American Journal of Botany 90(12): 1758–1776. 2003.

ANGIOSPERM PHYLOGENY BASED ON *matK* SEQUENCE INFORMATION[1]

KHIDIR W. HILU,[2,14] THOMAS BORSCH,[3] KAI MÜLLER,[3]
DOUGLAS E. SOLTIS,[4] PAMELA S. SOLTIS,[5] VINCENT SAVOLAINEN,[6]
MARK W. CHASE,[6] MARTYN P. POWELL,[6] LAWRENCE A. ALICE,[7]
RODGER EVANS,[8] HERVÉ SAUQUET,[9] CHRISTOPH NEINHUIS,[10]
TRACEY A. B. SLOTTA,[2] JENS G. ROHWER,[11]
CHRISTOPHER S. CAMPBELL,[12] AND LARS W. CHATROU[13]

[2]Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 USA; [3]Botanisches Institut, Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany; [4]School of Biological Sciences, Washington State University, Pullman, Washington 99164 USA; [5]Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida 32611 USA; [6]Molecular Systematics Section, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK; [7]Department of Biology, Western Kentucky University, Bowling Green, Kentucky 42101 USA; [8]Biology Department, Acadia University, 24 University Avenue, Wolfville, Nova Scotia, Box 1X0, Canada; [9]Muséum National d'Histoire Naturelle, 16, rue Buffon, 75005 Paris, France; [10]Institut für Botanik, Zellescher Weg 22, 01062 Dresden, Germany; [11]Institut für Allgemeine Botanik, Universität Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany; [12]Department of Biological Sciences, University of Maine, Orono, Maine 04469-5722 USA; and [13]National Herbarium of the Netherlands, Utrecht University Branch, Heidelberglaan 2, 3584 CS Utrecht, Netherlands

Plastid *matK* gene sequences for 374 genera representing all angiosperm orders and 12 genera of gymnosperms were analyzed using parsimony (MP) and Bayesian inference (BI) approaches. Traditionally, slowly evolving genomic regions have been preferred for deep-level phylogenetic inference in angiosperms. The *matK* gene evolves approximately three times faster than the widely used plastid genes *rbcL* and *atpB*. The MP and BI trees are highly congruent. The robustness of the strict consensus tree supercedes all individual gene analyses and is comparable only to multigene-based phylogenies. Of the 385 nodes resolved, 79% are supported by high jackknife values, averaging 88%. *Amborella* is sister to the remaining angiosperms, followed by a grade of Nymphaeaceae and Austrobaileyales. Bayesian inference resolves *Amborella* + Nymphaeaceae as sister to the rest, but with weak (0.42) posterior probability. The MP analysis shows a trichotomy sister to the Austrobaileyales representing eumagnoliids, monocots + Chloranthales, and *Ceratophyllum* + eudicots. The *matK* gene produces the highest internal support yet for basal eudicots and, within core eudicots, resolves a crown group comprising Berberidopsidaceae/Aextoxicaceae, Santalales, and Caryophyllales + asterids. Moreover, *matK* sequences provide good resolution within many angiosperm orders. Combined analyses of *matK* and other rapidly evolving DNA regions with available multigene data sets have strong potential to enhance resolution and internal support in deep level angiosperm phylogenetics and provide additional insights into angiosperm evolution.

Key words: angiosperms; Bayesian inference; *matK*; phylogeny; systematics.

Phylogenetic analysis of gene sequences has significantly impacted views of angiosperm relationships (Dahlgren, 1980; Takhtajan, 1987; Cronquist, 1988; Thorne, 1992). Consequently, the overall phylogeny of angiosperms has been radically revised at all levels. Some subclasses, such as Dilleniidae and Hamamelidae, have been shown to be polyphyletic with their constituent families now placed (APG, 1998; APG II, 2003) in several distantly related clades. The composition of other groups has also been altered to varying degrees, e.g., Rosidae, Asteridae, Ericales, Cornales, and Saxifragales. Contributions toward this reassessment of angiosperm phylogeny have come primarily from large data sets of individual genes or combined analyses of these data sets (e.g., Chase et al., 1993; Qiu et al., 1998, 1999, 2000; Hoot et al., 1999; Soltis et al., 1999, 2000, 2003; Olmstead et al., 2000; Savolainen et al., 2000a, b; Zanis et al., 2002). In addition, extensive analyses of morphological, anatomical, and phytochemical characters from across angiosperm families (Nandi et al., 1998) have also contributed to modern views of angiosperm relationships. Consequently, a new concept for the overall phylogeny of flowering plants has emerged, depicting a basal grade of Amborellaceae, Nymphaeaceae (sensu APG II, 2003), and Austrobaileyales, followed by eumagnoliids (sensu APG II, 2003, to include Canellales, Laurales, Magnoliales, and Piperales), monocots, Ceratophyllales, Chloranthaceae, and eudicots. However, a number of questions remain unanswered due to variable or unresolved positions and weak support for various lineages. This situation is particularly true for the eudicots, which constitute about 75% of angiosperm species diversity (Drinnan et al., 1994). Among eudicots, the basal grade lacks convincing bootstrap (BS)/jackknife (JK) support (Qiu et al., 1998; Hoot et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003). Moreover, relationships among the major clades of core eudicots (i.e., Berberidopsidaceae/Aextoxicaceae, Saxifraga-

[1] Manuscript received 27 March 2003; revision accepted 3 July 2003.

The authors thank Stefan Wanke of University of Bonn for the sequence of *Lactoris*, Prof. Wilhelm Barthlott for his continuous support of this project and molecular systematics at the Botanical Institute in Bonn, and the Deep Time Research Coordination Network for facilitating interaction among coauthors. The work is supported in part by DFG grants BO 1815-1/1 to T. B. and Ro700/3 to J. G. R., NSF grants DEB 9634231 to K. W. H. and DEB 9806945 to C. S. C., and the Netherlands Organisation for Scientific Research grant S85-324 to L. W. C.

[14] E-mail: hilukw@vt.edu.

les, Caryophyllales, rosids, asterids, and Santalales) remain uncertain (Hoot et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003). Phylogenetic relationships among rosids also remain unclear (Savolainen et al., 2000a; Soltis et al., 2000, 2003). For basal angiosperms, there are still questions concerning the position of eumagnoliids, monocots, Ceratophyllaceae, and Chloranthaceae. In addition, alternative hypotheses, albeit with weak support, for the position of *Amborella* as sister to all other angiosperms have emerged, depicting waterlilies alone or along with *Amborella* in that position (Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002). Resolving relationships among these groups is not only essential for a comprehensive systematic treatment of angiosperms, but also for understanding patterns of species diversification and character evolution.

Angiosperm phylogenetic studies based on individual genes have faced two difficulties: limited resolution and low internal support for major clades and topological incongruence (Olmstead and Sweere, 1994; Soltis et al., 1997, 2000, 2003; Mathews and Donoghue, 1999, 2000; Savolainen et al., 2000a). Combining data sets in multigene analyses improved resolution and internal support (Soltis et al., 2000, 2003; Parkinson et al., 1999; Graham and Olmstead, 2000; Qiu et al., 2000; Savolainen et al., 2000a, b; Zanis et al., 2002; Sauquet et al., 2003). Combined analyses of genes from different subcellular compartments are considered to be a good method to estimate organismal phylogeny (e.g., Donoghue and Sanderson, 1992; Hillis, 1996, 1998; Kim, 1998), a view supported by empirical studies (Qiu et al., 1999, 2000; Soltis et al., 2000, 2003; Zanis et al., 2002). Consequently, the consensus tree based on combined *rbcL, atpB,* and 18S rDNA sequences of Soltis et al. (2000; henceforth referred to as the three-gene analysis) may be considered the most reliable overall angiosperm phylogeny so far available.

In the majority of broad angiosperm phylogenetic studies, authors have emphasized using sequence information from slowly evolving genes based on the notion that the number of multiple hits and levels of homoplasy are expected to be relatively low (Farris, 1977; Swofford et al., 1996; Olmstead et al., 1998; Graham et al., 2000). However, use of slowly evolving genomic regions can result in severe limitations in taxon sampling due to need for sequencing a large number of nucleotides per species to obtain sufficient number of variable characters. Consequently, it restricts the number of taxa that can reasonably be sequenced and analyzed cladistically, introducing a new set of phylogenetic problems as pointed out in several recent studies (e.g., Graybeal, 1998; Rannala et al., 1998; Pollock et al., 2002). For example, Graham and Olmstead (2000) sequenced 13.4 kilobases (kb) of slowly evolving cpDNA genes, but as a result could only include 19 taxa in a study of basal angiosperms. This raises the problem of taxon density, an issue addressed by several authors (e.g., Graybeal, 1998; Hillis, 1998; Bremer et al., 1999; Zwickl and Hillis, 2002). Therefore, genomic regions that can provide sufficient signal in deep level phylogeny reconstruction without compromising taxon representation are essential for accurate assessment of evolutionary histories. The rapidly evolving *matK* gene satisfies these prerequisites.

The *matK* gene is ~1600 base pairs (bp) in most angiosperms, located within the *trnK* intron, and functionally may be involved in splicing group II introns coding for tRNALys (UUU; Neuhaus and Link, 1987; Ems et al., 1995). Believed to code for a maturase based on structural similarities to other such genes (Neuhaus and Link, 1987; Mohr et al., 1993), *matK* is the only maturase of higher plant plastids (Vogel et al., 1997). The *trnK* intron, including the *matK* exon, is transcribed in one piece (Chiba et al., 1996) and is expressed at the protein level in *Solanum* (Du Jardin et al., 1994). The *matK* open reading frame (ORF) is maintained intact except at the 3′ end where frameshift substitutions slightly alter the length with apparently minimal impact on function (Hilu and Alice, 1999). These data and the analysis of the RNA-binding activity of a *trnK*-encoded polypeptide from *Sinapis* (Liere and Link, 1995) further support a *matK* function in splicing group II introns. The presence of *matK* as a free-standing ORF in the plastid genome of the parasitic *Epifagus virginiana* (Ems et al., 1995), which has lost ~65% of its genes (Wolfe et al., 1992), also points to the functional significance of *matK*.

The *matK* gene stands out among genes used in angiosperm systematics in its substantially greater number of: (1) nucleotide substitutions, (2) nonsynonymous mutations, and (3) insertion/deletion events or indels (Johnson and Soltis, 1994, 1995; Olmstead and Palmer, 1994; Hilu and Liang, 1997; Soltis and Soltis, 1998; K. W. Hilu, K. Müller, and T. Borsch, unpublished data). The gene also exhibits a relatively high proportion of transversions, with the transition/transversion ratio (ti/tv) approaching unity (Olmstead and Palmer, 1994; Hilu and Liang, 1997). The percentage amino acid substitution for *matK* between the monocot rice and the eudicot tobacco is up to sixfold higher than for *rbcL* and *atpB* (41% vs. 7–8%; Olmstead and Palmer, 1994). Among-site rate variability for the three codon positions shows that *matK* is not skewed toward the third position as is the case in most protein-coding genes used in angiosperm systematics. Substitution rates in the first and second codon positions in *matK* approach those of the third position (Johnson and Soltis, 1994, 1995; Hilu and Liang, 1997; Hilu et al., 1999), a situation that elevates the rate of nonsynonymous changes. These data point to either a low correlation between structure and function with a rather small core being functionally important (e.g., domain X; Mohr et al., 1993; Hilu and Liang, 1997), or that the enzyme's function as a maturase might require a particular stereochemistry in which the actual amino acid sequence is of reduced importance. Therefore, *matK* has evolutionary patterns and tempo that distinguish it from most genes used in angiosperm phylogeny reconstruction (Olmstead and Palmer, 1994; Hilu and Liang, 1997).

Some of these attributes of *matK* may have discouraged researchers from using *matK* sequences in broad studies such as overall angiosperm relationships. Another reason for infrequent use of *matK* at broad levels may be that taxon-specific primers are usually required. The location of *matK* within the *trnK* intron and its close proximity to *psbA* provide nearly universal primers for its amplification, and the need to design primers for sequencing is counterbalanced by the quality of the data provided. Effective sequencing strategies for *matK* are discussed in Materials and Methods.

This analysis provides an angiosperm tree based on the largest data set so far compiled for *matK*. We compare the topology obtained with this gene to previously published topologies based on single gene and multigene data sets. We also examine patterns of variability in *matK*. A parsimony approach has been chosen for data analysis to allow for direct comparison with the three-gene analysis of Soltis et al. (2000). We evaluated the effect of including Gnetales as an outgroup on the

topology; the angiosperms and Gnetales represent the two most divergent groups of seed plants (e.g., Bowe et al., 2000). In addition, a Bayesian analysis (Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2002) was performed.

MATERIALS AND METHODS

Taxon sampling and plant material—This study includes representatives of 374 angiosperm genera from 240 families and all orders recognized by APG II (2003) and 12 gymnosperm genera (Appendix 1; see Supplemental Data accompanying the online version of this article). Large families are represented wherever possible by more than one genus. A large proportion of the *matK* sequences was generated specifically for this study, and additional sequences were taken from GenBank (Appendix 1).

DNA isolation, polymerase chain reaction (PCR) amplification, and sequencing—Total cellular DNA was isolated from fresh, silica-dried, or herbarium specimens using the hexadecyltrimethylammonium bromide (CTAB) procedure of Doyle and Doyle (1987) or its modification (Borsch et al., 2003). Because a large number of the sequences available cover a region from around position 400 through the stop codon of *matK*, this region (~1200 bp) became the focus of this study to avoid potential problems associated with large amounts of missing data. For gene amplification, either the entire *trnK* intron was amplified using primers trnK3914F and trnK2R (Johnson and Soltis, 1995) or, in most cases, the 3′-two-thirds of the *trnK* intron was amplified with a forward primer located approximately 480 bp into the coding region and trnK2R (for information on primers see Appendix 2 in Supplemental Data accompanying the online version of this article). By amplifying this region, the PCR primers could also be used for sequencing because primer annealing was guaranteed within the otherwise rather variable coding region of *matK*. Primer NYmatK480F (Borsch et al., 2003), originally designed for *Nymphaea*, turned out to be useful for many angiosperms. For highly divergent taxa, such as *Gnetum* and *Welwitschia*, the whole *trnK* intron was amplified and sequenced first with the amplification primers; specific internal sequencing primers were subsequently designed by ''walking'' into the region. For some taxa, internal primers 390F and 1326R were used (Johnson and Soltis, 1994, 1995). Cycle sequencing was performed using a Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, California, USA), and extension products were electrophoresed on ABI 310, 373, and 377 automated sequencers (Applied Biosystems).

Sequence alignment and phylogenetic analysis—Sequences were aligned using QuickAlign (Müller, 2003) or ClustalX (Thompson et al., 1997) followed by manual adjustments. All sequences were translated into amino acids and their ORFs checked. Several in-frame gaps were inserted to align the sequences. Frame-shift mutations near the stop codon formed a mutational hot spot 15 bp upstream of the stop codon; this section was excluded from the analysis. Due to differences in amplification procedures used by the collaborators, the sequences of some taxa lack ca. 200 bp at the 3′ end of the *matK* gene.

All aligned positions were given equal weight, and gaps were treated as missing data. Parsimony analyses were conducted using PAUP* *4.0b6* (Swofford, 2001) and PRAT (Müller, 2002). PRAT is a program written for this study; it generates command files that execute parsimony ratchet searches (Nixon, 1999) using PAUP*. Program options include random addition cycles of the ratchet and parsimony jackknifing, applying the ratchet in each replicate. In this study, 10 random addition cycles of 150 ratchet iterations each were used. Each iteration is comprised of two rounds of tree-bisection-reconnection (TBR) swapping, one on a randomly reweighted data set and the other on the original matrix, saving one minimum-length tree. Random upweighting affected 25% of the positions. Because each random addition cycle soon converged on the same tree score, cycles were not extended beyond 150 iterations and further cycles were not added. Shortest trees collected from the different tree islands were subjected to a final TBR swapping with 5000 saved trees, from which a strict consensus tree was computed. To estimate internal support, parsimony jackknifing with 500 cycles was carried out according to the approach and parameters suggested by Farris et al. (1996) for large data sets, with TBR swapping on five saved trees per cycle. The deleted fraction of characters was e^{-1}, which means that bootstrap frequencies agree with jackknife frequencies (Farris et al., 1996). This allows us to compare jackknife support values obtained here with bootstrap values reported in other studies. We also compare jackknife values from studies that used the same deletion percentage, e.g., by employing the program JAC (J. S. Farris, unpublished program). Two searches were performed on the *matK* data set; one on a matrix that included Gnetales among the outgroup taxa (matrix A) and the other on a matrix that excluded this order (matrix B). The second analysis was performed to evaluate the potential effects of Gnetales on the analysis.

Bayesian inference used the program MrBayes (Huelsenbeck and Ronquist, 2001). Calculations of likelihood were based on a general time reversible model of nucleotide substitution, assuming different stationary nucleotide frequencies and site-specific rate categories for each codon position. The posterior probability (PP) was estimated by sampling trees from the PP distribution, using Metropolis-coupled Markov chain Monte Carlo simulations. Four chains were run for 500 000 generations, starting with one of the shortest trees found with the parsimony ratchet, and the temperature of heated chains was set to 0.2. Chains were sampled every 10 generations. Likelihood scores converged on a stable value after generation 100 000 (the ''burn in'' of the chain), and calculations of PP were based upon the trees sampled after this generation.

Number of steps and consistency, retention, and rescaled consistency indices (CI, RI, and RC, respectively;) (Kluge and Farris, 1969; Farris, 1989) for the three codon positions of *matK* were calculated with PAUP*. Lists of steps for each codon position were subjected to the nonparametric Mann-Whitney *U* test to evaluate differences in nucleotide substitutions at these positions. Because the underlying sample distribution is largely unknown, no parametric test was applied.

RESULTS

Sequence variability and substitution patterns—The ~1200-bp sequenced region of *matK* resulted in 1749 aligned characters due to the insertion of gaps. Except for the 15-bp region upstream of the stop codon, all indels occurred in multiples of three nucleotides (up to 9 bp in length). Three genera in Caryophyllales (*Anredera*-Basellaceae, *Halophytum*-Halophytaceae, and *Rhipsalis*-Cactaceae) had an inversion 6–24 bp in length. Due to its location in a palindromic region, the actual size of the inversion could not be determined. Inversions are often associated with such palindromic motifs (see Graham et al., 2000; Kelchner, 2000). Of the aligned characters, 1221 (70%) are variable and 1083 (62%) are potentially parsimony-informative (based on matrix A). The distribution of variable sites among codon positions is 414, 386, and 421 for the first, second, and third codon positions, respectively. The overall nucleotide *p* distance is 0.216 and translates into an amino acid *p* distance of 0.339 using MEGA (Kumar et al., 2001). Thus, amino acid variation in *matK* is higher than nucleotide variability. The *p* distance at synonymous sites is 0.351, which is twice as high as the *p* distance at nonsynonymous sites (0.176). Based on unambiguous transitions and transversions traced on a single tree from matrix A, the ti/tv ratio in *matK* is 1.275. According to Holmquist (1983), a ratio of 0.4 and below is an indication of highly saturated sequences, a ratio that is certainly not reached here. However, saturation is a complex issue, and its magnitude may differ depending on nucleotide and codon positions along a genomic region; a more complete analysis will be presented elsewhere. Nevertheless, the ratio obtained in this analysis does not point towards a high level of saturation in *matK*. Base substitutions are fairly evenly distributed across the length of *matK* (Fig. 1). The low nucleotide variability depicted for the end of the

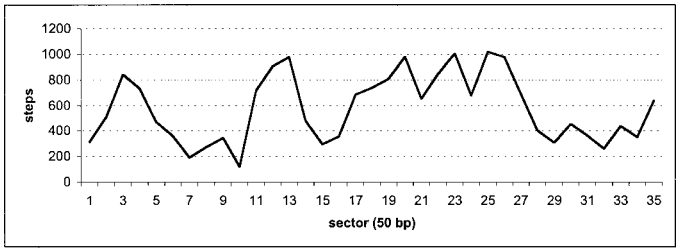

Fig. 1. Distribution of substitutional changes along the coding region, starting with position ~500 in *matK* (begin sector 1) and ending at the stop codon (end sector 36). Sectors represent 50 bp.

sequenced region corresponds in part to the conserved domain *X* (Mohr et al., 1993; Hilu and Liang, 1997). However, this decrease in variability may be accentuated by the amount of missing data at the end (see Material and Methods).

Measured on one of the shortest trees, the number of steps is greater at third positions compared to first and second positions (Table 1). Moreover, in third codon positions, the overall level of homoplasy is higher than in second positions, but lower than in first. The *U* test shows that equality of the distributions for steps at third positions vs. those at first and second positions can be rejected (Table 1, $P < 0.0001$).

Phylogenetic results—Parsimony trees were 20 646 and 20 192 steps in length for matrices A and B, respectively. The CI, RI, and RC values were identical to two decimal places in both searches (CI = 0.14, RI = 0.64, RC = 0.09). In matrix A, the performance of the ratchet using PAUP* and PRAT was compared to the strategy of random addition replicates, saving a limited number of trees per cycle (maxtrees = 1000, nreps = 1000). The latter approach resulted in minimum-length trees of 20 651 steps after several weeks of computation on a 350 Mhz Macintosh G4. In contrast, shorter trees were encountered after only 3 min when PAUP* executed a PRAT command file on the same computer; tree collection from 1500 islands was completed in about 22 h. A summary tree containing the major angiosperm lineages based on the strict consensus of 5000 trees from matrix A is provided in Fig. 2, with detailed strict consensus trees depicted in Figs. 4–12. A total of 305 of 385 nodes (79%) receive jackknife support greater

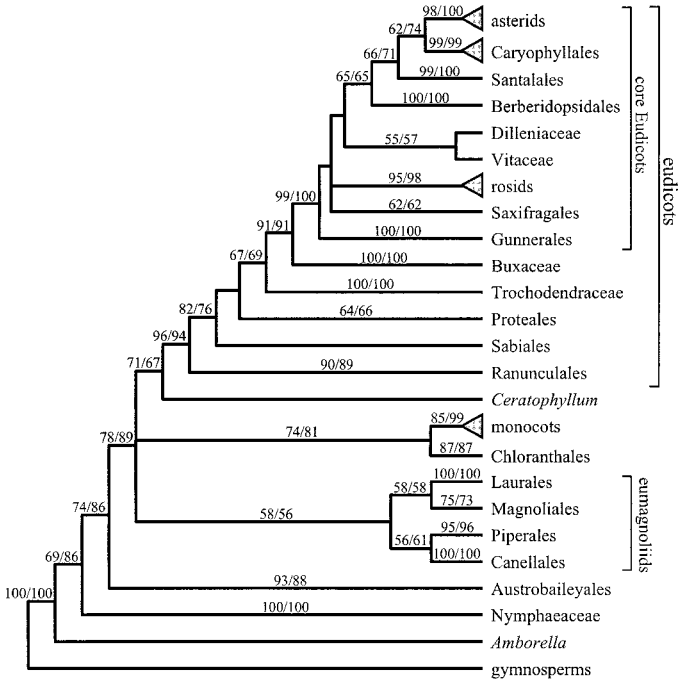

Fig. 2. Summary of angiosperm strict consensus tree based on parsimony analysis of *matK* gene sequences using gymnosperms as the outgroup. Numbers above branches are jackknife values derived from heuristic-based searches on matrices A and B (A/B, Gnetales included or excluded, respectively, from outgroup). Large clades are indicated by triangles.

than 50%, and support levels average 88%. Deletion of Gnetales (matrix B) increased support for some clades (Fig. 2). Homoplasy levels are comparable to those obtained with the analyses of other large matrices. For example, Soltis et al. (2000) reported a CI of 0.12 for their combined three-gene data set (567 taxa) compared to a CI of 0.14 here (374 taxa).

Overall internal support for a tree based on third codon positions is higher than for trees produced from analyzing first or second positions only (Table 1). In the former case, 234 nodes received jackknife support greater than 50%, whereas first and second positions yield 223 and 180 supported nodes, respectively. These data underscore the phylogenetic utility of

TABLE 1. Characteristics of the different codon positions in *matK*. Values are based on the first shortest tree (20 192 steps) found in search B. Jackknife support for phylogenies based on individual codon positions was estimated as described in text, using 100 replicates. Support values in the "Total" column are based on the 500 jackknife replicates of search A. Pi, parsimony informative; CI, consistency index; RI, retention index; RC, rescaled consistency index; *U*, *U* statistic derived from the *U* test, comparing steps at first and second codon positions with those at third codon positions; *P*, corresponding probabilities ($P < 0.0001$ in both tests).

Codon	Position 1	Position 2	Position 1 + 2	Position 3	Total
Characters	583	583	1166	583	1749
Variable characters	414	386	800	421	1221
% Variable	71%	66%	69%	72%	70%
Pi	362	334	696	387	1083
% Pi	62%	57%	60%	66%	62%
CI	0.149	0.176	0.161	0.114	0.140
RI	0.638	0.646	0.642	0.634	0.638
RC	0.095	0.114	0.103	0.072	0.089
Steps	6459	4958	11 417	9384	20 801
U (vs. 3rd)	31 075	25 327	—	—	—
P values	2.204×10^{-5}	7.048×10^{-10}	—	—	—
Supported nodes	223	180	273	234	305
% Average support	81.2	82.6	85.3	83.6	88.4

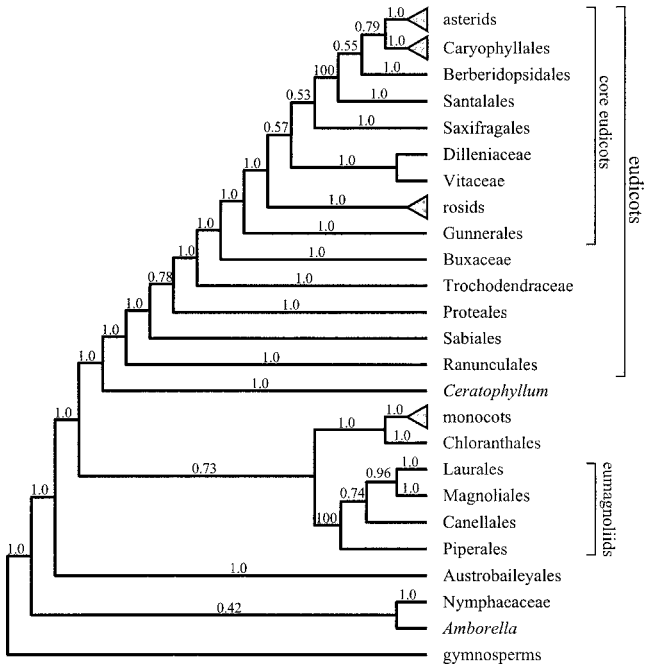

Fig. 3. Summary of angiosperm phylogeny based on Bayesian inference from the *matK* gene sequences using gymnosperms as the outgroup. Numbers above branches are posterior probabilities. Large clades are indicated by triangles.

third codon positions as demonstrated by Chase et al. (1995), Källersjö et al. (1999), Savolainen et al. (2000a), and others and are correlated with the number of informative sites at the different positions (66% in third vs. 62% and 57% in first and second positions). In addition, levels of homoplasy (CI, Table 1) are not on average higher in 3rd codon positions than in first and second positions. Therefore, differences in numbers of steps, variable sites, and phylogenetic structure are not as high among codon positions in *matK* as they are in more conserved genes such as *rbcL* (Källersjö et al., 1999; K. W. Hilu, K. Müller, and T. Borsch, unpublished data).

Analyses of matrices A and B result in consensus trees with identical topologies, although jackknife support for the basal nodes increases when Gnetales are excluded (Fig. 2). The Bayesian analysis produced a tree that is similar in topology to the MP tree (Fig. 3). Noteworthy differences in the Bayesian tree compared to the parsimony tree are the sister group relationship of *Amborella* + Nymphaeaceae to the remaining angiosperms in the former, though support is low (0.42 PP), and the positions of Canellales/Piperales and Saxifragales (Figs. 2, 3). The discussion is based on the MP matrix A; results from both matrix B and Bayesian analysis will be contrasted only when relevant. Bootstrap and/or jackknife support of 50–74% is considered low, 75–84% moderate, and >85% high (Chase et al., 2000). Posterior probabilities for nodes in the Bayesian tree are interpreted as reliable when above 0.95. In cases for which considerable differences exist in JK percentages between matrices A and B, both percentages are reported as A/B.

The MP tree depicts the angiosperms as monophyletic (100%) with the New Caledonian *Amborella trichopoda* (Amborellaceae) as sister to the rest of the flowering plants (69/86%), followed successively by Nymphaeaceae and an *Austrobaileya-Illicium-Schisandra* (Austobaileyales) clade as sister to the remaining angiosperms. These lineages correspond to the ANITA grade sensu Qiu et al. (2000). Monocots are (85/99%) sister to *Chloranthus* (74/81%). A weakly supported eumagnoliid clade (58/56%) consists of Piperales + Canellales (56/61%) and Laurales + Magnoliales (58%). *Ceratophyllum* is sister to eudicots (71%). The eumagnoliids, monocots + Chloranthaceae, and *Ceratophyllum* + eudicots diverge after the Austrobaileyales and their relationships are unresolved in the strict consensus tree. Eudicots are strongly supported (96%) and include a basal grade of Ranunculales, Sabiaceae, Proteales, Trochodendraceae, and Buxaceae (including Didymelaceae; APG II, 2003) that are subsequent sister to the core eudicots. The sister group relationship of Ranunculales to the remaining eudicots and Buxaceae to the core eudicots receive the highest support yet (82% and 91%, respectively). Within the core eudicots, Gunnerales are sister to a trichotomy of Saxifragales, rosids, and a clade comprising Dilleniaceae + Vitaceae, Berberidopsidaceae + Aextoxicaceae, Santalales, and Caryophyllales + asterids. Internal support is low for many core eudicot lineages.

DISCUSSION

Early-diverging angiosperms—Phylogenetic relationships among early-diverging (basal) angiosperms have been well studied due to their importance in understanding character evolution and early diversification in angiosperms (Parkinson et al., 1999; Qiu et al., 1999, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002; Borsch et al., 2003). Most recent phylogenetic analyses of angiosperms, including this study, have converged on a basal assemblage of lineages characterized by monosulcate or monosulcate-derived pollen. This is in contrast to the eudicots, a group that comprises the remaining angiosperms defined by their triaperturate or triaperturate-derived pollen. Amborellaceae, Nymphaeaceae, and Austrobaileyales have emerged with strong support as successive sisters to other angiosperms (Figs. 2–3). In contrast, relationships among eumagnoliids, monocots, Chloranthaceae, and Ceratophyllaceae remain uncertain.

The position of *Ceratophyllum* has varied in previous analyses from sister to all angiosperms, monocots, Chloranthaceae, eudicots, or in an unresolved position (Chase et al., 1993; Qiu et al., 2000; Savolainen et al., 2000a; Zanis et al., 2002). Similarly, Chloranthaceae has appeared as sister to either monocots or eudicots or in an unresolved position with other lineages (Chase et al., 1993; Mathews and Donoghue, 1999; Graham and Olmstead, 2000; Savolainen et al., 2000a; Soltis et al., 2000; Zanis et al., 2002). Increased number of characters has not enhanced our understanding of relationships among these lineages. The eumagnoliids, monocots, *Ceratophyllum*, and Chloranthaceae were unresolved in the five-gene study of Qiu et al. (2000). The combined 11-gene data set (Zanis et al., 2002) showed *Ceratophyllum* + monocots diverging after Illiciales (= Austrobaileyales), followed by Chloranthaceae, but with only 52% BS as sister to a clade comprising eumagnoliids + eudicots. A recent analysis of combined sequence data for basal angiosperms from the relatively fast-evolving noncoding *trnT-trnF* plus *matK* (K. W. Hilu, K. Müller, and T. Borsch, unpublished data) shows 99–100% JK support for the Amborellaceae, Nymphaeaceae, and Austrobaileyales grade, but weak support for the relationships among the rest. Therefore, resolving the relationships among these enigmatic lineages re-

mains one of the major challenges in angiosperm phylogenetics.

Amborellaceae, Nymphaeaceae, and Austrobaileyales—A general consensus exists on the branching pattern of these three most basal nodes (Parkinson et al., 1999; Qiu et al., 1999, 2000; Soltis et al., 1999, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002; Borsch et al., 2003). Nevertheless, alternative relationships depicting Nymphaeaceae alone or *Amborella* + Nymphaeaceae as sister to the rest have been recovered by some methods of analysis (Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Qiu et al., 2000; Zanis et al., 2002). The position of *Amborella* as the first branching angiosperm is supported here with 69% JK value in MP (matrix A, Fig. 4). Exclusion of Gnetales from the outgroup (matrix B) increases support for this node to 86%, comparable to analyses based on five genes (Qiu et al., 1999, 2000) and 11 genes (Zanis et al., 2002). Similar JK results were obtained for the Nymphaeaceae as successive sister to the rest (Fig. 2). The three-gene analysis of Soltis et al. (2000) provided only moderate jackknife support for the same topology. The jackknife support achieved by *matK* is substantial considering that the number of variable characters in this *matK* data set is less than one quarter of those analyzed in the three-gene matrix. Among other individual genes, only 18S rDNA (Soltis et al., 2000; <50% JK), *atpB* (Savolainen et al., 2000a; <50% BS), and *trnT-trnF* (Borsch et al., 2003; 94–100% JK) identify *Amborella* and Nymphaeaceae in these positions.

In contrast with the MP topology, BI places *Amborella* + Nymphaeaceae as sister to all other angiosperms, although support for this clade is very low (0.42 PP, Fig. 3). The meaning of such a low posterior probability is particularly dubious given the recent simulations of Suzuki et al. (2002). Kishino-Hasegawa tests (Kishino and Hasegawa, 1989) carried out by Parkinson et al. (1999) and Qiu et al. (2000) on other matrices of combined genes could not reject this position for *Amborella* + Nymphaeaceae. Neighbor-joining analyses of the six- and nine-gene matrices (Barkman et al., 2000) in which ''noisy'' positions were removed with relative apparent synapomorphy analysis (RASA; Lyons-Weiler et al., 1996) provided high bootstrap support value for an *Amborella* + Nymphaeaceae clade. However, recent studies suggest that RASA introduces errors when used in removing noisy sites (Farris, 2002; Simmons et al., 2002). Analyses of an expanded 11-gene data set (Zanis et al., 2002) showed high bootstrap and posterior probabilities for *Amborella* as sister to all other angiosperms, whereas *Amborella* + Nymphaeaceae as sister to all other angiosperms could be rejected in two of the three tests. The *matK* data favor the hypothesis that *Amborella* is sister to all other angiosperms.

Eumagnoliids—As recognized by APG II (2003), eumagnoliids include Canellales, Laurales, Magnoliales, and Piperales. Evidence from *matK* is in line with combined multigene analyses (Graham and Olmstead, 2000; Qiu et al., 2000; Zanis et al., 2002), the strict consensus trees based on analyses of *phyA* + *phyC* (Mathews and Donoghue, 2000), and *trnT-trnF* sequences (Borsch et al., 2003) in supporting this definition of eumagnoliids.

Savolainen et al. (2000a) and Soltis et al. (2000) recognized the magnoliids to include Chloranthaceae and monocots in addition to these four orders but with low support; we will refer to this clade here as eumagnoliids sensu lato (s.l.). Although the MP analysis of *matK* shows monocots + Chloranthaceae as an unresolved lineage with eumagnoliids and *Ceratophyllum* + eudicots (Fig. 4), the Bayesian analysis (Fig. 3) recovers eumagnoliids s.l. with low probability (0.73), encompassing monocots + Chloranthaceae (<0.5 PP) and eumagnoliids (1.0 PP). Eumagnoliids s.l. was also recovered with similar topology in a combined *matK* + *trnT-trnF* data analysis, but with <50% JK support (K. W. Hilu, K. Müller, and T. Borsch, unpublished data). Whereas stronger evidence points to the composition of eumagnoliids to include only the Laurales-Magnoliales-Canellales-Piperales, the hypothesis of an expanded eumagnoliids to include monocots and Chloranthaceae cannot be disregarded. Accepting the hypothesis of a eumagnoliid s.l. clade implies that carpel evolution in Chloranthaceae is secondarily ascidiate as pointed out by Doyle and Endress (2000) and Endress and Igersheim (2000).

Although MP support for the Magnoliales and Laurales sister group relationship was weak (58/56% JK), the Bayesian approach infers the same node but with 0.96 PP. This relationship has been inferred from a number of molecular data sets (Mathews and Donoghue, 1999, 2000; Qiu et al., 1999, 2000; Barkman et al., 2000; Graham and Olmstead, 2000; Graham et al., 2000; Zanis et al., 2002; Borsch et al., 2003). In addition, Sauquet et al. (2003) found particularly good support for this relationship from parsimony analysis of molecular and combined molecular and morphological data (decay index 14 and 17, BS 99%) based on unrooted trees in which all four orders were represented. However, support for this relationship from morphological data alone has been either lacking (Doyle and Endress, 2000) or weak (Sauquet et al., 2003). The internal structure of Magnoliales inferred from *matK* differs from the multigene-based results of Sauquet et al. (2003) in the arrangement of the five families above Myristicaceae; however, this inconsistency is not well supported (JK ≤ 51%) and may be explained by extremely short branches. The internal structure of Laurales based on *matK* is congruent with the minimum evolution tree in Renner and Chanderbali (2000), but not with the trees retrieved using MP in the same study or in Renner (1999).

The MP analysis shows a Piperales + Canellales relationship, but with low support. However, Bayesian analysis (Fig. 3) depicts Piperales as sister to Canellales (74% PP) and Laurales + Magnoliales. The sister group relationship of Piperales to Canellales is consistent with other molecular analyses (Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Qiu et al., 2000; Zanis et al., 2002; Borsch et al., 2003). The internal structure of Piperales in the *matK* tree shows Aristolochiaceae paraphyletic to Lactoridaceae and Saururaceae-Piperaceae in both parsimony (Fig. 4) and Bayesian approaches (tree not shown). A similar situation also occurred in analysis of *atpB* (Savolainen et al., 2000a). Phylogeny reconstruction in Piperales is complicated by the presence of several long branches. Results of higher sampling density using sequences of the complete *trnK* intron and other genomic regions are congruent with those of *matK* in showing a Piperaceae + Saururaceae clade sister to an Aristolochiaceae + Lactoridaceae clade (S. Wanke, University of Bonn, personal communication).

Monocots—The *matK* sequences provide good support (85% JK, Fig. 5) for the monophyly of monocots, with *Acorus* followed by Alismatales as sisters to other monocots. Individual gene analyses, such as *rbcL* and *atpB* (Chase et al., 1993, 1995; Savolainen et al., 2000a) and 18S (Soltis et al., 1997),

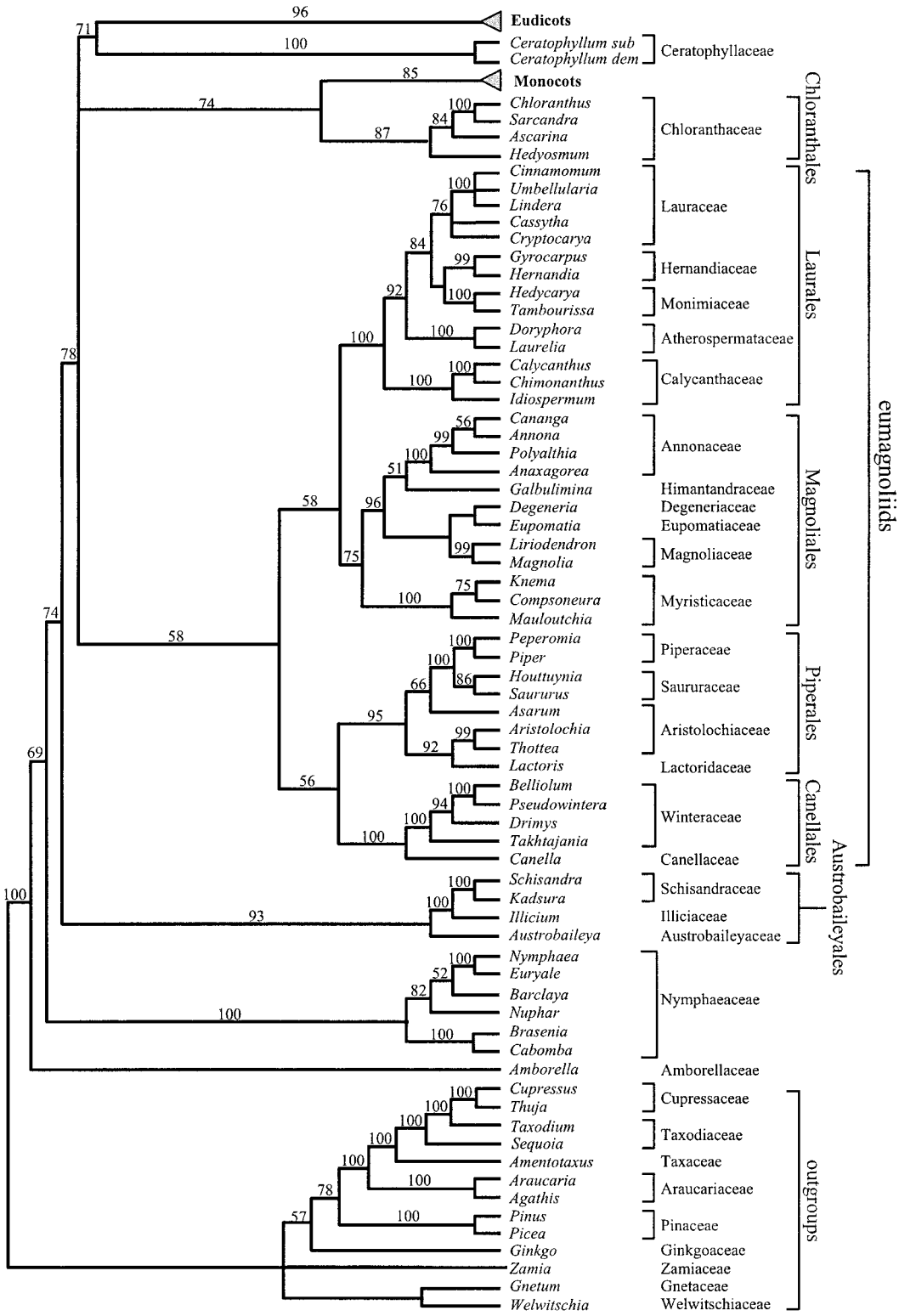

Fig. 4. Strict consensus tree highlighting relationships among basal angiosperm lineages. Numbers above branches are jackknife values derived from heuristic searches of matrix A. Large clades are indicated by triangles.

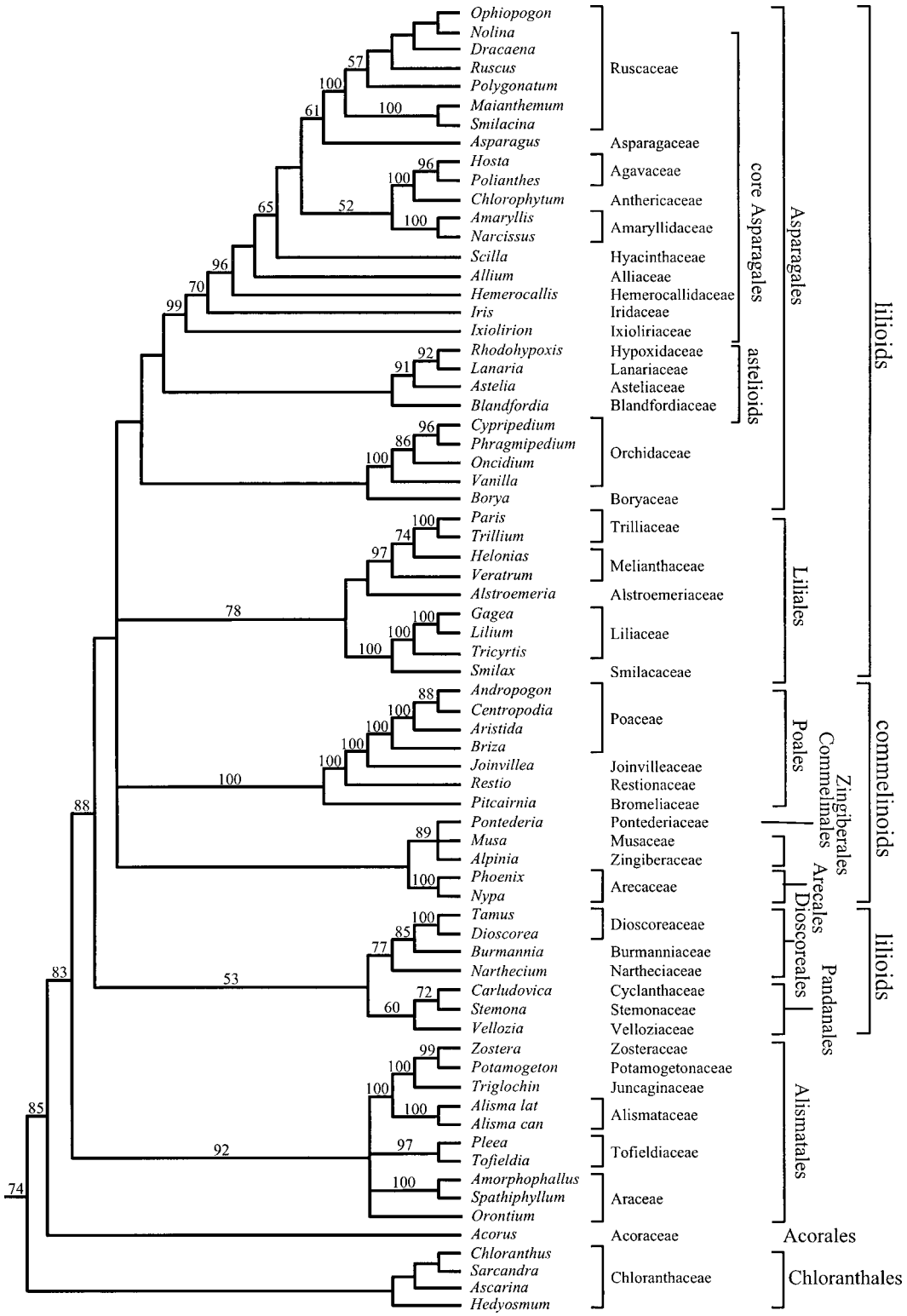

Fig. 5. Strict consensus tree highlighting relationships among monocots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

have indicated such a topology but with low or <50% support. Support for these relationships with *matK* is comparable to data sets that combined three or more genes (Table 2; Chase et al., 2000; Soltis et al., 2000; Zanis et al., 2002). The circumscription and internal structure of most monocot orders, particularly Liliales and Dioscoreales, revealed by *matK* is congruent with those suggested by the three-gene analysis, with similar or higher JK/PP support (Figs. 5–6).

TABLE 2. Comparative bootstrap/jackknife values for various nodes in angiosperm trees based on single and combined genes analyses. Large-scale analyses that covered most major angiosperm lineages are compared. Support value for the monophyly of a clade is reported when its name is cited alone, whereas support values for the others are given for a specified position in the tree. In a grade, the support by which a taxon is excluded from remaining clades (i.e., its sister group) is that at the node representing the common ancestor of the clade from which the taxon is excluded. For example, support for *Amborella* as the first clade in angiosperms refers to the node that separates *Amborella* from remaining angiosperms. ''nr'' (not resolved) denotes unresolved node, whereas ''—'' refers to taxa/clades that were not sampled. Data for *rbcL*, *atpB*, and *rbcL*/*atpB* were obtained from Savolainen et al. (2000a), 18S rDNA from Soltis et al. (1997), *rbcL*/*aptB*/18S from Soltis et al. (2000), and *rbcL*/*atpB*/18S/26S rDNA from Soltis et al. (2003). Savolainen et al. (2000a) used bootstrap, whereas remaining studies employed jackknife. For comparability see Materials and Methods.

Node	*rbcL*	*atpB*	18S	*rbcL*/*atpB*	*rbcL*/*aptB*/18S	*rbcL*/*atpB*/18S/26S	*matK*
Angiosperms	53	60	100	100	100	—	100
Amborella first clade in angiosperms	nr	<50	nr	nr	65	—	69
Nymphaeaceae second clade in angiosperms	nr	<50	nr	nr	72	—	74
Austrobaileyales third clade in angiosperms	nr	<50	nr	nr	71	—	78
Eumagnoliids s.str.	nr	<50	nr	<50	nr	—	58
Piperales sister to Canellales	nr	<50	nr	nr	nr	—	56
Monocots	59	<50	<50	86	95	—	85
Eudicots	72	<50	<50	89	99	100	96
Ranunculales first clade in eudicots	<50	<50	nr	67	59	87	82
Ranunculales	51	<50	<50	94	98	100	90
Sabiales second clade in eudicots	nr	nr	nr	nr	100	nr	<50
Proteales	<50	<50	nr	60	84	73	64
Core eudicots	<50	<50	<50	91	100	100	99
Gunnerales first clade in core eudicots	<50	nr	nr	<50	nr	84	<50
Gunnerales	57	nr	—	80	75	85	100
Rosids	<50	<50	nr	61	99	79	95
Eurosids I	<50	<50	nr	<50	77	<50	52
Eurosids II	<50	<50	nr	<50	95	88	nr
Asterids	<50	66	nr	92	99	99	98
Euasterids I	<50	<50	nr	<50	56	58	<50
Euasterids II	<50	<50	nr	51	88	87	91
Caryophyllales	84	74	<50	97	100	100	99
Saxifragales	<50	<50	68	<50	98	100	62
Santalales	<50	<50	66	86	100	100	99
Dilleniaceae/Vitaceae	nr	nr	nr	nr	nr	nr	55
Berberidopsidales	<50	95	—	97	100	100	100
Cornales	52	74	nr	96	98	100	99
Ericales	<50	<50	nr	97	98	100	98
Supported nodes (>50)	24%	14%	7%	55%	83%	82%	83%

Acorus has been inferred as sister to *Ceratophyllum* (Savolainen et al., 2000a, b, with *atpB* alone), sister to remaining monocots (Soltis et al., 2000), or placed within Alismatales (Qiu et al., 2000); monocot sampling was sparse in both studies and support in both cases was weak or lacking. Compared to three-gene analyses (Chase et al., 2000; Soltis et al., 2000), *matK* provides better support for the Alismatales as sister to the commelinoid/lilioid clades (88% vs. 78% JK) and for the monophyly of the Alismatales (92% vs. 75% JK).

Following Alismatales, the *matK* consensus tree depicts (53% JK; Fig. 5) Dioscoreales and Pandanales (lilioids) in a clade sister to remaining monocots that appear in a polytomy of two commelinoid lineages (Arecales + Zingiberales and Poales) and the two remaining lilioid clades (Liliales and Asparagales). Clarification of the phylogenetic relationships among the lilioid orders Asparagales, Dioscoreales, Liliales, and Pandanales, as well as within the commelinoids (Poales, Commelinales, Zingiberales, and Arecales), has been a major problem because previous molecular studies, including the detailed three-gene analysis of monocots (Chase et al., 2000), did not yield topologies with high internal support. The emergence in the MP analysis of a Dioscoreales + Pandanales clade (53% JK; Fig. 5) as sister to the remaining lilioid and commelinoid lineages is supported by 1.0 PP in the BI analysis (Fig. 6). Therefore, *matK* data also indicate paraphyly of the lilioids to the commelinoids, a hypothesis that requires further evaluation. Significantly, Liliales and most Asparagales uniquely share a particular type of epicuticular wax (parallel platelets; Barthlott et al., 2003). The commelinoids do not form a clade in the MP analysis of the *matK* data. In contrast, this lineage was resolved with 1.0 PP in the Bayesian tree (Fig. 6), in line with the three-gene analyses of Chase et al. (2000) and Soltis et al. (2000) that resolved this group with weak support. The sister group relationship between Commelinales and Zingiberales as inferred earlier by the three-gene studies with moderate support is well supported with *matK* sequences (89% JK, 1.0 PP). Moreover, *matK* places Arecales as sister group to the Commelinales-Zingiberales clade (Figs. 5, 6; <50% MP, 0.88 BI).

Chase et al. (1995, 2000) defined Asparagales in a broad sense to include the astelioids, Boryaceae and Orchidaceae. However, the combined *rbcL*, *atpB*, and 18S rDNA analyses of Soltis et al. (2000) and Chase et al. (2000) only recovered the core Asparagales (sensu Chase et al., 1995) as a monophyletic group with high support. Hypoxidaceae, Asteliaceae, Boryaceae, Blandfordiaceae, and Orchidaceae appeared in their strict consensus tree as unresolved at the base of the Asparagales s.l. clade. The *matK* study recovers core Asparagales (Ixoliriaceae to Convallariaceae, Figs. 5, 6) with better support (99% JK, 1.0 PP). At the base of Asparagales, two

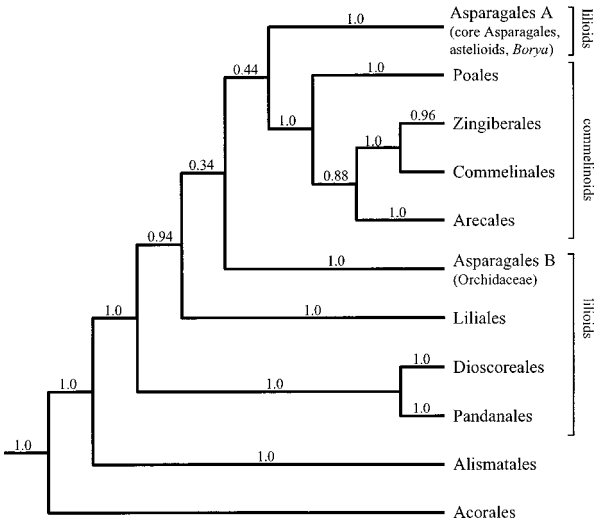

Fig. 6. Summary of monocot relationships inferred from Bayesian analysis. Numbers above branches are posterior probabilities.

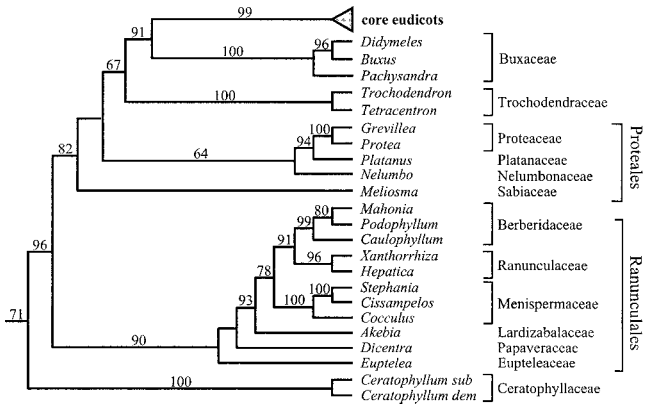

Fig. 7. Strict consensus tree highlighting relationships among basal eudicots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

clades, Boryaceae + Orchidaceae and Blandfordiaceae + Asteliaceae + Lanariaceae + Hypoxidaceae (the latter group was termed the astelioid clade by Fay et al., 2000; Fig. 5) were found successive sisters to the rest in MP but with <50% JK support (annotated as lower Asparagales in Fig. 5). This topology is identical to the one inferred from the second search of the three-gene data set by Chase et al. (2000), although none of the relevant nodes received good support. The *matK* places Boryaceae as sister to Orchidaceae in the MP tree with <50% JK and shows it nested within the astelioid clade in the Bayesian tree with high PP (the sucessive branching order is *Blandfordia*, *Borya*, *Astelia*, and *Rhodohypoxis* plus *Lanaria*; all nodes receive 1.0 PP; details are not shown in Fig. 6). The orchids appear as sister to the commelinoids in the Bayesian tree (Fig. 6). A recent combined *rbcL* + *atpB* + *trnL-F* parsimony analysis of Asparagales (Fay et al., 2000) showed Boryaceae sister to the astelioid clade, which are in turn sister to the core Asparagales. Orchidaceae appear basal in the latter analysis, but none of these basal nodes within lower Asparagales are well supported. As a result, the order of the first branches within the Asparagales and their broader circumscription to include the orchids are still in need of further testing.

Eudicots—The eudicots include over 200 000 species and comprise about 75% of all angiosperm species (Drinnan et al., 1994). This clade has consistently been recovered in all recent single-gene and multiple-gene phylogenetic analyses of angiosperms (e.g., Chase et al., 1993; Hoot et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Qiu et al., 2000; Savolainen et al., 2000a, b; Borsch et al., 2003; Soltis et al., 2003) and in nonmolecular phylogenetic analyses (Drinnan et al., 1994; Nandi et al., 1998). The strong support (100%) for eudicot monophyly achieved by combining genes from several genomes is corroborated by *matK* (Figs. 2, 3). The eudicot clade encompasses a basal grade and a strongly supported core clade that includes the large rosid and asterid clades and the smaller Gunnerales, Caryophyllales, Santalales, Berberidopsidaceae/Aextoxicaceae, and Saxifragales clades. Eudicots have tricolpate or tricolpate-derived pollen (e.g., tricolporate, pantoporate; e.g., Nandi et al., 1998). Morphologically similar pollen having three apertures in Austrobaileyales have been identified as trisyncolpate (Takahashi, 1994), but these are apparently convergent.

Early-diverging eudicots—The same lineages of early-diverging eudicots (Ranunculales, Proteales, Sabiaceae, Trochodendraceae including Tetracentron, and Buxaceae including Didymelaceae) have been consistently recovered (e.g., Chase et al., 1993; Soltis et al., 1997, 2000, 2003; Hoot et al., 1999; Savolainen et al., 2000a), although relationships among them are unclear. In addition, support for the nodes is low in several cases (Table 2). Both MP and BI analyses of our *matK* data yield identical topologies with good support for a grade of Ranunculales/Sabiaceae/Proteales/Trochodendraceae (Fig. 7). The *matK* tree is for the most part in agreement with those obtained from the combined three- and four-gene analyses (Soltis et al., 2000, 2003). However, *matK* data provide considerably higher support for relationships in this group than in previous studies (Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000), particularly for the position of Ranunculales as sister to all other eudicots (82/76% JK, 1.0 PP). Within Ranunculales, the exact position of Eupteleaceae and Papaveraceae at the base is method dependent in *matK*. *Euptelea* is sister to all other Ranuncuales though without support in MP, whereas Papaveraceae were basal in MB, excluded with 75% PP from the rest of Ranunculales. Family relationships among remaining Ranunculales are identical in both approaches (Bayesian not shown) and in agreement with the results from the multigene analyses of Hoot et al. (1999) and Soltis et al. (2000, 2003).

The divergence of Sabiaceae after Ranunculales gains <50% JK support but 1.0 PP. The position of Sabiaceae and its inferred affinity to Proteales has varied (Hoot et al., 1999; Qiu et al., 2000; Savolainen et al., 2000a; Soltis et al., 2000, 2003), but with weak or no support. Although *matK* presents a potential position for Sabiaceae branching after Ranunculales, the hypothesis of a Sabiaceae + Proteales clade (see Soltis et al., 2000) cannot be excluded.

In line with previous studies (Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000), the *matK* tree places *Nelumbo* as the sister to the rest of Proteales. Also evident is the high support for a sister group relationship of Proteaceae and Platanaceae (94% JK, 1.0 PP). Several anatomical and morpho-

logical synapomorphies have been identified for Proteales (see Nandi et al., 1998; Savolainen et al., 2000a).

Buxaceae and Trochodendraceae exchanged positions in the basal eudicot grade in previous studies; these nodes always received low support (Table 2). This *matK* study shows for the first time good support for a Trochodendraceae/Buxaceae/core eudicot including Gunnerales (Figs. 2, 3, 7) successive branching. The support is particularly strong for Buxaceae being sister to core eudicots (91% JK, 1.0 PP).

Core eudicots—Core eudicots, composed of Gunnerales, Caryophyllales, Berberidopsidales, Saxifragales, Santalales, Vitaceae, Dilleniaceae, asterids, and rosids, received 100% JK and 1.0 PP support with *matK*. The core eudicots were recovered with high support in the combined *rbcL*/*atpB* analysis, and 100% support was achieved with the addition of 18S and 26S rDNA sequences (Hoot et al., 1999; Soltis et al., 2000, 2003). Thus, the core eudicots stand now as one of the best-supported major clades of angiosperms (Figs. 2, 3). Soltis et al. (2003) discussed the implications for character evolution of Gunnerales sister to the rest of the core eudicots.

The *matK* data provide more structure for the major branches within core eudicots than previous single or combined gene analyses (Figs. 8, 9; Table 2). A crown group comprising Berberidopsidales/Santalales/Caryophyllales/asterids is inferred here (65% JK and 1.0 PP; Figs. 2, 3, 8). The sister group relationship of asterids and Caryophyllales is weakly supported (62% JK, 0.79 PP). The study based on 18S rDNA sequences (Soltis et al., 1997) showed Caryophyllales nested within asterids (BS $<$ 50%), but combined analysis of 18S rDNA and *rbcL* sequences placed the order within rosids (Soltis and Soltis, 1997); sparse sampling in the latter study might have affected Caryophyllales placement. All previous analyses have indicated affinities of Caryophyllales to Dilleniaceae (see discussion under Dilleniaceae/Vitaceae), but never with high support. Santalales and Berberidopsidaceae/Aextoxicaceae change positions depending on method of analysis. This unstable position is reflected by the low posterior probability (0.55) for the node uniting Berberidopsidaceae/Aextoxicaceae and asterids/Caryophyllales.

Gunnerales—Support for the monophyly of Gunnerales (i.e., the sister group relationship of *Gunnera* and *Myrothamnus*) with *matK* is also high compared to all previous analyses (1.0 PP and 100% JK; Figs. 2, 3, 8), exceeding the 85% JK support achieved by the four-gene analysis (Soltis et al., 2003). The *matK* analysis provides strong support (99/100% JK and 1.0 PP) for the inclusion of Gunnerales in core eudicots (Figs. 2, 8). However, Gunnerales being sister to the remaining core eudicots receives $<$50% JK support in MP, yet is highly probable (1.0 PP) in the Bayesian analysis. This position for Gunnerales in the core eudicot clade is in agreement with studies based on individual and combined sequence data analyses (Chase et al., 1993; Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000). The solidified relationship of the *Gunnera* + *Myrothamnus* clade is important for future analyses of the remarkable evolutionary diversification of these two genera. *Myrothamnus* comprises two species of small, poikilohydric shrubs on African inselbergs, whereas *Gunnera* (40 species) includes hemicryptophytes, living in symbiosis with the cyanobacterium *Nostoc* for nitrogen fixation and inhabiting more or less wet habitats, mostly in the Southern Hemisphere.

Saxifragales—In Saxifragales, four well-supported clades of interest are resolved with *matK*. One (97%) includes Iteaceae as sister to Grossulariaceae/Saxifragaceae (95%). These three families were part of a large polytomy in trees based on individual and combined analyses of *rbcL, atpB,* and 18S rDNA (Savolainen et al., 2000a; Soltis et al., 2000). The inclusion of the 26S rDNA sequence data improved resolution for Saxifragales with the sister group relationship of Iteaceae (plus Pterostemonaceae) and Saxifragaceae receiving 100% JK. A clade with similar topology and support was recovered in a combined analysis of data from five nuclear and plastid genes (Fishbein et al., 2001). A second clade resolved here includes Cercidiphyllaceae and Daphniphyllaceae (78% JK), which were unresolved in previous large-scale analyses. A third clade includes Haloragaceae, Tetracarpaeaceae, and *Aphanopetalum* (formerly included in Cunoniaceae; 95%). A similar clade was recovered in the five-gene study of Fishbein et al. (2001). A fourth, weakly supported clade (61% JK) encompasses Altingiaceae (*Rhodoleia* and *Altingia*). *Rhodoleia* has been placed in Rhodeliaceae, Hamamelidaceae, or Altingiaceae.

Vitaceae/Dilleniaceae—Dilleniaceae is sister here to Vitaceae, although support for this relationship varies with the method applied (weak using MP, but 1.0 PP in BI). The phylogenetic positions of these families among core eudicots have been difficult to assess and differ in each analysis. Floral and endosperm characters were proposed as potential synapomorphies for the relationship of Dilleniaceae to Caryophyllales and Vitaceae to rosids (Savolainen et al., 2000a). In contrast, several features are shared by Dilleniaceae and Vitaceae, such as calcium oxalate raphides (Metcalfe and Chalk, 1950), an endotesta containing radially elongate cells and a tracheidal endotegmen (Corner, 1976). Similarities of Dilleniaceae to Vitaceae have also been pointed out by Nandi et al. (1998).

Berberidopsidales—The sister group relationship of Berberidopsidaceae and Aextoxicaceae is another case in which *matK* provides the greatest support (100% JK; Fig. 8) among single-gene analyses. The results are in close agreement with trees inferred from multigene data sets (e.g., Savolainen et al., 2000a; Soltis et al., 2000, 2003).

Santalales—The strong support (99% JK; Fig. 8) for the monophyly of Santalales and internal relationships are in agreement with previous molecular studies. This analysis provides the highest support (66/71% JK and 1.0 PP) for the sister group relationship of Santalales to Caryophyllales + asterids. The position of Santalales has varied considerably in previous molecular studies. Soltis et al. (2000) noted that Santalales and Caryophyllales may be related and in turn be sister to the asterids; both relationships receive weak support in this study.

Caryophyllales—The *matK* sequence data strongly support the broadly circumscribed Caryophyllales (99% JK, 1.0 PP) as first indicated by analyses of *rbcL* alone (Savolainen et al., 2000b) and in later studies (Savolainen et al., 2000a; Soltis et al., 2000, 2003; Cuénoud et al., 2002). Two Caryophyllales clades, here named Caryophyllales I and II, are recognized with *matK* (Fig. 8).

In Caryophyllales I, *Rhabdodendron* (Rhabdodendraceae) and *Simmondsia* (Simmondsiaceae) emerge in a polytomy with a clade containing the rest of Caryophyllales I (Fig. 8), which is in agreement with Cuénoud et al. (2002). Core Caryophyl-

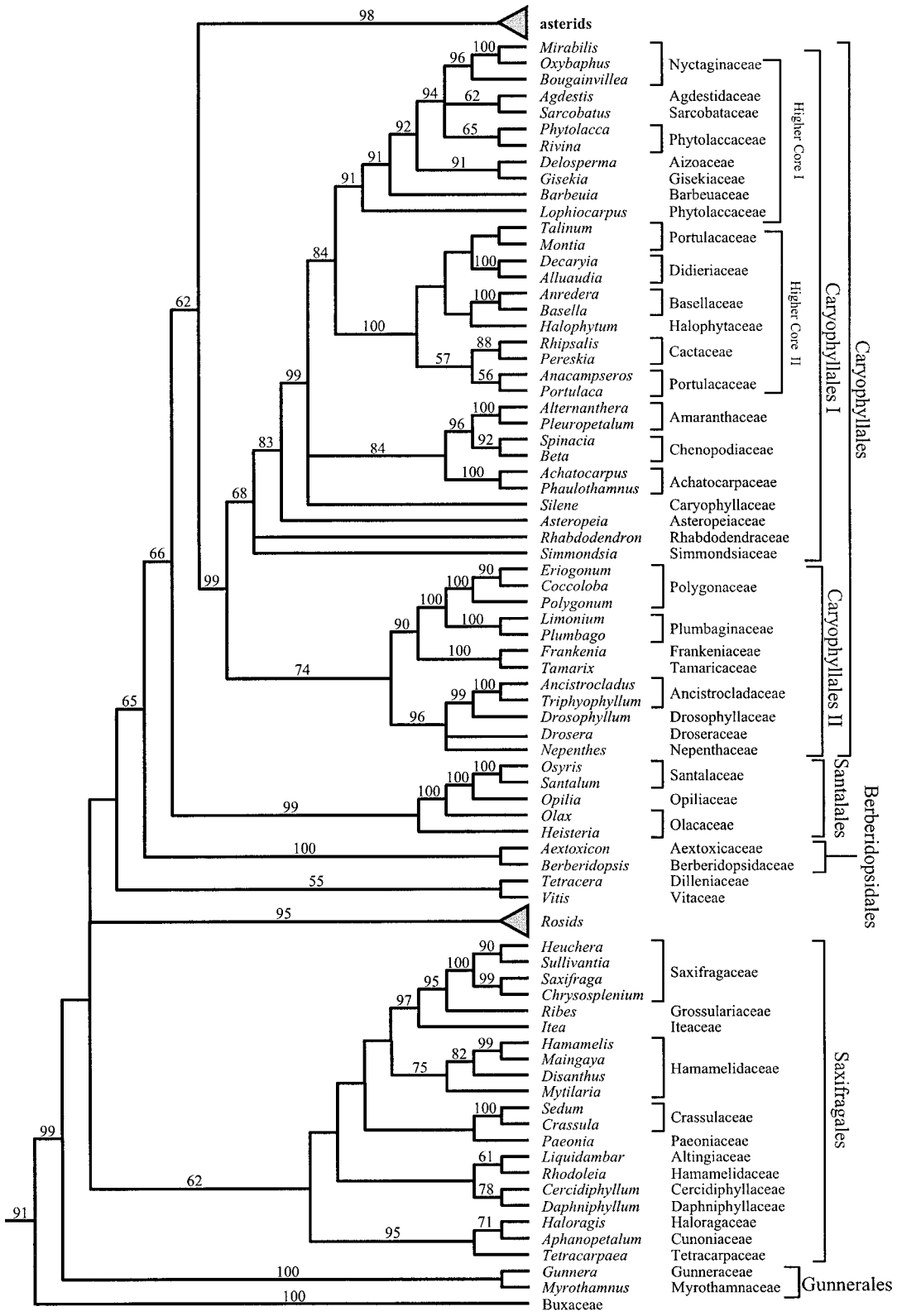

Fig. 8. Strict consensus tree highlighting relationships among core eudicots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

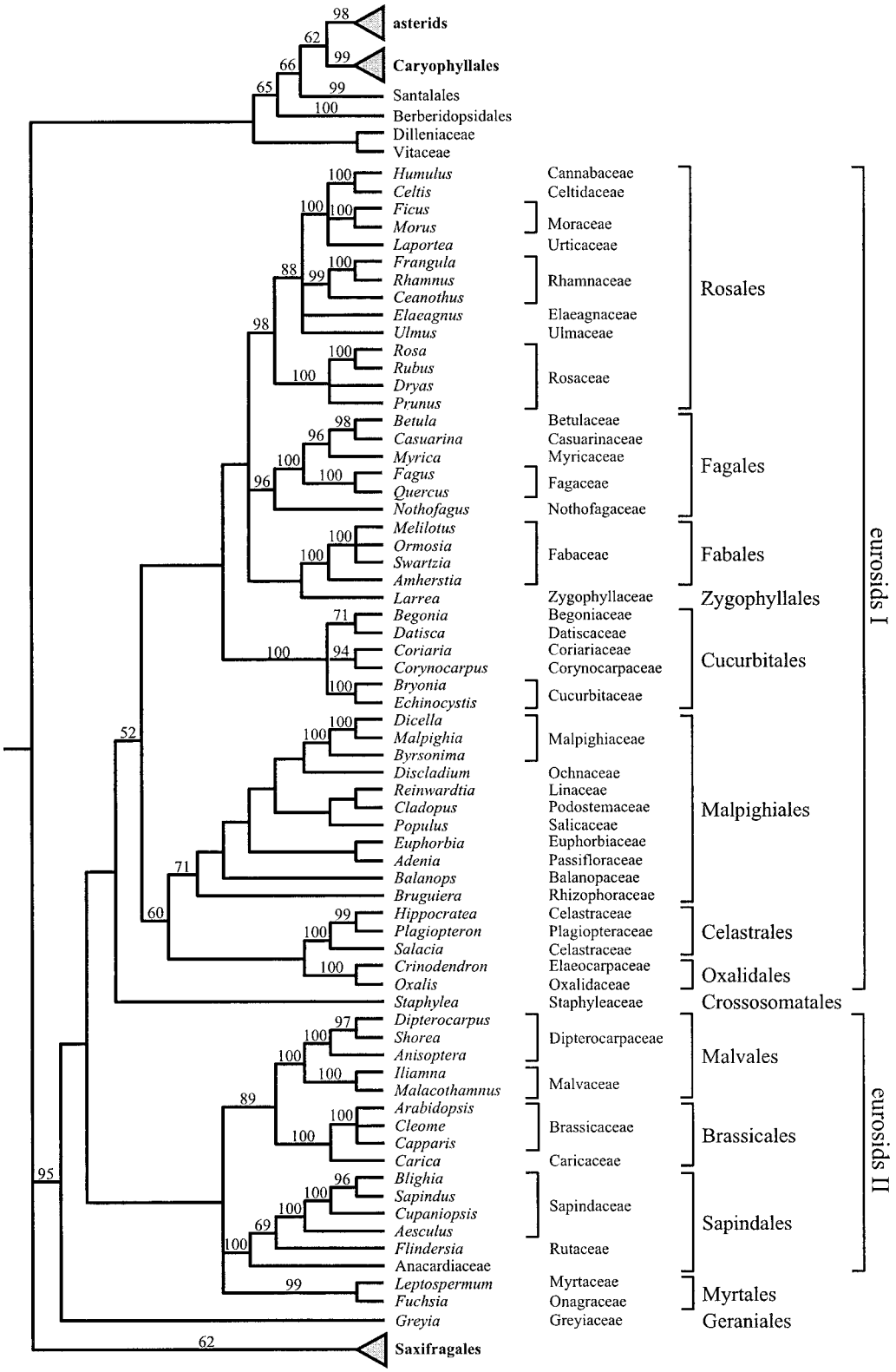

Fig. 9. Strict consensus tree highlighting relationships among rosids. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

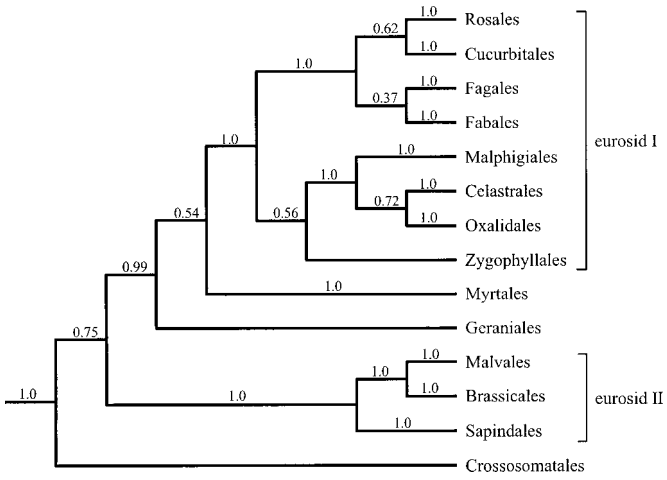

Fig. 10. Summary of relationships among rosids based on Bayesian inference. Numbers above branches are posterior probabilities.

lales (83% BS, 1.0 PP) sensu Cuénoud et al. (2002) includes *Asteropeia* (Asteropeiaceae) as sister to a polytomy comprising Caryophyllaceae, an Amaranthaceae-Chenopodiaceae-Achatocarpaceae clade that was also resolved in Kadereit et al. (in press), and a higher core Caryophyllales clade (Fig. 8). The latter group in turn consists of two major clades, higher core Caryophyllales I and II. Caryophyllaceae appears unresolved in MP, but emerges sister to the rest in BI (0.59 PP), a position also reflected by the *matK*-MP analysis of Cuénoud et al. (2002) and *rbcL* + *atpB* analysis of Savolainen et al. (2000a). Relationships revealed here for Caryophyllales are in general agreement with Cuénoud et al. (2002).

In Caryophyllales II, two subclades are recovered. One encompasses the carnivorous families Ancistrocladaceae, Drosophyllaceae, Droseraceae, and Nepenthaceae in a subclade (96% JK, 1.0 PP) and the Polygonaceae (including Coccolobaceae), Plumbaginaceae, Frankeniaceae, and Tamaricaceae in another (90% JK, 1.0 PP).

Rosids—The rosid clade (not including Vitaceae) corresponds to ''eurosids'' of Soltis et al. (2000). The strong support for rosids with *matK* (95% JK and 1.0 PP) is only comparable to that obtained from the combined *atpB, rbcL,* and 18S rDNA sequence data (99% JK; Soltis et al., 2000). However, the addition of 26S rDNA data to the three-gene data set (Soltis et al., 2003) decreased the JK value for rosids to 79%. Among single data sets, only *rbcL* (Chase et al., 1993; Savolainen et al., 2000a, b) and 18S rDNA (Soltis et al., 1997) demonstrated a rosid clade of similar circumscription but with <50% support. In this circumscription, the rosids include eurosids I and II plus Myrtales, Crossosomatales, and Geraniales (Fig. 9).

Relationships among eurosid I, eurosid II, Crossosomatales, Geraniales, and Myrtales varied among all previous single gene (Chase et al., 1993; Soltis et al., 1997; Savolainen et al., 2000a, b) and multigene (Soltis et al., 2000, 2003) analyses. In this study, topological differences are evident between the MP and Bayesian trees (Figs. 9, 10), reflecting low levels of internal support. The Geraniales/Myrtales plus eurosid II/Crossosomatales/eurosid I grade found in the MP analysis is contrasted with the BI tree that resolves Crossosomatales as sister to the remaining rosids (0.75 PP), followed by eurosid II (excluded from the rest by 0.99 PP), Geraniales, with Myrtales and eurosid I forming the terminal clade (0.54 PP). Low support for such relationships was also evident in previous molecular studies. In contrast to the ambiguous relationships among the major rosid lineages, the rosid orders are generally well supported by *matK* data. All orders receive 1.0 PP (Fig. 10) and 96–100% JK except for Malpighiales (71%). For most orders, this level of confidence was achieved only by two or more genes (Savolainen et al., 2000a; Soltis et al., 2000, 2003).

Eurosid I—Weak support (52%) is obtained for this clade in the MP analysis. In contrast, support in BI is high (1.0 PP). Previous molecular studies mostly yielded <50% support for the eurosid I clade except for the combined three-gene analysis (77% JK; Soltis et al., 2000). However, support declined to <50% when 26S rDNA sequences were added. Within eurosids I, *matK* reveals two major clades. The first comprises Celastrales, Oxalidales, and Malpighiales, which are weakly supported by MP (60%, Fig. 9) but with 1.0 PP. The other eurosid I clade includes Rosales, Fagales, Fabales, and Cucurbitales (the nitrogen-fixing clade; Soltis et al., 1995) as well as Zygophyllales. This clade recently received some support in the three-gene analysis (68% JK; Soltis et al., 2000). The Bayesian tree shows 1.0 PP support for this clade (Fig. 10). The MP topology differs from that of the corresponding Bayesian tree by depicting Zygophyllales as sister to Fabales, albeit with <50% JK support. Anthroquinones are a potential synapomorphy for Zygophyllaceae and the nitrogen-fixing clade (Sheahan and Chase, 2000). Because none of the previous topologies demonstrated good internal support, the position of Zygophyllales remains questionable.

Malpighiales is currently recognized to include at least 30 families, among which relationships have been difficult to establish. The two analyses of *matK* (Figs. 9, 10) agree upon the arrangement at the basal nodes, with Rhizophoraceae and Balanopaceae as successive sisters to the rest (0.60 and 0.87 PP, respectively; <50% JK). In the most extensive sampling of the order thus far (Savolainen et al., 2000b; *rbcL* only), Rhizophoraceae were sister to Erythroxylaceae (not included here) and the pair sister to the rest. No resolution, however, was provided for such relationships with three genes (Soltis et al., 2000). For other clades in Malpighiales, support is only achieved for Passifloraceae + Salicaceae in the BI tree (1.0 PP).

Eurosid II—A eurosid II clade is not recovered in the *matK* MP tree, but instead its components appear in two well-supported clades that form a polytomy with Myrtales (Figs. 9, 10). The three- and four-gene analyses (Soltis et al., 2000, 2003) recovered eurosids II, but relationships among its members lacked internal support. The eurosid II taxa resolved with *matK* appear in two subclades, Sapindales (100% JK) and Brassicales + Malvales (89% JK). In contrast, the Bayesian analysis inferred the three orders in a clade with 1.0 PP support (Fig. 1). The *matK* data provide high internal support for a sister group relationship of Brassicales and Malvales (89% JK, 1.0 PP); such a relationship was obtained in the *atpB* and *rbcL*/*atpB* analyses, although support in those cases was <50% and 62%, respectively (Savolainen et al., 2000a). In contrast, the three- and four-gene analyses (Soltis et al., 2000, 2003) weakly or moderately depicted Sapindales and Malvales

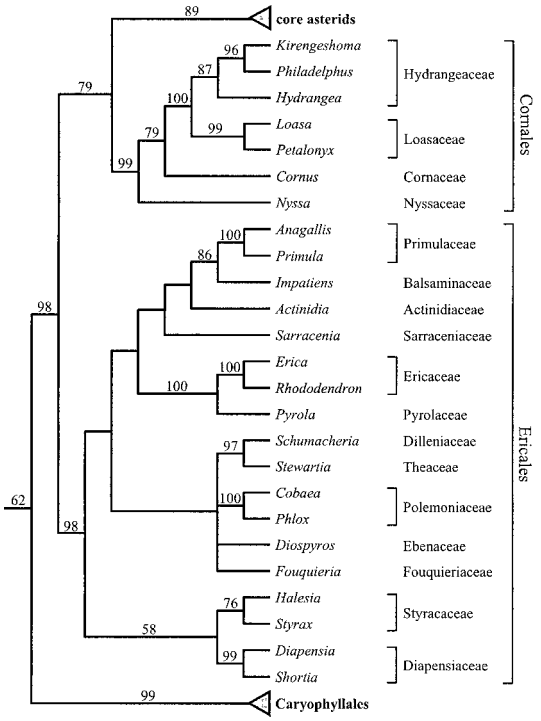

Fig. 11. Strict consensus tree highlighting relationships within Cornales and Ericales of the asterid clade. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

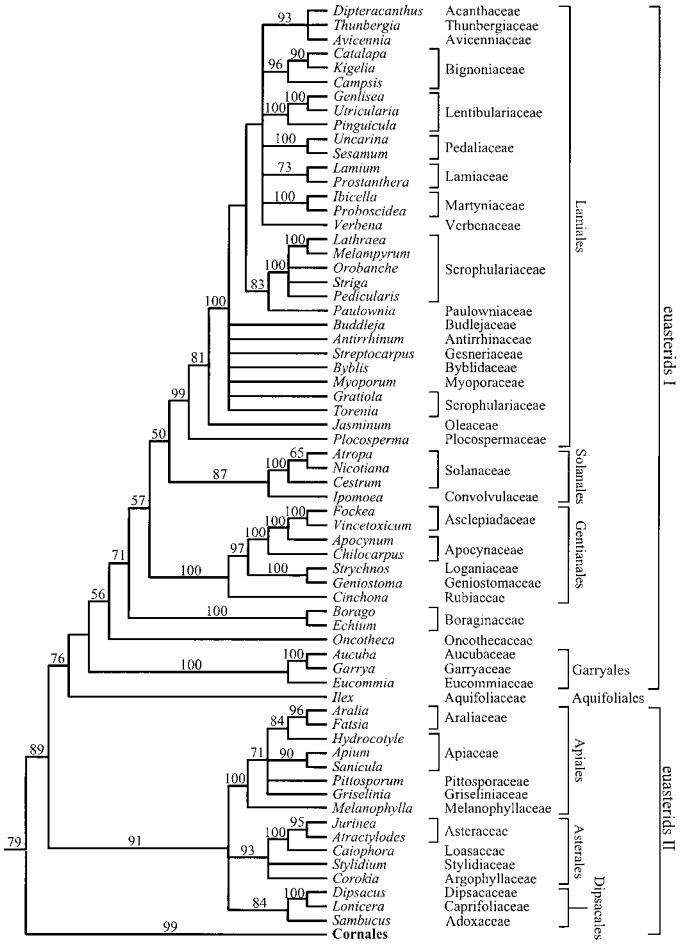

Fig. 12. Strict consensus tree depicting relationships among higher asterids. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

in a clade. No clear morphological synapomorphies for this clade are evident (Judd et al., 1994; Gadek et al., 1996).

Asterids—This study reveals a strongly supported (98% JK, 1.0 PP) asterid clade sensu Olmstead et al. (1992, 1993) that includes four major lineages: Cornales, Ericales, and euasterid I and II (Figs. 2, 11). Bremer et al. (2002) referred to euasterid I and II as Lamids and Campanulids, respectively. Internal support within asterids using *matK* is surpassed only by the analysis of six-genomic region by Bremer et al. (2002) and is similar to what was obtained in the combined three- and four-gene analyses (Soltis et al., 2000, 2003; Albach et al., 2001). Our *matK* data alone show Ericales as sister to remaining asterids (79% JK; Fig. 11). In contrast, most other studies depict Cornales (Albach et al., 2001, *ndhF* tree; Bremer et al., 2002; Soltis et al., 2003) or Cornales + Ericales (Savolainen et al., 2000a; Soltis et al., 2000) in this position.

Ericales/Cornales—The monophyly of Ericales and of Cornales each receives strong support in the *matK* tree (98–99% JK and 1.0 PP; Fig. 11). The backbone of Ericales remains unresolved in this and all other molecular studies including Bremer et al. (2002) and Anderberg et al. (2002). In Cornales, Nyssaceae and Cornaceae are successive sisters to other members of the order, which is in agreement with Xiang et al. (1998), Soltis et al. (2000), and Albach et al. (2001).

Euasterid I (Lamids)—The euasterid I clade (consisting of Garryales, Oncothecaceae, Boraginales, Gentianales, Solanales, and Lamiales including Plocospermataceae) has a 1.0 PP but <50% JK support with MP (Fig. 12). Lack of support in MP is in line with other studies, including multigene data sets (e.g., Soltis et al., 2000, 2003; Albach et al., 2001; Bremer et al., 2002). The *matK* data provide moderate support for a sister position of Aquifoliales to euasterids I in both MP and BI. The sister group relationship of Garryales and Oncothecaceae to remaining euasterid I in the *matK* tree (Fig. 12) is in agreement with Bremer et al. (2002); however, BI shows 1.0 PP support compared to the low MP support (56–>50% JK) in both studies. Support for relationships among the remaining orders of euasterid I (Boraginaceae, Gentianales, Lamiales, and Solanales) is weak, a situation that has been encountered in all previous studies (Olmstead et al., 1992, 1993, 2000; Chase et al., 1993; Cosner et al., 1994; Savolainen et al., 2000a; Soltis et al., 2000, 2003; Albach et al., 2001; Bremer et al., 2002).

Gentianales are well resolved and strongly supported as monophyletic (mostly 100% JK, 1.0 PP) with Rubiaceae being sister to the rest. Support within Gentianales has been either very weak or moderate, except for the *ndhF* study of Olmstead et al. (2000) and the six-gene-region study of Bremer et al. (2002). A sister-group relationship between Lamiales and Solanales is inferred by *matK* (Fig. 12), receiving 0.8 PP (tree not shown), but only 50% JK.

Monophyly of Lamiales is supported by 99% JK, and the first branching position of Plocospermataceae in the Lamiales

receives strong support here (81% JK, 1.0 PP) and in Bremer et al. (2002). This position for Plocospermataceae was suggested by a broad sampling of *rbcL* (Savolainen et al., 2000a), but with only 56% BS. Plocospermataceae is a small family (one genus, three species) from Central America. The sister group relationship of *Paulownia* to a clade comprising parasitic and hemi-parasitic tribes of the former Scrophulariaceae, as well as the holoparasitic former family Orobanchaceae sensu stricto (s.s.), is congruent with the topology found by Olmstead et al. (2001) using *ndhF, rbcL,* and *rps2.* The Bayesian approach provides more resolution in Lamiales (see Müller et al., in press).

Euasterid II (Campanulids)—Within euasterids II, the Apiales, Asterales, and Dipsacales form a strongly supported clade (91% JK, 1.0 PP; Fig. 12) that basically corresponds to euasterids II excluding Aquifoliales. This alliance of Aquifoliales with euasterids II based on *matK* data is not well supported (76% JK, 0.81 PP; Fig. 12). Aquifoliales appeared as the first branching lineage in euasterids II in most previous studies with highest support achieved in combined analyses of 3–6 genomic regions (Soltis et al., 2000, 2003; Bremer et al., 2002). The relationships among the three euasterids II orders remain unclear.

PROSPECTS OF USING *matK* IN ANGIOSPERM PHYLOGENETICS

Sequence information from the *matK* gene produce an angiosperm tree that is considerably more robust than any previous single gene tree. Congruence is high between our *matK* tree and those based on multiple genes representing one, two, or all three genomes (Qiu et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003; Zanis et al., 2002). The analyses of Qiu et al. (1999) and Zanis et al. (2002) were based on 8733 (five genes) and over 15 000 (11 genes) nucleotides, respectively, and thus represent approximately eight and 13 times the number of characters used here. Congruence between our *matK* phylogenies and the various multigene/multigenome phylogenies of angiosperms underscores the utility of *matK* in angiosperm phylogenetics.

When clades from the backbone phylogeny of angiosperms are compared in various molecular phylogenetic studies (Table 2), 83% received jackknife support >50% with *matK* compared with 7–24% for individual analyses of *rbcL, atpB,* and 18S rDNA. Relationships revealed by *matK* data are more robust than those derived from combining *rbcL* and *atpB* sequences (Savolainen et al., 2000a). The number of nodes receiving >50% support with *matK* is in the same range as the combined analyses of 3–4 genes from two genomic compartments (also see Table 2). Examples where *matK* stands out in terms of support are the backbone of the angiosperms, basal eudicots, core eudicots, asterids, eurosid II, and Cornales.

The topology of the angiosperm tree was not influenced by the exclusion of the Gnetales from the outgroup taxa (matrix B), but JK support generally increased at various nodes, implying a higher level of homoplasy introduced by Gnetales relative to other outgroup taxa. In contrast, the Bayesian approach, although it yields results largely congruent with the most parsimonious trees, provides alternative hypotheses for some relationships. However, alternative topologies were confined to areas of the tree at which internal support (JK or BS) has always been low.

Patterns of molecular evolution in *matK* that make it notable among other genes used in studying plant phylogenetics are quantity of information (number of parsimony-informative sites/rate of change at variable positions) and quality of characters (signal vs. noise). The *matK* gene differs from coding genes used in phylogenetic reconstruction in the nearly equitable rates of nucleotide substitution among its three codon positions and the high relative rate of nonsynonymous substitution. This evolutionary mode was previously demonstrated in smaller-scale analyses (Olmstead and Palmer, 1994; Johnson and Soltis, 1995; Hilu and Liang, 1997; Soltis and Soltis, 1998; Cuénoud et al., 2002). Such a pattern would imply relatively relaxed selection on amino acid composition in relation to function as determined by physiochemical and structural properties. In an initial analysis comparing different plastid regions in basal angiosperms, K. W. Hilu et al. (unpublished manuscript) demonstrated that purifying selection was determined to be less significant in *matK* than in other protein coding genes, whereas phylogenetic signal at informative positions was found to be highest.

Although progress has been achieved in understanding angiosperm relationships in this study, several parts of the tree remain unresolved or unsupported. Outstanding among these are the positions of monocots, Chloranthaceae, eudicots, and *Ceratophyllum* among basal angiosperms. Within eudicots, relationships among the major lineages of the core eudicots remain for the most part unclear. Combining *matK* sequences with other gene sequences has strong potential to provide more information for inference of angiosperm phylogeny. Using additional rapidly evolving genomic regions is desirable to provide insight needed to improve our understanding of angiosperm evolution.

LITERATURE CITED

ALBACH, D. C., P. S. SOLTIS, D. E. SOLTIS, AND R. G. OLMSTEAD. 2001. Phylogenetic analysis of asterids based on sequences of four genes. *Annals of the Missouri Botanical Garden* 88: 163–212.

ANDERBERG, A. A., C. RYDIN, AND M. KÄLLERSJÖ. 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany* 89: 677–687.

APG (ANGIOSPERM PHYLOGENY GROUP). 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85: 531–553.

APG II. 2003. An update of APG classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society* 141: 399–436.

BARKMAN, T. J., G. CHENERY, J. R. MCNEAL, J. LYONS-WEILER, W. ELLISENS, G. MOORE, A. D. WOLFE, AND C. W. DEPAMPHILIS. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences, USA* 97: 13 166–13 171.

BARTHLOTT, W., I. THEISEN, T. BORSCH, AND C. NEINHUIS. 2003. Epicuticular waxes and vascular plant systematics: integrating micromorphological and chemical data. *In* T. F. Stuessy, V. Mayer, and E. Hörandl [eds.], Deep morphology: toward a renaissance of morphology in plant systematics, 189–206. *Regnum Vegetabile 141.* Ganter Verlag, Rugell, Liechtenstein.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Non-coding plastid *trnT-trnF* sequences reveal a well supported phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

BOWE, L. M., G. COAT, AND C. W. DEPAMPHILIS. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales closest relatives are conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4092–4097.

BREMER, B., K. BREMER, N. HEIDARI, P. ERIXON, R. G. OLMSTEAD, A. A.

ANDERBERG, M. KALLERSJO, AND E. P. BARKHORDARIAN. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxoninc levels. *Molecular Phylogeny and Evolution* 24: 274–301.

BREMER, B., R. K. JANSEN, B. OXELMAN, M. BACKLUND, H. LANTZ, AND K.-J. KIM. 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.

CHASE, M. W., M. R. DUVALL, H. G. HILLS, J. G. CONRAN, A. V. COX, L. E. EGUIARTE, J. HARTWELL, M. F. FAY, L. R. CADDICK, K. M. CAMERON, AND S. HOOT. 1995. Molecular phylogenetics of Lilianae. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution, 109–137. Royal Botanic Gardens, Kew, UK.

CHASE, M. W., D. E. SOLTIS, P. S. SOLTIS, P. J. RUDALL, M. F. FAY, W. H. HAHN, S. SULLIVAN, J. JOSEPH, T. GIVNISH, K. J. SYTSMA, AND C. PIRES. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. *In* K. K. L. Wilson and D. Morrison [eds.], Proceedings of monocots II: the second international symposium on the comparative biology of the monocotyledons, 3–16. CSIRO Press, Sydney, Australia.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.

CHIBA, T., T. HARADA, S. GOTO, R. ISHIKAWA, AND M. NIIZEKI. 1996. Transcription of tRNA genes from large-scale plastid DNA deletion clearly reveals the action of nuclear-encoded RNA polymerase in the plastid. *Journal of Plant Physiology* 148: 652–656.

CORNER, E. J. H. 1976. The seeds of dicotyledons, vols. 1, 2. Cambridge University Press, Cambridge, UK.

COSNER, M. E., R. K. JANSEN, AND T. G. LAMMERS, 1994. Phylogenetic relationship in the Campanulales based on *rbcL* sequences. *Plant Systematics and Evolution* 190: 79–95.

CRONQUIST, A. 1988. The evolution and classification of flowering plants, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

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

DAHLGREN, R. M. T. 1980. A revised classification of the angiosperm. *Botanical Journal of the Linnean Society* 80: 91–124.

DONOGHUE, M. J., AND M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, 340–368. Chapman and Hall, New York, New York, USA.

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

DOYLE, J. A., AND P. K. ENDRESS. 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. *International Journal of Plant Sciences* 161(Supplement): 121–153.

DRINNAN, A. N., N. R. CRANE, AND S. B. HOOT. 1994. Patterns of floral evolution in the early diversification of non-magnoliid dicotolydons (eudicots). *Plant Systematics and Evolution (Supplement)* 8: 93–122.

DU JARDIN, P., D. PORTETELLE, L. HARVENGT, M. DUMONT, AND B. WATHELET. 1994. Expression of intron-encoded maturase-like polypeptides in potato chloroplasts. *Current Genetics* 25: 158–163.

EMS, S. C., C. W. MORDEN, AND K. W. DIXON. 1995. Transcription, splicing, and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* 29: 621–733.

ENDRESS, P. K., AND A. IGERSHEIM. 2000. Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Sciences* 161 (Supplement): 211–213.

FARRIS, J. S. 1977. Phylogenetic analysis under Dollo's law. *Systematic Zoology* 26: 77–88.

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

FARRIS, J. S. 2002. RASA attributes highly significant structure to randomized data. *Cladistics* 18: 334–353.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.

FAY, M. F., P. J. RUDALL, S. SULLIVAN, K. L. STOBART, A. Y. BRUIJN, G. REEVES, F. QAMARUZ-ZAMAN, W. P. HONG, J. JOSEPH, W. J. HAHN, J. G. CONRAN, AND M. W. CHASE. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. *In* K. K. L. Wilson and D. Morrison [eds.], Proceedings of monocots II: the second international symposium on the comparative biology of the monocotyledons, 347–359. CSIRO Press, Sydney, Australia.

FISHBEIN, M., C. HIBSCH-JETTER, D. E. SOLTIS, AND L. HUFFORD. 2001. Phylogeny of Saxifragales (angiosperms, eudicots): analysis of a rapid, ancient radiation. *Systematic Biology* 50: 817–847.

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

GRAHAM, S. W., AND R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS, AND R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161(Supplement): 83–96.

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

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

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy and investigator bias. *Systematic Biology* 47: 3–8.

HILU, K. W., AND L. A. ALICE. 1999. Systematic and evolutionary implications of *matK* indels in Poaceae. *American Journal of Botany* 86: 1735–1741.

HILU, K. W., L. A. ALICE, AND H. LIANG. 1999. Phylogeny of Poaceae inferred from *mat*K sequences. *Annals of the Missouri Botanical Garden* 86: 835–851.

HILU, K. W., AND H. LIANG. 1997. The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.

HOLMQUIST, R. 1983. Transitions and transversions in evolutionary descent: an approach to understanding. *Journal of Molecular Evolution* 19: 134–144.

HOOT, S. B., S. MAGALLÓN, AND P. R. CRANE. 1999. Phylogeny of basal eudicots based on three molecular datasets: *atpB, rbcL,* and 18S nuclear ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 86: 1–32.

HUELSENBECK, J. P., B. LARGET, R. E. MILLER, AND F. RONQUIST. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51: 673–688.

HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes. Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.

JOHNSON, L. A., AND D. E. SOLTIS. 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* 19: 143–156.

JOHNSON, L. A., AND D. E. SOLTIS. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. *Annals of the Missouri Botanical Garden* 82: 149–175.

JUDD, W. S., R. W. SAUNDERS, AND M. J. DONOGHUE. 1994. Angiosperm family pairs: preliminary phylogenetic analyses. *Harvard Papers in Botany* 5: 1–51.

KADEREIT, G., T. BORSCH, K. WEISING, AND H. FREITAG. In press. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences.*

KÄLLERSJÖ, M., V. A. ALBERT, AND J. S. FARRIS. 1999. Homoplasy increases phylogenetic structure. *Cladistics* 15: 91–93.

KELCHNER, S. A. 2000. The evolution of noncoding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

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

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 29: 170–179.

KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18: 1–32.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, AND M. NEI. 2001. MEGA 2.

Molecular evolutionary genetics analysis software. Arizona State University, Tempe, Arizona, USA.

LIERE, K., AND G. LINK. 1995. RNA-binding activity of the *matK* protein encoded by the chloroplast *trnK* intron from mustard (*Sinapis alba* L.). *Nucleic Acids Research* 23: 917–921.

LYONS-WEILER, J., G. A. HOELZER, AND R. J. TAUSCH. 1996. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. *Molecular Biology and Evolution* 13: 749–757.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MATHEWS, S., AND M. J. DONOGHUE. 2000. Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. *International Journal of Plant Sciences* 161(Supplement): 41–55.

METCALFE, C. R., AND L. CHALK. 1950. Anatomy of the dicotyledons. Leaves, stem, and wood in relation to taxonomy with notes on economic uses, 1st ed., vols. 1–2, Clarendon Press, Oxford, UK.

MOHR, G., P. S. PERLMAN, AND A. M. LAMBOWITZ. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Research* 21: 4991–4997.

MÜLLER, K. 2002. PRAT: computer program for phylogenetic analyses of large data sets. Program distributed by the author, Botanical Institute, University of Bonn, Bonn, Germany. http://www.botanik.uni-bonn.de/system/download/PRAT

MÜLLER, K. 2003. QuickAlign: a new alignment editor. *Plant Molecular Biology Reporter* 21: 5.

MÜLLER, K., T. BORSCH, L. LEGENDRE, S. POREMBSKI, I. THEISEN, AND W. BARTHLOTT. In press. Evolution of carnivory in Lamiales and Lentibulariaceae. *Plant Biology*.

NANDI, O. I., M. W. CHASE, AND P. L. ENDRESS. 1998. A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Annals of the Missouri Botanical Garden* 85: 137–212.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast tRNALys (UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase related polypeptide. *Current Genetics* 11: 251–257.

NIXON, K. 1999. The parsimony ratchet: a rapid means for analyzing large data sets. *Cladistics* 15: 407–414.

OLMSTEAD, R. G., B. BREMER, K. SCOTT, AND J. D. PALMER. 1993. A parsimony analysis of the Asteridae *sensu lato* based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 80: 700–722.

OLMSTEAD, R., G. C. W. DEPAMPHILIS, A. D. WOLFE, N. D. YOUNG, W. J. ELISONS, AND P. A. REEVES. 2001. Disintegration of the Scrophulariaceae. *American Journal of Botany* 88: 348–361.

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

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

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

OLMSTEAD, R. G., P. A. REEVES, AND A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II, 164–187. Kluwer, Boston, Massachusetts, USA.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular datasets in the Solanaceae. *Systematic Biology* 43: 467–481.

PARKINSON, C. L., K. L. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488.

POLLOCK, D. D., D. J. ZWICKL, J. M. MCGUIRE, AND D. M. HILLIS. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664–671.

QIU, Y.-L., M. W. CHASE, S. B. HOOT, E. CONTI, P. R. CRANE, K. J. SYTSMA, AND C. R. PARKS. 1998. Phylogenetics of the Hamamelidae and their allies: parsimony analyses of nucleotide sequences of the plastid gene *rbcL*. *International Journal of Plant Science* 159: 891–905.

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

QIU, Y.-L., J. LEE, F. BERNASCONI-QUADRONI, D. E. SOLTIS, P. S. SOLTIS, M. ZANIS, E. A. ZIMMER, Z. CHEN, V. SAVOLAINEN, AND M. W. CHASE. 2000. Phylogeny of basal angiosperms: analyses of five genes from three genomes. *International Journal of Plant Sciences* 161: S3–S27.

RANNALA, B., J. P. HUELSENBECK, Z. YANG, AND R. NIELSEN. 1998. Taxon sampling and the accuracy of large phylogenies. *Systematic Biology* 47: 702–710.

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

RENNER, S. S., AND A. S. CHANDERBALI. 2000. What is the relationship among Hernandiaceae, Lauraceae, and Monimiaceae, and why is this question so difficult to answer? *International Journal of Plant Sciences* 161(Supplement): S109–S119.

SAUQUET, H., J. A. DOYLE, T. SCHARASCHKIN, T. BORSCH, K. W. HILU, L. W. CHATROU, AND A. LE THOMAS. 2003. Phylogenetic relationships in the Magnoliales and Myristicaceae: a multi-level approach based on morphology and molecules. *Botanical Journal of the Linnean Society* 142: 125–186.

SAVOLAINEN, V., M. W. CHASE, S. B. HOOT, C. M. MORTON, D. E. SOLTIS, C. BAYER, M. F. FAY, A. Y. DE BRUIJN, S. SULLIVAN, AND Y.-L. QIU. 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* 49: 306–362.

SAVOLAINEN, V., ET AL. 2000b. Phylogeny of the eudicots: a nearly complete familial analysis of the *rbcL* gene sequences. *Kew Bulletin* 55: 257–309.

SHEAHAN, M. C., AND M. W. CHASE. 2000. Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions, with special emphasis on Zygophylloideae. *Systematic Botany* 25: 371–384.

SIMMONS, M. P., C. P. RANDLE, J. V. FREUDENSTEIN, AND J. W. WENZEL. 2002. Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. *Molecular Biology and Evolution* 19: 14–23.

SOLTIS, D. E., A. E. SENTERS, M. J. ZANIS, S. KIM, J. D. THOMPSON, P. S. SOLTIS, L. R. DE CRAENE, P. K. ENDRESS, AND J. S. FARRIS. 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany* 90: 461–470.

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

SOLTIS, D. E., AND P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, vol. II, 1–42. Kluwer, Boston, Massachusetts, USA.

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

SOLTIS, D. E., ET AL. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.

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

SOLTIS, P. S., ET AL. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SUZUKI, Y., G. V. GLAZKO, AND M. NEI. 2002. Overcredibility of molecular phylogenetics obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences, USA* 99: 16 138–16 143.

SWOFFORD, D. L. 2001. PAUP: phylogenetic analysis using parsimony, ver. 4.0b6. Sinauer, Sunderland, Massachusetts, USA.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. HILLIS. 1996. Phylogenetic inference. *In* D. M. Hillis, C. Moritz, and B. K. Mable [eds.], Molecular systematics, 2nd ed., 407–514. Sinauer, Sunderland, Massachusetts, USA.

TAKAHASHI, M. 1994. Exine development in *Illicium religiosum* Sieb. et Zucc. (Illiciaceae). *Grana* 33: 309–312.

TAKHTAJAN, A. 1987. System of Magnoliophyta. Academy of Sciences, USSR, Leningrad, Russia.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

THORNE, R. F. 1992. An updated classification of the flowering plants. *Aliso* 13: 365–389.

VOGEL, J., T. HÜBSCHMANN, T. BÖRNER, AND W. R. HESS. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for *matK* as an essential splice factor. *Journal of Molecular Biology* 270: 179–187.

WOLFE, A. D., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Science, USA* 89: 10 648–10 652.

XIANG, Q.-Y., D. E. SOLTIS, AND P. S. SOLTIS. 1998. Phylogenetic relationships of Cornaceae and close relatives inferred from m*atK* and *rbcL* sequences. *American Journal of Botany* 85: 285–297.

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

ZWICKL, D. J., AND D. M. HILLIS. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.