PHYLOGENETICS OF SEED PLANTS: AN ANALYSIS OF NUCLEOTIDE SEQUENCES FROM THE PLASTID GENE rbcL[1]

Mark W. Chase,[2] Douglas E. Soltis,[3] Richard G. Olmstead,[4] David Morgan,[3] Donald H. Les,[5] Brent D. Mishler,[6] Melvin R. Duvall,[7] Robert A. Price,[8] Harold G. Hills,[2] Yin-Long Qiu,[2] Kathleen A. Kron,[2] Jeffrey H. Rettig,[9] Elena Conti,[10] Jeffrey D. Palmer,[8] James R. Manhart,[9] Kenneth J. Sytsma,[10] Helen J. Michaels,[11] W. John Kress,[12] Kenneth G. Karol,[10] W. Dennis Clark,[13] Mikael Hedrén,[14] Brandon S. Gaut,[7] Robert K. Jansen,[15] Ki-Joong Kim,[15] Charles F. Wimpee,[5] James F. Smith,[12] Glenn R. Furnier,[16] Steven H. Strauss,[17] Qiu-Yun Xiang,[3] Gregory M. Plunkett,[3] Pamela S. Soltis,[3] Susan M. Swensen,[18] Stephen E. Williams,[19] Paul A. Gadek,[20] Christopher J. Quinn,[20] Luis E. Eguiarte,[7] Edward Golenberg,[21] Gerald H. Learn, Jr.,[7] Sean W. Graham,[22] Spencer C. H. Barrett,[22] Selvadurai Dayanandan,[23] and Victor A. Albert[2]

ABSTRACT

We present the results of two exploratory parsimony analyses of DNA sequences from 475 and 499 species of seed plants, respectively, representing all major taxonomic groups. The data are exclusively from the chloroplast gene rbcL, which codes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO or RuBPCase). We used two different state-transformation assumptions resulting in two sets of cladograms: (i) equal-weighting for the 499-taxon analysis; and (ii) a procedure that differentially weights transversions over transitions within characters and codon positions among characters for the 475-taxon analysis. The degree of congruence between these results and other molecular, as well as morphological, cladistic studies indicates that rbcL sequence variation contains historical evidence appropriate for phylogenetic analysis at this taxonomic level of sampling. Because the topologies presented are necessarily approximate and cannot be evaluated adequately for internal support, these results should be assessed from the perspective of their predictive value and used to direct future studies, both molecular and morphological. In both analyses, the three genera of Gnetales are placed together as the sister group of the flowering plants, and the anomalous aquatic Ceratophyllum (Ceratophyllaceae) is sister to all other flowering plants. Several major lineages identified correspond well with at least some recent taxonomic schemes for angiosperms, particularly those of Dahlgren and Thorne. The basalmost clades within the angiosperms are orders of the apparently polyphyletic subclass Magnoliidae sensu Cronquist. The most conspicuous feature of the topology is that the major division is not monocot versus dicot, but rather one correlated with general pollen type: uniaperturate versus triaperturate. The Dilleniidae and Hamamelidae are the only subclasses that are grossly polyphyletic; an examination of the latter is presented as an example of the use of these broad analyses to focus more restricted studies. A broadly circumscribed Rosidae is paraphyletic to Asteridae and Dilleniidae. Subclass Caryophyllidae is monophyletic and derived from within Rosidae in the 475-taxon analysis but is sister to a group composed of broadly delineated Asteridae and Rosidae in the 499-taxon study.

Current assessments of higher-level relationships in seed plants are based largely on informed judgments of the relative value of various reproductive and vegetative characters (including secondary chemistry) and to some extent on historical precedent. Authors of recent taxonomic schemes (for example, Dahlgren, 1980; Takhtajan, 1980, 1987; Cronquist, 1981; Thorne, 1983, 1992) have synthesized an enormous amount of information. Nevertheless, their taxonomic decisions have been guided by estimations of which characters are reliable indicators of relationships. These differing judgments are responsible for radical differences in delimitation and relative ranks of taxa in each system of classification. Recently, a number of explicit, cladistic hypotheses have been developed at inclusive hierarchical levels (e.g., Crane, 1985, 1988; Dahlgren & Bremer, 1985; Dahlgren et al., 1985; Doyle & Donoghue, 1986, 1992; Bremer et al., 1987; Donoghue & Doyle, 1989; Loconte & Stevenson, 1991; Martin & Dowd, 1991; Hamby & Zimmer, 1992; Hufford, 1992; Olmstead et al., 1992; Taylor & Hickey, 1992). Such cladistic studies have previously been limited in scope; some data matrices contain significant taxonomic gaps, and in others characters for some taxa are missing. Both of these factors may have unpredictably misleading effects (Nixon & Davis, 1991; Platnick et al., 1991). Despite a great deal of investigation and analysis, seed-plant phylogenetics is, at best, in a preliminary stage of investigation and knowledge.

Molecular data, specifically DNA sequences, have received a great deal of attention as a potential source of "phylogenetically informative" characters that are putatively less ambiguous than non-molecular characters. Such pronouncements suffer from the limitation that, at higher taxonomic levels, no extensive sampling and phylogenetic description of DNA sequence variation has taken place. The most taxonomically comprehensive analysis of nucleic acid sequences published so far on plants (rRNA; Hamby & Zimmer, 1992) used only 60 taxa, and a number of these were partial sequences. Martin & Dowd (1991), using nucleic acid sequences of the small subunit of RuBisCO (rbcS) inferred from amino acid sequences, studied 335

taxa from 135 families, still less than a third of angiosperm families. If insufficient sampling of taxa or characters (i.e., sequence length, acknowledged as a problem with the rbcS data studied by Martin & Dowd, 1991) are indeed factors, then no valid investigation of the "informativeness" of a given gene sequence exists for seed plants. Thus we are left with only an unfounded assessment of sequence data as having the *potential* to aid in estimating higher-level relationships.

Suggestions that the chloroplast gene rbcL, which codes for the large subunit of ribulose-1,5-bis-phosphate carboxylase/oxygenase (RuBisCO or RuBPCase), was an appropriate locus to use in phylogenetic studies began with Ritland & Clegg (1987) and Zurawski & Clegg (1987). Initial attempts to evaluate relationships used only a dozen or so sequences representing all land plants (Palmer et al., 1988; Giannasi et al., 1992). Other recent studies have been restricted to single families (Doebley et al., 1990; Kim et al., 1992) or putatively closely related families (Soltis et al., 1990; Les et al., 1991; Donoghue et al., 1992; Olmstead et al., 1992; Rettig et al., 1992). Most of the latter studies began the process of incorporating significantly greater sampling to enhance their phylogenetic perspective. The use of rbcL was spurred by G. Zurawski, who generously made available a set of internal sequencing primers. The advent of temperature cyclers and high-temperature-resistant DNA polymerases (sometimes termed Polymerase Chain Reaction or PCR) has greatly enhanced rates at which gene sequence data are accumulating, so that effects on phylogenetic estimates of more intensive sampling of sequence variation can now be investigated.

We wish to examine here the degree to which a representative sampling of sequence variation for rbcL contains evidence of the evolutionary history of seed plants. In this study, we address the quality of evidence present in rbcL sequences for all major seed-plant lineages (roughly 265 families, the exact number depending on the taxonomic scheme followed). To a limited extent, we will compare our phylogenetic hypotheses with recent schemes, but such comparisons are difficult and must be considered heuristic because sister-group relationships ex-

pressed in cladograms are difficult to relate directly to diagrams and statements of progenitor/descendant relationships used in many taxonomic schemes. Dahlgren et al. (1985) and Dahlgren & Bremer (1985) have published analyses most similar to the ones presented here but they are not of a similar scope. Although some specific topological components can be compared to other cladistic studies (Conti et al., 1993; Rodman et al., 1993; Qiu et al., 1993; and several of the other studies in this issue), morphological phylogenetic studies at this level with similar, broad taxonomic sampling do not exist. The computational difficulties of evaluating internal support (e.g., the bootstrap, Felsenstein, 1985; decay analysis, Bremer, 1988) or departure from matrix randomness (Archie, 1989; Faith & Cranston, 1991; Källersjö et al., 1992) for such large numbers of taxa likewise prevent us from addressing extensively these issues here. We do present an analysis of families traditionally referred to Hamamelidae as an example of how internal support for specific clades might be examined. (See also other papers in this issue that address internal and external support for subsets of the general results.) Thus, the broad relationships described here can be used to focus more restricted (with fewer taxa), and therefore more rigorous, investigations.

We expect that patterns presented here will change somewhat as sequences of more species are added or if methodological improvements permit exact solutions (for a discussion of progress, see Penny et al., 1992). These results nonetheless have great value, both from heuristic and methodological perspectives, although the preliminary nature of these studies precludes a detailed examination of implications for seed-plant taxonomy and character evolution, as well as investigations of gene and protein evolution.

We have generally followed the taxonomic circumscriptions of Cronquist (1981) for dicots because this system conforms closely to those used in most textbooks and floras. For monocots, we have adopted the system of Dahlgren et al. (1985) but have changed the superorder ending "-iflorae" used by Dahlgren to the more appropriate "-anae" (Thorne, 1992).

MATERIALS AND METHODS

Selection of taxa for this study was not guided by a specific plan. Close examination of genera included in the analysis will reveal an uneven taxonomic distribution; some groups are well represented (e.g., Asteraceae and Asteridae in general,

Cornaceae, Ericales, Magnoliaceae, Zingiberanae), whereas others are poorly sampled (e.g., dilleniid orders, especially Violales and Theales). Despite lack of a coordinated effort, all subclasses and orders have at least some representatives. All species that are used in this issue of the *Annals of the Missouri Botanical Garden* are listed alphabetically by family in a final Appendix along with other information concerning voucher status, sequence gaps, literature citations for published sequences, and figures of this paper in which each taxon occurs. Some taxa that were included in the 475-taxon data set were excluded from the 499-taxon matrix. Thus, in the second analysis certain families have fewer representatives, but the overall representation of lineages is greater.

Due to the large number of laboratories that contributed unpublished sequences, no standardized procedure was used to produce the sequences analyzed here. A generalization would be the following: a fragment containing *rbc*L was amplified from a total DNA extract using primers that flank or are near the ends of the coding region; this fragment was then directly sequenced using one of several different procedures or was cloned using standard recombinant DNA techniques; dideoxy sequencing generally included both strands for at least $\frac{4}{5}$ of the minimally 1 428-bp gene. Some workers used more closely spaced primers to sequence only one strand of DNA; either strategy appears to provide reasonably error-free sequences. Most extraction protocols relied on fresh or freshly dried leaf samples, but some samples were amplified from DNA extracted from herbarium specimens as old as 20 years.

Amplification of *rbc*L from some taxa produced two different products. Some of these products were different enough in size to observe on an agarose gel, whereas others were detected initially because multiple bands occurred at the same points in autoradiograms, indicating that more than one template was present. Nearly all cycads produced two loci (Hills & Chase, unpublished), which were separately cloned and sequenced to characterize both copies. In cycads examined, one copy contained deletions that disrupted the reading frame, an indication that this copy may represent a "pseudogene." In Convolvulaceae, Olmstead (unpublished) detected two copies of *rbc*L, one copy of which contained deletions. In *Canella* (Canellaceae; Qiu et al., 1993) and *Galphimia* (Malpighiaceae; M. W. Chase, H. G. Hills & W. R. Anderson, unpublished), two size-conserved copies of *rbc*L were also encountered. In *Galphimia*, one copy clearly contained substitutions at numerous sites otherwise conserved among angiosperms, suggesting that this

copy was a pseudogene. In *Canella*, no such un-usual substitutions were observed in either copy. Two, reading-frame-intact copies of *rbc*L are also reported in Ulmaceae (E. Conti & K. J. Sytsma, unpublished). For *Canella*, we included both se-quences in all analyses, but because the two se-quences are always each other's sister we show only the position of "*Canella*" (in fact, there are two terminals here; the complete matrix used in Search II thus has 500 terminals but only 499 taxa).

Sequences of *rbc*L were easily aligned by sight. Among land plants, the coding region contains little size variation through base position 1428 (num-bered from the first nucleotide of methionine in the start codon, AUG). Positions 1426, 1427, and 1428 are the most common stop-codon among land plants; longer reading frames, up to 1452 in some monocots and 1458 in the Asteraceae (Kim et al., 1992), appear to be due to insertions, often of a short repeating sequence. Most laboratories have sequenced past this codon, but for phylogenetic analysis we have terminated all sequences at po-sition 1428 to be confident that we have analyzed homologous regions (the portion of the gene in-cluded in this analysis for each species is also included in the Appendix). All sequences were en-tered into a text file in NEXUS format (used by PAUP 3.0; Swofford, 1991) as complete sequences and then analyzed directly in nucleotide form. Ma-trices used in both searches are available from the first author upon receipt of request and a diskette for each matrix.

TREE-SEARCH STRATEGY

Parsimony-based methods permit direct exam-ination of hypothesized character-state changes on the reconstructed tree, and this information can be used in studies of molecular evolution. (These are, however, likely to be underestimates of se-quence change and could be misleading for this reason.) Numerous empirical studies have shown that not all classes of substitutions are equally like-ly, and this kind of information may be incorpo-rated into weightng schemes for nucleotide data (Swofford & Olsen, 1990; Albert & Mishler, 1992; among others). Various models of molecular evo-lution exist, and appropriately circumspect use of these may assist in the separation of historical signal from homoplasy. The character-state weighting model of Albert et al. (1993, this issue) uses prob-ability formulae to calculate weights for different classes of molecular change. We have used their method in the 475-taxon analysis presented here.

Although the assumptions of the Albert et al. model are admittedly simplistic, we nonetheless support the investigation of weighting approaches to nu-cleotide data and view the model used here as a justifiable first approximation. It must be recog-nized that giving all categories of molecular change equal transformational weight is also an assump-tion, but one that the investigation of Albert et al. (1993) found to be adequate if sampling effects were not a factor. Actual weights applied in this study fall within such a narrow and minimally asymmetrical range that deviation of results from those using the "equal weighting" criterion of Fitch (1971) is likely to be slight (see Albert et al., 1993).

Recently, concern has focused on the probability that "islands" of equally parsimonious trees exist, particularly in large data sets (Maddison, 1991). Because of their enormous size, our analyses do not use methods of multiple random taxon-addition, which have been suggested to uncover such dis-junct, equally optimal islands. This topic is ad-dressed in most empirical papers in this number.

The search for parsimonious trees consisted of several separate but linked heuristic searches using either PAUP 3.0r (**Search I**; 475 taxa) or 3.0s (**Search II**; 499 taxa). All searches included the full data matrix (all codon positions). Search I was performed on a SPARC II (Sun, Inc.) workstation (PAUP 3.0 for non-Macintosh computers is avail-able only by special arrangement with D. Swofford). For Search II, a Macintosh Quadra 800 with 20 MB RAM was used. (Although slower than a Sun computer, the more interactive nature of a search on a PC is preferred by many workers.)

In **Search I**, an initial heuristic search with character-state changes given equal weight (i.e., "unordered" status), SIMPLE data addition se-quence, STEEPEST DESCENT, and NNI (nearest-neighbor interchange) branch swapping algorithm was used to find a single tree (MULPARS option deactivated). The second phase used this single tree obtained from phase one as a starting tree for another heuristic search, this time using the SPR (subtree pruning-regrafting) branch swapping al-gorithm with MULPARS deactivated again. The third phase paralleled the first and second; the single SPR tree from phase two was used as the starting topology for a heuristic search using the TBR (tree bisection-reconnection) branch swapping algorithm, again with MULPARS deactivated.

The fourth phase used the single TBR tree as the starting point for a heuristic search employing a character-state weighting criterion (with a dif-ferent step matrix for each codon position; cf. Swofford, 1991; Albert & Mishler, 1992; Albert et al., 1993). This time MULPARS was activated.

Transversion substitutions were weighted over transitions differentially by codon position (Albert et al., 1993). The specific weights used were the following: for transitions, 5520 (first positions), 6368 (second positions), 4039 (third positions); for transversions, 6620 (first positions), 7470 (second positions), 5127 (third positions). These weights were calculated from empirically derived parameters (see Albert et al., 1993, for a complete description and justification). Because use of step matrices is CPU-intensive, this search was executed with the simplest branch-swapping algorithm, NNI. Additionally, because of dynamic RAM limitations (a tree of 475 terminals occupies a great deal of memory), we restricted our search to a maximum of 500 trees. Although the initial TBR tree was by default a member of a single island, we hoped to provide a bridge to shorter trees through use of the STEEPEST DESCENT option (see Maddison, 1991; Swofford, 1991). This final step yielded 500 equally parsimonious, weighted trees. Because the maximum prespecified number of trees was found, many others probably exist at the same length. Search I required approximately 200 hours to complete.

Search II was performed for three reasons. First, the likely existence of other islands of equal or greater parsimony prompted us to use a strategy that would be more likely to find shorter trees and perhaps a different topology. Second, we were concerned about the effects of the Albert et al. (1993) weighting scheme upon the resulting topology. Third, we wished to examine positions of additional taxa (and make use of updated sequences) that became available after Search I was completed; many of these belonged to previously unrepresented lineages. Differences between these two sets of trees could thus be due to different taxon sampling, corrections or completions of sequences after Search I was finished, or search strategy. We did not intend these two searches to be controlled, direct tests (i.e., with only one variable differing between them); we show them both, rather than simply the one that we judge to be better (Search II), because their similarities, despite variation in taxon composition and search strategy, are considerable. They each represent results of searches that in their own context are worthy of publication, and their differences should be viewed as reasons to be skeptical of both and as cause for future study with more *rbc*L data as well as other characters.

In Search II, we were able to save more trees at the shortest length found: 3,900 rather than only 500 trees. The initial starting tree was pro-

duced by using the CLOSEST addition sequence with the HOLD option set for five trees (this in effect permitted initial swapping on several different starting topologies). Approximately 120 hours were required merely to add all taxa in this manner. The initial search (i) used NNI swapping and STEEPEST DESCENT with MULPARS off. The shortest single tree found was then swapped on using (ii) TBR, which generally found a shorter tree, at which time (iii) NNI (with MULPARS) was used. When use of MULPARS resulted in 3,900 trees, which used up available RAM, a single tree was randomly selected (iv) to swap on with TBR (MULPARS off). If this resulted in a shorter tree being found, the search was then stopped and restarted (iii, again) using NNI and MULPARS and this shorter tree as starting point (iii and iv were repeated until no shorter trees were found). The shortest tree length found with this method was 16,305. Three randomly selected trees from the 3,900 saved at this length were swapped in succession to completion with TBR (no MULPARS), and no shorter trees were found. A strict consensus tree was computed, and branch lengths for one randomly selected tree were calculated using the ACCTRAN optimization. Search II thus used no relative weighting; it required approximately four weeks to complete.

CAVEATS

Methodologically, these searches suffer from (i) uncertainty about maximum parsimony, (ii) unquestionable absence of many trees at the same level of optimality, (iii) identification of only a single topology, and (iv), in Search I, incomplete branch swapping on any of the shortest trees found. We would never recommend these search strategies for smaller data sets, but several options were seriously restricted by the number of taxa included (these are reputedly the largest PAUP analyses attempted to date). Specific sections of the general topologies, when analyzed in a more "localized" manner, provide different sets of relationships (see Michaels et al., 1993; Morgan & Soltis, 1993; both this issue). The broader taxon distribution of the general analysis (thus with far greater outgroup information) may be assessing character-state change on the immediate branch leading to a specific ingroup differently than in more restricted analyses. Greater outgroup information could "improve" the ingroup analysis or add spuriousness to it; it is generally best to implement tree searches both with and without outgroups to examine their effects. If more restricted analyses differ from those pre-

sented here, we would certainly favor the former because of increased confidence in finding parsimonious solutions. We nonetheless view these analyses as instructive about seed-plant relationships and the utility and limits of rbcL information (see Discussion).

The major limiting factor in our studies is clearly matrix size. When confronted with the largest molecular database relevant to seed plants, computational trade-offs inevitably arose. The amount of time spent on these approximations may not be directly proportional to the time taken to generate all the rbcL sequences, but no method exists to predict how many trees at how many steps shorter might have been found after additional months or even years of continuous computation. We have opted for the approximations presented here rather than commit ourselves to an open-ended experiment.

Potential effects of errors in autoradiogram reading and data entry should be considered. We detected a number of internal stop-codons in sequences used in these analyses, and other workers have reported errors of various kinds. Most of these were corrected in the matrix used in Search II. Certainly the potential for errors is present, but the effects of such mistakes should not be extensive because they are likely to be random.

Identification of some taxa in this analysis has been questioned due to peculiarity of placement. One example of this has been now identified among the species used in Search I. We have not expunged it from the illustrations; the results represent the outcome of real tree searches and are instructive for that reason. The sequence analyzed was an actual member of the clade into which it was placed, but it was not the species to which it had been attributed. The material sent by a botanical garden and labeled as *Kirengeshoma* was evidently misidentified. It was almost certainly a member of the *Parnassia* group (Saxifragaceae sensu lato; the position it occupies in Fig. 11A) rather than a member of Hydrangeaceae. Sequencing of another specimen of *Kirengeshoma* and subsequent data analysis revealed the expected placement of this genus with other Hydrangeaceae (Xiang & Soltis, unpublished). Other "surprising" results also have been checked by obtaining another sample of the taxon in question and re-sequencing rbcL. The original sequence of purportedly saxifragaceous *Montinia* (which nested in Solanales) was checked and found to be accurate. Still others (such as *Sargentodoxa* among Fabaceae) are still being reassessed. Most workers have tried to make vouchers for each species in this study (see Appendix). Many tissue samples were provided by botanical gardens, and, if vouchers are not included with samples, investigators are dependent upon identification of these plants by the respective organizations (see Goldblatt et al., 1992; it is critical that a voucher sample taken from the same plant used for DNA extraction be included at the time of collection; accession numbers or vouchers taken previously are subject to later events, such as lost labels or collector mistakes). Because a number of the samples used in these two analyses are unvouchered (see Appendix), reproducibility is compromised. In cases with multiple species from a family, we should be able to recognize grossly misidentified samples (but this is true for only 37% of the 265 families represented).

RESULTS

For display purposes, we show here the combinable component consensus (Bremer, 1990) of the 500 equally parsimonious trees found in Search I (Figs. 1–15, A series). Because these are character-state weighted trees, tree lengths and tree statistics (e.g., consistency index, etc.) are not comparable to those of Fitch trees and are not given here. For Fitch trees found in Search II, the length was 16,305 with a consistency index (C.I., excluding unique substitutions and constant characters) of 0.102 and a retention index (R.I.) of 0.632. The branch lengths, again with ACCTRAN optimization, are shown on one of the 3,900 trees selected at random (Figs. 1–15, series B). Branches that collapse in the strict consensus tree are indicated by arrows on the B series (Figs. 3–15B).

First we will summarize the topology found in Search I ("A" series of figures). The results of Search II ("B" series) have been interdigitated with those of Search I to facilitate comparisons. After describing the results of Search I, we briefly examine major differences between the two results. Note that all figures in the A series are from the combinable component consensus tree, whereas in the B series, Figures 1B and 2B are the strict consensus tree (branch lengths in Fig. 2B, however, were taken from a single tree) and the remainder (Figs. 3B–15B) are a single tree randomly selected from 3,900 equally parsimonious Fitch trees (this may be confusing; for example, whereas the polytomy among monosulcate clades in Figs. 1B and 2B is due to variation among the 3,900 trees for branches in this portion, the topology shown in 3B is resolved because it is resolved in the single tree selected). To indicate branches of the B series that are absent in the strict consensus tree of the 3,900,

we have placed arrows to their right. The branch lengths presented in the B series should under no circumstances be interpreted as meaningful measures of support; thus, in the example of Hamamelidae provided, branches that decay at one step less parsimonious have lengths that range from 2 to 16 steps, whereas those that decay at four steps less parsimonious range from 7 to 14 steps (Fig. 16). The sole reason for providing branch lengths in the B series is to permit readers to estimate roughly relative degrees of divergence and to identify cases in which long terminal branches are connected to short internal branches (a situation in which adding related taxa often radically alters hypothesized relationships).

Search I. Unless stated otherwise, we have used the taxonomic circumscriptions of Cronquist (1981) for dicots and Dahlgren et al. (1985) for monocots, although we acknowledge that other recent systems fit these results better. We arranged the unrooted trees of both searches with cycads sister to all other seed plants (Figs. 1–3) in accord with recent results of several non-molecular cladistic studies (Crane, 1985, 1988; Doyle & Donoghue, 1986, 1992). In Search I, conifers are paraphyletic, but some trees (not shown) found in Search II have a monophyletic conifer lineage; the strict consensus tree from Search II is unresolved regarding conifers (Figs. 1B, 2B; we have cited in this section the B series of figures along with the A series if they include the same general set of taxa). The three genera of Gnetales, *Ephedra*, *Welwitschia*, and *Gnetum*, are highly divergent from all other seed plants but were nonetheless identified as sister of the angiosperms (Fig. 3A, B), within which *Ceratophyllum* (Ceratophyllaceae) alone is sister to and highly divergent from the rest (Fig. 4A, B). The major feature of flowering plants (exclusive of *Ceratophyllum*) is their separation into two major groups; these correspond well with distributions of the two major angiospermous pollen types, uniaperturate (monosulcate and monosulcate-derived) and triaperturate (tricolpate and tricolpate-derived). *Ceratophyllum* has inaperturate pollen (Cronquist, 1981). The major exception to this split is the presence of tricolpate pollen in Illiciaceae and Schisandraceae, which fall among the monosulcate taxa (Fig. 4A, B; see Qiu et al., 1993, this issue). No morphological support for monophyly of the monosulcate clade has been recognized in the literature (their pollen type exists among nonflowering seed plants and thus must be considered plesiomorphic).

Three monophyletic lineages within uniaperturate magnoliids were identified, and these correspond closely, although not exactly, to (i) Mag-

noliales, (ii) Laurales, and (iii) "paleoherbs" (here defined as composed of Aristolochiales, Piperales, and Nymphaeales; Fig. 4A). Monocots (also with uniaperturate pollen and sometimes included in "paleoherbs"; Donoghue & Doyle, 1989) represent a fourth member of this clade. Among the paleoherbs are also nested several problematic families: Illiciaceae, Schisandraceae (both Illiciales), Amborellaceae (Laurales), and Austrobaileyaceae (Magnoliales). Chloranthaceae are also allied cladistically with the paleoherbs, but *Chloranthus* does not form a monophyletic group with other Piperales.

Monocots are a well supported monophyletic group (see Duvall et al., 1993, and Qiu et al., 1993, both this issue) and are derived from within monosulcate Magnoliidae; the paleoherbs are their immediate sister group (Fig. 4A). Within monocots (Fig. 5A, B), Aranae plus *Pleea* (of polyphyletic Melanthiaceae; Lilianae) are basal-most, followed by Alismatanae plus *Burmannia* (Burmanniaceae) and *Aletris* (Melanthiaceae). Lilianae form a paraphyletic series of three lineages that correspond well to Dioscoreales, Liliales, and Asparagales of Dahlgren et al. (1985), except in the placement of certain families (Iridaceae, Orchidaceae, and Smilacaceae) and genera (*Chamaelirium* of Melanthiaceae: Melanthiales). *Vellozia* (Velloziaceae: Bromelianae), *Freycinetia* (Pandanaceae: Pandananae), and *Sphaeradenia* (Cyclanthaceae: Cyclanthanae) together form a monophyletic clade that is collectively sister of the Liliales. The "commelinoid" group of monocots (Fig. 6A, B) incorporates all of those that Harris & Hartley (1980) found to exhibit fluorescing cell-wall phenolics. In both searches, this commelinoid clade includes monophyletic Arecanae and Zingiberanae and polyphyletic Bromelianae and Commelinanae (Fig. 6A, B). Cyclanthaceae have the same phenolic biosynthetic pathway but do not accumulate end products; they are not members of the commelinoid assemblage (Clark et al., in prep.).

The two orders of Magnoliidae with triaperturate pollen, Ranunculales and Papaverales, form a clade that is sister to the rest of "eudicots" (Fig. 7A, B). The term "eudicot" has been variously defined in the literature, but we use it here to refer to all angiosperms with triaperturate or triaperturate-derived pollen (Donoghue & Doyle, 1989; Doyle & Holton, 1991). This is one of the best supported clades among angiosperms (Qiu et al., 1993, this issue). Two other basal clades within the eudicots (Fig. 7A, B) consist of some Hamamelidae (Trochodendraceae and Tetracentraceae) and Platanaceae, Sabiaceae, Nelumbonaceae, and Proteaceae.

Within eudicots, two large sister clades are iden-

tified, one that corresponds roughly to Asteridae and the other to Rosidae (Figs. 1, 2). Membership in both lineages is considerably expanded with respect to their circumscription by Cronquist (1981), although less so with respect to the circumscriptions of Dahlgren & Bremer (1985) and Thorne (1992). These two major clades (Fig. 2A, B) reflect the division of eudicots into two major groups (Young & Watson, 1970): sympetalous/tenuinucellate and polypetalous/not tenuinucellate (asterids and rosids, respectively, in our figures). The "crassinucellate" condition actually consists of several different states. Exceptions to this generalization exist, but these traits appear much less homoplastic here than in most systems of classification. The basalmost lineage within Rosidae includes Saxifragaceae sensu stricto and Crassulaceae besides lower hamamelids such as *Cercidiphyllum* and *Hamamelis* (Fig. 8A; see Morgan & Soltis, 1993, this issue, for a discussion of Saxifragaceae sensu lato). The next-most-basal group contains Caryophyllidae (including Plumbaginaceae and Polygonaceae; Rettig et al., 1992) plus Droseraceae, Nepenthaceae (Nepenthales, Dilleniidae; see Albert et al., 1992b), Dilleniaceae (Dilleniidae), and Vitaceae (Fig. 9A).

The remaining Rosidae are split into two large sister groups (Figs. 2A, 10A, 11A). In one (rosid I) are several families of higher Hamamelidae, Euphorbiales, Fabales, Linales, Polygalales, and Rosales (Fig. 11A). Other members of this clade include a number of dilleniid families: Ochnaceae (Theales), Datiscaceae, Passifloraceae, and Violaceae (all Violales). Ordinal boundaries of this group of Rosidae (sensu Cronquist, 1981) are largely unsupported; this assemblage is particularly heterogeneous. The largest polytomy in the consensus tree from Search I occurs at the base of this group, and sampling of the families that potentially belong to this clade is the most sparse in this analysis. The other major lineage of Rosidae (rosid II; Fig. 10A) includes orders Myrtales (see Conti et al., 1993, this issue) and Sapindales, for which ordinal boundaries are reasonably intact. Malvaceae (Malvales; Dilleniidae) and all but one of the mustard-oil families (those in Capparales plus others in Dilleniidae; see Rodman et al., 1993, this issue) are associated with Sapindales. Geraniaceae are also members of this clade, although other members of Geraniales appear elsewhere (Oxalidaceae with a group of families in Rosales, Fig. 11A; Balsaminaceae with Ebenales, Fig. 13A, B; and mustard-oil-producing Limnanthaceae and Tropaeolaceae with Capparales, Fig. 10A, B; Price & Palmer, 1993, this issue). Two members of Rosales, *Greyia* (Greyiaceae) and *Francoa* (Saxifragaceae), appear derived from within Geraniaceae, if *Viviania* is

included in Geraniaceae (see also Price & Palmer, 1993, and Morgan & Soltis, 1993, this issue).

The tenuinucellate/sympetalous clade that terminates in Asteridae sensu Cronquist (1981) was also identified by Olmstead et al. (1992) in their efforts to circumscribe subclass Asteridae using *rbc*L sequences. This study greatly expands upon their sampling and identifies as members of the asterid clade two lineages of often polypetalous Rosidae (Figs. 2A, 12A): (i) Santalales plus Paeoniaceae and Gunneraceae and (ii) some families of Cornales plus Hydrangeaceae (see Xiang et al., 1993, this issue). The sister group of Asteridae is a clade (Figs. 2A, B, 13A, B; see Kron & Chase, 1993, this issue) that contains the dilleniid orders Ebenales, Ericales, Primulales, Diapensiales, plus some members of Theales (Actinidiaceae and Theaceae). Sarraceniaceae (Nepenthales; Dilleniidae) and *Roridula* (but not *Byblis* of Byblidaceae of Rosales: Rosidae) are also members of this lineage (Albert et al., 1992b). Polemoniaceae (Solanales: Asteridae) and Balsaminaceae (Geraniales: Rosidae) also belong to this ericalean/ebenalean group. (See also Olmstead et al., 1993, this issue, for a treatment of the Asteridae sensu lato.)

Asteridae sensu Cronquist split into two major sister groups. In one of these (asterid II; Fig. 14A, B) are families of Asterales, Calycerales, Campanulales, Dipsacales, and some Solanales (Menyanthaceae). Rosid taxa that are members of this clade include Apiales, Aquifoliaceae (Celastrales), some Cornaceae (Cornales), Pittosporaceae and Grossulariaceae (both Rosales). In the other major clade (Fig. 15A, B) fall orders Callitrichales, Gentianales, Lamiales, Rubiales, Scrophulariales, and most Solanales, although these ordinal limits are not always supported (see Olmstead et al., 1993, this issue). In this clade (Fig. 15A, B) is a group that includes rosids *Aucuba* (Cornaceae: Cornales) and *Garrya* (Garryaceae: Cornales) and hamamelid *Eucommia* (Eucommiaceae: Eucommiales), all of which accumulate aucubin (Cronquist, 1981) and share distinctive anatomical wood characteristics (E. Wheeler, pers. comm.). Thus the suite of floral characteristics that have been interpreted as support for the monophyly of Asteridae sensu Cronquist (1981) appears either (i) to have twice arisen independently from ancestors with rosalean and cornalean floral traits or (ii) to have undergone reversals in groups traditionally included in Rosidae (sensu Cronquist; Donoghue et al., 1992; Olmstead et al., 1992, 1993, this issue).

Search II. A number of taxa from Search I were removed from Search II (all marked with a "†" in the A series of figures) to accommodate the representatives of additional lineages in Search II.

Most of the removed taxa were from monophyletic families in which six or more species were present in Search I (for example, Asteraceae, Ericaceae sensu lato, Magnoliaceae, and Poaceae). Two others, *Burmannia* (Burmanniaceae) and *Hydrolea* (Hydrophyllaceae), were removed because they are highly sequence divergent from all other taxa; in separate, smaller analyses, these two appear to be involved in "branch attractions" and attach in radically different positions as other taxa are added or removed (other similarly divergent genera, for example, *Paeonia* and *Gunnera*, do not cause these problems and were kept; we admit to being relatively arbitrary in removing only these two taxa). *Burmannia* presents an interesting case. Members of Burmanniaceae are often achlorophyllous; in spite of being green, *Burmannia* may still derive a great deal of its nutrition through its mycorrhizal associate. In such cases in other families, a number of protein loci exhibit higher rates of sequence divergence (as measured by relative branch lengths when compared to completely autotrophic members of their lineages; C. dePamphilis, pers. comm.). Thus the high levels of sequence divergence, which make it difficult to place these taxa accurately, are a product of or at least associated with their partial heterotrophy.

For Search II, additional taxa were available (these species are marked with asterisks in the B series of figures). The placements of these are described first, and then different arrangements of taxa included in both searches are identified (the clades involved with major shifts of position are marked with an "§" in Fig. 2B). This last section describes only the major shifts of position, but many "minor" shifts also occur within clades (for example, 12 "minor" shifts take place among the taxa in Fig. 15, a clade in which only a few new species were added). What constitutes a "major" versus a "minor" alteration is, of course, a matter of personal perspective. We would therefore advise readers to examine carefully the trees from both searches for taxa of specific interest, which is one reason we presented and intercalated the results of both analyses.

Positions of additional taxa. *Ginkgo* (Ginkgoaceae) intercalates between cycads and conifers (Figs. 1–3B). *Taxus* (Taxaceae), *Cephalotaxus* (Cephalotaxaceae), and *Sciadopitys* (Taxodiaceae or in its own family) are members of the non-Pinaceae clade of conifers, with *Sciadopitys* sister to the rest of that clade (Fig. 3B). Within magnoliid angiosperms, *Lactoris* (Lactoridaceae) is sister to *Aristolochia* (Aristolochiaceae). Other new mag-

noliid species added in Search II represent additional members of families already present in Search I, and all of these form monophyletic units with other members of their respective families (Fig. 4B).

Sister to all other monocots is *Acorus* (Araceae; see Duvall et al., 1993, this issue). *Calochortus* (Calochortaceae) is sister to Liliaceae, *Hemerocallis* (Hemerocallidaceae) is sister to *Chlorophytum* (Anthericaceae), and *Lomandra* (Dasypogonaceae) is a member of the often arborescent clade composed of Agavaceae, Asphodelaceae, and Xanthorrhoeaceae within Lilianae (Fig. 5B). Additional members of families in Search I are all placed as sister taxa to their respective family representatives. Among the commelinoid clade identified in Search I (Fig. 6A, B) are the following additional families: *Sparganium* (Sparganiaceae) sister to *Typha* (Typhaceae) and *Lachnocaulon* (Eriocaulaceae) among the graminoids.

Among eudicots (Fig. 7A, B), *Pachysandra* (Buxaceae) is sister to Trochodendrales (*Trochodendron* and *Tetracentron*). Composition of the various rosid clades is somewhat different in Search II (Fig. 2B; see below), but the following additional taxa are placed roughly among the rosid II (Fig. 10B) clade identified in Search I: *Shorea* (Dipterocarpaceae), *Theobroma* (Sterculiaceae), *Tilia* (Tiliaceae), and *Bombax* (Bombacaeae) are members of a malvalean clade (represented only by *Gossypium* in Search I). *Akania* (Akaniaceae) and *Bretschneidera* (Bretschneideraceae), both Sapindales, are members of the mustard-oil clade. Three additional members of Capparaceae, *Cleome*, *Koeberlinia*, and *Setchellanthus*, along with *Capparis*, create a polyphyletic Capparaceae (Fig. 10B; see also Rodman et al., 1993, this issue). *Hypseocharis* (Oxalidaceae) is sister to a clade containing many Geraniaceae (Fig. 10B; see Price & Palmer, 1993, this issue). Among members of rosid I (Fig. 11B, C) are representatives of the following new families: *Reinwardtia* (Linaceae) is sister to *Viola* (Violaceae); *Sargentodoxa* (Sargentodoxaceae) is imbedded in Fabaceae; *Humulus* (Cannabinaceae), *Trema* (Ulmaceae), and *Boehmeria* (Urticaceae) fall into an urticalean clade with *Morus* and *Ficus* (both Moraceae), but Ulmaceae are paraphyletic to Cannabaceae; *Coriaria* (Coriariacaeae), *Begonia* (Begoniaceae), and three genera of Cucurbitaceae are placed with Datiscaceae; *Betula* (Betulaceae) and *Carya* (Juglandaceae) are members of the clade containing Fagaceae and Casuarinaceae; *Mouriri* and *Osbeckia* (Melastomataceae), *Punica* (Punicaceae), *Trapa* (Trapa-

ceae), and *Heteropyxis* (Myrtaceae) are members of a clade of (mostly Myrtales) that includes Onagraceae and Combretaceae (Fig. 11C). In asterid lineages, two additional families have been added: *Byblis* (Byblidaceae) and *Vahlia* (Saxifragaceae) are placed among Schrophulariales (Fig. 15B).

Changes in placements of major clades. At higher levels among the seed plants, Search II produced relatively few "major" differences in taxon placements from the topology of Search I. The conifers in some, but not all, 3,900 trees are monophyletic. Piperales–Aristolochiales were sister to all monosulcate angiosperms (Figs. 2B, 4B). The "paleoherb" clade of Search I (Fig. 4A) was thus split in two with the portion containing Chloranthaceae, Illiciales, Austrobaileyaceae, Amborellaceae, and Nymphaeales situated as sister to Magnoliales. Although these may seem to represent major shifts, extremely short branches separate these clades (Fig. 4B); thus neither topology has much internal support (Qiu et al., 1993, this issue). Within monocots, only shifts among the groups in Aranae, Alismatanae, and Lilianae occurred (Fig. 5B): Aranae (minus *Acorus*) are sister to Alismatanae; the clade containing Pandanaceae, Cyclanthaceae, and Velloziaceae is isolated and no longer sister to Liliales, and Dioscoreales are sister to the commelinoid taxa (the position of Asparagales in the trees from Search I). Among commelinoid taxa (Fig. 6B), *Stegolepis* (Rapateaceae) is sister to Bromeliaceae, Typhales are sister to Juncaceae–Cyperaceae, and *Flagellaria* is sister to the other graminoid clade.

Among eudicots, three major shifts of taxa present in Search I occurred. The first series of rearrangements involves the heterogeneous asterid V clade (Fig. 12A). *Gunnera* (Gunneraceae) intercalates as an isolated lineage (Fig. 7B) between Trochodendrales (plus *Pachysandra* of Buxaceae) and higher eudicots (asterids, caryophyllids, and rosids). Santalales, *Phoradendron* (Viscaceae), *Schoepfia* (Olacaceae), and *Osyris* (Santalaceae) become sister to Caryophyllidae–Droseraceae–Nepenthaceae (Fig. 9B), and this larger caryophyllid clade is shifted from a position within the rosid clade (rosid III; Figs. 2A, 9A) to sister to the larger asterid–rosid clade (Figs. 2B, 9B). The remaining member of asterid V, *Paeonia* (Paeoniaceae), is deeply imbedded within rosid III clade (Fig. 8B) as sister to Crassulaceae. We attach little significance to these shifts of position; internal branches of these groups are among the shortest supporting positions of major clades.

The second series of shifts occurs within rosids.

Families of Myrtales (plus *Qualea* of Vochysiaceae, Polygalales) and the clade containing *Viviania*, *Wendtia* (both Geraniaceae), *Greyia* (Greyiaceae), and *Francoa* (Saxifragaceae) are shifted from rosid II to rosid I (Fig. 11B, C), and relationships of intermediate-level clades (containing several families, i.e., those of Urticales, Fagales, etc.) are somewhat modified from their position in Search I.

The third shift involves the two taxa that were sister to the expanded caryophyllid clade from Search I (Fig. 9A): *Dillenia* (Dilleniaceae) and *Vitis* (Vitaceae), which become sister to the rest of the larger asterid clade (Fig. 12B; *Vitis* and *Dillenia* in a sense exchanged positions with Santalales). Again, these groups have short internal branches, and these shifts would require little change of overall parsimony.

In several instances, the topology of Search II (which we favor because it was a more complete search) is more similar to that found in investigations of restricted nature (for example, outgroup relationships of Asteraceae are identical with those found by Michaels et al., 1993, this issue, and Olmstead et al., 1992, whereas those of Search I deviated in several ways).

DISCUSSION

Although *rbc*L sequences for several groups of spore-bearing plants are available (true mosses, hornworts, liverworts, *Equisetum*, *Isoetes*, *Lycopodium*, *Psilotum*, and both eu- and leptosporangiate ferns), their use as outgroups is complicated by extensive sequence divergence relative to that in the seed-plant ingroup. No other extant lineages of land plants are likely to have shared a common ancestor with seed plants for well over 350 million years, and a great deal of sequence change, much of it in the form of multiple, unrecoverable substitutions, has occurred. Analysis of these other land plant sequences produces topologies (not shown, but see Hamby & Zimmer, 1992, for similar results) that are radically different from all previous hypotheses of relationships (e.g., Bremer et al., 1987). In contrast, seed-plant relationships presented here are at least congruent in gross aspect with comparable morphological studies (Crane, 1985, 1988; Doyle & Donoghue, 1986, 1992; Loconte & Stevenson, 1991). Addition of highly sequence-divergent outgroups could be expected to increase ingroup homoplasy with unpredictable topological results (Felsenstein, 1978). In these analyses, we have chosen to use the more conservative approach of an unrooted ingroup analysis

of seed plants, which are almost certainly mono-phyletic (Doyle & Donoghue, 1986, 1992).

Likewise, effects of missing groups upon topol-ogies may be profound (Donoghue et al., 1989) and unpredictable, and we expect the absence of the numerous extinct lineages of early land plants in molecular data matrices to pose potentially se-rious problems for elucidating relationships of ex-tant lineages. Perhaps more conservative genes sampled for more of the sequence variation present within extant groups may be able to "bridge" gaps caused by extinction. Features of genome organi-zation (such as gene and intron content and gene order; Downie & Palmer, 1992; Raubeson & Jan-sen, 1992) may offer more robust hypotheses than gene sequences for such questions, but these are likely to be too few to provide a fully resolved tree by themselves. A great deal more experimentation with combined data sets of morphological and mo-lecular characters is obviously needed, and a num-ber of these studies are underway (both with rbcL and rRNA/rDNA data). Consideration of these problems here is premature.

Sister-group status of angiosperms and Gnetales (Figs. 2A, B and 3A, B) is corroborated by other cladistic studies (Crane, 1985, 1988; Doyle & Donoghue, 1986, 1992; Loconte & Stevenson, 1991). The isolated position of Ceratophyllum as sister to all other angiosperms has been argued previously (Les, 1988; Les et al., 1991). In studies by Qiu et al. (1993, this issue) in which non-flowering seed plants were removed, this arrange-ment was made equivocal by the existence of an-other equally parsimonious island in which Cera-tophyllum occurred in a radically different position. Hamby & Zimmer (1992; rRNA) also found a yet different placement for Ceratophyllum (but we suspect that the sparser sampling of their study may be responsible for most of the differences from those found with rbcL). In instances in which a taxon's morphology and anatomy are as divergent and potentially modified as those of Ceratophyl-lum, its position becomes difficult to address ade-quately in cladistic studies. Ceratophyllum has been absent from many morphological cladistic studies, such as those of Doyle & Donoghue (1986, 1992), so corroboration is currently precluded.

The general groupings of angiosperms (exclusive of Ceratophyllum) identified in these two analyses are highly similar to each other and to those of most recent taxonomic schemes, particularly those of Dahlgren (1980), Dahlgren et al. (1985), and Thorne (1992; this last has admittedly incorporated results of several molecular investigations). Fur-thermore, results of rbcS (Martin & Dowd, 1991)

and ribosomal studies (Hamby & Zimmer, 1992) are quite similar to ours as well. How these larger groupings (clusters of families and in some cases orders) are inter-related is the point at which their similarity diverges. Although having quite different implications for angiosperm origins and evolution, the preferred hypothesis in one of these studies is not vastly different in relative parsimony from those favored in other investigations. For example, with mostly morphological data, Doyle & Donoghue (1986) discovered a "paleoherb rooting" at one step less parsimonious. Constraining a paleoherb rooting for angiosperms, at Nymphaeales (as in Hamby & Zimmer, 1992), was also only slightly less parsimonious in the subset of angiosperm rbcL sequences studied by Qiu et al. (1993, this issue). When examined to address basal angiosperm re-lationships, all these data appear to lack a strong historical signal.

A number of phenomena have been suggested to be capable of confounding molecular phyloge-netic studies. We consider below several of these factors and examine their potential to affect studies of rbcL sequence variation and then address some additional concerns about future directions of mo-lecular systematic study.

EFFECTS OF PARALLEL NUCLEOTIDE
SUBSTITUTIONS IN INDEPENDENT LINEAGES

An effect of unrecoverable (due to extinction) or unobserved (due to insufficient sampling) char-acter-state changes is the introduction of spurious similarities, which may result in treatment of in-dependently derived nucleotides at a given base position as homologous (see Albert & Mishler, 1992). Such mistaken interpretations of indepen-dent events can lead to "branch attractions" if an analysis includes a great number of such assess-ments (Felsenstein, 1978). Adequate taxon sam-pling, sometimes referred to as appropriate "taxon density," is one means of reducing potentially in-accurate assessments of similarity, but determining at what point sampling is sufficient has so far only been addressed in an a posteriori manner.

The improvement afforded to assessments of character-state change by increased taxon sam-pling is counterbalanced by a decrease in com-putational speed and ability to ascertain how near results are to maximum parsimony. Intrafamilial studies are not as likely to be affected severely by these problems because, in general, numbers of taxa are not as great and evenness of sampling is better. At higher taxonomic levels within seed plants, a paradoxical impediment to progress arises: if taxon number is great enough to assess character-

state changes accurately, then one reasonably can expect to reach only suboptimal phylogenetic solutions, but if taxon number is restricted enough to gain confidence of the maximum parsimony of the trees found, spurious assessments of character-state identity could obscure all but the closest relationships.

To evaluate relationships of Plumbaginaceae and Polygonaceae to families of Caryophyllales, Giannasi et al. (1992) used a number of phenetic and "phylogenetic" methods, including a Fitch-Margoliash dendrogram based upon genetic distances (Kimura, 1981; Felsenstein, 1990), maximum parsimony (PHYLIP, Felsenstein, 1990; PAUP, Swofford, 1991), and maximum likelihood (ML; Felsenstein, 1981). All three methods produced the same lack of resolution concerning relationships of these families; to state that Plumbaginaceae and Polygonaceae are "not closely related" (meaning "closely similar") to Caryophyllales does not preclude them nonetheless from being closest relatives. From the perspective of results presented here (Fig. 9A, B), spurious similarities in the data analyzed by Giannasi et al. (1992) appear to affect equally results of all three tree-building methods: *Gossypium* simultaneously attracts higher asterids and magnoliids (in our trees *Gossypium* is well imbedded among rosids; rosid II, Fig. 10A, B). (See Olmstead et al., 1992, 1993, this issue, for an example and discussion of the effects of taxon sampling.)

Prospects for improvements of tree-building methods with greater numbers of taxa exist (Penny et al., 1992). In the example of Giannasi et al. (1992), none of the methods employed succeeded in eliminating what we interpret as branch attractions due to the small number of taxa sampled and use of distantly related outgroups. Phylogenetic studies using morphological characters for a closely related group of rosid or asterid families would likely be affected adversely by outgroups of magnoliids or monocots. This phenomenon is perhaps even more probable with nucleotide data in which homology is initially assessed only by nucleotide position and character states are restricted to the same four alternatives. Character "homology" for distantly related taxa can be easily determined with *rbc*L data (i.e., a given nucleotide position in this size-conserved gene; "primary homology," de Pinna, 1991), but this does not mean that assessments of character-state homology (synapomorphy) are less subject to homoplasy than with other data.

If phylogenetic analyses of nucleotide data cannot be expected to reveal relationships when sampled with manageable numbers of taxa (from the standpoint of computation of minimal trees) and if sampling with sufficient numbers of taxa precludes assessing the parsimony of results, then we have reached an impasse until improved methods of analysis are developed. An appreciation of this problem has led us to be skeptical of the overall topologies presented (Figs. 1, 2), and competing ideas of relationships should not be overlooked when performing more restricted analyses of these and other gene sequences.

INTERNAL SUPPORT FOR THE BROAD TOPOLOGY: AN AD HOC ANALYSIS OF SUBCLASS HAMAMELIDAE

For purposes of suggesting an appropriate manner to examine internal support of families traditionally recognized as a natural group, we selected one of the most controversial subclasses, Hamamelidae sensu Cronquist (1981), for which we have data from 18 of 24 families. Numerous phylogenetic and systematic studies of the Hamamelidae have been completed (see various authors in Crane & Blackmore, 1989). The morphological features suggesting a close relationship among these families are largely those associated with the temperate amentiferous syndrome, and these clearly could be the result of parallel modification in unrelated lineages. In their developmental characteristics and wood anatomy, families of Hamamelidae are particularly heterogeneous (Cronquist, 1981; Crane & Blackmore, 1989). In performing the broad analysis, we sought to avoid a priori ideas about what constituted monophyletic subgroupings of angiosperms, but for this heuristic example we have accepted ad hoc the outgroup relationships found in the general study.

We selected 72 species that included all members of Hamamelidae and their immediate sister taxa as identified in trees from both searches. No attempt was made to select species that would reproduce the particulars of the general topologies. A tree search under the Fitch (equal weights) criterion using 2,000 random sequence additions, MULPARS, STEEPEST DESCENT, and NNI branch-swapping (but permitting only 10 trees to be held at each step) found only one island at maximum parsimony (i.e., all trees could be found by single branch swaps using any one of them as a starting tree; cf. Maddison, 1991). Additional as well as shorter islands could still exist but are unlikely after 2,000 repetitions (this type of search required about 24 hours to complete on a Macintosh Quadra 950 with 20 MB of RAM). After random addition searches were completed, the trees

found were used as starting points in a single analysis with MULPARS on and TBR swapping to completion; this process should find all equally parsimonious trees in the single island identified. Our search found 36 equally parsimonious Fitch trees, three of which were optimal under the weighting criterion of Albert et al. (1993, this issue; see Materials and Methods). One of these was randomly selected and is shown with Fitch branch lengths (ACCTRAN optimization; Fig. 16). The tree length was 2,234 steps, the C. I. was 0.288 (excluding unique characters), and the R. I. was 0.532. The maximum parsimony trees also were used to search for trees up to five steps less parsimonious under the Fitch criterion; the FILTER TREES option was used to identify trees at each length, and a strict consensus tree at each step was computed. The number of steps less parsimonious at which each topological component decayed was recorded (Fig. 16; "decay values" are shown below the branches: "d0" indicates that the branch is a polytomy in the strict consensus of the maximum parsimony trees, "d1" indicates that the branch is a polytomy in the consensus tree at one step less parsimonious, etc.).

This analysis identified eight lineages into which members of Hamamelidae fell (Fig. 16): *Eucommia*, clade A, sister to *Aucuba* (Cornaceae) and *Garrya* (Garryaceae) and nested within an asterid clade; Hamamelidaceae, Cercidiphyllaceae, etc., clade B, in a series paraphyletic to Saxifragaceae–Grossulariaceae; Fagales, etc., clade C, sister to Fabaceae–Polygalaceae; Urticales, clade D, sister to Rosaceae; *Leitneria*, clade E, nested within families of Sapindales; isolated Trochodendrales, clade F, sister to Buxaceae; *Platanus*, clade G, situated in a heterogeneous group; and the last, *Euptelea*, clade H, situated among Ranunculales. Many of these clades are well supported internally, decaying at three or more steps less parsimonious (some, such as Trochodendrales–Buxaceae and the *Platanus* assemblage are weakly supported as monophyletic lineages but clearly are not members of other well supported groups, leaving them in isolated positions apart from other Hamamelidae).

Results of this restricted analysis are congruent with the topology found in the broad searches (Figs. 1, 2) and indicate the level of internal support demonstrated by *rbc*L data. Future studies could combine morphological data with this molecular matrix and perform constraint experiments in which various rearrangements of these taxa are examined for their relative degrees of parsimony. The general conclusion from this example is that the Hamamelidae do not form a monophyletic lineage; they are shown to be grossly polyphyletic. Besides the amentiferous syndrome, the major trait of Hamamelidae is the presence of tannins, which is likewise compatible with a relationship to the other tannin-containing families, Fabaceae, Rosaceae, Saxifragaceae sensu stricto, and Crassulaceae. Many authors (e.g., in Crane & Blackmore, 1989) have discussed these families in terms of which are "lower" and which are "higher" families. This distinction finds some support from these results; the "lower" groups either stand in an isolated position near the base of the eudicots (Eupteleaceae, Platanaceae, and Tetracentraceae–Trochodendraceae) or basal within rosids (rosid IV, Fig. 2A; Daphniphyllaceae, Hamamelidaceae, and Cercidiphyllaceae), whereas most "higher" hamamelids (Casuarinaceae, Fagaceae, Moracae, Ulmaceae, and Urticaceae) demonstrate a well supported relationship to Fabaceae or Rosaceae (d > 5 on two branches at the base of the largest clade in Fig. 16). The position of *Leitneria* in Sapindales near Burseraceae is corroborated by a shared suite of secondary compounds and presence of intercellular resin canals.

The example presented above is not intended to be more than a superficial phylogenetic treatment of families traditionally referred to Hamamelidae. It is meant to serve as an example of how the general topology, which itself is suspected of being suboptimal and presently cannot be examined by decay analysis because of its size, may identify a relevant analysis within which questions of optimality and relative support can be addressed. We are pleased that general relationships found in the broad analyses hold up well when addressed in this and other more restricted investigations, none of which have found vastly different topologies.

ON THE INFORMATIVENESS OF ALL SUBSTITUTIONS

The majority of character-state changes in protein-coding genes have been demonstrated to occur at third positions within codons, and numerous empirical studies have shown third position substitutions to be more abundant in this and other data sets. Some workers have experimented with discarding third position substitutions from their analyses or analyzing nucleotide sequences inferred from amino acid data (which standardizes all synonymous substitutions; Martin & Dowd, 1991). In several studies (Conti et al., 1993, this issue; Donoghue et al., 1992; Kim et al., 1992; Smith et al., 1993, this issue), all three codon positions have been found to exhibit similar levels of homoplasy (and perhaps similar rates of change per site as

To gain a measure of the robustness of the topology, decay of parsimony (Bremer, 1988) was examined. All maximum parsimony trees of both islands were used as starting trees to search for trees up to five steps less parsimonious; the FILTER TREES option was used to identify trees at each length, and a strict consensus tree was computed at each step. The number of steps less parsimonious at which a branch collapsed was recorded as the decay index. The larger the number, the more robust a branch. A decay index of 1 indicates that the branch is present in all maximum parsimony trees but collapses at maximum parsimony plus one step.

Search D. There has been great interest in recent cladistic studies to determine the basalmost lineage of angiosperms. The Nymphaeales are often claimed as a sister group to all other angiosperms (Donoghue & Doyle, 1989; Hamby & Zimmer, 1992), but our results did not place them in such a position. To evaluate the hypothesis of Nymphaeles being the basalmost angiosperms, we performed a topology constraint experiment using the same set of taxa as in Search B, but with the Nymphaeales constrained to be sister to all other angiosperms (using the CONSTRAINTS option of PAUP). The search was conducted under the Fitch criterion using 2,000 random sequence additions in the same manner as described in Search B. This is a more accurate means to determine the loss of parsimony associated with an alternate topology; merely constraining the topology based on the shortest trees found overall without performing branch swapping is likely to result in an inaccurate assessment of tree length because it does not permit character transformations to be optimized on different branches than in the unconstrained trees.

RESULTS

Search A. A strict consensus tree was computed from 3,900 equally parsimonious trees at a length of 16,305 steps (Chase et al., 1993). These trees have a consistency index (C.I.; for potentially synapomorphous characters) of 0.102 and a retention index (R.I.) of 0.638; a summary of this tree is shown in Figure 1 (for the detailed version of the strict consensus tree, see Chase et al., 1993). Exclusive of *Ceratophyllum*, which is sister to all other flowering plants, the angiosperms are split into two clades that correspond to the two general pollen types, monosulcate and tricolpate (for the sake of convenience, we refer to angiosperms with monosulcate and monosulcate-derived pollen types

as the "monosulcates" and those with tricolpate and tricolpate-derived pollen types as the "eudicots," the latter in accord with Doyle & Hotton's, 1991, use of that term). Five major lineages are identified among the Magnoliidae; they are roughly equal to the Magnoliales, Laurales, Piperales (including Aristolochiaceae and Lactoridaceae), Nymphaeales (excluding Nelumbonaceae but including Amborellaceae, Austrobaileyaceae, Chloranthaceae, and Illiciales), and Ranunculales (excluding Coriariaceae, Sargentodoxaceae, and Sabiaceae but including Eupteleaceae, Fumariaceae, and Papaveraceae). (We will use these ordinal circumscriptions throughout the rest of this paper; i.e., under Piperales we will include Aristolochiaceae and Lactoridaceae, unless stated otherwise.) The four magnoliid lineages with monosulcate pollen types, (i) Magnoliales, (ii) Laurales, (iii) Piperales, and (iv) Nymphaeales, plus the monocots (with the same type of pollen), form a clade. The Illiciales, which have anomalous tricolpate pollen (see below), also fall into this monosulcate clade. The Piperales are sister to all other monosulcates; Magnoliales/Nymphaeales, Laurales, and monocots form an unresolved trichotomy in the strict consensus tree. The Ranunculales, Nelumbonaceae, and Sabiaceae, all of which have tricolpate pollen, together with other angiosperms possessing the same pollen type, form a monophyletic group. This large clade corresponds to Walker & Doyle's (1975) nonmagnoliid dicots or Doyle & Hotton's (1991) "eudicots." The Ranunculales are sister to the rest of the eudicots. In another basal lineage of eudicots are Nelumbonaceae and Sabiaceae (with Platanaceae and Proteaceae; hamamelid I, Fig. 1). The Coriariaceae and Sargentodoxaceae are not affiliated with the Magnoliidae, the former being sister to the clade of Cucurbitaceae/Begoniaceae/Datiscaceae and the latter imbedded in the Fabaceae (both in rosid I, Fig. 1; also see fig. 11B in Chase et al., 1993).

Search B. A single island of 46 trees at a length of 3,414 steps with C.I. of 0.275 (unique substitutions excluded) and R.I. of 0.566 was found. One of these maximum parsimony trees favored by the weighting criterion (Albert et al., 1993) is shown with Fitch branch lengths optimized on it (ACCTRAN optimization; Fig. 2). Even though far fewer taxa (82 species) were used in Search B, the same general topology as Search A was obtained.

Search C. Two islands of 89 and 3 trees respectively, at a length of 2,674 steps with C.I. of 0.346 (unique substitutions excluded) and R.I. of 0.531 were found. For illustration, one tree from each island was selected using the weighting criterion of Albert et al. (1993). The trees in the

using relative rate tests (Gaut et al., 1992) demonstrated five-fold variation for rbcL, which still falls within the range cited above. Although 5–10-fold differences in rates would appear potentially to contribute enormous branch-length differences, the small overall rates involved mean that great divergence times would be a more important factor. In addition, differences in rates appear to be highly lineage-correlated (for example, in graminoids; Gaut et al., 1992) rather than random, and if more extensive sampling at lower taxonomic levels is possible in these "fast" clades, then effects of rate differences can be offset.

Most if not all rbcL rates will probably exist within a relatively narrow "window," perhaps approximating the range illustrated above (10^{-10}–10^{-11}, V. A. Albert, M. W. Chase & J. F. Wendel, unpublished). If that is correct, lineage-specific rate inequalities are unlikely to be a primary factor in branch attraction; rather, asymmetrical divergence times would be implicated because the product of rate and time is the central parameter in considerations of potential systematic errors (see Albert et al., 1993, this issue). Thus uneven sampling or extinction of lineages may present greater problems than do differences in rates.

EFFECTS OF LATERAL GENOME TRANSFER, ANCESTRAL
POLYMORPHISMS, AND DIFFERENT
MODES OF INHERITANCE

The trees presented here represent only information from a single gene, and factors peculiar to its evolution could lead to erroneous results. Several of these phenomena are discussed below, but we feel that their impact is likely minimal. Genome transfers would result in all or parts of a genome being phylogenetically coherent (transferred as a unit) at the time transit occurs (Doyle, 1992; Rieseberg & Soltis, 1991), but before and after movement most characters within genomes evolve independently (although still linked if on the same chromosome) and ought to be expected to contain historical evidence. Hybridization is unlikely to influence phylogenetic analyses except at lower taxonomic levels. Even matings between divergent parents still occur within portions of families (and usually among closely related species in a genus) rather than between families. Parental taxa at the time of an ancient hybridization also were likely closely related, and in these genomes highly conserved loci, such as rbcL, would have been similar or even identical.

Ancestral polymorphisms (Pamilo & Nei, 1988; Wu, 1991; Doyle, 1992) also affect only closely related taxa, and, like the effects of ancient hybridization, those of ancient polymorphisms would likely be minor relative to subsequent genetic divergence. If lineages that contain polymorphisms diverged in a closely spaced manner and uneven sorting did take place, it is unlikely that any evidence of such variation within a progenitor could be identified as such over the great amounts of time involved in this study. Furthermore, polymorphisms are short-lived and are undocumented for conservative, single-copy loci, such as rbcL. At most these effects would be highly localized among groups of closely related terminal taxa and, with adequate taxon sampling, would not be expected to perturb greatly the results.

Lateral transfers not involving exchange of gametes (by unknown mechanisms) may have occurred between major lineages in the past (i.e., of rbcL from a purple bacterium to a red algal ancestor; Morden et al., 1992). Such transfers presently appear rare among land plants and seed plants in particular; furthermore, artificial transformations are relatively difficult and often have a destabilizing or transient effect on transformed plants. While we must admit that this is an unknown area that could have played a role in certain anomalous placements in the rbcL trees (e.g., *Montinia* and *Vahlia* of Saxifragaceae sensu lato among asterids; Fig. 15), at the same time we are not prepared to advocate it as a scenario until trees based on other data demonstrate a pattern consistent with such hypotheses (as in the example of Morden et al., 1992). Using evidence from studies of secondary chemistry and development, Morgan & Soltis (1993, this issue) build strong cases for the highly dispersed groupings found in their study of Saxifragaceae sensu lato. Furthermore, many of their findings were also congruent with recent investigations of non-molecular characters. Rather than resorting to explanations involving lateral transfers of chloroplasts by mechanisms about which we can only speculate, we would prefer first to examine specific cases from a cladistic perspective rather than from that of current taxonomic schemes.

An rbcL tree is not solely a maternal tree. Although chloroplast transmission is principally maternal and uniparental, several groups exhibit a paternal (e.g., conifers) or biparental pattern (Neale et al., 1986; Szmidt et al., 1987; Whatley, 1982; Wagner et al., 1987; Corriveau & Coleman, 1988; White, 1990; Owens & Morris, 1991). Others that have been thought most probably maternal, such as *Liriodendron* and *Magnolia* (Corriveau & Coleman, 1988), consistently exhibit 5–15% paternal inheritance (Sewell et al., 1993). The potential

effects of a mixed pattern of inheritance on gene trees are also unlikely to affect studies at interfamilial levels and above.

GENES VERSUS TAXA

One of the more persistent controversies surrounding molecular phylogenetic studies has centered on whether results, at a given taxonomic level, will be "improved" more by adding additional taxa sequenced for the same gene or by adding sequence analyses of additional genes for the same set of taxa (analyzed simultaneously or each performed independently for assessments of congruence; Pamilo & Nei, 1988; Wu, 1991). From the standpoint of corroboration, phylogenetic studies of other data sets are absolutely crucial. Nevertheless, it seems quite clear from our work on these data sets that ideas of relationships have changed considerably as more taxa have been added. We suspect that no single gene sequence can provide reasonable hypotheses of relationships of seed plants when sampled uncritically and superficially, but perhaps erroneous results from one locus would be "corrected" by stronger signals present in the others. Data to evaluate this most critical question do not exist: would separately analyzing 20 gene sequences for the same set of 25 to 30 taxa produce well supported relationships?

We are convinced that adding representatives of the still numerous and diverse families absent from this analysis has the potential to enhance assessments of relationships at all levels (despite obvious computational complications). A great deal of the variation present within the gene still remains to be sampled taxonomically. Studies in several groups of plants indicate that *rbc*L often can be valuable at rather low taxonomic levels (Albert, in press, in the slipper orchids; S. Graham, B. Morton & S. Barrett, unpublished, in *Eichhornia*; R. Price & J. Palmer, unpublished, in *Pelargonium*; S. Williams & M. Chase, submitted, in *Drosera*; Xiang et al., 1993, this issue, in *Cornus*). What can be concluded from these studies is that although it is absolutely critical that more genes be studied to provide corroboration, it is unlikely that any of these studies will make a contribution to understanding seed-plant relationships if the level of sampling is too sparse. Only empirical studies will ultimately resolve the "genes versus taxa" controversy. We estimate that by the time this paper is in print, more than 1,200 sequences of *rbc*L from seed plants will exist, and future studies will undoubtedly benefit from this enormous database.

An additional implication of adding more taxa is that the results presented here may not be stable or reflective of those of an analysis of two, three, or four times as many sequences. We point to two regions of the general topology that are morphologically quite heterogeneous and that we believe are possibly generated by undersampling: hamamelid II (Figs. 2A, B, 7A, B) and rosid I (Figs. 2A, B, 11A, B, C). In the former, the critical taxa needed to provide more appropriate relationships could be some of the still-unsampled families, but these groups (represented by *Lambertia*, *Nelumbo*, *Platanus*, and *Sabia*) could just as likely represent the isolated relicts of now largely extinct lineages. In the case of rosid I, only a small percentage of the families has been sampled (for instance, only 7 of 24 families in the Violales: Dilleniidae), and it is in such a case that we might predict "curious" sister-group relationships. Topological instability often occurs in cases where long terminal branches are next to short internal branches, that is, hypothesized relationships are drastically altered by the addition of related taxa that "break up" long branches. Two surprising pairs of sister taxa from Search I, *Erythroxylum–Viola* and *Ochna–Drypetes*, are quite divergent and connected to each other by relatively short branches (Fig. 11A; branch lengths shown only in 11B). In Search II, the addition of *Reinwardtia* displaced *Erythroxylum* from *Viola* to *Drypetes*, leaving *Ochna* to stand isolated from any other taxon (Fig. 11B). In such situations, assessments of relationships are difficult, but, as more closely related species are added, distinguishing synapomorphies for families and groups of families as distinct from autapomorphies for individual species will become more reliable. We would not argue that the relationships found for rosid I are "better" in Search I than in Search II (or vice versa), but rather point these out as areas of the trees that require additional sampling and in which many of the relationships suggested by these analyses of *rbc*L have little or no morphological support.

VALUE OF THE BROAD ANALYSIS

We view the relatively robust internal support found by evaluations of portions of the broad analysis as an indication that *rbc*L sequences contain information relevant to the evolutionary history of angiosperms (Fig. 16, on Hamamelidae and most of the other papers in this issue; in particular see Conti et al., 1993, and Rodman et al., 1993, for which comparable cladistic analyses of non-molecular data are also available and compared). Well-

characterized families and groups of families (as evidenced in several more restricted studies of morphology, anatomy, and secondary chemistry; e.g., Rodman, 1991a, b; Hufford, 1992) are largely congruent with our results. Other evidence supports the monophyly of the angiosperms (Doyle & Donoghue, 1986, 1992), monocots (Dahlgren et al., 1985), and eudicots (Donoghue & Doyle, 1989). Likewise, the position of *Ceratophyllum* as sister to the rest of the angiosperms compares favorably with the fossil record (Les, 1988; fossil fruits from 120 million years ago are identical to extant fruits, D. Dilcher, pers. comm.). The status of the families of the Papaverales–Ranunculales collectively as sister to the rest of the eudicots (Figs. 1, 2) also finds ample external support (Donoghue & Doyle, 1989, among others).

The results reported here do find some "surprising" relationships. Several families, such as Chenopodiaceae (Fig. 9A, B) and Berberidaceae with only two representatives each in our studies, are paraphyletic to other families. These results do not surprise us because sampling has been demonstrated repeatedly to be a major factor, and insufficient or uneven sampling can generate anomalous relationships. Concomitantly, our results emphasize the need for studies of familial limits to include much better taxon sampling than is generally the case in these two analyses. Some of these instances of paraphyly may be accurate; many pairs of temperate, herbaceous/woody, tropical families have long been suspected of being unnatural (for example, Lamiaceae are derived within Verbenaceae and Brassicaceae within Capparaceae in our studies).

Other "major" findings of this study, although discordant when viewed from the perspective of various taxonomic treatments, find support from recent studies of non-molecular characters. For example, the placement of Ericales as sister to asterids (Figs. 1, 2) was also found by Hufford (1992), and the monophyly of most mustard-oil families (Fig. 10A, B) was previously suggested by Rodman (1991b). Other sets of relationships are unique to this analysis and require more thorough morphological and molecular studies. These include: (i) the position of several families with polypetalous corollas and supposed affinities to Saxifragaceae among each major lineage of Asteridae (requiring a hypothesized reversal of the sympetalous condition in *Escallonia, Montinia, Phyllonoma,* and *Vahlia* (all Saxifragaceae sensu lato); (ii) nesting of dilleniid orders Capparales, Malvales, Theales, and Violales among rosid clades (the systems of both Dahlgren, 1980, and Thorne, 1992,

treated these groups in a manner somewhat similar to our topology); (iii) relationships of Nepenthaceae and Droseraceae to Caryophyllidae sensu lato (Fig. 9A, B); and (iv) specific associations of numerous problematic genera like *Dillenia* (among rosids near Caryophyllidae, Fig. 9A, or near the base of the asterids, Fig. 12B; see Olmstead et al., 1993, this issue), *Impatiens* (near members of Ebenales and Ericales; Fig. 13A, B), and *Nelumbo* and *Lambertia* (among lower hamamelids; Fig. 7A, B). Placement of most of these genera and families has varied substantially among recently proposed classifications (although no one has suggested the relationships found here), and their positions in this study will undoubtedly add to the controversies.

Because additional lineages were present in Search II and different methods were used to construct the trees, it is impossible to evaluate whether the topology found in Search II represents a different island of trees. Indeed, multiple islands of equally parsimonious trees were found in other studies in this issue: Morgan & Soltis (1993), Olmstead et al. (1993), and Qiu et al. (1993). Certainly shifts of some taxa, especially *Paeonia* from a basal asterid to sister of Crassulaceae and Saxifragaceae sensu stricto, suggest a radically different explanation of the distribution of at least some characters. *Paeonia* is well supported internally in its new position (its sister status to *Ribes* does not decay even at five steps less parsimonious; Fig. 16). Shifts of *Dillenia, Gunnera,* Santalales, and *Vitis* seem, at first glance, to be major alterations of position, but branches are so short near the split between asterids and rosids that these could not involve many additional steps in either topology (Fig. 2B).

Although the trees of Search II are preferred to those of Search I because some of them were swapped on to completion and therefore are more likely to represent at least a local optimum, taxonomic conclusions based on either search are untimely. When faced with the fact that large numbers of angiosperm families are still unrepresented in the *rbc*L data set, we would argue that a valid assessment of the most appropriate positions of many taxa, such as *Dillenia, Gunnera,* and *Vitis,* grossly premature. Some conclusions, polyphyly of Hamamelidae and Dilleniidae, for example, seem well supported now.

This study is noteworthy not only for its scope but also for the large number of contributors whose unpublished sequences made up the bulk of the data analyzed. This wide collaboration was advantageous to all workers; many found that taxa sequenced by other laboratories supposedly working

on distantly related groups (in recent taxonomic schemes) fell into or near their group of interest. Prime examples of this are the close phylogenetic relationship of Nepenthaceae and Droseraceae to families of Caryophyllidae, Pittosporaceae to Apiaceae–Araliaceae, Malvales and Capparales (both "dilleniids") to Sapindales, and *Corokia* (Cornaceae) to Asteraceae.

Broad analyses are thus important in providing evaluations of a priori assumptions about appropriate sets of study taxa for more focused and rigorous studies; they should be formalized so that someone takes the initiative to perform them. Governmental funding agencies should facilitate studies, such as the one presented here, that are well beyond the scope of individual laboratories. This study demonstrates the potential of this kind of analysis, but no single individual or laboratory could have received formal support for sampling this diverse set of taxa and performing the phylogenetic analysis; it would have been deemed by reviewers too broad and too unfocused. Although extramural funds supported most of the other studies in this issue, no support was received specifically for the broad analysis. It is the investment in individual studies that justifies a further expenditure to support syntheses that supply an essential overall perspective, even though they may be necessarily approximate.

The benefits of performing this largest-yet phylogenetic study of seed plants lie not only in support of specific relationships hypothesized by other studies (e.g., the sister group status of the Ericales and higher asterids proposed by Hufford, 1992) and in identification of previously unhypothesized monophyletic groups. This study also presents a comprehensive, explicit hypothesis for higher level relationships, permitting and encouraging initiation of studies evaluating other character systems that may show congruence with the major lineages of seed plants described here. We have performed none of the essential experiments that these results suggest (topological constraints, removal of characters, combining morphological with molecular data, etc.) and have not developed implications these topologies may have for specific character transformations in seed plants or molecular evolution of *rbc*L or RuBisCO. Believing that serious consideration of the significance of these general topologies is best handled at the more manifest level of other papers published here (and elsewhere), we have chosen instead to emphasize reasons for caution. We are content to present the findings of these studies as examples of potential benefits and pitfalls of such exercises. At the least,

these analyses represent an attempt to improve both our understanding of seed-plant evolution and methods of phylogenetic inference. Corroboration by other data sets analyzed in a similar fashion is by far the most significant measure of relationships proposed here, and we hope this process of evaluation will be innervated by our efforts.

LITERATURE CITED

ALBERT, V. A. Phylogeny of the slipper orchids (Cypripedioideae: Orchidaceae) from congruent morphological and molecular data sets. Nordic J. Bot. (in press).
——— & B. D. MISHLER. 1992. On the rationale and utility of weighting nucleotide sequence data. Cladistics 8: 73–83.
———, M. W. CHASE & B. D. MISHLER. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. Ann. Missouri Bot. Gard. 80: 752–766.
———, B. D. MISHLER & M. W. CHASE. 1992a. Character-state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. Pp. 369–403 *in* P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics in Plants. Chapman & Hall, New York.
———, S. E. WILLIAMS & M. W. CHASE. 1992b. Carnivorous plants: Phylogeny and structural evolution. Science 257: 1491–1495.
ARCHIE, J. W. 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38: 219–252.
BOUSQUET, J., S. H. STRAUSS, A. D. DOERKSEN & R. A. PRICE. 1992. Extensive variation in evolutionary rate of *rbc*L gene sequences among seed plants. Proc. Natl. Acad. U.S.A. 89: 7844–7848.
BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42: 795–803.
———. 1990. Combinable component consensus. Cladistics 6: 369–372.
———, C. J. HUMPHRIES, B. D. MISHLER & S. P. CHURCHILL. 1987. On cladistic relationships in green plants. Taxon 36: 339–349.
CONTI, E., A. FISCHBACH & K. J. SYTSMA. 1993. Tribal relationships in Onagraceae: Implications from *rbc*L sequence data. Ann. Missouri Bot. Gard. 80: 672–685.
CORRIVEAU, J. L. & A. W. COLEMAN. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Amer. J. Bot. 75: 1443–1458.
CRANE, P. R. 1985. Phylogenetic analysis of seed plants and the origin of the angiosperms. Ann. Missouri Bot. Gard. 72: 716–793.
———. 1988. Major clades and relationships in the "higher" gymnosperms. Pp. 218–272 *in* C. B. Beck (editor), Origin and Evolution of Gymnosperms. Columbia Univ. Press, New York.
——— & S. BLACKMORE. 1989. Evolution, Systematics, and Fossil History of the Hamamelidae. Systematics Assoc., Clarendon Press, Oxford.
CRONQUIST, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

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

———— & K. BREMER. 1985. Major clades of the angiosperms. Cladistics 1: 349–368.

————, H. T. CLIFFORD & P. F. YEO. 1985. The Families of the Monocotyledons: Structure, Evolution, and Taxonomy. Springer-Verlag, New York.

DOEBLEY, J., M. DURBIN, E. M. GOLENBERG, M. T. CLEGG & D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097–1108.

DONOGHUE, M. J. & J. A. DOYLE. 1989. Phylogenetic studies of seed plants and angiosperms based on morphological characters. Pp. 181–193 in B. Fernholm, K. Bremer & H. Jörnvall (editors), The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis. Elsevier Science Publishers, Amsterdam.

————, R. G. OLMSTEAD, J. F. SMITH & J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on rbcL sequences. Ann. Missouri Bot. Gard. 79: 333–345.

————, J. A. DOYLE, J. GAUTHIER, A. G. KLUGE & T. ROWE. 1989. The importance of fossils in phylogeny reconstruction. Ann. Rev. Ecol. Syst. 20: 431–460.

DOWNIE, S. R. & J. D. PALMER. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. Pp. 14–35 in P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics in Plants. Chapman & Hall, New York.

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

———— & ————. 1992. Fossils and seed plant phylogeny reanalyzed. Brittonia 44: 89–106.

———— & L. J. HICKEY. 1976. Pollen and leaves from the mid-Cretaceous Potomac Group and their bearing on early angiosperm evolution. Pp. 139–206 in C. B. Beck (editor), Origin and Early Evolution of Angiosperms. Columbia Univ. Press, New York.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Syst. Bot. 17: 144–163.

———— & HOLTON. 1991. Diversification of early angiosperm pollen in a cladistic context. Pp. 169–195 in S. Blackmore & S. H. Barnes (editors), Pollen and Spores. Clarendon Press, Oxford.

DUVALL, M. R., M. T. CLEGG, M. W. CHASE, W. D. CLARK, W. J. KRESS, H. G. HILLS, L. E. EGUIARTE, J. F. SMITH, B. S. GAUT, E. A. ZIMMER & G. H. LEARN. 1993. Phylogenetic hypotheses for the monocotyledons constructed from rbcL sequence data. Ann. Missouri Bot. Gard. 80: 607–619.

FAITH, D. & P. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone? Cladistics 7: 1–28.

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

————. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Molec. Evol. 17: 368–376.

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

————. 1990. PHYLIP manual, version 3.3. Univ. of California, University Herbarium, Berkeley, California.

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

GAUT, B. S., S. V. MUSE, W. D. CLARK & M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. J. Molec. Evol. 35: 292–303.

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

GOLDBLATT, P., P. C. HOCH & L. M. McCOOK. 1992. Documenting scientific data: The need for voucher specimens. Ann. Missouri Bot. Gard. 79: 969–970.

HAMBY, R. K. & E. A. ZIMMER. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. Pp. 50–91 in P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics in Plants. Chapman & Hall, New York.

HARRIS, P. J. & R. D. HARTLEY. 1980. Phenolic constituents of cell walls of monocotyledons. Biochem. Syst. Ecol. 8: 153–160.

HENDY, M. D. & D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38: 297–309.

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

JANSEN, R. K. & J. D. PALMER. 1988. Phylogenetic implications of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). Amer. J. Bot. 75: 751–764.

KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE & C. BULT. 1992. Skewness and permutation. Cladistics 8: 275–287.

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

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Molec. Evol. 16: 111–120.

————. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. U.S.A. 78: 454–458.

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

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

————, D. K. GARVIN & C. F. WIMPEE. 1991. Molecular evolutionary history of ancient aquatic angiosperms. Proc. Natl. Acad. U.S.A. 88: 10119–10123.

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

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

MARTIN, P. G. & J. M. DOWD. 1991. Studies of angiosperm phylogenies using protein sequences. Ann. Missouri Bot. Gard. 78: 296–337.

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 rbcL sequence variation. Ann. Missouri Bot. Gard. 80: 742–765.

MORDEN, C. W., C. F. DELWICHE, M. KUHSEL & J. D. PALMER. 1992. Gene phylogenies and the endosymbiotic origin of plastids. Biosystems 28: 75–90.

MORGAN, D. R. & D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on rbcL sequence data. Ann. Missouri Bot. Gard. 80: 631–660.

NEALE, D. B., N. C. WHEELER & R. W. ALLARD. 1986. Paternal inheritance of chloroplast DNA in Douglas-fir. Canad. J. Forest Res. 16: 1152–1154.

NIXON, K. C. & G. I. DAVIS. 1991. Polymorphic taxa, missing values, and cladistic analyses. Cladistics 7: 233–241.

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 rbcL. Ann. Missouri Bot. Gard. 79: 249–265.

———, B. BREMER, K. M. SCOTT & J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on rbcL sequences. Ann. Missouri Bot. Gard. 80: 700–722.

OWENS, J. N. & S. J. MORRIS. 1991. Cytological basis for cytoplasmic inheritance in Pseudotsuga menziesii. II. Fertilization and proembryo development. Amer. J. Bot. 78: 1515–1527.

PALMER, J. D., R. K. JANSEN, H. J. MICHAELS, M. W. CHASE & J. M. MANHART. 1988. Analysis of chloroplast DNA variation. Ann. Missouri Bot. Gard. 75: 1180–1206.

PAMILO, P. & M. NEI. 1988. Relationships between gene trees and species trees. Molec. Biol. Evol. 5: 568–583.

PENNY, D., M. D. HENDY & M. A. STEEL. 1992. Progress with methods for constructing evolutionary trees. Trends Ecol. Evol. Biol. 7: 73–79.

DE PINNA, M. C. C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7: 367–394.

PLATNICK, N. I., C. E. GRISWOLD & J. A. CODDINGTON. 1991. On missing entries in cladistic analyses. Cladistics 7: 337–343.

PRICE, R. A. & J. D. PALMER. 1993. Phylogenetic relationships of the Geraniaceae and Geraniales from rbcL sequence comparisons. Ann. Missouri Bot. Gard. 80: 661–671.

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

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

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

RIESEBERG, L. H. & D. E. SOLTIS. 1991. Phylogenetic

consequences of cytoplasmic gene flow in plants. Evol. Trends Pl. 5: 65–84.

RITLAND, K. & M. T. CLEGG. 1987. Evolutionary analyses of plant DNA sequences. Amer. Naturalist 130: S74–S100.

RODMAN, J. 1991a. A taxonomic analysis of glucosinolate-producing plants, Part 1: Phenetics. Syst. Bot. 16: 598–618.

———. 1991b. A taxonomic analysis of glucosinolate-producing plants, Part 2: Cladistics. Syst. Bot. 16: 619–629.

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

SCHWARZWALDER, R. & D. L. DILCHER. 1991. Systematic placement of the Platanaceae in the Hamamelidae. Ann Missouri Bot. Gard. 78: 962–969.

SEWELL, M., M. W. CHASE, Y.-L. QIU & C. R. PARKS. 1993. Trace paternal inheritance of chloroplast DNA in Liriodendron and Magnolia (Magnoliaceae). Amer. J. Bot. 80 (in press).

SMITH, J. F., W. J. KRESS & E. A. ZIMMER. 1993. Phylogenetic analysis of the Zingiberales based on rbcL sequences. Ann. Missouri Bot. Gard. 80: 620–630.

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

SWOFFORD, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0r. Computer program distributed by the 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 Associates, Sunderland, Massachusetts.

SZMIDT, A. E., T. ALDÈN & J.-E. HALLGREN. 1987. Paternal inheritance of chloroplast DNA in Larix. Pl. Molec. Biol. 9: 59–64.

TAKHTAJAN, A. 1980. Outline of the classification of flowering plants (Magnoliophyta). Bot. Rev. 46: 225–359.

———. 1987. Systema Magnoliophytorum. Nauka, Leningrad. [In Russian.]

TAYLOR, D. W. & L. J. HICKEY. 1992. Phylogenetic evidence for the herbaceous origin of angiosperms. Pl. Syst. Evol. 180: 137–156.

THORNE, R. F. 1983. Proposed new realignments in the angiosperms. Nordic J. Bot. 3: 85–117.

———. 1992. An updated classification of the flowering plants. Aliso 13: 365–389.

WAGNER, D. B., G. R. FURNIER, M. A. SAGHAI-MAROOF, S. M. WILLIAMS, D. B. DANCIK & R. W. ALLARD. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. U.S.A. 84: 2097–2100.

WENDEL, J. F. & V. A. ALBERT. 1992. Phylogenetics of the cotton genus (Gossypium): Character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17: 115–143.

WHATLEY, J. M. 1982. Ultrastructure of plastid inheritance: Green algae to angiosperms. Bot. Rev. 57: 527–569.

WHITE, E. E. 1990. Chloroplast DNA in Pinus mon-

ticola. 2. Survey of within-species variability and detection of heteroplasmic individuals. Theoret. Appl. Gen. 79: 251–255.

WILSON, M. A., B. GAUT & M. T. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family. Molec. Biol. Evol. 7: 303–314.

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

WU, C.-L. 1991. Inference of species phylogeny in relation to segregation of ancient polymorphisms. Genetics 127: 429–435.

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.

ZURAWSKI, G. & M. T. CLEGG. 1987. Evolution of higher-plant chloroplast DNA-coded genes: Implications for structure-function and phylogenetic studies. Ann. Rev. Pl. Phys. 38: 391–418.

[2] Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280, U.S.A. Current address for MWC: Laboratory of Molecular Systematics, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom. Current address for KAK: Department of Biology, Wake Forest University, Winston-Salem, North Carolina 27109, U.S.A. Current address for VAA: Institutionen För Systematisk Botanik, Uppsala Universitet, Uppsala S-751 21, Sweden.

[3] Department of Botany, Washington State University, Pullman, Washington 99164-4238, U.S.A.

[4] Department of E. P. O. Biology, University of Colorado, Boulder, Colorado 80309-0334, U.S.A.

[5] Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin 53201, U.S.A. Current address for DHL: Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269-3042, U.S.A.

[6] Department of Botany, Duke University, Durham, North Carolina 27708-0338, U.S.A. Current address: Department of Integrative Biology, University of California, Berkeley, California 94720, U.S.A.

[7] Department of Botany and Plant Sciences, University of California, Riverside, California 92521-0124, U.S.A.

Current address for LEE: Centro de Ecología, Universidad Nacional Autonóma de México, Apartado Postal 70-275, C.U., C.P. 04510, México, D.F., Mexico. Current address for BSG: Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A. Current address for MRD: Office of Biotechnology, Iowa State University, Ames, Iowa 50011-3260, U.S.A.

[8] Department of Biology, Indiana University, Bloomington, Indiana 47405, U.S.A. Current address for RAP: Department of Botany, University of Georgia, Athens, Georgia 30602, U.S.A.

[9] Department of Biology, Texas A & M University, College Station, Texas 77843, U.S.A. Current Address for JHR: Department of Biology, College of the Ozarks, Point Lookout, Missouri 65726, U.S.A.

[10] Department of Botany, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

[11] Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio 43403, U.S.A.

[12] Department of Botany, Smithsonian Institution, Washington, D.C. 20560, U.S.A. Current address for JFS: Department of Biology, Boise State University, Boise, Idaho 83725, U.S.A.

[13] Department of Botany, Arizona State University, Tempe, Arizona 85287, U.S.A.

[14] Institutionen För Systematisk Botanik, Uppsala Universitet, Uppsala S-751 21, Sweden.

[15] Department of Botany, University of Texas, Austin, Texas 78713, U.S.A.

[16] Department of Forest Resources and Plant Biology, University of Minnesota, St. Paul, Minnesota 55108-8112, U.S.A.

[17] Department of Forest Science, FSL 020, Oregon State University, Corvallis, Oregon 97331-7505, U.S.A.

[18] Department of Botany, University of Tennessee, Knoxville, Tennessee 37996-1100, U.S.A. Current address: Rancho Santa Ana Botanic Garden, Claremont, California 91711-3101, U.S.A.

[19] Department of Biology, Lebanon Valley College, Annville, Pennsylvania 17003-0501, U.S.A.

[20] School of Biological Sciences, University of New South Wales, P.O. Box 1, Kensington, New South Wales, Australia.

[21] Department of Biological Sciences, Wayne State University, Detroit, Michigan 48202, U.S.A.

[22] Department of Botany, University of Toronto, Toronto, Ontario M5S 3B2, Canada.

[23] Department of Biology, Boston University, Boston, Massachusetts 02215, U.S.A.

1A

Asteridae -- including
Ericales, Primulales,
Ebenales, Santalales,
Apiales, Cornales, and
some Rosales

Rosidae -- including
Violales, Malvales,
mustard-oil families,
higher Hamamelidae,
and Caryophyllidae

hamamelids
ranunculids
paleoherbs

monocots

Laurales
Magnoliales
Ceratophyllaceae
gnetophytes
Pinaceae
other conifers
cycads
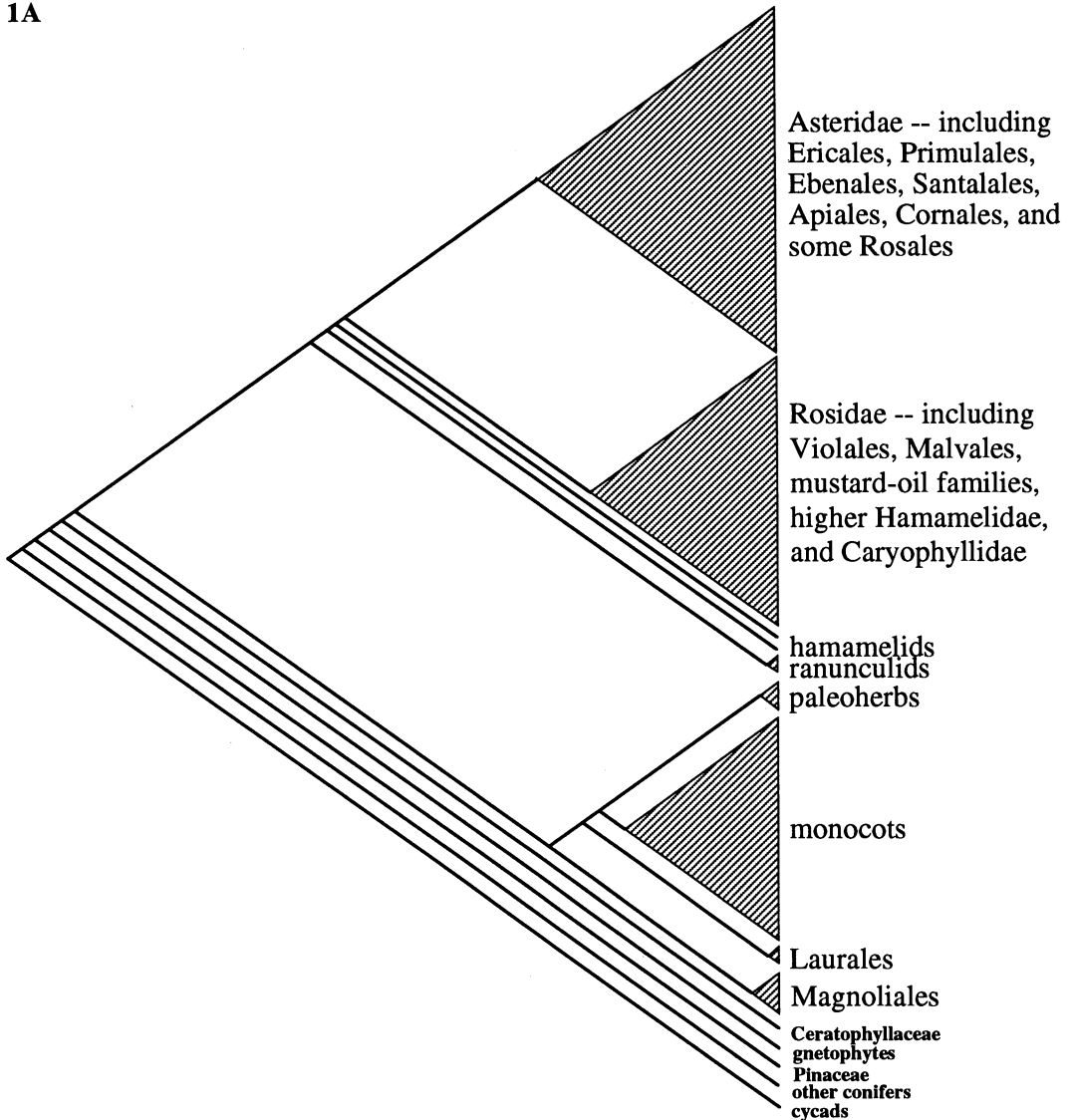
FIGURE 1. Summaries of the major clades identified in: (A) the combinable component consensus tree of 500 equally parsimonious trees found for 475 taxa using the character-state weighting method of Albert et al. (1993, this issue); and (B) the strict consensus tree of 3,900 equally parsimonious trees for 499 taxa found using the Fitch (even weights) criterion. These are ingroup networks arranged arbitrarily with the cycads sister to all other seed plants.

1B

Asteridae -- including
Ericales, Primulales,
Ebenales, Santalales,
Apiales, Cornales, and
some Rosales

Rosidae -- including
Violales, Malvales,
mustard-oil families,
and higher
Hamamelidae

caryophyllids
Gunneraceae
hamamelids
ranunculids
paleoherbs II
Magnoliales
Laurales

monocots

paleoherbs I
Ceratophyllaceae
gnetophytes
other conifers
Pinaceae
Ginkgo
cycads

2A

asterid I
asterid II
asterid III
asterid IV
asterid V
rosid IV
rosid III
rosid II
rosid I
hamamelid II
hamamelid I
ranunculids
paleoherbs
monocots
Laurales
Magnoliales
Ceratophyllum
Gnetales
Pinaceae
other conifers
cycads

FIGURE 2.   Summaries of the same topologies as in Figure 1. In B, Fitch branch lengths are optimized from a single tree; optimization on consensus trees is likely to overestimate branch lengths. Names of the specific clades identified do not conform to the composition of families used in most taxonomic schemes, but rather are designated with respect to the components of the major lineages (i.e., by the subclass name for the majority of taxa included, except for the heterogeneous hamamelid I, which is so designated because of its position and inclusion of Platanaceae). Names of each clade correspond to groups shown in Figures 3–15. Clades marked with a "§" in B are those that differ significantly in position or composition from A.

2B
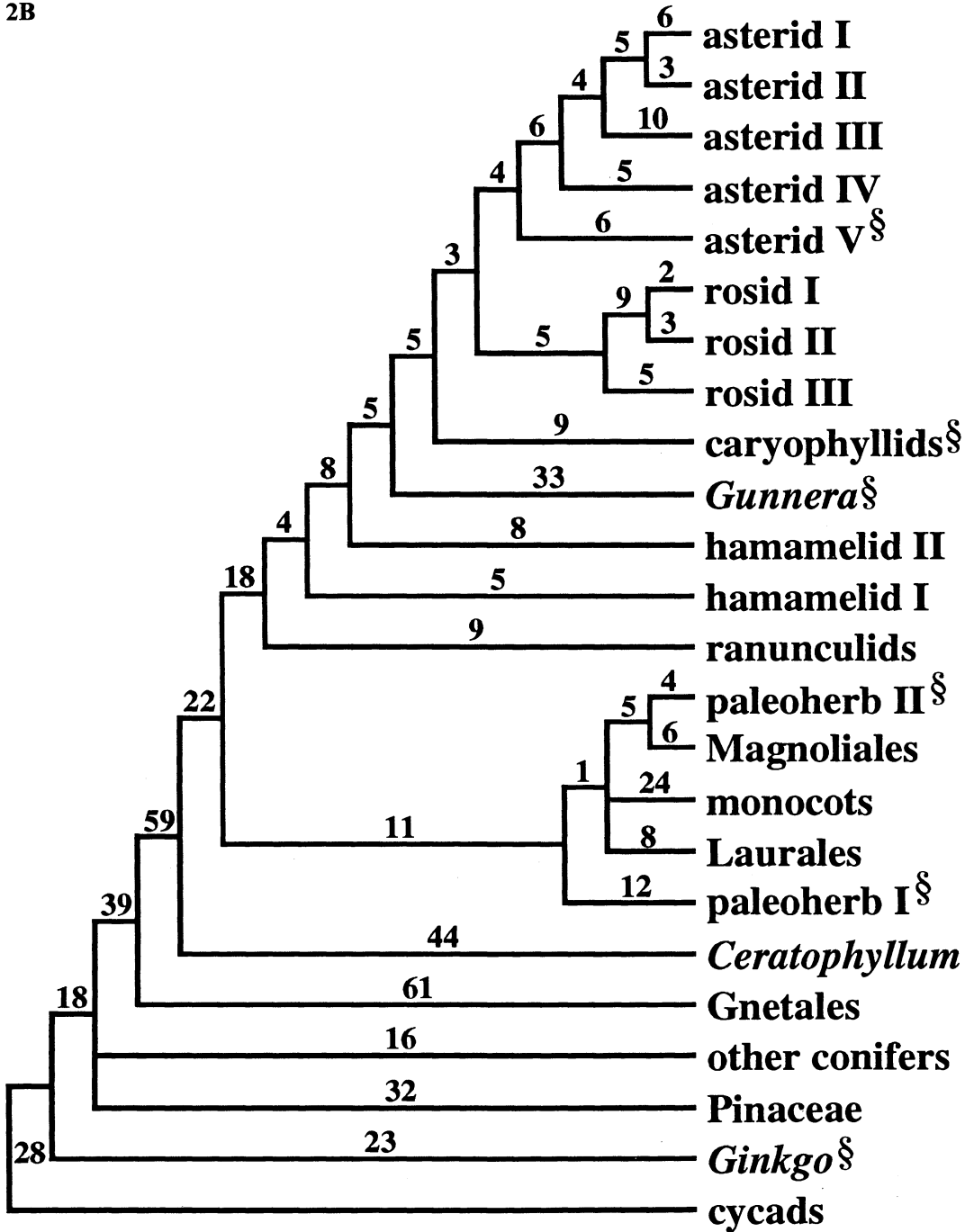
3A
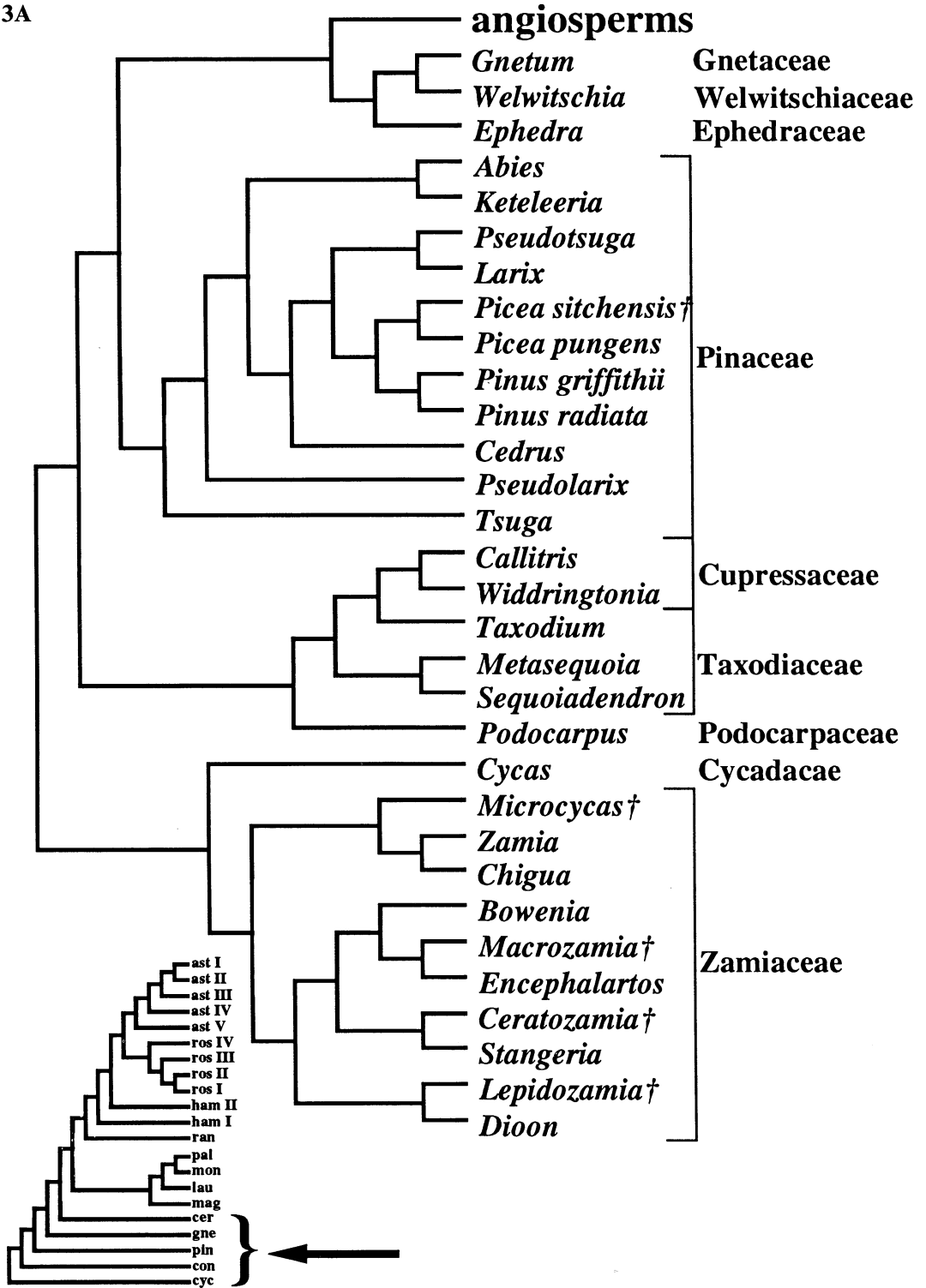
FIGURE 3. A portion of the overall analysis showing the "gymnosperms." (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow. Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I.

3B

eudicots

Euryale†
Victoria
Nymphaea
Nuphar
Barclaya
Brasenia
Cabomba
Amborella
Schisandra
Illicium
Austrobaileya
Chloranthus
Aristolochia
Saruma
Asarum
Houttuynia
Saururus
Peperomia

monocots

Hernandia
Hedycarya
Idiospermum
Calycanthus chinensis
Calycanthus floridus†
Chimonanthus
Cinnamomum
Persea
Canella
Drimys
Belliolum
Magnolia macrophylla†
Magnolia salicifolia
Michelia
Manglietia
Talauma singaporensis
Talauma ovata†
Magnolia tripetala†
Magnolia hypoleuca
Liriodendron
Degeneria
Annona
Asimina
Galbulimima
Eupomatia
Knema
Ceratophyllum

Nymphaeaceae[a]
Barclayaceae[a]
Cabombaceae[a]
Amborellaceae[c]
Schisandraceae[f]
Illiciaceae[f]
Austrobaileyaceae[b]
Chloranthaceae[d]
Aristolochiaceae[e]
Saururaceae[d]
Piperaceae[d]

Hernandiaceae[c]
Monimiaceae[c]
Idiospermaceae[c]
Calycanthaceae[c]
Lauraceae[c]
Canellaceae[b]
Winteraceae[b]

Magnoliaceae[b]

Degeneriaceae[b]
Annonaceae[b]
Himantandraceae[b]
Eupomatiaceae[b]
Myristicaceae[b]
Ceratophyllaceae[a]

paleoherbs

Laurales

Magnoliales

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

[a]Magnoliidae, Nyphaeales
[b]Magnoliidae, Magnoliales
[c]Magnoliidae, Laurales
[d]Magnoliidae, Piperales
[e]Magnoliidae, Aristolochiales
[f]Magnoliidae, Illiciales

FIGURE 4. Basal portion of the overall analysis showing the positions of Ceratophyllaceae (inaperturate pollen), monocots (uniaperturate pollen), eudicots (dicots with triaperturate pollen), and the three clades of "primitive" dicots (monosulcate pollen). Note that, exclusive of *Ceratophyllum*, the angiosperms form two sister groups marked by the general pollen aperture number (one versus three). (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. The "*Canella*" in this figure represents the position of two sequences amplified from a total cellular DNA template (see Materials & Methods). Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

4B

18 — **eudicots**

3	6	*Victoria*
4	4	*Nymphaea*
	1 5	*Nuphar*
42	5	*Barclaya*
12	10 6	*Brasenia*
	1	*Cabomba*
14	35 6	*Amborella*
15 14	*Schisandra*
14 15	*Illicium*
24	*Austrobaileya*
14 29	*Hedyosmum**
10 12	*Chloranthus*
15	*Sarcandra**

Nymphaeaceae[a]

Barclayaceae[a]
Cabombaceae[a]

Amborellaceae[c]
Schisandraceae[f]
Illiciaceae[f]
Austrobaileyaceae[b]

Chloranthaceae[d]

43 — *Canella* — Canellaceae[b]
7
2 7 *Drimys*
19 14 *Belliolum*
5 *Tasmannia**
Winteraceae[b]

Magnolia salicifolia
Michelia
Talauma ovata
Magnolia hypoleuca
Manglietia
Talauma singaporensis
Liriodendron tulipifera
Liriodendron chinensis*

Magnoliaceae[b]

13 *Degeneria* — Degeneriaceae[b]
3 24 *Galbulimima* — Himantandraceae[b]
24 *Eupomatia* — Eupomatiaceae[b]
22 *Annona*
13 *Asimina*
20 *Cananga**
Annonaceae[b]

21 — *Knema* — Myristicaceae[b]

13 6 *Cinnamomum*
3 27 *Persea*
20 *Hedycarya*
13 *Idiospermum*
15 6 *Calycanthus*
9 2 *Chimonanthus*
10 33 *Hernandia*
28 *Gyrocarpus**
Lauraceae[c]
Monimiaceae[c]
Idiospermaceae[c]
Calycanthaceae[c]
Hernandiaceae[c]

24 — **monocots**

17 11 *Asarum*
4 *Saruma*
11 39 *Aristolochia*
18 38 *Lactoris**
Aristolochiaceae[e]
Lactoridaceae[b]

14 *Saururus*
12 15 *Houttuynia*
27 18 *Piper**
39 34 *Peperomia*
Saururaceae[d]
Piperaceae[d]

44 — *Ceratophyllum* — Ceratophyllaceae[a]

paleoherb II

Magnoliales

Laurales

paleoherb I

[a] Magnoliidae, Nyphaeales
[b] Magnoliidae, Magnoliales
[c] Magnoliidae, Laurales
[d] Magnoliidae, Piperales
[e] Magnoliidae, Aristolochiales
[f] Magnoliidae, Illiciales

5A

commelinoids

Genus	Family
Xanthorrhoea	Xanthorrhoeaceae[J]
Kniphofia *Aloe* *Haworthia*	Asphodelaceae[J]
Sansevieria	Dracaenaceae[J]
Danae	Ruscaeae[J]
Nolina recurvata *Nolina lindheimerina†*	Nolinaceae[J]
Bowiea	Hyacinthaceae[J]
Chlorophytum	Anthericaceae[J]
Clivia	Amaryllidaceae[J]
Iris *Anomatheca*	Iridaceae[i]
Cyanastrum	Tecophilaeaceae[J]
Curculigo *Hypoxis*	Hypoxidaceae[J]
Neuwiedia *Oncidium*	Orchidaceae[i]
Smilax	Smilacaceae[e]
Lilium *Medeola*	Liliaceae[i]
Alstroemeria	Alstroemeriaceae[i]
Colchicum *Burchardia*	Colchicaceae[i]
Chamaelirium	Melanthiaceae[b]
Vellozia	Velloziaceae[h]
Freycinetia	Pandanaceae[g]
Sphaeradenia	Cyclanthaceae[f]
Dioscorea	Dioscoreaceae[e]
Tacca	Taccaceae[e]
Aletris	Melanthiaceae[b]
Burmannia†	Burmanniaceae[d]
Sagittaria graminea *Sagittaria latifolia* *Alisma*	Alismataceae[c]
Potamogeton	Potamogetonaceae[c]
Pleea	Melanthiaceae[b]
Spathiphyllum *Gymnostachys*	Araceae[a]

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

[a]Aranae
[b]Lilianae, Melanthiales
[c]Alismatanae
[d]Lilianae, Burmanniales
[e]Lilianae, Dioscoreales
[f]Cyclanthanae
[g]Pandananae
[h]Bromelianae, Velloziales
[i]Lilianae, Liliales
[J]Lilianae, Asparagales

FIGURE 5. The five basalmost lineages of the monocots, composed of the aroids, alismatids, and lilioid groups. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

5B

Cypripedium* — 8
Neuwiedia — 27
Oncidium — 6, 25
Orchidaceae[d]

Lomandra* — 27 — Dasypogonaceae[i]
Hemerocallis* — 19, 38 — Hemerocallidaceae[i]
Chlorophytum — 17 — Anthericaceae[i]
Xanthorrhoea — 17 — Xanthorrhoeaceae[i]
Kniphofia — 21
Aloe — 25, 13, 6
Haworthia — 1
Asphodelaceae[i]
Scilla* — 18, 10, 20
Bowiea — 10
Hyacinthaceae[i]
Sansevieria — 21
Danae — 11, 5, 8
Nolina
Dracaenaceae[i]
Ruscaceae[i]
Nolinaceae[i]
Clivia — 11 — Amaryllidaceae[i]
Iris — 21, 19
Anomatheca
Iridaceae[d]
Cyanastrum — 19 — Tecophilaeaceae[i]
Curculigo — 23, 21
Hypoxis
Hypoxidaceae[i]

commelinoids — 6

Dioscorea — 27, 28
Tacca — 27
Dioscoreaceae[d]
Taccaceae[d]
Pandanus* — 5, 14, 16
Freycinetia — 10
Pandanaceae[h]
Sphaeradenia — 40 — Cyclanthaceae[g]
Vellozia — Velloziaceae[f]
Smilax — 30 — Smilacaceae[e]
Calochortus* — 19
Lilium — 14, 17, 21
Medeola
Calochortaceae[d]
Liliaceae[d]
Colchicum — 18, 24, 16
Burchardia — 39
Alstroemeria
Colchicaceae[d]
Alstroemeriaceae[d]
Veratrum* — 36, 23
Chamaelirium
Melanthiaceae[c]
Aletris — 57

Pleea — 22
Sagittaria graminea — 18, 30, 8
Sagittaria latifolia — 34, 20
Alisma — 44, 56
Alismataceae[b]
Potamogeton — Potamogetonaceae[b]
Spathiphyllum — 13, 41
Pistia* — 23, 17, 39
Lemna*
Araceae[a]
Lemnaceae[a]
Gymnostachys — 23
Acorus* — 45
Araceae[a]

[a] Aranae
[b] Alismatanae
[c] Lilianae, Melanthiales
[d] Lilianae, Liliales
[e] Lilianae, Dioscoreales
[f] Bromelianae, Velloziales
[g] Cyclanthanae
[h] Pandananae
[i] Lilianae, Asparagales

6A

Avena
Puccinella †
Triticum
Hordeum †
Aegilops
Cenchrus
Pennisetum †
Neurachne
Oryza
Elegia
Juncus
Oxychloe
Cyperus
Carex
Prionium
Flagellaria
Typha
Stegolepis
Aechmea †
Ananas
Glomeropitcairnia †
Catopsis †
Tillandsia
Puya
Hechtia †
Tradescantia pallida
Tradescantia zebrina
Tradescantia soconuscana
Pontederia
Philydrum
Anigozanthos
Ravenala
Strelitzia
Phenakospermum
Maranta
Calathea
Hedychium
Riedelia
Zingiber
Globba
Costus
Tapeinocheilos
Musa
Orchidantha
Heliconia
Caryota †
Nypa
Phoenix †
Drymophloeus
Chamaedorea
Calamus †
Serenoa

Poaceae[i]

Restionaceae[i]
Juncaceae[j]
Cyperaceae[j]
Juncaceae[j]
Flagellariaceae[i]
Typhaceae[h]
Rapateaceae[f]

Bromeliaceae[g]

Commelinaceae[f]
Pontederiaceae[e]
Philydraceae[d]
Haemodoraceae[c]

Strelitziaceae[b]

Marantaceae[b]

Zingiberaceae[b]

Costaceae[b]
Musaceae[b]
Lowiaceae[b]
Heliconiaceae[b]

Arecaceae[a]

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

[a]Arecanae
[b]Zingiberanae
[c]Bromelianae, Haemodorales
[d]Bromelianae, Philydrales
[e]Bromelianae, Pontederiales
[f]Commelinanae, Commelinales
[g]Bromelianae, Bromeliales
[h]Bromelianae, Typhales
[i]Commelinanae, Poales
[j]Commelinanae, Cyperales

FIGURE 6.   The terminal lineages of the monocots, composed of the palms, gingers, and commelinoids. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

6B

Poaceae[i]	
Eriocaulaceae[f]	
Restionaceae[i]	
Flagellariaceae[i]	
Juncaceae[j]	
Cyperaceae[j]	
Juncaceae[j]	
Sparganiaceae[h]	
Typhaceae[h]	
Rapateaceae[f]	
Bromeliaceae[g]	
Commelinaceae[f]	
Pontederiaceae[e]	
Philydraceae[d]	
Haemodoraceae[c]	
Strelitziaceae[b]	
Marantaceae[b]	
Zingiberaceae[b]	
Costaceae[b]	
Musaceae[b]	
Lowiaceae[b]	
Heliconiaceae[b]	
Arecaceae[a]	

[a] Arecanae
[b] Zingiberanae
[c] Bromelianae, Haemodorales
[d] Bromelianae, Philydrales
[e] Bromelianae, Pontederiales
[f] Commelinanae, Commelinales
[g] Bromelianae, Bromeliales
[h] Bromelianae, Typhales
[i] Commelinanae, Poales
[j] Commelinanae, Cyperales

7A

higher eudicots

Trochodendron	Trochodendraceae[f]
Tetracentron	Tetracentraceae[f]
Sabia	Sabiaceae[c]
Lambertia	Proteaceae[e]
Nelumbo	Nelumbonaceae[d]
Platanus	Platanaceae[b]
Akebia	Lardizabalaceae[c]
Mahonia	Berberidaceae[c]
Caulophyllum	
Ranunculus	
Xanthorhiza	Ranunculaceae[c]
Caltha	
Cocculus	Menispermaceae[c]
Euptelea	Eupteleaceae[b]
Dicentra	Fumariaceae[a]
Papaver	Papaveraceae[a]
Sanguinaria	

ham. II   ham. I   ranunculids

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

[a] Magnoliidae, Papaverales
[b] Hamamelidae, Hamamelidales
[c] Magnoliidae, Ranunculales
[d] Magnoliidae, Nymphaeales
[e] Rosidae, Proteales
[f] Hamamelidae, Trochodendrales

FIGURE 7. The basalmost lineages of the eudicots, composed of Ranunculales–Papaverales, Trochodendrales, and a heterogeneous lineage (plus *Gunnera* in B). (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

7B

higher eudicots

Taxon	Family	
Gunnera	Gunneraceae[h]	
Trochodendron	Trochodendraceae[g]	
Tetracentron	Tetracentraceae[g]	
Pachysandra*	Buxaceae[f]	
Sabia	Sabiaceae[c]	
Lambertia	Proteaceae[e]	
Nelumbo	Nelumbonaceae[d]	
Platanus	Platanaceae[b]	
Akebia	Lardizabalaceae[c]	
Mahonia	Berberidaceae[c]	
Caulophyllum		
Ranunculus	Ranunculaceae[c]	
Xanthorhiza		
Caltha		
Cocculus	Menispermaceae[c]	
Euptelea	Eupteleaceae[b]	
Dicentra	Fumariaceae[a]	
Papaver	Papaveraceae[a]	
Sanguinaria		

[a]Magnoliidae, Papaverales
[b]Hamamelidae, Hamamelidales
[c]Magnoliidae, Ranunculales
[d]Magnoliidae, Nymphaeales
[e]Rosidae Proteales
[f]Rosidae, Euphorbiales
[g]Hamamelidae, Trochodendrales
[h]Rosidae, Haloragales

8A

FIGURE 8. The basalmost lineage of rosid dicots, which includes a number of lower hamamelids. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. Note that the positions of this clade and that of the Caryophyllidae (Fig. 9A, B) differ significantly in the results of Searches I and II (see Figs. 1, 2). Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

8B

*Pterostemon** — Grossulariaceae [c]
Itea — Grossulariaceae [c]
Boykinia — Saxifragaceae [c]
Heuchera — Saxifragaceae [c]
Saxifraga — Saxifragaceae [c]
Tetracarpaea — Grossulariaceae [c]
Myriophyllum — Haloragaceae [e]
Penthorum — Saxifragaceae [c]
Paeonia — Paeoniaceae [d]
Sedum — Crassulaceae [c]
Dudleya — Crassulaceae [c]
Kalanchoe — Crassulaceae [c]
Crassula — Crassulaceae [c]
Ribes — Grossulariaceae [c]
Rhodoleia — Hamamelidaceae [a]
Cercidiphyllum — Cercidiphyllaceae [a]
Daphniphyllum — Daphniphyllaceae [b]
Hamamelis
*Altingia** — Hamamelidaceae [a]
*Liquidambar**

← **rosid III**

[a]Hamamelidae, Hamamelidales
[b]Hamamelidae, Daphniphyllales
[c]Rosidae, Rosales
[d]Dilleniidae, Dilleniales
[e]Rosidae, Haloragales

9A

Genus	Family
Dillenia	Dilleniaceae[f]
Dianthus	Caryophyllaceae[e]
Atriplex	Chenopodiaceae[e]
Spinacia	
Amaranthus	Amaranthaceae[e]
Basella	Basellaceae[e]
Portulaca	Portulacaceae[e]
Alluaudia	Didiereaceae[e]
Mollugo	Molluginaceae
Phytolacca	Phytolaccaceae[e]
Mirabilis	Nyctaginaceae[e]
Rivina	Phytolaccaceae[e]
Trianthema	Aizoaceae[e]
Stegnosperma	Phytolaccaceae[e]
Nepenthes	Nepenthaceae[b]
Plumbago	Plumbaginaceae[d]
Rheum	Polygonaceae[c]
Drosera	Droseraceae[b]
Vitis	Vitaceae[a]

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pln
con
cyc

rosid III

[a]Rosidae, Rhamnales
[b]Dilleniidae, Nepenthales
[c]Caryophyllidae, Polygonales
[d]Caryophyllidae, Plumbaginales
[e]Caryophyllidae, Caryophyllales
[f]Dilleniidae, Dilleniales

FIGURE 9. The lineage that includes the Caryophyllidae. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with an asterisk in B were not available for Search I. Note that the positions of this clade and that of rosid IV (Fig. 8A, B) differ significantly in the results of Searches I and II (see Figs. 1, 2). Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

9B

Taxon	Family
Dianthus	Caryophyllaceae [e]
Atriplex	Chenopodiaceae [e]
Spinacia	
Amaranthus	Amaranthaceae [e]
Basella	Basellaceae [e]
Portulaca	Portulacaceae [e]
Alluaudia	Didiereaceae [e]
Mollugo	Molluginaceae [e]
Phytolacca	Phytolaccaceae [e]
Mirabilis	Nyctaginaceae [e]
Rivina	Phytolaccaceae [e]
Trianthema	Aizoaceae [e]
Stegnosperma	Phytolaccaceae [e]
Plumbago	Plumbaginaceae [d]
Rheum	Polygonaceae [c]
Nepenthes	Nepenthaceae [b]
Drosera filiformis	
Drosera spathulata *	
Drosera petiolaris *	Droseraceae [b]
Dionaea *	
Phoradendron	Viscaceae [a]
Schoepfia	Olacaceae [a]
Osyris	Santalaceae [a]

caryophyllids

[a] Rosidae, Santalales
[b] Dilleniidae, Nepenthales
[c] Caryophyllidae, Polygonales
[d] Caryophyllidae, Plumbaginales
[e] Caryophyllidae, Caryophyllales

10A

Gossypium	Malvaceae[j]
Leitneria	Leitneriaceae[i]
Ailanthus	Simaroubaceae[h]
Poncirus	Rutaceae[h]
Bursera	Burseraceae[h]
Acer	Aceraceae[h]
Cupaniopsis	Sapindaceae[h]
Schinus	Anacardiaceae[h]
Tropaeolum	Tropaeolaceae[a]
Brassica	Brassicaceae[e]
Capparis	Capparaceae[e]
Reseda	Resedaceae[e]
Tovaria	Tovariaceae[e]
Batis	Bataceae[g]
Limnanthes	Limnanthaceae[a]
Floerkea	
Carica	Caricaceae[f]
Moringa	Moringaceae[e]
Qualea	Vochysiaceae[d]
Clarkia	
Oenothera	
Epilobium	
Lopezia	Onagraceae[c]
Hauya	
Fuchsia†	
Circaea	
Ludwigia	
Lythrum	Lythraceae[c]
Quisqualis	Combretaceae[c]
Crossosoma	Crossosomataceae[b]
Pelargonium	
Geranium	Geraniaceae[a]
Monsonia	
Greyia	Greyiaceae[b]
Francoa	Saxifragaceae[b]
Viviania	Geraniaceae[a]

rosid II

[a]Rosidae, Geraniales [f]Dilleniidae, Violales

[b]Rosidae, Rosales [g]Dilleniidae, Batales

[c]Rosidae, Myrtales [h]Rosidae, Sapindales

[d]Rosidae, Polygalales [i]Hamamelidae, Leitneriales

[e]Dilleniidae, Capparales [j]Dilleniidae, Malvales

FIGURE 10. One of the two "higher" rosid lineages. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera mar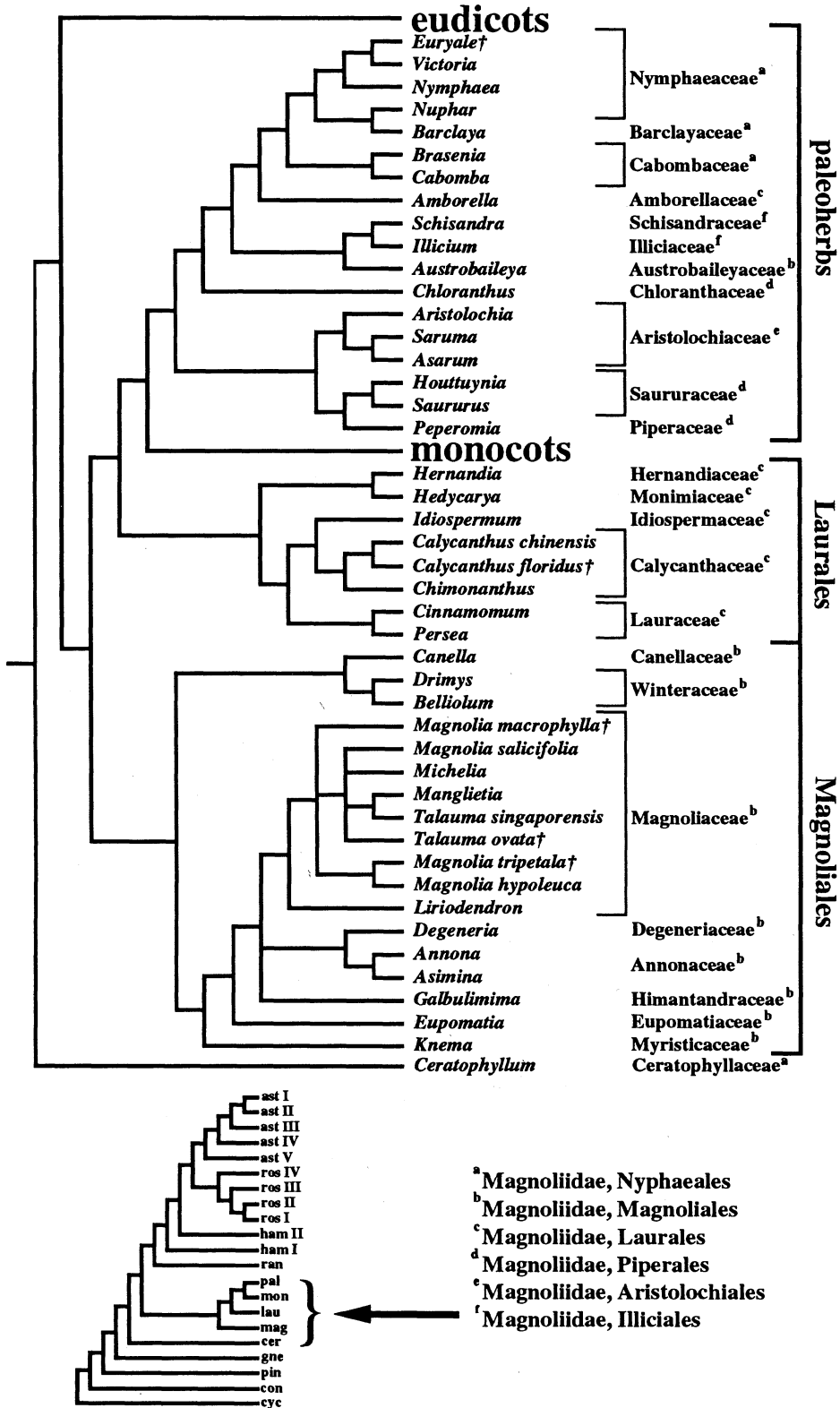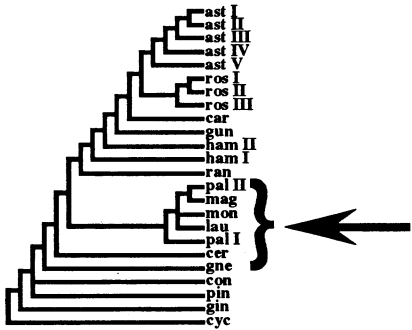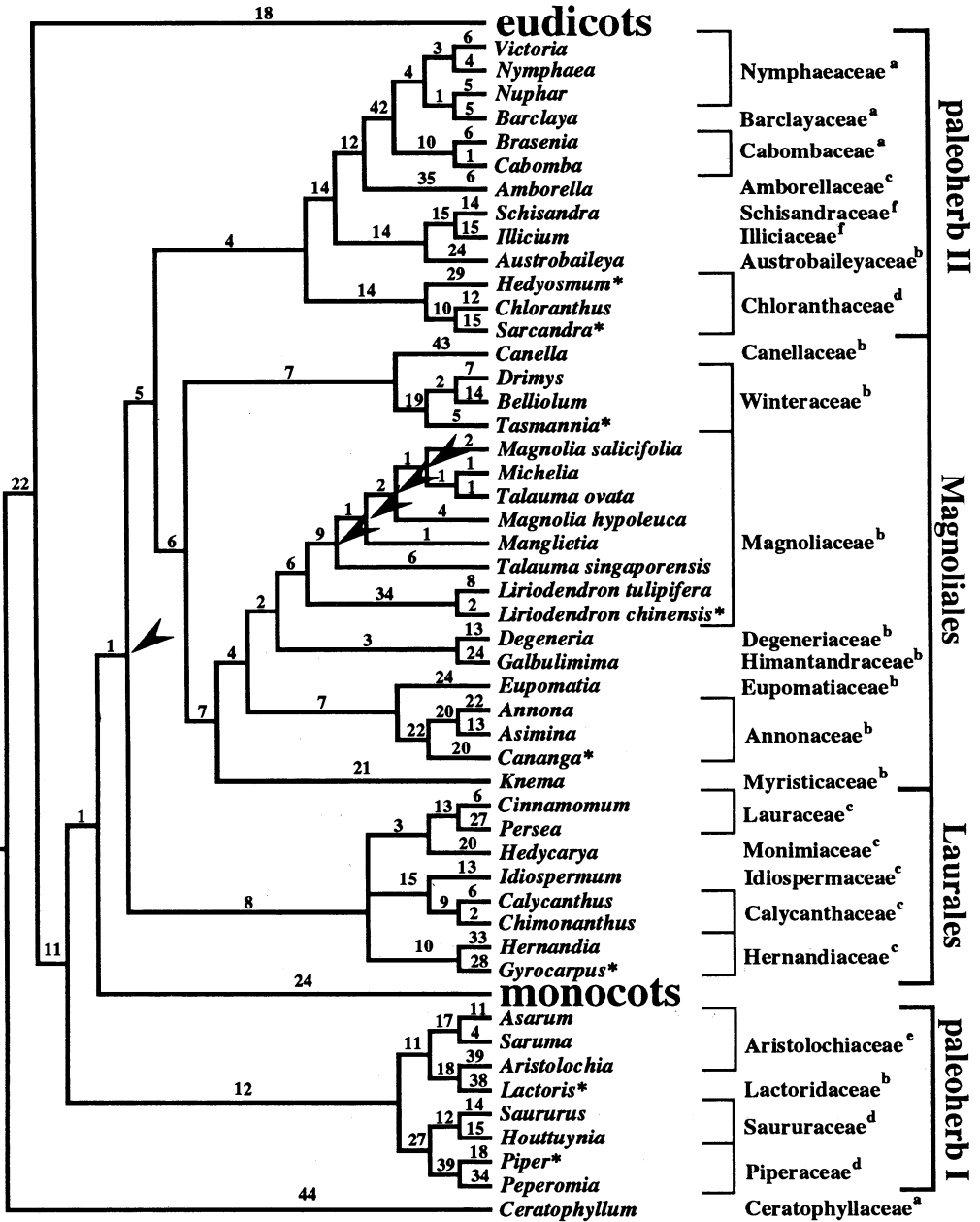ked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

10B

21	*Leitneria*	Leitneriaceae [i]
13 16	*Ailanthus*	Simaroubaceae [c]
15 44	*Poncirus*	Rutaceae [c]
10 21	*Bursera*	Burseraceae [c]
24	21 *Acer*	Aceraceae [c]
17 20	*Cupaniopsis*	Sapindaceae [c]
9 19	*Schinus*	Anacardiaceae [c]
9	5 *Bombax**	Bombacaceae [h]
21 22	*Tilia**	Tiliaceae [h]
8 14	*Theobroma**	Sterculiaceae [h]
14 4 10	*Gossypium*	Malvaceae [h]
13 10	*Thespesia**	
40 12	*Shorea zeylanica**	Dipterocarpaceae [g]
11	*Shorea stipularis**	

7 12	*Moringa*	Moringaceae [d]
32	*Carica*	Caricaceae [f]
20 15	*Stanleya**	Brassicaceae [d]
6 16	*Brassica*	
13 18	*Cleome**	Capparaceae [d]
6 12	*Capparis*	
9 38	*Reseda*	Resedaceae [d]
10 16	*Tovaria*	Tovariaceae [d]
12 29	*Koeberlinia**	Capparaceae [d]
8 56	*Batis*	Bataceae [e]
20 31 14	*Limnanthes*	Limnanthaceae [a]
3	*Floerkea*	
44	*Setchellanthus**	Capparaceae [d]
5 8	*Bretschneidera**	Bretschneideraceae [c]
10 2	*Akania**	Akaniaceae [c]
17	*Tropaeolum*	Tropaeolaceae [a]
25	*Crossosoma*	Crossosomataceae [b]
8 25	*Pelargonium*	
20 31	*Geranium*	Geraniaceae [a]
26 19 27	*Monsonia*	
33	*Hypseocharis**	Oxalidaceae [a]

14		
3 8		

rosid II

[a] Rosidae, Geraniales
[b] Rosidae, Rosales
[c] Rosidae, Sapindales
[d] Dilleniidae, Capparales
[e] Dilleniidae, Batales
[f] Dilleniidae, Violales
[g] Dilleniidae, Theales
[h] Dilleniidae, Malvales
[i] Hamamelidae, Leitneriales

(small cladogram tip labels:)
ast I
ast II
ast III
ast IV
ast V
ros I
ros II
ros III
car
gun
ham II
ham I
ran
pal II
mag
mon
lau
pal I
cer
gne
con
pin
gin
cyc

11A

A phylogenetic tree diagram with the following tip labels and family names:

Genus	Family
Celtis	Ulmaceae[l]
Morus	Moraceae[l]
Rhamnus	Rhamnaceae[n]
Krameria	Krameriaceae[c]
Guaiacum	Zygophyllaceae[m]
Spiraea	Rosaceae[b]
Photinia	Rosaceae[b]
Prunus	Rosaceae[b]
Geum	Rosaceae[b]
Myrica	Myricaceae[o]
Fagus	Fagaceae[j]
Casuarina	Casuarinaceae[k]
Chrysolepis	Fagaceae[j]
Datisca	Datiscaceae[f]
Octomeles	Datiscaceae[f]
Trigonia	[c]
Licania	Chrysobalanaceae[b]
Chrysobalanus	Chrysobalanaceae[b]
Erythroxylum	Erythroxylaceae[h]
Viola	Violaceae[f]
Ochna	Ochnaceae[i]
Drypetes	Euphorbiaceae[g]
Humiria	Humiriaceae[h]
Euphorbia	Euphorbiaceae[g]
Passiflora	Passifloraceae[f]
Acridocarpus	Malpighiaceae[c]
Galphimia	Malpighiaceae[c]
Thryallis	Malpighiaceae[c]
Bunchosia	Malpighiaceae[c]
Mascagnia	Malpighiaceae[c]
Dicella	Malpighiaceae[c]
Byrsonima	Malpighiaceae[c]
Euonymus	Celastraceae[e]
Brexia	Grossulariaceae[b]
Lepuropetalon	Saxifragaceae[b]
Kirengeshoma†	Hydrangeaceae[b]
Parnassia	Saxifragaceae[b]
Pisum	Fabaceae[d]
Medicago	Fabaceae[d]
Albizia	Fabaceae[d]
Bauhinia	Fabaceae[d]
Securidaca	Polygalaceae[c]
Polygala	Polygalaceae[c]
Cephalotus	Cephalotaceae[b]
Platytheca	Tremandraceae[c]
Bauera	Cunoniaceae[b]
Ceratopetalum	Cunoniaceae[b]
Eucryphia	Eucryphiaceae[b]
Oxalis	Oxalidaceae[a]

rosid I

Legend tree with tips: ast I, ast II, ast III, ast IV, ast V, ros IV, ros III, ros II, ros I, ham II, ham I, ran, pal, mon, lau, mag, cer, gne, pin, con, cyc

[a] Rosidae, Geraniales
[b] Rosidae, Rosales
[c] Rosidae, Polygalales
[d] Rosidae, Fabales
[e] Rosidae, Celastrales
[f] Dilleniidae, Violales
[g] Rosidae, Euphorbiales
[h] Rosidae Linales
[i] Dilleniidae, Theales
[j] Hamamelidae, Fagales
[k] Hamamelidae, Urticales
[l] Hamamelidae, Casuarinales
[m] Rosidae, Sapindales
[n] Rosidae, Rhamnales
[o] Hamamelidae, Myricales

FIGURE 11. The second "higher" rosid lineage. (Numbers above the branches in B and C, see foldout, are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B and C were not available for Search I. The taxon labeled "*Kirengeshoma*" was, subsequent to Search I, discovered to be misidentified and was removed from Search II; its identity is unknown. Note that A is the consensus tree of Search I, whereas B and C are a single tree with branches not present in the strict consensus of Search II marked by an arrow.
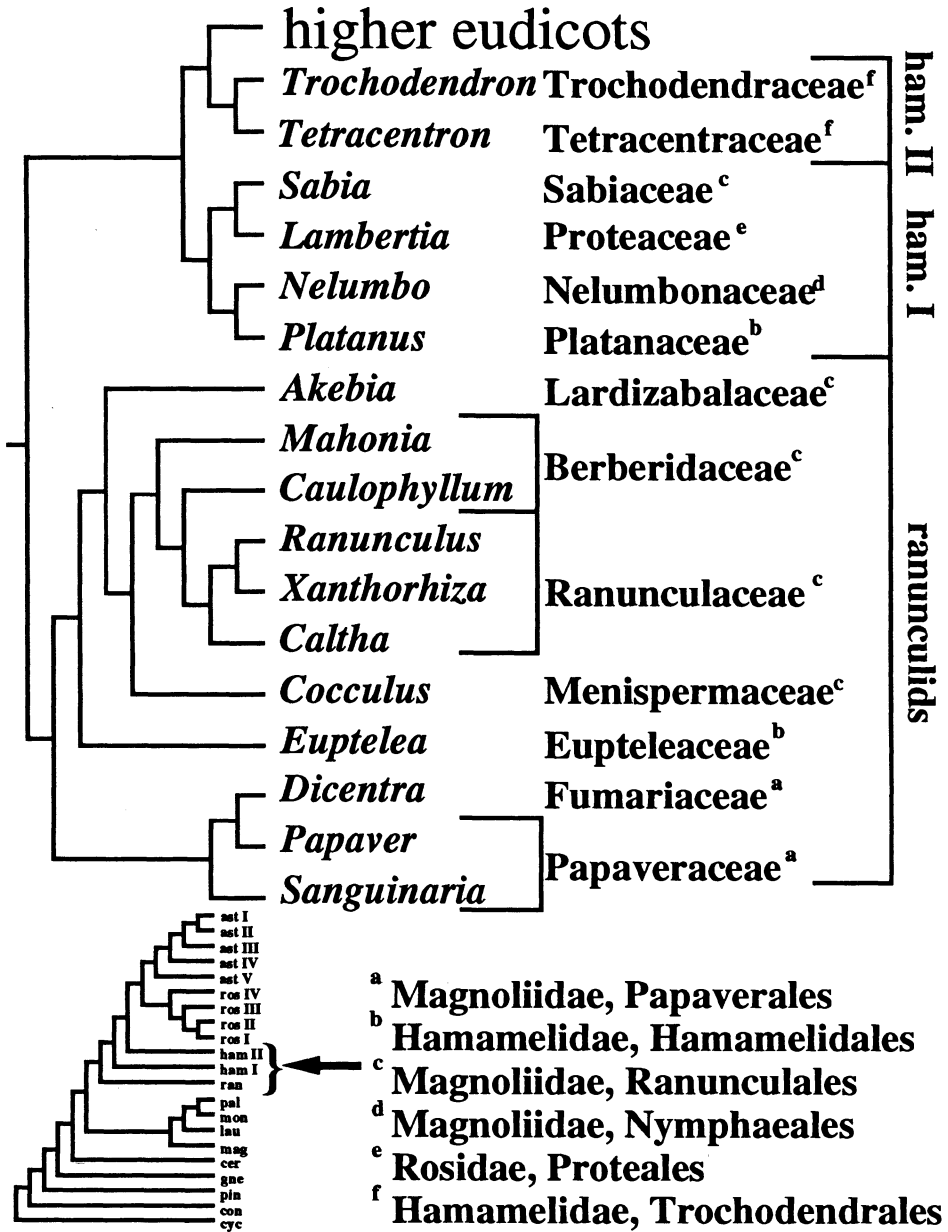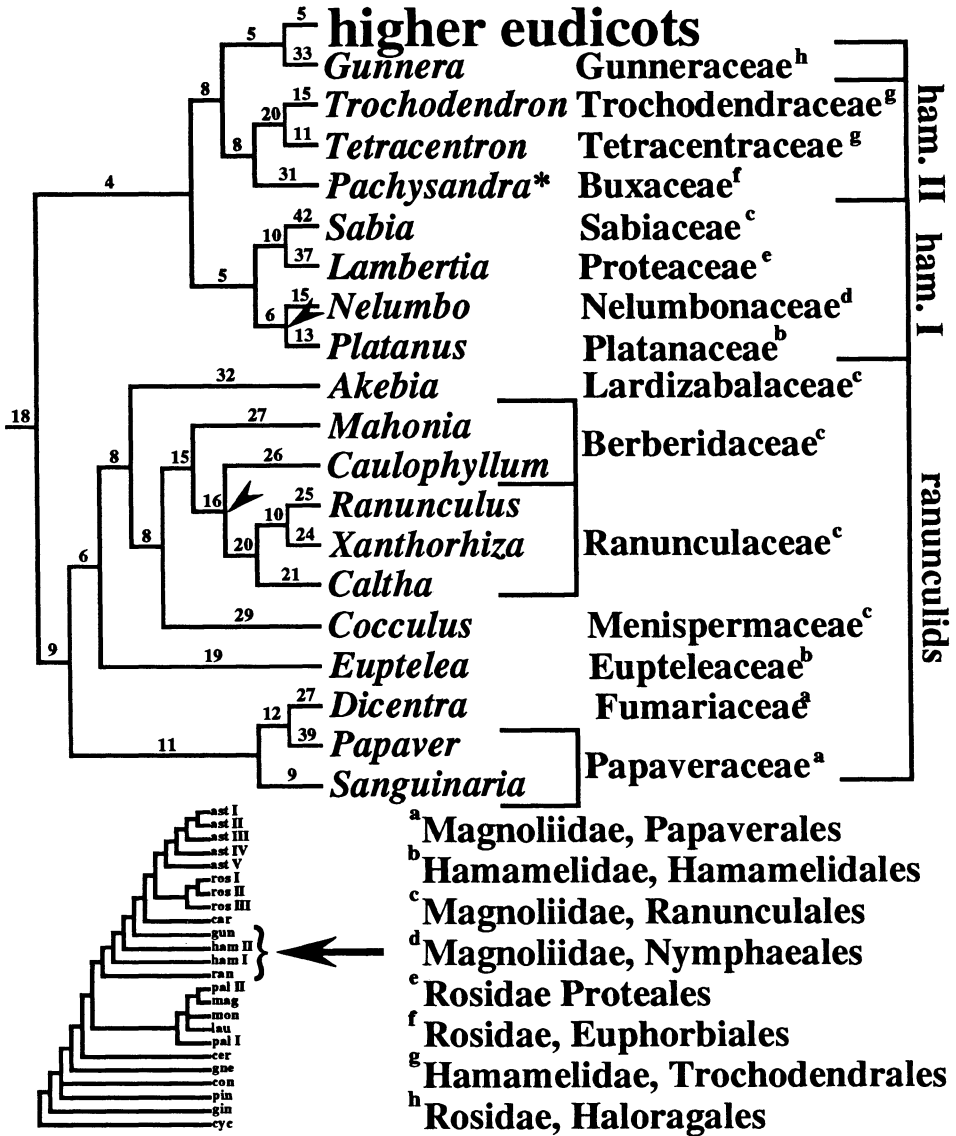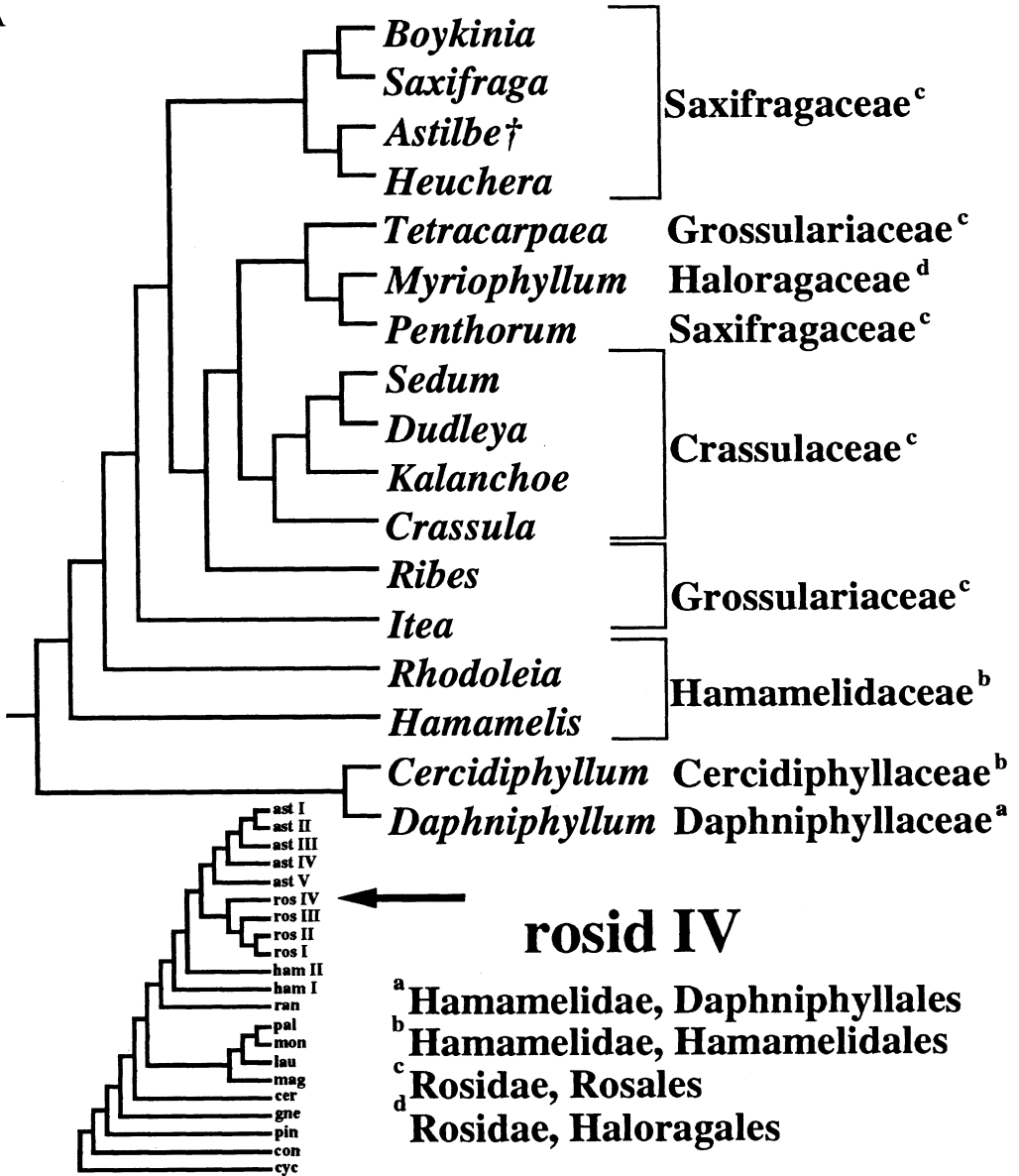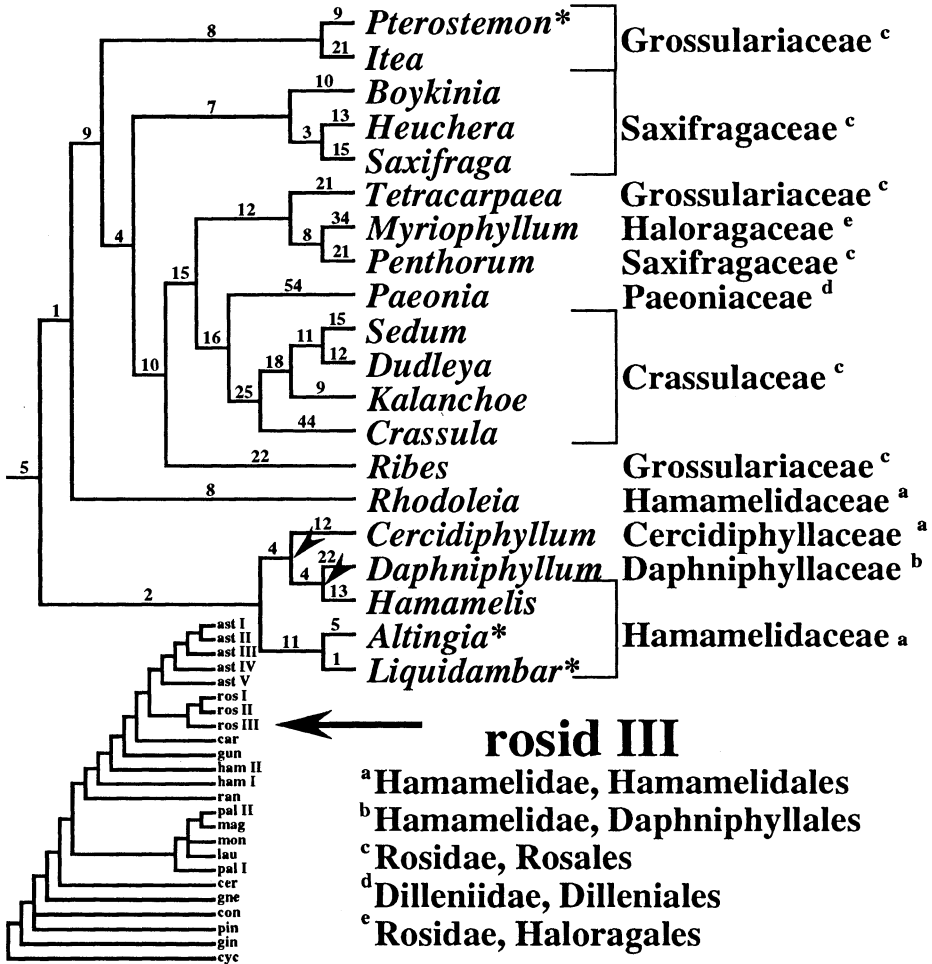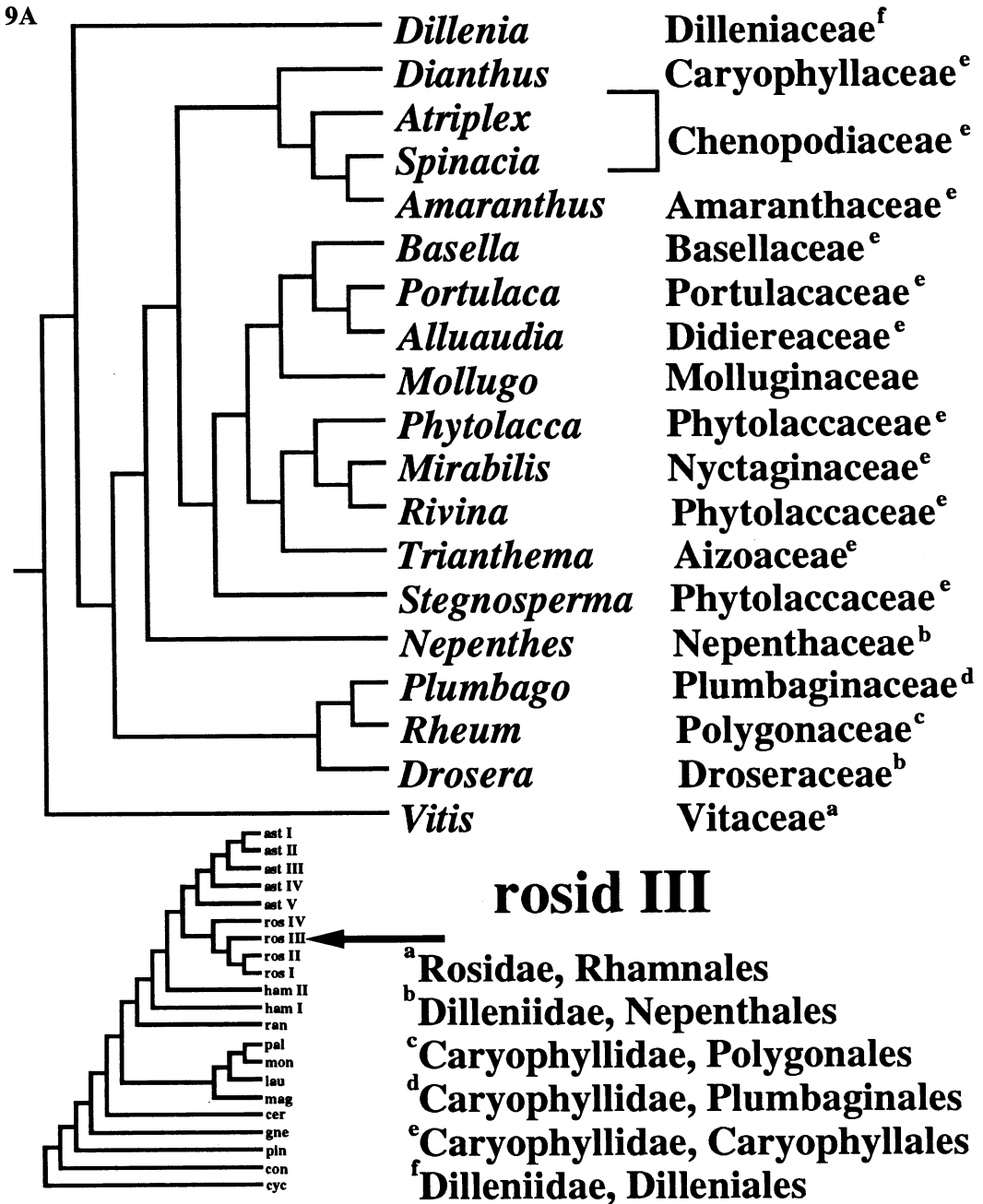
2		other rosid I	
26		*Cephalotus*	Cephalotaceae [b]
25		*Platytheca*	Tremandraceae [d]
21		*Bauera*	Cunoniaceae [b]
8		*Ceratopetalum*	
4		*Eucryphia*	Eucryphiaceae [b]
46		*Oxalis*	Oxalidaceae [a]
12		*Averrhoa**	

other rosid I

Cephalotus — Cephalotaceae [b]
Platytheca — Tremandraceae [d]
Bauera — Cunoniaceae [b]
Ceratopetalum
Eucryphia — Eucryphiaceae [b]
Oxalis — Oxalidaceae [a]
*Averrhoa**

Oenothera — 13
Clarkia — 14
Epilobium — 31
Circaea — 24
Lopezia — 17
Hauya — 10
Ludwigia — 18
Onagraceae [c]

Lythrum — 33 — Lythraceae [c]
*Trapa** — 30 — Trapaceae [c]
*Punica** — 16 — Punicaceae [c]
*Heteropyxis** — 15 — Myrtaceae [c]

Qualea — 31 — Vochysiaceae [d]
*Mouriri** — 26 — Melastomataceae [c]
*Osbeckia** — 40
Quisqualis — 14 — Combretaceae [c]
*Terminalis** — 11

Greyia — 4 — Greyiaceae [b]
Francoa — 8 — Saxifragaceae [b]
Viviania — 29 — Geraniaceae [a]
*Wendtia** — 56

3 — rosid II

ast I
ast II
ast III
ast IV
ast V
ros I
ros II
ros III
car
gun
ham II
ham I
ran
pal II
mag
mon
lau
pal I
cer
gne
con
pin
gin
cyc

rosid I (in part)

[a] Rosidae, Geraniales
[b] Rosidae, Rosales
[c] Rosidae, Myrtales
[d] Rosidae, Polygalales

11B

rosid I (in part)

[a] Hamamelidae, Fagales
[b] Hamamelidae, Juglandales
[c] Hamamelidae, Casuarinales
[d] Hamamelidae, Urticales
[e] Dilleniidae, Violales
[f] Magnoliidae, Ranunculales
[g] Rosidae, Rosales
[h] Rosidae, Sapindales

[i] Rosidae, Polygalales
[j] Rosidae, Rhamnales
[k] Hamamelidae, Myricales
[l] Rosidae, Fabales
[m] Rosidae, Celastrales
[n] Rosidae, Euphorbiales
[o] Rosidae Linales
[p] Dilleniidae, Theales

12A

FIGURE 12.   The two basalmost lineages of the general asterid clade. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Note the different composition of asterid V in Searches I and II. Species marked with a "†" in A were omitted from Search II. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.
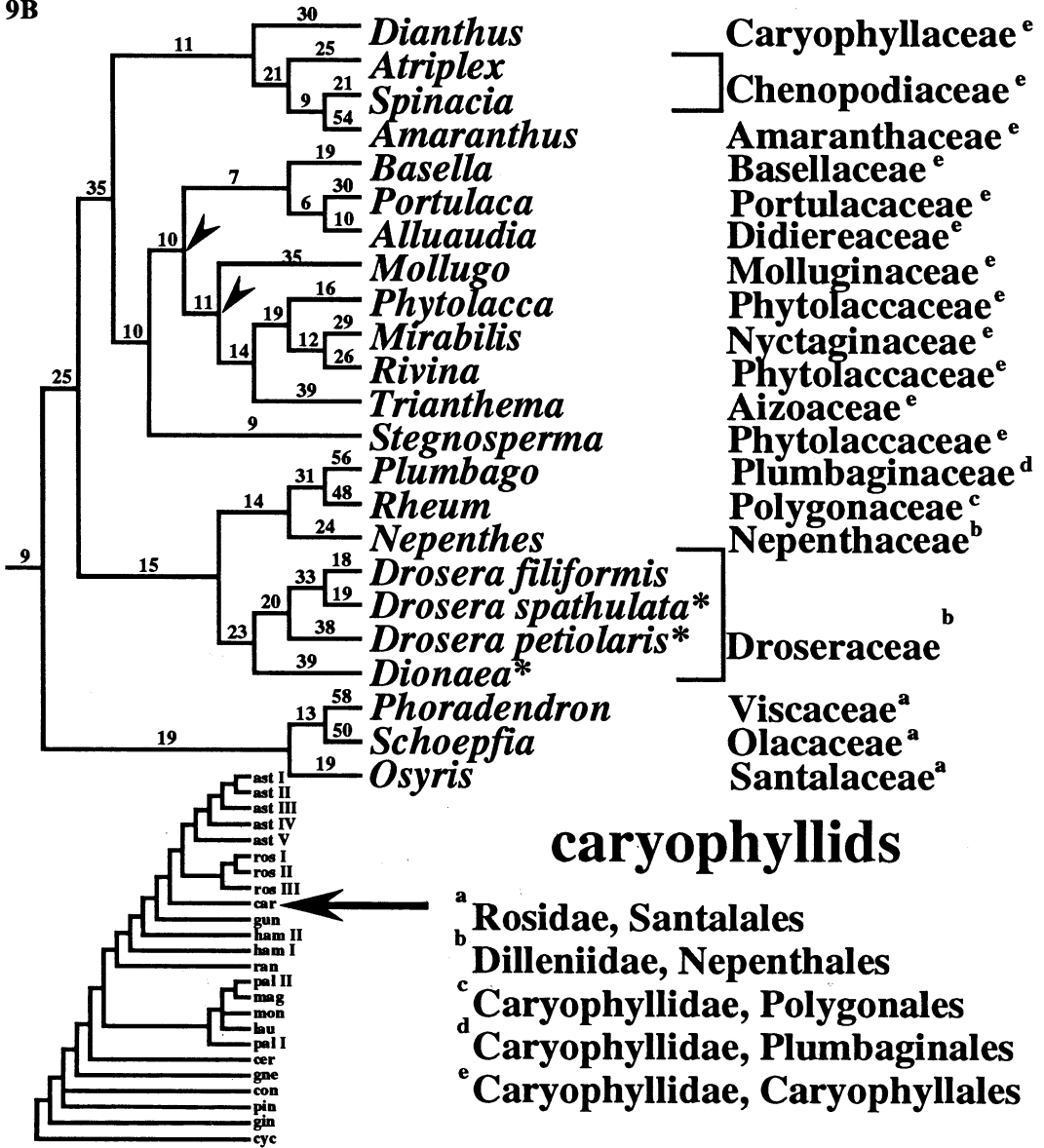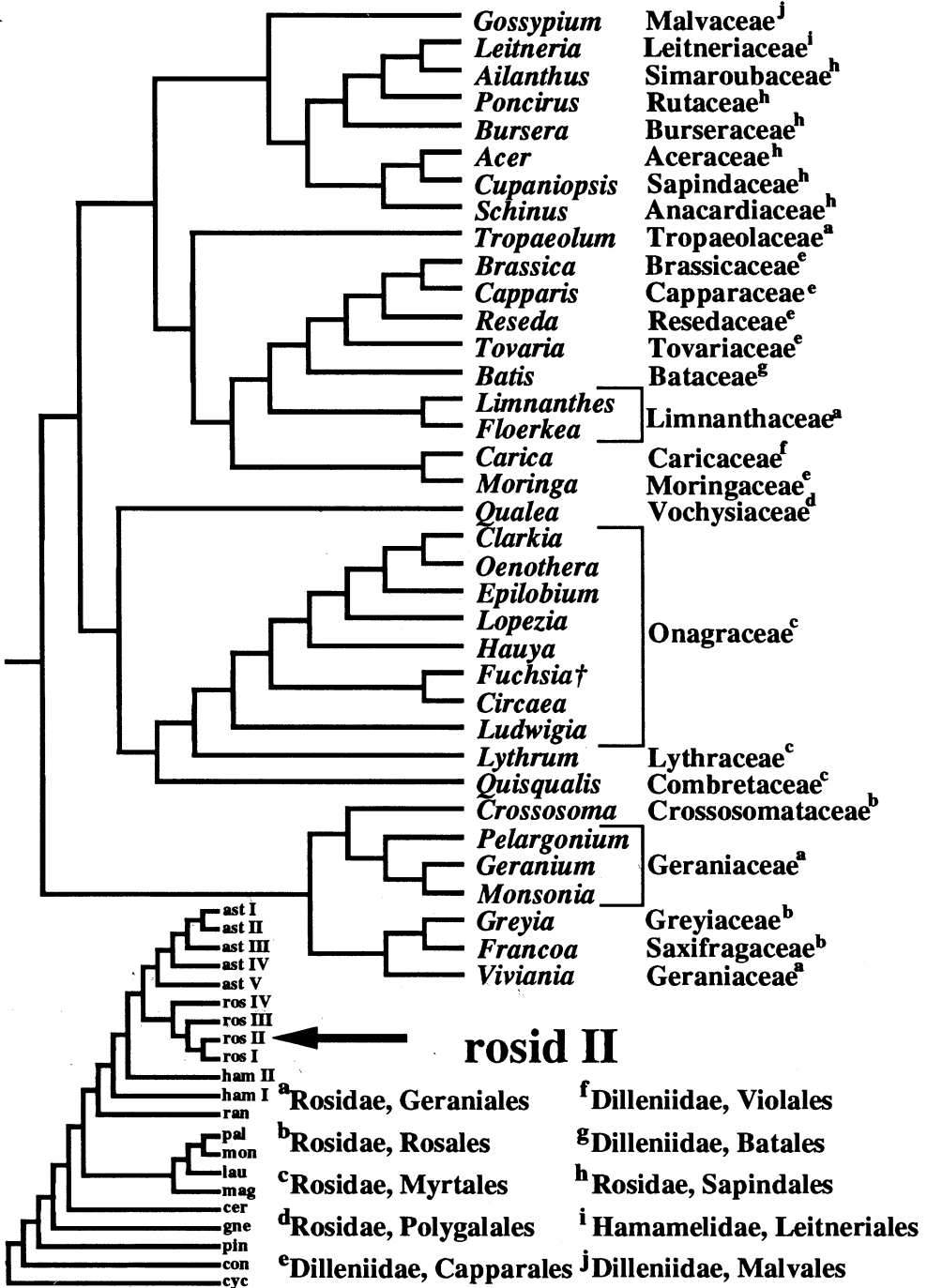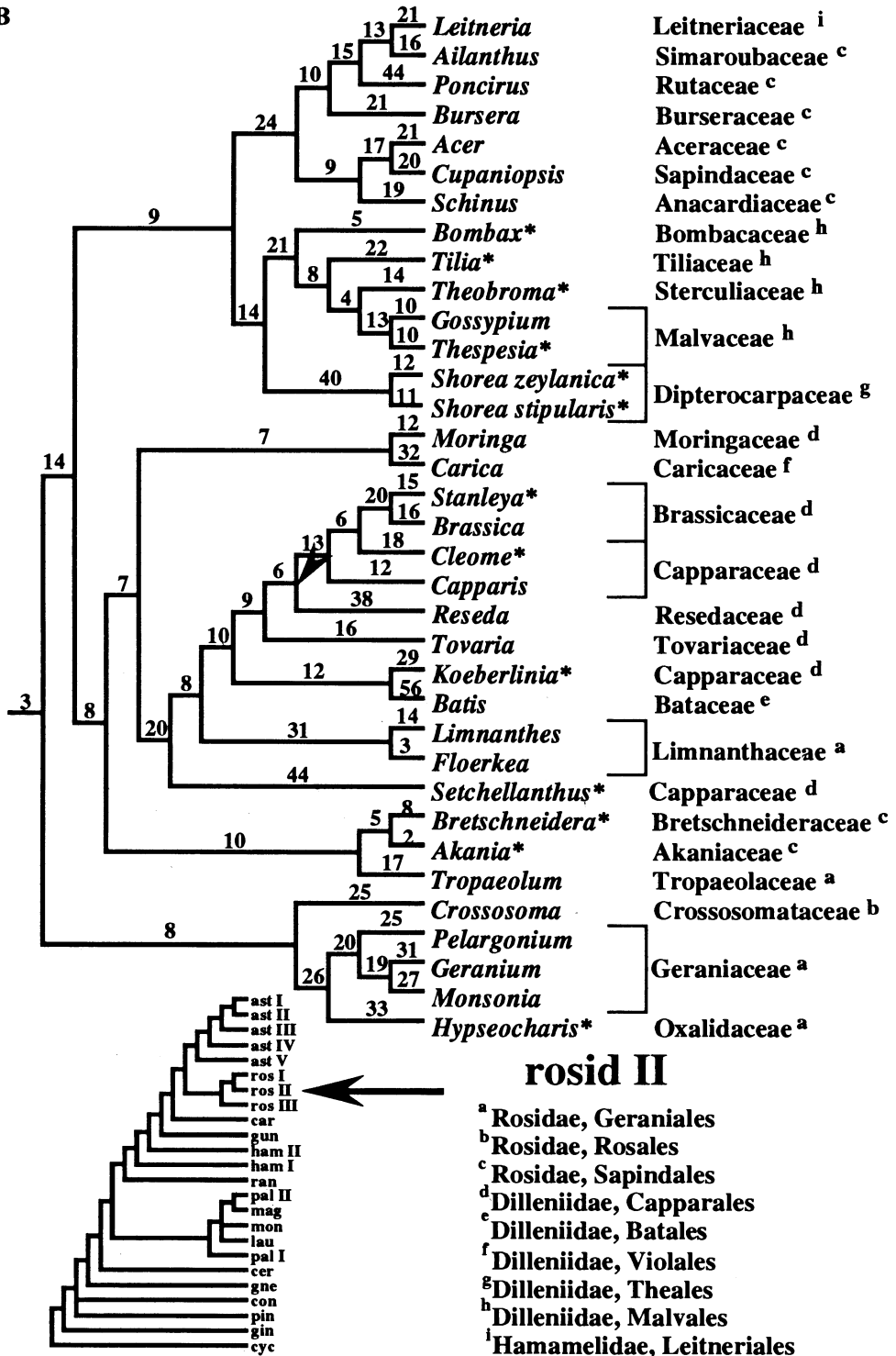
12B

higher asterids

Branch	Taxon	Family
26	*Cornus kousa*	
26	*Cornus canadensis*	Cornaceae[d]
11	*Cornus walteri*	
13	*Cornus florida*	
27	*Alangium*	Alangiaceae[d]
11	*Nyssa*	Nyssaceae[d]
9	*Davidia*	
7	*Diplopanax*	Araliaceae[e]
10	*Camptotheca*	Nyssaceae[d]
12	*Decumaria*	
12	*Hydrangea*	
8	*Carpenteria*	Hydrangeaceae[c]
11	*Philadelphus*	
21	*Deutzia*	
43	*Dillenia*	Dilleniaceae[b]
35	*Vitis*	Vitaceae[a]

asterid IV

asterid V

ast I
ast II
ast III
ast IV
ast V
ros I
ros II
ros III
car
gun
ham II
ham I
ran
pal II
mag
mon
lau
pal I
cer
gne
con
pin
gin
cyc

[a] Rosidae, Rhamnales
[b] Dilleniidae, Dilleniales
[c] Rosidae, Rosales
[d] Rosidae, Cornales
[e] Rosidae, Apiales

13A

Clavija	Theophrastaceae[j]
Anagallis	Primulaceae[j]
Ardisia	Myrsinaceae[j]
Diospyros	Ebenaceae[b]
Diapensia	Diapensiaceae[i]
Polemonium	Polemoniaceae[h]
Symplocus	Symplocaceae[b]
Impatiens	Balsaminaceae[g]
Camellia	Theaceae[f]
Calluna†	
Erica	
Befaria	Ericaceae[c]
Elliottia†	
Leiophyllum†	
Ceratiola	Empetraceae[c]
Rhododendron	
Cassiope	
Chamaedaphne	
Zenobia†	
Daboecia	Ericaceae[c]
Leucothoë†	
Gaultheria	
Vaccinium	
Cyathodes†	
Pentachondra	
Leucopogon†	Epacridaceae[c]
Dracophylllum†	
Epacris	
Arbutus	Ericaceae[c]
Arctostaphylos†	
Pyrola	Pyrolaceae[c]
Enkianthus	Ericaceae[c]
Actinidia	Actinidiaceae[f]
Sarracenia	Sarraceniaceae[e]
Roridula	Byblidaceae[d]
Cyrilla	Cyrillaceae[c]
Styrax	Styracaceae[b]
Clethra	Clethraceae[c]
Manilkara	
Chrysophyllum	Sapotaceae[b]
Fouquieria	Fouquieriaceae[a]

asterid III

[a] Dilleniidae, Violales
[b] Dilleniidae, Ebenales
[c] Dilleniidae, Ericales
[d] Rosidae, Rosales
[e] Dilleniidae, Nepenthales
[f] Dilleniidae, Theales
[g] Rosidae, Geraniales
[h] Asteridae, Solanales
[i] Dilleniidae, Diapensiales
[j] Dilleniidae, Primulales

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

FIGURE 13. The immediate sister lineage to the clade composed of traditional asterids. (Numbers above the branches are the numbers of substitutions optimized on the general semi-strict consensus tree in A and one tree randomly selected from the 3,900 saved in B.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.
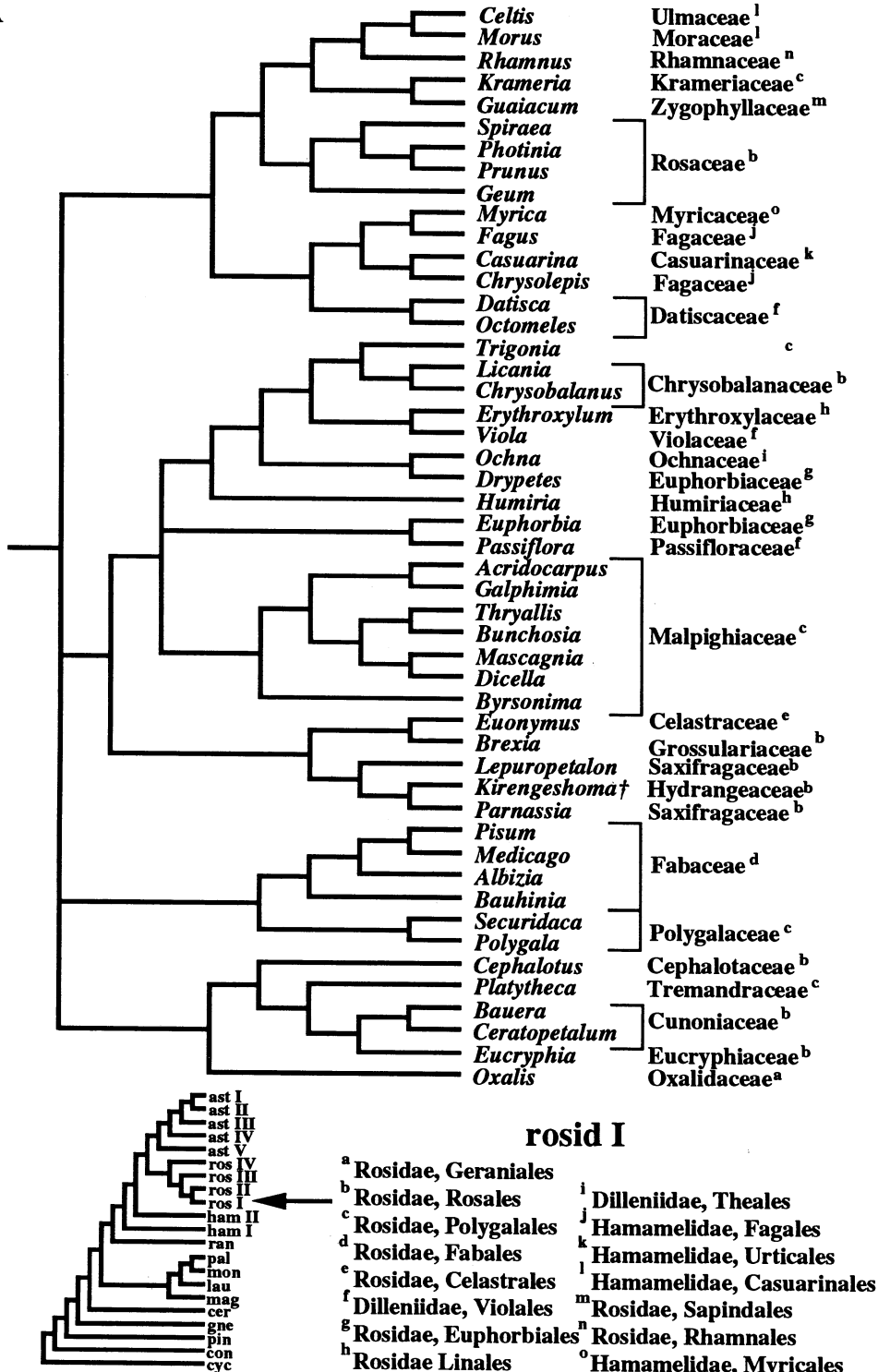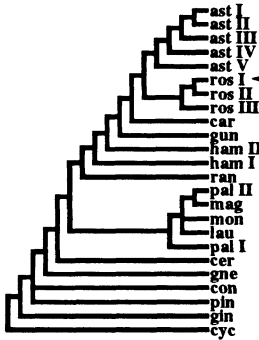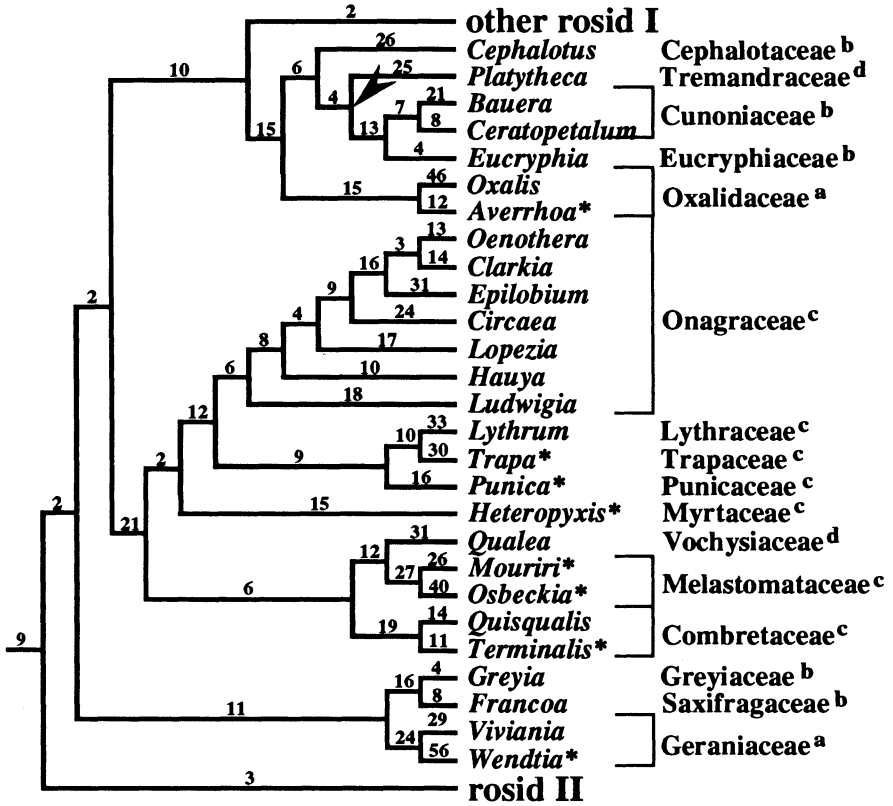
13B
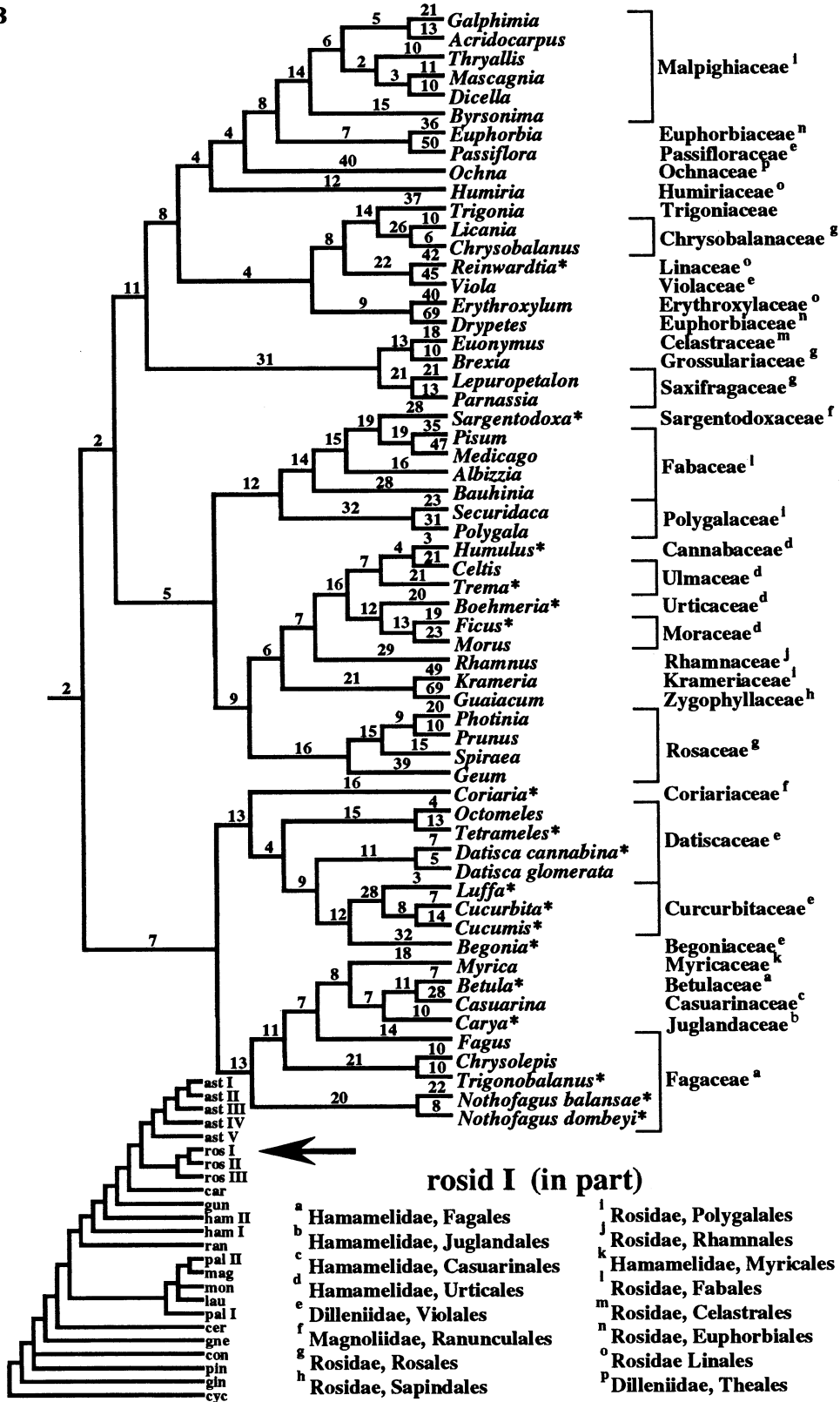
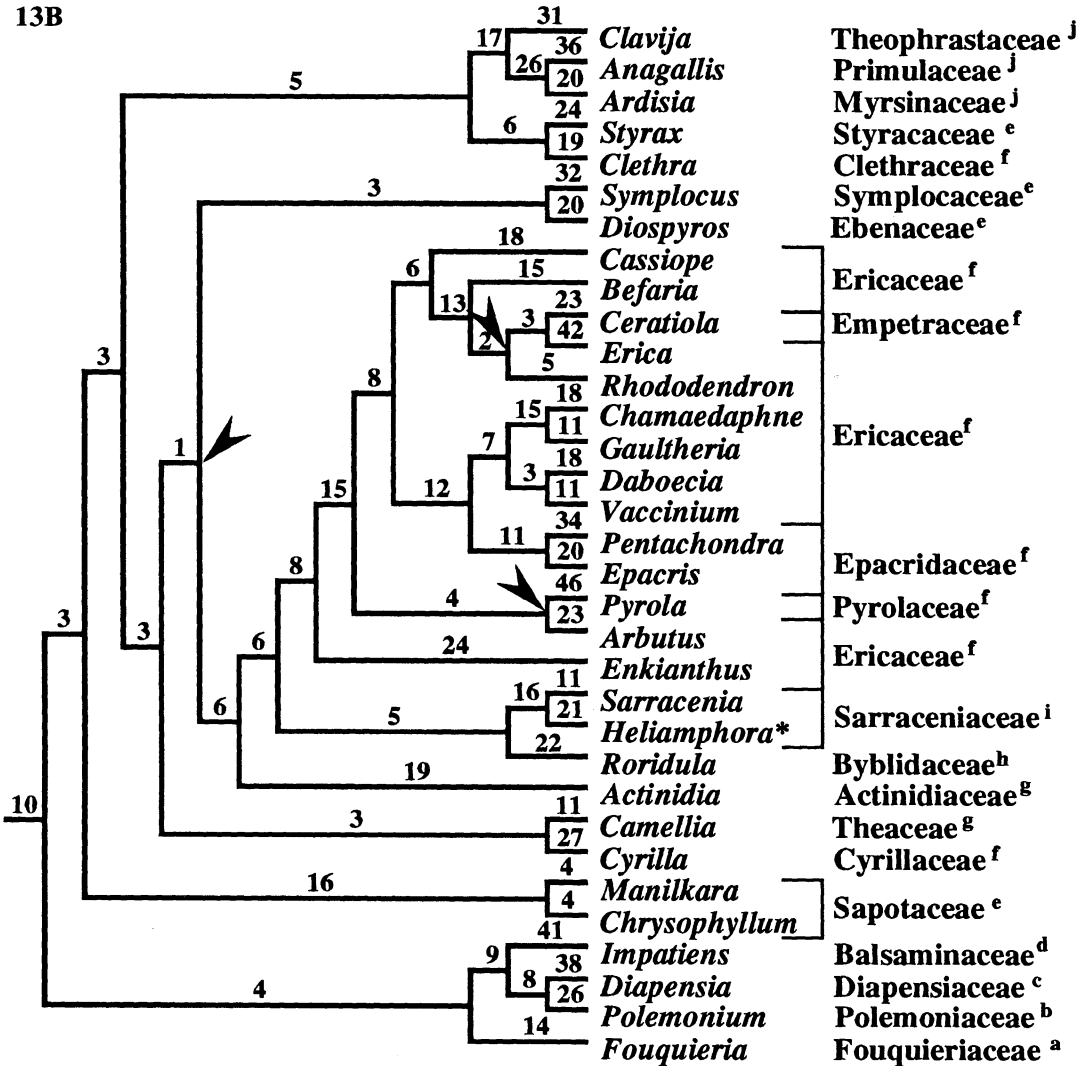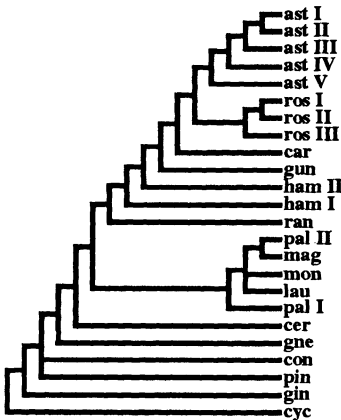| Taxon	Family
Clavija	Theophrastaceae [j]
Anagallis	Primulaceae [j]
Ardisia	Myrsinaceae [j]
Styrax	Styracaceae [e]
Clethra	Clethraceae [f]
Symplocus	Symplocaceae [e]
Diospyros	Ebenaceae [e]
Cassiope	Ericaceae [f]
Befaria	
Ceratiola	Empetraceae [f]
Erica	
Rhododendron	
Chamaedaphne	
Gaultheria	Ericaceae [f]
Daboecia	
Vaccinium	
Pentachondra	
Epacris	Epacridaceae [f]
Pyrola	Pyrolaceae [f]
Arbutus	Ericaceae [f]
Enkianthus	
Sarracenia	Sarraceniaceae [i]
Heliamphora*	
Roridula	Byblidaceae [h]
Actinidia	Actinidiaceae [g]
Camellia	Theaceae [g]
Cyrilla	Cyrillaceae [f]
Manilkara	Sapotaceae [e]
Chrysophyllum	
Impatiens	Balsaminaceae [d]
Diapensia	Diapensiaceae [c]
Polemonium	Polemoniaceae [b]
Fouquieria	Fouquieriaceae [a]

asterid III

[a] Dilleniidae, Violales
[b] Asteridae, Solanales
[c] Dilleniidae, Diapensiales
[d] Rosidae, Geraniales
[e] Dilleniidae, Ebenales
[f] Dilleniidae, Ericales
[g] Dilleniidae, Theales
[h] Rosidae, Rosales
[i] Dilleniidae, Nepenthales
[j] Dilleniidae, Primulales

ast I
ast II
ast III
ast IV
ast V
ros I
ros II
ros III
car
gun
ham II
ham I
ran
pal II
mag
mon
lau
pal I
cer
gne
con
pin
gin
cyc

14A

Symphoricarpos   Caprifoliaceae[i]
Valeriana   Valerianaceae[i]
Dipsacus   Dipsacaceae[i]
Sambucus   Caprifoliaceae[i]
Adoxa   Adoxaceae[i]
Viburnum   Caprifoliaceae[i]
Boöpis   Calyceraceae[h]
Lobelia   Campanulaceae[g]
Campanula
Scaevola   Goodeniaceae[g]
Stokesia†
Piptocarpha
Vernonia
Gerbera
Tragopogon
Cichorium
Lactuca†
Gazania†
Cacosmia
Echinops†
Carthamnus
Dimorphotheca
Felicia
Achillea   Asteraceae[f]
Chrysanthemum†
Blennospermum†
Senecio
Tagetes
Flaveria†
Chromolaena
Coreopsis†
Eupatorium
Helianthus†
Barnadesia†
Dasyphyllum†
Villarsia
Menyanthes   Menyanthaceae[e]
Corokia   Cornaceae[c]
Hedera
Aralia   Araliaceae[d]
Pittosporum   Pittosporaceae[b]
Coriandrum
Apium
Conium†   Apiaceae[d]
Sanicula
Griselinia   Cornaceae[c]
Berzelia   Bruniaceae[b]
Escallonia   Grossulariaceae[b]
Helwingia   Cornaceae[c]
Phyllonoma   Grossulariaceae[b]
Ilex vomitoria†   Aquifoliaceae[a]
Ilex crenata

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

asterid II

[a] Rosidae, Celastrales
[b] Rosidae, Rosales
[c] Rosidae, Cornales
[d] Rosidae, Apiales
[e] Asteridae, Solanales
[f] Asteridae, Asterales
[g] Asteridae, Campanulales
[h] Asteridae, Calycerales
[i] Asteridae, Dipsacales

14B

FIGURE 14.   One of the two clades of traditional Asteridae. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.
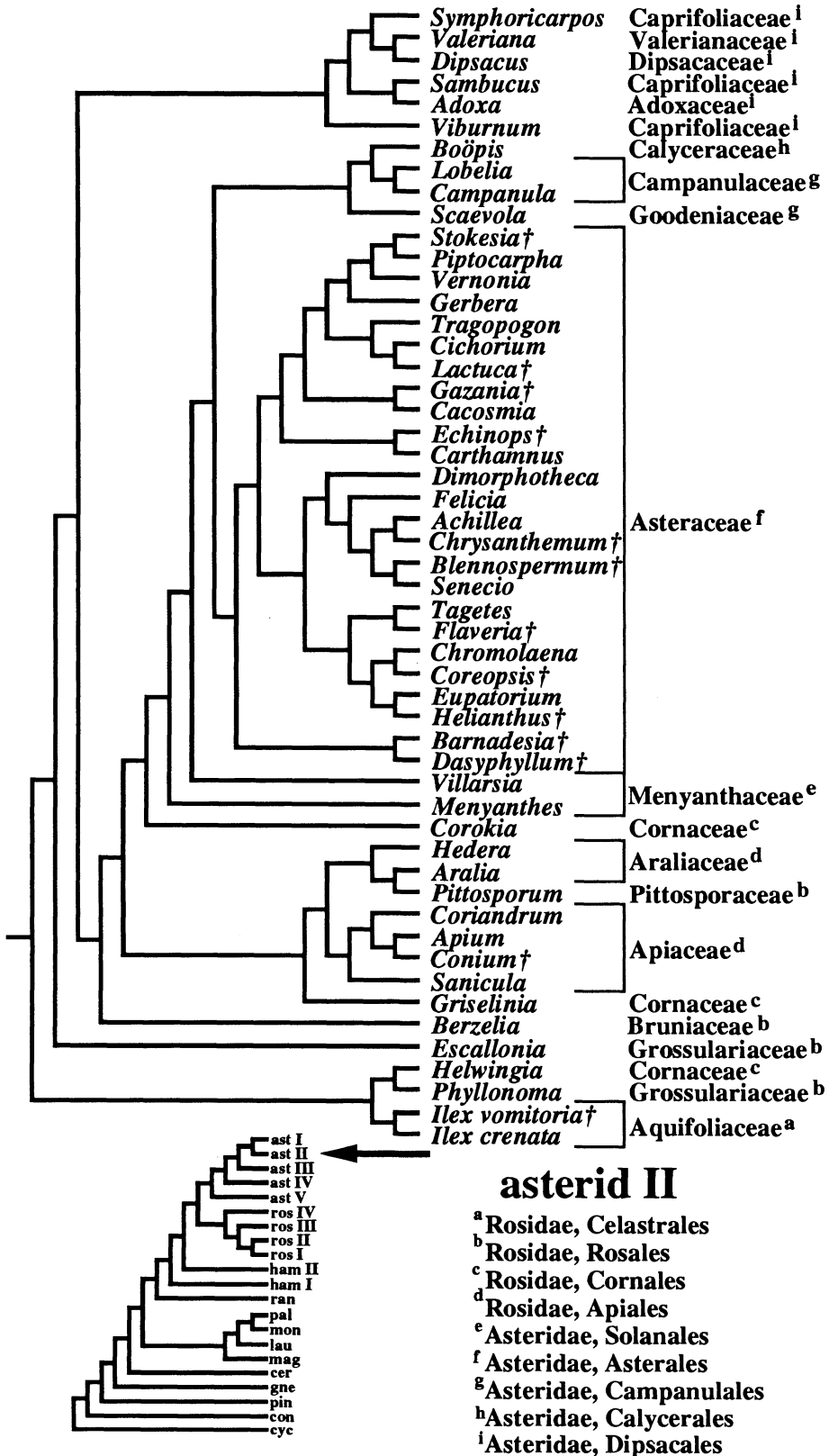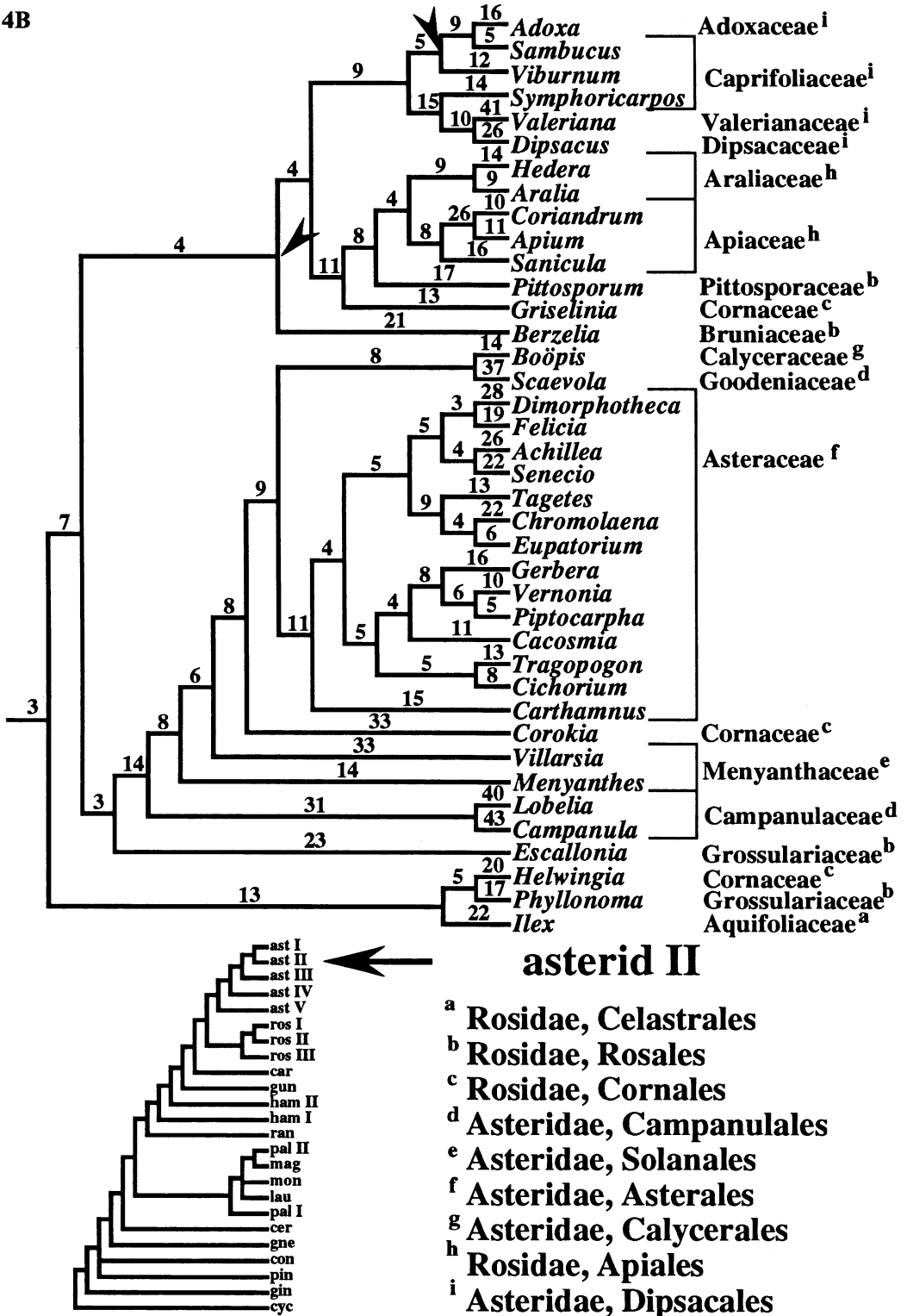
15A

Nicotiana
Lycopersicon
Petunia   Solanaceae[c]
Convolvulus
Ipomoea   Convulvulaceae[c]
Montinia   Grossulariaceae[i]
Borago
Heliotropium   Boraginaceae[g]
Hydrophyllum
Eriodictyon   Hydrophyllaceae[c]
Callitriche   Callitrichaceae[h]
Digitalis
Antirrhinum   Schrophulariaceae[f]
Nelsonia   Nelsoniaceae[f]
Justicia americana†
Justicia odora
Hypoëstes
Ruttya   Acanthaceae[f]
Aphelandra
Acanthus
Thunbergia   Thunbergiaceae[f]
Phaulopsis†
Ruellia
Barleria   Acanthaceae[f]
Lepidagathis
Sesamum   Pedaliaceae[f]
Verbena   Verbenaceae[g]
Teucrium   Lamiaceae[g]
Clerodendrum   Verbenaceae[g]
Prostanthera
Scutellaria
Salvia   Lamiaceae[g]
Glechoma†
Physostegia
Pogostemon
Callicarpa   Verbenaceae[g]
Harpagophytum   Pedaliaceae[f]
Proboscidea
Buddleja   Buddlejaceae[f]
Catalpa   Bignoniaceae[f]
Streptocarpus   Gesneriaceae[f]
Pinguicula
Utricularia   Lentibulariaceae[f]
Jasminum
Ligustrum   Oleaceae[f]
Gentiana
Exacum   Gentianaceae[e]
Apocynum   Apocynaceae[e]
Asclepias   Asclepiadaceae[e]
Spigelia
Gelsemium   Loganiaceae[e]
Pentas   Rubiaceae[d]
Hydrolea†   Hydrophyllaceae[c]
Aucuba   Cornaceae[b]
Garrya   Garryaceae[b]
Eucommia   Eucommiaceae[a]

ast I
ast II
ast III
ast IV
ast V
ros IV
ros III
ros II
ros I
ham II
ham I
ran
pal
mon
lau
mag
cer
gne
pin
con
cyc

asterid I

[a] Hamamelidae, Eucommiales
[b] Rosidae, Cornales
[c] Asteridae, Solanales
[d] Asteridae, Rubiales
[e] Asteridae, Gentianales
[f] Asteridae, Schrophulariales
[g] Asteridae, Lamiales
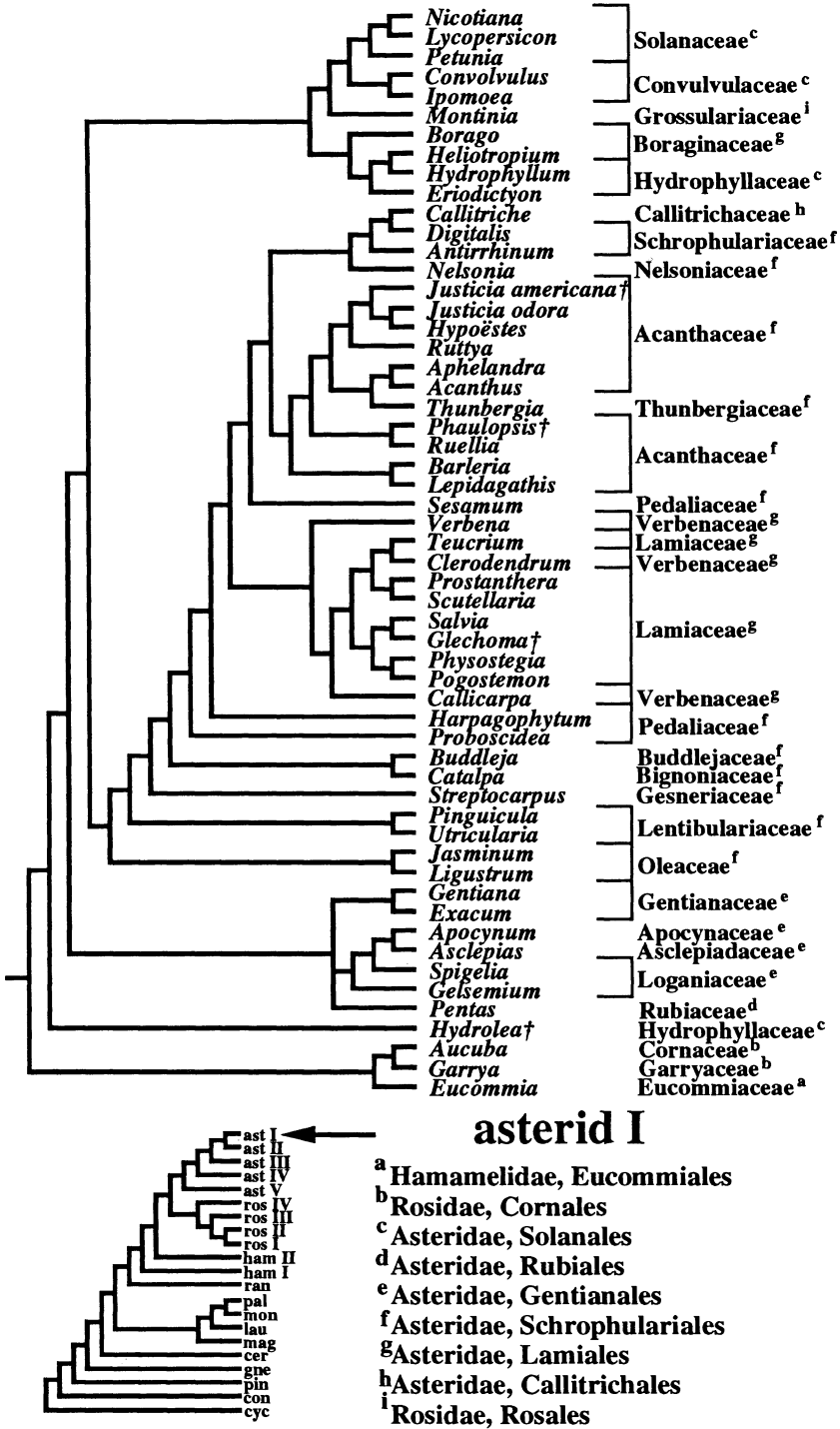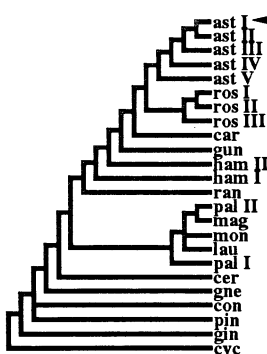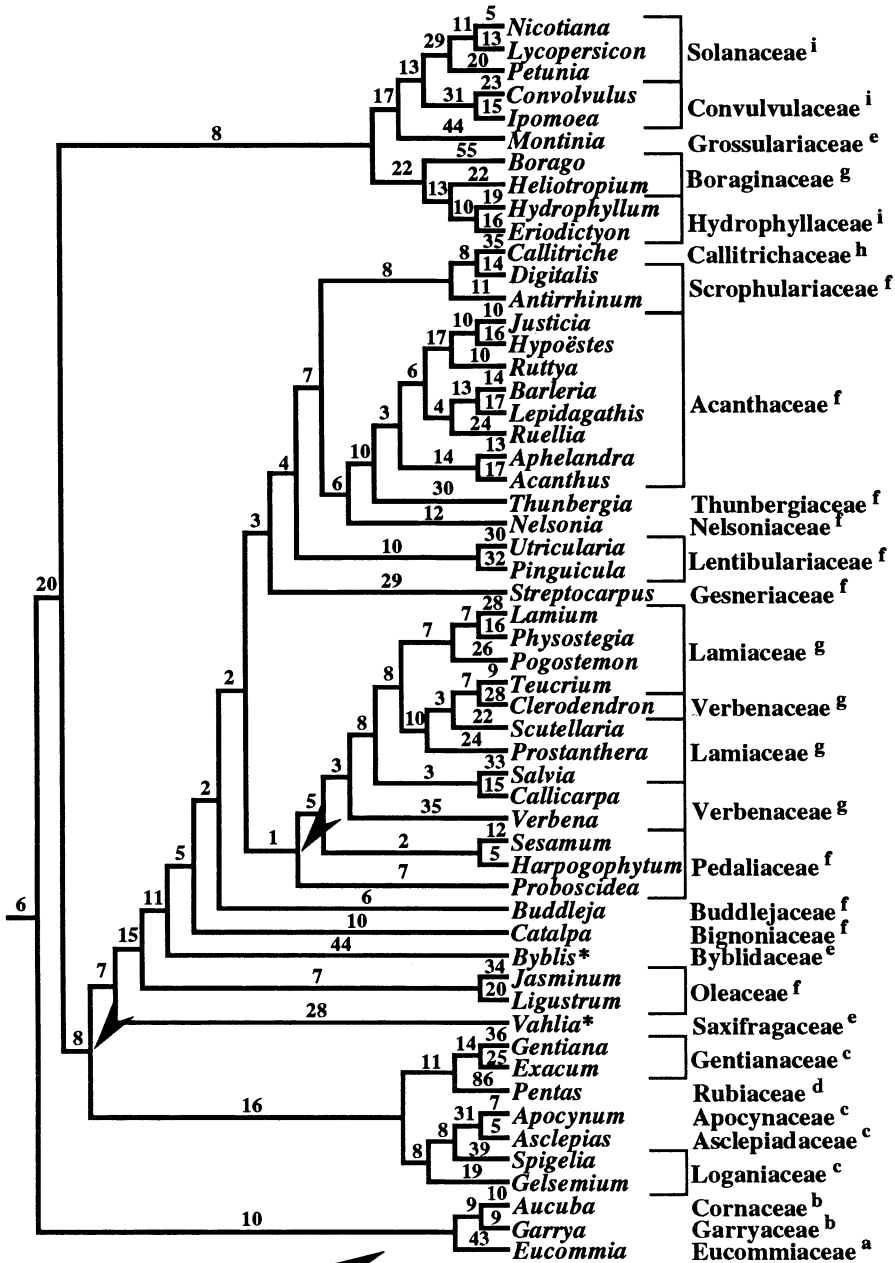[h] Asteridae, Callitrichales
[i] Rosidae, Rosales

FIGURE 15. The second clade of traditional Asteridae. (Numbers above the branches in B are the numbers of substitutions optimized onto one tree randomly selected from the 3,900 saved in Search II.) Genera marked with a "†" in A were omitted from Search II; genera marked with an asterisk in B were not available for Search I. Note that A is the consensus tree of Search I, whereas B is a single tree with branches not present in the strict consensus of Search II marked by an arrow.

15B

asterid I

- [a] Hamamelidae, Eucommiales
- [b] Rosidae, Cornales
- [c] Asteridae, Gentianales
- [d] Asteridae, Rubiales
- [e] Rosidae, Rosales
- [f] Asteridae, Schrophulariales
- [g] Asteridae, Lamiales
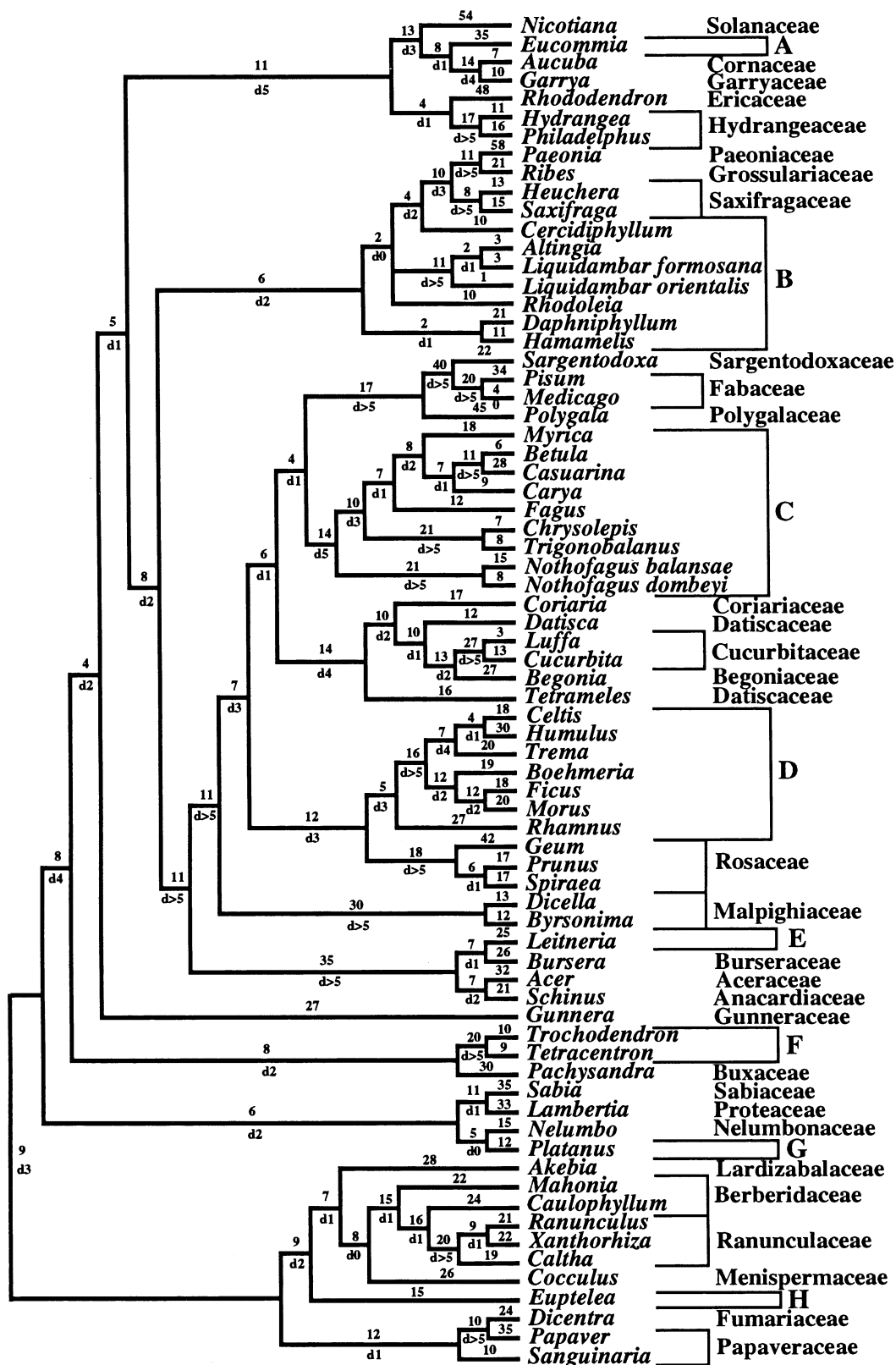- [h] Asteridae, Callitrichales
- [i] Asteridae, Solanales

FIGURE 16. An example of using the general analysis to focus a narrower study of internal support for a polyphyletic Hamamelidae. Numbers above the horizontal lines indicate the number of substitutions optimized (ACCT-RAN) onto one of the most parsimonious trees found using character-state weighting (i.e., these are Fitch steps, equal weighting). The numbers below the horizontal lines (preceded with a "d") are the number of steps less parsimonious at which a branch becomes a polytomy with the branch interior to it. Branches that are not present in the strict consensus of most-parsimonious Fitch trees are indicated by "d0," signifying that they "decay" at maximum parsimony. Groups of taxa often considered to be members of Hamamelidae are bracketed and lettered to designate putatively independent lineages (for further explanation see text).