

Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots

Michael J. Moore^{a,1}, Pamela S. Soltis^b, Charles D. Bell^c, J. Gordon Burleigh^d, and Douglas E. Soltis^d

^aBiology Department, Oberlin College, Oberlin, OH 44074; ^bFlorida Museum of Natural History, University of Florida, Gainesville, FL 32611; ^cDepartment of Biological Sciences, University of New Orleans, New Orleans, LA 70148; and ^dDepartment of Biology, University of Florida, Gainesville, FL 32611

Edited* by Michael J. Donoghue, Yale University, New Haven, CT, and approved January 26, 2010 (received for review July 14, 2009)

Although Pentapetalae (comprising all core eudicots except Gunnerales) include $\approx 70\%$ of all angiosperms, the origin of and relationships among the major lineages of this clade have remained largely unresolved. Phylogenetic analyses of 83 protein-coding and rRNA genes from the plastid genome for 86 species of seed plants, including new sequences from 25 eudicots, indicate that soon after its origin, Pentapetalae diverged into three clades: (i) a “superrosid” clade consisting of Rosidae, Vitaceae, and Saxifragales; (ii) a “superasterid” clade consisting of Berberidopsidales, Santalales, Caryophyllales, and Asteridae; and (iii) Dilleniaceae. Maximum-likelihood analyses support the position of Dilleniaceae as sister to superrosids, but topology tests did not reject alternative positions of Dilleniaceae as sister to Asteridae or all remaining Pentapetalae. Molecular dating analyses suggest that the major lineages within both superrosids and superasterids arose in as little as 5 million years. This phylogenetic hypothesis provides a crucial historical framework for future studies aimed at elucidating the underlying causes of the morphological and species diversity in Pentapetalae.

Angiosperm Tree of Life | Pentapetalae | plastid genome

The Eudicotyledoneae (sensu (1), or eudicots, comprise $\approx 75\%$ of all angiosperm species (2) and encompass enormous morphological, biochemical, and ecological diversity. More than 90% of eudicot species diversity is found within the clade Pentapetalae (1), which includes major clades such as Rosidae, Caryophyllales, Saxifragales, Asteridae, and Santalales, as well as smaller lineages such as Berberidopsidales and Dilleniaceae (3–8). Previous analyses of multigene data sets have failed to resolve relationships among the major clades of Pentapetalae (6, 9). The inability to resolve these relationships suggests that the major lineages of Pentapetalae diverged rapidly, a hypothesis supported by the fossil record (10, 11). However, our understanding of the origins and evolution of Pentapetalae diversity, and consequently much of angiosperm diversity, is hindered by the lack of a well-supported phylogenetic hypothesis.

Phylogenetic analyses based on complete plastid genome sequences have resolved several enigmatic relationships within angiosperms (12, 13). However, these analyses have not included data for many crucial eudicot clades. To resolve relationships among the major clades of Eudicotyledoneae (with a focus on Pentapetalae), we performed phylogenetic analyses using a data set composed of 83 genes derived from 86 complete plastid genome sequences, 25 of which were eudicot sequences generated for this study. To date, this is the largest plastid genome data set used for phylogenetic inference and includes representatives of nearly all (37 of 42) orders of eudicots sensu Angiosperm Phylogeny Group (APG) III (14). The resulting phylogenetic hypothesis helps to clarify the diversification of Pentapetalae and provides an improved framework for investigating evolutionary processes that accompanied this radiation.

Results

Phylogenetic Analyses. Maximum-likelihood (ML) analyses of the 83-gene alignment yielded similar trees and bootstrap support

(BS) values regardless of the partitioning strategy (Fig. 1 and Figs. S1–S3). The ML topologies were also similar to the maximum parsimony (MP) topology (Fig. S4). Relationships among early-diverging angiosperms and among many clades of Mesangiospermae agree with those from the largest previous plastid genome analyses (12, 15). Below, we focus on results regarding relationships within the eudicots, with an emphasis on the ML topologies.

Within Eudicotyledoneae, a basal grade emerged, with most nodes receiving 100% ML BS support. In all ML analyses, Ranunculales were sister to all remaining eudicots (BS = 100%), followed by a clade of Sabiaceae + Proteales (BS = 70–80%, depending on partitioning strategy) as sister to a strongly supported clade (BS = 100%) of Trochodendraceae + Buxaceae + Gunneridae (Gunneridae = Gunnerales + Pentapetalae; Fig. 1 and Figs. S1–S3). However, the approximately unbiased (AU) test did not reject alternative relationships among Sabiaceae, Proteales, and remaining eudicots (Table S1). In all but one ML analysis, Buxaceae were recovered as sister to Gunneridae with BS = 52–70%, depending on partitioning strategy (Fig. 1 and Figs. S1 and S2). In the CodonPartBL partition, Trochodendraceae were sister to Gunneridae, but with less than 50% bootstrap support (Fig. S2).

Within the strongly supported (BS = 100%) Gunneridae, Gunnerales were sister to Pentapetalae (BS = 100%). Pentapetalae comprised three major clades in the ML trees: (i) a “superrosid” clade (BS = 96–100%) of Saxifragales, Vitaceae, and Rosidae; (ii) a “superasterid” clade (BS = 92–100%) of Santalales, Berberidopsidales, Caryophyllales, and Asteridae; and (iii) Dilleniaceae (Fig. 1 and Figs. S1–S3). The superrosid and superasterid clades were also recovered in the MP trees (Fig. S4), with one major difference in composition. Whereas ML placed *Dillenia* (the single representative of Dilleniaceae in the data set) sister to superrosids (Fig. 1 and Figs. S1–S3), MP placed *Dillenia* sister to Caryophyllales, with 97% BS support (Fig. S4). Still, the MP analysis also supported a superrosid clade of Vitaceae + Saxifragales + Rosidae with 99% BS. When *Dillenia* was constrained to be sister to the superrosid clade, the best MP tree was 44 steps longer than the best unconstrained tree. A Templeton test (16) failed to reject this alternative topology ($n = 712$; $P = 0.099$). For ML, AU tests strongly rejected topologies in which *Dillenia* was sister to Caryophyllales (Table S1); however, they did not reject the placement of *Dillenia* as sister to all remaining Pentapetalae or as sister to the superasterids.

Parametric bootstrapping provided evidence of bias in the MP analyses against placing *Dillenia* as sister to the superrosid clade.

Author contributions: P.S.S. and D.E.S. designed research; M.J.M. performed research; M.J.M., C.D.B., J.G.B., and D.E.S. analyzed data; and M.J.M., P.S.S., C.D.B., J.G.B., and D.E.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data Deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GQ996966–GQ998871).

¹To whom correspondence should be addressed. E-mail: michael.moore@oberlin.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907801107/DCSupplemental.

trees, *Bulnesia* formed a clade with all other Fabidae. BS for the *Bulnesia* + Fabidae clade was at least 98% except in the CodonGenePartBL analyses, in which BS = 52% (Fig. 1 and Figs. S1–S3). The MP topology was strongly rejected by an AU test ($P < 0.001$), as was an alternative topology in which *Bulnesia* was constrained to be sister to Malvidae.

Within the superasterids, the ML analyses placed Santalales, Berberidopsidales, and Caryophyllales as successive sisters to Asteridae, with each node receiving 99–100% BS support, except for the clade of Caryophyllales + Asteridae (BS = 77–88% depending on partitioning strategy; Fig. 1 and Figs. S1–S3). Some relationships within Asteridae, such as the positions of *Lonicera* and *Cornus*, were in strongly supported but incongruent positions in MP and ML trees (Fig. 1 and Fig. S4). However, MP placements for these taxa were rejected by AU tests (Table S1). Finally, the relationships among the basal clades of Lamiidae (represented in our data set by Boraginales, Gentianales, Lamiales, and Solanales) were generally supported by BS values less than 55% with ML (Fig. 1 and Figs. S1–S3).

Divergence Time Estimates. The crown group 95% highest posterior density (HPD) age estimates for the major lineages of Pentapetalae were as follows: superasterids (107–98 mya), Dilleniaceae + superrosids (112–103 mya), superrosids (111–103 mya), Santalales (91–99 mya), Caryophyllales (71–63 mya), Asteridae (89–80 mya), Rosidae (110–103 mya), and Saxifragales (103–94 mya) (Fig. S5 and Table S2). The split between Vitaceae and Saxifragales was dated to 112–101 mya (Fig. S5 and Table S2).

Discussion

Phylogeny of Eudicots. Analyses of complete plastid genome sequences help to resolve the relationships among the major lineages of Pentapetalae with the strongest support yet obtained, providing a unique perspective on the origin and early evolution of a large proportion of angiosperm species diversity. Soon after the split between Pentapetalae and Gunnerales (Fig. 1), Pentapetalae diverged into three major subclades: (i) a superrosid clade of Saxifragales, Vitaceae, Rosidae; (ii) a superasterid clade of Berberidopsidales, Santalales, Caryophyllales, and Asteridae; and (iii) Dilleniaceae (Fig. 1). This finding differs from previous molecular phylogenetic analyses in providing strong support for these clades. Although Berberidopsidales, Santalales, and Caryophyllales have often been placed with Asteridae, and Saxifragales with Rosidae and Vitaceae, these relationships were only weakly supported (3, 5, 6, 8, 9, 17, 18).

All ML analyses support the position of Dilleniaceae as sister to superrosids; however, we cannot reject alternate hypotheses (Table S1). Recent multigene analyses (MP, ML, or Bayesian) with greater taxon sampling, although far fewer genes, have placed Dilleniaceae sister to Caryophyllales, although again without strong support (5, 9, 17). Our AU tests reject this hypothesis, and our parametric bootstrapping experiment suggests that, at least in MP analyses, the placement of Dilleniaceae with Caryophyllales may be the result of systematic error. Instead, our data suggest a position for Dilleniaceae nearer the base of Pentapetalae, but additional taxon and genomic sampling will be necessary to place Dilleniaceae definitively.

The close relationship of Saxifragales, Vitaceae, and Rosidae is perhaps not surprising considering that these clades formed the core of Cronquist's (19) and Takhtajan's (20) Rosidae. All have articulated anthers (21, 22), stipules (22), nuclear endosperm development, and micropyle formation that involves the outer or both integuments (21). Still, relationships among these lineages remain unclear. Most previous molecular analyses have placed Vitaceae sister to Rosidae (with these groups comprising the rosid clade sensu APG III) (14), although often without strong support (5, 17, 23). Vitaceae and Rosidae are maintained as distinct by

Stevens (22), which we follow here. Our ML analyses weakly support Rosidae as sister to a clade of Vitaceae + Saxifragales (Fig. 1), but we cannot reject any of the alternative relationships among Saxifragales, Vitaceae, and Rosidae. Thus, further genomic and taxon sampling will likely be necessary to resolve the relationships among these three clades with confidence.

Within Rosidae, relationships among major clades were almost entirely congruent with those obtained from a recent analysis of 12 targeted genes and the chloroplast inverted repeat for 104 species of rosids by Wang et al. (23). The only difference was the position of Celastrales vs. Oxalidales and Malpighiales, but the relationships among these clades have long been difficult to reconstruct (reviewed in ref. 23). The overall congruence between this study and Wang et al. (23) strengthens our confidence in basal Rosidae relationships and suggests that increased taxon sampling would not affect our findings for this clade.

The plastid analyses provide strong support for the superasterid clade as well as the basal split within superasterids (Fig. 1). Previous molecular analyses have linked Santalales, Caryophyllales, Berberidopsidales, and Asteridae, but the relationships among them have never received strong support (5, 17). Several putative morphological synapomorphies may characterize the superasterid clade. Psilate or granulate pollen exine structure and the absence of craspedodromous venation may be synapomorphies of the superasterids (21). However, the latter character appears to be evolutionarily labile, and it is also widespread in Rosidae. The presence of leaf sclereids, isomery in the androecium (this trait also unites Rosidae as an independent synapomorphy), and fused carpels and styles unites Santalales, Caryophyllales, and Asteridae (21). Likewise, the absence of stipules may be a synapomorphy for Santalales, Caryophyllales, and Asteridae (Berberidopsidaceae have stipules, but Aextoxicaceae lack stipules—hence the ancestral state for Berberidopsidales, and therefore the superasterids, is unclear). To understand whether these and other non-DNA characters are in fact synapomorphies will require both more extensive morphological analyses and a tree with denser taxon sampling, but our findings bring us significantly closer to realizing the phylogenetic framework necessary for the detailed study of character evolution in eudicots.

Rapid Diversification Within Eudicotyledoneae. Our molecular dating analyses imply that the major subclades of Pentapetalae likely diversified rapidly. Our analyses suggest that the initial splits among Dilleniaceae, superrosids, and superasterids may have occurred as soon as 1 million years following the divergence of Pentapetalae from Gunnerales (Fig. S5 and Table S2). Furthermore, both superrosids and superasterids display early and rapid diversifications, with the lineages leading to Vitaceae + Saxifragales and Rosidae arising within a window of only about 5 my, and the Berberidopsidales, Caryophyllales, and Asteridae lineages splitting in a similarly short timeframe. Furthermore, our inability to resolve either the diversification within Lamiidae or the diversification among Trochodendrales, Buxales, and Gunneridae, even with >66,000 bp, suggests that these lineages may have also diverged very rapidly (Fig. S5).

Other recent studies employing character-rich plastid DNA data sets have inferred rapid radiations in other major groups of angiosperms, including the basal lineages of Mesangiospermae (12), Rosidae (23), and Saxifragales (24). Our results provide additional examples of rapid diversification within angiosperms and are consistent with earlier suggestions that angiosperm evolution has been characterized by a series of nested, rapid radiations (25–27).

Sources of Error in Phylogenetic and Dating Analyses. Although plastid genome data sets have shown unprecedented power to resolve difficult phylogenetic questions across angiosperms and

date the divergences of major lineages (12, 13, 15, 28), analyses of genome-scale phylogenetic matrices are still susceptible to error, which may be masked by high BS values. For example, using models that fail to account for the heterogeneous patterns of sequence evolution that are inherent in such large data sets can mislead ML analyses (29). Although it is difficult to demonstrate such model misspecification errors, we obtained very similar topologies and support values in ML analyses across all partitioning schemes, suggesting that the analyses are robust to gene-specific and codon position-specific heterogeneity in molecular evolution. Furthermore, when we simulated plastid genome data sets, allowing for different substitution model parameters and branch lengths for each gene, the unpartitioned ML analysis of the simulated data recovered the “true” (simulated) topology.

Error may also arise from inadequate taxon sampling. Matrices with many characters and few taxa may result in incorrect topologies with high support (13, 30), and plastid genome-scale analyses have been particularly susceptible to such errors (31). Although we find evidence that systematic error may affect the placement of *Dillenia* in the MP analysis, this study is the largest whole-plastid genome analysis to date, with taxon sampling designed to mitigate as much as possible against long-branch attraction and related systematic errors. Thus, our analysis likely does not suffer from the errors that plagued several previous plastid genome analyses. Still, we cannot rule out possible changes in topology or support if additional taxa were included in our 83-gene data set.

We also examined if any of our results might be driven by a small number of genes. Conventional MP and ML bootstrapping analyses estimate the sampling variance among characters, or sites in the alignment, implicitly assuming that the different genes compose a homogeneous sample. We performed a two-stage bootstrapping experiment (sampling 83 genes with replacement and then sampling sites within the sampled genes with replacement) to account for sampling variance both among genes and among sites within genes (32). The support values for both MP and ML two-stage bootstrapping were very similar to the conventional bootstrap support values (Figs. 1 and Figs. S1–S4), suggesting that there is not extensive sampling variance among genes.

During periods of rapid cladogenesis, incomplete lineage sorting may produce incongruence among gene tree and species tree topologies (33). In such cases, the gene tree topologies might not reflect the species phylogeny. Incongruence among gene trees may provide evidence of incomplete lineage sorting, but given that the plastid genome represents a single, nonrecombining chromosome, we would not expect to detect evidence of incomplete lineage sorting from the plastid genome alone. Finding topological congruence between multiple unlinked nuclear loci and our plastid tree would provide strong independent evidence in support of the plastid ML topology. Future analyses of genome-scale nuclear data sets will enable such phylogenetic comparisons.

Although our divergence time estimates (Fig. S5 and Table S2) are similar to those of several other recent dating analyses (12, 34, 35), we stress that estimates of diversification dates are also susceptible to error, even when based on large data sets. Perhaps the most important potential source of error involves the use of fossil constraints and taxon sampling (34–37). Age estimates for individual clades in a molecular dating analysis can be strongly affected by the number of independent fossil constraints employed, as well as their proper assignment to nodes in the tree being analyzed, which in turn depends heavily on taxon sampling (11, 35, 38, 39). Because our taxon sampling was limited by the availability of plastid genomes, the number of fossil constraints that we were able to employ was similarly limited. Consequently, our age estimates should be interpreted with caution, and we do not suggest that they supersede those of other recent analyses that included greater taxon sampling and many more fossil constraints (11, 40, 41). In fact, for a given clade,

different ages inferred in our analyses and those of others may be due in large part to differences in taxon sampling and constraints. For example, the age of Dipsacales, in Campanulidae, was estimated to be >100 my in a recent analysis of 30 taxa of Dipsacales (42), whereas in our analyses, the crown group Campanulidae itself dates back only ~75 mya. Nevertheless, our analyses further show the rapidity with which many groups of eudicots, and in particular basal Pentapetalae, appear to have diverged.

Conclusion

Phylogenetic analyses of complete plastid genome sequences provide much-improved confidence in the relationships among major lineages of Pentapetalae and also provide a framework with which to investigate the evolutionary processes that produced a large proportion of extant angiosperm diversity. In light of these phylogenetic results, we are now challenged to identify characters that are unique to the superasterid or superrosid clades and those that arose in parallel in the two clades, and then explore their evolutionary implications. Understanding the morphological evolution of Pentapetalae more fully will require both careful morphological studies and rigorous reconstructions of morphology using this and future well-supported phylogenetic hypotheses. This improved phylogenetic framework can also be used to test whether specific non-DNA characters are linked to diversification and to assess whether these characters involve the same or independently co-opted underlying genetic mechanisms.

Materials and Methods

Gene/Taxon Sampling and DNA Sequence Alignment. The character matrix for all phylogenetic analyses consisted of the nucleotide sequences of all 79 protein-coding genes and all four ribosomal RNA genes (Table S3) that are known from angiosperm plastid genomes (43). To produce this data set, we modified the 81-gene, 64-taxon alignment of Jansen et al. (15) by adding portions of two genes (*accD* and *ycf1*) and new sequences from 29 species, 25 of which were eudicots we sequenced for this study, and by removing sequences from the well-sampled Poaceae and Solanaceae (Table S4). Gene sequences for the 25 newly included eudicots were derived from complete plastid genome sequences and are available in GenBank. These plastid genomes were sequenced using both the Genome Sequencer 20 and FLX Systems (454 Life Sciences Corp.) following Moore et al. (12, 44). Genomes were also assembled, corrected, and annotated following Moore et al. (12, 44).

The addition of new sequences required manual realignment of some genes. Several short regions of the more quickly evolving genes (e.g., *matK*, *ndhF*, and *rpoC2*) were difficult to align and were therefore excluded from analyses, as were all sequence insertions present in only one or a few taxa. The 3' portion of *ycf1* that is typically located in the small single-copy region was also excluded due to alignment difficulties. The length of the 83-gene alignment used for all analyses was 66,741 bp (Table S3). The total amount of missing data, after excluding ambiguously aligned regions, was 4.1%, caused primarily by the lack of specific genes in some taxa (e.g., *infA* is missing in many Rosidae; some or all *ndh* genes are absent in parasitic taxa). The aligned data set is available from the authors on request.

Phylogenetic Analyses. Both MP and ML analyses were performed on the concatenated 83-gene data set, using a number of different partitioning strategies for ML analyses. First, we performed an unpartitioned ML analysis, which estimated a single set of nucleotide substitution model and branch length parameters for all characters across the 83-gene alignment. Next, we performed two analyses that partitioned the data based on gene region. In the first analysis, we partitioned the data set by gene and estimated substitution model parameters for each gene while maintaining a single set of branch lengths across all genes, whereas in the second analysis we estimated both substitution parameters and branch lengths for each gene.

We also performed four ML analyses that partitioned protein-coding genes by codon position. The CodonPart analysis created a partition for each of the three codon positions and estimated substitution rate matrix parameters separately for each partition. The CodonPartBL analysis used the same partitioning scheme, but it additionally estimated separate branch lengths for each codon partition. Finally, the CodonGenePart analysis estimated substitution model parameters separately for each codon position in each gene, and the CodonGenePartBL analysis additionally estimated separate branch lengths for each codon position in each gene. Both the

CodonGenePart and CodonGenePartBL analyses required us to recompile RAXML to enable it to handle the large number of parameters.

Analyses proceeded as follows. For MP, we used PAUP* version 4.0b10 (45) to implement heuristic searches that consisted of TBR branch swapping, starting from 1,000 trees built by random taxon stepwise addition, and saving all optimal trees. To assess uncertainty in the MP topology, we performed 200 nonparametric BS replicates, each consisting of TBR branch swapping from 100 random taxon addition starting trees, saving all optimal trees from each replicate. ML analyses were implemented in RAXML version 7.0.2 (46) and included 10 runs to find an optimal tree and 200 nonparametric BS replicates (47). All ML analyses used the general time-reversible (GTR) (48) model of evolution with among-site rate variation modeled using the CAT discrete rate categories option in RAXML. For all ML bootstrap analyses, we first generated bootstrap data sets using HYPHY version 0.9920061107beta for Linux and version 1.0020080508beta for Mac (49) and performed a single run on each bootstrap data set in RAXML following the same protocol as used for the original empirical data set. For the unpartitioned analysis, we obtained the bootstrap data sets by sampling with replacement the characters from the entire alignment. For each replicate in the two partitioned analyses, we created a bootstrap data set for each gene separately and then concatenated the single-gene bootstrap data sets into a single alignment.

To account for any potential gene-specific effects on the phylogenetic inference, we also performed a two-stage bootstrap analysis (32). By resampling with replacement genes and nucleotides within the genes, the two-stage bootstrap procedure seeks to estimate sampling variance from the genes as well as the nucleotides that compose each gene. Note that because the gene alignments vary in length, the two-stage bootstrap alignments also vary among replicates. We created 200 two-stage bootstrap data sets using a series of Perl scripts (available upon request) and HYPHY, which was again used to create single-gene bootstrap data sets. MP and unpartitioned ML analyses were performed on the two-stage bootstrap data sets using the protocols that were used in the conventional bootstrap analyses.

Parametric Bootstrapping. Parametric bootstrapping (simulation) can be an effective tool for revealing biases associated with particular phylogenetic methods (50). We performed a parametric bootstrap experiment to examine the different placement of *Dillenia* in the optimal ML and MP trees. Specifically, we asked if the optimal MP or ML tree were true, how often would we expect the MP or ML analyses to place *Dillenia* correctly? For both the optimal MP and unpartitioned ML topologies, we simulated 200 replicate data sets using HYPHY. To do this, for each gene, we first estimated the nucleotide substitution parameters and branch lengths using the GTR substitution model with rate variation among sites modeled using a discrete gamma distribution with four rate categories (51), and then simulated a replicate data set using these parameters. After performing simulations for each gene, we concatenated the simulated single-gene data sets to make a single alignment. Thus, the simulated data sets account for heterogeneous processes of evolution among genes. Also, because there are many differences in taxon sampling among genes due to gene loss, the single-gene simulations reflect the true patterns of missing data in the empirical matrix. After the simulations, we performed both MP and unpartitioned ML analyses on each simulated data set using the same protocols as the original analyses. Based on the parametric bootstrap analyses, we estimated how often we would expect *Dillenia* to be placed as sister to the superrosid clade

or sister to the Caryophyllales in MP and ML analyses when the MP topology is true and the ML topology is true.

AU Tests. To assess whether certain alternative relationships among clades of eudicots could be statistically rejected, we simultaneously performed AU tests (52) for the best ML topology and 34 alternative topologies using CONSEL version 0.1i (53). All alternative topologies tested involved areas of eudicot phylogeny where BS values were weak in the ML tree and/or where alternative relationships have been suggested in the past (e.g., all 14 possible alternative topologies were tested for the four major clades of Lamiidae, and three alternative positions for *Dillenia* were tested). A complete list of alternative topologies tested is provided in Table S1. Because we cannot compare likelihoods for topologies calculated with CAT rate variation among sites (46), individual site likelihoods were estimated in RAXML under the GTR + Γ model using both an unpartitioned model and a partitioned model in which different substitution model parameters were estimated for each gene region.

Molecular Dating Analyses. Given the lack of rate constancy among lineages [based on a likelihood ratio test (54); $P < 0.001$ for all data sets], divergence times were estimated under a relaxed molecular clock. An uncorrelated log-normal (UCLN) model implemented in BEAST version 1.4.7 (55) was used to infer divergence times. We performed two BEAST analyses: one assumed a single common model of molecular evolution across all nucleotide positions, whereas the other was partitioned by gene, with separate rates and rate-change parameters for each partition. We used the GTR + Γ model of sequence evolution. For each analysis, we initiated four independent MCMC analyses from starting trees that included all ingroup and outgroup taxa, with branch lengths that satisfied the respective priors on divergence times. Convergence of each chain to the target distribution was assessed using Tracer version 1.4 (56) and by plotting time series of the log posterior probability of sampled parameter values. Post-MCMC analyses using Tracer suggested that all chains had reached stationarity and that there was convergence among the independent runs. All parameters and statistics had an effective sample size of greater than 200, as calculated with Tracer. After convergence was achieved, each chain was sampled every 1,000 steps until 50,000 samples were obtained. Model fit of the different UCLN-relaxed clock models was assessed using Bayes factors as implemented in Tracer version 1.4. Bayes factors overwhelmingly favored the model in which the data set was partitioned by gene. The BEAST XML file is available from the authors on request.

The estimation of absolute divergence times requires calibrating (or constraining) the age of one or more nodes. The UCLN model allows uncertainty in the age of calibrations to be represented as prior distributions rather than as strict calibration/fixing points. We therefore constrained several of the nodes in the tree to prior probability distributions based on fossil data; these constraints are discussed in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank the editor and three anonymous reviewers for their helpful comments on earlier versions of this manuscript. We also thank Bob Jansen and Jim Leebens-Mack for access to unpublished genome sequences, the University of Washington Arboretum for tissue of *Staphylea* and *Trochodendron*, Rob Ferl and Beth Laughner for laboratory space and general laboratory assistance, and the University of Florida Genetics Institute for access to the Fisher computer cluster. This study was carried out as part of the Angiosperm Tree of Life Project, National Science Foundation Grant EF-0431266 (to D.E.S. and P.S.S.).

- Cantino PD, et al. (2007) Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* 56:822–846.
- Drinnan AN, Crane PR, Hoot SB (1994) Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). *Plant Syst Evol* 8(Suppl): 93–122.
- Chase MW, et al. (1993) Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann Mo Bot Gard* 80:526–580.
- Soltis DE, et al. (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann Mo Bot Gard* 84:1–49.
- Soltis DE, et al. (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbcl*, and *atpB* sequences. *Bot J Linn Soc* 133:381–461.
- Soltis DE, et al. (2003) Gunnerales are sister to other core eudicots: Implications for the evolution of pentamerism. *Am J Bot* 90:461–470.
- Hoot SB, Magallon S, Crane PR (1999) Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcl*, and 18S nuclear ribosomal DNA sequences. *Ann Mo Bot Gard* 86:1–32.
- Savolainen V, et al. (2000) Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences. *Syst Biol* 49:306–362.
- Burleigh JG, Hilu KW, Soltis DE (2009) Inferring phylogenies with incomplete data sets: A 5-gene, 567-taxon analysis of angiosperms. *BMC Evol Biol* 9:61.
- Magallón-Puebla S, Crane PR, Herendeen PS (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Ann Mo Bot Gard* 86:297–372.
- Magallón S, Castillo A (2009) Angiosperm diversification through time. *Am J Bot* 96: 349–365.
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA* 104:19363–19368.
- Leebens-Mack J, et al. (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Mol Biol Evol* 22:1948–1963.
- Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161:105–121.
- Jansen RK, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104:19369–19374.
- Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* 37:221–244.

17. Soltis DE, Gitzendanner MA, Soltis PS (2007) A 567-taxon data set for angiosperms: The challenges posed by Bayesian analyses of large data sets. *Int J Plant Sci* 168:137–157.
18. Soltis PS, Soltis DE, Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402:402–404.
19. Cronquist A (1981) *An Integrated System of Classification of Flowering Plants* (Columbia Univ Press, New York).
20. Takhtajan AL (1997) *Diversity and Classification of Flowering Plants* (Columbia Univ Press, New York).
21. Nandi OI, Chase MW, Endress PK (1998) A combined cladistic analysis of angiosperms using *rbcl* and nonmolecular data sets. *Ann Mo Bot Gard* 85:137–212.
22. Stevens PF (2001) Angiosperm Phylogeny, Version 9. Available at <http://www.mobot.org/MOBOT/research/APweb/>.
23. Wang H, et al. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc Natl Acad Sci USA* 106:3853–3858.
24. Jian S, et al. (2008) Resolving an ancient, rapid radiation in Saxifragales. *Syst Biol* 57:38–57.
25. Davies TJ, et al. (2004) Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proc Natl Acad Sci USA* 101:1904–1909.
26. Pennisi E (2009) Origins. On the origin of flowering plants. *Science* 324:28–31.
27. Soltis PS, Soltis DE, Chase MW, Endress PK, Crane PR (2004) *Assembling the Tree of Life*, eds Cracraft J, Donoghue MJ (Oxford Univ Press, Oxford), pp 154–170.
28. Stefanović S, Rice DW, Palmer JD (2004) Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? *BMC Evol Biol* 4:35.
29. Kolaczowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431:980–984.
30. Soltis DE, Soltis PS (2004) *Amborella* not a “basal angiosperm”? Not so fast. *Am J Bot* 91:997–1001.
31. Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *amborella* is not a basal angiosperm. *Mol Biol Evol* 20:1499–1505.
32. Seo T-K, Kishino H, Thorne JL (2005) Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. *Proc Natl Acad Sci USA* 102:4436–4441.
33. Rannala B, Yang Z (2008) Phylogenetic inference using whole genomes. *Annu Rev Genomics Hum Genet* 9:217–231.
34. Sanderson MJ, Thorne JL, Wikstrom N, Bremer K (2004) Molecular evidence on plant divergence times. *Am J Bot* 91:1656–1665.
35. Magallón SA, Sanderson MJ (2005) Angiosperm divergence times: The effect of genes, codon positions, and time constraints. *Evolution* 59:1653–1670.
36. Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbcl* and 18S rDNA data. *Am J Bot* 88:1499–1516.
37. Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG (2002) Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils. *Proc Natl Acad Sci USA* 99:4430–4435.
38. Magallón SA (2004) Dating lineages: Molecular and paleontological approaches to the temporal framework of clades. *Int J Plant Sci* 165:57–S21.
39. Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 46:239–257.
40. Anderson CL, Bremer K, Friis EM (2005) Dating phylogenetically basal eudicots using *rbcl* sequences and multiple fossil reference points. *Am J Bot* 92:1737–1748.
41. Bell CD, Soltis DE, Soltis PS (2005) The age of the angiosperms: A molecular timescale without a clock. *Evolution* 59:1245–1258.
42. Bell CD, Donoghue MJ (2005) Dating the Dipsacales: Comparing models, genes, and evolutionary implications. *Am J Bot* 92:284–296.
43. Raubeson LA, Jansen RK (2005) *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, ed Henry RJ (CABI, Cambridge, MA), pp 45–68.
44. Moore MJ, et al. (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol* 6:17.
45. Swofford DL (2000) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4 (Sinauer Associates, Sunderland, MA).
46. Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
47. Felsenstein J (1985) Confidence limits on phylogeny: An approach using the bootstrap. *Evolution* 39:783–791.
48. Tavaré S (1986) *Lectures on Mathematics in Life Sciences*, ed Muir RM (American Mathematics Society, Providence, RI), pp 57–86.
49. Pond SL, Frost SDW, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
50. Huelsenbeck JP, Hillis DM, Jones R (1996) *Molecular Zoology: Advances, Strategies, and Protocols*, eds Ferraris JD, Palumbi SR (Wiley-Liss, New York), pp 19–45.
51. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* 39:306–314.
52. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
53. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
54. Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368–376.
55. Drummond AJ, Rambaut A (2006) BEAST Version 1.4. Available at <http://beast.bio.ed.ac.uk/>.
56. Drummond AJ, Rambaut A (2003) TRACER Version 1.4. Available at <http://evolve.zoo.ox.ac.uk/>.