Supporting Information

Wang *et al.* 10.1073/pnas.0813376106

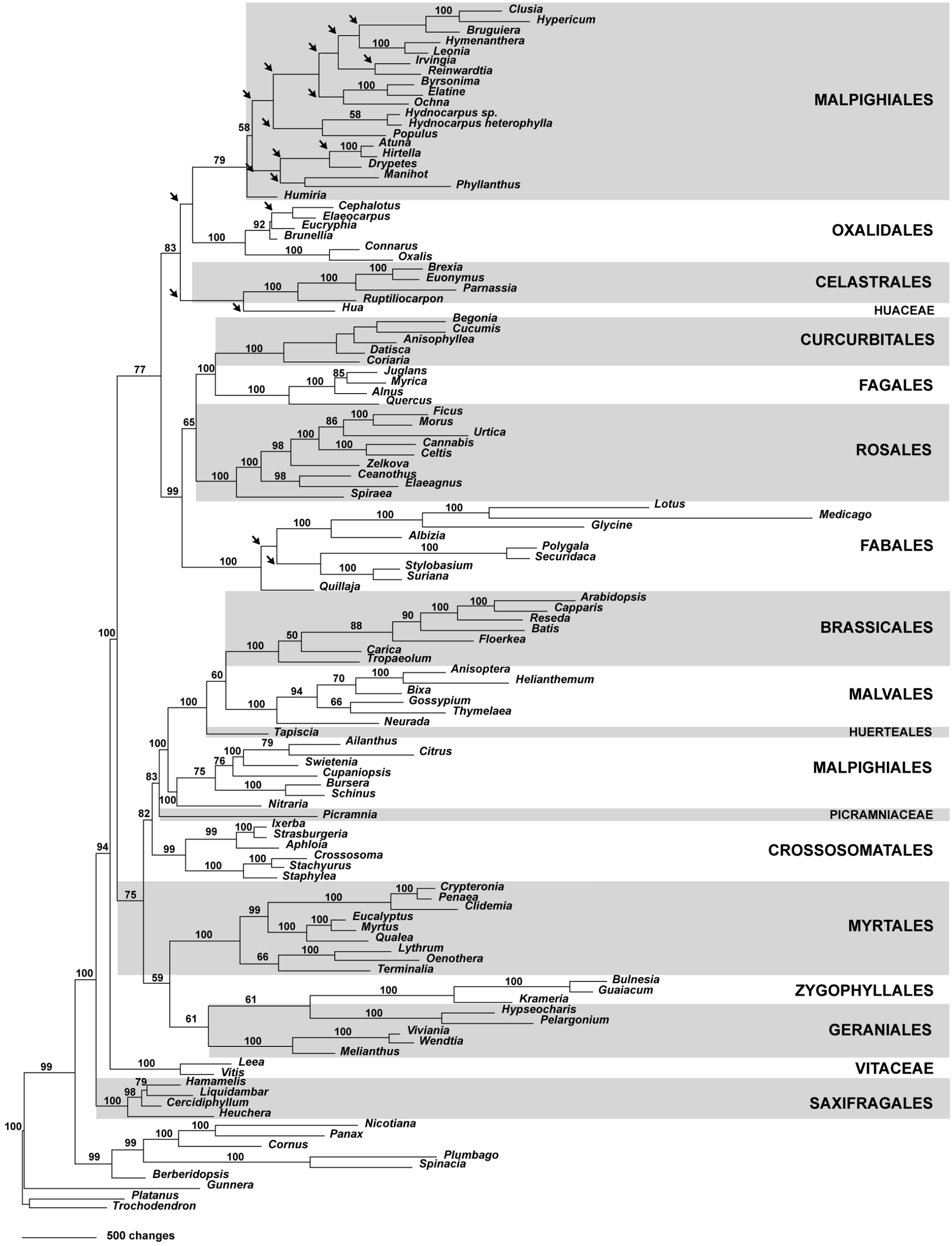

Fig. S1. Phylogram of 1 of 9 trees resulting from parsimony analysis of 12-gene dataset (2 nuclear and 10 plastid) and IR for 117 members of the rosid clade and outgroups (length 79,613, consistency index = 0.369, retention index = 0.433). Arrows designate nodes that collapse in the strict consensus. Numbers above branches are bootstrap values.

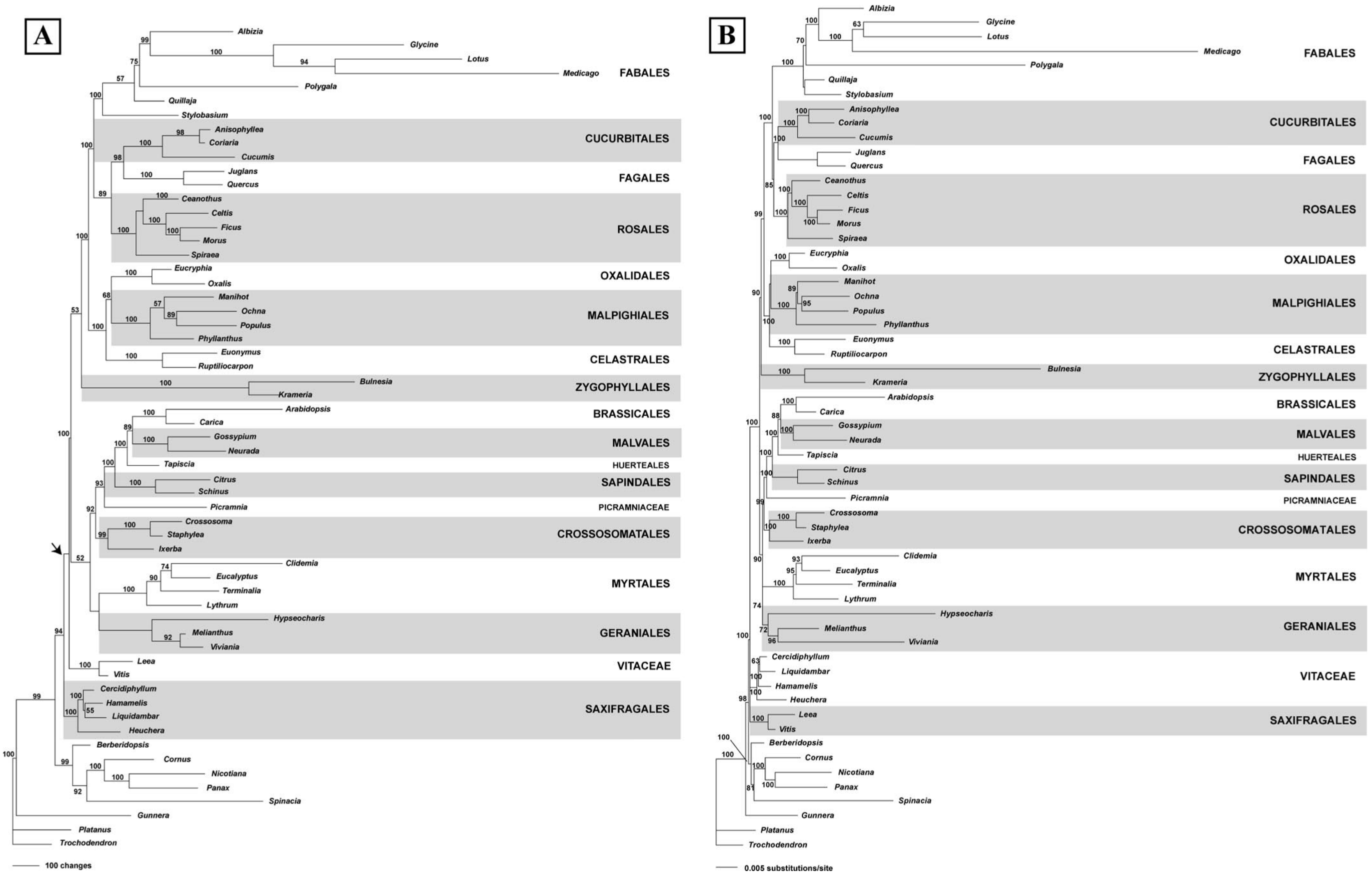

Fig. S2. Phylogenetic analyses of IR dataset for 59 taxa of the rosid clade and outgroups. (*A*) Phylogram of 1 of 4 most parsimonious trees. Bootstrap values are shown above branches. Arrows indicate nodes that collapse in the strict consensus. (*B*) Phylogram of ML tree.

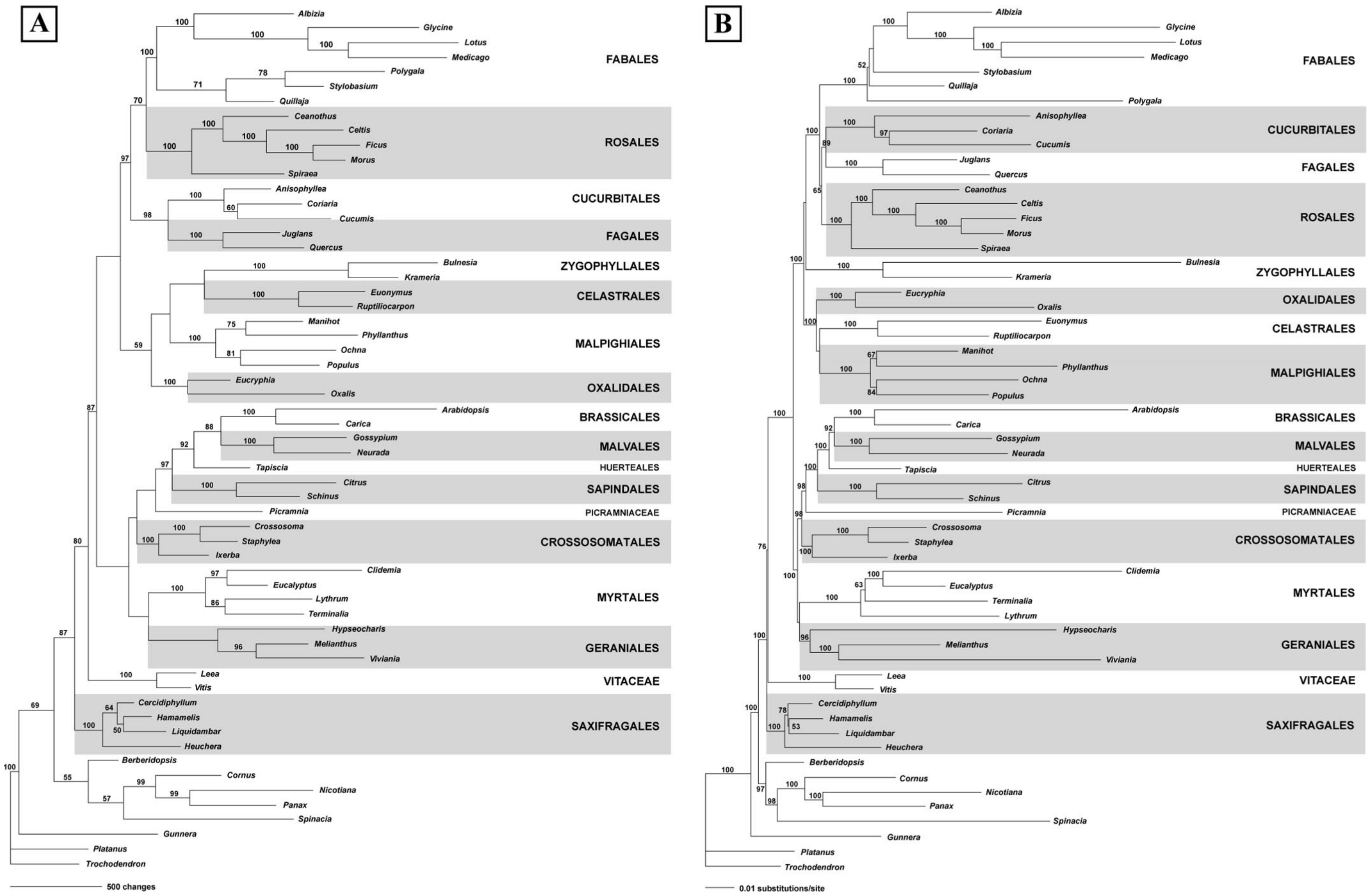

Fig. S3. Phylogenetic analyses of the 12-gene dataset (2 nuclear and 10 plastid) for 59 members of the rosid clade and outgroups. (*A*) Phylogram of the single tree resulting from MP analysis (length 34,843, consistency index = 0.374, retention index = 0.368). Numbers above branches are bootstrap values. (*B*) Phylogram of the ML tree.

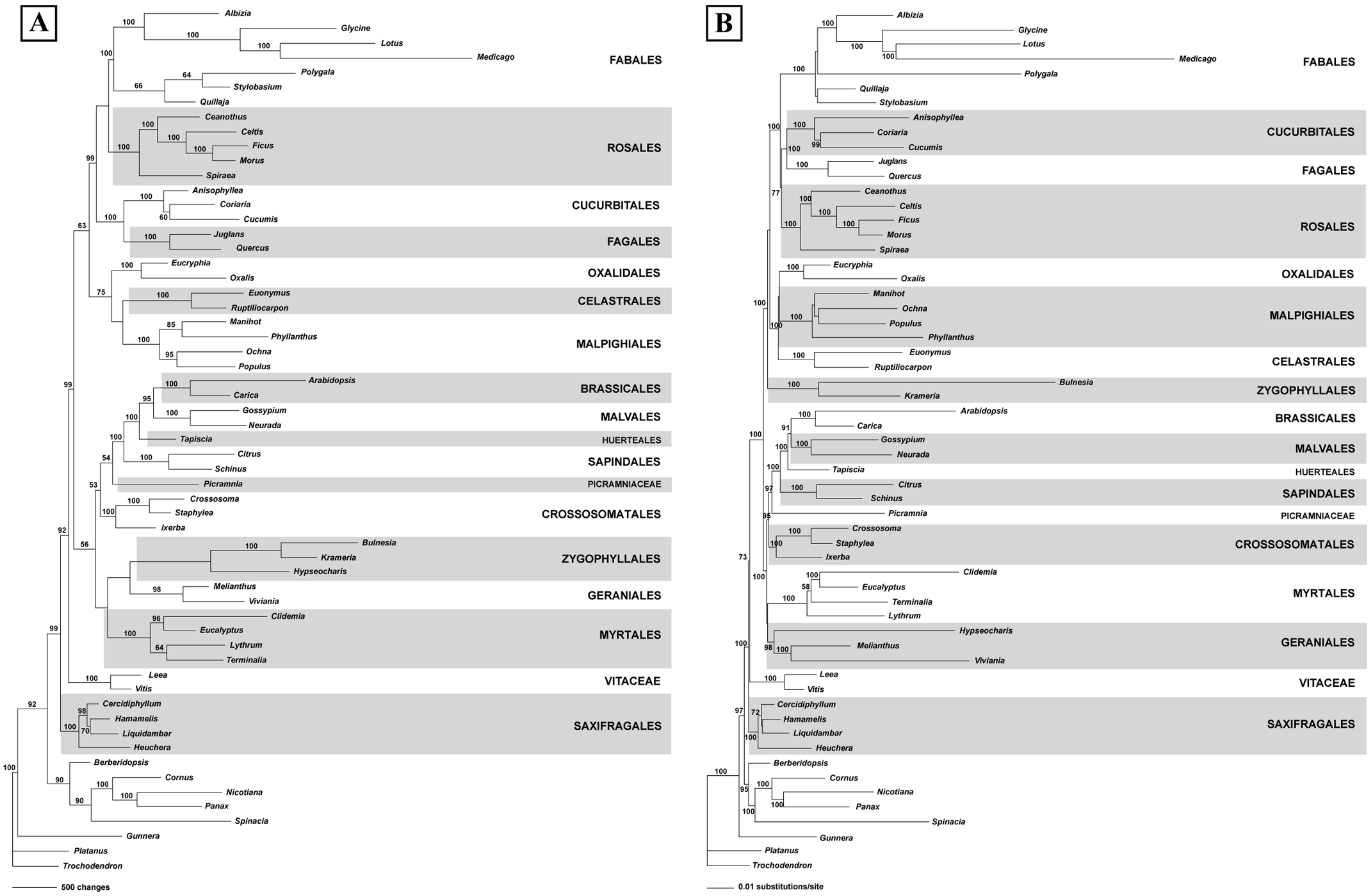

Fig. S4. Phylogenetic analyses of the total evidence dataset (2 nuclear genes, 10 plastid genes, and the IR) for 59 members of the rosid clade and outgroups. (*A*) Phylogram of the single tree resulting from MP analysis (length 53,418, consistency index = 0.476, retention index = 0.410). Numbers above branches are bootstrap values. (*B*) Phylogram of the ML tree.

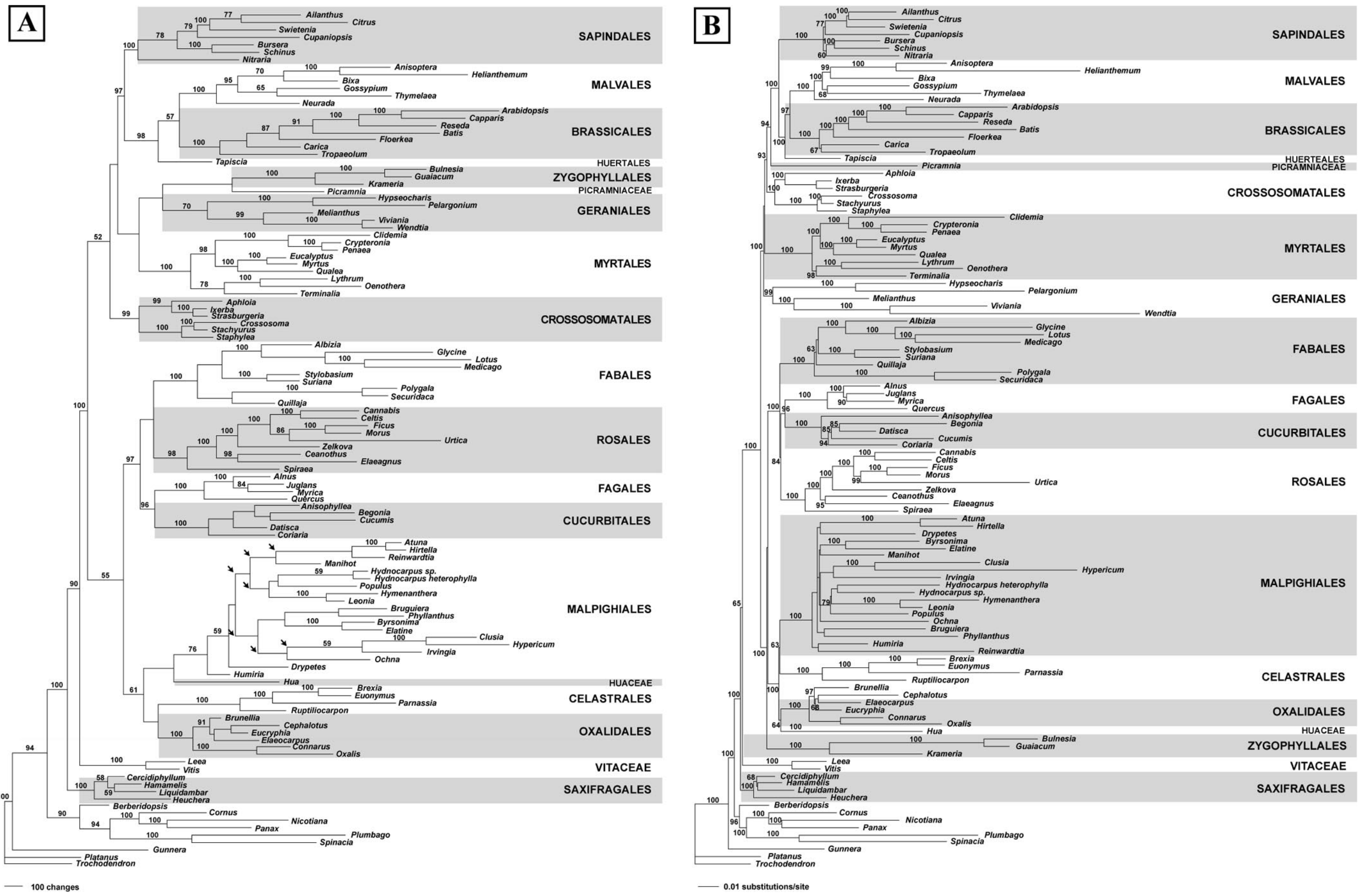

Fig. S5. Phylogenetic analyses of the 12-gene dataset (2 nuclear and 10 plastid) for 117 members of the rosid clade and outgroups. (*A*) Phylogram of 1 of 3 trees resulting from MP analysis (length 61,042, consistency index = 0.278, retention index = 0.417). Arrows designate nodes that collapse in the 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

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

CI, consistency index; RI, retention index.

Table S2. Primers used in this study

Gene	Primer name	Primer sequence	Ref.
26S rDNA			1
matK			2
ndhF	1F, 274F, 536F, 803F, 972F, 1318F; 536R, 803R, 972R, 1318R, 1603R, 2110R		3
	73F	CCC TTC ATT CCR CTT CCA GTT CCT	Designed for this study
	2153R	GGA ATT CCA TCA ATK ATT CGT YTA TC	Designed for this study
psbBTNH	60F, B60F, B61F, 63F, 65F, 67F, 69F; 61R, B64R, 66R, B66R, 68R, 70R, 71R; B71R		4
rpoC2	80F	TAG ATC ACT TCG GAA TGG C	Designed for this study
	750F	GGT CGT GTA TTA GCR GAC G	Designed for this study
	1000F	GGT ATT ATT GCG GGT CAA TC	Designed for this study
	1400F	AGG GRG ARA TGG CAY TGG AG	Designed for this study
	1950F	GAT CCT CGA TAC AGA AGA AAG AGT TC	Designed for this study
	2100F	TCA TTC CCG AGG AAG TRC AT	Designed for this study
	2600F	CAA GTA TTC AAT TAG TTC GGA CTT G	Designed for this study
	3000F	CAT CTA ATT GTT TTC GAA TGG GTC C	Designed for this study
	3400F	CAT TYA ATT TGA ATT GGT ATT TTC TSC	Designed for this study
	1000R	GAT TGA CCC GCA ATA ATA CC	Designed for this study
	1400R	CTC CAR TGC CAT YTC YCC CT	Designed for this study
	1950R	GAA CTC TTT CTT CTG TAT CGA GGA TC	Designed for this study
	2100R	ATG YAC TTC CTC GGG AAT GA	Designed for this study
	2600R	CAA GTC CGA ACT AAT TGA ATA CTT G	Designed for this study
	3000R	GGA CCC ATT CGA AAA CAA TTA GAT G	Designed for this study
	3400R	GGA GAA AAT ACC AAT TCA AAT TYA ATG	Designed for this study
	4200R	TTT CAG GCC TTT YAR CCA RTC	Designed for this study
rps4	1F, trnR		5
	48F	GGC TTT ACC RGG ACT AAC	Designed for this study

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

1. Soltis DE, *et al.* (2000). Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. *Bot J Linn Soc* 133:381–461.
2. Hilu KW, *et al.* (2003) Inference of angiosperm phylogeny based on *matK* sequence information. *Am J Bot 90:1758*–1776
3. Olmstead RG, Sweere JA (1994) Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. *Syst Biol* 43:467–481.
4. Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am J Bot* 87:1712–1730.
5. Souza-Chies TT, *et al.* (1997) Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene *rps4*. *Plant Syst Evol* 204:109–123.

Table S3. Gene/partition characteristics

Gene/partition	Number of missing taxa	Total aligned length	Analyzed aligned length	Number of parsimony-informative characters
18S rDNA	12	1832	1707	244
26S rDNA	39	3477	3213	526
total nuclear	9	5309	4920	770
atpB	9	1528	1431	527
matK	23	1845	1632	1,066
ndhF	8	2532	2124	1221
psbB	9	1530	1464	486
psbH	9	237	141	54
psbN	9	132	129	38
psbT	9	129	111	28
rbcL	0	1434	1395	495
rpoC2	7	5037	4143	2,302
rps4	7	810	591	292
IR	58	25549	25549	4,300
total plastid targeted genes (no IR)	0	40763	38710	6,509
total targeted genes (no IR)	0	20523	18081	7,279
total evidence	0	46072	43630	11,579

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 *rps4/trnV* spacer regions were sequenced but were not included in analyses, and so these regions are excluded from the numbers in this table.

Table S4. Fossil treatments

Clade	Assigned minimum age, Mya	Anchor fossil (and ref.)	UCLN-uniform	UCLN-lognormal
Simaroubaceae-Rutaceae-Meliaceae	E. Eocene, 50	*Ailanthus* (1)	*U*(50, 125)	*Ln*(0,1)
Burseraceae-Anacardiaceae	E. Eocene, 50	*Bursericarpum*, *Choerospondias* (2)	*U*(50, 125)	*Ln*(0,1)
Malvaceae s.l.	Maastrichtian, 65.5–70.6	*Parabombacaceoxylon, Javelinoxylon* (3, 4)	*U*(70.6, 125)	*Ln*(0,1)
Myrtales	Paleocene, 60	*Paleomyrtinaea* (5)	*U*(60, 125)	*Ln*(0,1)
Fagales	Santonian, 85	*Bedelia, Caryanthus, Antiquocupula* (6, 7, 8)	*U*(85, 125)	*Ln*(0,1)
Celtis-Cannabis	Paleocene, 60	*Celtis* (9)	*U*(60, 125)	*Ln*(0,1)
Clusia/*Hypericum*	Turonian, 90	*Paleoclusia* (10)	*U*(90, 125)	*Ln*(0,1)

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 *U*(minimum, maximum); UCLN-lognormal corresponds to the BEAST analysis that treated fossils as being drawn from a lognormal distribution *Ln*(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: *Paleoclusia*, placed as a constraint relative to *Clusia*/*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.

1. Corbett SR, Manchester SR (2004) Phytogeography and fossil history of *Ailanthus* (Simaroubaceae). *Int J Plant Sci* 165:671–690.
2. Collinson ME, Cleal CJ (2001) Introduction to the Mesozoic and Tertiary palaeobtaony of Great Britain, in *Mesozoic and Tertiary palaeobotany of Great Britain*, eds Cleal CJ, Thomas BA, Batten DJ, Collinson ME (Geological Conservation Review Series No. 22, Peterborough: Joint Nature Conservation Committee), pp 187–226.
3. Wheeler EA, Lee M, Matten LC (1987) Dicotyledonous woods from the Upper Cretaceous of Southern Illinois. *Bot J Linn Soc* 95:77–100.
4. Wheeler EA, Lehman TM, Gasson PE (1994) *Javelinoxylon*, a new genus of malvalean tree from the Upper Cretaceous and Paleocene of San Juan Basin, New Mexico, USA. *Amer J Bot* 81:703–710.
5. Pigg KB, Stockey RA, Maxwell SL (1992) *Paleomyrtinaea*, a new genus of permineralized myrtaceous fruits and seeds from the Eocene of British Columbia and Paleocene of North Dakota. *Can J Bot* 71:1–9.
6. Sims HJ, Herendeen PS, Lupia R, Christopher RA, Crane PR (1999) Fossil flowers with Normapolles pollen from the Late Cretaceous of southeastern North America. *Rev Palaeo and Palynol* 106:131–151.
7. Friis EM (1983) Upper Cretaceous (Senonian) floral structures of juglandalean affinity containing Normapolles pollen. *Rev Paleo Palynol* 39:161–188.
8. Sims HJ, Herendeen PS, Crane PR (1998) A new genus of fossil Fagaceae from the Campanian (Late Cretaceous) of central Georgia, U.S.A. *Int J Plant Sci* 159:391–404.
9. Manchester SR, Akhmetiev MA, Kodrul T (2002) Leaves and fruits of *Celtis aspera* (Newberry) comb. nov. (Celtidaceae) from the Paleocene of North America and eastern Asia. *Int J Plant Sci* 163:725–736.
10. Crepet WL, Nixon KC (1998) Fossil Clusiaceae from the Late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *Amer J Bot* 85: 1122–1133.

Other Supporting Information

SI Appendix