A Parsimony Analysis of the Asteridae Sensu Lato Based on rbcL Sequences

Richard G. Olmstead, Birgitta Bremer, Kathy M. Scott, Jeffrey D. Palmer

Annals of the Missouri Botanical Garden, Volume 80, Issue 3 (1993), 700-722.

Stable URL:
http://links.jstor.org/sici?sici=0026-6493%281993%2980%3A3%3C700%3AAPAOTA%3E2.0.CO%3B2-L

A PARSIMONY ANALYSIS OF THE ASTERIDAE SENSU LATO BASED ON *rbc*L SEQUENCES[1]

Richard G. Olmstead,[2] *Birgitta Bremer,*[3] *Kathy M. Scott,*[4] *and Jeffrey D. Palmer*[4]

ABSTRACT

A parsimony analysis of 156 representative sequences of the Asteridae sensu lato and 28 outgroup sequences was conducted using a two-tiered approach. First, an analysis of the entire group, including 105 sequences, examined relationships among major lineages within the Asteridae s.l.; subsequently, several clades within the larger group were examined individually in greater detail by including more sequences for the group in question. The search strategy was designed to discover multiple islands of equal parsimony using the heuristic search routine in PAUP. In the broad search and in each more detailed search of subclades, multiple islands were found that imply substantially different relationships. The results suggest a monophyletic Asteridae s.l., comprising the Ericales, Primulales, Ebenales and relatives of the Dilleniidae sensu Cronquist; Cornales, Apiales, and Hydrangeaceae of the Rosidae sensu Cronquist; and the conventionally circumscribed Asteridae. Within the Asteridae s.l., 11 groups were congruent between islands and are designated as follows: (1) Cornales, (2) ericalean clade, (3) *Garrya* clade, (4) *Ilex* clade, (5) Apiales, (6) Dipsacales, (7) Asterales s.l., (8) Gentianales, (9) Solanales, (10) Boraginales, and (11) Lamiales s.l. The only grouping between the level of these 11 clades and the whole Asteridae s.l. that is congruent between islands is the clade consisting of the Gentianales, Solanales, Boraginales, and Lamiales s.l., i.e., the Lamiidae of Takhtajan.

Most recent constructs of the subclass Asteridae (e.g., Cronquist, 1981; Takhtajan, 1987) represent reduced groups of the former Gamopetalae (de Candolle, 1813), which was characterized primarily by fused corollas (see Wagenitz, 1977, 1992, for a more complete review of the development of ideas about the Asteridae). A series of traits interferes with the usefulness of the simple dichotomy of fused versus free petals in defining major groups in the "higher" dicots (i.e., the tricolpate clade of Donoghue & Doyle, 1989), including number of stamen whorls and their position relative to corolla lobes (Dahlgren, 1983); embryological characters, such as one versus two integuments and the crassi- versus tenuinucellar condition in the ovule (Dahlgren, 1975); and, more recently, the distribution of secondary chemicals, particularly iridoids (Jensen, 1992). A "group" defined on the basis of any one of these traits will be incongruent with a "group" defined on the basis of any other (Wagenitz, 1977).

Conventional approaches to classification often proceed by *defining* groups on the basis of one or a few clearly defined traits, with additional traits used to refine groups and establish subgroups. In contrast, phylogenetic classification relies on the identification of derived traits to *discover* groups that are hypothesized to be monophyletic. Molecular approaches to systematics have become popular, in part, because they usually are well suited to phylogenetic analysis. Another reason for the popularity of molecular systematics is that phylogenetic hypotheses generated by molecular data generally are independent of the morphological traits in which many evolutionary biologists are interested.

With the combination of phylogenetic systematics and molecular methods, plant systematics has entered a new period of discovery. With respect to the Asteridae, we are not simply testing hypotheses set forth by previous workers (e.g., Cronquist, 1981; Dahlgren, 1980; Takhtajan, 1987; Thorne, 1992); instead we are discovering groups, and those groups may or may not reflect traditional classifications. However, we are indebted to pre-

[1] This research has been supported by NSF grants BSR-8717600 to J. Palmer and BSR-9107827 to R. Olmstead and NFR (Swedish Natural Science Research Council) grant B-BU 1487 to B. Bremer. Thanks are extended to D. Swofford and D. Maddison for helpful discussions of PAUP search strategies, to Steven Wagstaff for critical reading of the manuscript, and to the several researchers who have contributed *rbc*L sequences of the Asteridae s.l., particularly K. Kron and M. Chase (Ericales and related groups), Q.-Y. Xiang, D. Morgan, and D. Soltis (Apiales, Cornales, Hydrangeaceae, and relatives), M. Hedren and M. Chase (Acanthaceae), and H. Michaels (Asterales).

[2] Department of E.P.O. Biology, University of Colorado, Boulder, Colorado 80309, U.S.A.

[3] Department of Systematic Botany, Uppsala University, S-752 36 Uppsala, Sweden.

[4] Department of Biology, Indiana University, Bloomington, Indiana 47405, U.S.A.

vious workers for providing a framework for sampling in molecular studies. It is fortunate that there has been disagreement among conventional angiosperm phylogeneticists, or else our sampling may have been biased. The possibility of biased sampling combined with the limited body of data yet produced by molecular systematists means that interpretations should be made with caution. Suggestions made today (e.g., based on *rbc*L sequences) may be viewed in the not-too-distant future as we view those of 19th-century botanists. The results reported here fall into three classes: (1) results congruent with traditional classifications, (2) results resolving relationships of taxa ambiguously placed in traditional classifications, and (3) novel results wherein the sequence data are in apparent conflict with conventional data. To be treated properly, these novel results should be considered as starting points for further studies, in which both morphological *and* molecular data are re-examined and supplemented with additional data.

Recent efforts at discovering a monophyletic Asteridae (Downie & Palmer, 1992; Hufford, 1992; Olmstead et al., 1992) identified a large clade containing the Asteridae sensu stricto and several elements generally excluded from traditional circumscriptions of the group, including the Apiales, Ericaceae, Fouquieriaceae, Cornaceae, Hydrangeaceae, Loasaceae, and Columelliaceae. Olmstead et al. (1992) referred to this clade as the Asteridae sensu lato and identified four major lineages comprising the group, with relationships unresolved among them. The four lineages were the (1) ericalean clade (Ericaceae, Polemoniaceae, Fouquieriaceae), (2) cornalean clade (Cornaceae, Hydrangeaceae), (3) Asteridae sensu Takhtajan (1987) plus Apiales (Asterales, Campanulales, Goodeniales, Menyanthaceae, Dipsacales, Apiales), and (4) Lamiidae sensu Takhtajan (1987) (Gentianales, Solanales, Boraginales, Lamiales, and Scrophulariales). Restriction site analysis of the cpDNA inverted repeat region (Downie & Palmer, 1992) for the Asteridae s.s. and putative relatives concurred in finding a cornalean group, a group consisting of Polemoniaceae and Fouquieriaceae (Ericaceae was not sampled), and the Lamiidae (Takhtajan, 1987). Hufford (1992) identified a large Corniflorae–asterid group including a cornalean clade, an ericalean clade (excluding Polemoniaceae), and a conventional asterid clade (including Pittosporaceae and Columelliaceae).

Current work presented here and in accompanying papers represents another step toward discovering the best circumscription of the Asteridae sensu lato and groups within it. This series of papers is a fitting follow-up to the "Phylogeny of Asteridae" symposium proceedings published in this journal (Ann. Missouri Bot. Gard. 79(2), 1992), containing as it does papers on groups belonging to the Asteridae s.s. (Michaels et al., 1993) and papers on groups now included in the Asteridae s.l. (Xiang et al., 1993; Morgan & Soltis, 1993; Kron & Chase, 1993), in addition to this one on the Asteridae s.l.

A detailed study examining which sequences belong in or out of a group as large as the Asteridae s.l. is a formidable task. Olmstead et al. (1992) established a series of goals for their molecular studies of the Asteridae and addressed the first two: (1) to evaluate the monophyly of the Asteridae and (2) to identify major lineages within the Asteridae. This study addresses the next two goals: (3) to circumscribe ordinal-level groups through greater sampling and (4) to determine relationships among them. We anticipate that this will bring us one step closer to the final goals of a revised classification of the Asteridae and of a better understanding of evolutionary patterns and trends within the group. By focusing on the Asteridae s.l., we examine in much greater detail the *rbc*L data for taxa suggested to belong to the group than is possible in the analyses of Chase et al. (1993, this issue). The coarse nature of the broad-scope analyses of Chase et al. (1993) weakens their utility as rigorous hypotheses of relationship, but they do serve a useful purpose by suggesting sets of taxa that merit inclusion in more detailed analyses such as this one.

The analysis and interpretation of large data sets are complicated by the probable existence of multiple islands of trees representing equally or nearly equally parsimonious solutions (Maddison, 1991). A single parsimony analysis, even one using a detailed search strategy, can identify only one island. Identifying and presenting separate islands of equal parsimony, or even within one or two steps of the shortest trees (especially when these trees may contain 2,000–4,000 steps), provide valuable insight into possible patterns of relationship and are essential when revised classifications or interpretations of character evolution are to be based on the outcome of a phylogenetic analysis. An inadequate analysis, in which a search for multiple islands is not carried out, may produce exaggerated resolution, lead to misleading conclusions, and deter potential lines of further research.

Materials and Methods

The sequences included in the analyses reported here have been published previously (e.g., Olm-

stead et al., 1992), are from various researchers and included in the seed plant analysis of Chase et al. (1993), or are reported for the first time here. All sequences and their sources are listed in the Appendix at the end of this issue. Some of the sequences were obtained as described previously (Olmstead et al., 1992). The remaining sequences were determined by first PCR-amplifying the entire gene from a total DNA extract by use of primers described in Olmstead et al. (1992). Single-stranded DNA for dideoxy sequencing was then obtained by a second round of PCR using 0.01 μl of the original PCR product as template and a 1:100 primer ratio with the excess primer in the same concentration as in the initial PCR run.

Nucleotide positions 27–1428 were included in the cladistic analyses, but the actual number of nucleotides varied among species depending on whether sequences were complete for this region (see Appendix at end of issue). Positions 1–26 represent the PCR primer sequence for most sequences, and after position 1428 sequence alignment becomes problematic.

Several subsets of the entire data set were compiled to examine the entire Asteridae s.l. generally and several internal clades specifically, as well as to examine the question of sister group to the Asteridae s.l. In all cases, care was taken to include a number of outgroups, which were selected following examination of the results of an analysis of 475 *rbc*L sequences ("Search I" of Chase et al., 1993).

Parsimony analyses were conducted using PAUP 3.0s (Swofford, 1992) on a Mac IIfx. All analyses were conducted with equal weights assigned to all changes. The general search strategy proceeded as follows (variations are discussed in Results). First, two or more replicates of 200–1,000 searches each were carried out using a random-order, sequential entry with NNI branch swapping (nearest neighbor interchanges), STEEPEST DESCENT on, and with the MULPARS option off. The goal of this phase of the search was to increase the chance of locating the island or islands containing the shortest trees by finding a reasonably short tree to initiate a more detailed search. The shortest tree or trees from each replicate was entered separately as the starting point for the second phase search for all equal-length trees reachable from that starting point. The second phase of the search proceeded by using TBR swapping (tree bisection-reconnection) with MULPARS on. Searches on each individual data set, except for the one examining sister group relationships to the Asteridae s.l., were pursued until two or more islands of trees were discovered.

Results

Analysis of Asteridae s.l.

An analysis of 184 *rbc*L sequences is reported here, including 156 representatives of the Asteridae s.l. and 28 outgroup sequences. To examine the entire Asteridae s.l., a set of 105 sequences was compiled. To reduce the complexity of the data set while maintaining the best-possible distribution of taxa, several well-supported families with multiple representatives (e.g., Asteraceae, Acanthaceae, Cornaceae) were thinned from the number of representatives in the analyses of Chase et al. (1993, this issue). Choice of outgroup sequences was guided by the results of "Search I" of Chase et al. (1993). Five sequences (*Gunnera, Paeonia, Phoradendron, Schoepfia, Osyris*) comprising a clade identified as the sister group to the Asteridae s.l. in that search were included. Six progressively more remote outgroups, *Heuchera* and *Cercidiphyllum* ("rosid IV" clade of Chase et al., 1993), *Dianthus* ("rosid III"), *Tetracentron* ("hamamelid II"), *Platanus* ("hamamelid I"), and *Caltha* ("ranunculid"), were also included. All outgroups are "higher" dicots (sensu Donoghue & Doyle 1989), to minimize the problem of long outgroup branches (Swofford & Olsen, 1990), and fall outside the broadest possible circumscription of the Asteridae s.l. based on traditional or molecular evidence. Most of the outgroups belong to the closest sister groups (rosids and caryophyllids) to the Asteridae s.l. as identified by the subsequent and more complete "Search II" of Chase et al. (1993).

The search for the shortest trees connecting 105 sequences of the Asteridae s.l. and outgroups represents a major computational effort and resulted in the discovery of three islands of trees of equal parsimony. Considering the amount of time and effort expended in acquiring a data set of this size, the data deserve the substantial effort dedicated to their analysis. A more detailed description of the approach used will be described for this data set than for subsequent searches. As with any search where the number of taxa is above the limit of a branch-and-bound search algorithm (i.e., ca. 20 taxa), there is no way to guarantee finding the shortest tree or trees.

The search was initiated by performing 200 random-order-entry replicate searches using NNI swapping with MULPARS off (this search required approximately 5 hours and found one tree of length 3,609). The resulting tree was then used as a start for a search using TBR swapping with MULPARS off (26 min., length = 3,605), which was used, in turn, as the start for a search using NNI with

MULPARS on (2 hours, 432 trees of length 3,601). These trees were then used as the starting point for a search using TBR with MULPARS on. This search was terminated prematurely without swapping on all trees when the computer RAM buffer reached a limit of 4,600 trees (approximately 4 days, length = 3,600). An attempt to find shorter trees was then made by dividing the tree file into many smaller files containing subsets of the 4,600 trees. TBR swapping was initiated on several Mac II computers simultaneously with MAXTREES set to the number of trees in each subset plus one. This strategy enables TBR swapping to proceed on all 4,600 trees saved, but will not save any more trees. The search for shorter trees from the 4,600 trees of length 3,600 lasted several days with as many as eight computers operating simultaneously. Two separate trees of length 3,600 each led to the same island (Maddison, 1991) of 102 equal-length trees (length = 3,597).

The resulting strict consensus tree (Fig. 1A) for island-102 (terminology of Maddison, 1991) shows a basal split in the Asteridae s.l. with one branch consisting of the Lamiidae and the small *Garrya* clade and the other consisting of the ericalean clade, the Cornales (including the Hydrangeaceae), Apiales, Dipsacales, Asterales s.l., and the small *Ilex* clade. In addition to the basal split into two major lineages, this tree differs from the results of Chase et al. (1993) in showing the Cornales as a more derived branch than the ericalean clade.

While this search was in progress, a second search was initiated on the same set of 105 sequences by performing 1,000 random-order-entry replicate searches using NNI swapping with MULPARS off (26 hours, one tree, length = 3,605). This tree was then input directly into a search using TBR swapping with MULPARS on. This search required approximately 10–12 weeks computer time and resulted in the discovery of a second island of length 3,597 containing 2,784 trees (Figs. 1B, 2). During this search >6,300 trees were found at one step longer, of which >3,500 were swapped before a shorter tree was found. The topology of the strict consensus tree (Fig. 1B) for island-2,784 is dramatically different from that derived from island-102 with respect to the relationships among the primary groups identified within the Asteridae s.l. These results unite the Lamiidae with the Asterales s.l., thus forming a clade equivalent to the Asteridae of Cronquist except for the exclusion of the Dipsacales. The Dipsacales and Apiales together are sister group to the above-mentioned clade, with the small *Ilex* and *Garrya* clades each forming progressively more basal branches, and the Cornales and ericalean clade producing a basal trichotomy.

After the first two islands were discovered using the search strategy outlined above, another approach to finding initial trees, possibly representing separate islands, was tried (suggested by D. Maddison, pers. comm.). One hundred replicate random-order-entry searches were conducted using TBR swapping with MULPARS on, but with the option to save no more than two trees greater than, or equal to, length "X," where "X" is a number small enough to be certain to be shorter than the shortest tree found. This approach enables each random start to be searched much more completely than the approach outlined above. One replicate yielded two trees of length 3,597 and which were topologically distinct from any tree belonging to island-102 or island-2,784. A search for all most parsimonious trees yielded 5,568 trees at that length. The initial search required 11 days and the search for all parsimonious trees in the island required an additional seven weeks to complete. This island contains exactly twice as many trees as island-2,784 and can be accounted for by one additional polytomy (within the Gentianales). Topologically, this island is identical with island-2,784, except that the Boraginales are united with the Gentianales instead of with the Solanales and the relationships among the five sequences representing the Gentianales are different (hence this island is not illustrated). Figure 3 shows the strict consensus of all trees discovered in the three islands.

Another analysis was performed to explore more fully the putative sister group to the Asteridae s.l. "Search I" of Chase et al. (1993, this issue) identified the clade mentioned above as sister group, whereas their "Search II" (Chase et al., 1993) identified a small clade of two sequences (*Dillenia* and *Vitis*) as sister group to the Asteridae s.l. For this analysis, 40 sequences of the Asteridae s.l. were removed and 17 additional dicot sequences were included, giving a total of 82 sequences. This analysis found 30 equal-length trees of length 3,322 (Fig. 4). This search was initiated with 300 random-order-entry replicate searches and was not replicated to find alternate islands, but the search strategy was equivalent to that used to find one of the shortest islands in other analyses reported here. This result is congruent with that based on the previous analysis in suggesting that the Santalales, *Paeonia*, and *Gunnera* do not belong together or with the Asteridae s.l. and more likely belong in the large clade identified as Rosidae s.l. in the analyses of Chase et al. (1993). The results do not place the *Dillenia* and *Vitis* sequences as sister

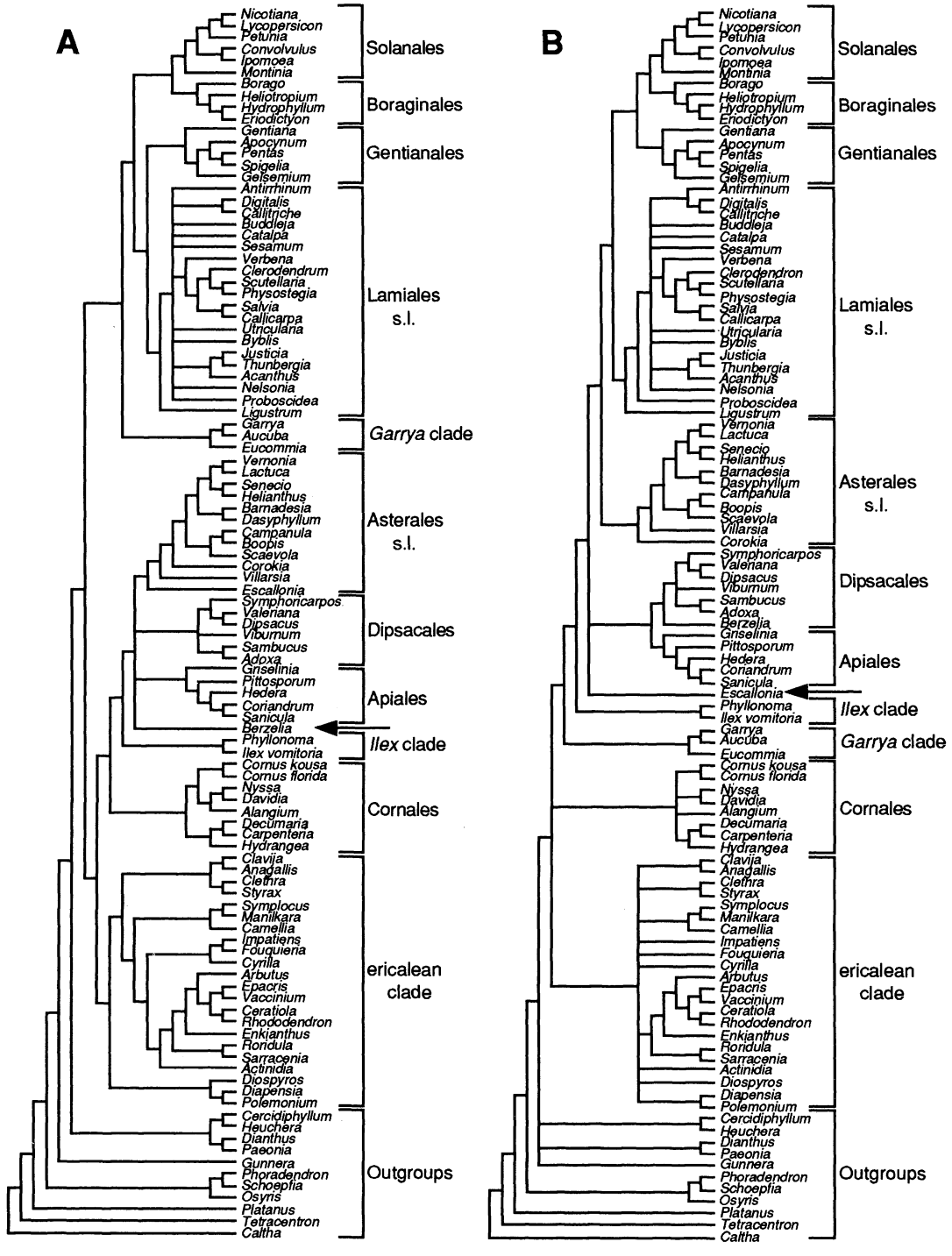

FIGURE 1. Strict consensus trees for two islands of equal parsimony in the Asteridae s.l. (105 *rbc*L sequences, including 11 outgroups; length = 3,597; C.I. = 0.25; R.I. = 0.51).—A. Island-102. Strict consensus of 102 trees.—B. Island-2,784. Strict consensus of 2,784 trees. Eleven clades that are congruent between the two islands and island-5,568 (not shown) are identified. Arrows identify sequences that resolve to a specific clade in one island, but not in the other.

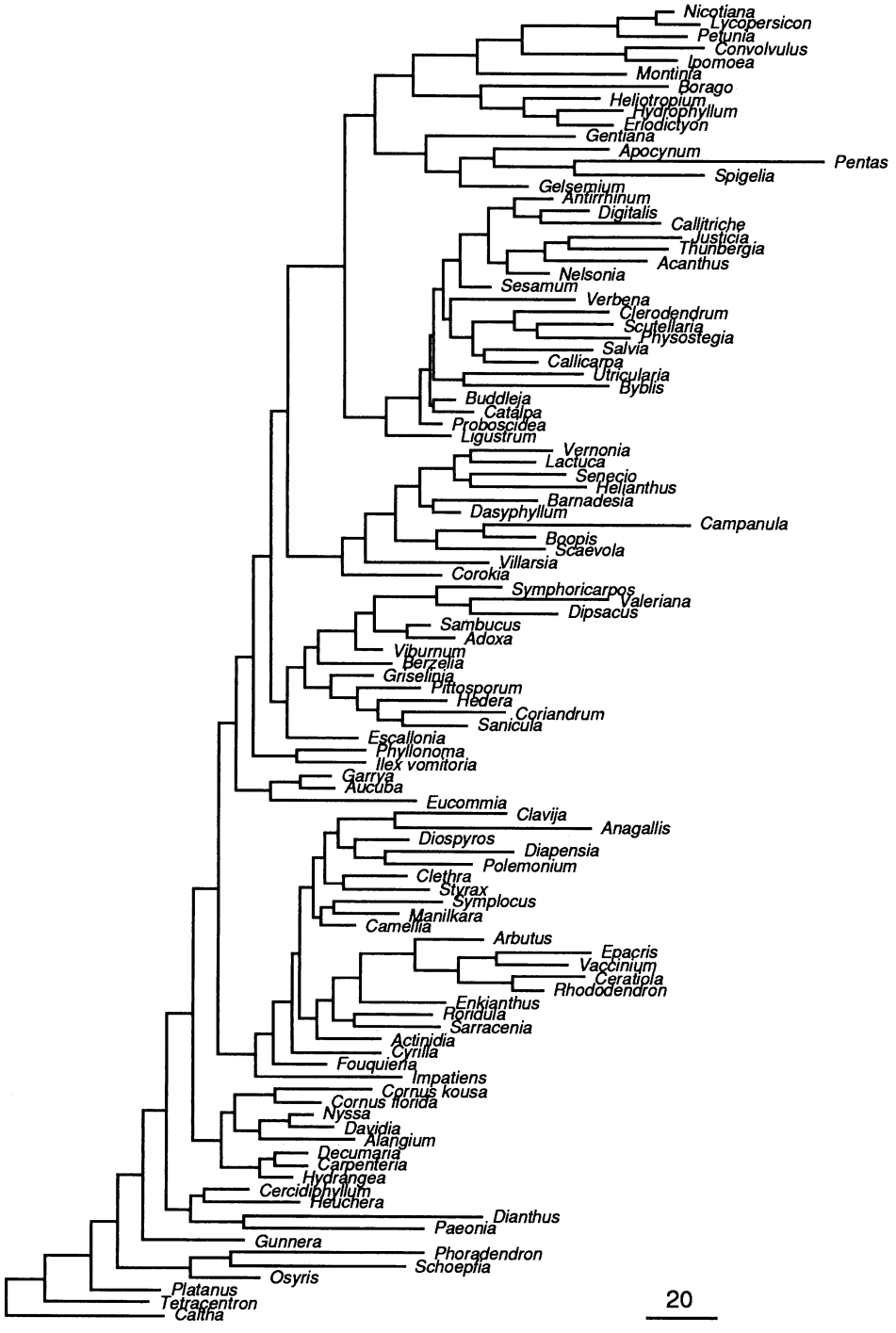

FIGURE 2. Majority rule consensus tree of island-2,784 selected as representative tree to illustrate branch lengths (ACCTRANS optimization). Particularly long branches are evident in *Pentas* (Rubiaceae) and *Campanula* (Campanulaceae); see discussion concerning incongruent results in different analyses containing these sequences (e.g., *Pentas* in Figs. 3 and 5). Note scale bar equal to 20 inferred nucleotide substitutions.

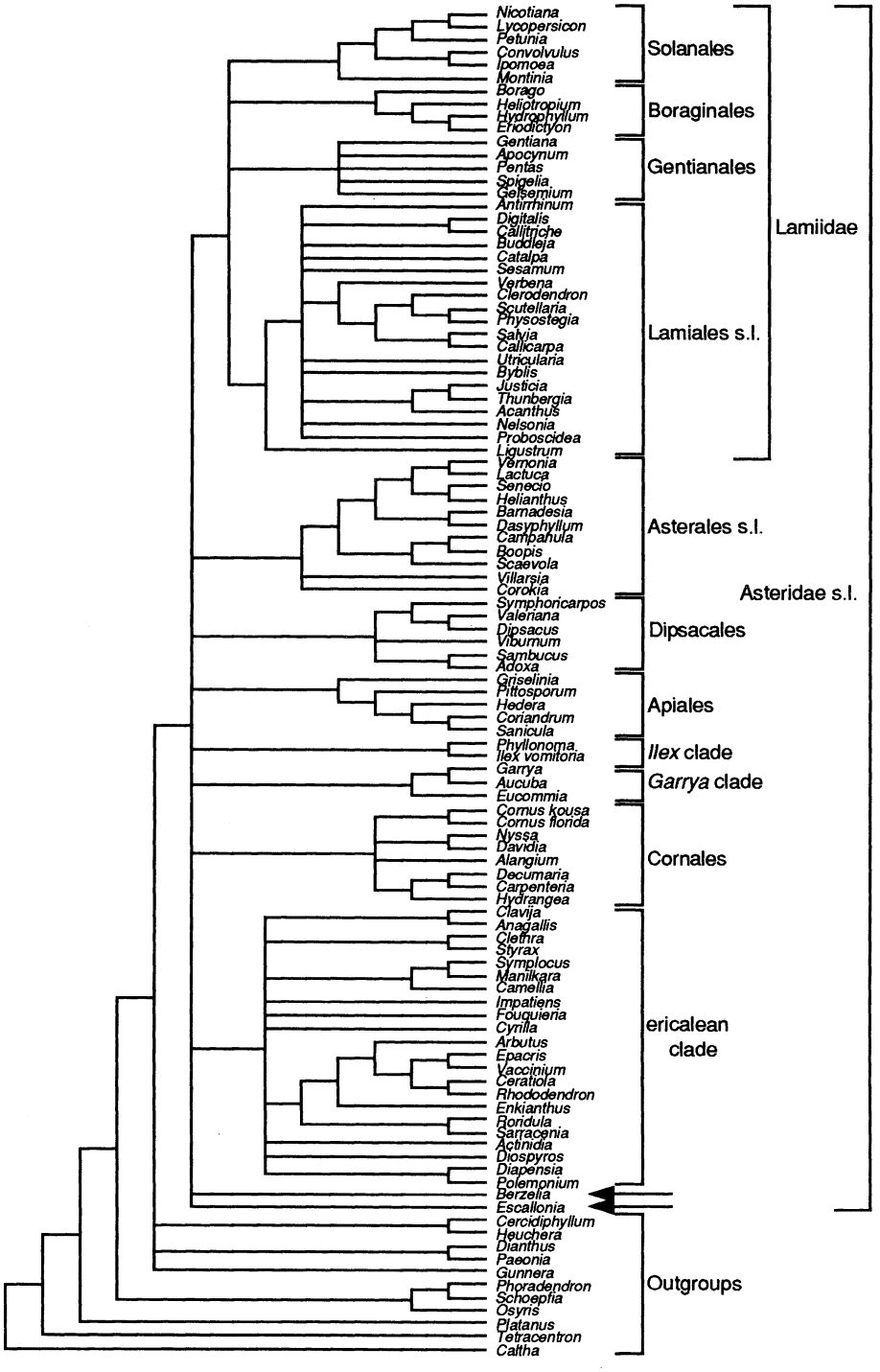

FIGURE 3. Strict consensus tree of 8,454 equal-length trees based on *rbc*L sequences. Trees from three islands (island-102, island-2,784, and island-5,568) are combined. Eleven terminal clades that are congruent among all three islands are identified. The Lamiidae sensu Takhtajan (1987) is the only group resolved between the level of the 11 congruent clades and the entire Asteridae s.l. *Berzelia* and *Escallonia* (indicated with arrows) are not resolved to any of the clades that are congruent between islands.

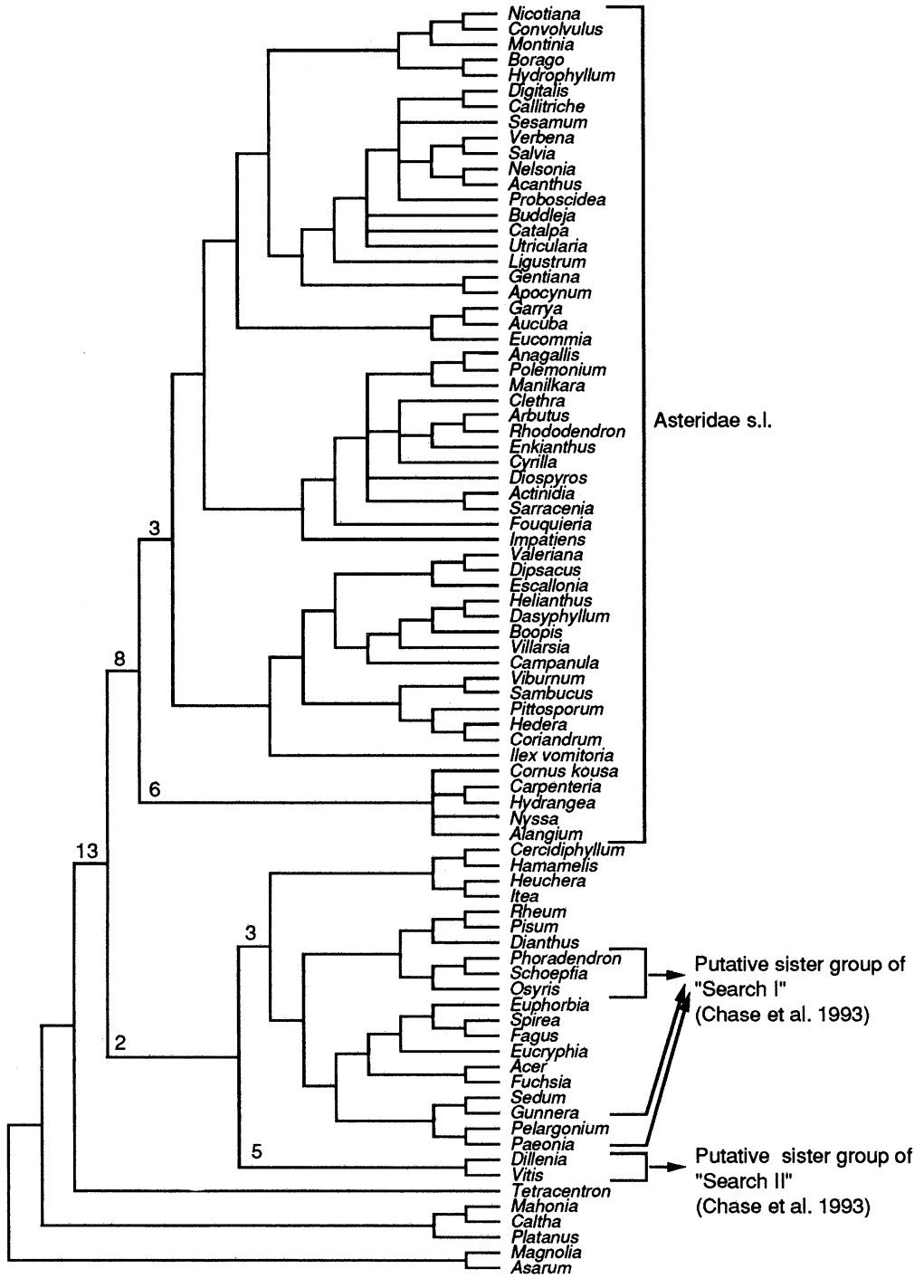

FIGURE 4. Strict consensus of 30 trees based on *rbc*L sequences resulting from the analysis designed to explore sister-group relationships to the Asteridae s.l. (length = 3,322, C.I. = 0.25, R.I. = 0.41). The sister group to the Asteridae s.l. indicated by this analysis is a large clade consisting of parts of the Rosidae and Dilleniidae sensu Cronquist (1981). Putative sister groups suggested by "Search I" and "Search II" of Chase et al. (1993) are indicated. Typical branch lengths in the vicinity of the node from which the Asteridae s.l. and their sister group diverge are indicated.

group to the Asteridae s.l. either. Instead, these sequences are placed at the base of the Rosidae s.l., suggesting that they represent a clade that diverged close to the divergence of the two larger clades. A search for multiple islands might find equal length trees congruent with the results of "Search II" of Chase et al. (1993) with respect to the placement of these two sequences.

ANALYSIS OF SUBCLADES

Detailed analyses were conducted on three subsets of the asterid *rbc*L data that were identified as clades by Olmstead et al. (1992) and in "Search I" of Chase et al. (1993). The primary purpose for the subclade analyses is to examine in detail more sequences within major clades than is possible in the more inclusive broad analysis of the Asteridae s.l. Thus, the subclade analyses are viewed in relationship to the Asteridae s.l. analysis as that analysis is to the seed plant analyses of Chase et al. (1993). Results from this nested hierarchical approach should be complementary; the local analyses better assess relationships within terminal clades and the broader analyses better assess relationships among those clades. Time considerations dictated that selection of these subsets of sequences precede the completion of "Search II" (Chase et al., 1993) and the search for multiple islands conducted on the 105-sequence Asteridae s.l. data set described above. The first subset analyzed is the Lamiidae sensu Takhtajan (1987). The second (hereafter referred to as the Asterales–Dipsacales–Apiales) contains the Asterales, Campanulales, Dipsacales, Apiales, and several taxa traditionally assigned to other orders in the Asteridae or Rosidae (sensu Cronquist, 1981). Both "Searches I & II" of Chase et al. (1993) identified a small clade containing *Garrya* (Garryaceae), *Aucuba* (Cornaceae), and *Eucommia* (Eucommiaceae) placed near the point of divergence of the two above-mentioned, larger clades. To examine the possible relationships of these three taxa, they were included in each analysis. The third clade examined in greater detail is one containing representatives of the Ericales, Ebenales, Primulales, and other taxa from the Rosidae, Dilleniidae, and Asteridae. The Cornales and Hydrangeaceae represent the only primary group within the Asteridae s.l. not included in a more detailed analysis here, because a detailed analysis of this group appears elsewhere (Xiang et al., 1993).

To examine the Lamiidae, a set of 72 sequences was compiled. This set included all of the sequences belonging to the group "Asteridae I" in "Search II" of Chase et al. (1993), except *Vahlia*, and included nine additional sequences from taxa traditionally assigned to the orders Gentianales (six) and Scrophulariales (three). Outgroup sequences included *Cornus* (Cornaceae), *Nyssa* (Nyssaceae), *Arbutus* (Ericaceae), *Polemonium* (Polemoniaceae), *Ilex* (Aquifoliaceae), *Escallonia* (Grossulariaceae), *Hedera* (Araliaceae), *Viburnum* (Caprifoliaceae), and *Dasyphyllum* (Asteraceae). The analysis of the Lamiidae data set (72 sequences) resulted in the discovery of four islands (although more may exist); two islands contain 135 (Fig. 5) and 1,350 trees (Fig. 6A), each of 2,189 steps, whereas the other two are one step longer (see below). Both islands identify as clades the Solanales (e.g., Solanaceae, Convolvulaceae, and *Montinia*), Boraginales (e.g., Boraginaceae, Hydrophyllaceae), Lamiales s.l. (e.g., Oleaceae, Bignoniaceae, Labiatae, Verbenaceae, Scrophulariaceae, Acanthaceae), and Gentianales (e.g., Rubiaceae, Gentianaceae, Apocynaceae, Loganiaceae), with a clade comprising the Solanales and Boraginales sister group to the others. The differences between the two islands fall entirely within the Lamiales s.l. clade, where several individual sequences or small groups of sequences occupy very different positions. Much resolution within the Lamiales s.l. is lost in the strict consensus of both islands combined (Fig. 6B). Both islands identify the *Garrya–Aucuba–Eucommia* clade as sister group to the Lamiidae. Two additional islands (not shown) containing 2,160 and 135 trees representing local optima were discovered at one step longer (2,190 steps). These trees are identical to the shorter islands with respect to the pattern of relationships among primary lineages within the Lamiidae, but differ in that the *Garrya* clade is not identified as their sister group.

To examine the Asterales–Dipsacales–Apiales, a set of 43 sequences was compiled. This set included all of the sequences belonging to the group "Asteridae II" in "Search II" of Chase et al. (1993), except that six representatives of the Asteraceae were included (rather than 14), and one additional sequence each of *Ilex* and of Apiaceae were added. *Garrya*, *Aucuba*, and *Eucommia* were included. Outgroup sequences included *Cornus*, *Nyssa*, *Arbutus*, *Polemonium*, *Nicotiana* (Solanaceae), *Gentiana* (Gentianaceae), and *Antirrhinum* (Scrophulariaceae). The analysis of this data set resulted in the discovery of two islands containing 80 (Fig. 7A) and 196 trees (Fig. 7B) each of 1,368 steps. The smaller island identifies a monophyletic Asterales–Dipsacales–Apiales, whereas the island of 196 trees (Fig. 7B) is unresolved with respect to these three orders and

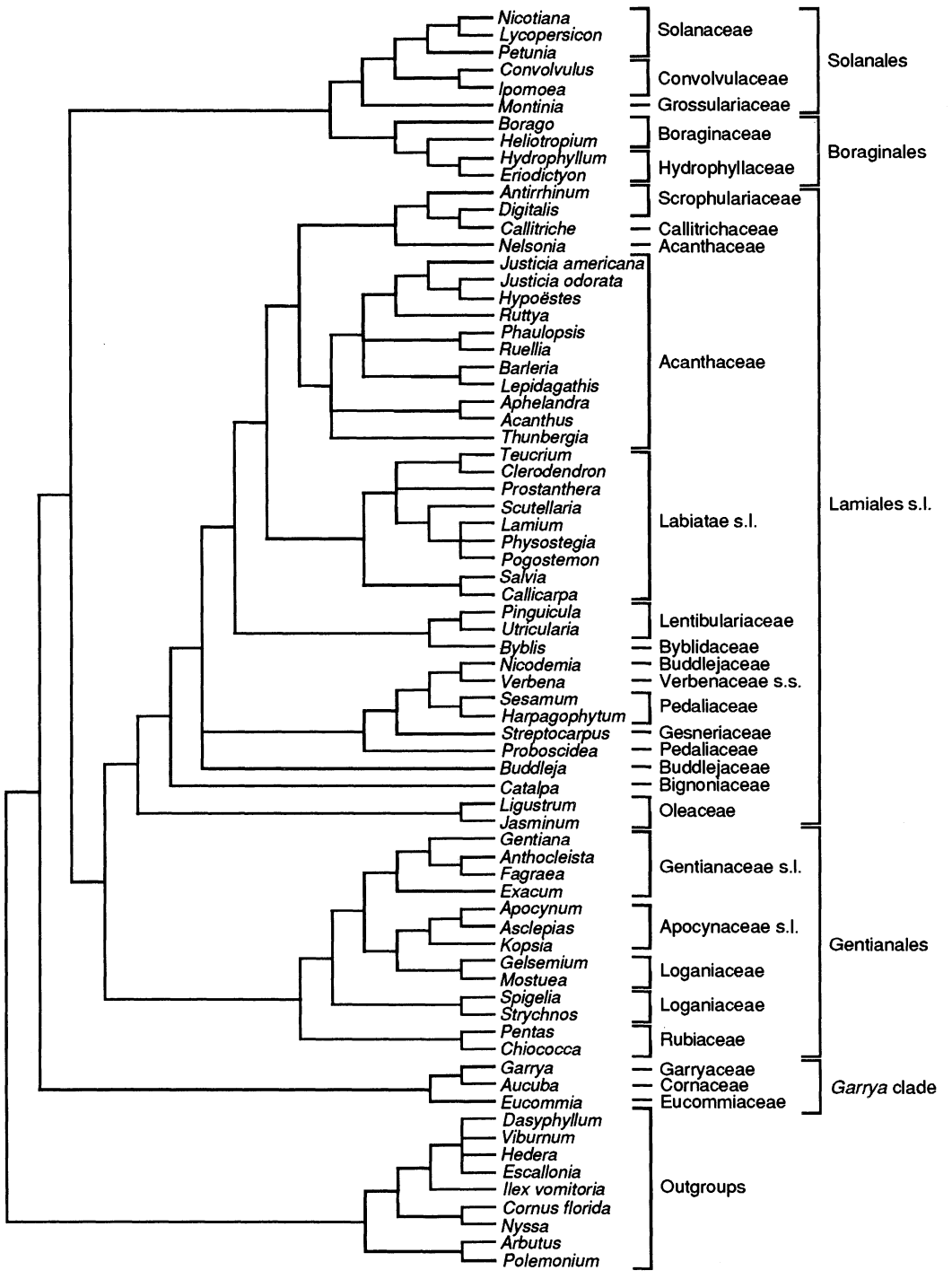

FIGURE 5. Lamiidae, island-135. Strict consensus of 135 trees based on *rbc*L sequences with increased taxon sampling in the Lamiidae (length = 2,189, C.I. = 0.31, R.I. = 0.57). Family designations follow Cronquist (1981), except for the Labiatae (Cantino et al., 1992), and do not imply necessarily that the entire family belongs where these representatives occur. Two additional islands each one step longer place the *Garrya* et al. clade among the other designated outgroups and not as sister group to the Lamiidae (not shown).

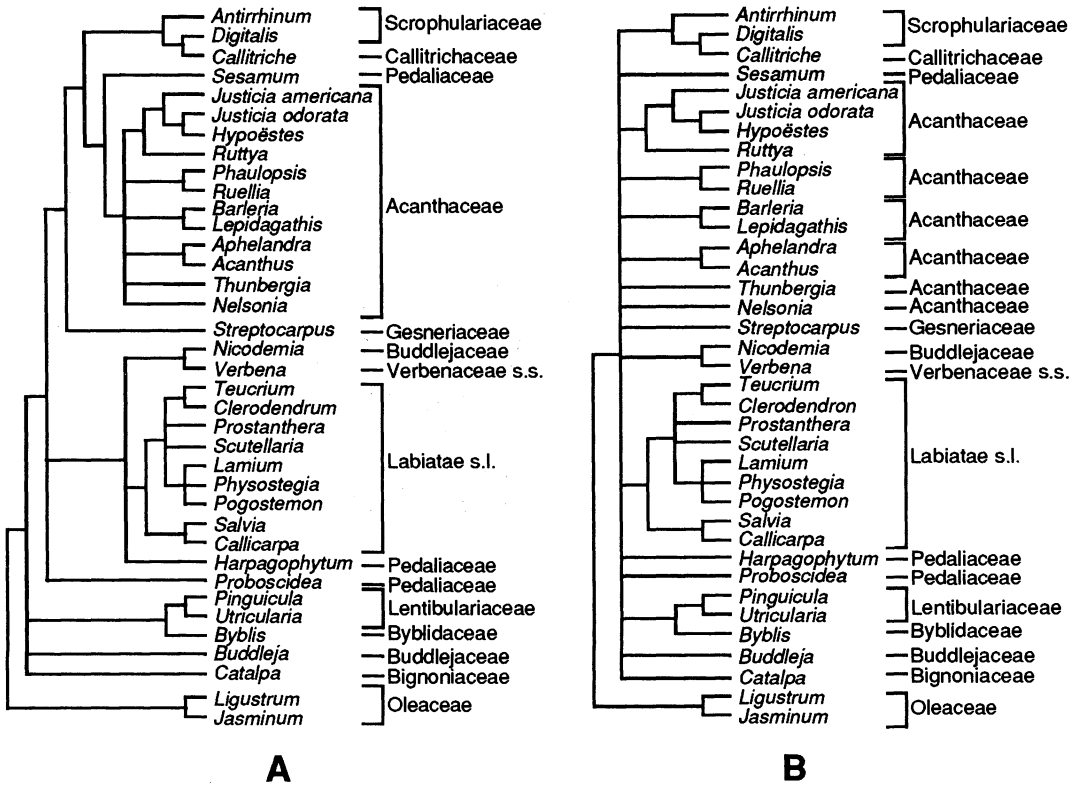

FIGURE 6. Strict consensus trees of the Lamiales s.l. based on *rbc*L sequences. These trees represent the portion of the Lamiidae analysis that is variable among islands; relationships among other members of the Lamiidae and outgroups are the same as in Figure 5.—A. Island-1,350. Strict consensus of 1,350 trees (length = 2,189, C.I. = 0.31, R.I. = 0.57).—B. Strict consensus of 1,485 equal-length trees from island-135 and island-1,350 combined. Family designations as in Figure 5.

several other groups (however, the Dipsacales and Apiales group with representatives of the Cornales in many trees). This is similar to the results of the Asteridae s.l. analysis, in which one island identifies this clade (Fig. 1A) and the other does not (Fig. 1B). Both islands identify the following clades: Dipsacales, Asterales s.l. (including Campanulales, Menyanthaceae, and *Corokia*), Apiales (including *Pittosporum* and *Griselinia*), and the group containing *Ilex*, *Phyllonoma*, and *Helwingia*.

To examine the ericalean clade, a set of 60 sequences was compiled, including all of the sequences belonging to the group "Asteridae III" in "Search II" of Chase et al. (1993) except *Heliamphora*. Eight additional sequences representing the Ericaceae were included (see Kron & Chase, 1993). Nineteen representatives of other clades in the Asteridae s.l. were included as outgroups. The analysis of this data set resulted in the discovery of two islands of trees, one containing 612 trees (Fig. 8B) of 1,889 steps and the other ,containing 357 trees (Fig. 8C) of 1,890 steps. Both islands identify a monophyletic Ericaceae s.l. (including the Epacridaceae, Empetraceae, and Pyrolaceae), in agreement with Kron & Chase (1993). All members of this clade belong to Cronquist's Dilleniidae, except *Impatiens*, *Roridula*, and *Polemonium* (see discussion below). The largest clade that is congruent between the two islands is the Ericaceae. The shorter island is congruent in most respects with the results of Kron & Chase (1993). The strict consensus tree of both islands combined indicates very little resolution outside the Ericaceae (Fig. 8A).

DISCUSSION

The goals of a study of this nature (i.e., reconstructing the phylogeny of the Asteridae) are lofty, but accomplishing them is problematic. A major problem with analyzing this and any other large and complex systematic data set is assessing the

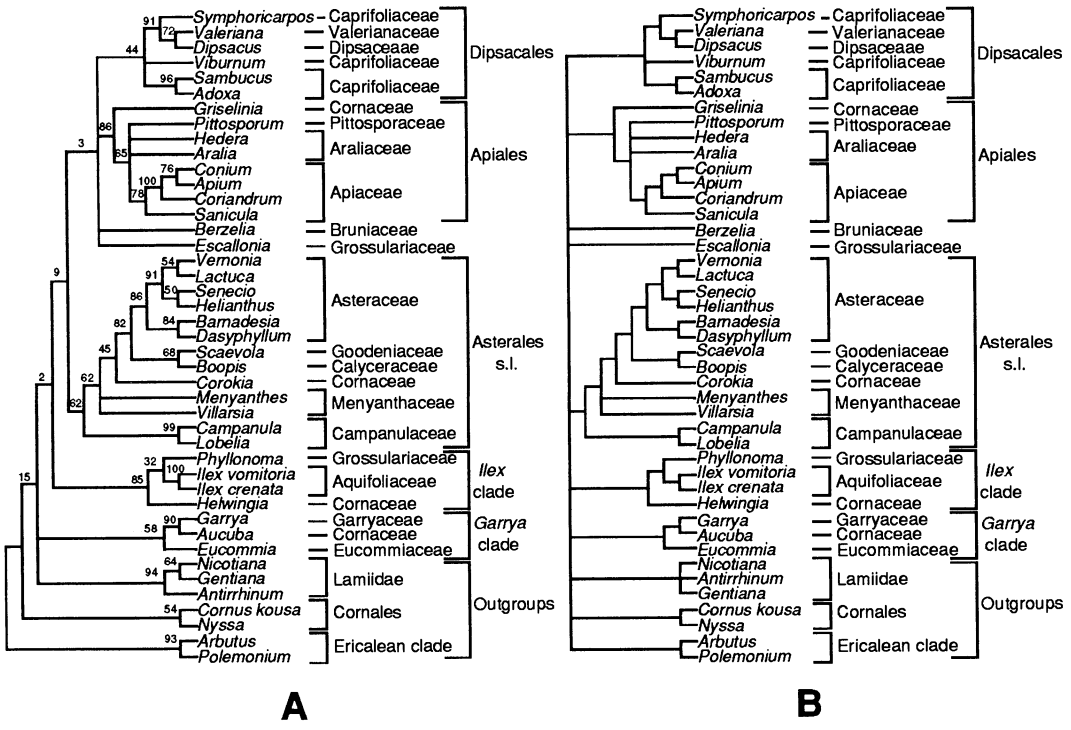

FIGURE 7.—Strict consensus trees for *rbc*L sequences representing the Asterales s.l., Apiales, and Dipsacales.—A. Island-80. Strict consensus of 80 trees (length = 1,368, C.I. = 0.39, R.I. = 0.51). Numbers above branches indicate percent occurrence of subtended clade in 100 bootstrap replicates.—B. Island-196. Strict consensus of 196 trees (length = 1,368, C.I. = 0.39, R.I. = 0.51). This consensus tree is the same as for the combined consensus of island-80 and island-196. Family designations follow Cronquist (1981) and do not imply necessarily that the entire family belongs where these representatives occur.

reliability of the results. How can one be sure that the results found represent the optimal solution for the data? How can one resolve incongruence between separate analyses conducted on similar data sets (e.g., differences between the analyses of Chase et al., 1993, this issue, and analyses presented here)? Will sampling more taxa, sampling more characters (e.g., nucleotides), or more detailed analyses of a subset of the data provide better results? It is not clear that there is a single best answer to any of these questions and when data of this nature are compiled for the first time for a group of organisms, as these are for the Asteridae (and angiosperms), there is no independent, external means for evaluating the results. Internal methods for evaluating trees (e.g., bootstrap or decay analysis) are extremely difficult to implement on a large data set. Bearing in mind that there is little in the way of rigorous, quantitative, phylogenetic hypotheses against which to evaluate these results, they should be considered preliminary, awaiting the availability of a substantially greater amount of data derived from other sources.

For the results of a cladistic analysis to provide strong support for a hypothesis, alternative hypotheses must be examined. This requires searching for multiple islands with sufficient diligence to reach other islands, if they exist. For one well-known data set of human mitochondrial DNA sequences (Vigilant et al., 1991), a reanalysis of the data yielded separate islands each having different implications with respect to the geographic origin of modern humans (Maddison et al., 1992). The initial analysis (Vigilant et al., 1991) was unable, by design, to find more than one island and, therefore, was predestined to yield one conclusion.

A total of approximately eight to ten months of computer time was spent on the several analyses reported here. Multiple islands were found whenever the appropriate search strategy was employed for each data set examined in this study. In all cases only two or three separate initial searches each of 200–2,000 random-order-entry starting points were required to find multiple islands, leaving open the possibility that additional islands, some perhaps shorter, remain to be found. Whereas it

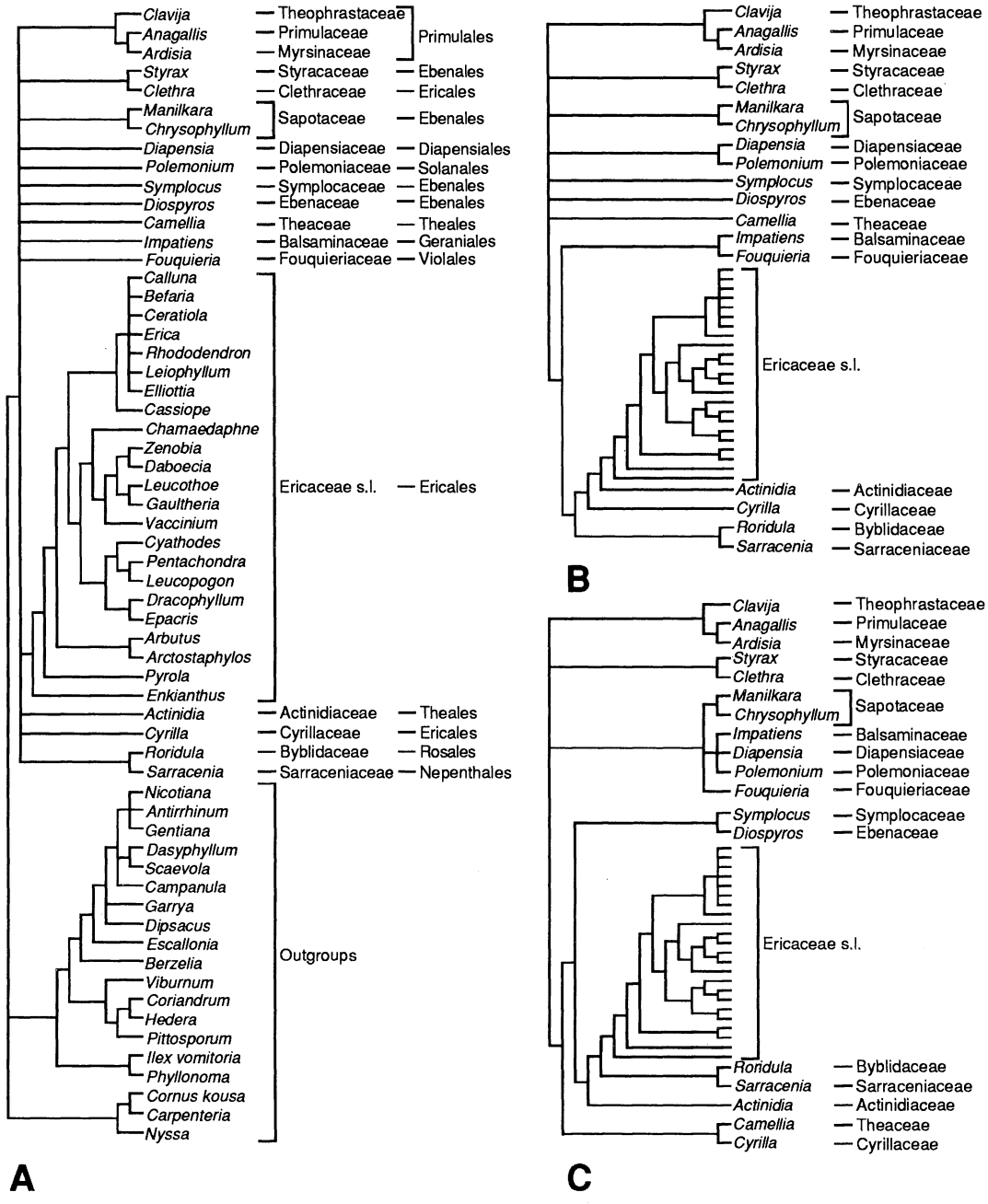

FIGURE 8. Strict consensus trees for the ericalean clade based on *rbc*L sequences.—A. Strict consensus of 969 trees combining island-612 and island-357. Resolution resides almost exclusively within the Ericaceae s.l.—B. Strict consensus of island-612 (612 trees length = 1,889, C.I. = 0.33, R.I. = 0.55).—C. Strict consensus of island-357 (357 trees length = 1,890, C.I. = 0.33, R.I. = 0.55). Relationships among outgroups and within Ericaceae s.l. are the same for both islands, thus are only indicated in A. Family designations follow Cronquist (1981) and do not imply necessarily that the entire family belongs where these representatives occur.

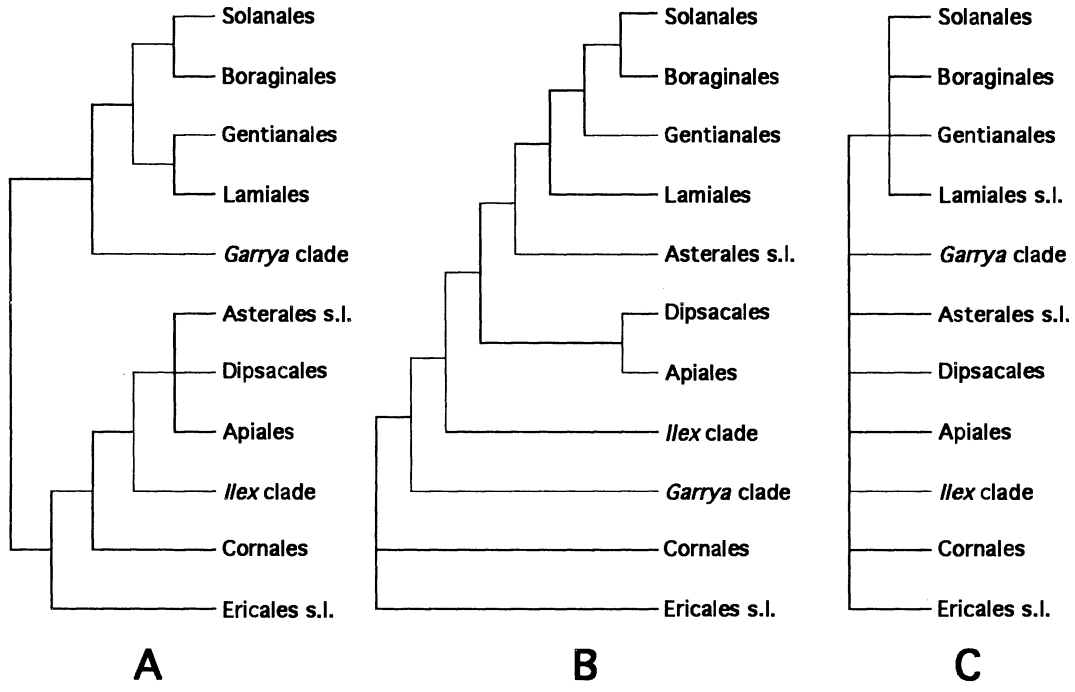

Figure 9.—Condensed cladograms representing the relationships among primary groups in the Asteridae s.l. based on *rbc*L sequences.—A. Island-102.—B. Island-2,784.—C. Combined consensus of island-102, island-2,784, and island-5,568.

is not possible to know if all islands within the Asteridae s.l. have been discovered, finding at least two islands provides a cautionary note with respect to the extent to which the findings can be used for evolutionary and systematic interpretations.

The discovery of multiple islands of trees of equal, or near-equal, length implies that there are regions of conflict in the hypotheses of relationship that derive from them. The conclusions in which we can have the most confidence are those that are shared between islands. The analysis of 105 sequences (designed to examine overall relationships within the Asteridae s.l.) and the analysis of 82 sequences (which includes numerous additional outgroup sequences, but reduced sampling within the Asteridae s.l.) both agree with the conclusions of Olmstead et al. (1992) and Chase et al. (1993) in the circumscription of a monophyletic Asteridae s.l. Within the Asteridae s.l., the three islands of trees discovered in the analysis of 105 sequences are congruent in the composition of 11 terminal groups (Fig. 9) and in the monophyly of one group comprising four of the terminal groups (Lamiidae sensu Takhtajan, 1987). Differences among islands exist both within and among these 11 groups, indicating that different cladistic patterns within monophyletic groups can influence the most parsimonious patterns among those groups. Two sequences (*Berzelia* and *Escallonia*) each fall into a specific clade in one island, but remain unresolved in the other island. Interestingly, these two sequences represent families that Hufford (1992) identified as individual branches at the base of the entire Asteridae s.l. They will be included in the discussion below with the clade to which they belong in at least one island.

The names used for groups identified here were chosen to reflect formally described taxonomic groups to which they are most closely similar. However, we consider these to be provisional names at this time, and they should not be viewed as constituting a formal classification. The 11 putatively monophyletic groups identified within the Asteridae s.l. are as follows (family, order, and subclass designations for included representatives follow Cronquist, 1981, unless otherwise noted):

(1) **Cornales.** This clade contains representatives of the Cornaceae, Hydrangeaceae, and Nyssaceae, all formerly placed in the subclass Rosidae. Several representatives placed within or near one or the other of these families in most angiosperm classification schemes are found elsewhere in the

*rbc*L tree of the Asteridae s.l. Three subgroups are identified within this clade: the Hydrangeaceae, *Cornus* (representing the Cornaceae s.s.), and the Nyssaceae. *Alangium* remains unresolved with respect to placement in either the Cornaceae s.s. or Nyssaceae (but see Xiang et al., 1993). Both Anderberg (1992) and Hufford (1992) identified this clade (including Loasaceae and Sarraceniaceae in Hufford, 1992) in their morphology-based cladistic analyses. Greater sampling and more detailed analysis of this clade shows substantial resolution based on *rbc*L sequences; see Xiang et al. (1993) and Morgan & Soltis (1993) for further discussion.

(2) **Ericalean clade.** This clade is mostly composed of former representatives of the Dilleniidae, including the orders Diapensiales, Ebenales, Ericales, Nepenthales, Primulales, Theales, and Violales (*Fouquieria*). The only non-dilleniid representatives in this clade are *Polemonium* (Polemoniaceae, Asteridae), *Roridula* (Byblidaceae, Rosidae), and *Impatiens* (Balsaminaceae, Rosidae). The clade identified here (see also Chase et al., 1993; Kron & Chase, 1993) has not been recognized in any formal classification. Within this group (Fig. 8A) representatives of the Ericaceae s.l. form a clade (Kron & Chase, 1993), as do a few pairs or triplets of sequences: *Clavija*, *Anagallis*, *Ardisia* (Primulales), *Styrax* (Styracaceae, Ebenales) with *Clethra* (Clethraceae, Ericales), *Manilkara* with *Chrysophyllum* (Sapotaceae, Ebenales), and *Sarracenia* (Sarraceniaceae, Nepenthales) with *Roridula* (Byblidaceae, Rosales). Most of the rest of the sequences in this clade are single representatives of families. This clade is noteworthy for the lack of resolution even within islands at the level of sampling represented here (Fig. 8B, C). *Actinidia* and *Roridula–Sarracenia* consistently occur with the Ericaceae (Fig. 8), each as sister group in one island, but do not appear as part of the clade in the consensus tree, because *Cyrilla* occurs in that clade in one island and not in the other. Considerably more work is needed before the picture of relationships within this clade can be brought into focus. See Kron & Chase (1993) for further discussion.

(3) ***Garrya* clade.** This small clade is represented by three sequences: *Garrya* (Garryaceae, Cornales), *Aucuba* (Cornaceae), and *Eucommia* (Eucommiaceae, Eucommiales, Hamamelidae). This represents a previously unrecognized group and includes species not formerly placed in the Asteridae. However, morphological and chemical features uniting these taxa and supporting their association with the Asteridae s.l. can be recognized (see below and Xiang et al., 1993).

(4) ***Ilex* clade.** This small clade is represented by four sequences: *Ilex crenata* Thunberg and *I. vomitoria* Aiton (Aquifoliaceae, Celastrales, Rosidae), *Helwingia* (Cornaceae), and *Phyllonoma* (Grossulariaceae, Rosales). This is another previously unrecognized clade, which does, nonetheless, have putative morphological synapomorphies (Morgan & Soltis, 1993) and is united with the Asteridae s.l. by the presence of unitegmic and tenuinucellate ovules. Our results, with greater sampling within the Asteridae s.l., differ from those of Morgan & Soltis (1993) in identifying *Ilex* and *Phyllonoma* as sister taxa (Fig. 7A, B). However, Morgan & Soltis (1993) emphasize that *Helwingia* and *Phyllonoma* share the unusual feature of epiphyllous inflorescences, suggesting that they may be more closely related than either is to *Ilex*.

(5) **Apiales.** This clade comprises representatives of the families Pittosporaceae, Araliaceae, Apiaceae, and *Griselinia* (Cornaceae), all formerly placed in the Rosidae. The Araliaceae and Apiaceae long have been recognized as closely related, and Dahlgren (1980) recognized the association of Pittosporaceae with the Apiales. Anderberg (1992) identified a monophyletic Apiales–Pittosporaceae in his morphology-based analysis of the Ericales and related families, whereas Hufford (1992) found no support for this union in his morphology-based analysis. The representatives of the Apiaceae included here form a monophyletic group (but see Plunkett et al., 1992), whereas *Pittosporum* and the representatives of the Araliaceae form an unresolved polychotomy with the Apiaceae. *Griselinia* is basal within this clade and represents one of several taxa formerly associated with the Cornaceae, but here found associated with various groups of Asteridae (see below and Morgan & Soltis, 1993).

(6) **Dipsacales.** This clade matches the order as circumscribed by Cronquist (1981). Within the order, *Sambucus* and *Adoxa* form a clade and *Dipsacus*, *Valeriana*, and *Symphoricarpos* form a clade. *Viburnum* forms a trichotomy with the other two clades. The Caprifoliaceae, as classically circumscribed (represented here by *Symphoricarpos*, *Viburnum*, and *Sambucus*), appear to be a paraphyletic, ancestral grade within the Dipsacales. Downie & Palmer (1992) did not identify a monophyletic Dipsacales; *Viburnum* and *Sambucus* occurred with the Asterales s.l. and the rest of the order with the Apiales (*Adoxa* was not included in their study). *Berzelia* (Bruniaceae, Rosales) occurs as the basal branch of this clade in island-2,784 (Figs. 1B, 2), but is unresolved at the base of a clade comprising the Dipsacales, Asterales s.l., and Apiales in island-102 (Fig. 1A) and in the Astera-

les–Dipsacales–Apiales analysis (Fig. 7). See Donoghue et al. (1992) for further discussion.

(7) **Asterales s.l.** This clade comprises the orders Asterales, Campanulales, and Calycerales, as well as the Menyanthaceae (Solanales), and *Corokia* (Cornaceae). In the analysis of 105 sequences of the Asteridae s.l., the Campanulaceae, Goodeniaceae, and Calyceraceae together form the sister group to the Asteraceae, whereas with greater sampling in the Asterales s.l., only the Goodeniaceae and Calyceraceae form the sister group (in agreement with "Search II" of Chase et al., 1993, and Michaels et al., 1993) with the Campanulaceae forming the basal lineage within the clade (Fig. 7). The very long terminal branch of *Campanula* (Fig. 2), when included in an analysis without the closely related sequence *Lobelia* (e.g., Fig. 1), raises the possibility of artifactual placement due to long-branch attraction (Felsenstein, 1978). When *Lobelia* also is included (Fig. 7), the long-branch problem is lessened. The Asterales s.l. have been the subject of extensive *rbc*L sequence analysis (Kim et al., 1992; Michaels et al., 1993). Michaels et al. (1993) report a detailed analysis of this group in which the Campanulaceae are excluded from both the sister clade of the Asteraceae (bootstrap value—86% and decay index—four steps) and from the clade comprising the Asteraceae and its sister group (84%, four steps). A bootstrap analysis was performed on the 43-sequence data set used in this study with similar results (Fig. 7A). The fact that "Search I" of Chase et al. (1993) included both *Campanula* and *Lobelia* in the sister group to the Asteraceae indicates that "Search I" probably did not reach maximum parsimony for this clade. Anderberg's (1992) morphological analysis identified the Campanulaceae as sister group to the Asteraceae, with the Goodeniaceae–Calyceraceae as the next branch. The association of the Menyanthaceae with the Asterales was not predicted by conventional classifications, but was identified by Olmstead et al. (1992) based on *rbc*L sequences and by Downie & Palmer (1992) based on analysis of cpDNA inverted repeat restriction sites, and has now been accepted by Thorne (1992). See Michaels et al. (1993) for further discussion. *Escallonia* (Grossulariaceae, Rosales) occurs as the basal branch of this clade in island-102 (Fig. 1A), but is unresolved at the base of a large clade comprising the Asterales s.l., Apiales, Dipsacales, and Lamiidae (sensu Takhtajan, 1987) in island-2,784 (Figs. 1B, 2) and island-5,568 (not shown).

(8) **Gentianales.** This clade comprises the Apocynaceae, Asclepiadaceae, Gentianaceae, Loganiaceae, and Rubiaceae. Of recent classifications, Thorne's (1992) is the only one fully in accord with this circumscription. Others have included Oleaceae and Menyanthaceae (in superorder Gentiananae; Dahlgren, 1980; Takhtajan, 1987), or excluded Rubiaceae (Cronquist, 1981). The monophyly of this group is supported by the analysis of restriction site variation in the cpDNA inverted repeat (Downie & Palmer, 1992). Within the clade, the representatives of the Rubiaceae and Apocynaceae s.l. (including the Asclepiadaceae) each form monophyletic groups, as does a group containing representatives of the Gentianaceae and two members of the Loganiaceae (*Anthocleista* and *Fagraea*). The Loganiaceae represent a heterogeneous assemblage, with members in several lineages of the Gentianales (Bremer & Struwe, 1992).

A comparison of the results of the analysis of the Asteridae s.l. data set with those of the Lamiidae data set reveals an interesting and possibly common artifact of analysis when a sequence generating a long terminal branch is included (as with *Campanula* above). In the analysis with fewer Gentianales (Figs. 1, 2), *Gentiana* is basal, with *Gelsemium* next, then *Apocynum*, and *Pentas* and *Spigelia* are terminal. *Pentas* (Rubiaceae) has a very long terminal branch (Figs. 2, 10). In the analysis with 13 Gentianales included (Figs. 5, 6), almost the exact reverse order is obtained with Rubiaceae basal (Rubiaceae is also basal in Downie & Palmer, 1992). A possible explanation for this difference is that in the analysis with fewer sequences, spurious similarity between one of the terminal taxa (*Gentiana*) and the sister group "pulled" that sequence to the base of the clade. Homologous similarity within the clade then resulted in a reverse branching order. In other words, the branching topology within the clade remains essentially the same, but the clade attaches to the rest of the tree at a different point. Adding more sequences along each lineage in the clade (e.g., including *Chiococca* in the Rubiaceae) increases the probability that parsimony analysis will correctly interpret homology, resulting in the correct attachment of the clade to the tree. A local analysis of the Gentianales was conducted to examine internal support within the clade. The 13 representatives of the Gentianales were included with seven outgroups (*Garrya*, *Eucommia*, *Nicotiana*, *Heliotropium*, *Antirrhinum*, *Catalpa*, and *Ligustrum*) and bootstrap and decay analyses were conducted. Seven equally parsimonious trees were found (Fig. 10), with the majority rule consensus tree and bootstrap majority rule consensus trees having the same topology as in Figures 5 and 6. The results support groups representing the Gentianaceae,

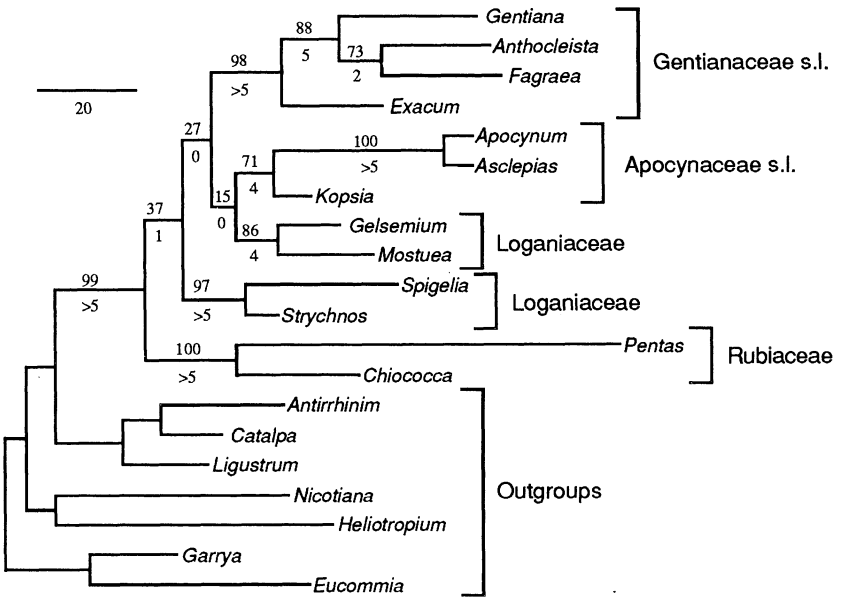

FIGURE 10. Majority-rule consensus of seven equal-length trees of the Gentianales based on *rbc*L sequences and rooted using seven outgroups from the Solanales, Boraginales, Lamiales s.l., and the *Garrya* clade. This tree is equivalent to the topology found in the analysis of the entire Lamiidae data set (Fig. 5) and represents the bootstrap majority-rule consensus. Internal support for groups within the Gentianales is indicated by branch lengths proportional to the scale bar in the upper left and by bootstrap values above and decay indices below internodes subtending clades. A decay index of zero indicates the node collapses in the strict consensus of the seven trees.

Apocynaceae s.l., Rubiaceae, and two separate groups of Loganiaceae (*Gelsemium–Mostuea* and *Spigelia–Strychnos*), but provide little support for any internal groupings.

(9) **Solanales.** This clade is comprised of the Convolvulaceae, Solanaceae, and *Montinia* (Grossulariaceae). Comparable groups exist in the classifications of Dahlgren (1980) and Thorne (1992), except for the inclusion of the Polemoniaceae by Dahlgren (1980). Both Takhtajan (1987) and Thorne (1992) included the Polemoniaceae in a more inclusive group with the Solanales and Boraginales. *Montinia*, not previously suggested to belong to this group on the basis of conventional characters, is basal to the Solanaceae and Convolvulaceae (see below and Morgan & Soltis, 1993). The Solanales and Boraginales form a clade in most analyses, but in island-5,568 (not shown) the Solanales are sister to a clade comprising the Boraginales and Gentianales.

(10) **Boraginales.** This clade comprises the Hydrophyllaceae and Boraginaceae. Groups comparable to this exist in most recent angiosperm treatments (Dahlgren, 1980; Takhtajan, 1987; Thorne, 1992; but not Cronquist, 1981). With the limited sampling currently available, the Boraginaceae appear paraphyletic and the Hydrophyllaceae a monophyletic derivative from them. However, the putatively primitive borages in the woody subfamilies Ehretioideae and Cordioideae have not been sampled for *rbc*L sequences. The Boraginales are sister group to the Solanales in most analyses reported here and in the analyses of Olmstead et al. (1992) and Chase et al. (1993), but are sister group to the Gentianales in island-5,568 (not shown).

(11) **Lamiales s.l.** This clade is comprised of the orders Scrophulariales, Lamiales, and Callitrichales, as well as *Byblis* (Byblidaceae, Rosales) and was recognized by Olmstead et al. (1992) and Downie & Palmer (1992) as monophyletic. This group is recognized by Dahlgren (1980), Takhtajan (1987), and Thorne (1992), but all three treatments maintain a distinction between the Lamiales s.s. and the Scrophulariales. As was pointed out previously (Olmstead et al., 1992), the Scrophulariales are not a monophyletic group. In addition, results presented here (Fig. 5) and results of current work (Wagstaff & Olmstead, unpublished data) suggest that the Lamiales s.s. also may not be monophyletic. Therefore, it seems appropriate not to subdivide this group at present. *Buddleja* and *Nicodemia* have been considered congeneric and placed either in the Buddlejaceae (e.g., Cronquist, 1981) or Loganiaceae (e.g., Mabberley, 1987). In

this analysis, they both clearly belong in the Lamiales s.l. rather than the Gentianales, but the two sequences do not appear to be closely related. Likewise, the Oleaceae have been placed either in the Scrophulariales (e.g., Cronquist, 1981), Gentianales (e.g., Stebbins, 1974), or in their own order, Oleales, allied with the Gentianales (Dahlgren, 1980; Takhtajan, 1987). Based on this *rbc*L analysis the Oleaceae appear to be the basal branch of the Lamiales s.l.

As with the ericalean clade, the strict consensus of the two islands discovered in the analysis of the Lamiidae reveals little resolution within the Lamiales s.l. However, unlike the ericalean clade, where there is little resolution within either island, in this group each island exhibits substantial resolution with particular sequences or small groups occurring in very different places in the two islands. This clade is more densely sampled than most other clades in the Asteridae s.l. Most of the groups that are congruent between islands consist of multiple representatives of large families, including the Acanthaceae, Scrophulariaceae, and Labiatae s.l. The Lamiales s.s. (Labiatae and Verbenaceae) may not be a monophyletic group, as was suggested previously based on a more limited sampling of *rbc*L sequences (Olmstead et al., 1992). Most of the representatives of the Lamiales s.s. form a monophyletic group with no clear distinction between the traditional Labiatae and Verbenaceae. In this respect the clade closely fits the circumscription of the family first suggested by Junell (1934) and recently revived by Cantino et al. (1992). However, *Verbena*, the sole representative of subfamily Verbenoideae in this analysis, is placed closest to *Nicodemia* in a position removed from the rest of the Lamiales in island-135 (Fig. 5) and basal to them in island-1,350 (Fig. 6A). The Lamiales s.l. are the subject of continuing research (Hedrén et al., in prep.; Scotland & Olmstead, in prep.; Wagstaff & Olmstead, unpublished).

These results are fully congruent with those of Olmstead et al. (1992) for relationships among clades within the Asteridae s.l. However, in that study the Dipsacales and Apiales form a clade and together with the Asterales s.l. form another, more inclusive, clade, whereas in this study relationships among the three groups remain unresolved. A bootstrap analysis (Olmstead et al., 1992) suggested moderately strong support (79%) for the Dipsacales–Apiales clade and weaker support for the Asterales–Dipsacales–Apiales (61%). The results of a bootstrap analysis conducted here (Fig. 7) indicate 11% (not indicated on Fig. 7) and 9% values, respectively, for the same two groups. These much lower values reflect the much greater taxon sampling density of this study (see Olmstead et al., 1992, for discussion of this effect). The Asterales–Dipsacales–Apiales clade also is found in both analyses of Chase et al. (1993). Differences in sampling and search strategies may account for the different results obtained in the three studies. A clade comprising the Dipsacales and Apiales minus *Viburnum* and *Sambucus* was found by Downie & Palmer (1992), but their study did not support a united Asterales–Dipsacales–Apiales. No clade comprising any combination of these groups was found by Hufford (1992).

The last four clades, Gentianales, Solanales, Boraginales, and Lamiales s.l., together form the only higher-order clade that is congruent between all three islands. These orders form Takhtajan's (1987) subclass Lamiidae. This group is one of the most strongly supported clades in bootstrap analyses of *rbc*L data (94%—Olmstead et al., 1992; 94%—Morgan & Soltis, 1993) and also was identified in the cpDNA inverted repeat restriction site analysis of Downie & Palmer (1992). Relationships among the four primary clades within the Lamiidae remain unresolved in Olmstead et al. (1992), Downie & Palmer (1992), and in the strict consensus of the 3,900 trees found in "Search II" of Chase et al. (1993). In the analysis of the Asteridae s.l., the Lamiales s.l. are basal in island-2,784 (Figs. 1B, 2) and island-5,568 (not shown), whereas a clade comprising the Solanales and Boraginales is basal in island-102 (Fig. 1A). In the Lamiidae analysis, with greater sampling within the group, the Solanales–Boraginales appear basal in both islands found. However, the strength of the latter conclusion is weakened by the reduced outgroup sampling in that analysis. Morgan & Soltis (1993), including a sequence for the saxifragaceous genus *Vahlia* at the base of the Lamiales s.l., identify the Gentianales as the basal branch within the Lamiidae.

The analyses of Chase et al. (1993) and the inverted repeat restriction site analysis of Downie & Palmer (1992) identify the *Garrya* clade as the sister group to the Lamiidae. That result also was obtained in island-102 and both of the shortest islands resulting from the analysis of the Lamiidae data set. However, island-2,784, island-5,568, and both of the islands one step longer than the shortest in the Lamiidae analysis do not support the sister group relationship between the Lamiidae and the *Garrya* clade.

"Search II" of Chase et al. (1993) identified the small clade of *Dillenia* and *Vitis* as sister group to the Asteridae s.l., but our results suggest that the sister group is a large clade comprising most

of the Rosidae and Dilleniidae (including *Dillenia* and *Vitis*). That this relationship remains uncertain is indicated by the short internode length between the divergence of this small branch from the main lineage of the Rosidae and in adjacent internodes (Fig. 4). Thus, the question of the sister group to the Asteridae s.l. remains unresolved, but the conclusion that this clade originated relatively early in the diversification of the higher dicots (Olmstead et al., 1992) is supported here and in the analyses of Chase et al. (1993).

These results and those of Olmstead et al. (1992) both leave the base of the Asteridae s.l. unresolved. Island-102 in this study identified the Lamiidae and *Garrya* clades as the basal branch, whereas island-2,784 and island-5,568 were unresolved at the base, with either the Cornales or the Cornales–ericalean groups as possible basal branches. The analyses of Chase et al. (1993) identify the Cornales as basal within the Asteridae s.l. and the ericalean clade the next branch. The cpDNA inverted repeat restriction site study of Downie & Palmer (1992) identified the ericalean group as basal, then the Dipsacales–Apiales, Asterales s.l., Cornales, and Lamiidae, but their study had few representatives of the ericalean group (Polemoniaceae and Fouquieriaceae) and fewer outgroups. Hufford (1992) identified three solitary families as individual basal branches (Bruniaceae, Alseuosmiaceae—not included in *rbc*L analyses, and Escalloniaceae), then the Cornales, ericalean group, and the traditional Asteridae. Anderberg (1992) identified the ericalean group as basal, with no resolution among the remaining groups, but his study also suffered from extremely limited outgroup sampling, which compromises any conclusions at this level.

These results represent a detailed, but by no means exhaustive, search for optimal solutions to the *rbc*L phylogeny of the Asteridae s.l. Our two-tiered approach of broadly sampling the entire Asteridae s.l. and sampling more densely within recognized clades offers opportunities to explore patterns of relationship in greater detail at various levels. The discovery of multiple islands in each of the subsets of the data in which a search for islands was conducted indicates that analysis of *rbc*L data (and probably most large complex DNA sequence data sets) must include an appropriate search strategy for finding multiple islands. The presence of multiple islands, which, by their nature, imply different patterns of relationship, indicates that the *rbc*L data by themselves cannot resolve relationships at all levels within the Asteridae s.l. Congruence between analyses at different hierarchical levels is encouraging, although it does not in any sense constitute a test of the results found in one analysis, because the same data form the basis for all analyses. In contrast, incongruence between analyses at different levels (e.g., the inverted topology of the Gentianales in one analysis relative to another) demands an explanation.

In addition to the two-tiered hierarchical analyses reported here for the Asteridae s.l., the analyses of Chase et al. (1993) present a third tier available for comparison (i.e., angiosperms and seed plants). The same 11 primary clades identified here in the Asteridae s.l. are identified in that analysis. However, the pattern of relationships among primary lineages within the Asteridae s.l. implied by the analyses of Chase et al. (1993) is incongruent with each island found in the analysis of the Asteridae s.l. data set. Two important differences exist between these analyses: (1) the analyses of Chase et al. (1993) have greater sampling of ingroup and outgroup sequences in a single analysis, and (2) the smaller data sets used in this study were analyzed more fully for shortest trees and multiple islands. During the search for multiple islands in the Asteridae s.l., each search became "stuck" for a long time on a plateau of trees longer than was ultimately found by that search. In one search, a plateau three steps longer was searched with over 4,600 trees found at that length; the consensus tree was congruent with the results reported by Chase et al. (1993) for the Asteridae s.l. The seed plant analysis was conducted with NNI swapping and a ceiling of 3,900 trees, of which only three were examined with TBR swapping and with MULPARS off. This search strategy is unlikely to enable one to get off a plateau such as those encountered in our searches. This suggests that, as was explicitly admitted in Chase et al. (1993) for their broad analyses in general, their results may represent a suboptimal plateau with respect to the Asteridae s.l.

CONGRUENCE WITH CONVENTIONAL CHARACTERS

The identification of biological traits, other than nucleic acid bases at particular positions in a DNA sequence, which are congruent with the results of a DNA sequencing study, provides potential synapomorphies and important validation for groupings suggested by the molecular analysis. The *rbc*L analysis of the Asteridae offers the opportunity to examine some of the critical characters that have been used in traditional classifications.

The most visible feature commonly associated with the Asteridae is the sympetalous corolla. It

formed the basis for a major group of dicots, the Gamopetalae (de Candolle, 1813), in early classifications. Many families characterized by sympetaly, which have been segregated into different subclasses in many recent classifications (e.g., Cronquist, 1981), are reunited in the *rbc*L tree in the clade designated Asteridae s.l. The *rbc*L tree, however, includes several polypetalous groups, including the Cornales (as broadly defined by Cronquist, with representatives found in several clades), Apiales, Grossulariaceae (*Escallonia* and *Phyllonoma*), Hydrangeaceae, and Pittosporaceae (*Pittosporum*). The distribution of these polypetalous taxa on the *rbc*L tree makes problematic the interpretation of the origin of sympetaly. The Cornales and the *Garrya* clade are the only groups of the 11 discussed above that are consistently polypetalous. Polypetalous groups are basal to otherwise predominantly sympetalous groups in many important clades. For examples of this pattern note *Corokia* with the Asterales s.l., *Vahlia* with the Lamiales s.l. (see Morgan & Soltis, 1993), and *Berzelia* with the Dipsacales. A few sequences represent families in which both sympetaly and polypetaly occur (e.g., *Camellia*—Theaceae, *Berzelia*—Bruniaceae). Some polypetalous taxa (e.g., *Sarracenia* and *Montinia*) are embedded within sympetalous clades (e.g., Ericales and Solanales, respectively). Hufford (1992) considered sympetaly to be a synapomorphy for his Corniflorae–asterid group (comparable to our Asteridae s.l.) exclusive of Bruniaceae (*Berzelia* in our study).

One possible explanation for the observed pattern of corolla fusion is that the Asteridae s.l. originally diversified as a polypetalous clade with subsequent evolution of sympetaly in several lineages. In this scenario, the polypetalous taxa near the base of many clades represent plesiomorphic relicts. However, an alternative hypothesis may be that sympetaly evolved early in the Asteridae s.l. and, during the early diversification of the clade, corolla fusion remained a developmentally labile trait before becoming rigidly canalized (Donoghue, 1989; Olmstead et al., 1992). In this scenario, polypetalous taxa represent the occasional reversion to polypetaly and would be expected to be found primarily near the base of clades. The pattern of corolla development in some groups (e.g., Apiales, Erbar, 1991) suggests that reversal from sympetaly to polypetaly has occurred (Olmstead et al., 1992). This is an example of an area in which morphology deserves renewed investigation as a result of hypotheses derived from a molecular cladogram. The distribution of corolla fusion among the groups and the existence of tree islands make difficult any assessment of corolla evolution. It appears that multiple origins are likely, but that some reversals may have occurred as well.

The presence of tenuinucellate ovules with a single integument has been associated with the Asteridae s.s. (e.g., Cronquist, 1981), but has many exceptions outside the narrow circumscription of the Asteridae. However, the broadly circumscribed Asteridae s.l. encompass the large majority of those exceptions. A single integument (the result of fusion or loss of one integument) characterizes nearly all groups in the Asteridae s.l., except some members of the ericalean clade (e.g., Primulales and some Ebenales). The tenuinucellate condition generally is considered to be derived from the crassinucellate condition in angiosperms and also characterizes most members of the Asteridae s.l.; exceptions include the *Garrya* clade and the Cornaceae and relatives (but not the Hydrangeaceae). The trait "crassinucellate ovule" encompasses a wide range of expression (Davis, 1966; Young & Watson, 1970) and is apparently derived from tenuinucellate ancestors where it occurs in the Asteridae s.l. Hufford (1992) identified single integument as a synapomorphy for his Corniflorae–asterid group and tenuinucellate ovules as a synapomorphy to the same group exclusive of the Bruniaceae.

The distribution of iridoid compounds has influenced heavily the classification of Dahlgren (1980), and the presence/absence of iridoids led Jensen et al. (1975) to reject the traditional Asteridae "as a natural group." Most discussions of the systematic significance of iridoids rely on their evolutionarily derived condition and the implication that the iridoids identify a monophyletic group (e.g., Jensen, 1992). This leads to conflicts with groupings based on other characters. The distribution of iridoids in the Asteridae s.l. suggests that iridoids form a synapomorphy for the entire clade comprising the Asteridae s.l. The distribution of iridoids in the ericalean clade combined with the lack of resolution in that clade makes it difficult to assess the basal state for the clade with respect to iridoids. The pattern of relationships depicted in island-2,784 (Figs. 1B, 2) includes the possibility that the ericalean clade is basal within the Asteridae, thereby leaving open the possibility that if that clade is non-iridoid-containing at its base then iridoids may have evolved twice in the Asteridae s.l. However, it seems that the presence of iridoids is the basal condition in the Asteridae s.l. and the absence of iridoids is best interpreted as losses in several lineages, including the Primulales, Ebenales and relatives (within the ericalean clade), the Apiales, the Campanulaceae (Asterales s.l.), the Asteraceae, and

the Solanales. Additional losses within the Asteridae s.l. are known (e.g., within the Gesneriaceae and Labiatae subfam. Nepetoideae). The presence of simple iridoids in some members of the Hamamelidae (e.g., *Liquidambar*, *Daphniphyllum*), which exhibit morphological, embryological, and *rbc*L evidence incongruent with a placement in the Asteridae s.l., suggests that iridoids have evolved more than once.

The distribution of iridoids in plants has proven to be problematic for systematics, but as phylogenetic hypotheses become more refined, the distribution of iridoids and other secondary compounds may make more sense. For example, Jensen (1992) presented an apparent contradiction in the Oleaceae, which contain both seco-iridoids (grouping them with the Gentianales) and verbascosides (grouping them with the Lamiales s.l.). The *rbc*L tree places the Oleaceae at the base of the Lamiales s.l. clade, suggesting that the seco-iridoids are a plesiomorphic trait shared with the Gentianales, whereas the verbascosides are a synapomorphy uniting them with the Scrophulariales and Lamiales.

PROBLEMS MERITING FURTHER INVESTIGATION

The *rbc*L tree of the Asteridae s.l. suggests several unexpected groups that merit further investigation. For example, *Polemonium* (Polemoniaceae, Asteridae) is grouped (e.g., Fig. 1A) with *Diapensia* (Diapensiaceae, Dilleniidae). These two taxa share many characters, including fused, pentamerous perianths, three fused carpels with numerous ovules, axile placentation, ovules that are unitegmic and tenuicellate, an annular disk at the base of the ovary, and tetrasporangiate anthers opening by longitudinal slits. There are some differences in the secondary compounds reported for the two families, but both are congruent in the absence of iridoids and the presence of calcium oxylate crystals. *Diapensia* differs from *Polemonium* in having a second staminal whorl, a character shared with many members of the Dilleniidae, but in *Diapensia* the whorl opposite the petals is reduced to staminodia. The loss of staminodia in the Diapensiaceae would result in flowers indistinguishable from Polemoniaceae and, based on traditional criteria, assignable to the Asteridae rather than the Dilleniidae (e.g., in the system of Cronquist, 1981). The distribution of the Diapensiaceae is circumboreal at high latitudes, extending into temperate Asia. This suggests that the Polemoniaceae may represent the temperate North American derivatives of the Diapensiaceae.

The clade composed of *Garrya*, *Aucuba*, and *Eucommia* is an unanticipated grouping. However, all three taxa are dioecious and produce the iridoid aucubin. *Garrya* and *Aucuba* have similar pollen and are alike in producing petroselinic acid and in having a single whorl of stamens alternate with the petals (a character shared with the Asteridae s.s. and relatives (e.g., Apiales)). *Eucommia* lacks a corolla, but contains iridoids, has a single integument, and has a gynoecial structure that is virtually identical to that of *Garrya*, a unilocular ovary with two pendulous ovules and two style branches (*Aucuba* also has a unilocular ovule).

In several clades, a single taxon stands apart from the others, with respect to many conventional characters. Two examples are illustrative. *Griselinia* (Cornaceae) is united with the Apiales, but differs in having crassinucellate ovules, iridoids, and a single ovule. *Montinia* is perhaps the most unusual anomaly in this respect. It falls out in the Solanales, from which it differs in many respects (see Morgan & Soltis, 1993). *Griselinia* is basal within its clade, suggesting that its distinctive characters may be either plesiomorphic or autapomorphies not shared with the rest of the clade. *Montinia* appears basal within the Solanales, but embedded well within the Lamiidae, thereby making the case for plesiomorphy more difficult. Cases of anomalous results such as these demand further study, both molecular and morphological, to confirm or reject the placements suggested by these results.

CONCLUSIONS

The results presented here and elsewhere in this issue (e.g., Xiang et al., 1993) represent another step toward a molecular-based phylogeny for the Asteridae s.l. It is clear from the existence of multiple equally parsimonious solutions, some different enough to belong to different islands using a TBR-swapping, heuristic search strategy, that we still have a long way to go to resolve many important questions. Conclusions that appear to be strong, based on the common occurrence of groups in different islands and in results of analyses at different hierarchical levels or of different data sets (Downie & Palmer, 1992; Hufford, 1992; Anderberg, 1992), include the following. (1) There exists a monophyletic Asteridae s.l. containing some elements not formerly assigned to the Asteridae. (2) The Lamiidae (sensu Takhtajan, 1987) are a monophyletic group comprising four separate clades, the Gentianales, Solanales, Boraginales, and Lamiales s.l. (including the Scrophulariales). (3) Eleven ter-

minal clades are congruent among all trees and all islands, including the four comprising the Lamiidae, the Asterales s.l., Dipsacales, Apiales, Cornales, an ericalean clade (including representatives of the Ebenales, Primulales, etc.), and two small clades herein referred to as the *Garrya* clade and *Ilex* clade. (4) With the exception of the Lamiidae, relationships among these primary lineages remain unresolved. (5) There is substantial resolution within many of these 10 clades (e.g., Apiales, Asterales s.l., Cornales, Dipsacales, Solanales), but not for others (e.g., ericalean clade, Gentianales, Lamiales s.l.). (6) Unitegmic and tenuinucellate ovules appear to be synapomorphies for the Asteridae s.l. (7) Sympetaly and the presence of iridoids also may be basal traits for the Asteridae s.l., but additional resolution is needed to distinguish multiple origins from a single origin with subsequent reversals.

Literature Cited

Anderberg, A. A. 1992. The circumscription of the ericales, and their cladistic relationships to other families of "higher" dicotyledons. Syst. Bot. 17: 660–675.

Bremer, B. & L. Struwe. 1992. Phylogeny of the Rubiaceae and the Loganiaceae: Congruence or conflict between morphological and molecular data? Amer. J. Bot. 79: 1171–1184.

Candolle, A. P. de. 1813. Théorie Elémentaire de la Botanique. Paris.

Cantino, P. D., R. M. Harley & S. Wagstaff. 1992. Genera of Labiatae: Status and classification. Pp. 511–522 *in* R. M. Harley & T. Reynolds (editors), Advances in Labiate Science. Royal Botanic Gardens, Kew.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedren, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

Cronquist, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

Dahlgren, R. 1975. The distribution of characters within an angiosperm system. I. Some embryological characters. Bot. Not. 128: 181–197.

———. 1980. A revised system of classification of the angiosperms. J. Linn. Soc., Bot. 80: 91–124.

———. 1983. General aspects of angiosperm evolution and macrosystematics. Nordic J. Bot. 3: 119–149.

Davis, G. L. 1966. Systematic Embryology of the Angiosperms. John Wiley & Sons, New York.

Donoghue, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43: 1137–1156.

——— & J. A. Doyle. 1989. Phylogenetic studies of seed plants and angiosperms based on morphological characters. Pp. 181–193 *in* B. Fernholm, K. Bremer & H. Jornvall (editors), The Hierarchy of Life. Elsevier Science, Amsterdam.

———, R. G. Olmstead, J. F. Smith & J. D. Palmer. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 79: 333–345.

Downie, S. R. & J. D. Palmer. 1992. Restriction site mapping of the chloroplast DNA inverted repeat: A molecular phylogeny of the Asteridae. Ann. Missouri Bot. Gard. 79: 266–283.

Erbar, C. 1991. Sympetaly—A systematic character? Bot. Jahrb. Syst. 112: 417–451.

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

Hufford, L. D. 1992. Rosidae and their relationships to other nonmagnoliid dicotyledons: A phylogenetic analysis using morphological and chemical data. Ann. Missouri Bot. Gard. 79: 218–248.

Jensen, S. R. 1992. Systematic implications of the distribution of iridoids and other chemical compounds in the Loganiaceae and other families of the Asteridae. Ann. Missouri Bot. Gard. 79: 284–302.

———, B. J. Nielsen & R. Dahlgren. 1975. Iridoid compounds, their occurrence and systematic importance in the angiosperms. Bot. Not. 128: 148–180.

Junell, S. 1934. Zur gynaceummorphologie und systematik der Verbenaceen und Labiaten. Symb. Bot. Upsal. 4: 1–219.

Kim, K.-J., R. K. Jansen, R. S. Wallace, H. J. Michaels & J. D. Palmer. 1992. Phylogenetic implications of *rbc*L sequence variation in the Asteraceae. Ann. Missouri Bot. Gard. 79: 428–445.

Kron, K. A. & M. W. Chase. 1993. Systematics of the Ericaceae, Empetraceae, Epacridaceae and related taxa based upon *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 735–741.

Mabberley, D. J. 1987. The Plant-book. Cambridge Univ. Press, Cambridge.

Maddison, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40: 315–328.

———, M. Ruvolo & D. L. Swofford. 1992. Geographic origins of human mitochondrial DNA: Phylogenetic evidence from control region sequences. Syst. Biol. 41: 111–124.

Michaels, H. J., K. M. Scott, R. G. Olmstead, T. Szaro, R. K. Jansen & J. D. Palmer. 1993. Interfamilial relationships of the Asteraceae: Insights from *rbc*L sequence variation. Ann. Missouri Bot. Gard. 80: 742–751.

Morgan, D. R. & D. E. Soltis. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 631–660.

Olmstead, R. G., H. J. Michaels, K. M. Scott & J. D. Palmer. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from

DNA sequences of *rbc*L. Ann. Missouri Bot. Gard. 79: 249–265.

PLUNKETT, G. M., D. E. SOLTIS & P. S. SOLTIS. 1992. Molecular phylogenetic study of Apiales (Apiaceae, Araliaceae, and Pittosporaceae). Amer. J. Bot. 79(suppl.): 158.

STEBBINS, G. L. 1974. Flowering Plants. Evolution Above the Species Level. Belknap Press, Cambridge, Massachusetts.

SWOFFORD, D. L. 1992. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0s. Illinois Natural History Survey, Champaign, Illinois.

——— & G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 *in* D. M. Hillis & C. Moritz (editors), Molecular Systematics. Sinauer, Sunderland, Massachusetts.

TAKHTAJAN, A. 1987. Systema Magnoliophytorum. Nauka, Leningrad. [In Russian.]

THORNE, R. T. 1992. Classification and geography of the flowering plants. Bot. Rev. 58: 225–348.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES & A. C. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science 253: 1503–1507.

WAGENITZ, G. 1977. New aspects of the systematics of Asteridae. Pl. Syst. Evol. Suppl. 1: 375–395.

———. 1992. The Asteridae: Evolution of a concept and its present status. Ann. Missouri Bot. Gard. 79: 209–217.

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

YOUNG, D. J. & L. WATSON. 1970. The classification of dicotyledons: A study of the upper levels of hierarchy. Austral. J. Bot. 18: 387–433.