nature plants

ARTICLES

https://doi.org/10.1038/s41477-019-0421-0

Origin of angiosperms and the puzzle of the Jurassic gap

Hong-Tao Li[1,15], Ting-Shuang Yi[1,15], Lian-Ming Gao[2,15], Peng-Fei Ma[1,15], Ting Zhang[1,15], Jun-Bo Yang[1,15], Matthew A. Gitzendanner[3,4,15], Peter W. Fritsch[5], Jie Cai[1], Yang Luo[2], Hong Wang[2], Michelle van der Bank[6], Shu-Dong Zhang[1], Qing-Feng Wang[7], Jian Wang[8], Zhi-Rong Zhang[1], Chao-Nan Fu[2,9], Jing Yang[1], Peter M. Hollingsworth[10], Mark W. Chase[11,12], Douglas E. Soltis[3,4,13,14], Pamela S. Soltis[3,13,14]* and De-Zhu Li[1,2,9]*

Angiosperms are by far the most species-rich clade of land plants, but their origin and early evolutionary history remain poorly understood. We reconstructed angiosperm phylogeny based on 80 genes from 2,881 plastid genomes representing 85% of extant families and all orders. With a well-resolved plastid tree and 62 fossil calibrations, we dated the origin of the crown angiosperms to the Upper Triassic, with major angiosperm radiations occurring in the Jurassic and Lower Cretaceous. This estimated crown age is substantially earlier than that of unequivocal angiosperm fossils, and the difference is here termed the 'Jurassic angiosperm gap'. Our time-calibrated plastid phylogenomic tree provides a highly relevant framework for future comparative studies of flowering plant evolution.

Angiosperms, that is, flowering plants, are critical components of nearly all terrestrial and many aquatic habitats, and knowledge of their origin and evolution can provide the framework for understanding the history and composition of major terrestrial ecosystems and general patterns of biodiversity[1,2]. The apparent rapid early diversification of angiosperms within a short geological time period was referred to by Darwin as "an abominable mystery"[3]. Their rise to dominance in ecosystems since the Lower Cretaceous subsequently promoted diversification of insects[4,5], amphibians[6], mammals[7], ferns[8] and many other organisms.

Uncertain relationships among major subclades of angiosperms have hindered a better understanding of patterns of angiosperm diversification and the evolution of key traits[1,9]. Over the past two decades, molecular phylogenetic studies have greatly advanced our knowledge of angiosperm evolution, as reflected in the recent update to the Angiosperm Phylogeny Group (APG) classification[10]. Amborellaceae are sister to all other extant angiosperms, followed by Nymphaeales and then Austrobaileyales[11–13]. These clades are, together, referred to as the ANA grade, and the clade comprising the remainder of the angiosperms is referred to as the mesangiosperms[14]. Despite the inclusion in many previous phylogenetic studies of representatives of the five clades of mesangiosperms, that is, Chloranthales, magnoliids, monocots, Ceratophyllales and eudicots, the relationships among these have remained unclear[9], with as many as 15 poorly to moderately supported topologies having been proposed[15]. Furthermore, the composition and relationships of some subclades recognized as orders and families have also remained weakly supported due to sparse gene sampling, low taxonomic coverage or both.

The plastid genome (plastome) has been the most important source of data for reconstruction of green plant phylogeny[9], and recent phylogenomic analyses with nuclear genes have generally supported previous plastid-based hypotheses[15,16]. To improve our current understanding of plastid phylogenomics, we generated and assembled a large DNA dataset comprising 80 genes from 2,881 plastomes and estimated divergence times using a validated set of 62 fossils spread across the angiosperm tree.

Results and Discussion

We assembled a dataset of 2,881 plastomes (Supplementary Table 1) including 2,694 plastomes of angiosperms representing 2,351 species from all 64 clades recognized as orders and 353 (85%) of the 416 APG IV families[10], as well as 187 plastomes from 163 species of gymnosperms as outgroups. To maximally capture the plastome and taxon diversity of angiosperms, particularly at the familial level, we included published angiosperm plastomes in National Center for Biotechnology Information (last accessed 1 January 2017) and generated plastome sequences for 1,659 species from 63 orders and 347 families (Supplementary Table 1), including 677 plastomes of 671 species from the OneKP Project[17]; those of six orders and

[1]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. [2]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China. [3]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. [4]Department of Biology, University of Florida, Gainesville, FL, USA. [5]Botanical Research Institute of Texas, Fort Worth, TX, USA. [6]Department of Botany & Plant Biotechnology, University of Johannesburg, Johannesburg, South Africa. [7]Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China. [8]Queensland Herbarium, Department of Environment and Science, Brisbane Botanic Gardens, Toowong, Queensland, Australia. [9]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China. [10]Royal Botanic Garden Edinburgh, Edinburgh, UK. [11]Royal Botanic Gardens, Kew, UK. [12]Department of Environment and Agriculture, Curtin University, Bentley, Western Australia, Australia. [13]Genetics Institute, University of Florida, Gainesville, FL, USA. [14]Biodiversity Institute, University of Florida, Gainesville, FL, USA. [15]These authors contributed equally: Hong-Tao Li, Ting-Shuang Yi, Lian-Ming Gao, Peng-Fei Ma, Ting Zhang, Jun-Bo Yang, Matthew A. Gitzendanner. *e-mail: psoltis@flmnh.ufl.edu; DZL@mail.kib.ac.cn

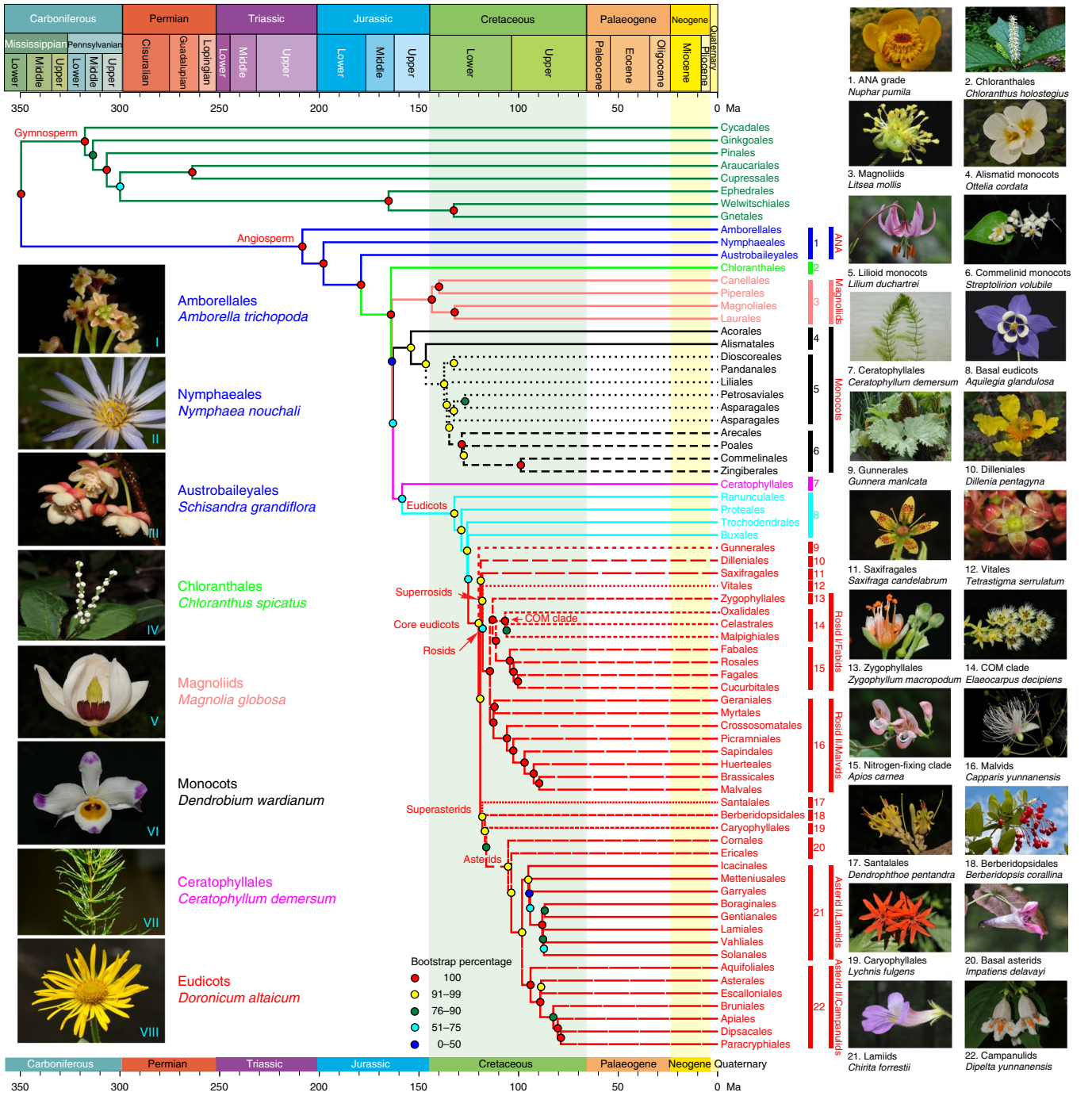

Fig. 1 | Dated phylogenetic tree showing relationships among orders of angiosperms based on a maximum likelihood partitioned analysis of 82,286 bp of DNA sequence from 80 genes of 2,881 plastomes. Representative images of eight major backbone clades and 22 major clades on the left and right, respectively. The colours of the clade branches and terminal taxon names represent eight major clades (dark blue, ANA grade; green, Chloranthales; light red, magnoliids; black, monocots; purple, Ceratophyllales; red/light blue, eudicots). The dashed lines represent major clades within monocots and core eudicots.

195 families are newly sequenced here. This study represents an extensive taxon sampling for plastid phylogenomic analyses of angiosperms[9,17], greatly expanding sampling density within the key clades, especially for those diverging early in angiosperm history.

Alignment of 80 genes for 2,881 plastomes produced a matrix of 82,286 base pairs with only ~5% missing data despite the loss of many genes in some heterotrophic taxa. Maximum likelihood analysis of this matrix resulted in a well-resolved plastid phylogenomic angiosperm (PPA) tree (Fig. 1 and Supplementary Fig. 1), with 79% of angiosperm nodes at or above the ordinal level and 83% at or above the familial level allocated bootstrap percentage ≥90.

We also performed maximum likelihood analysis on the complete 80-gene matrix with 24% of the most rapidly evolving sites removed (Supplementary Table 2), and the results were almost identical to the tree based on all sites. Although this approach may remove sites with valuable signal as well as noise[18], there are no conflicts (bootstrap percentage ≥70) at the familial and ordinal levels between the two trees (Supplementary Fig. 2). In comparison to the

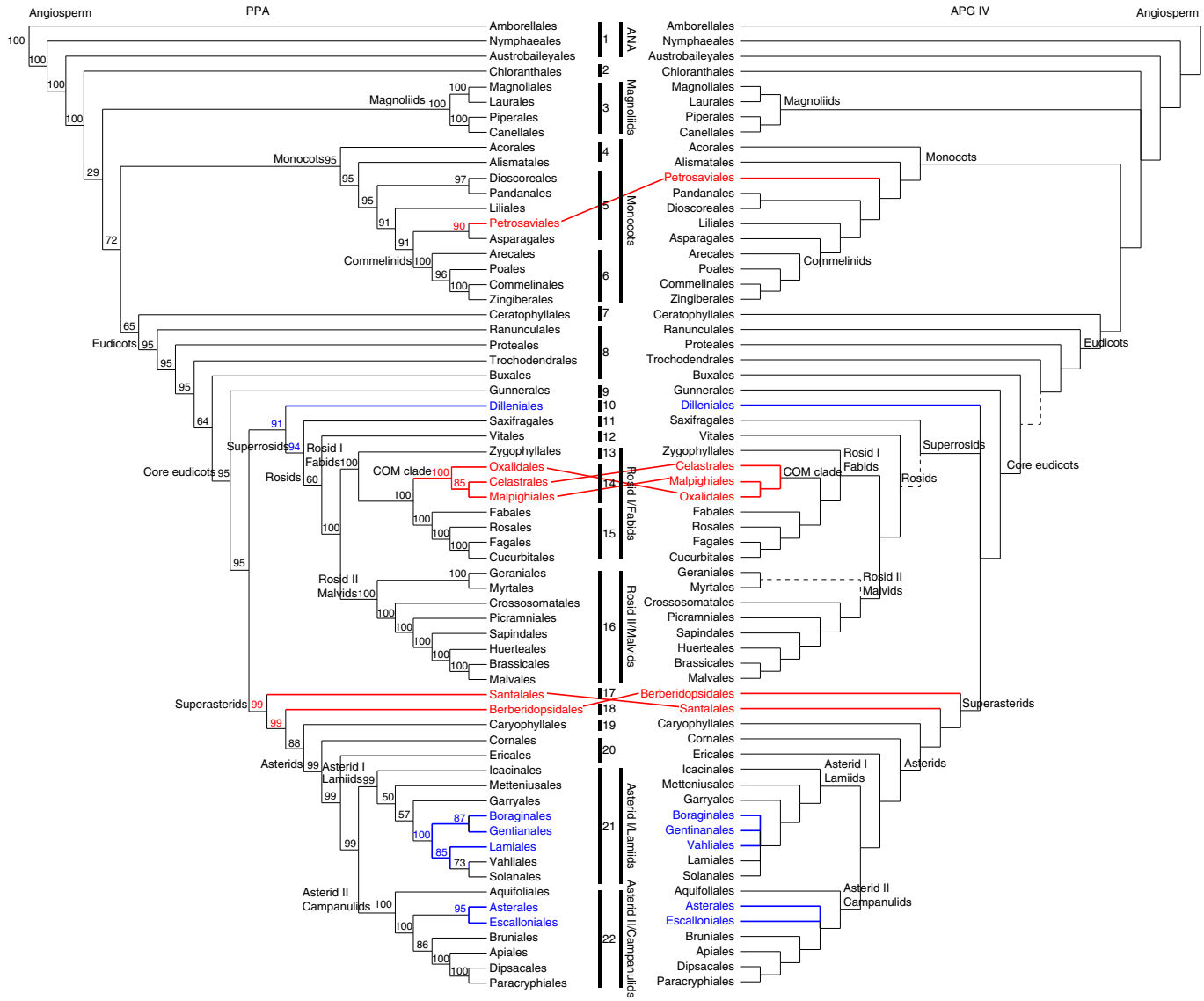

Fig. 2 | Angiosperm ordinal phylogenetic relationships in PPA (based on the complete 80-gene matrix) versus APG IV. Dashed lines indicate clades with bootstrap percentage <50. Red, difference in phylogenetic position between PPA and APG IV trees with bootstrap percentage ≥85 in PPA; blue, phylogenetic relationships resolved in PPA with bootstrap percentage ≤85 but not in APG IV.

tree based on the complete 80-gene matrix, the bootstrap percentages for phylogenetic placements of six families increased to >70, whereas those for six other families decreased to <70. Given the similarity between these two trees, we discuss phylogenetic relationships inferred from the complete matrix.

The general PPA relationships are nearly identical to those of the APG IV summary tree, which is a synthesis of numerous published phylogenetic studies (Fig. 2). The phylogenetic (including unresolved) placements of clades corresponding to 12 orders and 51 families, however, differed between the PPA tree and the APG IV summary tree as represented on the Angiosperm Phylogeny website (www.mobot.org/MOBOT/research/APweb/) (Supplementary Fig. 3 and Supplementary Information). Among these orders, six are placed as unresolved in the APG IV summary tree and the other six conflict (Fig. 2). Most new PPA phylogenetic positions of these orders exhibit bootstrap percentage ≥90. Petrosaviaceae grouped in Asparagales as sister to Orchidaceae in the PPA tree (bootstrap percentage, 90; Fig. 2), but are sister to all monocots except Acorales and Alismatales in the APG IV summary tree. Other notable phylogenetic results obtained here include the strong support (bootstrap percentage, 91) for Dilleniales as sister to the superrosids. In addition, an Asterales + Escalloniales clade is recovered (bootstrap percentage, 95), whereas that relationship was not resolved in the APG IV summary tree. Santalales, rather than Berberidopsidales, are resolved with strong support as sister to the remaining superasterids. Celastrales and Malpighiales are strongly supported as sisters within the Celastrales–Oxalidales–Malpighiales clade. Core lamiids also received moderate bootstrap percentage here, from 73–87 (Supplementary Information).

At the familial level, 42 of the 51 families with phylogenetic relationships differing between the PPA and APG trees were allotted bootstrap percentage ≥90 in our tree (Supplementary Information). In all, the PPA tree provides insights into the phylogenetic relationships of numerous specific clades (Supplementary Information) and can serve as the most comprehensive overall plastid-based phylogenetic assessment, better focusing a multitude of future studies on flowering plants.

Along the backbone of the angiosperm tree, Amborellales, Nymphaeales and Austrobaileyales (ANA grade; Fig. 1) were

Table 1 | Details of the 62 fossils used as calibration point prior settings in divergence time analyses of angiosperms

Calibrated node	Fossils	Age (Ma)
CG Spermatophyta	Fossils of Cordaitales	350
CG Pinopsida	*Cordaixylon iowensis*	307
SG Cycadaceae	*Primocycas chinensis*	273
CG Podocarpaceae–Araucariaceae	*Araucarites rudicula*	227
CG Pinaceae	*Compsostrobus neotericus*	201.3
CG Taxaceae–Cupressaceae	*Palaeotaxus rediviva*	201.3
CG Cupressaceae	*Austrohamia minuta; Sewardiodendron laxum; Austrohamia acanthobractea; Hughmillerites juddii*	163.5
SG *Araucaria*	*Araucaria mirabilis*	157.3
CG Angiosperms	Unnamed	132.9
CG Eudicots	Tricolpate pollen grains	125
CG Winteraceae–Canellaceae	*Walkeripollis gabonensis*	125
CG Trimeniaceae–Schisandraceae	*Anacostia teixeirae*	125
CG Chloranthaceae	*Asteropollis* sp.	121
SG *Pinus*	*Pityostrobus bernissartensis*	118.5
CG Monocots	*Liliacidites* sp.	113
CG Magnoliaceae–Annonaceae–Degeneriaceae–Eupomatiaceae–Himantandraceae	*Endressinia brasiliana*	113
CG Ranunculales	*Teixeiraea lusitanica*	113
CG Nymphaeaceae	*Monetianthus mirus*	113
CG Laurales	*Virginianthus calycanthoides*	104.7
CG Proteales	*Platanocarpus brookensis*	104.7
SG Buxales	*Spanomera marylandensis*	100.5
CG Monimiaceae–Lauraceae–Hernandiaceae	*Mauldinia mirabilis*	100.5
CG Pentapetalae	Pentamerous flower	100.5
CG Fagales	*Caryanthus triasseris*	93.9
SG Brassicales	*Dressiantha bicarpelata*	89.8
CG Saxifragales	*Microaltingia apocarpela*	89.8
CG Myrtales	*Esgueiria* sp.	89.8
CG Cornales	*Hironoia fusiformis*	89.8
CG Ericales	*Paleoenkianthus sayrevillensis*	89.8
SG Clusiaceae	*Paleoclusia chevalieri*	89.8
SG Annonaceae	*Futabanthus asamigawaensis*	89.8
CG Arecales	*Sabal magothiensis*	86.3
SG Myrtaceae	Flower number 3	83.6
SG Juglandaceae–Myricaceae	*Caryanthus* sp.	83.6
CG Actinidiaceae	*Parasaurauia allonensis*	83.6
CG Hamamelidoideae	*Allonia decandra; Androdecidua endressi*	83.6
SG Paracryphiaceae	*Silvianthemum suecicum*	78.9
SG Cunoniaceae	*Platydiscus peltatus*	78.9
SG Zingiberales	*Spirematospermum chandlerae*	77.8
CG Asteraceae	*Tubulifloridites lilliei*	72.1
SG Rhamnaceae	*Coahuilanthus belindae*	72.1
CG Amaranthaceae	*Polyporina cribraria*	66
CG Cannabaceae	*Aphananthe cretacea*	66
CG Sabiaceae	*Meliosma prealba; Sabia microsperma; Sabia praeovalis*	66
CG Rutaceae	*Rutaspermum biornatum*	66
CG Juglandaceae	*Polyptera manningi*	64.4
CG Elaeocarpaceae	*Sloanea ungeri*	61.6

Continued

Table 1 | Details of the 62 fossils used as calibration point prior settings in divergence time analyses of angiosperms (continued)

Calibrated node	Fossils	Age (Ma)
SG *Thuja*	*Thuja Polaris*	59.2
CG Torricelliaceae	*Toricellia bonesii*	56
CG Sapindaceae	*Aesculus hickeyi*	56
CG Vitaceae	*Ampelocissus parvisemina*	56
CG Myrtaceae	*Paleomyrtinaea princetonensis*	56
SG *Dioon*	*Dioon praespinulosum*	56
SG Puelioideae	Unnamed	56
SG *Rhododendron*	*Rhododendron newburyanum*	56
SG *Physalis*	*Physalis infinemundi*	52.2
SG *Prunus*	*Prunus cathybrownae*	49.4
CG Meliaceae	*Toona sulcata*	47.8
SG *Anacardium*	*Anacardium germanicum*	47.8
CG Tapisciaceae	*Tapiscia occidentalis*	37.8
CG Araliaceae	*Paleopanax oregonensis*	37.8
SG *Lepidozamia*	*Lepidozamia foveolata*	33.9

CG, crown group; SG, stem group.

recovered as the successive sister groups to all other angiosperms; the remaining angiosperms (mesangiosperms) were again recovered as monophyletic. However, bootstrap percentages for relationships among the five main mesangiosperm clades were generally low (Fig. 2), with only moderate support for the monocots–Ceratophyllales–eudicots relationships, despite extensive taxon and gene sampling. These relationships have long been debated, having previously been contingent on taxon sampling and data source[13,15,16]. To evaluate these results further, a data subset (82 taxa representing all 64 orders) was analysed by delimiting various data partitions. Although resolution and support were again poor, the node connecting Ceratophyllales and eudicots was moderately supported (bootstrap percentage, 43–93; Supplementary Figs. 4 and 5). Moreover, the five component clades of mesangiosperms all have relatively long subtending branches connected by short internodes (Supplementary Fig. 5), indicating a rapid radiation of the major clades of mesangiosperms. Thus, even with 80 plastid genes and 2,881 samples, resolution of relationships among the five mesangiosperm clades proved challenging. Together with past geological evidence[2] and dating analyses here, we suggest that rapid radiations may have occurred during the early stages of mesangiosperm evolution[13,19].

We selected 62 fossils that reliably represent the oldest record of major seed plant clades, orders, families and some intra-familial taxa (Table 1 and Supplementary Information) as constraints for calibrating the age of our angiosperm tree. We chose fossils with the reproductive structures of flowers, fruits and seeds over those with only vegetative structures or pollen. Most fossils were selected from three well-documented studies[20–22], which included detailed descriptions of fossil ages, locality, type, morphological characters and relationships with taxa in the tree. We supplied detailed information of the remaining fossils based on descriptions, illustrations and their similarity to extant taxa as described in the original publications (Supplementary Information). The absolute age assigned to each fossil was used. In most cases, the absolute ages of fossils were equal to the uppermost boundary of the stratigraphic interval to which the fossils could be assigned. Ages of stratigraphic boundaries were converted to absolute ages with the International Chronostratigraphic Chart (International Commission on Stratigraphy v.2017/02).

Age estimates of the angiosperm crown group exhibit considerable variation and have been heavily influenced by differences in taxon sampling, gene sampling and numbers, fossil calibrations and evolutionary models[20,21]. We investigated the temporal evolution of angiosperms with a matrix of unprecedented size and a set of well-documented fossil calibrations (Table 2). Our study suggests that crown-group angiosperms arose in the Rhaetian in the Upper Triassic ~209 Ma (267–187 Ma, 95% highest posterior density; Figs. 1, 3 and 4 and Supplementary Table 3). This estimate is generally congruent with several estimates from other relaxed molecular clock analyses based on DNA and/or genomic data (Table 2)—for example, 209 (246–186) Ma (ref. [15]), 194 (210–162) Ma (ref. [22]), 214 (238–190) Ma (ref. [23]), 221 (251–192) Ma (ref. [21]) and a range of 149–256 Ma[20], although younger estimates such as 139 Ma (ref. [24]) or 147 (154–141) Ma (ref. [25]) and older estimates such as 232 (256–210) Ma (ref. [26]) and 243 Ma (ref. [27]) have also been suggested. The estimates of divergence times for other selected major groups of angiosperms are summarized in Table 2.

Our estimated age of the angiosperm crown group also corresponds closely with a transcriptomic-based estimate for the origin of several major clades of Insecta[4]—for example, Megaloptera, Neuroptera, Orthoptera, the clade of Trichoptera and Lepidoptera, and Zygentoma. Nearly all estimated crown angiosperm ages based on molecular data are substantially older than the age estimate of the unequivocal fossil pollen found in the Valanginian to early Hauterivian of the Lower Cretaceous (~139–131 Ma)[1,28], with a gap of ~70 (48–128) million years (Table 2). We refer to this difference between the age estimates, based on molecular and morphological data, as the Jurassic angiosperm gap. Nevertheless, angiosperm-like pollen grains have been reported from the Middle Triassic (~247–242 Ma)[29] and may represent the stem group of angiosperms[23].

Although no unequivocal evidence of pre-Cretaceous angiosperms thus far exists, the morphology of some fossils from the early Cretaceous suggests close relationships to Austrobaileyales, Chloranthales, Nymphaeales and magnoliids[30,31]. These clades had already diversified in numbers and morphology in the early Cretaceous, implying that their ancestor is older. Estimated ages of the major clades of angiosperms that are clearly older than their putative earliest fossils may indicate an incomplete fossil record, a bias in molecular dating analysis or both[20,21]. However, support for a

Table 2 | Comparison of angiosperm divergence time estimates from previous studies to those from our study

Study	Calibration point	Oldest fossil (calibration node/age)	Molecular dataset		Age of clade (crown group)				
			No. of taxa (angiosperms)	No. of loci	Angiosperms	Mesangiosperms	Magnoliids	Monocots	Eudicots
Ref. [62]	5	Eudicots/125	71 (71)	4 (2-pt, 1-mt, 1-nuc)	180–140	N/A	N/A	133–99	125–93
Ref. [13]	6	Spermatophyta/310–290	45 (43)	61 (61-pt)	170	144	130	129	125
Ref. [63]	49	Spermatophyta/350	267 (265)	5 (3-pt, 2-nuc)	242–241	N/A	N/A	N/A	N/A
Ref. [25]	36	Angiospermae/350–132	567 (560)	3 (2-pt, 1-nuc)	183 (199–167)	146 (156–139)	122 (138–108)	N/A	130 (139–123)
Ref. [64]	33	Tracheophyta/377	154 (113)	3 (2-pt, 1-nuc)	217 (257–182)	174 (200–153)	155 (181–136)	156 (167–139)	137 (147–128)
Ref. [65]	17	Embryophyta/1,042	18 (8)	7 (7-pt)	205 (240–175)	152 (179–133)	N/A	N/A	94 (115–83)
Ref. [22]	28	Embryophyta–streptophyte divergence/912	81 (44)	5 (5-pt)	194 (210–162)	152 (165–140)	143 (155–131)	134 (145–125)	125 (129–120)
Ref. [27]	39	Spermatophyta/320–290	639 (631)	11 (11-pt)	243	194	147	171	173
Ref. [15]	2	Spermatophyta/310–290	61 (60)	59 (59-nuc)	209 (246–186)	170 (171–150)	136 (150–122)	138 (149–127)	126 (126–115)
Ref. [24]	137	Spermatophyta/330	799 (792)	5 (3-pt, 2-nuc)	139–136	137–135	134–130	135–132	133–130
Ref. [26]	24	Spermatophyta/317	125 (121)	4 (3-pt, 1-nuc)	256–210	N/A	195–160	181–149	170–142
Ref. [21]	37	Spermatophyta/350	195 (193)	76 (76-pt)	221 (251–192)	177 (197–159)	150 (171–130)	161 (176–141)	145 (154–136)
Ref. [23]	2	Eudicots/125	37 (34)	1,175 (1,175-nuc)	214 (238–190)	N/A	N/A	N/A	98 (109–87)
Ref. [20]	52	Spermatophyta/337	644 (632)	83 (77-pt, 4-mt, 2-nuc)	256–149	210–138	190–128	181–123	188–129
This study	62	Spermatophyta/350	2,541 (2,378)	80 (80-pt)	209 (267–187)	164 (193–146)	144 (155–136)	154 (184–131)	132 (161–125)

Times in millions of years ago (Ma). Ranges correspond to 95% highest posterior density. pt, plastid; mt, mitochondrial; nuc, nuclear; N/A, not available.

pre-Cretaceous origin of angiosperms is strengthened by nearly all molecular divergence time estimates pointing to a Jurassic or even Triassic origin[13,21,22,26]. The possibility of a Triassic origin of angiosperms is echoed by the recent suggestion of a cryptic early history of angiosperms[20].

Apart from Amborellales, the earliest divergences within angiosperms based on our analyses are those of Nymphaeales, followed by Austrobaileyales from the rest of the angiosperms at ~198 (236–180) Ma and ~179 (210–163) Ma, respectively, in the Lower Jurassic (Fig. 3 and Supplementary Table 3). The five major clades of mesangiosperms diverged rapidly from the late Middle Jurassic ~164 (193–146) Ma to early Upper Jurassic ~159 (188–140) Ma. Magnoliids and (monocots + (Ceratophyllales + eudicots)) diverged at ~164 (192–146) Ma, and monocots and (Ceratophyllales + eudicots) split at ~163 (192–145) Ma (Fig. 3 and Supplementary Table 3).

From our analyses, several radiations occurred among monocots and major core eudicot clades during the Cretaceous (~138–133 Ma: Dioscoreales-Pandanales, Liliales, commelinids, Asparagales, Dioscoreales and Pandanales; ~126–116 Ma: Trochodendrales, Buxales, Gunnerales, core eudicots, Dilleniales, superrosids, Saxifragales, Santalales, superasterids, Vitales, rosids, Berberidopsidales and Caryophyllales; ~107–103 Ma: Oxalidales, Celastrales, Malpighiales, Crossosomatales, Cornales, asterids, Fabales, Ericales, Picramniales and Rosales; ~95–94 Ma: Icacinales, and lammiids asterid I, Metteniusales, Garryales, Aquifoliales and campanulids asterid II; 90–87 Ma: Brassicales, Malvales, Asterales-Escalloniales, Asterales, Escalloniales, Boraginales-Gentianales, Lamiales, Vahliales, Solanales, Boraginales and Gentianales) (Figs. 3 and 4 and Supplementary Fig. 6). All 64 angiosperm clades recognized as orders appeared before the mid-Campanian (~79 Ma) in the Upper Cretaceous (Figs. 1, 3 and 4). Co-evolution among insects and angiosperms is likely to have driven angiosperm diversification and radiation (representing orders and some families) during the Cretaceous[32], but other ecological factors were probably also involved[2,33].

Lower Cretaceous fossils related to monocots and magnoliids support the early divergence of these clades among angiosperms[1,30,31]. Within the monocots, Acorales are sister to the remaining monocots (~154 (184–131) Ma), followed by Alismatales (~147 (177–120) Ma) at the end of the Jurassic or in the Lower Cretaceous (Figs. 1, 3 and 4). Radiations of the major monocot clades occurred in the Valanginian in the Lower Cretaceous (~137–133 Ma). Commelinales and Zingiberales diverged much later in the Cenomanian (~99 Ma, Upper Cretaceous). The four magnoliid orders diverged in the Lower Cretaceous (~140–132 Ma).

We inferred an age of 132 (161–125) Ma for the crown eudicots. The first appearance of tricolpate pollen grains at the Barremian–Aptian boundary (125 Ma) has long been recognized as the oldest fossil evidence for eudicots[1], and this age has been widely used in calibrating divergence time analyses as the soft maximum, minimum or fixed age for the eudicots[21–24]. The date is one of the

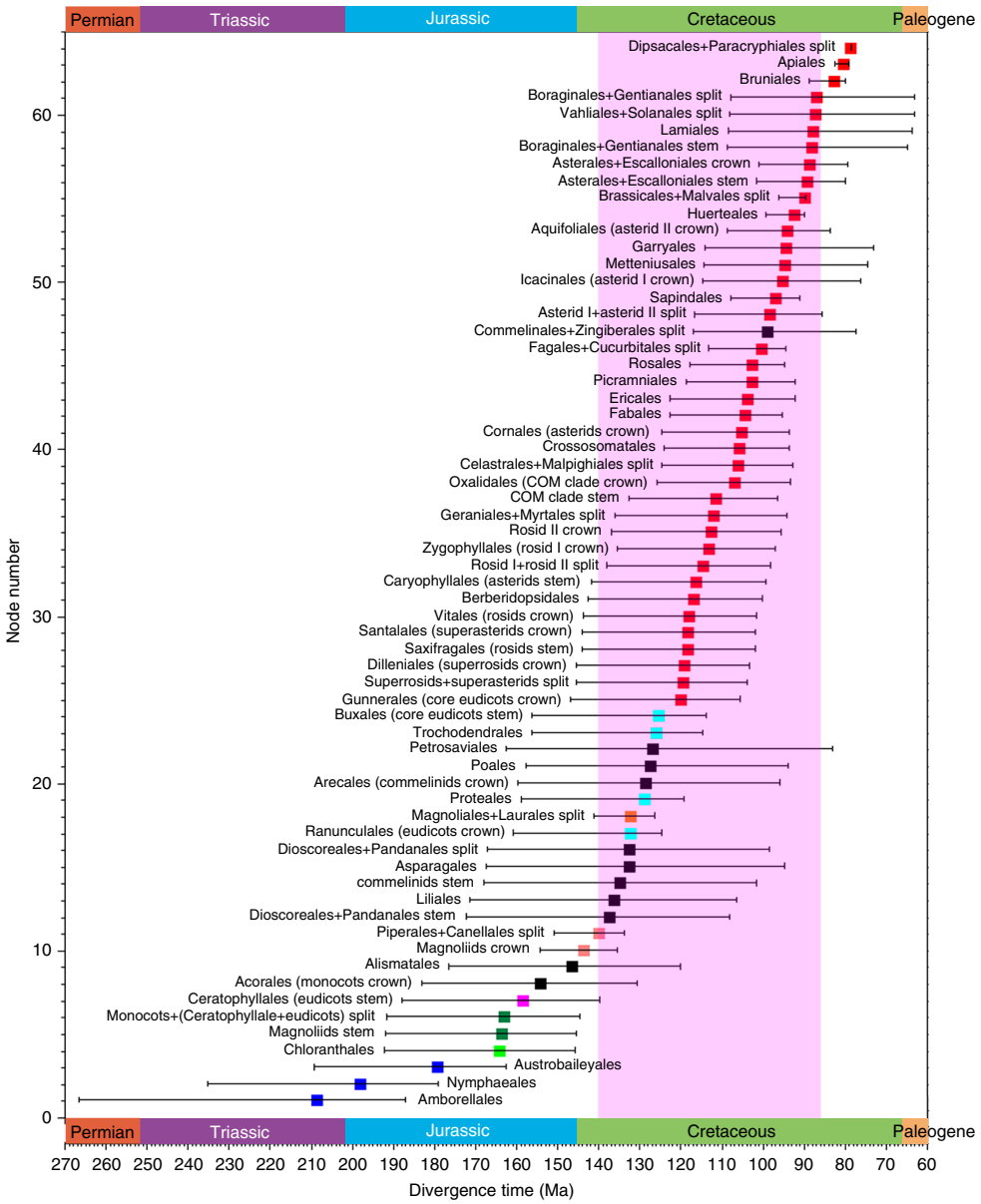

Fig. 3 | Sorted ordinal and interordinal node age estimates using TreePL based on the phylogenetic trees of 80 genes of 2,881 angiosperm plastomes with maximum likelihood analysis. For each node age estimate, the median and range of the upper and lower confidence intervals are depicted (*x* axis indicates divergence time (Ma)) (n = 1,000 bootstrap replications). The colour coding represents the major clades as in Fig. 1. The pink shaded region represents the period (140–86 Ma) in which 80% of the ordinal and interordinal angiosperm clades diversified.

firmest in the palaeontological record, with tricolpate pollen abundant globally at this point but not before; however, the global distribution of tricolpate pollen suggests that eudicots should have originated before this date to allow sufficient time for their spread across the continents. The lower boundary of this estimate is also close to the origin and diversification of bees at 124 (147–93) Ma[5]. Within eudicots, Ranunculales diverged in the Hauterivian, Lower Cretaceous, followed by Proteales ~129 (159–120) Ma, Trochodendrales ~126 (157–115) Ma and Buxales ~125 (157–114) Ma. This narrow range of ages suggests a rapid diversification of the early eudicots, in agreement with the results from other studies[20,21]. Moreover, the rapid diversification of mesangiosperms corresponds with the contemporary radiation of pollinators, herbivores and their predators during the Cretaceous[20,34,35]. For example, some major clades of holometabolous insects, such as Diptera (true flies), Hymenoptera (bees, ants and pollen wasps) and Lepidoptera (butterflies and moths), underwent expansions from the Upper Jurassic to the Lower Cretaceous[4,5,36,37].

In our analyses, clades corresponding to approximately 90% (36/40) of the core eudicot orders originated from the earliest Aptian (125 Ma) to the Coniacian–Santonian boundary (86.3 Ma) in the Cretaceous (Fig. 3), an interval that corresponds closely with the extensive angiosperm fossil record from this period (Table 1). Eleven of the 39 (28%) core eudicot clades at the ordinal level or higher diversified in the Aptian (125–113 Ma). Four campanulid orders, that is, Bruniales, Apiales, Dipsacales and Paracryphiales, are the most recent angiosperm orders, diverging in the mid-Campanian between 82.7 and 78.9 Ma (Fig. 3 and Supplementary Table 3). By this time, all angiosperm clades corresponding to orders and 139 (39%) of the 353 sampled families had appeared (Supplementary Tables 4, 5 and Supplementary Fig. 6). This period is also thought to have been critical for the

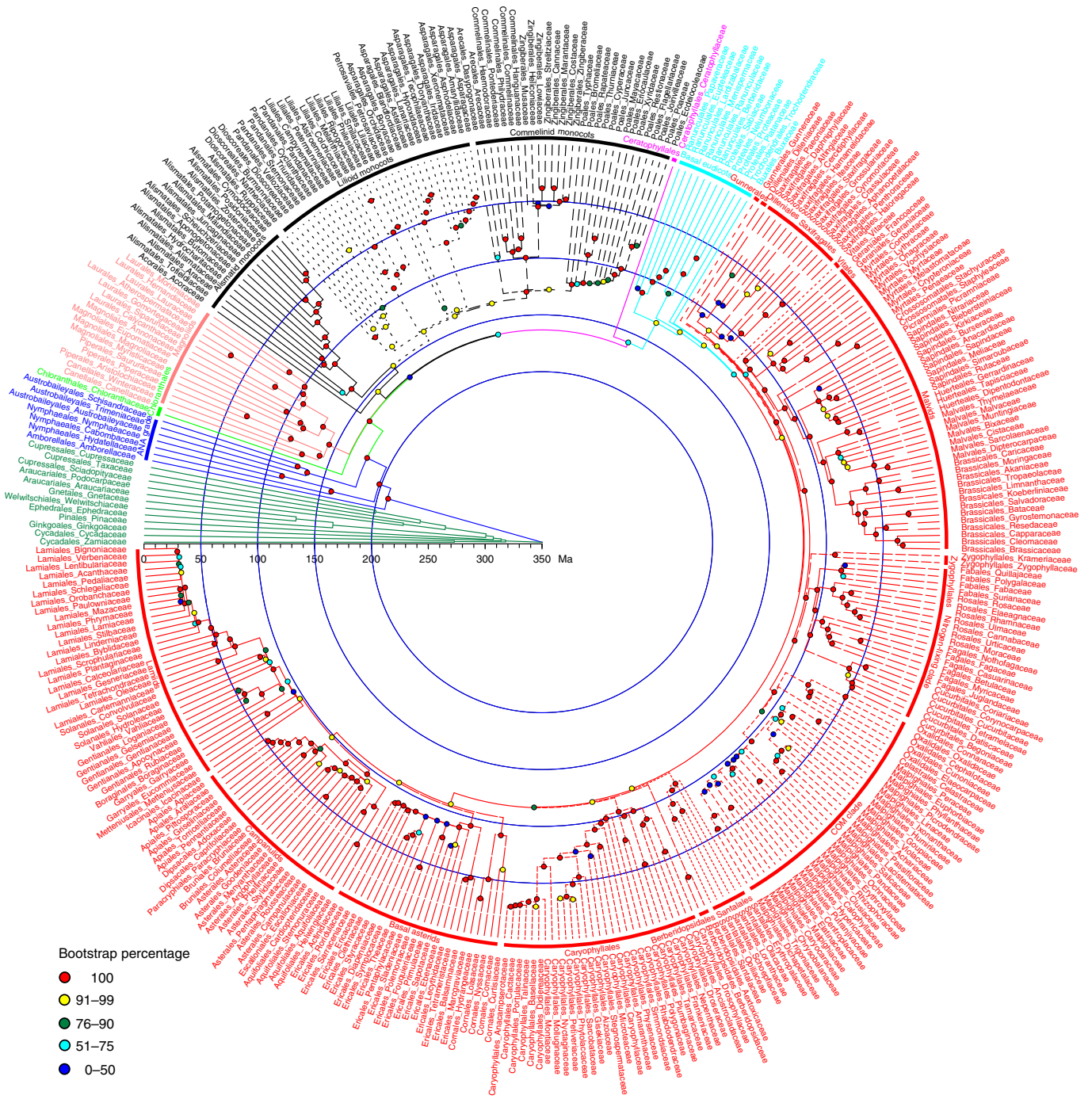

Fig. 4 | Dated family-level angiosperm phylogenetic tree based on 80 genes of 2,881 plastomes with maximum likelihood analysis. Colour coding represents the major clades, as in Fig. 1. The dashed lines indicate major clades within monocots and core eudicots.

development of the ecological dominance of angiosperms in terrestrial ecosystems, together with the co-diversification of insects, birds, and mammals that are key components of contemporary terrestrial biodiversity[2,35–39].

Emergence of the angiosperms was a major biotic transition in the Earth's history due to the intrinsic innovations of this highly successful lineage[1], as well as extrinsic ecological opportunities—for example, drastic climate change in the Triassic and Cretaceous driven by the break-up of Pangaea[40]. Our dating of the origin of the angiosperms provides an estimate in the Upper Triassic (209 Ma). The end of the Triassic witnessed mass extinction due to increased volcanism associated with the opening of the Atlantic Ocean[41], a fourfold increase in atmospheric CO_2 levels[42] and a consequent rise in global temperatures of between 3 and 6 °C (ref. [43]). In considering other organisms, familial-level diversity of insects increased steadily from the Triassic to the Jurassic[32]. Although angiosperm fossils reported from the Triassic and Jurassic need to be critically evaluated[29,30,44], the dense species and gene sampling in the PPA tree permits a more robust overall evaluation of the origin and early evolution of angiosperms than in trees from previous studies, and the dates derived from this analysis reveal a Jurassic gap between the fossil record and molecular-based dates of divergence.

This gap could be due to the relatively sparse occurrence of early angiosperms in the vegetation, their unsuitability for preservation as fossils[28], the minuteness of their flowers as seen in fossils from the early Cretaceous[31] or a combination of these factors. Although the Jurassic angiosperm gap needs to be further verified with nuclear sequence data, even greater taxon sampling and possibly alternative methods, the pattern of molecular ages older than fossil ages is fairly typical[4,45], and we therefore expect the ultimate recognition of angiosperms in the Jurassic or earlier through discovery of credible palaeobotanic evidence. Recent fossil evidence confirms that some conifer lineages are much older than indicated previously by the fossil record, in agreement with older, molecular-based ages[46]. Such fossil discoveries therefore indicate that molecular estimates may often be valid, and we hope that novel discovery or further studies of pre-Cretaceous fossils that share clearer synapomorphies with angiosperms[29] could fill the Jurassic gap.

Methods

Taxon sampling. To maximally capture the diversity of angiosperms, particularly at the ordinal and familial levels, 2,881 plastomes from 2,514 species representing all 72 orders, 365 families and 1,467 genera of seed plants were used for analysis. The analysis included published angiosperm plastomes in NCBI (last accessed 1 January 2017). For most of the monospecific families and/or orders (that is, those comprising a single species), two or more individuals per species were included. We included 2,694 samples representing 2,351 species and 1,390 genera from all 64 orders and 353 (85%) of the 416 flowering plant families recognized by APG IV (see Supplementary Table 1). In all, 163 species (187 samples) representing eight orders, 12 families and 77 genera of gymnosperms formed the outgroup. We generated sequences for 1,659 species from 63 orders and 347 families (see Supplementary Table 1), together with 677 plastomes from the OneKP Project[17] (onekp.com) (see Supplementary Table 1).

Plastome sequencing and data assembly. Fresh plant material (mostly young leaves) was collected and immediately transferred to the molecular biology laboratory of the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, for genomic DNA isolation. Four methods were used to obtain plastome data. (1) Total genomic DNA was extracted from 100 mg of actively growing fresh leaves with a modified hexadecyltrimethylammonium bromide (CTAB) method[47], for which 4% rather than 2% CTAB was used and 1% polyvinyl polypyrrolidone and 0.2% DL-dithiothreitol were added. The plastomes were then amplified with the long-range PCR method given in ref. [48] or ref. [49] for sequencing. (2) Total genomic DNA was used directly for sequencing. (3) Total DNA enriched for plastome extraction from 100 g fresh leaves was obtained as in ref. [50] for sequencing. (4) Young leaves were collected and frozen at −80 °C, and then total RNA was isolated and complement DNA synthesized for transcriptome sequencing[17,51]. Purified cDNA (6 μg) was fragmented and used to construct short-insert (500 bp) libraries as per the manufacturer's manual (Illumina). cDNA from each individual was indexed with tags, and cDNA samples were pooled for sequencing with either the Illumina HiSeq 2000 or Illumina MiSeq. High-quality Illumina sequencing reads were assembled into scaffolds with de novo sequence assembly software Spades[52], SOAPdenovo[53] and CLC Genomics Workbench v.6.5 (CLC Bio), respectively. Plastomes were assembled and annotated as in ref. [48].

Phylogenetic analysis. Because high-sequence variability of *infA*, *ycf1* and *ycf2* among seed plants made reliable alignment difficult, these genes were excluded from phylogenetic analysis. The exons of all other protein-coding genes and the ribosomal RNA genes were extracted from plastid genome sequences. We aligned these 80 genes with MAFFT[54] and MUSCLE[55]. Some poorly aligned regions were manually adjusted in GENEIOUS v.9.1 (ref. [56]). All aligned genes were concatenated into a supermatrix. To test the effect of the most rapidly evolving sites on phylogenetic relationships, we examined the 80-gene matrix and removed the most rapidly evolving sites using Gblocks 0.91 b (ref. [57]) with the default setting. Details for the sites removed for each gene are presented in Supplementary Table 2.

Maximum likelihood analyses were performed with RAxML v.8 (ref. [58]) under the GTRGAMMA model for a partitioned supermatrix. Searches for the best trees were conducted by starting from 1,000 random trees, and bootstrap percentages were obtained with 1,000 non-parametric bootstrap replicates.

To further evaluate the phylogenetic relationships of the backbone tree of angiosperms, especially among the five major clades (magnoliids, Chloranthales, monocots, Ceratophyllales and eudicots) of mesangiosperms, we assembled a smaller data subset of 82 taxa (five gymnosperms and 77 angiosperms) derived from the complete taxon sampling. These 77 angiosperms were obtained by selecting one representative from each of the 64 orders and up to three taxa from the early-diverging clades of angiosperms (Supplementary Fig. 4). The sequences of plastid protein-coding exons and rRNA genes were aligned and manually adjusted as above, with a final alignment of 60,853 bp. Maximum likelihood analyses were conducted with RAxML under different partitioning schemes: unpartitioned, partitioned by different genes (76 partitions, 75 unique protein-coding genes and rRNA genes forming one further partition), and partitioned optimally as determined with PartitionFinder v.2.0 (ref. [59]) software. A total of 40 partitions was selected in a heuristic search by running PartitionFinder. We also performed unpartitioned maximum likelihood analyses on the combined first and second codon positions, third codon positions and amino acid sequences. The GTRGAMMA model was used for all nucleotide analyses, and the DUMMY2 model for amino acid alignment.

Divergence time estimates. Penalized likelihood dating analysis was conducted in TreePL[60] with the optimal maximum likelihood tree and 1,000 bootstrap replicates. All bootstrap replicates were analysed with RAxML to obtain branch lengths, followed by dating with TreePL. Sixty-two calibration points were identified from fossil and geological data obtained from the literature (Table 1). To identify the appropriate level of rate heterogeneity in the phylogram, cross-validation with TreePL was conducted in a test of 37 values for the smoothing parameter from 10^{18} to 10^{-18}, and resulted in an optimal smoothing parameter of 0.001. The configuration file for running TreePL is provided in the Supplementary Information. Confidence intervals for the dating estimates were calculated from the 1,000 bootstrap replicates, with branch lengths generated by RAxML by using TreeAnnotator as implemented in BEAST[61].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequence alignments underlying analyses and all trees are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.bq091cg.

Received: 25 August 2018; Accepted: 2 April 2019;
Published online: 06 May 2019

References

1. Friis, E. M., Crane, P. R. & Pedersen, K. R. *Early Flowers and Angiosperm Evolution* (Cambridge Univ. Press, 2011).
2. Benton, M. J. The origins of modern biodiversity on land. *Phil. Trans. R. Soc. Lond. B* **365**, 3667–3679 (2010).
3. Darwin, C. in *More Letters of Charles Darwin* Vol. 12 (eds Darwin, F. & Seward, A. C.) Vol. 2, 12–13 (John Murray, 1903).
4. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
5. Peters, R. S. et al. Evolutionary history of the Hymenoptera. *Curr. Biol.* **27**, 1013–1018 (2017).
6. Roelants, K. et al. Global patterns of diversification in the history of modern amphibians. *Proc. Natl Acad. Sci. USA* **104**, 887–892 (2007).
7. Bininda-Emonds, O. R. P. et al. The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
8. Schneider, H. et al. Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557 (2004).
9. Davis, C. C., Xi, Z. & Mathews, S. Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biol.* **12**, 11 (2014).
10. Angiosperm Phylogeny Group IV. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
11. Qiu, Y.-L. et al. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
12. Soltis, D. E. et al. Angiosperms phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **98**, 704–730 (2011).
13. Moore, M. J., Bell, C. D., Soltis, P. S. & Soltis, D. E. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl Acad. Sci. USA* **104**, 19363–19368 (2007).
14. Cantino, P. D. et al. Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* **56**, 822–846 (2007).
15. Zeng, L. et al. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Comm.* **5**, 4956 (2014).
16. Wickett, N. J. et al. Phylotranscriptomic analysis of the origin and early diversification land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868 (2014).
17. Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Ruhfel, B. R. & Solits, D. E. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *Am. J. Bot.* **105**, 291–301 (2018).
18. Drew, B. T. et al. Another look at the root of the angiosperms reveals a familiar tale. *Syst. Biol.* **63**, 368–382 (2014).
19. Coiro, M., Doyle, J. A. & Hilton, J. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytol.* https://doi.org/10.1111/nph.15708 (2019).

20. Barba-Montoya, J., Dos Reis, M., Schneider, H., Donoghue, P. C. J. & Yang, Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of Cretaceous terrestrial revolution. *New Phytol.* **218**, 819–834 (2018).
21. Foster, C. S. P. et al. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* **66**, 338–351 (2017).
22. Magallón, S., Hilu, K. W. & Quandt, D. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am. J. Bot.* **100**, 556–573 (2013).
23. Murat, F., Armero, A., Pont, C., Klopp, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* **49**, 490–496 (2017).
24. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
25. Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-revisited. *Am. J. Bot.* **97**, 1296–1303 (2010).
26. Beaulieu, J. M., O'Meara, B. C., Crane, P. & Donoghue, M. J. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Syst. Biol.* **64**, 869–878 (2015).
27. Zanne, A. E. et al. Three keys to the radiation of angiosperms into freezing enviroments. *Nature* **506**, 89–92 (2014).
28. Doyle, J. A. Molecular and fossil evidence on the origin of angiosperms. *Annu. Rev. Earth Planet. Sci.* **40**, 301–326 (2012).
29. Hochuli, P. A. & Feist-Burkhardt, S. Angiosperm-like pollen and *Afropollis* from the Middle Triassic (Anisian) of the Germanic Basin (northern Switzerland). *Front. Plant Sci.* **4**, 344 (2013).
30. Herendeen, P. S., Peter, B. G. & Snapp, S. S. Palaeobotanical redux: revisiting the age of the angiosperms. *Nat. Plants* **3**, 17015 (2017).
31. Friis, E. M., Crane, P. R., Pedersen, K. R., Stampanoni, M. & Marone, F. Exceptional preservation of tiny embryos documents seed dormancy in early angiosperms. *Nature* **528**, 551–554 (2015).
32. Labandeira, C. C. in *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life* (ed. Pontarotti, P.) 261–299 (Springer, 2014).
33. Davies, T. J. et al. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc. Natl Acad. Sci. USA* **101**, 1904–1909 (2004).
34. Dilcher, D. Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. *Proc. Natl Acad. Sci. USA* **97**, 7030–7036 (2000).
35. Meredith, R. W. et al. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
36. Cardinal, S. & Danforth, B. N. Bees diversified in the age of eudicots. *P. R. Soc. B* **280**, 20122686 (2013).
37. Moreau, C. S., Bell, C. D., Vila, R., Archibald, S. B. & Pierce, N. E. Phylogeny of the ants: diversification in the age of angiosperms. *Science* **312**, 101–104 (2006).
38. Augusto, L., Davies, T. J., Delzon, S. & De Schrijver, A. The enigma of the rise of angiosperms: can we untie the knot? *Ecol. Lett.* **17**, 1326–1338 (2014).
39. Wang, H. C. et al. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl Acad. Sci. USA* **106**, 3853–3858 (2009).
40. Chaboureau, A. C., Sepulchre, P., Donnadieu, Y. & Franc, A. Tectonic-driven climate change and the diversification of angiosperms. *Proc. Natl Acad. Sci. USA* **111**, 14066–14070 (2014).
41. Deenen, M. H. L. et al. A new chronology for the end-Triassic mass extinction. *Earth Planet. Sci. Lett.* **291**, 113–125 (2010).
42. Huynh, T. T. & Poulsen, C. J. Rising atmospheric CO_2 as a possible trigger for the end-Triassic mass extinction. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **217**, 223–242 (2005).
43. McElwain, J. C., Beerling, D. J. & Woodward, F. I. Fossil plants and global warming at the Triassic–Jurassic boundary. *Science* **285**, 1386–1390 (1999).
44. Ren, D. Flower-associated Brachycera flies as fossil evidence for Jurassic angiosperm origins. *Science* **280**, 85–88 (1998).
45. Nel, A. et al. The earliest known holometabolous insects. *Nature* **503**, 257–261 (2013).
46. Blomenkemper, P., Kerp, H., Hamad, A. A., DiMichele, W. A. & Bomfleur, B. A hidden cradle of plant evolution in Permian tropical lowlands. *Science* **362**, 1414–1416 (2018).
47. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
48. Yang, J.-B., Li, D.-Z. & Li, H.-T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol. Ecol. Resour.* **14**, 1024–1031 (2014).
49. Zhang, T., Zeng, C.-X., Yang, J.-B., Li, H.-T. & Li, D.-Z. Fifteen novel universal primer pairs for sequencing whole chloroplast genomes and a primer pair for nuclear ribosomal DNAs. *J. Syst. Evol.* **54**, 219–229 (2016).
50. Zhang, Y.-J., Ma, P.-F. & Li, D.-Z. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE* **6**, e20596 (2011).
51. Huang, C.-H. et al. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* **33**, 394–412 (2016).
52. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
53. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
54. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
55. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
56. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
57. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
58. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
59. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
60. Smith, S. A. & O'Meara, B. C. TreePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).
61. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
62. Bell, C. D., Soltis, D. E. & Soltis, P. S. The age of the angiosperms: a molecular timescale without a clock. *Evolution* **59**, 1245–1258 (2005).
63. Magallón, S. & Castillo, A. Angiosperm diversification through time. *Am. J. Bot.* **96**, 349–365 (2009).
64. Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl Acad. Sci. USA* **107**, 5897–5902 (2010).
65. Clarke, J. T., Warnock, R. C. M. & Donoghue, C. J. Establishing a time-scale for plant evolution. *New Phytol.* **192**, 266–301 (2011).

Acknowledgements

We thank the Germplasm Bank of Wild Species at the Kunming Institute of Botany (KIB) for facilitating this study; the curators and staff of the Beijing Botanical Garden (BG), Blue Mountains BG, Brisbane BG, Kunming BG, Missouri BG, Wuhan BG, Royal BG Edinburgh, RBG Kew, RBG Sydney, RBG Victoria (both Melbourne and Cranbourne), San Francisco BG, Shanghai Chenshan BG, South China BG, UC Berkeley BG, Xianhu BG Shenzhen, Xishuangbanna Tropical BG, Yinchuan BG and O. Maurin (Johannesburg, now Kew), J. R. Shevock (California), Y.-M. Shui (Kunming), and N. Zamora (Costa Rica) for samples; and S. R. Manchester (Florida) for critical discussion on fossil selection and calibration. This work was funded by the Strategic Priority Research Programme of the Chinese Academy of Sciences (CAS) (grant No. XDB31000000 to D.-Z.L.), CAS' Large-scale Scientific Facilities (grant No. 2017-LSF-GBOWS-02 to D.-Z.L. and J.-B.Y.), KIB's iFlora initiative (grant No. 2014-4-11 to D.-Z.L.) and the National Natural Science Foundation of China (grant No. 31570333 to H.-T.L.). P.-F.M. was supported by CAS' Youth Innovation Promotion Association (grant No. 2015321) and P.S.S. was supported by the Ten Thousand Talents Programme of China and the Yunling International High-end Experts Programme of Yunnan Province.

Author contributions

D.-Z.L, J.-B.Y., H.-T.L., T.-S.Y., D.E.S. and P.S.S. conceived the project and designed the research. T.-S.Y., T.Z., J.C., L.-M.G. and S.-D.Z. designed and carried out field collection work. Q.-F.W., J.W., P.W.F., M.v.d.B., P.M.H. and M.W.C. provided and/or collected samples. J.-B.Y., H.-T.L., Z.-R.Z., C.-N.F. and J.Y. performed DNA laboratory work. M.A.G., D.E.S. and P.S.S. prepared the OneKP dataset. H.-T.L., L.-M.G., T.-S.Y., P.-F.M., D.E.S. and P.S.S. designed and coordinated computational analyses. H.-T.L., T.Z., J.C., Y.L. and H.W. prepared the Figures and Tables. T.-S.Y., Y.L., L.-M.G., P.-F.M., D.-Z.L. and H.W. wrote the supplementary information. D.-Z.L., T.-S.Y., L.-M.G., P.-F.M. and P.W.F. wrote the first manuscript draft with input from all co-authors, particularly P.S.S., M.W.C., D.E.S. and P.M.H.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-019-0421-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.S.S. or D.-Z.L.

Journal peer review information: *Nature Plants* thanks Jennifer Mandel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): De-Zhu Li, Pamela S. Soltis

Last updated by author(s): Mar 25, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used in data collection.
Data analysis	Softwares used for data analysis include SPAdes, SOAPdenovo 2, CLC Genomics Workbench v. 6.5, MAFFT v. 5, MUSCLE, GENEIOUS v.9.1, Gblocks v. 0.91b, RAxML v. 8, PartitionFinder v.2.0, TreePL, and BEAST 2.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence alignments underlying analyses and all phylogenetic trees in the study are available on DRYAD (doi:10.5061/dryad.bq091cg).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	To maximally capture the diversity of angiosperms, we sampled 2514 species from 72 orders and 365 families for analyses.
Research sample	We clearly described our study samples and the details for these 2881 samples of 2514 species are included in Supplementary Table 1.
Sampling strategy	We conducted sampling base on the previous studies and attempted to cover as many samples of angiosperms as possible at the family level. Our sampling covered all orders and 85% families of extant angiosperms that are sufficient for questions addressed in this study.
Data collection	Data collection was based on our own samples and the available angiosperms plastid genomes in the GenBank.
Timing and spatial scale	We addressed geological historical timing (up to the Triassic) and world-wide spatial scale to address the questions.
Data exclusions	No data were excluded.
Reproducibility	The phylogenetic results and divergence time estimates can be reproducible from the deposited data with the published methods.
Randomization	Sampling was designed to capture the diversity of angiosperm families, while species sampling is random.
Blinding	Blinding is not relevant to our study as we performed phylogenetic/evolutionary analyses of plants.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	We collected plant specimens with small-amount (5 g) of young leaves for DNA isolation where field conditions are not relevant.
Location	We sampled species worldwide and the location information of samples can be found in Supplementary Table S1.
Access and import/export	As sampling was non-damaging (5 g of young leaves) and collection permits in botanic gardens and/or natural habitats were obtained where necessary.
Disturbance	Only limited plant material (5 g of young leaves) were collected for DNA isolation with a voucher specimen. And disturbance to natural habitats and/or cultivated plants was minimized..

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging