Rosid radiation and the rapid rise of angiosperm-dominated forests

Hengchang Wang, Michael J. Moore, Pamela S. Soltis, Charles D. Bell, Samuel F. Brockington, Roolse Alexandre, Charles C. Davis, Maribeth Latvis, Steven R. Manchester, and Douglas E. Soltis

The rosid clade (70,000 species) contains more than one-fourth of all angiosperm species and includes most lineages of extant temperate and tropical forest trees. Despite progress in elucidating relationships within the angiosperms, rosids remain the largest and most diverse. Deep relationships within the rosids are particularly enigmatic. Based on parsimony and maximum likelihood (ML) analyses of separate and combined 12-gene (10 plastid genes, 2 nuclear) datasets for >100 rosid species, we provide a greatly improved understanding of rosid phylogeny. Vitaceae are sister to all other rosids, which in turn form 2 large clades, each with a ML bootstrap value of 100%: (i) euasters I (Fabidae) include the nitrogen-fixing clade, Celastrales, Huaeeae, Zygophyllales, Malphighiales, and Oxlidiales; and (ii) euasters II (Malvidae) include Tapisciaceae, Brassicaceae, Malvales, Sapindales, Geraniaceae, Myrtiales, Crossosomatales, and Picramniaceae. The rosid clade diversified rapidly into these major lineages, possibly over a period of <15 million years, and perhaps as little as 4 to 5 million years. The timing of the inferred rapid radiation of rosids [108 to 91 million years ago (Mya) and 107–83 Mya for Fabidae and Malvidae, respectively] corresponds with the rapid rise of angiosperm-dominated forests and the concomitant diversification of other clades that inhabit these forests, including amphibians, ants, placentals, and mammals. and ferns. Community assembly | divergence time estimates | phylogeny | rapid radiation

Great progress has been made in elucidating deep-level angiosperm relationships during the past decade. The euaster clade, with >75% of all angiosperm species, comprises several major subclades: rosids, asterids, Saxifragales, Santalales, and Caryophyllales (1–3). Investigations have converged on the branching pattern of the basal angiosperms, revealing that Amborellaceae, Nymphaeales [in the sense of APG II (3) and including Hydatellaceae (4)], and Austrobaileyales are sister groups to all other extant angiosperms (reviewed in ref. 2). Analyses of complete plastid genome sequences have resolved other problematic deep-level relationships, suggesting that Chloranthaceae and magnoliids are sister to all other rosids, which in turn form 2 large clades, each with a ML bootstrap value of 100%: (i) euasters I (Fabidae) include the nitrogen-fixing clade, Celastrales, Huaeeae, Zygophyllales, Malphighiales, and Oxlidiales; and (ii) euasters II (Malvidae) include Tapisciaceae, Brassicaceae, Malvales, Sapindales, Geraniaceae, Myrtiales, Crossosomatales, and Picramniaceae. The rosid clade diversified rapidly into these major lineages, possibly over a period of <15 million years, and perhaps as little as 4 to 5 million years. The timing of the inferred rapid radiation of rosids [108 to 91 million years ago (Mya) and 107–83 Mya for Fabidae and Malvidae, respectively] corresponds with the rapid rise of angiosperm-dominated forests and the concomitant diversification of other clades that inhabit these forests, including amphibians, ants, placentals, and mammals. and ferns.

Rosids exhibit enormous heterogeneity in habit, habitat, and life form, comprising herbs, shrubs, trees, vines, aquatics, succulents, and parasites. The rosid clade also contains novel biochemical pathways, such as the machinery necessary for symbiosis with nitrogen-fixing bacteria (nitrogen-fixing clade) and defense mechanisms such as glucosinolate production (Brassicaceae) and cyanogenic glycosides (2). Many important crops, including legumes (Fabaceae) and fruit crops (Rosaceae), are rosids. Furthermore, 4 of the 5 published complete angiosperm nuclear genome sequences are rosids (with 2 other rosids, Manihot and Ricinus, well under- way): Arabidopsis (Brassicaceae), Carica (Caricaceae), Populus (Salicaceae), and Vitis (Vitaceae, sister to other rosids). The variation in morphological, chemical, and ecological features and the importance of rosids as genetic and genomic models require a phylogenetic perspective for interpreting large-scale evolutionary patterns across the clade. Finally, most lineages of extant temperate and tropical forest trees are rosids (e.g., Betulaceae, Celastraceae, Fabaceae, Fagaceae, Malvaceae, Sapindaceae, and Ulmaceae). In temperate North America >71% of the forest tree species are rosids (15). Similarly, rosids often constitute >50% of tropical tree species diversity. For example, they comprise 59% of the tree species on Barro Colorado Island, Panama (16), 60% of the tree flora of Paraguay (17), 60% on Puerto Rico and the Virgin Islands (18), and 63% of the tree species of Brazil (19). Analysis of tropical forest plots from around the world again suggests a major role of rosid trees (20). Half the species richness of the Neotropical plots is made up of just 11 families, 4 of which are rosids, with Fabaceae the most important (Moraceae, Meliaceae, and Euphorbiaceae are other major rosids in the Neotropics); Fabaceae are also the most important family of trees in Africa, but are replaced as number one by another rosid family, Dipterocarpaceae, in Southeast Asia (20). The important ecological role of many rosid families further argues for greater resolution of rosid phylogeny. Although studies agree on the composition of the rosid clade, deep-level relationships within the rosids remain enigmatic. Vitaceae usually appear as sister to all other rosids, although support for this placement is generally low (reviewed in refs. 2 and 10). Most remaining rosids appear in 2 large subclades (1–3): euasters I [Fabidae (21)] and euasters II [Malvidae (21)], with jackknife support of 77% and 95%, respectively. Fabidae include (i) the nitrogen-fixing clade (Rosales, Fabales, Cucurbitales, and Fagales); (ii) Zygophyllales; and (iii) a weakly supported clade of Celastrales, Oxlidiales, and Malphighiales [COM group (22)]. Malvidae include Brassicaceae, Malvaceae, Sapindaceae, and Tapiaciaceae. Some rosid clades (Crossoomatales, Geraniaceae, and Myrtales), however, do not fall into either Fabidae or Malvidae, and their relationships remain unclear. Myrtales are often resolved as sister...
to Fabidae, whereas Crossosomatales and Geraniales appear with Malvidae, but without strong support (1, 2, 10). Furthermore, the placements of several problematic families—Aphloiaceae, Apodanthaceae, Geissolomataceae, Huaceae, Ixerbaeaceae, Picrammiaceae, and Strasburgeriaceae—also need to be ascertained.

Rosid diversification, like angiosperm phylogeny as a whole, is characterized by repeating patterns of radiation from deep to shallow levels (2). Lack of resolution at the base of the rosid phylogeny in previous studies suggests an early radiation of crown-group rosids into major clades. Similar polytomies are evident within both Fabidae and Malvidae and within clades further nested within Fabidae (i.e., Malpighiales). However, it is unclear whether the apparent radiations within the rosids are real or merely artifacts due to poorly resolved relationships. Improved analysis of phylogeny, coupled with dating of diversification events, could help to distinguish between these alternatives.

To resolve deep relationships within the rosids and evaluate hypotheses of repeated radiations, we constructed a dataset for >100 rosids using sequences of 12 genes (two nuclear and 10 plastid) representing >18,000 bp. We also constructed a dataset using the entire plastid inverted repeat (IR; 24 genes plus inter-spacers; > 25,000 bp), a region with great utility at deep levels in the angiosperms (23); these datasets were also combined (>43,000 bp). Our goals were to (i) elucidate deep relationships among the major lineages of rosids; (ii) determine the placement of large, problematic lineages (e.g., Crossosomatales, Geraniales, Myrtales), and enigmatic families (Aphloiaceae, Huaceae, Ixerbaeaceae, Picrammiaceae, and Strasburgeriaceae) not placed in either Fabidae or Malvidae in previous studies; (iii) date the early diversification of the rosid clade; and (iv) reconsider the fossil record in light of this resolved phylogeny.

Results

Parasimony Analyses. Maximum parsimony (MP) analysis of the total evidence dataset (12 targeted genes plus IR for 117 taxa) produced 9 shortest trees (Fig. S1) that differ mainly in the relationships recovered among some Malpighiales. There are no instances of strongly supported conflicting relationships (i.e., > 80% BS support) among trees from the individual genes (trees not shown; tree statistics in Table S1). Single-gene analyses generally recovered clades that correspond to angiosperm orders (in the sense of ref. 3), but, with few exceptions, failed to resolve interordinal relationships. MP analyses of the combined nuclear genes (18S and 26S rDNA; trees not shown) also provided limited resolution and support of relationships. In contrast, MP analyses of combined plastid and nuclear datasets (12 targeted genes) yielded trees with much greater resolution (Figs. S2–S5).

Maximum Likelihood Analyses/Comparison with Parsimony. Maximum likelihood (ML) analyses of 10 targeted plastid genes and 10 plastid plus 2 nuclear genes (Figs. S2–S5) yielded topologies nearly identical to the ML total evidence tree for 117 taxa (Fig. 1). In ML analyses of the total evidence dataset (12 targeted genes plus IR), Zygodephyllales were placed in Malvidae as sister to Geraniales, whereas with ML Zygodephyllales were sister to Fabidae (Fig. 1). A more minor difference is that in MP analyses, Fabales were sister to Rosales whereas in ML analyses Fabales were sister to a clade of Rosales, Cucurbitales, and Fagales. With both ML and MP, the slowly evolving IR region alone yielded a topology (Fig. S2) identical to that retrieved by ML analysis of the total evidence dataset (Fig. 1). The conflicts between MP and ML analyses are likely due to the inherent difficulties that long branches pose for parsimony (see below); indeed both Fabales and Zygodephyllales contain lineages that are on long branches in these trees. Therefore, for the remainder of this article we will focus on the ML topology of the total evidence tree. In this tree (Fig. 1), Vitaceae are sister to all other rosids, which in turn form 2 large clades, Fabidae and an expanded Malvidae, each with a bootstrap value of 100%.

The approximately unbiased (AU) topology test indicated that the sister relationship of Zygodephyllales to Malvidae observed in MP can be rejected (P < 0.001) in favor of the ML placement of Zygodephyllales as sister to an expanded Fabidae using both the total evidence dataset and the 12 targeted genes dataset.

Divergence Time Estimates. Cross-validation determined a smoothing value of 1,000 for each of the penalized likelihood (PL) analyses. In general, when the root of the tree was fixed to an age of 125 Mya, age estimates were an average of 5–10 million years older than when the root was constrained to a maximum age of 125 Mya (Table 1). Likewise, based on the bootstrap estimates of variation, the 2 different methods usually did not fall within one standard deviation of one another. Our estimates for the origin of crown group rosids from the 2 analyses ranged from 110 (+ 6) to 93 (+ 6) Mya, in the Early to Late Cretaceous, followed by rapid diversification into the Fabidae and expanded Malvidae clades (+108 (+ 6) to 91 (+ 6) Mya and 107 (+ 6) to 83 (+ 7) Mya, respectively (Table 1).

The divergence times estimated under the uncorrelated lognormal (UCLN) relaxed-clock model with the data partitioned by genome, treating fossils as drawn from a uniform distribution (UCLN-uniform), estimated a coefficient of variation of 0.815 [95% HPD (highest posterior density): 0.777, 0.863] and covariance of 0.136 (95% HPD: 5.62 × 10⁻², 0.237) for the plastid partition and a coefficient of variation of 0.634 (95% HPD: 0.543, 0.736) and covariance of 0.067 (95% HPD: 1.62 × 10⁻², 0.147) for the nuclear partition. Parameter values for the UCLN relaxed-clock model with fossils treated as being drawn from a lognormal distribution (UCLN-lognormal) were almost identical. Accordingly, these data appear suited to divergence-time estimation methods that assume a positive autocorrelation of substitution rate variation (e.g., refs. 24–28). However, for the Bayesian relaxed-clock models, the partitioned model with fossils drawn from a uniform distribution was determined to be the best-fitting model based on Bayes factors (mean for the log of posterior equaled −3.76 × 10⁵ for the “uniform” model and −3.78 × 10⁵ for the “lognormal” model). Our estimates for the origin of crown group rosids from the 2 fossil treatments ranged from 115 (119–112) to 113 (117–111) Mya, once again in the Early Cretaceous, followed by rapid diversification into the Fabidae and Malvidae crown groups +112 (115–110) to 110 (113–109) Mya and 109 (113–105) to 106 (111–102) Mya, respectively. In general, the variance on nodes tended to be smaller for the analysis that treated fossils as being drawn from a lognormal distribution.

Discussion

Phylogenetic Relationships. MP and ML analyses of the total evidence dataset yielded similar topologies. However, the primary difference between the MP and ML trees is significant, with Zygodephyllales appearing as either part of an expanded Malvidae with MP or sister to the expanded Fabidae with ML. MP bootstrap values for Fabidae and Malvidae (77% and 75%, respectively) are lower than ML bootstrap values (both 100%). The AU test results provide statistical support for the ML total evidence topology by rejecting the MP position of Zygodephyllales as sister to Malvidae. IR data alone (below) also support the ML topology. In addition, previous studies (MP and Bayesian) of a 3-gene dataset with more rosid taxa also placed Zygodephyllales in Fabidae, but in an unresolved position (2, 10). Zygodephyllales are characterized by a long branch, which could affect the placement of the clade (particularly with parsimony); the branches within Zygodephyllales (to Krameria, to Guaiaucum, and to Bulnesia) are particularly long (Figs. S2–S5). Zygodephyllales are highly enigmatic and share few non-DNA traits with any other rosid lineage (2).

With both ML and MP (Fig. S2), the IR dataset alone yielded a topology identical to that retrieved by ML analysis of the total evidence dataset.
evidence dataset (Fig. 1). These results parallel those reported for Saxifragales (23), illustrating the value of the slowly evolving IR in deep-level phylogeny reconstruction. Despite far fewer parsimony-informative sites than “fast” genes, the slowly evolving IR genes provide higher resolution and support for relationships, providing a topology identical to the total evidence ML tree.

Fig. 1. ML tree resulting from GARLI analysis of total evidence data set (2 nuclear genes, 10 plastid genes) for 117 members of the rosid clade and outgroups and the IR for 59 taxa. Numbers above branches are bootstrap values.
We will base our discussion on the ML total evidence tree for 117 taxa (Fig. 1) for several reasons. Nodes in the ML tree receive high bootstrap support, whereas with MP, many crucial relationships (including placement of Zygophyllales) have low bootstrap values. Parsimony has also had difficulty resolving other deep-level phylogenetic problems in angiosperms that appear to be rapid radiations and that are similarly characterized by a combination of short and long branches (e.g., refs. 6 and 23). Furthermore, ML and Bayesian approaches are better than MP under heterogenous evolution involving lineage- and gene-specific rate variation (29–31).

In previous studies, Fabidae were not well supported, and relationships were unresolved within the clade (1, 2, 10). Our analysis not only provides strong support for Fabidae (100% bootstrap with ML), but also indicates that Zygophyllales are sister to 2 major subclades within Fabidae (each with BS = 100%): the nitrogen-fixing clade (Cucurbitales, Fagales, Fabales, Rosales) and Celastrales, Oxalidales, and Malpighiales [COM group (22)], plus Huaceae.

Previous studies placed Tapisciaceae, Brassicales, Malvales, and Sapindales in Malvidae, but with little support for relationships among these clades (reviewed in refs. 2 and 10). Furthermore, Geraniales, Myrtales, and Crossosomatales were not placed in either Fabidae or Malvidae in previous analyses. We provide the first strong evidence that these 3 clades, and Picramniaceae [placed by some (12) within Simaroubaceae, Sapindales], are part of an expanded Malvidae (100% bootstrap with ML). We propose that Malvidae be formally expanded to include Geraniales, Myrtales, Crossosomatales, and Picramniaceae. Furthermore, Crossosomatales should be expanded to include Ixeribaceae and Strasburgeriaceae (2).

This hypothesis of rosid relationships is greatly improved compared with trees based on approximately ~4,700 bp (8) or ~8,400 bp (32). This study and other recent analyses (5, 6, 23) suggest that with enough data, many, if not most, remaining deep-level “radiations” in the flowering plants can be resolved. For example, a putative rapid radiation in Saxifragales was resolved, despite long-branch attraction problems, with >25,000 base pairs (23). Highly problematic deep-level relationships involving basal angiosperm lineages have similarly been resolved through analysis of ~42,000 bp (6).

Divergence Times. Our estimates for the origin of crown group rosids ranged from 115 to 93 Mya (late Aptian to early Turonian), in the Early to Late Cretaceous, followed by rapid diversification into the Fabidae and Malvidae crown groups ~112 to 91 Mya (Albian to Coniacian), and 109 to 83 Mya (Cenomanian to Santonian), respectively (Table 1). With the exception of the PL analysis that...
constrained the root to be a maximum age of 125 Mya, age estimates were quite consistent. For the most part, associated errors in age estimates overlapped across all nodes in the tree so that we cannot precisely resolve the actual timing and rates of these successive divergences.

The oldest confirmed fossils of Vitacea are Paleocene, ~60 Mya (33), which is anomalously young in view of the mid- to late-Cretaceous ages known for members of Fabidae and Malvidae. We predict, based on divergence time estimates presented above, that Vitacea should have a Cretaceous fossil history. However, the distinctive seeds of Vitacea have not been detected in the pre-Tertiary fossil record.

Our estimate for the time of origin of the crown group rosid clade is similar to those of both Wikström et al. (34) (117–108 Mya, their node 15) and Davies et al. (35) (~115–110 Mya). Inspection of the ages of rosid lineages obtained here revealed no general pattern to these node ages being older or younger, when compared with Wikström et al. (34). Regardless of analyses, the rosid clade appears to have diversified rapidly into several major lineages, over a period of ~15 million years, and perhaps as quickly as 4 to 5 million years.

To place this in perspective, this narrow window of initial diversification of ~4–5 million years represents a timeframe comparable to the rapid radiation of the Hawaiian silversword alliance (Asteraceae), which arose from a North American ancestor 5 Mya (36). Saxifragales (23), Malpighiales (34, 37), and Mesangiospermae (6) also each appear to have arisen and diversified over a very short period.

**Rise of Angiosperm-Dominated Forests.** We hypothesize that the bursts in diversification in the rosids correspond to the rapid rise of angiosperm-dominated forests (38, 39). Fossil leaf assemblages of late Albian to middle Cenomanian age (~104–97 Mya) have been interpreted as “an explosive increase in the structural diversity of flowering plants” (ref. 39, p. 259), not only magnoloids and platanoids, but also new types of rosids with “pinnately compound leaves and simple leaves showing evidence for derivation from a compound-leaved ancestor” (ref. 39, p. 259). This interval partially coincides with our inferred dates of the divergence for the rosid crown group and of the Fabidae and Malvidae. The results of these diversifications are already well-marked in the Late Cretaceous, by which time taxa related to Saxifragales, Rosales, Malpighiales, and Fagales are common, as confirmed by well-preserved charcoalified floral remains (37, 40, 41).

Our estimates of the timing of the rapid diversification of these rosid lineages are comparable to published values (34, 35), and those provided in more focused analyses within rosid clades. For example, Davis et al. (37) concluded that Malpighiales are very old and began to diversify during the mid-Cretaceous (~112–94 Mya), perhaps as small to mid-sized trees in tropical rain forest understories. A similar window of diversification is seen in woody species in clades other than the rosids, such as the woody Saxifragales [e.g., Altingiaceae (23)]. Within asterids, Cornales also appear to have diversified during this same window of time (e.g., 35), as did Ericales (34, 35).

The diversification of rosids also corresponds to that of several major insect groups. For example, the diversification of major ant lineages is attributed to the “rise in angiosperm-dominated forests” (42) and corresponds to the time period estimated here for the rosid radiation. This period also corresponds to the radiation of other major herbivores, such as beetles and hemipterans (43, 44). Similarly, the “principal splits” underlying the diversification of the extant lineages of placental mammals occurred in a similar time-frame, from 100 to 85 Mya (45). Diversification in amphibians occurred slightly later (80–85 Mya) and is similarly attributed to the rise of angiosperm forests (46), especially given that 82% of amphibian species live in these forests. The majority of extant ferns similarly resulted from a Cretaceous diversification (initiated ~100 Mya) coupled with the rise of angiosperm forests; divergence time estimates suggest that ferns diversified “in the shadow of the angiosperms” (47). The rise of all of these lineages appears to have closely tracked the rise of angiosperm-dominated forests, and most of these key forest lineages occur within the rosid clade.

Hence, the radiations we have detected in rosids largely represent the rapid rise of angiosperm-dominated forests and associated codiversification events.

**Materials and Methods**

**Taxon Sampling.** A total of 117 species (including 104 species of rosids) was sampled for 12 genes (see DNA Amplification and Sequencing); a subset of 59 taxa was sampled for the plastid IR. We sampled broadly across the rosids, including several exemplars of most orders (SI Appendix). For larger orders, multiple families spanning the phylogenetic diversity of the clade were included. We also broadly sampled orders and families whose placements remain uncertain, including Picramniaceae, Huaceae, Aplioiaceae, Ixerpaceae, and Strasburgeriaceae (3).

For most taxa, the same DNA was used throughout the study. For a few genera, sequences for 1 or 2 of the genes sequenced in earlier studies represent a different congeneric species from that used here. We selected non-rosids representing the other major lineages of eudicots as outgroups (Fig. 1). Species names and voucher collection information are given in SI Appendix.

**DNA Amplification and Sequencing.** We targeted 10 plastid [rbcL, atpB, matK, the psbB/psbH region (~4 genes), rpoC2, ndhF, and rdB] and 2 nuclear genes (18S and 26S rDNA) for all taxa. Primers for amplification and sequencing are provided in Table S2. PCR protocols follow ref. 23. We sequenced the entire IR (see ref. 23) for 59 of the rosids used in our 12-gene analyses (due to expense, we did not sequence the IR for all taxa). We also extracted IR sequences for all publicly available complete plastid genome sequences for rosids, and for several outgroups (Fig. 1 and SI Appendix). Sequences were generated on an ABI 3730 XL DNA sequencer.

**Alignment.** Preliminary alignments of the 12 targeted genes were obtained independently for each locus using Clustal X (48) and adjusted manually. To assist in the alignment of protein-coding regions, sequences were also aligned by amino acid. Alignment of coding sequences for all genes except ndhF and rpoC2 was straightforward; the latter 2 genes were more problematic because they possess many indels and some regions of high sequence variability. Regions that were difficult to align, and short incomplete regions at the beginnings and ends of genes, were excluded from the analyses. The entire IR was aligned as in ref. 23. The aligned, analyzed lengths of the genes included are given in Table S3.

**Phylogenetic Analysis.** We analyzed 4 datasets: (i) 12 targeted genes for 117 taxa; (ii) the IR for 59 taxa; (iii) 12 targeted genes plus IR for 117 taxa (with those taxa not sequenced for the IR scored as having missing data); and (iv) 12 targeted genes plus IR for those taxa sequenced for both (59 taxa). We also analyzed each of the 12 targeted genes individually.

Gaps were treated as missing data. MP and ML analyses were used to infer the phylogeny. With the exception of the nuclear data partition, all MP analyses were conducted using heuristic searches with 1,000 replicates of random taxon addition with tree-bisection-reconnection (TBR) branch swapping, saving all shortest trees per replicate. For the nuclear partition, MP heuristic searches involved 3,000 random addition replicates, saving no more than 1,000 trees per replicate. Bootstrap support was estimated following the methods of ref. 6 except for the nuclear partition, in which only 10,000 shortest trees per replicate were saved for the bootstrap analysis.

For ML analyses we used GARLI v. 0.942 (49). MrModelTest 2.2 (people.scs.fsu.edu/~nylander/mrmodeltest2/mrmodeltest2.html) and the Akaike information criterion was used to determine the appropriate evolutionary model (GTR + I + I). GARLI runs, including ML bootstrap analyses, were conducted following ref. 6.

To test the statistical significance of the differing positions of Zygophyllales in MP and ML analyses, we performed AU tests of topology (50) using the total evidence (all taxa) and the 12 targeted genes datasets. For both datasets we tested the total evidence ML topology against the topology with the highest ML score constrained to have Zygophyllales form a monophyletic group with Malvidae. To find the latter tree topology, we performed ML searches of both datasets using GARLI with G. 951 (49) with the topological constraint of Zygophyllales + Malvidae enforced and a GTR + I + I + I model. The AU test was performed in CONSEL v. 0.11 (51).

**Divergence Time Estimates.** Given the lack of rate constancy among lineages (based on a likelihood ratio test; *P* ~ 0.001), for all datasets divergence times were
estimated under a relaxed molecular clock. We used both penalized likelihood (PL, in r8s v.1.70: 27, 53) and a Bayesian method (BEAST v.1.4.7) (53-55). For PL analyses the smoothing parameter (λ) was determined by cross-validation. The ML topology was used. The basal polytomy involving the outgroups (Fig. 1) was resolved by rooting the tree with Platanaceae. The tree was then imported into PAUP* (56), where branch lengths were reestimated under a GTR + I + I model for use in PL analyses. To quantify errors in divergence time estimates, we used a 3-step hierarchical method (57).

The uncorrelated lognormal (UCLN) model implemented in BEAST (53) was used to infer divergence times. The underlying model of molecular evolution was specified to be GTR + I + I. The estimation of absolute divergence times requires calibrating (or constraining) the age of 1 or more nodes. For each analysis, we initiated 4 independent MCMC analyses from starting trees with branch lengths calibrated (or constraining) the age of 1 or more nodes. For each analysis, we used to infer divergence times. The underlying model of molecular evolution was for use in PL analyses. To quantify errors in divergence time estimates, we used a 3-step hierarchical method (57). The smoothing parameter (λ) was determined by cross-validation. The tree was then imported into PAUP* (56), where branch lengths were reestimated under a GTR + I + I model for use in PL analyses. To quantify errors in divergence time estimates, we used a 3-step hierarchical method (57).

ACKNOWLEDGMENTS. We thank D. Swoford for access to computational resources (supported by National Science Foundation Grant EF-0321495); M. Gitzendanner for help with some of the phylogenetic analyses; and B. Moore, S. Magallon, and an anonymous reviewer for helpful comments on the manuscript. This work was supported by Assembling the Tree of Life (ATOL) Grants EF-0431266 and EF-0431242 (National Science Foundation) and Deep Time Research Coordination Networks Grant NSF DEB-0090283.
SI Appendix

Voucher information and GenBank accession numbers for species used in this study. Voucher specimens are deposited in the following herbaria: FLAS = Florida Museum of Natural History, IND = Indiana University; K = Royal Botanic Gardens, Kew; MICH = University of Michigan; MO = Missouri Botanical Garden; SIU = Southern Illinois University; WS = Marion Ownbey Herbarium, Washington State University; WTU = University of Washington. Dashes indicate missing data.


EUROSID I

Geraniales: Hypseocharis bilobata Killip — —, —, —, —, L14699 [Hypseocharis sp.], EU002234, EU002421, EU002511, EU002333; M. W. Chase 2785 (K).


Myrtales

Combretaceae: Terminalia cattapa L.—AF207037, AF479147 [T. boivinii Tul.], AF209686, —, AF206826, EU002267, EU002459, EU002549, EU002370; E. Conti 103 (WIS). Crypteroniaceae: Crypteronia paniculata Blume — —, EU002152, —, —, EU002154, —, AY151566, AY078153, EU002217, EU002493, EU002313; M. W. Chase 1235 (K). Lythraceae: Lythrum flagellare Shuttlw. ex Chapman — AF206955 [L. salicaria L.], AF479240 [L. salicaria L.], AF209621 [L. salicaria L.], EU002183, —, EU002240, EU002430, EU002520, EU002342; D. Soltis 2704 (FLAS). Melastomaceae: Clidemia dentata D. Don — —, —, AJ235777, EU002173, AF215535 [C. rubra G. Don. F], EU002211, EU002396, EU002487, EU002307; D. Soltis 2708 (FLAS). Myrtaceae: Eucalyptus globulus Labill.—AM235528 [E. lehmanii (Schauer) Benth.], AM235600 [E. lehmanii (Schauer) Benth.], NC_008115, NC_008115, NC_008115, NC_008115, NC_008115, NC_008115, Myrtus communis L.— —, —, —, EU002154, —, AY151589, —, —, EU002352; D. J. McDonald (MO accession # 5708989). Vochysiaceae: Qualea parviflora Mart.—AF207002 [Q. sp.], —, AF209662 [Q. sp.], AF368216 [Q. grandiflora Mart.], U02730 (Q. sp.), EU002253, EU002444, EU002534, EU002356; M. W. Chase 168 (K).
Crossosomatales

**Crossosomataceae:** *Crossosoma californicum* Nutt.—U42529, AF479131, AF209571, DQ443456 [C. bigelovii S. Watson], AY101844, EU002216, EU002400, EU002492, EU002312; cult. at Rancho Santa Ana Bot Gard. **Stachyuraceae:** *Stachyurus praecox* Siebold & Zucc.—AF207025, —, AJ235609, DQ443457, AJ235794, EU002260, EU002452, EU002363; M. W. Chase 800 (K).

**Staphyleaceae:** *Staphylea trifoliata* Payer —AJ235978, AF479133, EU002168, EU002189, EU002285, EU002261, EU002453, EU002543, EU002364; Univ. of Washington Arboretum accession # 294-38.

**Fabales**

**Fabaceae:** *Albizia julibrissin* Durazz.—U42536, X66754, AF209524, AY386855, Z70147, EU002195, EU002379, EU002288; D. Soltis 2633 (FLAS).

**Medicago truncatula** Gaertn.—AF093506, Z11498 [M. sativa L.], NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119, NC_003119. **L. japonicus Regel) K.Larsen.—**—, —, —, —, —, —, —, —, —, —; Bruce Barrett 238 (MO).

**Polygala cruciata** L.—U42797 [P. pauciflora Thunb.], AF479233, AJ235568, AY386842 [P. californica Nutt. ex Torr. & A.Gray], L01945, —, EU002443, EU002533, EU002355; M. W. Chase 155 (K).

**Quillajaceae:** *Quillaja saponaria* Molina — —, —, —, AY386950, U07680, EU002264, EU002456, EU002367; D. Soltis R86 (FLAS). **Surianaceae:** *Stylobasium spathulatum* Desf.—AF207030, —, AF209681, —, U06828, EU002263, EU002455, EU002254, EU002256; Brunmitt, George & Oliver 21242 (K).

**Suriana maritima** L.— —, —, —, —, —, EU002456, EU002367; D. Soltis 2699 (FLAS). **Fabaceae:** *Alnus sinuata* Rydb.—AY263900, AF479106 [A. glutinosa Medik.], AY147101, AB060053 [A. firma Siebold & Zucc.], AY263926, EU002196, EU002380, EU002469, EU002289; R. G. Olmstead 2005-5 (WTU). **Fagaceae:** *Quercus nigra* L.—AY147108 [Q. multínervis (W. C. Cheng & T. Hong) Govaerts], AY094177 (Q. suber Kotschy), EU002167, EU002188, EU002224, EU002254, EU002445, EU002355, EU002357; M. J. Moore 311 (FLAS). **Juglandaceae:** *Juglans nigra* L.—AF206943, AF479105, AF209609, AF118036, AF206785, EU002237, EU002425, EU0022515, EU002337; D. Soltis 2520 (WS). **Myricaceae:** *Myrica cerifera* L.—AF206967, AF479247, AJ235537, AY191715 [M. gale L.], AF119179, EU002243, EU002433, EU002253, EU002345; D Soltis 2699 (FLAS). **Cucurbitales**

**Anisophylleaceae:** *Anisophyllea fallax* Scott-Elliot—AY968390, AF479106 [A. glutinosa Medik.], AY968401, AY968424, AY968444, AF027109, AY968487, —, EU002470, —; L. Gautier 3256 (MO). **Begoniaceae:** *Begonia cucullata* Willd.—AF008950 [B. oxyloba Welw. ex Hook. f.], AY968403 [B. oxyloba Welw. ex Hook. f.], AY968426 [B. oxyloba Welw. ex Hook.
Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).

Malpighiales


Cucurbitaceae: Cucumis sativa L. — AF206894, —, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144, NC_007144.

Datiscaceae: Datisca cannabina L. — AF008952, AY968410, AY788203, —, AY425054, L21939, EU002219, EU002403, EU002495, EU002315; M. W. Chase 2745 (K).
Salicaceae: *Populus alba* L.—[DQ371807, AF479118].

Violaceae: *Hymenanthera alpina* (T. Kirk) W. R. B. Oliv.—[AF206933, —, AJ235499, —, Z75692, EU002232, EU002419, EU002509, EU002331; M. W. Chase 501 (K) (CCD 352)].

*Leonia glycyarpa* Ruiz & Pav.—[AY674613, —, Z75693 [L. glycyarpa Ruiz & Pav.], —, Z75693 [L. glycyarpa Ruiz & Pav.], EU425065, EU002428, EU002518, EU002340; Pennington 13852 (K)].

Celastrales

Celastraceae: *Brexia madagascarensis* Thouars ex Ger. Gawl.—[U42543, AF479112, AJ235419, AY935899, AY935911, AY935932, EU002170, EU002277, EU002298, EU002320; D. Soltis 2714 (FLAS)].

**Oxalidaceae**: *Eucryphia lucida* (Labill.) Baill.—[U42533, AF036494, AF209584, EU002176, EU002186, EU002282, EU002248, EU002438, EU002528, EU002320; Strybing Arboretum accession no. 74.0135].

**Elaeocarpaceae**: *Elaeocarpus angustifolius* Blume.—[AF206906 [E. sphaericus Schum.], AF479128 [E. sphaericus Schum.], EU002174, U06798, EU002213, EU002398, EU002489, EU002309; JLD.]

**Elaeagnaceae**: *Elaeagnus sp.—*[L24090 [E. umbellata Thunb.], AF479235, AJ235462, AY257529 [E. umbellata Thunb.], AJ235462, EU002221, EU002405, EU002496, EU002317; Nickrent 2898 (SIU)].

*Celtidaceae*: *Celtis occidentalis* L.—[U42818 [C. yunnanensis C. K. Schneid.], AF479098 [C. yunnanensis C. K. Schneid.], EU002237, EU002419, EU002494, EU002485, EU002305; D. Soltis 2701 (FLAS)].

**Rhamnaceae**: *Ceanothus...*

Zygophyllales

Unplaced in eurosid I
Huaceae: Hua gabonii Pierre ex De Wild. — AY929345, AY935796, AY935830, AY935903, AY935726, EU002230, EU002417, EU002508, EU002329; J. L. Doucet 456 (MO).

EUROSID II
Brassicaceae

Malvales
Bixaceae: Bixa orellana L. — AF206868, AF479229, AF035897, —, Y15139, EU002202, EU002386, EU002476, EU002296; D. Soltis 2695 (FLAS). Cistaceae: Helianthemum grandiflorum DC. — AF206926, —, AF035907, DQ092969, Y15141,
Dipterocarpaceae: *Anisoptera marginata* Korth.—AF206849, —, AF035918, AB006371 [*A. oblonga* Dyer.], Y15144, EU002197, EU002381, EU002471, EU002290; *M. W. Chase* 2486 (K).

Malvaceae: *Gossypium barbadense* L.—, —, NC_008641, NC_008641, NC_008641, NC_008641, NC_008641, NC_008641.

Neuradaceae: *Neurada procumbens* L.—AF206970, AF479225, AF209637, —, U06814, EU002435, EU002525, EU002347; Collenette 8193 (K).

Thymelaeaceae: *Thymelaea sanamunda* All.—AF207041 [*T. hirsuta* Endl.], AF479234 [*T. hirsuta* Endl.], AJ233097 [*T. hirsuta* Endl.], EU002191, Y15151 [*T. hirsuta* Endl.], EU002268, EU002460, EU002550, EU002371; *M. W. Chase* 1882 (K).

Sapindales

Anacardiaceae: *Schinus terebinthifolius* Raddi.—AF207015 [*S. molle* L.], —, AF035914 [*S. molle* L.], AY491645 [*S. sp.*], U39270, EU002257, EU002449, EU002539, EU002361; *D. Soltis* 2705 (FLAS). Burseraceae: *Bursera simaruba* (L.) Sarg.—AF206877 [*B. inaguensis* Britton.], AY177421 [*B. inaguensis* Britton.], AF035899, —, L01890 [*B. inaguensis* Britton.], EU002206, EU002389, EU002479, EU002300; *D. Soltis* 2649 (FLAS). Meliaceae: *Swietenia macrophylla* King.—AF207031, AF479241, AJ235616, AY128200, AY128241, EU002265, EU002457, EU002547, EU002368; *M. W. Chase* 250 (K). Nitrariaceae: *Nitraria retusa* Asch.—, EU002155, —, EU002185, U39278, EU002246, EU002436, EU002526, EU002348; *M. W. Chase* 597 (K).

Rutaceae: *Citrus sinensis* Osbeck.—AY177420 [*C. aurantium* L.], NC_008334, NC_008334, NC_008334, NC_008334, NC_008334, NC_008334. Sapindaceae: *Cupaniopsis anacardioides* Radlk.—AF206896, AF479139, AF035903, AY724283, L13182, EU002218, EU002402, EU002494, EU002314; *M. W. Chase* 217 (K). Simaroubaceae: *Ailanthus altissima* (Mill.) Swingle.—AF206842, —, AF035895, AY128208, AY128247, EU002194, EU002378, EU002467, EU002287; *M. W. Chase* 126 (K).

Unplaced in eurosid II

Tapisciaceae: *Tapiscia sinensis* Oliver.—AF207034, AF479146, AF09685, EU002190, —, EU002266, EU002458, EU002548, EU002369; *M. W. Chase* 1201 (K).

Unplaced in rosids in APGII

Aphloiaceae: *Aphloia theiformis* (Vahl) Benn.—AF206851, AF479132, AF209528, —, AF206735, —, EU002472, EU002291; *C. C. Davis* 6-01 (CCD 387). Ixerbaeae: *Ixerba brexioides* A. Cunn.—AF084476, —, AF209606, EU002181, —, EU002235, EU002423, EU002513, EU002335; Small s.n. (WS). Picramniaceae: *Picramnia pentandra* Sw.—, —, AJ235559, —, AF127025 [*P. polyantha* Planch.], EU002251, EU002441, EU002531, EU002286; *M. W. Chase* 2571 (K). Strasburgeriaceae: *Strasburgeria robusta* (Vieill. ex Panch. & Sebert) Guillaumin.—AF502596, —, AF502597, —, AJ403007, EU002262, EU002454, EU002544, EU002365; *Pillon Y. et al.* 60 (K).

Vitaceae

Saxifragales
Cercidiphyllaceae: Cercidiphyllum japonicum Siebold & Zucc.—U42518, AF274639, AF092112, AF274608, AF274596, EF207460, EF207484, EF207509, EF207534.


Hamamelis japonica Siebold & Zucc.—AF094551 (H. virginiana L.), AF036495 (H. virginiana L.), AF093380 (H. virginiana L.), AF274617 (H. virginiana L.), L01922 (H. mollis Oliv.), EF207467, EF207491, EF207516, EF207541.

OUTGROUPS
Apiales
Araliaceae: Panax ginseng C. A. Mey—D83275, AF479193 (P. quinquefolius L.), NC_006290, NC_006290, NC_006290, NC_006290, NC_006290, NC_006290.

Caryophyllales
Plumbaginaceae: Plumbago auriculata Lam.—U42795, AF036492, EU002166, EU002187, EU002283, EU002252, EU002442, EU002532, EU002354; M. J. Moore 306 (FLAS).

Cornales
Cornaceae: Cornus florida Hook.—X17370, AF297532, EU002157, EU002175, EU002276, EU002215, EU002377, EU002491, EU002311; M. J. Moore 328 (FLAS).

Berberidopsidales
Flacourtiaceae: Berberidopsis corallina Hook. f.—AF206866, AF389242, EU002158, EU002171, EU002274, EU002201, EU002385, EU002466, EU002295; M. J. Moore 326 (FLAS).

Gunnerales
Gunneraceae: Gunnera manicata Linden.—U43787, AF389250, EU002162, EU002179, EU002279, EU002226, EU002413, EU002504, EU002325; M. J. Moore 325 (FLAS).

Proteales
Platanaceae: Platanus occidentalis L. —U42794, AF274662, NC_008335, NC_008335, NC_008335, NC_008335, NC_008335, NC_008335, NC_008335, NC_008335.
Solanales
Solanaceae: *Nicotiana tabacum* L. — AJ236016, AF479172, NC_001879, NC_001879, NC_001879, NC_001879, NC_001879, NC_001879, NC_001879, NC_001879.

Trochodendrales
Trochodendraceae: *Trochodendron aralioides* Siebold & Zucc.— U42816, AF274671, EU002169, —, L01958, EU002269, EU002461, EU002551, EU002372; Univ. of Washington Arboretum accession # 581-4C.
Fig. S1. Phylogram of 1 of 9 trees resulting from parsimony analysis of 12-gene dataset (2 nuclear and 10 plastid) and LRT for 117 members of the rosid clade and outgroups (length 79,613, consistency index = 0.369, retention index = 0.433). Arrows designate nodes that collapse in the strict consensus. Numbers above branches are bootstrap values.
Fig. S2. Phylogenetic analyses of IR dataset for 59 taxa of the rosid clade and outgroups. (A) Phylogram of 1 of 4 most parsimonious trees. Bootstrap values are shown above branches. Arrows indicate nodes that collapse in the strict consensus. (B) Phylogram of ML tree.
Fig. S3. Phylogenetic analyses of the 12-gene dataset (2 nuclear and 10 plastid) for 59 members of the rosid clade and outgroups. (A) Phylogram of the single tree resulting from MP analysis (length 34,843, consistency index = 0.374, retention index = 0.368). Numbers above branches are bootstrap values. (B) Phylogram of the ML tree.
Fig. S4. Phylogenetic analyses of the total evidence dataset (2 nuclear genes, 10 plastid genes, and the IR) for 59 members of the rosid clade and outgroups. (A) Phylogram of the single tree resulting from MP analysis (length 53,418, consistency index = 0.476, retention index = 0.410). Numbers above branches are bootstrap values. (B) Phylogram of the ML tree.
Fig. S5. Phylogenetic analyses of the 12-gene dataset (2 nuclear and 10 plastid) for 117 members of the rosid clade and outgroups. (A) Phylogram of 1 of 3 trees resulting from MP analysis (length 61,042, consistency index = 0.278, retention index = 0.417). Arrows designate nodes that collapse in the strict consensus. Numbers above branches are bootstrap values. (B) Phylogram of the ML tree.
Table S1. MP and ML tree statistics for the different data sets employed in this study

<table>
<thead>
<tr>
<th>Dataset (No. of Taxa)</th>
<th>No. of MP Trees</th>
<th>MP tree length</th>
<th>CI</th>
<th>RI</th>
<th>GARLI ML score (MP score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Evidence (117)</td>
<td>9</td>
<td>79613</td>
<td>0.369</td>
<td>0.433</td>
<td>468184.55 (79700)</td>
</tr>
<tr>
<td>Total Evidence (59)</td>
<td>1</td>
<td>53418</td>
<td>0.476</td>
<td>0.410</td>
<td>341883.36 (53460)</td>
</tr>
<tr>
<td>Plastid + Nuclear (117)</td>
<td>3</td>
<td>61042</td>
<td>0.278</td>
<td>0.417</td>
<td>316305.06 (61186)</td>
</tr>
<tr>
<td>Plastid + Nuclear (59)</td>
<td>1</td>
<td>34843</td>
<td>0.374</td>
<td>0.368</td>
<td>190348.63 (34901)</td>
</tr>
<tr>
<td>Inverted Repeat (59)</td>
<td>4</td>
<td>18525</td>
<td>0.671</td>
<td>0.526</td>
<td>146570.57 (18559)</td>
</tr>
<tr>
<td>Nuclear (117)</td>
<td>35,466</td>
<td>5398</td>
<td>0.331</td>
<td>0.343</td>
<td>34291.19 (5476)</td>
</tr>
</tbody>
</table>

CI, consistency index; RI, retention index.
Table S2. Primers used in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>26S rDNA</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>matK</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>ndhF</td>
<td>1F, 274F, 536F, 803F, 972F, 1318F; 536R, 803R, 972R, 1318R, 1603R, 2110R</td>
<td>CCC TTC ATT CCR CTT CCA GTT CCT</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>2153R</td>
<td>GGA ATT CCA TCA ATK ATT CGT YTA TC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td>rpoC2</td>
<td>80F</td>
<td>TAG ATC ACT TCG GAA TGG C</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>750F</td>
<td>GGT CGT GTA TTA GCR GAC G</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1000F</td>
<td>GGT ATT ATT GGG GGT CAA TC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1400F</td>
<td>AGG GGG ARA TGG CAY TGG AG</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1950F</td>
<td>GAT CCT CGA TAC AGA AGA AAG AGT C</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>2100F</td>
<td>TCA TTC CGG AGG AAG ATC AT</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>2600F</td>
<td>CAA GTA TCC AAT TAG TTC GGA CTT G</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>3000F</td>
<td>CAT CTA ATT GTT TTC GAA TGG GTC C</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>3400F</td>
<td>CAT TYA ATT TGA ATT GAT TTCC TSC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1000R</td>
<td>GAT TGA CCC GCA ATA ATC CC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1400R</td>
<td>CTC CAR TGC CAT YTC YCC CT</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>1950R</td>
<td>GAA CTC TTT CTT CTG TAT CGA GGA TC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>2100R</td>
<td>ATG YAC TTC CTC GGG GAT GA</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>2600R</td>
<td>CAA GTC CGA ACT AAT TGA ATA CTG G</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>3000R</td>
<td>GGA CCC ATT CGA AAA CAA TTA GAT G</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>3400R</td>
<td>GGA GAA AAT ACC AAT TCA AAT TYA ATG</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>4200R</td>
<td>TTT CAG GCC TTT YAR CCA RTC</td>
<td>Designed for this study</td>
</tr>
<tr>
<td>rps4</td>
<td>1F, trnR</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>48F</td>
<td>GGC TTT ACC RGG ACT AAC</td>
<td>Designed for this study</td>
</tr>
</tbody>
</table>

Primer sequences are given only for primers that were designed specifically for this study.

Table S3. Gene/partition characteristics

<table>
<thead>
<tr>
<th>Gene/partition</th>
<th>Number of missing taxa</th>
<th>Total aligned length</th>
<th>Analyzed aligned length</th>
<th>Number of parsimony-informative characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S rDNA</td>
<td>12</td>
<td>1832</td>
<td>1707</td>
<td>244</td>
</tr>
<tr>
<td>26S rDNA</td>
<td>39</td>
<td>3477</td>
<td>3213</td>
<td>526</td>
</tr>
<tr>
<td>total nuclear</td>
<td>9</td>
<td>5309</td>
<td>4920</td>
<td>770</td>
</tr>
<tr>
<td>atpB</td>
<td>9</td>
<td>1528</td>
<td>1431</td>
<td>527</td>
</tr>
<tr>
<td>matK</td>
<td>23</td>
<td>1845</td>
<td>1632</td>
<td>1,066</td>
</tr>
<tr>
<td>ndhF</td>
<td>8</td>
<td>2532</td>
<td>2124</td>
<td>1,221</td>
</tr>
<tr>
<td>psbB</td>
<td>9</td>
<td>1530</td>
<td>1464</td>
<td>486</td>
</tr>
<tr>
<td>psbH</td>
<td>9</td>
<td>237</td>
<td>141</td>
<td>54</td>
</tr>
<tr>
<td>psbN</td>
<td>9</td>
<td>132</td>
<td>129</td>
<td>38</td>
</tr>
<tr>
<td>psbT</td>
<td>9</td>
<td>129</td>
<td>111</td>
<td>28</td>
</tr>
<tr>
<td>rbcL</td>
<td>0</td>
<td>1434</td>
<td>1395</td>
<td>495</td>
</tr>
<tr>
<td>rpoC2</td>
<td>7</td>
<td>5037</td>
<td>4143</td>
<td>2,302</td>
</tr>
<tr>
<td>rps4</td>
<td>7</td>
<td>810</td>
<td>591</td>
<td>292</td>
</tr>
<tr>
<td>IR</td>
<td>58</td>
<td>25549</td>
<td>25549</td>
<td>4,300</td>
</tr>
<tr>
<td>total plastid targeted genes (no IR)</td>
<td>0</td>
<td>40763</td>
<td>38710</td>
<td>6,509</td>
</tr>
<tr>
<td>total targeted genes (no IR)</td>
<td>0</td>
<td>20523</td>
<td>18081</td>
<td>7,279</td>
</tr>
<tr>
<td>total evidence</td>
<td>0</td>
<td>46072</td>
<td>43630</td>
<td>11,579</td>
</tr>
</tbody>
</table>

Number of missing taxa refers to the number of taxa in the total evidence data set that lack sequence for a given gene or partition; analyzed aligned length refers to the total aligned length minus excluded characters. The psbB/psbT/psbN/psbH and rpoC2/trnV spacer regions were sequenced but were not included in analyses, and so these regions are excluded from the numbers in this table.
Information on fossils used in divergence time estimation. The assigned minimum ages were used as the zero offset in the BEAST analysis. UCLN-uniform corresponds to the BEAST analysis that treated fossils as being drawn from a uniform distribution \((\text{minimum, maximum})\); UCLN-lognormal corresponds to the BEAST analysis that treated fossils as being drawn from a lognormal distribution \((\text{mean, standard deviation})\). We treated the age of the root node as follows: For PL analyses, we treated it as (i) a fixed age calibration (at 125 million years, based on the Barremian fossil record of tricolpate pollen indicative of eudicots), and (ii) as a maximum age constraint of 125 million years. All other fossils were treated as minimum age constraints. For the UCLN analyses, we treated distribution of the age of the root node as a uniform distribution between 90 (the oldest fossil used in our analyses: \(\text{Paleoclusia}\), placed as a constraint relative to \(\text{Clusia}/\text{Hypericum}\)) and 125 million years \((U(90,125))\). The UCLN model allows uncertainty in the age of calibrations to be represented as prior distributions rather than as strict calibration/fixed points. We therefore constrained several of the nodes in the tree to prior probability distributions. Fossils were treated as a uniform distribution between the first occurrence of the fossil and the maximum age of the clade (i.e., 125 million years). In addition, MCMC runs were performed in which we modeled the fossil priors as lognormal distribution. This explicitly assumes that the actual divergence event is more likely to have occurred at some time prior to the minimum age of the fossil. In each case the “zero offset” value in BEAST was set to the minimum age of the fossil. The selection of the mean and standard deviation is somewhat subjective. For our fossils, we set the mean to zero and the standard deviation to one, to insure that 95% of the distribution fell within the maximum set age of the clade.


Other Supporting Information

*SI Appendix*