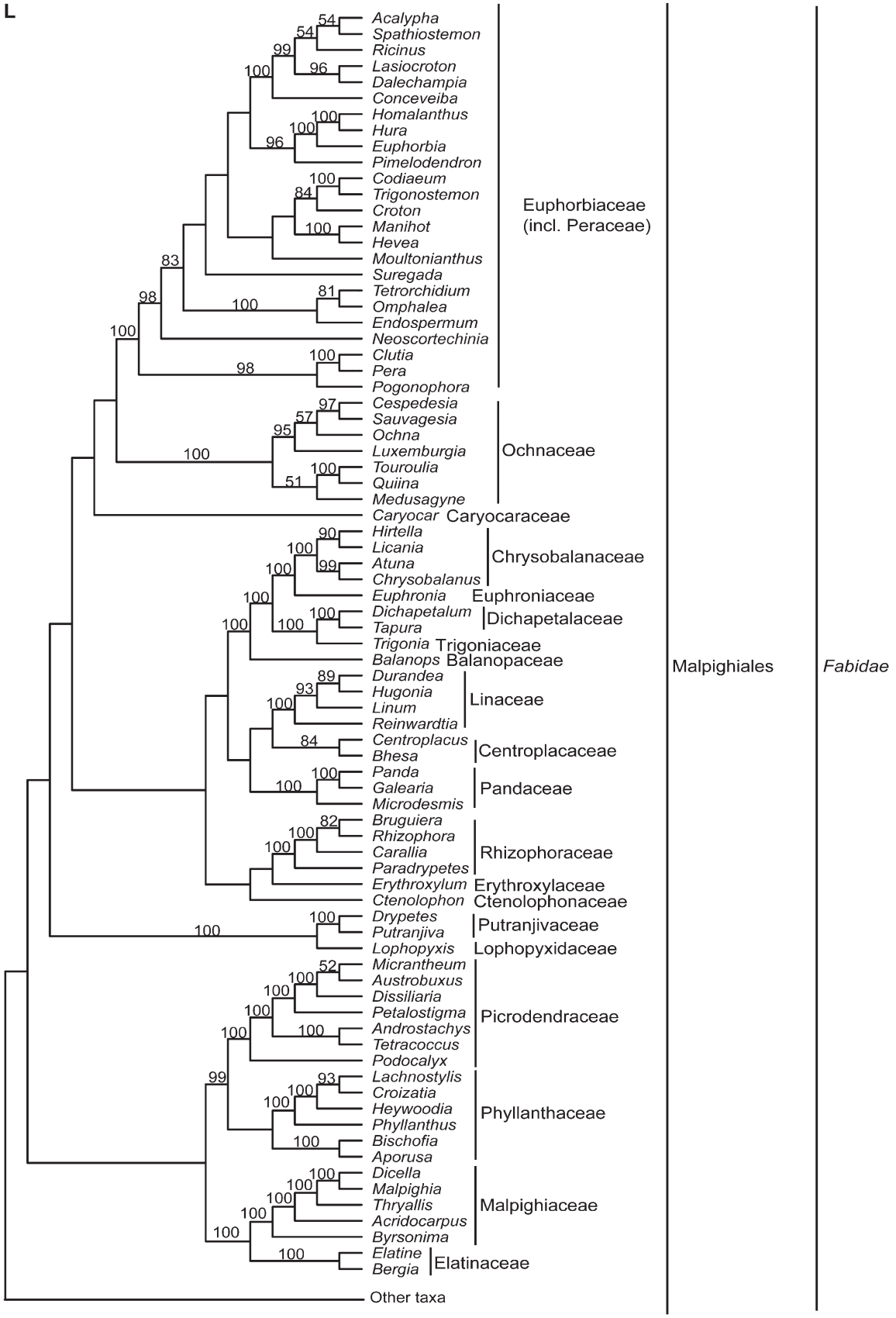

Fig. 2. Continued.

(1) Burseraceae + Anacardiaceae, and (2) a clade (also without BS >50%) of Sapindaceae sister to Meliaceae + (Simaroubaceae + Rutaceae). These relationships largely agree with other analyses (e.g., Gadek et al., 1996; Muellner et al., 2007); different relationships were reported by H. Wang et al. (2009), albeit with poor sampling and low support.

In Crossosomatales, there are two well-supported (BS = 100%) clades: Staphyleaceae + [Stachyuraceae + Crossosomataceae] and Aphloiaceae + Strasburgeriaceae (including Ixerbaceae). In Myrtales, two major clades were recovered: one comprising Onagraceae + Lythraceae and the other comprising the large families Myrtaceae and Melastomataceae and associated smaller families. These results are consistent with all previous studies of the order (e.g., Conti et al., 1997; Sytsma et al., 2004). The placement of Combretaceae, however, is not strongly supported, a result that is in agreement with previous studies.

Fabidae—Fabidae is well supported (BS = 99%), with Zygophyllales sister (albeit with low support, BS = 57%) to two major clades (each with BS = 100%): the nitrogen-fixing clade (Cucurbitales, Fabales, Fagales, and Rosales) and a clade of Celastrales, Oxalidales (including Huaceae; see APG III), and Malpighiales (the COM group; Endress and Matthews, 2006). These results agree with H. Wang et al. (2009).

In the COM clade, Celastrales (BS = 100%) are sister to a clade (BS = 59%) of Malpighiales (BS =100%) + Oxalidales (BS = 100%). This topology agrees with recent analyses (e.g., Soltis et al., 2007; H. Wang et al., 2009; Wurdack and Davis, 2009). Alternative topologies in which Malpighiales and Oxalidales did not form a clade have been strongly rejected (Wurdack and Davis, 2009).

Nitrogen-fixing clade—Within the nitrogen-fixing clade (BS = 100%), Fabales, Fagales, Rosales, and Cucurbitales all have BS = 100%. Fabales appear as sister with strong support (BS = 84%) to the remainder of the clade; Rosales are then sister to a clade (BS = 78%) of Fagales + Cucurbitales. These results are in agreement with recent studies (e.g., Soltis et al., 2005, 2007; H. Wang et al., 2009).

In Rosales (Barbeyaceae and Dirachmaceae were not sampled here), Rosaceae are sister to the remainder of the clade (BS = 100%), which forms two subclades: (1) a well-supported clade (BS = 100%) of Ulmaceae sister to (Cannabaceae + (Urticaceae + Moraceae)) and (2) a clade (BS = 73%) of Elaeagnaceae + Rhamnaceae. These results agree with other recent studies (e.g., Sytsma et al., 2002; H. Wang et al., 2009).

Relationships in Fagales also agree with broad (Soltis et al., 2000, 2007) as well as focused analyses (Manos and Steele, 1997; Li et al., 2002). Nothofagaceae and Fagaceae are successive sisters to the remaining Fagales, which form two clades: (1) Casuarinaceae + Betulaceae and (2) Myricaceae + Juglandaceae (Ticodendraceae and Rhoipteleaceae were not sampled here).

Our results for Cucurbitales are similar to recent analyses (with Corynocarpaceae and Tetrameleaceae not sampled here) (e.g., Zhang et al., 2006; Soltis et al., 2007). Anisophyllaceae and Coriariaceae are subsequent sisters to a clade (BS = 78%) of (Datiscacaee + Begoniaceae) + Cucurbitaceae.

In Fabales, Polygalaceae + Quillajaceae (BS support < 50%) are sister to a clade (BS = 62%) of Surianaceae + Fabaceae. The topologies for Fabales in recent analyses have varied greatly, but most of these studies have involved only a few DNA regions (see Wojciechowski et al., 2004, Banks et al., 2008; Bruneau et al., 2008; Bello et al., 2009).

COM clade—In Celastrales, Lepidobotryaceae are sister to a broadly defined Celastraceae, following other recent focused analyses (Simmons et al., 2001a, b; Zhang and Simmons, 2006). Within Oxalidales, Huaceae are sister to the remainder of the clade, which then forms two well-supported (BS = 100%) subclades: (1) Connaraceae + Oxalidaceae, and (2) Brunelliaceae, Cephalotaceae, Cunoniaceae, and Elaeocarpaceae. Within the second clade, Brunelliaceae are sister to Elaeocarpaceae + (Cephalotaceae + Cunoniaceae). These relationships differ only slightly from previous analyses. In Zhang and Simmons (2006), results were well supported and identical to those here, but Cephalotaceae were not sampled. In Wurdack and Davis (2009) a weakly supported (BS = 59%) Cephalotaceae + Cunoniaceae were unresolved with Brunelliaceae and Elaeocarpaceae. In H. Wang et al. (2009), Cunoniaceae were sister to (Cephalotaceae + Elaeocarpaceae), albeit without BS support > 50%; the current study also has increased representation of families within Oxalidales over the H. Wang et al. (2009) analysis of *Rosidae*.

Malpighiales are well represented in the current study, and relationships largely agree with recent analyses (e.g., Davis and Wurdack, 2004; Davis et al., 2001, 2005; Wurdack and Davis, 2009; Ruhfel et al., 2011). The strong agreement between our results and those of Wurdack and Davis (2009) is not surprising given that both studies share considerable data. We overview the major features of this clade below.

We recovered a well-supported (BS = 99%) parietal-placentation clade (sensu Wurdack and Davis, 2009). Within this clade, Lacistemataceae are the well-supported sister to Salicaceae, and Goupiaceae are weakly supported (BS = 68%) as sister to Violaceae. The placement of Goupiaceae was unresolved in Wurdack and Davis (2009). The clusioid clade sensu Wurdack and Davis (2009) was well supported (BS = 100%), and within this clade Calophyllaceae were sister (BS = 74%) to Podostemaceae + Hypericaceae. These relationships are also in agreement with Ruhfel et al. (2011). Balanopaceae are sister to a clade of Trigoniaceae + Dichapetalaceae, all of which is sister to Chrysobalanaceae + Euphroniaceae. Other well-supported (BS = 100%) sister-group relationships recovered here include: Erythroxylaceae + Rhizophoraceae, Elatinaceae + Malpighiaceae, Lophopyxidaceae + Putranjivaceae, and Phyllanthaceae + Picrodendraceae. The recently recognized (Wurdack and Davis, 2009) Euphorbiaceae segregate Peraceae are recovered here as sister to Euphorbiaceae s.s. APG III (2009) has deferred recognition of Peraceae pending additional support from other non-mtDNA gene regions for placement of Rafflesiaceae within Euphorbiaceae + Peraceae. Similar to recent findings (i.e., Davis et al., 2007; Wurdack and Davis, 2009), our preliminary 17-gene analyses that sampled Rafflesiaceae (later excluded with most other parasites in our final analyses, but available upon request) placed that holoparasitic lineage with this clade. Despite some advances in our understanding of Malpighiales here, relationships of many constituent clades still remain unclear (see Wurdack and Davis, 2009).

Superasteridae—Superasteridae (BS = 87%) consists of Berberidopsidales, Santalales, *Caryophyllales*, *Asteridae*, and possibly Dilleniaceae (see below). The 17-gene analysis placed Santalales as sister to a weakly supported (BS < 50%) clade of (Berberidopsidales + *Caryophyllales*) + *Asteridae*. The sister grouping of Berberidopsidales + *Caryophyllales* received BS = 75%.

Complete plastid genome data (Moore et al., 2010) and IR sequence data (Moore et al., in press) similarly recovered a

well-supported *Superasteridae* clade; hence, we formally name this clade here (see Appendix 2). However, these three analyses differ in the relationships suggested among members of *Superasteridae.* Complete plastid genome sequence data place Santalales as sister to the remainder (BS = 100%); *Berberidopsis* is then sister to a clade (BS = 88%) of *Caryophyllales* + *Asteridae*. IR data suggest Santalales as sister to the well-supported (BS = 98%) remaining taxa, in which *Caryophyllales* are sister to a clade (BS = 68%) of Berberidopsidales + *Asteridae*. Hence, all three studies agree in the placement of Santalales as sister to the rest, but the precise placement of the enigmatic Berberidopsidales within *Superasteridae* remains unclear.

Dilleniales—Dilleniaceae were recovered as sister to the remaining *Superasteridae* (BS = 97%). This position is much better supported than other recent placements of this enigmatic family (e.g., Moore et al., 2010, in press), but despite this high support, we remain cautious about its true placement because of differing placements in recent studies. Using complete plastid genome sequence data, Moore et al. (2010) placed Dilleniaceae as sister to *Superrosidae* (BS = 64%), but topology tests did not reject alternative positions of Dilleniaceae as sister to *Asteridae* or to all remaining *Pentapetalae*. Analysis of IR sequences for 244 taxa placed Dilleniaceae sister to *Superrosidae* + *Superasteridae* (BS = 79%) (Moore et al., in press). Because of these varied placements, Dilleniaceae are only tentatively placed here in *Superasteridae* based on their position in Fig. 1. Additional research is required to place Dilleniaceae definitively.

Santalales—Within Santalales (BS = 100%), the two sampled genera of "Olacaceae", a family now regarded as nonmonophyletic (see Malécot and Nickrent, 2008; Nickrent et al., 2010), are subsequent sisters to the well-supported (BS = 100%) remainder of the clade. Santalaceae and Opiliaceae are successive sisters to a clade (BS = 79%) of Schoepfiaceae + Loranthaceae. Representation of Santalales is sparse, notably missing Erythropalaceae, which may be sister to the rest of Santalales, and the holoparasitic Balanophoraceae, which have recently been recognized as part of this clade. Relationships among major clades of Santalales have not been resolved with strong BS support in recent analyses, and many segregate families have been newly recognized (e.g., Malécot, 2002; Nickrent et al., 2010).

Caryophyllales—The topology resolved here is similar to that of Brockington et al. (2009). Two major subclades were recovered: core Caryophyllales (Caryophyllineae) and noncore Caryophyllales (Polygonineae) (clade names of Judd et al., 2008). Polygonineae in turn comprise two clades: (1) Plumbaginaceae + Polygonaceae and Tamaricaceae + Frankeniacae, and (2) Droseraceae + Nepenthaceae sister to Drosophyllaceae + (Ancistrocladaceae + Dioncophyllaceae).

Within Caryophyllineae, Rhabdodendraceae, Simmondsiaceae, and Asteropeiaceae + Physenaceae are subsequent sisters to the rest of the clade. Within the latter, Caryophyllaceae + (Amaranthaceae + Achatocarpaceae) are sister to the remainder. In the remaining clade, Stegnospermataceae followed by Limeaceae are sister to (1) Barbeuiaceae, Aizoaceae, and Gisekiaceae as subsequent sisters to a clade of Phytolaccaceae followed by Sarcobataceae as sisters to Petiveriaceae (as represented by *Rivinia*) + Nyctaginaceae; and (2) Molluginaceae followed by Montiaceae sister to two clades: Halophytaceae sister to Basellaceae + Didiereaceae, and Talinaceae sister to Portulacaceae + Cactaceae. Portulacaceae and Phytolaccaceae, as broadly circumscribed, are not monophyletic in our analyses, in agreement with previous studies (e.g., Applequist and Wallace, 2001; Cuénoud et al., 2002; Brockington et al., 2009; Nyffeler and Eggli, 2010), leading to the inclusion of *Calyptrotheca* in Didiereaceae and recognition of Montiaceae and Talinaceae (as in APG III, 2009; Nyffeler and Eggli, 2010) and of Petiveriaceae (see Judd et al., 2008; Nyffeler and Eggli, 2010).

Asteridae—Within *Asteridae*, Cornales are sister to the remaining taxa (BS = 97%) within which Ericales are sister to a clade (BS = 100%) consisting of two subclades: *Lamiidae* (BS = 100%) + *Campanulidae* (BS = 100%). Recognition of *Asteridae* comprising these four groups was first based on a molecular phylogenetic analysis (Olmstead et al., 1992), and the relationships among these groups are congruent with most prior molecular studies (Olmstead et al., 2000; Albach et al., 2001; Bremer et al., 2002), although with stronger support.

Cornales—Within Cornales (BS = 100%), a weakly supported clade (BS = 76%) of *Cornus* + *Nyssa* (Cornaceae s.l.) is sister to a weakly supported (BS = 54%) clade of Curtisiaceae + (Hydrangeaceae + Loasaceae). Only the sister pair of Hydrangeaceae + Loasaceae is well supported (BS = 100%), in agreement with other analyses (Fan and Xiang, 2003). Two families, Grubbiaceae and Hydrostachyaceae, were not included here; the precise placement of the latter family within Cornales has been particularly problematic (Olmstead et al., 2000; Albach et al., 2001; Fan and Xiang, 2003).

Ericales—The monophyly of Ericales is well supported (BS = 100%), but as in other recent studies, resolution within the clade is more problematic. Our study retrieves a well-supported clade (BS = 100%) of Tetrameristaceae + (Marcgraviaceae + Balsaminaceae) (see Schönenberger et al., 2010) as sister to the remaining Ericales, which are also well supported as monophyletic (BS = 95%), but within which relationships are even more uncertain, mirroring earlier analyses by Schönenberger et al. (2005) based on 11 genes. Lecythidaceae are strongly supported (BS = 95%) as sister to the remainder, although in the 11-gene study with much greater taxon sampling (Schönenberger et al., 2005), Lecythidaceae were placed in a polytomy at this node. Three other well-supported clades were recovered: (1) Polemoniaceae + Fouquieriaceae (BS = 92%), (2) a clade (BS = 98%) of Actinidiaceae + Sarraceniaceae sister to a clade (BS = 100%) of Clethraceae + (Cyrillaceae + Ericaceae) (BS = 100%), and (3) a Primulaceae s.l. clade (BS = 100%) in which *Maesa* is sister to *Clavija* + Primulaceae s.s. (BS = 100%). As in the 11-gene analysis (Schönenberger et al., 2005), the relationships of Diapensiaceae, Ebenaceae, Sapotaceae, Ternstroemiaceae, and Theaceae are not well supported here. A recent 23-gene analysis with approximately 90 taxa of Ericales resolves many of these latter placements (K. Sytsma et al., unpublished manuscript).

Lamiidae—Lamiidae are well supported (BS = 100%). Within this clade, Icacinaceae + Garryales, and Vahliaceae are subsequent sisters to the remaining lamiids, which comprise four major clades: Boraginaceae, Lamiales, Solanales, and Gentianales, among which there is no well-supported resolution. The position of Vahliaceae as immediate sister to the rest of *Lamiidae* is consistent with the parsimony analysis of N. F. Refulio-Rodriguez and R. G. Olmstead (unpublished manuscript); however, in their ML and Bayesian analyses, *Vahlia* is sister to Solanales.

A more thorough analysis of the clade is provided by N. F. Refulio-Rodriguez and R. G. Olmstead (unpublished manuscript) with a total of 129 species of *Lamiidae*, including all recognized families (APG III, 2009). Their results differ mostly in places where resolution is weak in the results presented here, such as the branching order among the early-diverging groups, Oncothecaceae, Icacinaceae, and Garryales, and among the four major clades of *Lamiidae*, Lamiales, Solanales, Gentianales, and Boraginaceae.

Lamiales—Lamiales (BS = 100%) comprise Plocospermataceae, Oleaceae, and Tetrachondraceae as well-supported subsequent sisters (each with BS = 100%) to the rest of Lamiales. *Peltanthera* (unassigned to family) plus (Gesneriaceae + Calceolariaceae) form a clade (BS = 97%) sister to the remaining Lamiales, which form a strongly supported core (BS = 100%). Within this core, few relationships have BS values >50% in our analyses, indicating the difficulties in resolving relationships in this clade. Relationships with support >70% include Verbenaceae + Thomandersiaceae (BS = 75%), and (((Phrymaceae + Orobanchaceae (BS = 95%)) + Paulowniaceae (BS = 99%)) + Lamiaceae (BS = 74%)). These results are consistent with those obtained by Refulio-Rodriguez and Olmstead (unpublished manuscript), except that Orobanchaceae and *Paulownia* are well-supported sister groups, followed by Phrymaceae in their results. However, most of these families are represented here by only a single species.

Solanales—The topology for Solanales agrees with other recent analyses, although most have not had complete family coverage of the order (e.g., Soltis et al., 2000). Complete family coverage was obtained in Bremer et al. (2002), although the current study provides much stronger internal support because of the large number of genes included and added sampling within Solanaceae and Convolvulaceae. We found Montiniaceae are sister to Spenocleaceae + Hydroleaceae; this well-supported clade (BS = 98%) is in turn sister to Convolvulaceae + Solanaceae (BS = 100%).

Gentianales—Rubiaceae are sister to the remainder of the clade, in agreement with other studies (e.g., Backlund et al., 2000; Frasier, 2009). Relationships among the remaining families have been unclear in most previous studies (e.g., Olmstead et al., 2000; Soltis et al., 2000; Bremer et al., 2002), but here Loganiaceae are weakly supported as sister to Gelsemiaceae + (Apocynaceae + Gentianaceae), which is congruent with a study of four plastid DNA regions (Frasier, 2009). Although the current study does reveal areas of strong support, sampling in the families of this order is low.

Campanulidae—A *Campanulidae* clade was recovered with strong support (BS = 100%), and relationships within largely agree with recent analyses, especially Tank and Donoghue (2010; also see Lundberg, 2001; Winkworth et al., 2008). This is not surprising given the very large overlap in data between the current study and that of Tank and Donoghue (2010), the major difference being the addition of mtDNA and nuclear rDNA data in the current study. Tank and Donoghue (2010) used their phylogenetic results to discuss floral evolution in *Campanulidae*, and analyses of biogeographic patterns, especially in relation to past continental connections, are underway.

Each of the four major lineages of *Campanulidae*—Apiales, Aquifoliales, Asterales, and Dipsacales—was recovered here with strong support (BS = 100%). The several smaller *Campanulidae* clades, including Escalloniales (= Escalloniaceae sensu Tank and Donoghue, 2010) and Paracryphiales (= Paracryphiaceae sensu Tank and Donoghue, 2010), also receive strong support in the current study (BS = 91% and 99%, respectively). Although recovered in the best ML tree (Figs. 1 and 2), bootstrap support for Bruniales (the Columelliaceae-Bruniaceae clade sensu Tank and Donoghue, 2010) is <50% in the total evidence analysis reported here.

Relationships among the primary *Campanulidae* clades also agree with previous analyses. We recovered the sister-group relationship of Aquifoliales and *Apiidae* (i.e., the remaining *Campanulidae* taxa; Cantino et al., 2007). Likewise, relationships among the six major lineages within *Apiidae*—Apiales, Asterales, Bruniales, Dipsacales, Escalloniaceae, and Paracryphiaceae—were identical to earlier results (e.g., Zhang et al., 2003; Winkworth et al., 2008; Tank and Donoghue, 2010). To facilitate the study of evolutionary patterns within *Apiidae*, Tank and Donoghue (2010) named several clades that they recovered with strong statistical support, including *Dipsapiidae* (the Apiales-Dipsacales-Paracryphiaceae clade) and *Dipsidae* (the Dipsacales-Paracryphiaceae clade). Both of these clades were recovered in the total evidence analysis, albeit with <50% bootstrap support.

Aquifoliales—Aquifoliales were resolved as the sister group to *Apiidae*. Within Aquifoliales, there are two major lineages: Cardiopteridaceae + Stemonuraceae and a clade of Aquifoliaceae + (Phyllonomaceae + Helwingiaceae). As did Tank and Donoghue (2010), we found strong support for *Phyllonoma-Helwingia* as sister to *Ilex* (Fig. 2). Although *Phyllonoma* and *Helwingia* share several conspicuous morphological apomorphies, including epiphyllous inflorescences, most earlier studies (based on far less data) placed *Helwingia* directly with *Ilex* to the exclusion of *Phyllonoma* (e.g., Morgan and Soltis, 1993; Soltis and Soltis, 1997; Olmstead et al., 2000).

Escalloniaceae—Escalloniaceae (= Escalloniales sensu APG III, 2009) were recovered with strong support (Figs. 1 and 2; BS = 91%); however, their relationship to the other lineages of *Campanulidae* remain uncertain. In addition, relationships among the genera are not well supported aside from a core Escalloniaceae clade composed of *Escallonia*, *Forgesia*, and *Valdivia* (Fig. 2). In our mtDNA analyses, which included *Escallonia*, *Eremosyne*, and *Polyosma*, none of the genes recovered a monophyletic Escalloniaceae. However, there is little support for relationships among any of the unresolved lineages in the mtDNA analyses (Appendix S6), with one exception: when mtDNA data are included for *Polyosma*, it is placed with strong support within Paracryphiaceae (= Paracryphiales sensu APG III, 2009). Without the mtDNA data, *Polyosma* is placed with BS = 91% in Escalloniaceae, as expected based on previous analyses (e.g., Lundberg, 2001; Winkworth et al., 2008; Tank and Donoghue, 2010) (compare Fig. 2 and Appendix S3; see below, *Comparison of mtDNA, rDNA, and plastid topologies*).

Asterales—Each of the major lineages of Asterales was recovered with strong support, and, for the most part, relationships among them are also well supported (Fig. 2). One exception is at the base of the Asterales, where the successive sister-group relationship of Pentaphragmataceae and the Roussaceae + Campanulaceae clade to the remainder of Asterales received <50% bootstrap support (Fig. 2). In earlier studies, these relationships were reversed, with the Roussaceae-Campanulaceae

clade sister to the rest of Asterales (Lundberg and Bremer, 2003; Winkworth et al., 2008; Tank and Donoghue, 2010). In our analyses of the plastid partition alone, we recovered the same relationships as the earlier studies with strong support (Appendix S5). By contrast, the mtDNA partition provides strong support for the alternative resolution of these relationships (Appendix S6). When the mtDNA gene trees are examined separately, *atp1* and *matR* recovered the mtDNA topology (BS < 50%), while *rps3* resolved the plastid topology (BS values < 50%) (Appendices S17–S20).

Bruniales—As noted above, Bruniales were recovered with strong support by Tank and Donoghue (2010), but were only weakly supported (<50% bootstrap) in our combined analysis. This case is discussed further below (see *Comparison of mtDNA, rDNA, and plastid topologies*). This clade is of biogeographic interest in that Columelliaceae are restricted to South America, while Bruniaceae are a primarily South African lineage; this appears to be an ancient vicariance event in *Campanulidae*, and the detailed historical biogeography of this clade is the subject of current research.

Dipsapiidae—Tank and Donoghue (2010) named this clade to emphasize the close relationship between Apiales and Dipsacales (along with Paracryphiaceae, = Paracryphiales sensu APG III)—a result that appeared in several earlier studies (e.g., Bremer et al., 2002; Winkworth et al., 2008) and is strongly supported in their plastid analyses (Tank and Donoghue, 2010). Surprisingly, this clade is not recovered with >50% bootstrap support in our total evidence analyses, although it does appear in the ML BS tree (Figs. 1 and 2). Within *Dipsapiidae*, Paracryphiaceae are resolved as sister to Dipsacales (= *Dipsidae* sensu Tank and Donoghue, 2010), albeit with <50% bootstrap support. As with *Dipsapiidae*, *Dipsidae* were strongly supported in the analyses of both Winkworth et al. (2008) and Tank and Donoghue (2010).

Relationships within Dipsacales largely agree with earlier studies (e.g., Bell et al., 2001; Donoghue et al., 2001, 2003; Bell, 2004; Winkworth et al., 2008; Tank and Donoghue, 2010), as do relationships within Apiales (e.g., Kårehed, 2001, 2003; Chandler and Plunkett, 2004; Winkworth et al., 2008; Tank and Donoghue, 2010), with only minor differences in areas with little to no bootstrap support (e.g., within Araliaceae).

Possible taxon sampling issues—Several relationships do not agree with more detailed phylogenetic studies focused on particular clades (see also Tank and Donoghue, 2010). For example, within Dipsacaceae (Dipsacales), *Pterocephalodes* is sister to *Dipsacus*, and *Scabiosa* is their subsequent sister. The sister group of *Pterocephalodes* and *Dipsacus* was very strongly supported in the analyses of Tank and Donoghue (2010), but is less strongly supported here (BS = 54%). This result is incongruent with the much more densely sampled analyses of Carlson et al. (2009), where *Pterocephalodes* emerged as sister to all other Dipsacaceae. In contrast, while Tank and Donoghue (2010) found a well-supported sister-group relationship between *Echinops* and *Gerbera* within Asteraceae, here we recovered the successive sister-group relationship of *Gerbera* and *Echinops* to the remainder of Asteraceae that appears in more detailed studies of Asteraceae (e.g., Panero and Funk, 2008). In general, these examples highlight that although statistical support may be high for particular relationships, taxon and gene sampling can both have major effects on the outcome, especially when the diversity of the group in question has been grossly undersampled.

Comparison of mtDNA, rDNA, and plastid topologies—In the present study, there is some conflict among trees derived from mtDNA, rDNA, and plastid data. However, resolution and support of the rDNA topology is so low that there are no instances of mutual incongruence with BS greater than 75% involving it and either of the other partitions.

One major case of incongruence involves relationships among major lineages of *Rosidae*. The analysis of mitochondrial genes shows that the weakly supported COM clade is sister to the strongly supported core members of *Malvidae* (Sapindales, Malvales, Brassicales, Huerteales, Picramniaceae) with 92% BS support. This result parallels results obtained in a recent mtDNA analysis of angiosperms (Qiu et al., 2010). An earlier analysis of *matR* also recovered the same relationship, albeit with lower bootstrap support (Zhu et al., 2007). In all previous large-scale analyses of angiosperms or eudicots, the COM clade is sister to the nitrogen-fixing clade (Chase et al., 1993; Savolainen et al., 2000a, b; Soltis et al., 2000; Hilu et al., 2003; Burleigh et al., 2009), occasionally with high support (e.g., BS = 89% in Burleigh et al., 2009). These analyses were based either entirely on plastid genes alone or in combination with nuclear rDNA (18S and 26S rDNA). Thus far, only one large-scale analysis of angiosperms has been performed on nuclear gene data, 18S rDNA (Soltis et al., 1997). In that study, clades corresponding to *Fabidae* and *Malvidae* were not recovered, but this has been a general problem with 18S and 26S rDNA data—limited resolution and low support due to low phylogenetic signal (see Soltis et al., 1997).

One line of evidence that lends some support to the possible sister relationship between the COM clade and *Malvidae* suggested by mtDNA data comes from a broad survey of floral structural characters (Endress and Matthews, 2006). Twenty-two COM families and 18 families of *Malvidae* share a type of ovule with a thicker inner integument than the outer one, a situation that is otherwise very rare in eudicots (with only one other occurrence, in Trochodendrales). That survey did not find any morphological synapomorphy supporting the monophyly of *Fabidae*, but this is not surprising in that morphological synapomorphies remain elusive for many major clades of angiosperms.

Likewise, within *Campanulidae* there are several instances where mtDNA and plastid analyses are in conflict. The most obvious example of conflict among data sets involved the placement of *Polyosma*. With mtDNA sequence data included, *Polyosma* (Escalloniaceae) is placed sister to *Quintinia* in the Paracryphiales (Appendix S2), but when mtDNA data for *Polyosma* are removed, it is placed in Escalloniaceae. Both alternative placements received high BS support, and nuclear rDNA data do not provide sufficient resolution of the placement of this taxon. These results suggest either a biological phenomenon (e.g., HGT of mtDNA from some member of Paracryphiaceae to the *Polyosma* lineage) or human error (e.g., incorrect labeling of material). All of the plastid sequences for *Polyosma* came from two accessions—one used in Bremer et al. (2002) and one a silica-gel collection obtained from MO (Tank and Donoghue, 2010). The individual plastid gene trees all agree with the placement of *Polyosma* in Escalloniaceae. The *atp1* sequence for *Polyosma* was obtained from our MO accession, while the other three mtDNA genes came from a separate accession (Kew M285-TL265). Because sequences from the MO accession used by Tank and Donoghue (2010) in their plastid analyses agree with previous studies based on a different accession (e.g., Bremer et al., 2002), it is possible that the M285-TL265 accession was misidentified, or tissue or tubes were mislabeled. Significantly, sequencing of the mitochondrial genes surveyed

here in a second accession of *Polyosma*, as well as a second sample of *Quintinia* (the sister of *Polyosma* when mtDNA are included), yielded nearly identical sequences and the same topologies. Thus, available data favor HGT as the more likely cause of incongruence involving *Polyosma*.

Further evidence of conflict between the mtDNA and plastid partitions in *Campanulidae* comes from the lack of bootstrap support in the combined analyses for several clades that were well supported in our analyses of the plastid data alone. In analyses of the plastid partition (Appendix S5), we recovered a Bruniales clade with 79% BS support, in agreement with Tank and Donoghue (2010). However, as noted above, in the combined analysis this relationship received <50% bootstrap support. These results suggest that there is enough conflict between the mtDNA partition and the plastid partition to collapse this relationship. We observed several other such cases, including the *Dipsapiidae* and *Dipsidae* clades and relationships at the base of Asterales, as noted above.

Impact of taxon density—One concern with the complete plastid genome analysis of angiosperms by Moore et al. (2010) is that only 86 taxa were employed, raising questions regarding the impact of taxon density on the topology. To test this, we trimmed the 17-gene data set from 640 taxa to the 86 taxa in the analysis by Moore et al. (2010).

The topologies of the 17-gene 640-taxon and reduced 86-taxon (Appendix S21) data sets are very similar, but with a few differences. Topological similarities include strong support for: Chloranthaceae + *Magnoliidae* (BS = 94%); *Monocotyledoneae* + (Ceratophyllaceae + *Eudicotyledoneae*) (BS = 97%); and *Trochodendron* followed by *Buxus* as subsequent sisters to *Gunneridae* (BS = 100%; 95%). However, there are noteworthy differences between the trees. First, in the reduced tree, *Amborella* is sister to Nymphaeales (BS = 97%), rather than *Amborella* followed by Nymphaeales as sister to all other extant angiosperms. Second, Nelumbonaceae and *Meliosma* are subsequent sisters to other angiosperms in the reduced tree, rather than a clade, as in the larger tree. Third, *Superrosidae* are not recovered in the reduced tree; instead, a modified *Superasteridae* (BS = 50%) includes Vitaceae (a member of *Superrosidae* in the complete analysis). Fourth, Dilleniaceae are sister to *Caryophyllales* (BS < 50%), rather than sister to all *Superasteridae*. This placement of Dilleniaceae as sister to *Caryophyllales* was also seen in the three-gene tree (e.g., Soltis et al., 2000), but not in the larger tree here. All of these placements not only differ from the 17-gene, 640-taxon analysis, but also from the complete plastid genome topology and from an analysis of the IR for 244 taxa (Moore et al., in press). These results further confirm that reduced taxon density can have a major impact on topology, even when numerous genes are employed. Furthermore, some of these differences could also reflect conflict between plastid and mtDNA data.

LITERATURE CITED

ALBACH, D. C., D. E. SOLTIS, M. W. CHASE, AND D. E. SOLTIS. 2001. Phylogenetic placement of the enigmatic angiosperm *Hydrostachys*. *Taxon* 50: 781–805.

ALVERSON, W. S., B. A. WHITLOCK, R. NYFFELER, C. BAYER, AND D. A. BAUM. 1999. Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. *American Journal of Botany* 86: 1474–1486.

APG III [Angiosperm Phylogeny Group III]. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

APPLEQUIST, W. L., AND R. S. WALLACE. 2001. Phylogeny of the portulacaceous cohort based on *ndhF* sequence data. *Systematic Botany* 26: 406–419.

BACKLUND, M., B. OXELMAN, AND B. BREMER. 2000. Phylogenetic relationships within the Gentianales based on *ndhF* and *rbcL* sequences, with particular reference to the Loganiaceae. *American Journal of Botany* 87: 1029–1043.

BANKS, H., B. B. KLITGAARD, F. CLAXTON, F. FOREST, AND P. R. CRANE. 2008. Pollen morphology of the family Polygalaceae (Fabales). *Botanical Journal of the Linnean Society* 156: 253–289.

BARKMAN, T. J., J. R. MCNEAL, S. H. LIM, G. COAT, H. B. CROOM, N. D. YOUNG, AND C. W. DEPAMPHILIS. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 7: 248.

BAYER, C., M. F. FAY, A. DE BRUIJN, V. SAVOLAINEN, C. M. MORTON, K. KUBITZKI, 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. *Botanical Journal of the Linnean Society* 129: 267–303.

BEAMAN, R. S., AND N. CELLINESE. 2010. TOLKIN: The Tree of Life Knowledge and Information Network. Website http://www.tolkin.org/.

BELL, C. D. 2004. Preliminary phylogeny of Valerianaceae (Dipsacales) inferred from nuclear and chloroplast DNA sequence data. *Molecular Phylogenetics and Evolution* 31: 340–350.

BELL, C. D., E. J. EDWARDS, S. T. KIM, AND M. J. DONOGHUE. 2001. Dipsacales phylogeny based on chloroplast DNA sequences. *Harvard Papers in Botany* 6: 481–499.

BELL, C. D., D. E. SOLTIS, AND P. SOLTIS. 2010. The age and diversification of angiosperms re-revisited. *American Journal of Botany* 97: 1296–1303.

BELLO, M. A., A. BRUNEAU, F. FOREST, AND J. A. HAWKINS. 2009. Elusive relationships within order Fabales: Phylogenetic analyses using *matK* and *rbcL* sequence data. *Systematic Botany* 34: 102–114.

BERGTHORSSON, U., K. L. ADAMS, B. THOMASON, AND J. D. PALMER. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197–201.

BOWE, L. M., AND C. W. DEPAMPHILIS. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. *Molecular Biology and Evolution* 13: 1159–1166.

BREMER, B., K. BREMER, N. HEIDARI, P. ERIXON, R. G. OLMSTEAD, A. A. ANDERBERG, M. KÄLLERSJÖ, AND E. BARKHORDARIAN. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution* 24: 274–301.

BROCKINGTON, S. F., R. ALEXANDRE, J. RAMDIAL, M. J. MOORE, S. CRAWLEY, A. DHINGRA, K. HILU, ET AL. 2010. Phylogeny of the Caryophyllales sensu lato: Revisiting hypotheses on pollination biology and perianth differentiation in the Core Caryophyllales. *International Journal of Plant Sciences* 171: 185–198.

BRUNEAU, A., M. MERCURE, G. P. LEWIS, AND P. S. HERENDEEN. 2008. Phylogenetic patterns and diversification in the caesalpinioid legumes. *Botany* 86: 697–718.

BURLEIGH, J. G., K. W. HILU, AND D. E. SOLTIS. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evolutionary Biology* 9: 61.

CANTINO, P. D., J. A. DOYLE, S. W. GRAHAM, W. S. JUDD, R. G. OLMSTEAD, D. E. SOLTIS, P. S. SOLTIS, AND M. J. DONOGHUE. 2007. Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* 56: 822–846.

CARLSON, S. E., V. MAYER, AND M. J. DONOGHUE. 2009. Phylogenetic relationships, taxonomy, and morphological evolution in Dipsacaceae (Dipsacales) inferred by DNA sequence data. *Taxon* 58: 1075–1091.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17: 540–552.

CHANDLER, G. T., AND G. M. PLUNKETT. 2004. Evolution in Apiales: nuclear and chloroplast markers together in (almost) perfect harmony. *Botanical Journal of the Linnean Society* 144: 123–147.

CHASE, M., D. E. SOLTIS, D. MORGAN, D. OLMSTED, M. DUVALL, R. PRICE, H. HILLS, ET AL. 1993. Phylogenetic relationships among seed plants based on *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80: 528–580.

CHASE, M. W., M. F. FAY, D. DEVEY, O. MAURIN, N. RØNSTED, J. DAVIES, Y. PILLON, ET AL. 2006. Multigene analyses of monocot relationships: a summary. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: Comparative biology and evolution. Excluding Poales. Rancho Santa Ana Botanical Garden, Claremont, California, USA. *Aliso* 22: 63–75.

CONTI, E., A. LITT, P. G. WILSON, S. A. GRAHAM, B. G. BRIGGS, L. A. S. JOHNSON, AND K. J. SYTSMA. 1997. Interfamilial relationships in Myrtales: Molecular phylogeny and patterns of morphological evolution. *Systematic Botany* 22: 629–647.

CUENOUD, P., V. SAVOLAINEN, L. W. CHATROU, M. POWELL, R. J. GRAYER, AND M. W. CHASE. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL, atpB*, and *matK* DNA sequences. *American Journal of Botany* 89: 132–144.

DAVIES, T. J., T. G. BARRACLOUGH, M. W. CHASE, P. S. SOLTIS, D. E. SOLTIS, AND V. SAVOLAINEN. 2004. A supertree of the angiosperms. *Proceedings of the National Academy of Sciences, USA* 101: 1904–1909.

DAVIS, C. C., W. R. ANDERSON, AND M. J. DONOGHUE. 2001. Phylogeny of Malpighiaceae: Evidence from chloroplast *ndhF* and *trnL-F* nucleotide sequences. *American Journal of Botany* 88: 1830–1846.

DAVIS, C. C., M. LATVIS, D. L. NICKRENT, K. J. WURDACK, AND D. A. BAUM. 2007. Floral gigantism in Rafflesiaceae. *Science* 315: 1812.

DAVIS, C. C., C. O. WEBB, K. J. WURDACK, C. A. JARAMILLO, AND M. J. DONOGHUE. 2005. Explosive radiation of Malpighiales supports a Mid-Cretaceous origin of modern tropical rain forests. *American Naturalist* 165: E36–E65.

DAVIS, C. C., AND K. J. WURDACK. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* 305: 676–678.

DONOGHUE, M. J., C. D. BELL, AND R. C. WINKWORTH. 2003. The evolution of reproductive characters in Dipsacales. *International Journal of Plant Sciences* 164 (5 supplement): S453–S464.

DONOGHUE, M. J., T. ERIKSSON, P. A. REEVES, AND R. G. OLMSTEAD. 2001. Phylogeny and phylogenetic taxonomy of Dipsacales, with special reference to *Sinadoxa* and *Tetradoxa* (Adoxaceae). *Harvard Papers in Botany* 6: 459–479.

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

DOYLE, J. A., H. EKLUND, AND P. S. HERENDEEN. 2003. Floral evolution in Chloranthaceae: Implications of a morphological phylogenetic analysis. *International Journal of Plant Sciences* 164: S365–S382.

EKLUND, H., J. A. DOYLE, AND P. S. HERENDEEN. 2004. Morphological phylogenetic analysis of living and fossil Chloranthaceae. *International Journal of Plant Sciences* 165: 107–151.

ENDRESS, P. K., AND M. L. MATTHEWS. 2006. First steps towards a floral structural characterization of the major rosid subclades. *Plant Systematics and Evolution* 260: 223–251.

EVANS, R. C., L. A. ALICE, C. S. CAMPBELL, T. A. DICKINSON, AND E. A. KELLOGG. 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution* 17: 388–400.

FAN, C., AND Q.-Y. XIANG. 2003. Phylogenetic analyses of Cornales based on 26S rRNA and combined 26S rRNA-*matK*-*rbcL* sequence data. *American Journal of Botany* 90: 1357–1372.

FELSENSTEIN, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22: 240–249.

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

FELSENSTEIN, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, Washington, USA. Website http://www.phylip.com/.

FRASIER, C. L. 2009. Evolution and systematics of the angiosperm order *Gentianales* with an in-depth focus on *Loganiaceae* and its species-rich and toxic genus *Strychnos*. PhD dissertation, Rutgers University, New Brunswick, New Jersey, USA.

FUNK, V. A., A. SUSANNE, T. F. STUESSY, AND R. J. BAYER, eds. 2009. Systematics, evolution, and biogeography of Compositae. International Association for Plant Taxonomy, Institute of Botany, University of Vienna. Vienna, Austria.

GADEK, P. A., E. S. FERNANDO, C. J. QUINN, S. B. HOOT, T. TERRAZAS, M. C. SHEAHAN, AND M. W. CHASE. 1996. Sapindales: Molecular delimitation and infraordinal groups. *American Journal of Botany* 83: 802–811.

GIVNISH, T. J., J. H. LEEBENS-MACK, M. AMES SEVILLANO, J. R. MCNEAL, P. R. STEELE, J. I. DAVIS, C. ANÉ, ET AL. 2010. Assembling the tree of the monocotyledons: Plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

GOLUBCHIK, T., M. J. WISE, S. EASTEAL, AND L. S. JERMIIN. 2007. Mind the gaps: Evidence of bias in estimates of multiple sequence alignments. *Molecular Biology and Evolution* 24: 2433–2442.

GRAHAM, S. W., J. M. ZGURSKI, M. A. MCPHERSON, D. M. CHERNIAWSKY, J. M. SAARELA, E. F. C. HORNE, S. Y. SMITH, ET AL. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: Comparative biology and evolution. Excluding Poales. Rancho Santa Ana Botanical Garden, Claremont, California, USA. *Aliso* 22: 3–21.

HALL, J. C., H. H. ILTIS, AND K. J. SYTSMA. 2004. Molecular phlyogenetics of core Brassicales, placement of orphan genera *Emblingia, Forchammeria, Tirania*, and character evolution. *Systematic Botany* 29: 654–669.

HILU, K. W., T. BORSCH, K. MULLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, ET AL. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

JANSEN, R. K., Z. CAI, L. A. RAUBESON, H. DANIELL, C. W. DEPAMPHILIS, J. LEEBENS-MACK, K. F. MÜLLER, ET AL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

JIAN, S., P. S. SOLTIS, M. GITZENDANNER, M. MOORE, R. LI, T. HENDRY, Y. QIU, ET AL. 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57: 38–57.

JIAO, Y., N. WICKETT, S. AYYAMPALAYAM, A. CHANDERBALI, L. LANDHERR, P. E. RALPH, P. S. SOLTIS, ET AL. In press. Phylogenomic analysis reveals ancient genome duplications in seed plant and angiosperm history. *Nature*.

JUDD, W. S., C. S. CAMPBELL, E. A. KELLOGG, P. F. STEVENS, AND M. J. DONOGHUE. 2008. Plant systematics: A phylogenetic approach, 3rd ed. Sinauer, Sunderland, Massachusetts, USA.

KÅREHED, J. 2001. Multiple origin of the tropical forest tree family Icacinaceae. *American Journal of Botany* 88: 2259–2274.

KÅREHED, J. 2003. The family Pennantiaceae and its relationships to Apiales. *Botanical Journal of the Linnean Society* 141: 1–24.

KATOH, K., AND H. TOH. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298.

KIM, S., D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, AND Y. SUH. 2004. Phylogenetic relationships among early-diverging eudicots: Evidence from *atpB, rbcL*, 18S rDNA, and 26S rDNA sequences. *Molecular Phylogenetics and Evolution* 31: 16–30.

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* 38: 7–25.

LECOINTRE, G., AND P. DELEPORTE. 2005. Total evidence requires exclusion of phylogenetically misleading data. *Zoologica Scripta* 34: 101–117.

LEEBENS-MACK, J., L. A. RAUBESON, L. CUI, J. V. KUEHL, M. H. FOURCADE, T. W. CHUMLEY, J. L. BOORE, ET AL. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963.

LI, R.-Q., Z.-D. CHEN, Y.-P. HONG, AND A.-M. LU. 2002. Phylogenetic relationships of the "higher" hamamelids based on chloroplast *trnL-F* sequences. *Acta Botanica Sinica* 44: 1462–1468.

LUNDBERG, J. 2001. Phylogenetic studies in the euasterids II with particular reference to Asterales and Escalloniaceae. Ph.D. dissertation, Uppsala University, Uppsala, Sweden.

LUNDBERG, J., AND K. BREMER. 2003. A phylogenetic study of the order Asterales using one morphological and three molecular data sets. *International Journal of Plant Sciences* 164: 553–578.

MALÉCOT, V. 2002. Histoire, classification et phylogénie des Olacaceae Brown (Santalales). Thèse de Doctorat de l'Université Paris 6, France.

MALÉCOT, V., AND D. L. NICKRENT. 2008. Molecular phylogenetic relationships of Olacaceae and related Santalales. *Systematic Botany* 33: 97–106.

MANOS, P. S., AND K. P. STEELE. 1997. Phylogenetic analysis of "higher" Hamamelididae based on plastid sequence data. *American Journal of Botany* 84: 1407–1419.

MOORE, M. J., C. D. BELL, P. S. SOLTIS, AND D. E. SOLTIS. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

MOORE, M. J., N. HASSAN, M. A. GITZENDANNER, R. A. BRUENN, M. CROLEY, A. VANDEVENTER, J. W. HORN, ET AL. In press. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences*.

MOORE, M. J., P. S. SOLTIS, C. D. BELL, J. G. BURLEIGH, AND D. E. SOLTIS. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

MORGAN, D. R., AND D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 80: 631–660.

MUELLNER, A. N., D. D. VASSILIADES, AND S. S. RENNER. 2007. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Plant Systematics and Evolution* 266: 233–252.

NICKRENT, D. L., A. BLARER, Y.-L. QIU, R. VIDAL-RUSSELL, AND F. E. ANDERSON. 2004. Phylogenetic inference in Rafflesiales: The influence of rate heterogeneity and horizontal gene transfer. *BMC Evolutionary Biology* 4: 40.

NICKRENT, D. L., V. MALÉCOT, R. VIDAL-RUSSELL, AND J. P. DER. 2010. A revised classification of Santalales. *Taxon* 59: 538–558.

NIXON, K. C. 1999. The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics* 15: 407–414.

NYFFELER, R., AND U. EGGLI. 2010. Disintegrating Portulacaceae: A new familial classification of the suborder Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* 59: 227–240.

OLMSTEAD, R. G., K. J. KIM, R. K. JANSEN, AND S. J. WAGSTAFF. 2000. The phylogeny of the Asteridae sensu lato based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 16: 96–112.

OLMSTEAD, R. G., H. MICHAELS, K. M. SCOTT, AND J. D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbc*L. *Annals of the Missouri Botanical Garden* 79: 249–265.

PANERO, J. L., AND V. A. FUNK. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution* 47: 757–782.

POTTER, D., T. ERIKSSON, R. C. EVANS, S.-H. OH, J. SMEDMARK, D. MORGAN, M. KERR, ET AL. 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* 266: 5–43.

POTTER, D., F. GAO, P. E. BORTIRI, S.-H. OH, AND S. BAGGETT. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Systematics and Evolution* 231: 77–89.

QIU, Y.-L., O. DOMBROVSKA, J. H. LEE, L. B. LI, B. A. WHITLOCK, F. BERNASCONI-QUADRONI, J. S. REST, ET AL. 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 166: 815–842.

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. J. ZANIS, E. A. ZIMMER, ET AL. 1999. The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genes. *Nature* 402: 404–407.

QIU, Y.-L., L. LI, B. WANG, J.-Y. XUE, T. A. HENDRY, R.-Q. LI, J. W. BROWN, ET AL. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391–425.

RENNER, S. S. 1999. Circumscription and phylogeny of Laurales: Evidence from molecular and morphological data. *American Journal of Botany* 86: 1301–1315.

RODMAN, J. E., K. G. KAROL, R. A. PRICE, AND K. J. SYTSMA. 1996. Molecules, morphology, and Dahlgren's expanded order Capparales. *Systematic Botany* 21: 289–307.

RODMAN, J. E., P. S. SOLTIS, D. E. SOLTIS, K. J. SYTSMA, AND K. G. KAROL. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997–1007.

RUDALL, P. J., D. D. SOKOLOFF, M. V. REMIZOWA, J. G. CONRAN, J. I. DAVIS, T. D. MACFARLANE, AND D. W. STEVENSON. 2007. Morphology of Hydatellaceae, an anomalous aquatic family recently recognized as an early-divergent angiosperm lineage. *American Journal of Botany* 94: 1073–1092.

RUHFEL, B. R., V. BITTRICH, C. P. BOVE, M. H. G. GUSTAFSSON, C. T. PHILBRICK, R. RUTIHAUSER, Z. XI, AND C. C. DAVIS. 2011. Phylogeny of the clusioid clade (Malpighiales): evidence from the plastid and mitochondrial genomes. *American Journal of Botany* 98: 306–325.

SAARELA, J. M., H. S. RAI, J. A. DOYLE, P. K. ENDRESS, S. MATHEWS, A. D. MARCHANT, B. G. BRIGGS, AND S. W. GRAHAM. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446: 312–315.

SANDERSON, M. J., M. M. MCMAHON, AND M. STEEL. 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. *BMC Evolutionary Biology* 10: 155.

SAUQUET, H., J. A. DOYLE, T. SCHARASCHKIN, T. BORSCH, K. W. HILU, L. W. CHATROU, AND A. L. THOMAS. 2003. Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: implications for character evolution. *Botanical Journal of the Linnean Society* 142: 125–186.

SAVOLAINEN, V., M. W. CHASE, S. B. HOOT, C. M. MORTON, D. E. SOLTIS, C. BAYER, M. F. FAY, ET AL. 2000a. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* 49: 306–362.

SAVOLAINEN, V., M. F. FAY, D. C. ALBACH, A. BACKLUND, M. VAN DER BANK, K. M. CAMERON, S. A. JOHNSON, ET AL. 2000b. Phylogeny of the eudicots: A nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bulletin* 55: 257–309.

SCHÖNENBERGER, J., A. A. ANDERBERG, AND K. J. SYTSMA. 2005. Molecular phylogenetics and patterns of floral evolution in the Ericales. *International Journal of Plant Sciences* 166: 265–288.

SCHÖNENBERGER, J., M. VON BALTHAZAR, AND K. J. SYTSMA. 2010. Diversity and evolution of floral structure among early diverging lineages in the Ericales. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 365: 437–448.

SIKES, D. S., AND P. O. LEWIS. 2001. PAUPRat: A tool to implement parsimony and likelihood ratchet searches using PAUP* [online computer program]. Website http://www.ucalgary.ca/~dsikes/software2.htm.

SIMMONS, M. P., C. C. CLEVINGER, V. SAVOLAINEN, R. H. ARCHER, S. MATHEWS, AND J. J. DOYLE. 2001a. Phylogeny of the Celastraceae inferred from phytochrome B gene sequence and morphology. *American Journal of Botany* 88: 313–325.

SIMMONS, M. P., V. SAVOLAINEN, C. C. CLEVINGER, R. H. ARCHER, AND J. I. DAVIS. 2001b. Phylogeny of the Celastraceae inferred from 26S nuclear ribosomal DNA, phytochrome B, *rbcL, atpB*, and morphology. *Molecular Phylogenetics and Evolution* 19: 353–366.

SMITH, S. A., J. M. BEAULIEU, AND M. J. DONOGHUE. 2009. Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 37.

SMITH, S. A., AND C. W. DUNN. 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24: 715–716.

SOLTIS, D. E., M. A. GITZENDANNER, AND P. S. SOLTIS. 2007. A 567-taxon data set for angiosperms: The challenges posed by Bayesian analyses of large data sets. *International Journal of Plant Sciences* 168: 137–157.

SOLTIS, D. E., AND P. S. SOLTIS. 1997. Phylogenetic relationships in Saxifragaceae sensu lato: A comparison of topologies based on 18S rDNA and rbcL sequences. *American Journal of Botany* 84: 504–522.

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.

SOLTIS, D. E., P. S. SOLTIS, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, ET AL. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SOLTIS, D. E., P. S. SOLTIS, T. G. COLLIER, AND M. L. EDGERTON. 1991. Chloroplast DNA variation within and among genera of the *Heuchera* group: cpDNA evidence for intergeneric chloroplast transfer and the paraphyly of *Mitella* and *Heuchera. American Journal of Botany* 78: 1091–1112.

SOLTIS, D. E., P. S. SOLTIS, P. K. ENDRESS, AND M. W. CHASE. 2005. Phylogeny and evolution of the angiosperms. Sinauer, Sunderland, Massachusetts, USA.

SOLTIS, D. E., P. S. SOLTIS, D. L. NICKRENT, L. A. JOHNSON, W. J. HAHN, S. B. HOOT, J. A. SWEERE, ET AL. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

STAMATAKIS, A., P. HOOVER, AND J. ROUGEMONT. 2008. A fast bootstrapping algorithm for the RAxML web-servers. *Systematic Biology* 57: 758–771.

STEEL, M., AND M. J. SANDERSON. 2010. Characterizing phylogenetically decisive taxon coverage. *Applied Mathematics Letters* 23: 82–86.

STEVENS, P. F. 2001 onward. Angiosperm Phylogeny Website, version 9, June 2008 [and more or less continuously updated since]. Website http://www.mobot.org/MOBOT/research/APweb/.

SWOFFORD, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4b10. Sinauer, Sunderland, Massachusetts, USA.

SYTSMA, K. J. 1994. DNA extraction from recalcitrant plants: Long, pure, and simple? *In* R. P. Adams, J. S. Miller, E. M. Golenberg, and J. E. Adams [eds.], Conservation of plant genes II: Intellectual property rights and DNA utilization, 69–81. Missouri Botanical Garden, St. Louis, Missouri, USA.

SYTSMA, K. J., A. LITT, M. L. ZJHRA, J. C. PIRES, M. NEPOKROEFF, E. CONTI, J. WALKER, AND P. G. WILSON. 2004. Clades, clocks, and continents: historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the southern hemisphere. *International Journal of Plant Sciences* 165 (4 supplement): S85–S105.

SYTSMA, K. J., J. MORAWETZ, J. C. PIRES, M. NEPOKROEFF, E. CONTI, M. ZJHRA, J. C. HALL, AND M. W. CHASE. 2002. Urticalean rosids: Circumscription, rosid ancestry, and phylogenetics based on *rbcL, trnL-F*, and *ndhF* sequences. *American Journal of Botany* 89: 1531–1546.

TALAVERA, G., AND J. CASTRESANA. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56: 564–577.

TANK, D. C., AND M. J. DONOGHUE. 2010. Phylogeny and phylogenetic nomenclature of the Campanulidae based on an expanded sample of genes and taxa. *Systematic Botany* 35: 425–441.

VON BALTHAZAR, M., AND J. SCHÖNENBERGER. 2009. Floral structure and organization in Platanaceae. *International Journal of Plant Sciences* 170: 210–225.

WANG, H., M. J. MOORE, P. S. SOLTIS, C. D. BELL, S. BROCKINGTON, R. ALEXANDRE, C. C. DAVIS, ET AL. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA* 106: 3853–3858.

WANG, W., A.-M. LU, Y. REN, M. E. ENDRESS, AND Z.-D. CHEN. 2009. Phylogeny and classification of Ranunculales: Evidence from four molecular loci and morphological data. *Perspectives in Plant Ecology, Evolution and Systematics* 11: 81–110.

WINKWORTH, R. C., J. LUNDBERG, AND M. J. DONOGHUE. 2008. Toward a resolution of campanulid phylogeny, with a special reference to the placement of Dipsacales. *Taxon* 57: 53–65.

WOJCIECHOWSKI, M. F., M. LAVIN, AND M. J. SANDERSON. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.

WURDACK, K. J., AND C. C. DAVIS. 2009. Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. *American Journal of Botany* 96: 1551–1570.

ZANIS, M. J., D. E. SOLTIS, P. S. SOLTIS, S. MATHEWS, AND M. J. DONOGHUE. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences, USA* 99: 6848–6853.

ZHANG, L.-B., AND S. S. RENNER. 2003. The deepest splits in Chloranthaceae as resolved by chloroplast sequences. *International Journal of Plant Sciences* 164 (5 Supplement): S383–S392.

ZHANG, L.-B., AND M. P. SIMMONS. 2006. Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Systematic Botany* 31: 122–137.

ZHANG, L.-B., M. P. SIMMONS, A. KOCYAN, AND S. S. RENNER. 2006. Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system evolution. *Molecular Phylogenetics and Evolution* 39: 305–322.

ZHANG, W.-H., Z.-D. CHEN, J. H. LI, H.-B. CHEN, AND Y.-C. TANG. 2003. Phylogeny of the Dipsacales s.l. based on chloroplast *trnL-F* and *ndhF* sequences. *Molecular Phylogenetics and Evolution* 26: 176–189.

ZHU, X.-Y., M. W. CHASE, Y.-L. QIU, H.-Z. KONG, D. L. DILCHER, J.-H. LI, AND Z. D. CHEN. 2007. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology* 7: 217.

APPENDIX 1. Phylocode description of *Superrosidae*

Superrosidae W. S. Judd, D. E. Soltis, and P. S. Soltis

Definition

The most inclusive crown clade that contains *Rosa cinnamomea* L. 1753 (*Rosidae*) but not *Aster amellus* L. 1753 (*Asteridae*). This is a branch-modified node-based definition. Abbreviated definition: >∇ *Rosa cinnamomea* L. 1753 ~ *Aster amellus* L. 1753.

Etymology

From the Latin *super* (above, over, or on top) and *Rosidae,* a converted clade name based on *Rosa*, the Latin name for rose (and probably originally from the Greek, *rhodon*), in reference to the fact that the *Superrosidae* is intended to apply to a crown clade "above *Rosidae*."

Reference phylogeny

The primary reference phylogeny is this paper, see Figs. 1 and 2; see also H. Wang et al. (2009) and Moore et al. (2010). *Rosa cinnamomea* is used as a specifier because it is the type species of Rosaceae (and thus *Rosidae*, under ICBN, which forms part of this clade name); its close relationship to the species of Rosaceae included in the primary reference phylogeny (i.e., *Spiraea betulifolia*, *S. vanhouttei*, and *Prunus persica*) is supported by the analyses of Evans et al. (2000) and Potter et al. (2002, 2007), in which Rosaceae are shown to be monophyletic. Similarly, *Aster amellus* is used as a specifier because it is the type species of Asteraceae (and thus *Asteridae*, under ICBN); its close relationships to the species of Asteraceae included in the primary reference phylogeny (i.e., *Barnadesia arborea*, *Gerbera jamesonii*, *Echinops bannaticus*, *Tragopogon dubius*, *T. porrifolius*, *Cichorium intybus*, *Lactuca sativa*, *Tagetes erecta*, *Guizotia abyssinica*, and *Helianthus annuus*) is supported by the series of phylogenetic studies of Asteraceae presented in Funk et al. (2009).

Composition

Rosidae (incl. Vitaceae), Saxifragales, and possibly Dilleniaceae.

Diagnostic apomorphies

No non-DNA synapomorphies are known.

Synonyms

There are no synonymous scientific names but the informal name "superrosids" was used for this clade in Moore et al. (2010), while the name "superrosids" is used in version 9 of Stevens (2001, onwards).

Comments

This is a recently discovered clade that is strongly supported by the extensive molecular analyses reported in this study, although we note that this clade was also well supported in both the 12-gene ML analysis of H. Wang et al. (2009; fig. 1), a study that focused only on relationships within the *Rosidae* and thus included many fewer sampled taxa, especially among nonrosid taxa, and the 83-gene ML and MP analyses of Moore et al. (2010; fig. 1), which also were based on a much smaller array of sampled taxa. The clade *Superrosidae* was recovered (but without strong support) in all most parsimonious trees in the three-gene analyses of Soltis et al. (2000; see figs. 1(B), 5, and 6) and in some of the most parsimonious trees resulting from the analysis of *atpB* and *rbcL* sequences in Savolainen et al. (2000a). The clade received weak support in the ML analyses of angiosperms based on five genes (Burleigh et al., 2009, see fig. 3 and the "full tree" included with their online supplemental information). Therefore, although the existence of the *Superrosidae* was suspected, and this clade was informally named in Moore et al. (2010), we have not previously provided a formal name for this clade. Only with the present results in hand are we sufficiently confident that *Superrosidae* represents a well-supported clade and thus is in need of a formal scientific name.

There is some disagreement, however, among recent phylogenetic analyses regarding the position of Dilleniaceae. In the 17-gene analysis (reported here), Dilleniaceae (represented by *Tetracera*, *Hibbertia*, and *Dillenia*) are strongly placed as sister to a clade comprising Santalales, *Caryophyllales*, Berberidopsidales, and *Asteridae*, while in an analysis of complete plastid genome sequence data (Moore et al., 2010) they are placed as sister to a Saxifragales + *Rosidae* clade. Finally, in an analysis of inverted repeat (IR) sequences (see Moore et al., 2010), Dilleniaceae are placed as the sister group to a large clade comprising *Rosidae*, Saxifragales, *Asteridae*, Berberidopsidales, *Caryophyllales*, and Santalales. The use of branch-modified, node-based definitions for both *Superrosidae* and *Superasteridae* accommodates placement of Dilleniaceae in either *Superrosidae* or *Superasteridae*, or positioned as the sister taxon to a *Superrosidae* + *Superasteridae* clade, within the *Pentapetalae* (which in turn is nested within the *Gunneridae*).

The essential feature of our concept of *Superrosidae* is its inclusion of everything that is more closely related to *Rosidae* (i.e., *Fabidae*, *Malvidae*, Vitaceae) than to *Asteridae* (i.e., *Lamiidae*, *Campanulidae*, Ericales, Cornales). This is the feature that we have tried to capture in our definition. Furthermore, when this definition is used in conjunction with our reciprocal definition of *Superasteridae* (see Appendix 2), it ensures that *Superrosidae* and *Superasteridae* are always mutually exclusive, regardless of the placement of Dilleniaceae.

APPENDIX 2. PhyloCode description of *Superasteridae*

Superasteridae W. S. Judd, D. E. Soltis, & P. S. Soltis

Definition

The most inclusive crown clade that contains *Aster amellus* L. 1753 (*Asteridae*) but not *Rosa cinnamomea* L. 1753 (*Rosidae*). This is a branch-modified node-based definition. Abbreviated definition: >∇ *Aster amellus* L. 1753 ~ *Rosa cinnamomea* L. 1753.

Etymology

From the Latin *super* (above, over, or on top) and *Asteridae*, a converted clade name based on *Aster* (derived from the Latin, *aster*, meaning star, so called because of the form of the radiate floral heads of these plants), in reference to the fact that *Superasteridae* is intended to refer to a crown clade "above *Asteridae*."

Reference phylogeny

The reference phylogeny is this paper; see Figs. 1 and 2. See also Burleigh et al. (2009, including online supplemental files 6 and 7). *Aster amellus* is used as a specifier because it is the type species of Asteraceae (and thus *Asteridae*, under ICBN, which forms part of this clade name); its close relationship to the species of Asteraceae included in the reference phylogeny (i.e., *Barnadesia arborea*, *Gerbera jamesonii*, *Echinops bannaticus*, *Tragopogon dubius*, *T. porrifolius*, *Cichorium intybus*, *Lactuca sativa*, *Tagetes erecta*, *Guizotia abyssinica*, and *Helianthus annuus*) is supported by the series of phylogenetic studies of Asteraceae presented in Funk et al. (2009). Similarly, *Rosa cinnamomea* is used as a specifier because it is the type species of Rosaceae (and thus *Rosidae*, under ICBN). The close relationship of *Rosa cinnamomea* to the species of Rosaceae included in the reference phylogeny (i.e., *Spiraea betulifolia*, *S. vanhouttei*, and *Prunus persica*) is supported by the analyses of Evans et al. (2000) and Potter et al. (2002, 2007), in which Rosaceae are shown to be monophyletic.

Composition

Santalales, Berberidopsidales, *Caryophyllales*, *Asteridae*, and possibly Dilleniaceae.

Diagnostic apomorphies

No non-DNA synapomorphies are known.

Synonyms

There are no synonymous scientific names, but the informal names "superasterids" (Moore et al., 2010) and "superasterids" Stevens (2001, onward, version 9) have been used for this clade.

Comments

This is a recently discovered clade that is strongly supported by the extensive molecular analyses reported in this study, although we note that this clade was also strongly supported in the 83-gene ML and MP analyses of Moore et al. (2010; see fig. 1), which were based on a much smaller array of sampled taxa. The *Superasteridae* received weak support in the ML analyses of angiosperms based on five genes (Burleigh et al., 2009, see fig. 3 and their "full tree" included with the online supplemental information). Therefore, although the existence of the *Superasteridae* was suspected, and this clade was informally named by Moore et al. (2010), we have not previously provided a formal name for this clade. Only with the present results in hand are we sufficiently confident that *Superasteridae* represents a well-supported clade, and thus is in need of a formal scientific name.

There is some disagreement among recent phylogenetic analyses regarding the position of Dilleniaceae. In the 17-gene analysis (reported here) Dilleniaceae (represented by *Tetracera*, *Hibbertia*, and *Dillenia*) are strongly placed as sister to a clade comprising Santalales, *Caryophyllales*, Berberidopsidales, and *Asteridae*, while in an analysis of complete plastid genome sequence data (Moore et al., 2010 they are placed as sister to a Saxifragales + *Rosidae* clade. Finally, in an analysis of IR sequences (see Moore et al., unpublished) Dilleniaceae are placed as the sister group to a large clade comprising *Rosidae*, Saxifragales, *Asteridae*, Berberidopsidales, *Caryophyllales*, and Santalales. The use of branch-modified node-based definitions for both *Superasteridae* and *Superrosidae* accommodates placement of Dilleniaceae in either *Superasteridae* or *Superrosidae*, or positioned as the sister taxon to a *Superasteridae* + *Superrosidae* clade, within the *Pentapetalae* (which is in turn nested within the *Gunneridae*).

The essential feature of our concept of *Superasteridae* that we have tried to capture in our definition is its inclusion of everything that is closer to Asteridae (i.e., *Lamiidae*, *Campanulidae*, Ericales, Cornales) than to *Rosidae* (i.e., *Fabidae*, *Malvidae*, Vitaceae). Furthermore, when this definition is used in conjunction with our reciprocal definition of *Superrosidae* (see Appendix 1), it ensures that *Superasteridae* and *Superrosidae* are always mutually exclusive, regardless of the placement of Dilleniaceae.