

Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants

MARI KÄLLERSJÖ, JAMES S. FARRIS, MARK W. CHASE, BIRGITTA BREMER,
MICHAEL F. FAY, CHRISTOPHER J. HUMPHRIES, GITTE PETERSEN,
OLE SEBERG, and KÅRE BREMER

Received October 16, 1997; in revised version March 16, 1998

Key words: Parsimony jackknifing, phylogenetic analysis, large data sets, *rbcL*, DNA sequences, green plants, land plants, seed plants, flowering plants.

Abstract: The ever-larger data matrices resulting from continuing improvements in DNA sequencing techniques require faster and more efficient methods of phylogenetic analysis. Here we explore a promising new method, parsimony jackknifing, by analyzing a matrix comprising 2538 sequences of the chloroplast gene *rbcL*. The sequences included cover a broad taxonomic range, from cyanobacteria to flowering plants. Several parsimony jackknife analyses were performed, both with and without branch-swapping and multiple random addition sequences: 1) including all positions; 2) including only first and second codon positions; 3) including only third positions; and 4) using only transversions. The best resolution was obtained using all positions. Removal of third positions or transitions led to massive loss of resolution, although using only transversions somewhat improved basal resolution. While branch-swapping improved both resolution and the support found for several groups, most of the groups could be recovered by faster simple analyses. Designed to eliminate groups poorly supported by the data, parsimony jackknifing recognizes 1400 groups on the basis of all *rbcL* positions. These include major taxa such as green plants, land plants, flowering plants, monocots and eudicots. We include appendices of supported angiosperm families, as well as larger groups.

The DNA sequence of the chloroplast encoded *rbcL* gene is the most widely used for phylogenetic reconstruction of plants. The gene codes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, commonly abbreviated RuBisCo, the enzyme responsible for carbon dioxide fixation in photosynthesis. After its initial use for phylogenetic reconstruction a decade ago (RITLAND & CLEGG 1987, ZURAWSKI & CLEGG 1987), it soon became the first choice in plant molecular systematics and today thousands of species have been sequenced. The best-known study is the analysis of 500 *rbcL* sequences compiled by a team of 42 systematists (CHASE & al. 1993). This and a number of simultaneously published studies

highlighted the problems of analyzing large data sets (e.g. DUVAL & al. 1993, OLMSTEAD & al. 1993).

Algorithms for finding most parsimonious trees, as implemented in Hennig 86 (FARRIS 1988), MEGA (KUMAR & al. 1993), NONA (GOLOBOFF 1993) and PAUP (SWOFFORD 1993), were developed for working on small data sets. For the ever-larger matrices that are now being produced, they are inefficient and unacceptably time-consuming. Since only superficial searches are possible with these programs, the resulting trees are often far from optimal, and the consensus is likely to be misleadingly well resolved, being based on only a few of the many possible solutions.

One of the problems is the frequent occurrence of multiple solutions, sometimes so numerous that they exceed what is possible to store and evaluate with available computer technology. This problem was encountered in the 500-sequence analysis by CHASE & al. (1993), which suffered from 'unquestionable absence of many trees at the same level of optimality'. Another problem in phylogenetic reconstruction is the assessment of support for individual clades. Two commonly used procedures for estimating branch support are bootstrapping (FELSENSTEIN 1985) and Bremer support analysis (BREMER 1988, 1994; KÄLLERSJÖ & al. 1992), but they are time-consuming and unrealistic for large data sets given currently available software.

Until recently, the only available solution to these problems has been to reduce the size of the data matrix, i.e. to discard information. A smaller sample of taxa can be selected to save time, or an attempt can be made to make structure more obvious by a priori removing putatively homoplastic characters such as third codon positions or transitions. Discarding information is risky, however. Analyses of data sets with a small number of taxa covering a wide taxonomic range, may give different results depending on exactly which taxa have been chosen (OLMSTEAD & al. 1993, FARRIS & al. 1996). Removing transitions or third positions means fewer informative positions and a potential loss of resolution.

Recently, a more promising approach has been suggested. The jackknifing procedure described by FARRIS & al. (1996) was developed to deal specifically with the problems of analyzing large data sets. The method combines jackknifing with fast, efficient parsimony algorithms, and a new way of storing trees. The multiple trees produced through jackknifing are combined to a single tree with well supported clades only, thereby solving the problem of multiple solutions and providing a fast method for assessment of branch support. This procedure is implemented in the parsimony jackknife program *Jac* (FARRIS 1996), with which it is possible to analyze matrices even larger than the 500-sequence matrix of CHASE & al. (1993).

Here we present a parsimony jackknife analysis of 2538 *rbcL* sequences, possibly the largest phylogenetic analysis ever undertaken. The purpose of this analysis is threefold, (1) to test the efficiency of parsimony jackknifing on an extremely large data set, (2) to explore the possibility of simultaneously analyzing phylogenetic relationships in widely different groups at widely different levels of the taxonomic hierarchy, and (3) to investigate the amount of support from *rbcL* data for different groups at different levels of the phylogeny of green plants, including the land plants, the seed plants, and the flowering plants in particular.

Data and methods

CHASE & al.'s (1993) original data set was used as a starting point. To this matrix were added sequences compiled from GenBank and from unpublished sequences at the Molecular Systematics Laboratories at Royal Botanic Gardens, Kew, the Swedish Museum of Natural History, Stockholm, the Department of Systematic Botany, Uppsala University, and the Botanical Institute, University of Copenhagen. The number of sequences was limited by computer memory, since there was an upper limit to the size of the data matrix that could be processed with the internal computer memory available (in this particular case, 32 Mb). The sequences were aligned using AssemblyLign, Oxford Molecular Group, Inc. Several sequences from cyanobacteria and green algae had to be omitted due to difficulties in alignment. Only sequences for which alignments were straightforward were included in the final data matrix, which comprises 2538 sequences, 1428 bp long. It should be noted that several sequences are not complete; many lack bases at the 5', or the 3' end. These absent bases have been replaced by the IUPAC code "N". The sequences have not been scrutinized for errors or stop codons. In some cases we have received the same sequence from different sources, e. g. sequences included in the original matrix of CHASE & al. (1993) were later extracted also from GenBank. Identical, superfluous sequences have been removed, but in a few cases there have been minor discrepancies between the two versions, and we have then retained both in the matrix.

Two versions of the *Jac* program (FARRIS 1996) were used. Both versions randomly delete a fraction of e^{-1} of the characters, perform a quick heuristic parsimony search, and reiterate the two steps a specified number of times (FARRIS & al. 1996). The original version uses a fast search, without branch-swapping. The modified, later version has an even faster tree-building algorithm, which allows a branch swapper to be used at each replication (FARRIS & al. 1998). Also, with this version it is possible to perform several random-addition sequences per replicate.

For the original *Jac* analysis, without branch-swapping, 1000 replications were used, except in the analyses of reduced data (see below). The output consists of a single tree comprising those clades present in at least 50% of the trees generated by the 1000 searches (see FARRIS & al. 1996 for additional explanations). The first complete analysis comprised the entire sequences with all positions (first, second, and third) and all changes (transversions and transitions) included and weighted equally. A series of analyses with reduced data sets were also performed. Since the first analysis revealed that jackknife frequencies did not change significantly after 500 replicates, these additional analyses were restricted to 500 rather than 1000 reiterations. The second analysis used only the first and second positions in the sequences and a third analysis only the third position. A fourth analysis utilized transversions only, transitions being ignored throughout the sequences (in all positions). All analyses were run on a 133 MHz Pentium.

Using the modified, new version of *Jac* with branch-swapping, the same analyses were performed. To ensure that the addition order of taxa did not influence the results, five random-addition sequences were performed for each replicate. Both Pentiums (133–200 MHz) and PowerPCs (66 MHz) were used for these analyses.

Results

Our initial analysis using parsimony jackknifing without branch-swapping required 99 hours on our 133 MHz Pentium. The *Jac* version used for the analysis with branch-swapping has a stop-restart function, making it possible to move the analysis between computers; our use of this feature precludes a precise time

estimate. A relevant comparison can, however, be made on the basis of per replicate speed. The original *Jac* required 356 seconds per replicate. With the new version with branch-swapping the time is increased to approximately 1304.5 seconds.

The tree resulting from analysis of all positions, with branch-swapping, includes 1400 supported groups. The tree has a total of 67482 steps in informative characters, of these 15602, 12335, and 39545 are distributed among the 1st, 2nd and 3rd positions, respectively. 46098 of the steps comprise transversions, whereas only 21384 are transitions.

When comparing the results from the original and the new *Jac* it is obvious that including a combination of branch-swapping and random-addition sequences does make a positive difference (Table 1). A comparison shows that the first tree has 83 groups not present in the second tree, whereas the second tree has 124 unique groups. Most of the differences concern groups with low jackknife frequencies (<60%). The groups of the first tree are generally small, 51% of them consist of two or three taxa. Only 11% have ten or more taxa, with a largest group of 27. In the second, branch-swapped tree 30% of the new groups have two or three taxa, and 41% ten or more. For groups present in both trees the jackknife frequencies are often approximately the same, especially for smaller groups. For deep nodes, setting off more ancient groups, branch-swapping tends to increase jackknife frequencies (see Table 1). Analyses of the reduced data sets (see below) show a

Table 1. Some results from the analyses using all, first and second positions, third position, or transversions only. “–” indicates support of <50%. ¹Combined with five random addition sequences per replicate

	All data	1st & 2nd positions	3rd position	Transversions only
Number of positions				
Total number of positions	1428	952	476	1428
Informative positions	1235	764	471	1064
Resolution of tree				
Number of supported clades				
without branch-swapping	1359	410	1284	655
with branch-swapping ¹	1400	431	1327	706
Support for major groups				
(jackknife frequencies without branch-swapping)				
Streptophytes	55%	64%	–	71%
Land plants	70%	–	58%	80%
Seed plants	75%	–	55%	69%
Flowering plants	87%	–	87%	61%
(jackknife frequencies with branch-swapping ¹)				
Streptophytes	62%	76%	–	80%
Land plants	96%	–	97%	90%
Seed plants	98%	85%	65%	99%
Flowering plants	100%	–	100%	99%

similar tendency, and in all of them branch-swapping led to more resolved trees. Since the trees resulting from branch-swapping analyses represent better hypotheses of phylogeny, we will limit our discussion to these below.

The reduced analyses comprising the first and second positions, the third position, or transversions only, yielded less resolution but no conflicting results, except for a few clades being weakly supported by low jackknife values (below 55%). As with the full data set, branch-swapping resulted in more resolved trees, with higher jackknife frequencies for deeper nodes. Because we feel there is no more phylogenetic information to be gained by these approaches, we restrict our discussion to the resulting lower resolution and support for some major groups (Table 1). Using only the first and second positions reduces resolution drastically (to less than 1/3), with 431 supported clades compared to 1400 using all positions. This approach also fails to recognize the major groups of land plants and flowering plants. Using only the third position reduces resolution marginally, from 1400 to 1327 supported clades. Streptophytes are among the lost clades. Using only transversions reduces resolution by half, to 706 supported clades. The major groups of streptophytes, land plants, seed plants and flowering plants are all recognized, but resolution at higher levels is poor.

Obviously, it is not possible to reproduce the entire tree of 2538 sequences here. Details of the tree may be obtained from the first author upon request. Figure 1 shows the basal branches of the tree, with the flowering plants (angiosperms) as a single branch. Resolution is not shown within genera and orders of green algae (*Volvocales* and *Charales*) and within several families of ferns, cycads, and conifers. The trees show clades with jackknife frequencies of at least 50% and provide also the number of sequences within each group. The whole tree is rooted between cyanobacteria and eukaryotes, a split supported at the 100% level. Streptophytes, i.e. the green algal lineage leading to land plants ('charophytes' in a wide sense) and the land plants, are supported by a jackknife value of 62.4%, land plants by 96.8%, seed plants by 98.9%, and flowering plants by the maximum value of 100%. Mosses are not supported as monophyletic, but the sampling of this group is extremely limited. Ferns and conifers are also not supported as clades, but the possibility that they represent monophyletic groups is not contradicted by the tree.

Several details of relationships within the flowering plants are given in Appendices 1–5. The flowering plants are represented by 2230 sequences. There are 401 families included, as listed in Appendices 1–4. Appendix 5 lists a number of monophyletic groups including two or more families. Several of these groups correspond to commonly recognized orders. A number of larger groups are also supported, for example, monocots, eudicots, rosids (*Rosidae*), and caryophyllids (*Caryophyllidae*, the families of the order *Caryophyllales* and several others). Asterids do not appear as a clade in the tree, but their possible monophyly is not contradicted by the tree.

Discussion

Analysis. Our study shows that parsimony jackknifing makes it perfectly feasible to work with large data sets. The second analysis of the complete data set,

using branch-swapping with 1000 replicates, each replicate with five random-addition replicates, required approximately two weeks of computer time. Improvements in algorithms, computer programming, and hardware developments will no doubt speed up the calculations even more. For example, in 1995, a parsimony jackknife re-analysis of CHASE & al.'s (1993) original data set was presented by VICTOR ALBERT at the San Diego AIBS meeting. That analysis took 3.5 hours to execute on a 133 MHz Pentium (1000 replicates, without branch-swapping). Today, with the new *Jac*, the same analysis takes 15 minutes on a 200 MHz Pentium Pro.

The efficiency of the new *Jac* allowed us to incorporate branch-swapping and random-addition sequences in the analysis. When applied to our 2538-sequence matrix some groups from the original tree were lost, whereas several new groups were recognized. Many of these groups have low support (50–55%), and their gain or loss may be attributed to sampling error. A closer comparison shows that the lost groups are often found at the tips of the tree, including only a few taxa, whereas gained nodes are more evenly distributed over the tree, and several represent more ancient groups. The net result is that branch-swapping in combination with random-addition sequences has resulted in a better resolved tree (1400 groups, compared to 1359 without branch-swapping). Among the new groups are monocots and eudicots (see below), and, as can be seen in Table 1, jackknife frequencies for ancient groups are considerably higher. The same effect can be seen in the analyses using subsets of the matrix. For instance, when analyzing only 1st and 2nd positions without branch-swapping, the group “seed plants” is not recognized, but with branch-swapping it is set off by a frequency of 85%.

Whether or not branch-swapping and/or random-addition sequences will make a difference depends on the data. In many cases they affect the jackknife frequencies marginally, but sometimes branch-swapping is necessary to obtain accurate results. BENGT OXELMAN (pers. comm.) has pointed out that the *hennig* algorithm used in the old *Jac* tends to underestimate support for ancient groups set off by few characters. This certainly seems to be the case in our present study. In rare cases it may also overestimate support (PABLO GOLOBOFF, pers. comm.). A detailed discussion of jackknifing and branch-swapping is presented elsewhere (FARRIS & al. 1998). We recommend the use of more exact calculations, if possible, since they improve accuracy and provide more robust hypotheses of relationships. However, for extremely large data sets (>5000 taxa), “simple” jackknife parsimony is still the only choice.

Our taxon sampling includes representatives from broadly divergent groups at widely different levels of the taxonomic hierarchy. When we started collecting data we feared that the diverse, unbalanced sampling would inevitably lead to increased homoplasy and loss of resolution. This has not been the case. Most of the sequences have been used in more restricted studies, of specific taxonomic groups, e.g. the *Asterales* (GUSTAFSSON & BREMER 1997) or the *Malpighiales* (FAY & al. 1997). When comparing well supported groups from those studies with our tree there is a general agreement both in numbers of groups and jackknife frequencies (Tables 2, 3). In fact, support often increases with addition of more information. Poor sampling may lead to spurious groupings, as has been discussed in earlier papers (CHASE & al. 1993, OLMSTEAD & al. 1993, FARRIS & al. 1996).

Table 2. Comparison of jackknife frequencies (no branch swapping applied) in compatible groups in the 2538-sequence analysis and in a 46-sequence analysis of the order *Asterales* (GUSTAFSSON & BREMER 1997). Numbers in bold represent exactly the same sequences in both analyses

Jackknife%		Taxa		Taxon
2538-tree	46-tree	2538-tree	46-tree	
100%	100%	2	2	Unnamed group
100%	99.9%	2	2	<i>Argophyllaceae</i>
100%	99.9%	3	3	<i>Alseuosmiaceae</i>
98.7%	96.9%	9	7	<i>Campanulaceae</i>
97.9%	99.0%	2	2	Unnamed group
97.7%	97.9%	2	2	<i>Stylidiaceae</i>
91.7%	82.4%	3	2	Unnamed group
89.6%	78.1%	10	2	Unnamed group
85.5%	67.9%	3	2	Unnamed group
84.0%	88.0%	8	6	Unnamed group
80.9%	70.8%	26	4	<i>Asteraceae</i>
76.0%	95.8%	23	2	Unnamed group
74.6%	81.9%	42	11	Unnamed group
64.3%	<50%	13	4	<i>Goodeniaceae</i>
60.5%	64.1%	2	2	Unnamed group
54.3%	56.0%	3	3	<i>Calyceraceae</i>
52.2%	<50%	68	36	<i>Asterales</i>
51.7%	<50%	5	4	<i>Menyanthaceae</i>
<50%	64.3%	29	7	Unnamed group

RICE & al. (1997: 559) have objected to parsimony jackknifing on the astonishing grounds that it requires “abandoning” the parsimony criterion! They seem to have arrived at this idea by misinterpreting the fact that a parsimony jackknife tree need not itself be a most-parsimonious tree. For ambiguous data, the parsimony jackknife tree instead represents a most-parsimonious tree from which poorly supported groups have been removed. However, retaining only better-founded groups hardly means that parsimony has been abandoned. Nor did RICE & al. (1997) really believe so themselves, for they counted among “reasons for preferring parsimony” that several most-parsimonious trees “can be reduced to a consensus tree to give an indication of uncertainty in an analysis.”

A consensus tree consists only of groups with positive Bremer support and need not itself be a most-parsimonious tree. RICE & al. (1997) who themselves employed a consensus, could then scarcely object to removing poorly-supported groups. To further undercut their own position, they went on to point out that, “Parsimony jackknifing yields a single, often unresolved tree containing clades that are expected to appear in the consensus of all most-parsimonious trees”. In their view, apparently, removing poorly-supported groups is beneficial – unless parsimony jackknifing is used to accomplish this more easily! Their opposition to resampling, however, is even weaker than that. Parsimony jackknifing provides an

Table 3. Comparison of jackknife figures (no branch swapping applied) in compatible groups in the 2538-sequence analysis and in a 31-sequence analysis of the order *Malpighiales* (FAY & al. 1997). Numbers in bold represent exactly the same sequences in both analyses

Jackknife%		Taxa		Taxon
2558-tree	31-tree	2538-tree	31-tree	
100%	100%	17	2	<i>Malpighiaceae</i>
100%	100%	8	2	Unnamed group
100%	99.7%	4	3	<i>Linaceae</i>
100%	99.6%	14	3	<i>Violaceae</i>
100%	99.6%	5	4	Unnamed group
100%	99.6%	4	4	Unnamed group
99.9%	99.9%	2	2	Unnamed group
99.9%	99.1%	2	2	<i>Ochnoideae</i>
99.7%	94.7%	6	8	Unnamed group
99.7%	90.5%	3	3	<i>Quinaceae</i>
93.2%	94.8%	2	2	Unnamed group
82.7%	72.1%	2	2	<i>Passifloraceae</i>
81.3%	95.9%	2	2	<i>Humiriaceae</i>
74.4%	96.4%	4	2	Unnamed group
66.8%	61.3%	2	2	Unnamed group
60.9%	77.9%	11	2	Unnamed group
<50%	53.5%	>50	3	Unnamed group
<50%	50.6%	6	3	Unnamed group

indication of the different degrees of support for groups that are retained, while the consensus tree by itself does not.

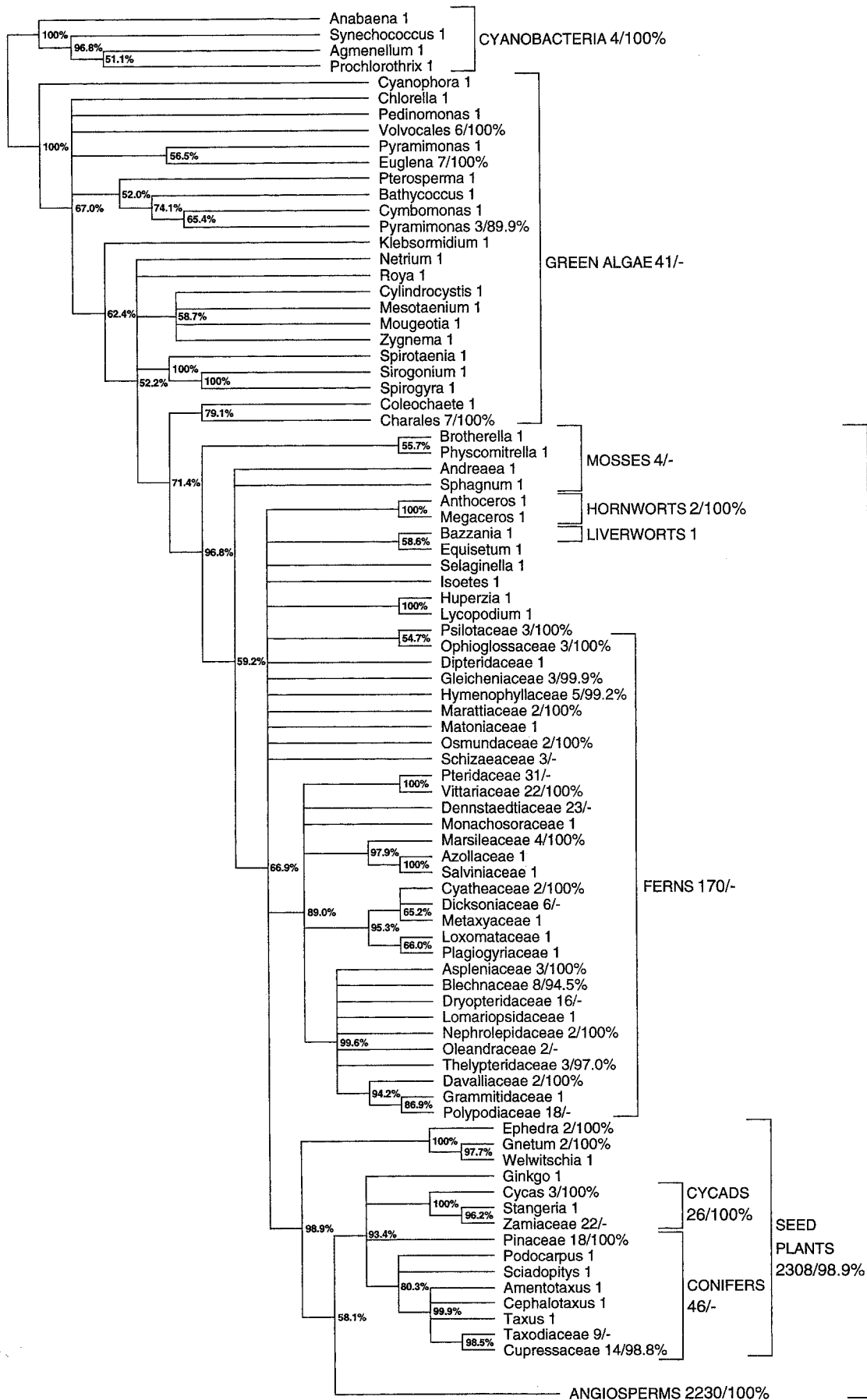
1st, 2nd, and 3rd positions. The total number of changes in informative characters, calculated from the parsimony jackknife tree is 67482. Of these 23.1, 18.3, and 58.6% occur in the 1st, 2nd, and 3rd position, respectively. It is frequently assumed that when changes occur more often in the third position, they are likely to contain more homoplasy. Following the same reasoning, if first and second positions are more conservative they should provide more phylogenetically reliable information. This assumption is not supported by our analysis. On the contrary, even though changes are more frequent in the third position, a large proportion of the groups found in analysis of the complete data set is lost when third positions are omitted (Table 1). Comparatively few groups are lost if the data are restricted to the third position only (from 1400 to 1327 clades). The phylogenetic information conveyed by the third position becomes even more obvious when the number of resolved clades is compared to the number of variable (informative) positions in the data. The first and second positions together contain 764 informative positions which support 431 clades, whereas 471 informative third positions support 1327 clades (Table 1). It seems that nothing is to be gained by excluding part of the data, especially not third positions.

Transversions and transitions. It is also often assumed that transitions should contain more homoplasy and that transversions should provide more reliable

phylogenetic information. The present study does not bear out this assumption. On our tree, transversions (46098 steps) occur more than twice as frequently as transitions (21384 steps). Further, an analysis with transversions only (1064 informative characters) results in much lower resolution (Table 1; from 1400 to 706 clades). There is a massive loss of resolution within the angiosperms. However, support for the very old, phylogenetically basal groups seems to increase if the data are restricted to transversions only, e.g. support for streptophytes increases from 62% to 80%. Hence, this may be a possible strategy for analyzing the oldest relationships, but not to be recommended in analyses restricted to land plants or subgroups of green algae.

Green plants. MANHART (1994) analyzed a set of 38 green plant *rbcL* sequences focusing on the green algal relatives of land plants and the pteridophytes (ferns and their allies). He concluded that analysis of these relationships with the *rbcL* gene seemed problematic, presumably due to limited phylogenetic information in the sequences but possibly also to sampling problems. Sampling of green algae is also rather limited in our data set and resolution among the green algae is low. There is good support for two well established orders, *Volvocales* and *Charales*. The latter order is one of the last algal branches in the green algal lineage leading to the land plants. This lineage is usually referred to as the 'charophytes' (MATTOX & STEWART 1984, MANHART 1994, MISHLER & al. 1994), and includes not only *Charales* but several other groups, e.g. *Coleochaete*, *Klebsormidium*, and the conjugates (*Netrium*, *Roya*, *Cylindrocystis*, *Mesotaenium*, *Mougeotia*, *Zygnema*, *Spirotaenia*, *Sirogonium*, and *Spirogyra* in Fig. 1). Charophytes and land plants together form the streptophytes (JEFFREY 1982, BREMER 1985). This group is recognized by the *rbcL* data, but supported with a comparatively low jackknife value of 62.4%. The sister group of the land plants according to our analysis is *Coleochaete* and *Charales* together, this latter pair being supported by a moderately high jackknife value of 79.1%.

Land plants. The old lineages of the land plants, i.e., the major bryophyte and pteridophyte groups, are not well represented in the data set. Resolution is also low, involving a polytomy of 15 land plant lineages (Fig. 1), and some are weakly supported (a jackknife frequency of less than 60%) and possibly spurious groups, such as the unexpected grouping of the two sequences representing liverworts (*Bazzania*) and sphenopsids (*Equisetum*). (In the tree based on transversions this grouping is not present. Instead, *Bazzania* appears as the sister to all other land plants.) Bryophyte and vascular land plant interrelationships were reviewed by MISHLER & al. (1994) but their review was based on other sources of data. A more comprehensive analysis of major bryophyte and pteridophyte interrelationships based on *rbcL* sequences remains to be done. Within the ferns, there are many more *rbcL* sequences available, and a comprehensive analysis was done by HASEBE & al. (1995). Their trees are more resolved than the jackknife tree, which does not support a fully resolved phylogeny of ferns based on *rbcL* sequences. There is, however, a general agreement with several larger, well supported groups, such as the water ferns (*Marsileaceae*, *Azollaceae*, *Salviniaceae*; 97.9%), the tree ferns together with three smaller families (*Cyatheaceae*, *Dicksoniaceae*, *Loxomataceae*, *Metaxyaceae*, *Plagiogyriaceae*; 95.3%), and the higher leptosporangiate ferns of ten families from *Aspleniaceae* to *Polypodiaceae* in Fig. 1 (99.6%; cf. HASEBE & al.



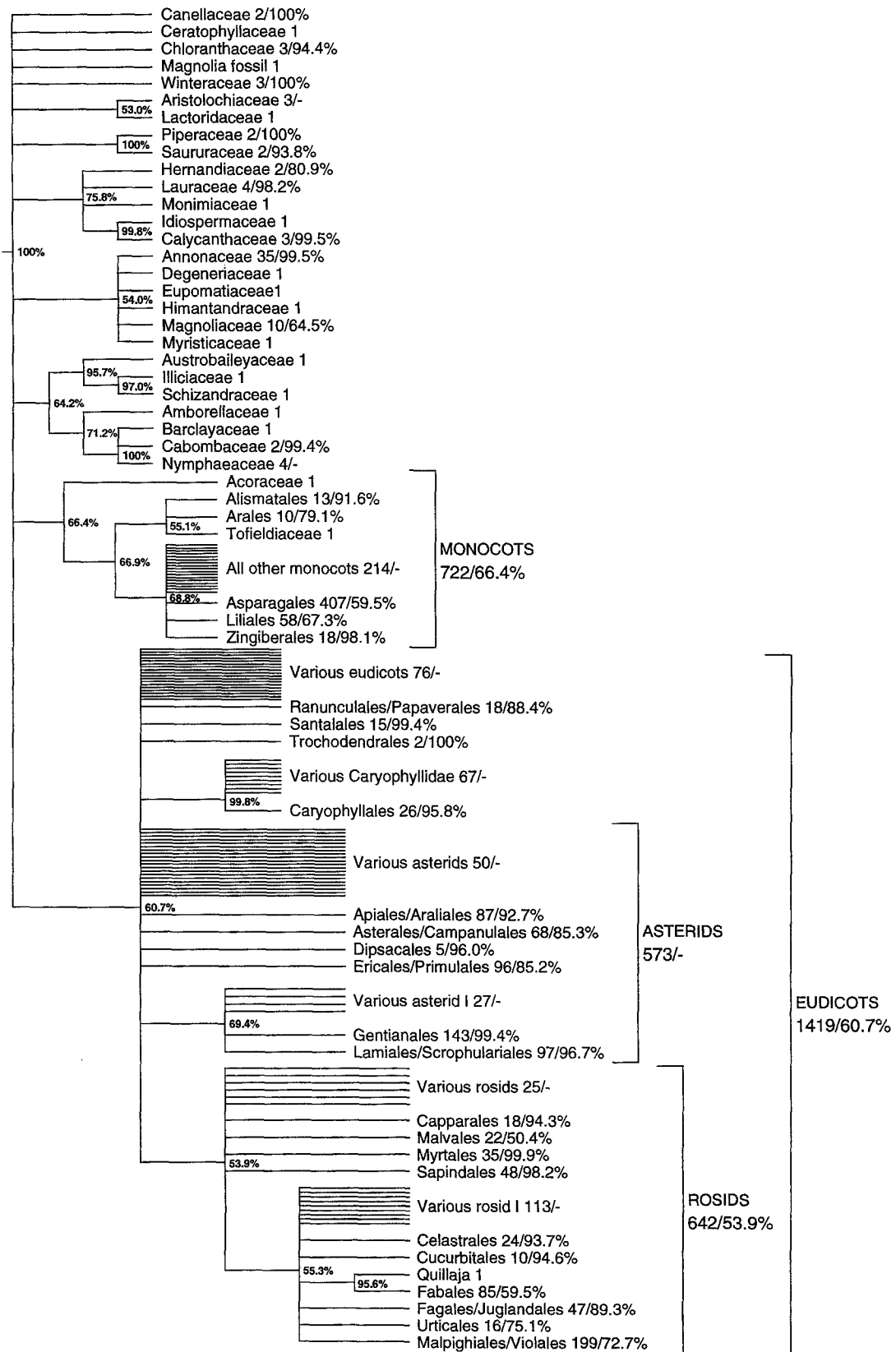
1995). Two of the families stand out as grossly paraphyletic, *Dryopteridaceae* and *Dennstaedtiaceae* (WOLF 1995). The former constitute a paraphyletic assemblage at the base of the higher leptosporangiate ferns, whereas the latter are a paraphyletic assemblage at the base of a more inclusive group, which includes also the water ferns and tree ferns and is supported by 89% of the replicates in the jackknife tree (Fig. 1).

Seed plant interrelationships have been the subject of much controversy (e.g. DOYLE & al. 1994, NIXON & al. 1994). Our jackknife analysis identifies three well supported monophyletic groups of extant seed plants, angiosperms (or flowering plants; 100%), *Gnetales* (*Ephedra*, *Gnetum*, *Welwitschia*; 100%), and more unexpectedly, all other seed plants together or 'gymnosperms s.str.', i.e. *Ginkgo*, cycads, and conifers, supported at 93.4%. Conifers are not supported as monophyletic but a resolution with *Pinaceae* as sister group to other conifers is not contradicted by the tree. *Taxus*, *Amentotaxus*, and *Cephalotaxus* are well nested in the conifers, in a group together with *Cupressaceae* and *Taxodiaceae* (99.9%). The latter family is paraphyletic. In the jackknife tree the *Gnetales* are sister to all other seed plants, angiosperms and gymnosperms s. str. being sister groups, but this grouping is supported by only 58.1% of the replicates. It is not supported by morphological data, which indicate a sister group relationship between *Gnetales* and angiosperms. The arrangement in the 2538-sequence jackknife tree, as well as that of the liverwort and *Equisetum* mentioned above, illustrates that groupings with low jackknife values (e.g. below 60%) may not be reliable.

During the compilation of the present matrix several preliminary analyses were performed while new data were being added. The groups occurring in these smaller data sets are consistent with the tree presented here, with one exception: the interrelationships of angiosperms, conifers, cycads and *Gnetales*. Depending on the taxon sampling, *Gnetales*, conifers, or conifers+*Ginkgo*+cycads have all appeared as the closest sister groups to the angiosperms, although often there is a basal polytomy of all of these groups. In the tree based on transversions only *Gnetales*+conifers+cycads+*Ginkgo* form the sister group of the angiosperms. At this point, it seems unlikely that sequences from *rbcL* alone will tell us which extant seed plants are most closely related to flowering plants.

Flowering plants. The angiosperms come out as a strongly monophyletic group, supported at the maximum 100% level. The two major groups of flowering plants, monocots and eudicots, are also supported as monophyletic, but with the somewhat low values of 66.4 and 60.7, respectively. Eudicots exclude the phylogenetically primitive flowering plants, which form ten unresolved branches at the flowering plant node of Fig. 2. Among them are the two commonly recognized orders *Laurales* and *Magnoliales*. The former is moderately well supported (75.8%) but the latter is only weakly supported (54.0%) and excludes *Canellaceae*

Fig. 1. The deep branches of the jackknife tree from parsimony analysis of 2538 *rbcL* sequences. At each node, the figure is the frequency of jackknife replicates containing that branch (branches appearing in less than 50% of the 1000 replicates are not retained). After each taxon, the first figure is the number of sequences included, the second is the frequency of replicates in which the group appears as a clade. A dash indicates an unsupported taxon that does not appear as a clade (not found in 50% of the 1000 replicates)



and *Winteraceae*, which are sometimes included in *Magnoliales*. The alleged *Magnolia* fossil, also placed outside *Magnoliales* in Fig. 2, is a sequence reported to have been obtained from a Miocene species (GOLENBERG & al. 1990). An interesting clade consists of the orders *Nymphaeales* (water lilies; *Nymphaeaceae*, *Barclayaceae*, *Cabombaceae*) and *Illiciales* (*Illiciaceae*, *Schizandraceae*) together with the two isolated families *Austrobaileyaceae* and *Amborellaceae* (64.2%). This clade was first identified by QIU & al. (1993) in their *rbcL* study of magnoliid groups. All clades in Fig. 2 are present in the trees reported by QIU & al., who also discussed various interrelationships among the clades and the sequence of the branching at the base of the tree. For example, in several *rbcL* analyses (e.g. CHASE & al. 1993, QIU & al. 1993, ALBERT & al. 1994), the genus *Ceratophyllum* (*Ceratophyllaceae*) appears as the sister group of all other flowering plants. This and other arrangements of the basal branches are not supported by our jackknife analysis. The *rbcL* sequences thus do not appear to be informative with respect to the difficult problem of resolving the deepest branches of the angiosperms.

The monocots have been the subject of a number of large-scale cladistic analyses using *rbcL* sequences, by DUVALL & al. (1993) and CHASE & al. (1995a, b). As expected, the clades supported by our jackknife analysis appear also in the trees reported by these authors, whereas many branches in their trees are absent in our tree; such branches lack support from the jackknife procedure. *Acorus* (*Acoraceae*) appears as the sister group of all other monocots, the latter being supported by a jackknife value of 66.9%, a moderately high value, just above the support for the monocots as a whole (including *Acorus*), which are maintained at 66.4%. The next clade consists of the two orders *Arales* (*Araceae* and *Lemnaceae*, *Lemna* being nested inside *Araceae*) and *Alismatales* (in a wide sense represented by the 11 families listed under *Alismatales* in Appendix 5) and a single sequence of *Tofieldiaceae* (from the genus *Pleea*, usually classified in the tribe *Tofieldieae* of *Melanthiaceae*; CHASE & al. 1995a, cf. TAKHTAJAN 1994). The *Arales/Alismatales/Tofieldiaceae* clade is not well supported (55.1%), but appears as the sister group of all remaining monocots (excluding *Acorus*). This major monocot clade, comprising all but *Acorus* (*Acoraceae*) and the *Arales/Alismatales/Tofieldiaceae* clade, is supported at 68.8%. The clade is not well resolved. It comprises a basal polytomy of 20 branches, three of them representing the large orders *Asparagales*, *Liliales* and *Zingiberales* (SMITH & al. 1993), the other 17 branches one or a few families each.

The two large orders *Asparagales* and *Liliales* were originally conceived as groups by HUBER (1969) and DAHLGREN & CLIFFORD (1982; also DAHLGREN & al.

Fig. 2. The major branches of flowering plants (angiosperms) in the jackknife tree from parsimony analysis of 2538 *rbcL* sequences. At each node, the figure is the frequency of jackknife replicates containing that branch (branches appearing in less than 50% of the 1000 replicates are not retained). After each taxon, the first figure is the number of sequences included, the second is the frequency of replicates in which the group appears as a clade. A dash indicates an unsupported taxon that does not appear as a clade (not found in 50% of the 1000 replicates). For further resolution of the terminal branches, see Appendix 5

1985), and later as a result of the *rbcL* analyses modified mainly by transfer of *Orchidaceae* and *Iridaceae* from *Liliales* to *Asparagales* (DUVALL & al. 1993, CHASE & al. 1995a, RUDALL & CUTLER 1995). *Asparagales* are here represented by no less than 407 sequences from 27 families. The order is weakly supported (59.5%) and interrelationships among the families are largely unresolved. *Liliales* with 58 sequences are supported by a jackknife frequency of 67.3%. The order comprises four strongly supported branches, (1) *Alstroemeriaceae*, *Colchicaceae*, *Luzuriagaceae* (in the strict sense comprising *Luzuriaga* only) and part of *Uvulariaceae* including the type genus, (2) *Melanthiaceae* (excluding *Aletris* of the tribe *Narthecieae* and *Pleea* of *Tofieldieae*=*Tofeldiaceae*) and *Trilliaceae*, (3) *Calochortaceae*, *Liliaceae*, *Philesiaceae*, *Smilacaceae* and several other genera of *Uvulariaceae*, and (4) *Campynemataceae*. The first three correspond to the *Liliales* subgroups a, b, and c of CHASE & al. (1995a) and they are here supported by the jackknife values of 100%, 85.2%, and 96.4%, respectively. At the family level, there is apparently considerable reclassification to be expected within *Asparagales* and *Liliales*, nine families of which are indicated to be non-monophyletic by our jackknife analysis (Appendix 3; CHASE & al. 1996, RUDALL & CHASE 1996). These orders comprise also several families well supported as monophyletic. Examples of such families represented by at least ten sequences are *Asphodelaceae*, *Orchidaceae* and *Themidaceae* (Appendix 4).

Except for the large monocot groups shown in Fig. 2 and discussed above, the jackknife analysis does not support any other supraordinal groups of monocots. The subclass *Commelinidae*, for example, is not recognized. There are, however, some suprafamilial groups not corresponding to currently recognized orders or subclasses (Appendix 5). One such group comprises *Cyclanthaceae*, *Pandanaceae*, *Stemonaceae* and *Velloziaceae* (*Stemonales* sensu CHASE & al. 1995b), together supported as monophyletic with a jackknife frequency of 93.0%. Within this group, a sister group relationship between *Cyclanthaceae* and *Pandanaceae* (*Pandanales*) is supported by 98.5%.

Relationships within the eudicots are not well resolved; the eudicot branch of the jackknife tree comprises a polytomy of 48 branches (Fig. 2). Several of these branches consist of single sequences or small groups of various phylogenetically archaic eudicots and asterids. The polytomy also contains several branches representing orders of these groups. In addition to the orders, there are three branches leading to other large, monophyletic groups, viz. the subclasses *Caryophyllidae* (93 sequences), *Rosidae* (642 sequences) and that part of *Asteridae* sometimes referred to as 'asterid I' (CHASE & al. 1993; 267 sequences). *Caryophyllidae* are strongly supported as monophyletic (99.8%), asterid I only moderately so (69.4%), and *Rosidae* are very weakly supported (53.9%).

Among the phylogenetically archaic eudicots (i.e. excluding rosids and asterids), the three orders *Ranunculales* s. l. (including *Papaveraceae*; 88.4%), *Santalales* (99.4%), and *Trochodendrales* (100%) are well supported as monophyletic. *Caryophyllales* has since long been recognized on its morphological and chemical features, and it is well supported also by the *rbcL* data (95.8%). The two families *Polygonaceae* and *Plumbaginaceae* are usually associated with *Caryophyllales*, and today often classified together with *Caryophyllales* in a subclass *Caryophyllidae* (e.g. CRONQUIST 1981). Molecular data have added several

families to this subclass, outside *Caryophyllales* but alongside *Polygonaceae* and *Plumbaginaceae* (ALBERT & al. 1992, CHASE & al. 1993; cf. Appendix 5). The latter two families are strongly supported as sister groups (99.8%). Two other small families of *Caryophyllidae*, *Asteropeiaceae* and *Physenaceae*, together form a well supported sister group of *Caryophyllales* (97.8%; MORTON & al. 1997a).

The rosids or *Rosidae* are a weakly supported clade in the jackknife tree (53.9%). Of the 642 sequences, 494 form a weakly supported subclade (55.3%), roughly corresponding to the 'rosid I' of CHASE & al. (1993; cf. Fig. 2 and Appendix 5), except that *Geraniaceae* and some related families are not included (cf. PRICE & PALMER 1993). Several large orders of rosids are recognized, *Capparales* (94.3%; RODMAN & al. 1993), *Myrtales* (99.9%; CONTI & al. 1996), *Fagales/Juglandales* (89.3%), *Sapindales* (98.2%; FERNANDO & al. 1995), *Celastrales* (93.7%; cf. SAVOLAINEN & al. 1997), *Cucurbitales* (94.6%), *Urticales* (75.1%), and *Malpighiales/Violales* (72.7%). *Urticales* form a group together with the families *Barbeyaceae*, *Elaeagnaceae*, *Rhamnaceae* and *Rosaceae* (77.1%), but interrelationships among these five taxa are unresolved or only weakly supported. *Malpighiales/Violales* are here taken in a very wide sense and include 29 families (Appendix 5) represented by 199 sequences in the jackknife tree. Among the families are the problematic *Flacourtiaceae*, which according to our analysis seem to be the most polyphyletic of all traditionally recognized flowering plant families. In the jackknife tree, the 20 *Flacourtiaceae* sequences appear as 18 separate branches, two of them even outside *Malpighiales/Violales*. *Salicaceae* are nested in a portion of *Flacourtiaceae*, supported by a jackknife frequency of 95.4%. This is consistent with the general view that *Flacourtiaceae* is a "garbage can" family.

Malvales are only weakly supported (50.4%), but the core of the order, with the four families *Bombacaceae*, *Malvaceae*, *Tiliaceae* and *Sterculiaceae*, is a well supported monophyletic group (98.2%). The extended concept of *Malvales* includes several small families (Appendix 5; cf. ALVERSON & al. 1998, BAYER & al. 1998). *Fabales* are here taken to include not only *Fabaceae* s. l. but also *Polygalaceae*, *Surianaceae* (Fernando & al. 1993) and *Xanthophyllaceae*. This clade of four families is weakly supported by 59.5% of the replicates, but their sister group is a single sequence of the genus *Quillaja* (traditionally in *Rosaceae*), and with *Quillaja* included, the *Fabales* clade is well supported by 95.6% replicates. As observed by MORGAN & al. (1994), *Quillaja* is not a member of *Rosaceae*. Apparently, it could be classified in a family of its own.

One unnamed, well supported rosid clade of eight families (97.4%) corresponds to what could be defined as an extended order *Oxalidales*, with *Cephalotaceae*, *Connaraceae*, *Cunoniaceae*, *Davidsoniaceae*, *Elaeocarpaceae*, *Eucryphiaceae*, *Oxalidaceae* and *Tremandraceae*. Two well supported pairs of sister families are *Krameriaceae/Zygophyllaceae* (96.3%; SHEAHAN & CHASE 1996) and *Geraniaceae/Hypseocharitaceae* (99.6%; PRICE & PALMER 1993). Another unnamed clade, possibly related to *Geraniaceae* (PRICE & PALMER 1993), consists of *Francoaceae*, *Greyiaceae*, *Ledocarpaceae*, *Melianthaceae* and *Vivianiaceae* (83.6%).

The asterids are not recognized as a monophyletic group, although their possible monophyly is not contradicted by the jackknife tree. They comprise 25

branches in the unresolved polytomy of eudicots. Of these branches, 20 represent single sequences or small groups usually classified as asterids, four are larger groups at the ordinal level, and one branch corresponds to the 'asterid I' of CHASE & al. (1993; OLMSTEAD & al. 1993), here represented by 267 sequences and supported by a jackknife frequency of 69.4%. This group commonly includes three major orders, *Gentianales* (BREMER 1996), *Lamiales/Scrophulariales* (OLMSTEAD & REEVES 1995) and *Solanales*. The first two are well supported, by 99.4 and 96.7%, respectively, but *Solanales* are not recognized in the tree. Instead, they are represented by three separate branches within the same polytomy of the asterid I clade, (1) *Boraginaceae*, (2) *Convolvulaceae* and *Solanaceae* as sister groups (95.2%) and (3) a clade comprising *Montiniaceae*, *Hydrophyllaceae* and *Sphenocleaceae* (59.9%).

The other well supported orders of asterids are *Apiales/Araliales* (92.7%; PLUNKETT & al. 1996), *Asterales/Campanulales* (85.3%; GUSTAFSSON & al. 1996, GUSTAFSSON & BREMER 1997), *Dipsacales* (BACKLUND & BREMER 1998), and *Ericales* s. l. including *Ebenales*, *Primulales* and some *Theales* (85.2%, MORTON & al. 1997b). This latter group includes many different families (cf. Appendix 5) traditionally classified in different orders, for example, *Ebenaceae* (MORTON & al. 1997b), *Ericaceae* (KRON & CHASE 1993), *Lecythidaceae* (MORTON & al. 1997c), *Primulaceae* and *Theaceae*. Two well supported subgroups within *Ericales* are a clade of *Balsaminaceae*, *Marcgraviaceae*, *Pellicieraceae* and *Tetrameristicaceae* (retained at the 100 % level), and the three families usually classified in a separate order *Primulales*, viz. *Primulaceae*, *Myrsinaceae* and *Theophrastaceae* (99.1%; ANDERBERG & al. 1998).

There are few interfamilial relationships among the 20 branches of asterid sequences outside the orders discussed above. One family pair is the *Aucubaceae* and *Garryaceae* clade, found in 98.3% of the replicates. Another group is the so called 'Ilex clade' (OLMSTEAD & al. 1993, MORGAN & SOLTIS 1993), consisting of *Aquifoliaceae*, *Helwingiaceae*, and *Phyllonomaceae* (92.8%).

Phylogenetic reconstruction and the *rbcL* gene. The *rbcL* gene proves to be extremely useful for phylogenetic reconstruction. It is not always informative, but may be so at widely different levels of the taxonomic hierarchy. Variation in the gene may support very old, major groups such as streptophytes, land plants, seed plants and flowering plants. It may also resolve relationships within families and even within genera. Included in the data matrix are, for example, 22 species of *Nothofagus* (*Nothofagaceae*) and 12 species of *Drosera* (*Droseraceae*), which are resolved into 11 and 9 supported clades, respectively. In other groups at various taxonomic levels, the *rbcL* gene is phylogenetically uninformative. This seems to be the case, for example, with respect to the problems of resolving the basal clades of major groups such as ferns, flowering plants and eudicots.

We thank, PAUL BYGRAVE, KEN CAMERON, ANTONY COX, PAUL KORES, ABRAHAM MUASYA, GAIL REEVES, JAMES RICHARDSON, VINCENT SAVOLAINEN, TEREZA TERRAZAS, DAVE WILLIAMS, MARK WHITTEN, PAUL G. WOLF and KENNETH WURDACK for DNA sequences and other useful information. MARIA BACKLUND prepared the illustrations. This work was supported by Swedish Natural Science Research Council grants to JAMES S. FARRIS, BIRGITTA BREMER, and KÄRE BREMER.

Appendix 1. The 184 flowering plant families represented by single sequences only.

Archaic angiosperms

Amborellaceae
Austrobaileyaaceae
Barclayaceae
Ceratophyllaceae
Degeneriaceae
Eupomatiaceae
Himantandraceae
Idiospermaceae
Illiciaceae
Lactoridaceae
Monimiaceae
Myristicaceae
Schizandraceae

Monocots

Acoraceae
Aphyllanthaceae
Aponogetonaceae
Blandfordiaceae
Burmanniaceae
Butomaceae
Calochortaceae
Campynemataceae
Cannaceae
Cymodoceaceae
Dioscoreaceae
Doryanthaceae
Dracaenaceae
Eriocaulaceae
Eriospermaceae
Flagellariaceae
Haemodoraceae
Hanguanaceae
Heliconiaceae
Hydrocharitaceae
Ixioliriaceae
Lemnaceae
Lowiaceae
Najadaceae
Potamogetonaceae
Rapateaceae
Restionaceae
Ruppiaceae
Ruscaceae
Scheuchzeriaceae
Sparganiaceae

Tofieldiaceae
Typhaceae
Zannichelliaceae
Zosteraceae

Eudicots

Achatocarpaceae
Aextoxicaceae
Amaranthaceae
Ancistrocladaceae
Asteropeiaceae
Basellaceae
Buxaceae
Cercidiphyllaceae
Daphniphyllaceae
Didiereaceae
Dioncophyllaceae
Eupteleaceae
Frankeniaceae
Grossulariaceae
Gunneraceae
Iteaceae
Leeaceae
Menispermaceae
Misodendraceae
Molluginaceae
Myrothamnaceae
Nepenthaceae
Olacaceae
Opiliaceae
Penthoraceae
Physenaceae
Platanaceae
Portulacaceae
Pterostemonaceae
Rhabdodendraceae
Simmondsiaceae
Tamaricaceae
Tetracarpaeaceae
Tetracentraceae
Trochodendraceae

Eudicots – rosids

Akaniaceae
Alzateaceae
Balanopaceae
Barbeyaceae

Bataceae
Begoniaceae
Bretschneideraceae
Brexaceae
Caricaceae
Caryocaraceae
Cephalotaceae
Cneoraceae
Coriariaceae
Corynocarpaceae
Crossosomataceae
Davidsoniaceae
Diegodendraceae
Eucryphiaceae
Francoaceae
Geissolomataceae
Greyiaceae
Gyrostemonaceae
Heteropyxidaceae
Hippocrateaceae
Hugoniaceae
Hypseocharitaceae
Irvingiaceae
Kirkiaceae
Koerberliniaceae
Krameriaceae
Lacistemataceae
Ledocarpaceae
Leitneriaceae
Lepuropetalaceae
Lophopyxidaceae
Lythraceae
Malesherbiaceae
Medusagynaceae
Melanthaceae
Memecylaceae
Moringaceae
Muntingiaceae
Neuradaceae
Nitrariaceae
Oliniaceae
Paracryphiaceae
Parnassiaceae
Penaeaceae
Plagiopteridaceae
Psiloxylaceae
Punicaceae
Resedaceae
Rhynchocalycaceae

Appendix 1 (continued)

<i>Salvadoraceae</i>	<i>Buddlejaceae</i>	<i>Myoporaceae</i>
<i>Sarcolaenaceae</i>	<i>Byblidaceae</i>	<i>Myrsinaceae</i>
<i>Scyphostegiaceae</i>	<i>Callitrichaceae</i>	<i>Oncothecaceae</i>
<i>Sphaerosephalaceae</i>	<i>Cardiopteridaceae</i>	<i>Pellicieraceae</i>
<i>Ticodendraceae</i>	<i>Carpodetaceae</i>	<i>Phellinaceae</i>
<i>Tiliaceae</i>	<i>Cyclocheilaceae</i>	<i>Phrymataceae</i>
<i>Tovariaceae</i>	<i>Desfontainiaceae</i>	<i>Phyllonomaceae</i>
<i>Trapaceae</i>	<i>Donatiaceae</i>	<i>Roridulaceae</i>
<i>Tremandraceae</i>	<i>Eremosynaceae</i>	<i>Sphenocleaceae</i>
<i>Trigoniaceae</i>	<i>Eucommiaceae</i>	<i>Tetrachondraceae</i>
<i>Tropaeolaceae</i>	<i>Garryaceae</i>	<i>Tetrameristaceae</i>
<i>Turneraceae</i>	<i>Griselinaceae</i>	<i>Vahliaceae</i>
<i>Vivianiaceae</i>	<i>Grubbiaceae</i>	<i>Valerianaceae</i>
<i>Xanthophyllaceae</i>	<i>Helwingiaceae</i>	
	<i>Hippuridaceae</i>	
Eudicots – asterids	<i>Hydrophyllaceae</i>	
	<i>Hydrostachyaceae</i>	
<i>Aucubaceae</i>	<i>Melanophyllaceae</i>	
<i>Avicenniaceae</i>	<i>Montiniaceae</i>	

Appendix 2. The 30 flowering plant families not supported as clades in the analysis, yet possibly monophyletic (grouping the sequences for each family into a single clade is not contradicted by the tree). The first figure is the number of sequences, the second is the number of clades in the tree.

Archaic angiosperms	<i>Eremolepidaceae</i> 2/2	<i>Sterculiaceae</i> 4/4
	<i>Hamamelidaceae</i> 7/3	
<i>Nymphaeaceae</i> 4/2	<i>Santalaceae</i> 2/2	Eudicots – asterids
	<i>Viscaceae</i> 3/3	
Monocots	Eudicots – rosids	<i>Acanthaceae</i> 12/4
<i>Colchicaceae</i> 10/2	<i>Anacardiaceae</i> 9/3	<i>Bignoniaceae</i> 3/3
<i>Convallariaceae</i> 9/7	<i>Bombacaceae</i> 2/2	<i>Dipsacaceae</i> 2/2
<i>Iridaceae</i> 37/2	<i>Cunoniaceae</i> 3/2	<i>Escalloniaceae</i> 2/2
<i>Marantaceae</i> 3/2	<i>Euphorbiaceae</i> 89/5	<i>Hydrangeaceae</i> 5/3
<i>Phormiaceae</i> 14/7	<i>Myrtaceae</i> 5/4	<i>Nyssaceae</i> 3/3
Eudicots	<i>Oxalidaceae</i> 2/2	<i>Pedaliaceae</i> 3/3
	<i>Rhamnaceae</i> 30/3	<i>Styracaceae</i> 7/2
<i>Droseraceae</i> 14/2	<i>Rutaceae</i> 9/4	<i>Theaceae</i> 6/2
		<i>Verbenaceae</i> 6/2

Appendix 3. The 29 flowering plant families not supported as clades in the analysis and indicated to be non-monophyletic (some sequences of each family form clades with sequences from other families). The first figure is the number of sequences, the second is the number of clades in the tree.

Archaic angiosperms	<i>Nolinaceae</i> 4/3	<i>Elaeocarpaceae</i> 5/4
	<i>Smilacaceae</i> 2/2	<i>Flacourtiaceae</i> 20/18
<i>Aristolochiaceae</i> 3/2	<i>Uvulariaceae</i> 16/3	<i>Ixonanthaceae</i> 3/2
		<i>Simaroubaceae</i> 7/5
Monocots	Eudicots	<i>Staphyleaceae</i> 2/2
<i>Agavaceae</i> 9/9	<i>Aizoaceae</i> 3/2	<i>Ulmaceae</i> 7/4
<i>Amaryllidaceae</i> 21/6	<i>Chenopodiaceae</i> 3/2	Eudicots – asterids
<i>Araceae</i> 9/8	<i>Loranthaceae</i> 5/2	<i>Apiaceae</i> 58/7
<i>Asteliaceae</i> 2/2	<i>Phytolaccaceae</i> 3/2	<i>Araliaceae</i> 22/10
<i>Hyacinthaceae</i> 54/3	Eudicots – rosids	<i>Cornaceae</i> 13/4
<i>Juncaceae</i> 8/5	<i>Capparaceae</i> 3/3	<i>Scrophulariaceae</i> 6/4
<i>Luzuriagaceae</i> 3/2	<i>Celastraceae</i> 17/11	
<i>Melanthiaceae</i> 11/3		
<i>Musaceae</i> 2/2		

Appendix 4. The 158 flowering plant families supported as clades in the analysis. The first figure is the number of sequences included, the second is the frequency of jackknife replicates in which the family appears as a clade. Figures in bold are comparatively high support values (>80% of jackknife replicates) for families represented by at least 10 sequences.

Archaic angiosperms	<i>Bromeliaceae</i> 6/100	<i>Xanthorrhoeaceae</i> 2/100
	<i>Commelinaceae</i> 4/99.9	<i>Zingiberaceae</i> 4/95.7
<i>Annonaceae</i> 35/ 99.5	<i>Costaceae</i> 3/67.5	Eudicots
<i>Cabombaceae</i> 2/99.4	<i>Cyanastraceae</i> 3/79.5	<i>Berberidaceae</i> 3/73.2
<i>Calycanthaceae</i> 3/99.5	<i>Cyclanthaceae</i> 2/98.0	<i>Cactaceae</i> 3/96.2
<i>Canellaceae</i> 2/100	<i>Cyperaceae</i> 73/ 88.1	<i>Caryophyllaceae</i> 4/100
<i>Chloranthaceae</i> 3/94.4	<i>Dasypogonaceae</i> 3/99.5	<i>Crassulaceae</i> 4/96.9
<i>Hernandiaceae</i> 2/80.9	<i>Hypoxidaceae</i> 5/83.8	<i>Dilleniaceae</i> 4/100
<i>Lauraceae</i> 4/58.8	<i>Joinvilleaceae</i> 2/100	<i>Haloragaceae</i> 2/98.0
<i>Magnoliaceae</i> 10/64.5	<i>Liliaceae</i> 9/57.3	<i>Lardizabalaceae</i> 5/98.7
<i>Piperaceae</i> 2/100	<i>Lomandraceae</i> 10/51.7	<i>Nelumbonaceae</i> 2/100
<i>Saururaceae</i> 2/93.8	<i>Orchidaceae</i> 163/ 99.8	<i>Nyctaginaceae</i> 2/93.9
<i>Winteraceae</i> 3/100	<i>Pandanaceae</i> 2/100	<i>Paeoniaceae</i> 2/100
Monocots	<i>Philesiaceae</i> 2/59.3	<i>Papaveraceae</i> 3/76.8
<i>Alismataceae</i> 3/100	<i>Philydraceae</i> 2/100	<i>Plumbaginaceae</i> 40/ 100
<i>Alliaceae</i> 16/76.3	<i>Poaceae</i> 71/ 100	<i>Polygonaceae</i> 4/99.9
<i>Alstroemeriaceae</i> 3/100	<i>Pontederiaceae</i> 14/ 100	<i>Proteaceae</i> 3/100
<i>Anthericaceae</i> 6/100	<i>Stemonaceae</i> 2/99.5	<i>Ranunculaceae</i> 5/94.5
<i>Areaceae</i> 8/99.7	<i>Strelitziaceae</i> 3/51.3	<i>Sabiaceae</i> 2/89.3
<i>Asparagaceae</i> 2/100	<i>Taccaceae</i> 2/100	<i>Saxifragaceae</i> 24/61.8
<i>Asphodelaceae</i> 14/ 100	<i>Tecophilaeaceae</i> 6/99.3	<i>Stegnospermataceae</i> 2/99.9
<i>Boryaceae</i> 2/96.0	<i>Themidaceae</i> 15/ 93.5	<i>Vitaceae</i> 12/62.5
	<i>Trilliaceae</i> 5/100	
	<i>Velloziaceae</i> 3/100	

Appendix 4 (continued)

Eudicots – rosids

Achariaceae 2/88.3
Betulaceae 6/66.4
Brassicaceae 2/99.7
Burseraceae 3/98.5
Cannabaceae 2/100
Casuarinaceae 5/99.9
Chrysobalanaceae 2/100
Clusiaceae 5/63.1
Combretaceae 4/100
Connaraceae 2/100
Cucurbitaceae 3/100
Datisceae 2/99.7
Dichapetalaceae 3/99.0
Dipterocarpaceae 2/99.6
Elaeagnaceae 3/100
Erythroxylaceae 2/99.9
Fabaceae 76/91.5
Fagaceae 8/66.5
Geraniaceae 13/100
Huaceae 2/100
Humiriaceae 2/81.3
Juglandaceae 3/100
Limnanthaceae 2/100
Linaceae 4/100
Malpighiaceae 17/100
Malvaceae 3/99.3
Melastomataceae 4/100
Meliaceae 8/80.8
Moraceae 3/60.2
Myricaceae 2/99.9
Nothofagaceae 22/100
Ochnaceae 2/99.9

Onagraceae 9/97.6
Pandaceae 2/99.7
Passifloraceae 2/82.6
Peganaceae 2/100
Polygalaceae 3/99.8
Quinaceae 3/99.7
Rhizophoraceae 2/99.9
Rosaceae 42/99.9
Salicaceae 14/81.0
Sapindaceae s.l. 7/99.9
Stackhousiaceae 2/86.3
Surianaceae 5/99.3
Tetramelaceae 2/96.3
Thymelaeaceae 5/97.6
Urticaceae 4/97.1
Violaceae 15/100
Vochysiaceae 3/99.9
Zygophyllaceae 17/77.9

Eudicots – asterids

Actinidiaceae 3/99.5
Adoxaceae s.l. 4/88.3
Alseuosmiaceae 3/100
Apocynaceae s.l. 25/80.1
Aquifoliaceae 4/99.6
Argophyllaceae 2/100
Asteraceae 26/85.4
Balsaminaceae 3/100
Boraginaceae 4/99.8
Bruniaceae 2/100
Calyceraceae 3/58.2
Campanulaceae 9/100
Caprifoliaceae 2/98.1

Clethraceae 2/99.3
Convolvulaceae 2/100
Cyrillaceae 2/98.4
Diapensiaceae 2/97.2
Ebenaceae 6/98.0
Ericaceae s.l. 24/89.0
Fouquieriaceae 2/100
Gelsemiaceae 3/86.3
Gentianaceae 8/97.4
Gesneriaceae 2/73.8
Globulariaceae 2/83.6
Goodeniaceae 13/77.0
Lamiaceae 45/91.7
Lecythidaceae 8/95.5
Lentibulariaceae 2/59.3
Loasaceae 6/96.4
Loganiaceae 4/99.8
Marcgraviaceae 2/95.6
Menyanthaceae 5/50.3
Oleaceae 4/65.8
Pentaphragmataceae 2/100
Pittosporaceae 5/100
Polemoniaceae 3/99.9
Primulaceae 6/100
Rubiaceae 103/91.7
Sapotaceae 8/99.1
Sarraceniaceae 3/99.4
Solanaceae 17/100
Stilbaceae 3/89.7
Stylidiaceae 2/99.3
Symplocaceae 3/100
Theophrastaceae 2/99.5

Appendix 5. Selected clades of more than one flowering plant family supported by the analysis. The first figure is the number of sequences included, the second is the frequency of replicates in which the group appears as a clade. Figures in bold are comparatively high support values (>80% of jackknife replicates) for groups represented by at least 10 sequences. A dash indicates an unsupported family that does not appear as a clade within the group.

Archaic angiosperms	<i>Barclayaceae</i> 1	<i>Boryaceae</i> 2/100
	<i>Cabombaceae</i> 2/99.4	<i>Convallariaceae</i> 9/-
<i>Magnoliales</i> 49/54.0	<i>Nymphaeaceae</i> 4/-	<i>Doryanthaceae</i> 1
<i>Annonaceae</i> 35/ 99.5		<i>Dracaenaceae</i> 1
<i>Degeneriaceae</i> 1	Monocots	<i>Eriospermaceae</i> 1
<i>Eupomatiaceae</i> 1		<i>Hyacinthaceae</i> 54/-
<i>Himantandraceae</i> 1	All monocots except <i>Acoraceae</i> 721/66.9	<i>Hypoxidaceae</i> 5/83.8
<i>Myristicaceae</i> 1		<i>Iridaceae</i> 37/-
<i>Magnoliaceae</i> 10/64.5		<i>Ixioliriaceae</i> 1
	All monocots except <i>Acoraceae</i> and next clade 697/68.8	<i>Luzuriagaceae</i> p.p. 2/100
<i>Laurales</i> 11/75.8		<i>Lomandraceae</i> 10/51.7
<i>Calycanthaceae</i> 3/99.5	Unnamed 24/55.1	<i>Nolinaceae</i> 4/-
<i>Hernandiaceae</i> 2/80.9	<i>Alismatales</i> 13/ 91.6	<i>Orchidaceae</i> 163/ 99.8
<i>Idiospermaceae</i> 1	<i>Arales</i> 10/79.1	<i>Phormiaceae</i> 14/-
<i>Lauraceae</i> 4/58.8	<i>Tofieldiaceae</i> 1	<i>Ruscaceae</i> 1
<i>Monimiaceae</i> 1		<i>Tecophilaeaceae</i> 6/99.3
	<i>Alismatales</i> 13/ 91.6	<i>Themidaceae</i> 15/ 93.5
Unnamed 4/53.0	<i>Alismataceae</i> 3/100	<i>Xanthorrhoeaceae</i> 2/100
<i>Aristolochiaceae</i> 3/-	<i>Aponogetonaceae</i> 1	
<i>Lactoridaceae</i> 1	<i>Butomaceae</i> 1	<i>Liliales</i> 58/67.3
	<i>Cymodoceaceae</i> 1	<i>Alstroemeriaceae</i> 3/100
<i>Piperales</i> 4/100	<i>Hydrocharitaceae</i> 1	<i>Calochortaceae</i> 1
<i>Piperaceae</i> 2/100	<i>Najadaceae</i> 1	<i>Campynemataceae</i> 1
<i>Saururaceae</i> 2/93.8	<i>Potamogetonaceae</i> 1	<i>Colchicaceae</i> 10/-
	<i>Ruppiaceae</i> 1	<i>Liliaceae</i> 9/57.3
Unnamed 11/64.2	<i>Scheuchzeriaceae</i> 1	<i>Luzuriagaceae</i> 1
<i>Amborellaceae</i> 1	<i>Zannichelliaceae</i> 1	<i>Melanthiaceae</i> 8/-
<i>Austrobaileyaceae</i> 1	<i>Zosteraceae</i> 1	<i>Philesiaceae</i> 2/59.3
<i>Illiciales</i> 2/97.0		<i>Smilacaceae</i> 2/-
<i>Nymphaeales</i> 7/100	<i>Arales</i> 10/79.1	<i>Trilliaceae</i> 5/100
	<i>Araceae</i> 9/-	<i>Uvulariaceae</i> 16/-
Unnamed 3/95.7	<i>Lemnaceae</i> 1	
<i>Austrobaileyaceae</i> 1		Unnamed in <i>Liliales</i> 20/ 100
<i>Illiciales</i> 2/97.0	<i>Asparagales</i> 407/59.5	<i>Alstroemeriaceae</i> 3/100
<i>Illiciaceae</i> 1	<i>Agavaceae</i> 9/-	<i>Colchicaceae</i> 10/-
<i>Schizandraceae</i> 1	<i>Alliaceae</i> 16/76.3	<i>Luzuriagaceae</i> 1
	<i>Amaryllidaceae</i> 20/-	<i>Uvulariaceae</i> p.p. 6/72.4
Unnamed 8/71.2	<i>Anthericaceae</i> 6/100	
<i>Amborellaceae</i> 1	<i>Aphyllanthaceae</i> 1	Unnamed in <i>Liliales</i> 13/ 85.2
<i>Nymphaeales</i> 7/100	<i>Asparagaceae</i> 2/100	<i>Melanthiaceae</i> 8/-
	<i>Asphodelaceae</i> 14/ 100	<i>Trilliaceae</i> 5/100
	<i>Asteliaceae</i> 2/-	
<i>Nymphaeales</i> 7/100	<i>Blandfordiaceae</i> 1	Unnamed in <i>Liliales</i> 24/ 96.4
		<i>Calochortaceae</i> 1

Appendix 5 (continued)

- Liliaceae* 9/57.3
Philesiaceae 2/59.3
Smilacaceae 2/-
Uvulariaceae p.p. 10/-
- Stemonales/Pandanales*
 9/93.0
Cyclanthaceae 2/98.0
Pandanaceae 2/100
Stemonaceae 2/99.5
Velloziaceae 3/100
- Unnamed in *Stemonales/Pandanales* 6/50.4
Cyclanthaceae 2/98.0
Pandanaceae 2/100
Stemonaceae 2/99.5
- Pandanales* 4/98.5
Cyclanthaceae 2/98.0
Pandanaceae 2/100
- Unnamed 3/100
Dioscoreaceae 1
Taccaceae 2/100
- Unnamed 18/68.8
Commelinaceae 4/99.9
Pontederiaceae 14/100
- Unnamed 81/791
Cyperaceae 73/88.1
Juncaceae 8/-
- Unnamed 74/83.5
Joinvilleaceae 2/100
Poaceae 71/100
Restionaceae 1
- Typhales* 2/63.8
Sparganiaceae 1
Typhaceae 1
- Zingiberales* 18/98.1
Cannaceae 1
Costaceae 3/53.0
Heliconiaceae 1
Lowiaceae 1
Marantaceae 3/-
Musaceae 2/-
- Strelitziaceae* 3/-
Zingiberaceae 4/91.0
- Eudicots**
- Unnamed 8/61.4
Crassulaceae 4/96.9
Haloragaceae 2/98.0
Penthoraceae 1
Tetracarpaeaceae 1
- Unnamed 2/811
Gunneraceae 1
Myrothamnaceae 1
- Unnamed 2/98.3
Iteaceae 1
Pterostemonaceae 1
- Unnamed 12/100
Leeaceae 1
Vitaceae 12/62.5
- Caryophyllidae* 93/99.8
Ancistrocladaceae 1
Asteropeiaceae 1
Dioncophyllaceae 1
Droseraceae 14/-
Frankeniaceae 1
Nepenthaceae 1
Physenaceae 1
Plumbaginaceae 40/100
Polygonaceae 4/99.9
Rhabdodendraceae 1
Simmondsiaceae 1
Tamaricaceae 1
Caryophyllales 26/95.8
- Unnamed in *Caryophyllidae*
 44/99.8
Plumbaginaceae 40/100
Polygonaceae 4/99.9
- Unnamed in *Caryophyllidae*
 28/97.8
Asteropeiaceae 1
Physenaceae 1
Caryophyllales 26/95.8
- Caryophyllales* 26/95.8
Achatocarpaceae 1
- Aizoaceae* 3/-
Amaranthaceae 1
Basellaceae 1
Cactaceae 3/96.2
Caryophyllaceae 4/100
Chenopodiaceae 3/-
Didiereaceae 1
Molluginaceae 1
Nyctaginaceae 2/93.9
Phytolaccaceae 3/-
Portulacaceae 1
Stegnospermataceae
 2/99.9
- Unnamed in *Caryophyllales*
 9/80.8
Achatocarpaceae 1
Amaranthaceae 1
Caryophyllaceae 4/100
Chenopodiaceae 3/-
- Unnamed in *Caryophyllales*
 17/94.5
Aizoaceae 3/-
Basellaceae 1
Cactaceae 3/96.2
Didiereaceae 1
Mulluginaceae 1
Nyctaginaceae 2/93.9
Phytolaccaceae 3/-
Portulacaceae 1
Stegnospermataceae
 2/99.9
- Ranunculales/Papaverales*
 18/88.4
Berberidaceae 3/66.0
Eupteleaceae 1
Lardizabalaceae 5/97.8
Menispermaceae 1
Papaveraceae 3/55.1
Ranunculaceae 5/83.5
- Santalales* 15/99.4
Eremolepidaceae 2/-
Olacaceae 1
Opiliaceae 1
Loranthaceae 5/-
Misodendraceae 1
Santalaceae 2/-
Viscaceae 3/-

Appendix 5 (continued)

<i>Trochodendrales</i> 2/100	<i>Malvales</i> 22/50.4	<i>Meliaceae</i> 8/80.8
<i>Tetracentraceae</i> 1	<i>Bombacaceae</i> 2/-	<i>Paracryphiaceae</i> 1
<i>Trochodendraceae</i> 1	<i>Diegodendraceae</i> 1	<i>Peganaceae</i> 2/100
Eudicots – rosids	<i>Dipterocarpaceae</i> 2/99.6	<i>Rutaceae</i> 9/-
	<i>Malvaceae</i> 3/99.3	<i>Sapindaceae</i> s.l. 7/99.9
	<i>Muntingiaceae</i> 1	<i>Simaroubaceae</i> p.p. 5/-
<i>Rosidae</i> 642/53.9	<i>Neuradaceae</i> 1	
<i>Crossosomataceae</i> 1	<i>Sarcolaenaceae</i> 1	<i>Unnamed in Sapindales</i>
<i>Francoaceae</i> 1	<i>Sphaerosephalaceae</i> 1	25/ 87.9
<i>Geissolomataceae</i> 1	<i>Sterculiaceae</i> 4/-	<i>Cneoraceae</i> 1
<i>Geraniaceae</i> 13/ 100	<i>Thymelaeaceae</i> 5/97.6	<i>Leitneriaceae</i> 1
<i>Greyiaceae</i> 1	<i>Tiliaceae</i> 1	<i>Meliaceae</i> 8/80.8
<i>Hypseocharitaceae</i> 1		<i>Paracryphiaceae</i> 1
<i>Ledocarpaceae</i> 1	<i>Unnamed in Malvales</i> 9/98.2	<i>Rutaceae</i> 9/-
<i>Melianthaceae</i> 1	<i>Bombacaceae</i> 2/-	<i>Simaroubaceae</i> p.p. 5/-
<i>Simaroubaceae</i> p.p. 2/100	<i>Malvaceae</i> 3/99.3	
<i>Staphyleaceae</i> 2/-	<i>Tiliaceae</i> 1	
<i>Vivianiaceae</i> 1	<i>Sterculiaceae</i> 4/-	
<i>Capparales</i> 18/ 94.3		<i>Rosid I</i> 494/55.3
<i>Malvales</i> 22/50.4		<i>Barbeyaceae</i> 1
<i>Myrtales</i> 35/ 99.9	<i>Myrtales</i> 35/ 99.9	<i>Cephalotaceae</i> 1
<i>Sapindales</i> 48/ 98.2	<i>Alzateaceae</i> 1	<i>Connaraceae</i> 2/100
<i>Rosid I</i> 494/55.3	<i>Combretaceae</i> 4/100	<i>Cunoniaceae</i> 3/-
	<i>Heteropyxidaceae</i> 1	<i>Davidsoniaceae</i> 1
<i>Unnamed</i> 5/83.6	<i>Lythraceae</i> 1	<i>Elaeagnaceae</i> 3/100
<i>Francoaceae</i> 1	<i>Melastomataceae</i> 4/100	<i>Elaeocarpaceae</i> 5/-
<i>Greyiaceae</i> 1	<i>Memecylaceae</i> 1	<i>Eucryphiaceae</i> 1
<i>Ledocarpaceae</i> 1	<i>Myrtaceae</i> 5/-	<i>Huaceae</i> 2/100
<i>Melianthaceae</i> 1	<i>Oliniaceae</i> 1	<i>Krameriaceae</i> 1
<i>Vivianiaceae</i> 1	<i>Onagraceae</i> 9/97.6	<i>Oxalidaceae</i> 2/-
<i>Unnamed</i> 14/ 99.6	<i>Penaeaceae</i> 1	<i>Rhamnaceae</i> 30/-
<i>Geraniaceae</i> 13/ 100	<i>Psiloxylaceae</i> 1	<i>Rosaceae</i> 42/ 99.9
<i>Hypseocharitaceae</i> 1	<i>Punicaceae</i> 1	<i>Tremandraceae</i> 1
<i>Capparales</i> 18/ 94.3	<i>Rhynchoalycaceae</i> 1	<i>Zygophyllaceae</i> 17/77.9
<i>Akaniaceae</i> 1	<i>Trapaceae</i> 1	<i>Celastrales</i> 24/ 93.7
<i>Bataceae</i> 1	<i>Vochysiaceae</i> 3/99.9	<i>Cucurbitales</i> 10/ 94.6
<i>Brassicaceae</i> 2/99.7	<i>Unnamed in Myrtales</i>	<i>Fabales/Juglandales</i>
<i>Capparaceae</i> 3/-	12/ 93.9	47/ 89.3
<i>Bretschneideraceae</i> 1	<i>Lythraceae</i> 1	<i>Urticales</i> 16/75.1
<i>Caricaceae</i> 1	<i>Onagraceae</i> 9/97.6	<i>Violales</i> 199/72.7
<i>Gyrostemonaceae</i> 1	<i>Punicaceae</i> 1	
<i>Koerberliniaceae</i> 1	<i>Trapaceae</i> 1	<i>Unnamed</i> 16/ 97.4
<i>Limnanthaceae</i> 2/100	<i>Sapindales</i> 48/ 98.2	<i>Cephalotaceae</i> 1
<i>Moringaceae</i> 1	<i>Anacardiaceae</i> 9/-	<i>Connaraceae</i> 2/100
<i>Resedaceae</i> 1	<i>Burseraceae</i> 3/98.5	<i>Cunoniaceae</i> 3/-
<i>Salvadoraceae</i> 1	<i>Cneoraceae</i> 1	<i>Davidsoniaceae</i> 1
<i>Tovariaceae</i> 1	<i>Kirkiaceae</i> 1	<i>Elaeocarpaceae</i> 5/-
<i>Tropaeolaceae</i> 1	<i>Leitneriaceae</i> 1	<i>Eucryphiaceae</i> 1
	<i>Nitrariaceae</i> 1	<i>Oxalidaceae</i> 2/-
		<i>Tremandraceae</i> 1

Appendix 5 (continued)

- Unnamed 12/**93.4**
Cephalotaceae 1
Cunoniaceae 3/-
Davidsoniaceae 1
Elaeocarpaceae 5/-
Eucryphiaceae 1
Tremandraceae 1
- Unnamed 5/991
Cunoniaceae 3/-
Davidsoniaceae 1
Eucryphiaceae 1
- Unnamed 4/94.9
Connaraceae 2/100
Oxalidaceae 2/-
- Unnamed 18/**96.3**
Krameriaceae 1
Zygophyllaceae 17/77.9
- Celastrales* 24/**93.7**
Brexiaceae 1
Celastraceae 17/-
Hippocrateaceae 1
Lepuropetalaceae 1
Parnassiaceae 1
Plagiopteridaceae 1
Stackhousiaceae 2/86.3
- Cucurbitales* 10/**94.6**
Begoniaceae 1
Coriariaceae 1
Corynocarpaceae 1
Cucurbitaceae 3/100
Datiscaceae 2/99.7
Tetramelaceae 2/96.3
- Unnamed 86/**95.6**
Quillaja 1
Fabales 85/59.5
- Fabales* 85/59.5
Fabaceae s.l. 76/**91.5**
Polygalaceae 3/99.8
Surianaceae 5/99.3
Xanthophyllaceae 1
- Fagales/Juglandales*
 47/**89.3**
Betulaceae 6/66.4
Casuarinaceae 5/99.9
Fagaceae 8/66.5
Juglandaceae 3/100
Myricaceae 2/99.9
Nothofagaceae 22/**100**
Ticodendraceae 1
- Unnamed 92/77.1
Barbeyaceae 1
Elaeagnaceae 3/100
Rhamnaceae 30/-
Rosaceae 42/**99.9**
Urticales 16/75.1
- Unnamed 50/51.6
Barbeyaceae 1
Elaeagnaceae 3/100
Rhamnaceae 30/-
Urticales 16/75.1
- Urticales* 16/75.1
Cannabaceae 2/100
Moraceae 3/60.2
Ulmaceae p.p. 7/-
Urticaceae 4/97.1
- Malpighiales/Violales*
 199/72.7
Achariaceae 2/88.3
Balanopaceae 1
Caryocaraceae 1
Chrysobalanaceae 2/100
Clusiaceae 5/63.1
Dichapetalaceae 3/99.0
Erythroxylaceae 2/99.9
Euphorbiaceae 89/-
Flacourtiaceae p.p. 18/-
Hugoniaceae 1
Humiriaceae 2/81.3
Irvingiaceae 1
Ixonanthaceae 3/-
Lacistemaceae 1
Linaceae 4/100
Lophopyxidaceae 1
- Malesherbiaceae* 1
Malpighiaceae 17/**100**
Medusagynaceae 1
Ochnaceae 2/99.9
Pandaceae 2/99.7
Passifloraceae 2/82.6
Quinaceae 3/99.7
Rhizophoraceae 2/99.9
Salicaceae 14/**81.0**
Scyphostegiaceae 1
Trigoniaceae 1
Turneraceae 1
Violaceae 15/**100**
- Unnamed in *Malpighiales/Violales* 4/80.9
Erythroxylaceae 2/99.9
Rhizophoraceae 2/99.9
- Unnamed in *Malpighiales/Violales* 24/**95.4**
Flacourtiaceae p.p. 9/-
Salicaceae 14/81.0
Scyphostegiaceae 1
- Unnamed in *Malpighiales/Violales* 5/100
Hugoniaceae 1
Linaceae 4/100
- Unnamed in *Malpighiales/Violales* 4/100
Malesherbiaceae 1
Passifloraceae 2/82.6
Turneraceae 1
- Unnamed in *Malpighiales/Violales* 6/99.7
Medusagynaceae 1
Ochnaceae 2/99.9
Quinaceae 3/99.7
- Eudicots – asterids**
 Unnamed 6/92.8
Aquifoliaceae 4/98.9
Helwingiaceae 1
Phyllonomaceae 1

Appendix 5 (continued)

Unnamed 2/98.3 <i>Aucubaceae</i> 1 <i>Garryaceae</i> 1	<i>Ericales/Primulales</i> 96/ 85.2 <i>Actinidiaceae</i> 3/99.5 <i>Balsaminaceae</i> 3/100 <i>Clethraceae</i> 2/99.3 <i>Cyrillaceae</i> 2/98.4 <i>Diapensiaceae</i> 2/97.2 <i>Ebenaceae</i> 6/98.0 <i>Ericaceae s.l.</i> 24/ 89.0 <i>Fouquieriaceae</i> 2/100 <i>Lecythidaceae</i> 8/99.5 <i>Marcgraviaceae</i> 2/95.6 <i>Pellicieraceae</i> 1 <i>Polemoniaceae</i> 3/99.9 <i>Roridulaceae</i> 1 <i>Sapotaceae</i> 8/99.1 <i>Sarraceniaceae</i> 3/99.4 <i>Styracaceae</i> 7/- <i>Symplocaceae</i> 3/100 <i>Tetrameristaceae</i> 1 <i>Theaceae</i> 6/- <i>Primulales</i> 9/99.1	<i>Lamiales/Scrophulariales</i> 97/ 96.7 Unnamed 19/ 95.2 <i>Convolvulaceae</i> 2/100 <i>Solanaceae</i> 17/ 100 Unnamed 3/59.9 <i>Hydrophyllaceae</i> 1 <i>Montiniaceae</i> 1 <i>Sphenocleaceae</i> 1 Unnamed 2/85.2 <i>Hydrophyllaceae</i> 1 <i>Sphenocleaceae</i> 1 <i>Gentianales</i> 143/ 99.4 <i>Apocynaceae s.l.</i> 25/ 80.1 <i>Gelsemiaceae</i> 3/86.3 <i>Gentianaceae</i> 8/97.4 <i>Loganiaceae</i> 4/99.8 <i>Rubiaceae</i> 103/ 91.7
<i>Apiales/Araliales</i> 87/ 92.7 <i>Apiaceae</i> 58/- <i>Araliaceae</i> 22/- <i>Griselinaceae</i> 1 <i>Melanophyllaceae</i> 1 <i>Pittosporaceae</i> 5/100	<i>Asterales/Campanulales</i> 68/ 85.3 <i>Alseuosmiaceae</i> 3/100 <i>Argophyllaceae</i> 2/100 <i>Asteraceae</i> 26/ 85.4 <i>Calyceraceae</i> 3/58.2 <i>Campanulaceae</i> 9/100 <i>Carpodetaceae</i> 1 <i>Donatiaceae</i> 1 <i>Goodeniaceae</i> 13/77.0 <i>Menyanthaceae</i> 5/50.3 <i>Pentaphragmataceae</i> 2/100 <i>Phellinaceae</i> 1 <i>Stylidiaceae</i> 2/99.3	<i>Lamiales/Scrophulariales</i> 97/ 96.7 <i>Acanthaceae</i> 12/- <i>Avicenniaceae</i> 1 <i>Bignoniaceae</i> 3/- <i>Buddlejaceae</i> 1 <i>Byblidaceae</i> 1 <i>Callitrichaceae</i> 1 <i>Cyclocheilaceae</i> 1 <i>Gesneriaceae</i> 2/73.8 <i>Globulariaceae</i> 2/83.6 <i>Hippuridaceae</i> 1 <i>Lamiaceae</i> 45/ 91.7 <i>Lentibulariaceae</i> 2/59.3 <i>Myoporaceae</i> 1 <i>Oleaceae</i> 4/65.8 <i>Pedaliaceae</i> 3/- <i>Phrymataceae</i> 1 <i>Scrophulariaceae</i> 6/- <i>Stilbaceae</i> 3/89.7 <i>Tetrachondraceae</i> 1 <i>Verbenaceae</i> 6/-
Unnamed in <i>Asterales</i> 47/ 68.3 <i>Asteraceae</i> 26/ 85.4 <i>Calyceraceae</i> 3/58.2 <i>Goodeniaceae</i> 13/77.0 <i>Menyanthaceae</i> 5/50.3	Unnamed in <i>Ericales s.l.</i> 7/100 <i>Balsaminaceae</i> 3/100 <i>Marcgraviaceae</i> 2/96.4 <i>Pellicieraceae</i> 1 <i>Tetrameristaceae</i> 1 <i>Primulales</i> 9/99.1 <i>Myrsinaceae</i> 1 <i>Primulaceae</i> 6/99.6 <i>Theophrastaceae</i> 2/98.4	<i>Asteridae</i> I 267/69.4 <i>Boraginaceae</i> 4/99.8 <i>Convolvulaceae</i> 2/100 <i>Hydrophyllaceae</i> 1 <i>Montiniaceae</i> 1 <i>Solanaceae</i> 17/ 100 <i>Sphenocleaceae</i> 1 <i>Vahliaceae</i> 1 <i>Gentianales</i> 143/ 99.4
Unnamed in <i>Asterales</i> 42/ 95.3 <i>Asteraceae</i> 26/ 85.4 <i>Calyceraceae</i> 3/58.2 <i>Goodeniaceae</i> 13/77.0	<i>Dipsacales</i> 5/96.0 <i>Caprifoliaceae</i> 2/98.1 <i>Dipsacaceae</i> 2/- <i>Valerianaceae</i> 1	

Appendix 6. Commonly recognized flowering plant families not represented in the analysis.

Basal angiosperms

Gomortegaceae
Trimeniaceae

Monocots

Anarthriaceae
Centrolepidaceae
Corsiaceae
Ecdeiocoleaceae
Hydatellaceae
Hydrocharitaceae
Juncaginaceae
Lilaeaceae
Limnocharitaceae
Mayacaceae
Posidoniaceae
Thurniaceae
Trichopodaceae
Triuridaceae
Xyridaceae

Eudicots

Anisophylleaceae
Aralidiaceae
Balanophoraceae

Biebersteiniaceae
Bixaceae
Brunelliaceae
Carlemanniaceae
Circaeasteraceae
Cistaceae
Cochlospermaceae
Columelliaceae
Crypteroniaceae
Cuscutaceae
Cynomoriaceae
Dialypetalanthaceae
Didymelaceae
Dipentodonaceae
Dirachmaceae
Duckeodendraceae
Elatinaceae
Goetzeaceae
Hoplostigmataceae
Hydnoraceae
Icacinaceae
Julianiaceae
Lennoaceae
Lepidobotryaceae
Lissocarpaceae
Martyniaceae
Medusandraceae
Mendonciaceae
Mitrastemonaceae
Morinaceae
Nesogenaceae
Pentadiplandraceae
Pentaphylacaceae
Peridiscaceae
Plantaginaceae
Plocospermataceae
Podostemaceae
Polyosmataceae
Pottingeriaceae
Ptaeroxylaceae
Rafflesiaceae
Rhoipteleaceae
Rousseaceae
Saccifoliaceae
Sargentodoxaceae
Sonneratiaceae
Sphenostemonaceae
Stachyuraceae
Symphoremataceae
Tepuianthaceae
Tetradiclidaceae
Torricelliaceae
Tribelaceae

References

- ALBERT, V. A., WILLIAMS, S. E., CHASE, M. W., 1992: Carnivorous plants: phylogeny and structural evolution. – *Science* **257**: 1491–1495.
- BACKLUND, A., BREMER, K., CHASE, M. W., MANHART, J., MISHLER, B. D., NIXON, K. C., 1994: Functional history and *rbcL* evidence for land plant phylogeny. – *Ann. Missouri Bot. Gard.* **81**: 534–567.
- ALVERSON, W. S., KAROL, K. G., BAUM, D.A., CHASE, M. W., SWENSEN, S. M., McCOURT, R., SYTSMA, K. J., 1998: Circumscription of the *Malvales* and relationship to other *Rosidae*: evidence from *rbcL* sequence data. – *Amer. J. Bot.* (in press).
- ANDERBERG, A. A., STÅHL, B., KÄLLERSJÖ, M., 1998: Phylogenetic interrelationships in the *Primulales* inferred from *rbcL* sequence data. – *Pl. Syst. Evol.* **211**: 93–102.
- BACKLUND, A., BREMER, B., 1998: Phylogeny of the *Asteridae* s.str. based on *rbcL* sequences, with particular reference to the *Dipsacales*. – *Pl. Syst. Evol.* **207**: 225–254.
- BAYER, C., CHASE, M. W., FAY, M. F., 1998: *Muntingiaceae*, a new family of dicotyledons with malvolean affinities. – *Taxon* **47**: 37–42.
- BREMER, B., 1996: Phylogenetic studies within *Rubiaceae* and relationships to other families based on molecular data. – *Opera Bot. Belg.* **7**: 33–50.
- BREMER, K., 1985: Summary of green plant phylogeny and classification. – *Cladistics* **1**: 369–385.

- 1988: The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. – *Evolution* **42**: 795–803.
 - 1994: Branch support and tree stability. – *Cladistics* **10**: 295–304.
- CHASE, M. W., SOLTIS, D. E., OLMSTEAD, R. G., MORGAN, D., LES, D. H., MISHLER, B. D., DUVAL, M. R., PRICE, R. Q., HILLS, H. G., QIU, Y.-L., KRON, K. A., RETTIG, J. H., CONTI, E., PALMER, J. D., MANHART, J. R., SYTSMA, K. J., MICHAELS, H. J., KRESS, W. J., KAROL, K. G., CLARK, W. D., HEDRÉN, M., GAUT, B. S., JANSEN, R. K., KIM, K.-J., WIMPEE, C. F., SMITH, J. F., FURNIER, G. R., STRAUSS, S. H., XIANG, Q.-Y., PLUNKETT, G. M., SOLTIS, P. S., SWENSEN, S., WILLIAMS, S. E., GADEK, P. A., QUINN, C. J., EGUIARTE, L. E., GOLENBERG, E., LEARN, G. H., JR., GRAHAM, S. W., BARRETT, S. C. H., DAYANANDAN, S., ALBERT, V. A., 1993: Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. – *Ann. Missouri Bot. Gard.* **80**: 528–580.
- DUVAL, M. R., HILLS, H. G., CONRAN, J. G., COX, A. V., EGUIARTE, L. E., HARTWELL, J., FAY, M. F., CADDICK, L. R., CAMERON, K. M., HOOT, S., 1995a: Molecular phylogenetics of *Liliana*. – In RUDALL, P. J., CRIBB, P. J., CUTLER, D. F., HUMPHRIES, C. J., (Eds): *Monocotyledons: systematics and evolution*, pp. 109–137. – Richmond: Royal Botanic Gardens, Kew.
 - STEVENSON, D. W., WILKIN, P., RUDALL, P. J., 1995b: Monocot systematics: a combined analysis. – In RUDALL, P. J., CRIBB, P. J., CUTLER, D. F., HUMPHRIES, C. J., (Eds): *Monocotyledons: systematics and evolution*, pp. 685–730. – Richmond: Royal Botanic Gardens, Kew.
 - RUDALL, P. J., CONRAN, J. G., 1996: New circumscriptions and a new family of asparagoid lilies: genera formerly included in *Anthericaceae*. – *Kew Bull.* **51**: 667–680.
- CONTI, E., LITT, A., SYTSMA, K. J., 1996: Circumscription of *Myrtales* and their relationships to other rosids: evidence from *rbcL* sequence data. – *Amer. J. Bot.* **83**: 221–233.
- CRONQUIST, A., 1981: An integrated system of classification of flowering plants. – New York: Columbia University Press.
- DAHLGREN, R. M. T., CLIFFORD, H. T., 1982: The monocotyledons: a comparative study. – London: Academic Press.
- YEO, P. F., CLIFFORD, H. T., 1985: The families of the monocotyledons. – Berlin, Heidelberg, New York: Springer.
- DOYLE, J. A., DONOGHUE, M. J., ZIMMER, E. A., 1994: Integration of morphological and ribosomal RNA data on the origin of angiosperms. – *Ann. Missouri Bot. Gard.* **81**: 419–450.
- DUVAL, M. R., CLEGG, M. T., CHASE, M. W., CLARK, W. D., KRESS, W. J., HILLS, H. G., EGUIARTE, L. E., SMITH, J. F., GAUT, B. S., ZIMMER, E. A., LEARN, G. H. JR., 1993: Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. – *Ann. Missouri Bot. Gard.* **80**: 607–619.
- FARRIS, J. S., 1988: Hennig86, version 1.5. Computer program and documentation. – Port Jefferson, New York.
- 1996: ‘Jac. Version 4.4.’ – Stockholm: Swedish Museum of Natural History.
 - ALBERT, V. A., KÄLLERSJÖ, M., LIPSCOMB, D., KLUGE, A. G., 1996: Parsimony jackknifing outperforms neighbor-joining. – *Cladistics* **12**: 99–124.
 - GOLOBOFF, P., KÄLLERSJÖ, M., OXELMAN, B., 1998: Improving the bootstrap. – *Cladistics* (in press).
- FAY, M. F., SWENSEN, S. M., CHASE, M. W., 1997: Taxonomic affinities of *Medusagyne oppositifolia* (*Medusagynaceae*). – *Kew Bull.* **52**: 111–120.
- FELSENSTEIN, J., 1985: Confidence limits on phylogenies: an approach using the bootstrap. – *Evolution* **39**: 783–791.
- FERNANDO, E. S., GADEK, P. A., CRAYN, D. M., QUINN, C. J., 1993: Rosid affinities of *Surianaceae*: molecular evidence. – *Molec. Phylogenet. Evol.* **2**: 344–350.

- QUINN, C. J., 1995: *Simaroubaceae*, an artificial construct: evidence from *rbcL* sequence variation. – *Amer. J. Bot.* **82**: 92–103.
- GOLENBERG, E. M., GIANNASI, D. E., CLEGG, M. T., SMILEY, C. J., DURBIN, M., HENDERSON, D., ZURAWSKI, G., 1990: Chloroplast DNA sequence from a miocene *Magnolia* species. – *Nature* **344**: 656–658.
- GOLOBOFF, P., 1993: Pee-Wee and NONA. Computer programs and documentation. – New York.
- GUSTAFSSON, M. H. G., BREMER, K., 1997: The circumscription and systematic position of *Carpodetaceae*. – *Austral. Syst. Bot.* **10**: 855–862.
- BACKLUND, A., BREMER, B., 1996: Phylogeny of the Asterales sensu lato based on *rbcL* sequences with particular reference to the *Goodeniaceae*. – *Pl. Syst. Evol.* **199**: 217–242.
- HASEBE, M., WOLF, P. G., PRYER, K. M., UEDA, K., ITO, M., SANO, R., GASTONY, G. J., YOKOYAMA, J., MANHART, J. R., MURAKAMI, N., CRANE, E. H., HAUFLE, C. H., HAUKE, W. D., 1995: Fern phylogeny based on *rbcL* nucleotide sequences. – *Amer. Fern J.* **85**: 134–181.
- HUBER H., 1969: Die Samenmerkmale und Verwandtschaftsverhältnisse der *Liliiflorae*. – *Mitt. Bot. Staatssamml. München* **8**: 219–538.
- JEFFREY, C., 1982: Kingdoms, codes and classification. – *Kew Bull.* **37**: 403–416.
- KÄLLERSJÖ, M., FARRIS, J. S., KLUGE, A. G., BULT, C., 1992: Skewness and permutation. – *Cladistics* **8**: 275–287.
- KRON, K. A., CHASE, M. W., 1993: Systematics of the *Ericaceae*, *Empetraceae*, *Epacridaceae* and related taxa based upon *rbcL* sequence data. – *Ann. Missouri Bot. Gard.* **80**: 735–741.
- KUMAR, S., TAMURA, K., NEI, M., 1993: MEGA: molecular evolutionary genetic analysis. Computer program and documentation. – University Park: Pennsylvania State University.
- MANHART, J. R., 1994: Phylogenetic analysis of green plant *rbcL* sequences. – *Molec. Phylogenet. Evol.* **3**: 114–127.
- MATTOX, K. R., STEWART, K. D., 1984: Classification of the green algae: a concept based on comparative cytology. – In IRVINE, D. E. G., JOHN, D., (Eds): *The systematics of green algae*, pp. 29–72. – London: Academic Press.
- MISHLER, B. D., LEWIS, L. A., BUCHHEIM, M. A., RENZAGLIA, K. S., GARBARY, D. J., DELWICHE, C. F., ZECHMAN, F. W., KANTZ, T. S., CHAPMAN, R. L., 1994: Phylogenetic relationships of the ‘green algae’ and ‘bryophytes’. – *Ann. Missouri Bot. Gard.* **81**: 451–483.
- MORGAN, D. R., SOLTIS, D. E., 1993: Phylogenetic relationships among members of *Saxifragaceae* sensu lato based on *rbcL* sequence data. – *Ann. Missouri Bot. Gard.* **80**: 631–660.
- SOLTIS, D. E., ROBERTSON, K. R., 1994: Systematic and evolutionary implications of *rbcL* sequence variation in *Rosaceae*. – *Amer. J. Bot.* **81**: 890–903.
- MORTON, C. M., CHASE, M. W., KAROL, K. G., 1997a: Taxonomic affinities of *Physena* (*Physenaceae*) and *Asteropeia* (*Asteropeiaceae*). – *Bot. Rev.* **63**: 231–239.
- KRON, K. A., SWENSEN, S. A., 1997b: A molecular evaluation of the monophyly of the order *Ebenales* based upon *rbcL* sequence data. – *Syst. Bot.* **21**: 567–586.
- MORI, S. A., PRANCE, G. T., KAROL, K. G., CHASE, M. W., 1997c: Phylogenetic relationships of *Lecythidaceae*: a cladistic analysis using *rbcL* sequences and morphological data. – *Amer. J. Bot.* **84**: 530–540.
- NIXON, K. C., CREPET, W. L., STEVENSON, D., FRIIS, E. M., 1994: A reevaluation of seed plant phylogeny. – *Ann. Missouri Bot. Gard.* **81**: 484–533.
- OLMSTEAD, R. G., REEVES, P. A., 1995: Evidence for the polyphyly of the *Scrophulariaceae* based on chloroplast *rbcL* and *ndhF* sequences. – *Ann. Missouri Bot. Gard.* **82**: 176–193.

- BREMER, B., SCOTT, K. M., PALMER, J. D., 1993: A parsimony analysis of the *Asteridae* sensu lato based on *rbcL* sequences. – *Ann. Missouri Bot. Gard.* **80**: 700–722.
- PLUNKETT, G. M., SOLTIS, D. E., SOLTIS, P. S., 1996: Higher level relationships of *Apiales* (*Apiaceae* and *Araliaceae*) based on phylogenetic analysis of *rbcL* sequences. – *Amer. J. Bot.* **83**: 399–415.
- PRICE, R. A., PALMER, J. D., 1993: Phylogenetic relationships of the *Geraniaceae* and *Geraniales* from *rbcL* sequence comparisons. – *Ann. Missouri Bot. Gard.* **80**: 661–671.
- QIU, Y.-L., CHASE, M. W., LES, D. H., PARKS, C. R., 1993: Molecular phylogenetics of the *Magnoliidae*: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. – *Ann. Missouri Bot. Gard.* **80**: 587–606
- RICE, K. A., DONOGHUE, M. J., OLMSTEAD, R. G., 1997: Analyzing large data sets: *rbcL* 500 revisited. – *Syst. Biol.* **46**: 554–563.
- RITLAND, K., CLEGG, M. T., 1987: Evolutionary analyses of plant DNA sequences. – *Amer. Naturalist* **130**: S74–S100.
- RODMAN, J. E., PRICE, R. A., KAROL, K., CONTI, E., SYTSMA, K. J., PALMER, J. D., 1993: Nucleotide sequences of the *rbcL* gene indicate monophyly of mustard oil plants. – *Ann. Missouri Bot. Gard.* **80**: 686–699.
- RUDALL, P. J., CUTLER, D. F., 1995: *Asparagales*: a reappraisal. – In RUDALL, P. J., CRIBB, P. J., CUTLER, D. F., HUMPHRIES, C. J., (Eds): *Monocotyledons: systematics and evolution*, pp. 157–168. – Richmond: Royal Botanic Gardens, Kew.
- CHASE, M. W., 1996: Systematics of *Xanthorrhoeaceae* sensu lato: evidence for polyphyly. – *Telopea* **6**: 629–647.
- SANDERSON, M. J., 1989: Confidence limits on phylogenies: the bootstrap revisited. – *Cladistics* **5**: 113–129.
- SAVOLAINEN, V., SPICIGER, R., MANEN, J.-F., 1997: Polyphyly of *Celastrales* deduced from a chloroplast noncoding DNA region. – *Molec. Phylogenet. Evol.* **7**: 145–157.
- SHEAHAN, M. C., CHASE, M. W., 1996: A phylogenetic analysis of *Zygophyllaceae* R. Br. based on morphological, anatomical and *rbcL* sequence data. – *Bot. J. Linn. Soc.* **122**: 279–300.
- SMITH, J. F., KRESS, W. J., ZIMMER, E. A., 1993: Phylogenetic analysis of the *Zingiberales* based on *rbcL* sequences. – *Ann. Missouri Bot. Gard.* **80**: 620–630.
- SWOFFORD, D., 1993: PAUP, phylogenetic analysis using parsimony, version 3.1.1. – Champaign, Illinois: Illinois Natural History Survey.
- TAKHTAJAN, A. L., 1994: New families of the monocotyledons. – *Bot. Zhurn.* **79**: 65–66.
- WOLF, P. G., 1995: Phylogenetic analyses of *rbcL* and nuclear ribosomal RNA gene sequences in *Dennstaedtiaceae*. – *Amer. Fern J.* **85**: 306–327.
- ZURAWSKI, G., CLEGG, M. T., 1987: Evolution of higher-plant chloroplast DNA-coded genes: implications for structure-function and phylogenetic studies. – *Ann. Rev. Pl. Physiol.* **38**: 391–418.

Addresses of the authors: MARI KÄLLERSJÖ, JAMES S. FARRIS, Molecular Systematics Laboratory, Swedish Museum of Natural History, P.O. Box 50007, S-104 05 Stockholm, Sweden. – MARK W. CHASE, MICHAEL F. FAY, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, U.K. – BIRGITTA BREMER, KÅRE BREMER, Department of Systematic Botany, Uppsala University, Villavägen 6, S-723 36 Uppsala, Sweden. – CHRISTOPHER J. HUMPHRIES, Department of Botany, The Natural History Museum, Cromwell Road, London SW7 5BD, U.K. – GITTE PETERSEN, OLE SEBERG, Botanical Institute, University of Copenhagen, Gothersgade 140, DK-1123 Copenhagen K, Denmark.