Lin *et al. BMC Biology (2025) 23:45*
https://doi.org/10.1186/s12915-025-02135-9

BMC Biology

RESEARCH

Open Access

Phylogenomics of angiosperms based on mitochondrial genes: insights into deep node relationships

Dongliang Lin[1,2,3], Bingyi Shao[1], Zhiyuan Gao[1], Jianwu Li[4], Zhanghai Li[1], Tingyu Li[2,5], Weichang Huang[6], Xin Zhong[6], Chao Xu[1,3], Mark W. Chase[7,8*] and Xiaohua Jin[1,3*]

Abstract

Background Angiosperms are the largest plant group and play an essential role in the biosphere. Phylogenetic relationships of many families and orders remain contentious, and, in an attempt to address these, we performed the most extensive sampling of mitochondrial genes to date.

Results We reconstructed a seed plant phylogenetic framework based on 41 mitochondrial protein-coding sequences (mtCDSs), representing 335 families and 63 orders with 481 angiosperm species. The results for major clades of angiosperms produced moderate to strong support (> 70% bootstrap) for more than 80% of nodes and strong support for most orders. Eight major nodes were supported, including the three paraphyletic ANA orders (Amborellales, Nymphaeales, and Austrobaileyales) and five major core-angiosperm clades. Chloranthales and Ceratophyllales are sister to the eudicots, whereas the monocots are sister to the magnoliids. Although well-supported, relationships within the asterids and rosids were in some cases unresolved or weakly supported, due to the low levels of variability detected in these genes.

Conclusions Our results indicated that mitochondrial genomic data were effective at resolving deep node relationships of angiosperm phylogeny and thus represent an important resource for phylogenetics and evolutionary studies of angiosperm.

Keywords Phylogenomics, Amborellales, Chloranthales, Ceratophyllales, Magnoliids, Mitochondrial genes, Nymphaeales

*Correspondence:
Mark W. Chase
m.chase@kew.org
Xiaohua Jin
xiaohuajin@ibcas.ac.cn
[1] State Key Laboratory of Plant Diversity and Speciality Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] China National Botanical Garden, Beijing, China
[4] Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan 666303, China
[5] Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China
[6] Shanghai Chenshan Botanical Garden, Chenhua Road 3888, Songjiang, Shanghai 201602, China
[7] The Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK
[8] Department of Environment and Agriculture, Curtin University, Bentley, WA 6102, Australia

Background

Angiosperms are the largest plant group, with complex morphologies and ecologies, and occupy almost all ecosystems, terrestrial and marine [1]. They play an essential role in the biosphere. Knowledge of angiosperm phylogenetics can provide a framework for understanding the history and composition of major terrestrial ecosystems and deciphering the general patterns of biodiversity [1–3].

To resolve angiosperm phylogenetic relationships, dozens of studies have focused on large-scale genomic datasets and broad taxonomic sampling, including plastome data, low-copy, homologous nuclear genes [4–7], and transcriptomes [8]. These studies have greatly improved our understanding of phylogenetic relationships for angiosperms, created a phylogenetic framework for other kinds of studies, and contributed significantly to the overall understanding of extant angiosperm diversity. Nevertheless, some relationships have been inconsistent among these studies [3, 6, 9–26], and it has been suggested that systematic errors (e.g., substitutional saturation) are an important factor influencing nucleus-plastid conflicts [3, 6, 9, 27–29]. Moreover, plastid genes exhibit codon-usage biases and the highest heterogeneity in GC content [6, 28, 30–32].

Based on the inferred phylogenetic relationships, it was often unclear whether the sister clade of extant angiosperms is Amborellales alone (e.g., [5, 6, 33–36]) or Amborellales + Nymphaeales (e.g., [20, 30, 32, 37–40]). Recent studies have identified Amborellales as the sister group to all other angiosperms and placed Nymphaeales and Austrobaileyales as successive sisters to the remaining extant angiosperms [3, 7, 22, 27, 41]. The remaining angiosperms, referred to as the mesangiosperms, a highly supported clade, were resolved into five clades: eudicots (Eud), monocots (Mon), magnoliids (Mag), Chloranthales (Chl), and Ceratophyllales (Cer). The relationships among these five clades of mesangiosperms has remained problematic and unsettled. Different topologies have been proposed based on different datasets (Additional file 1: Table S1, Additional file 2: Fig. S1): (Mag-Chl)-(Mon-(Eud-Cer)) [30, 31, 37, 42, 43]; Chl-(Mag-(Mon-(Eud-Cer)), Li et al. [3] based on 2881 plastid genomes; (Cer-Chl)-(Mag-(Mon-Eud)), Qiu et al. [20] based on four mitochondrial genes; Mon-(Mag-(Eud-(Cer-Chl))) and Mon-((Mag-Chl)-(Eud-Cer)) [22] based on transcriptomic data from 1124 species; and (Chl-Mag)-(Cer-(Mon-Eud)), Zuntini et al. [26] based on multiple nuclear genes in the Angiosperms353 probe set.

Early diversification of angiosperms occurred over a short geological period, leading to a rapid increase in angiosperms during the Early Cretaceous and their gradual emergence as a major component of terrestrial ecosystems [1, 44, 45], a phenomenon often referred to as Darwin's 'abominable mystery' [46, 47]. According to recent molecular dating studies, angiosperms arose in the Upper Triassic (ca. 237.0–201.3 million years ago [Ma]) [3, 6, 48–50], and the Mesangiospermae in the Jurassic (201.3–145 Ma)) [3, 6, 50]. Alternatively, other estimates have been reported for the origin of angiosperms: 242–241 Ma, 194 Ma [51], 170 Ma [32], and 137–135 Ma [52].

The mitochondrial genome (mitogenome) is one of the three genomes in plants, and mitochondrial protein-coding genes are characterized by a slow evolutionary rate [53–55]. Mitogenomes are a rich source of evolutionary data for well-sampled phylogenetic analyses [28]. In addition, because of rare duplications [56] and slow evolutionary rate, mitochondrial genes exhibit clear orthologous relationships [20, 28, 56]. Yet, the phylogenomic analysis of mitochondrial data has rarely been included in plant studies, possibly because the mitochondrial genome has a complex structure, and is more challenging to assemble than the relatively simple structure of the plastid genome. Only a few mitochondrial genes/sequences, e.g., *matR*, *atp1*, *mtSSU*, *mtLSU*, *nad5*, and *rps3*, have been used for phylogenetic studies to date [13, 20, 40, 57, 58]. The use of mitochondrial DNA in angiosperm phylogenetic analyses has been gradually increasing [28, 56, 59]. In this study, we analyzed 41 mitochondrial protein-coding sequences (mtCDSs) for 481 species representing 355 families and 63 orders of angiosperms, sensu APG IV (2016), representing the largest sampling of angiosperm mitochondrial genes dataset to date. We aimed to (1) reconstruct angiosperm phylogenetic relationships, (2) determine if there is conflict between the plastid and mitochondrial datasets, and (3) improve our understanding of the temporal evolution of angiosperms.

Results

MTCDS datasets

A total of 481 mitochondrial genomic data of angiosperms, representing 335 families and 63 orders of angiosperms, were included in this study, covering the three ANA and five major core angiosperm clades and almost all angiosperm orders. Additionally, five mitochondrial genomic data from five families representing five major clades of gymnosperms had been included as outgroups [60]. This represents the largest sampling of angiosperm mitochondrial genes dataset to date. We newly sequenced 91 species from 87 families (Additional file 1: Tables S2, S3) and assembled the mtCDSs, and we assembled those of 203 species in 176 families based on data downloaded from NCBI (http://www.ncbi.nlm.nih.gov) (Additional file 1: Table S2). The mtCDSs of 192 species (77 families, including five species of gymnosperms) were collected from the mitochondrial genome sequences

available at NCBI (Additional file 1:Table S2). A total of 17,338 mtCDSs were obtained, with approximately 13% gaps/missing data.

Phylogenetic analyses

Phylogenetic analyses yielded well-supported and largely congruent results across the two datasets and methods (Fig. 1; Additional file 2: Figs. S2-S7). With IQ-TREE, support was medium to high (BP > 70%) for more than 80% of the nodes in the 41-MTCDS concatenation analysis (Additional file 2: Fig. S7) and was higher in RAxML tree for most nodes. Support also increased in the 24-MTCDS analysis (Additional file 2: Figs. S2, S5). Fifty-seven orders were monophyletic with high support (BP > 80%). Five orders (Proteales, Saxifragales, Bruniales, Geraniales, and Oxalidales) were not monophyletic in the 41-MTCDS tree, but Proteales and Saxifragales were monophyletic in the 24-MTCDS tree (Additional file 2: Fig. S2). Here, we show the 41-mtCDS tree (Fig. 1; Additional file 2: Fig. S3) and illustrate the conflicts between 41-mtCDS and 24-MTCDS trees (Fig. 2).

Amborellales were sister to all other extant angiosperms, followed successively by Nymphaeales and Austrobaileyales (the ANA grade; Fig. 1). A different outcome was observed in the relative positions of Nymphaeales and Amborellales in 24-MTCDS tree (Fig. 2; Additional file 2: Fig. S2), but these positions were only weakly supported. The remaining angiosperms (known as mesangiosperms) were monophyletic with high support and divided into five major clades. The 41-MTCDS trees and the ASTRAL tree (Fig. 1; Additional file 2: Fig. S6) consistently placed Ceratophyllales as sister to Chloranthales, whereas the monocots were sister to the magnoliids in the 41-MTCDS tree (Additional file 2: Fig. S3). The monocots and magnoliids were successively sister to Ceratophyllales + Chloranthales in the ASTRAL tree (Additional file 2: Fig. S4). Ceratophyllales and Chloranthales formed a sister clade to magnoliids, which together were sister to monocots (Additional file 2: Fig. S4).

In magnoliids, Piperales, Canellales, and Laurales were successively sister to Magnoliales in the 41-MTCDS (RAxML) and ASTRAL trees (Fig. 1, Additional file 2: Figs. S3, S4), consistent with the results of previous studies [8]. Inconsistently, Piperales and Canellales were sister to Laurales + Magnoliales with high support in the 24-MTCDS tree (Additional file 2: Figs. S2, S7).

In monocots, Acorales, Alismatales, and Petrosaviales were successively sister to the remaining monocots along the monocot backbone (Fig. 1). Pandanales and Dioscoreales were sister groups with high support, followed successively by Asparagales, Liliales, and commelinids along the monocot spine. Among the commelinids, Arecales and Poales were successive sisters to the clade formed by Commelinales and Zingiberales (Fig. 1). By contrast, Arecales were resolved as sister to Poales with weak support in the 24-MTCDS tree (Fig. 2).

Among eudicots, Ranuncunales, Proteales, Trochodendrales, and Buxales were successive sisters to the remaining eudicots (known as core eudicots). In the 41-MTCDS tree (by both RAxML and IQ-TREE), Proteales were not supported as monophyletic (Fig. 1; Additional file 2: Figs. S3, S7). Nelumbonaceae formed an isolated clade with low support (BP = 26% by RAxML or 65% by IQ-TREE) along the angiosperm backbone (Additional file 2: Figs. S3, S7). In contrast, Proteales were monophyletic in the 24-MTCDS and ASTRAL trees (Fig. 2; Additional file 2: Figs. S2, S4). Circaeasteraceae were separated from Ranunculales as a clade along the angiosperm backbone in the ASTRAL tree (Additional file 2: Figs. S4, S6). Among the core eudicots, Berberidopsidales and Dilleniales + Gunnerales were the first and second successive branches, respectively, along the backbone, with moderate to strong support (Additional file 2: Fig. S3). Santalales were resolved as a third branch of core eudicots in the 41-MTCDS (RAxML) tree with weak support (Additional file 2: Fig. S3) and with increased support by IQ-TREE (Additional file 2: Fig. S7). However, Francoaceae were resolved as the third branch (with medium support) and Santalales as the fourth branch (with weak support) in the 24-MTCDS tree (Fig. 2; Additional file 2: Figs. S2). The ASTRAL tree differed and placed Circaeasteraceae and Berberidopsidales as the first and second branches, respectively (Additional file 2: Figs. S4, S6). Vitales were sister to Saxifragales, and together they were sister to the remaining core eudicots, which were divided into two major groups: rosids (17 orders) and asterids (16 orders).

In asterids, Cornales and Ericales were resolved as the first and second successive clades, respectively, with moderate support (Fig. 1). Dipsacales-Apiales-Paracryphiales-Escalloniales-Columelliaceae-Bruniales-Asterales together formed the third branch with moderate support, and Icacinales, Garryales, and Aquifoliales formed the fourth successive branch with low support. Metteniusales, Solanales, Boraginales, and Gentianales were successive sisters to Lamiales with high support (Fig. 1).

In rosids, phylogenetic relationships among orders generally received low support (Additional file 2: Figs. S2-S4). The Celastrales-Oxalidales-Malpighiales (COM) clade was recovered as monophyletic (Additional file 2: Fig. S3). Inconsistently, Oxalidales were paraphyletic, with Huaceae placed in various isolated positions away from other Oxalidales with strong support (Fig. 1; Additional file 2: Figs. S2, S3, S5). Caryophyllales were sister to Geraniales away from the asterids with weak support and were an early branching asterid clade in the 24-MTCDS and ASTRAL analyses.

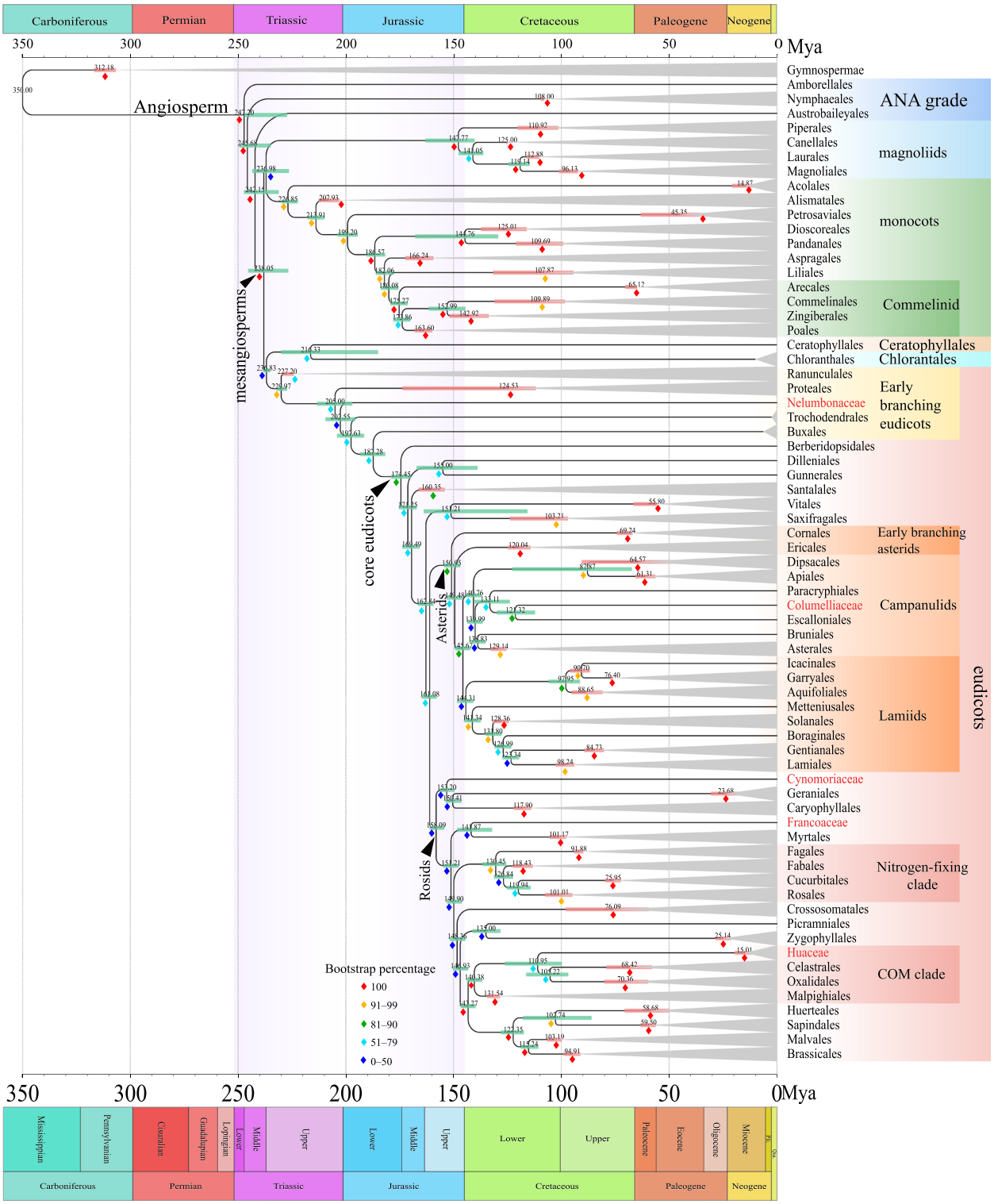

Fig. 1 Time-calibrated tree of angiosperms based on 41 mitochondrial protein-coding sequences (MTCDSs) from 486 sampling. Diamonds under the nodes represent the range of maximum-likelihood (ML) bootstrap percentages (BP). Taxa that separated from the orders recognized by the APG IV classification system are indicated in red font. Values above the branches represent node ages (million years ago [Ma]). Horizontal bars represent the 95% confidence intervals of node ages, with red bars indicating the node ages of orders and cyan bars indicating the node ages of the backbone

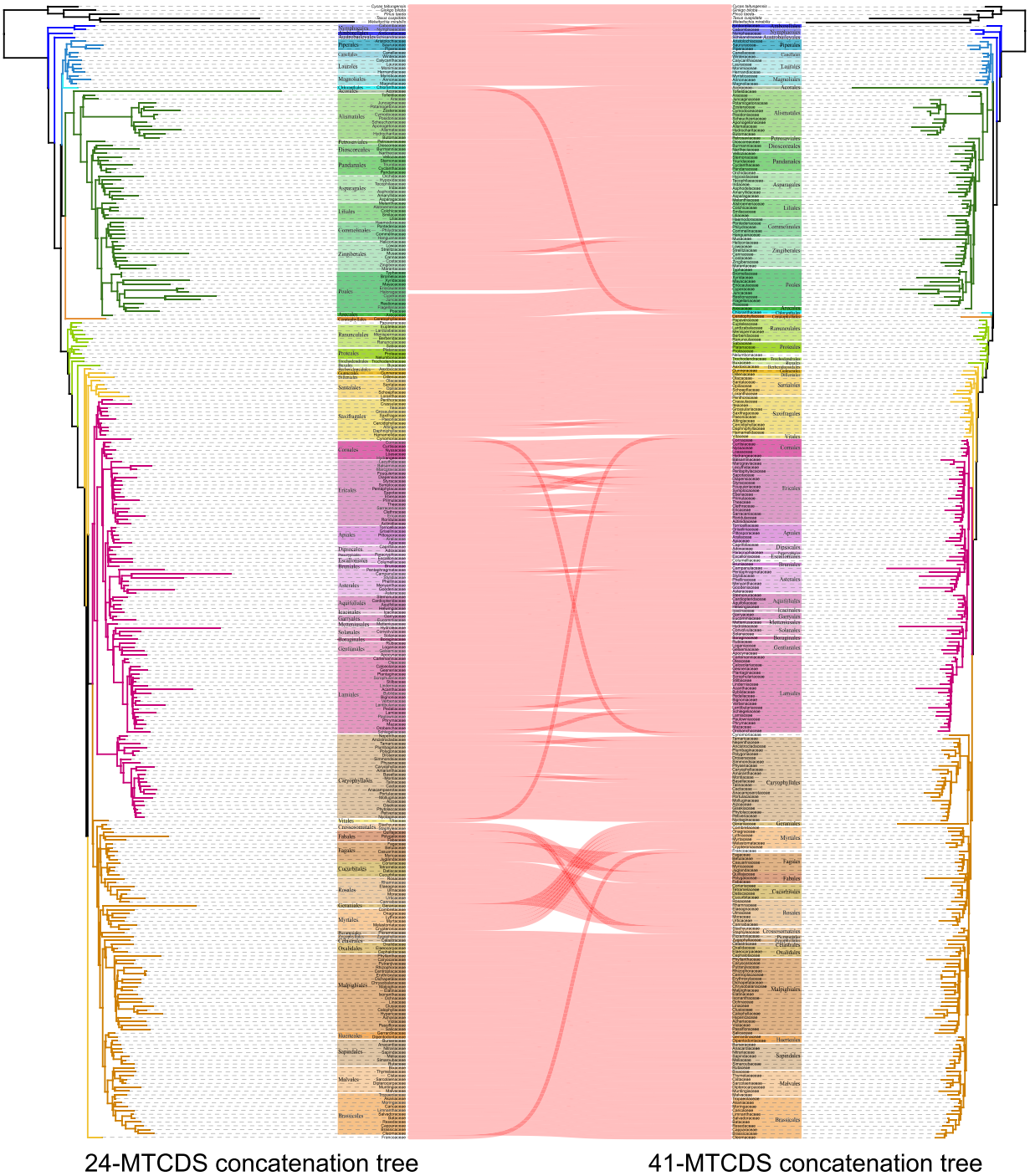

Fig. 2 Discordance at family level in 41 MTCDSs and 24 MTCDSs from 486 sampling. Curved lines between tip labels from in the two trees phylogenies link the same families. Families are represented by a randomly chosen sample from each family

Phylogenetic conflict

To detect phylogenetic conflicts in phylogenetic analysis, especially for nodes with low support, we calculated the internode certainty (ICA) metric for all nodes in each result. ICA measures the degree of certainty based on the frequency of all conflicting bipartitions, and ICA values

of −1, 0, and 1 indicate bipartition with no concordance, equal support, and strong concordance, respectively, with one or more conflicting bipartitions. Monophyly of most orders in 41-MTCDS concatenation tree was supported by ICA (Fig. 3; Additional file 2: Figs. S9-S15). Columelliaceae were highly supported (ICA = 1) as distinct from Bruniales as sister to Escalloniales, Huaceae was distinct from Oxalidales with ICA > 0.5. Few nodes had ICA values near zero, which implies equal support for one or more conflicting bipartitions among these nodes; for example, the ICA values for the positions of Circaeasteraceae, Francoaceae, Cynomoriaceae, and Nelumbonaceae were close to zero.

The proportion of informative quartets was high (mean QI = 0.93; range = 0.89–1.0) and that of concordant quartets was also high (mean QC = 0.54; interquartile range [IQR) = 0.16–1.0) (Additional file 2: Fig. S12). Additionally, the mean QD was low (0.37; IQR = 0.01–0.68), while the mean QF (Quartet Fidelity) was high (0.72; IQR = 0.66–0.8). Like the ICA results, the nodes connecting several major mesangiosperm clades along the backbone received low support (Fig. 3); notably, QC values less than −0.05 occurred in 17.6% (63) of nodes (such as magnoliids + monocots), indicating conflicts.

Discordance between PPA tree and 41-MTCDS concatenation tree was shown by topological differences (Fig. 4). The topological incongruence was 0.212 with information-based generalized Robinson–Foulds distances (from full concordance to full discordance express as information-based generalized Robinson–Foulds distances from 0 to 1). Six major phylogenetic inconsistencies were identified between PPA tree and 41-MTCDS concatenation tree, including relative position of five clades (especially Chl and Cer), early branches of core eudicots, and phylogenetic positions of Caryophyllales, Columelliaceae, Nelumbonaceae, and Huaceae (Fig. 4). Phylogenetic inconsistencies of two orders in monocots and 13 orders in core eudicots were also identified (Fig. 4).

Divergence time estimates

Estimation of angiosperm divergence times is based on the phylogenetic tree of 41 MTCDSs (Fig. 1; Additional file 2: Fig. S16). Our results indicated that the crown

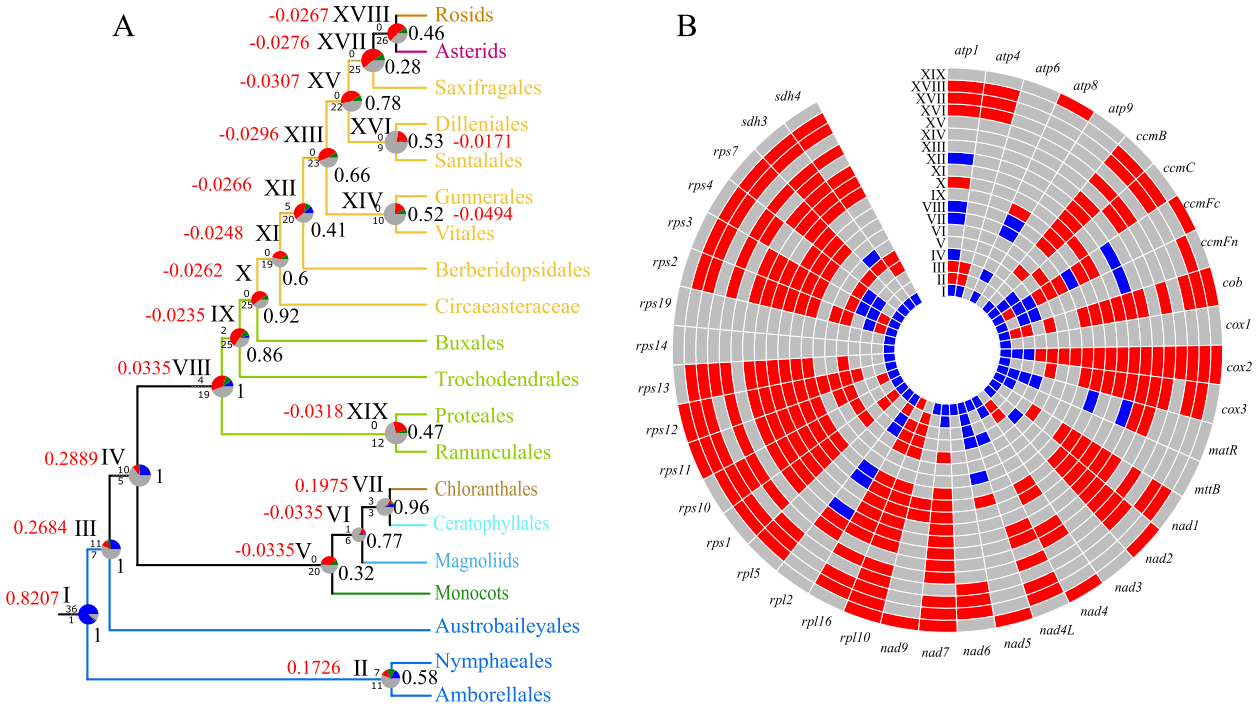

Fig. 3 ASTRAL phylogram based on 41 single gene trees. **A** Backbone of angiosperms inferred by ASTRAL using 41 single gene trees. Roman numerals denote the serial number of nodes. Numbers on the right side of the nodes represent posterior probability (PP) values. On the left side of the nodes, the top number indicates the number of the single gene concordant with the ASTRAL tree at that node, and the bottom number indicates the number of the single gene in conflict with that clade in the ASTRAL tree. Numbers with red font represent the internode certainty (ICA) scores for nodes. In the pie chart at each node, the blue area represents the proportion of a single gene that supports that clade; green and red areas represent the two alternative (main and remaining) relationships, respectively; and gray area represents the missing information (BP < 50%). **B** Roman numerals correspond to the serial numbers of nodes in panel **A**; the outer ring lists the 41 MTCDSs; blue block means the supporting for present clade (mapping the blue area of pie charts in panel **A**), red block represents alternative quartets (mapping the green and red areas of pie charts in panel **A**), and gray block represents the missing information (mapping the gray area of pie charts in panel **A**)

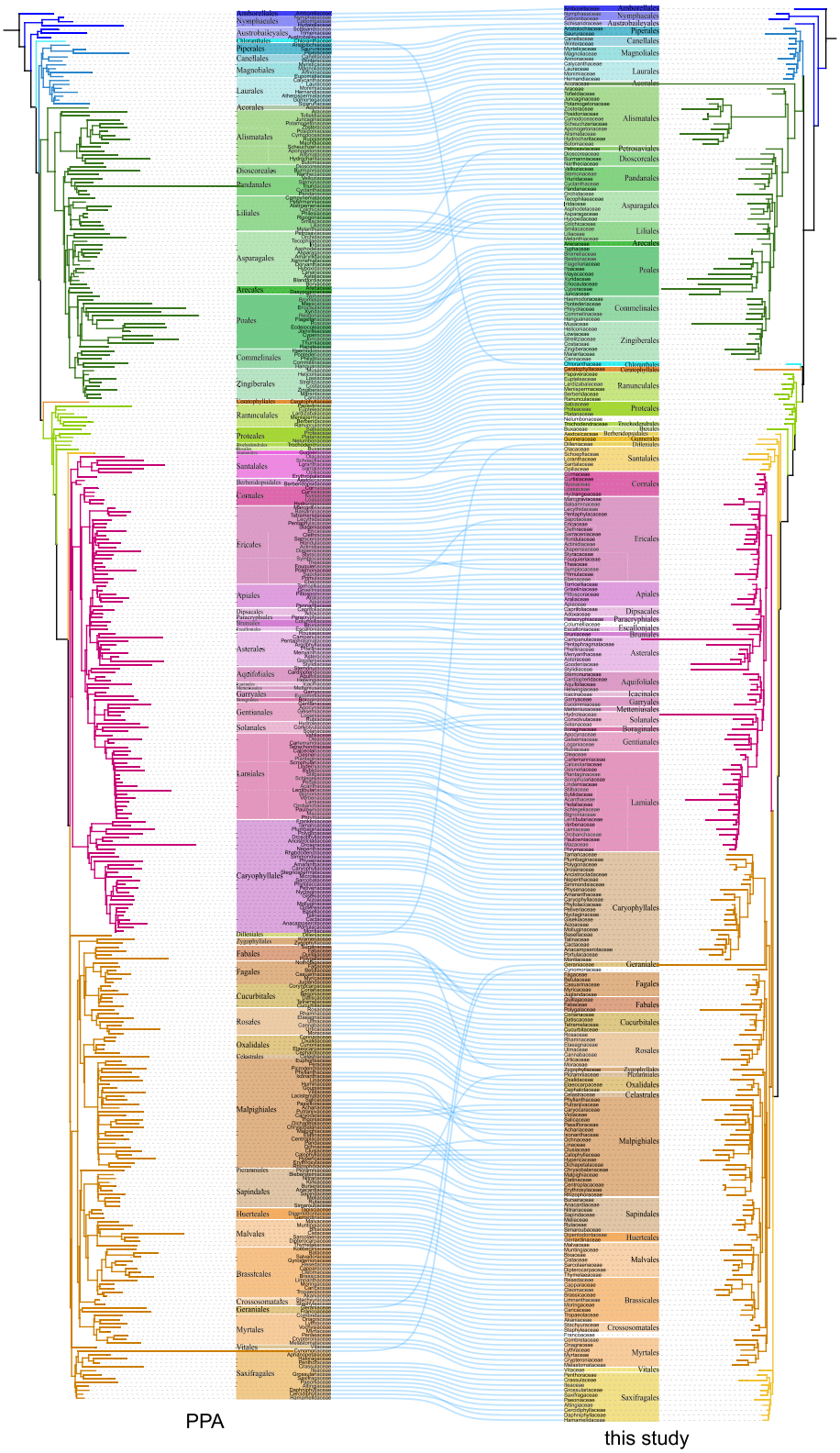

Fig. 4 Discordance at family level in 41 MTCDSs from the sampling of 486 accessions (right) and the plastid phylogenomic tree II (PPA II; left). Curved lines between tip labels of the two trees link the same families. Families are represented by a randomly chosen sample from each family

angiosperms arose at the boundary of Anisina and Olenekian in the Lower to Middle Triassic (ca. 247.17 Ma), with a 95% confidence interval (CI) of 227.2–247.2 Ma. Crown mesangiosperm divergence originated 237.87 Ma (95% CI:226.72–245.32 Ma), with the five major clades of mesangiosperms diverging over the following approximately 30 million years (Myr) (Figs. 1; Additional file 2: Figs. S16, S17). Magnoliids diverged from monocots approximately 236.82 Ma (95% CI: 226.40–243.45 Ma), with crown cluster divergence times of 149.71 Ma (95% CI: 140.28–163.14 Ma) for magnoliids and 226.48 Ma (95% CI: 222.26–229.81 Ma) for monocots. We inferred the time of divergence between [Ceratophyllales-Chloranthales] and eudicots as 236.65 Ma (95% CI: 235.12–238.29 Ma), followed by the onset of divergence of crown eudicots (229.68 Ma; 95% CI: 227.18–232.10 Ma) and the separation of Ceratophyllales and Chloranthales (213.05 Ma; 95% CI: 185.02–230.02 Ma). The crown monocots diverged in the Upper Triassic (266.85 Ma), which was followed by the monocot orders Alismatales (213.72 Ma; 95% CI: 209.85–217.73 Ma) and Petrosaviales (199.0 Ma; 95% CI: 194.47–203.81 Ma). Diversification of stem groups of orders of monocots was largely completed in the Lower Jurassic. The crown-eudicots diverged at 229.97 Ma (95% CI: 227.18–232.10 Ma), with Ranunculales, Proteales, Trochodendrales, and Buxales diverging at 205.0 Ma (95% CI: 197.02–213.40 Ma), 202.55 Ma (95% CI: 195.65–209.39 Ma), 197.63 Ma (95% CI: 191.52–204.10 Ma), and 187.28 Ma (95% CI: 181.68–193.31 Ma), respectively. The crown group of core eudicots diverged at 174.45 Ma (95% CI: 169.96–175.47 Ma) in the Middle Jurassic. The first and second diverging clades of core eudicots and the rosids and asterids all diverged at roughly the same time,161.08 Ma (95% CI: 157.92–164.90 Ma), during the Middle to Upper Jurassic (Fig. 1).

Discussion

With dense sampling across angiosperms, we obtained well-supported deep-node relationships of the angiosperms, which generally agreed with previous studies employing mitochondrial data (see [20, 29]) but also provided insights at deep nodes. Our results indicated that mitochondrial genomic data are effective at resolving deep-node relationships, provided insights into angiosperm phylogenetic relationships, and are an important resource for phylogenetic and evolutionary studies. We reconstructed the phylogenetic relationship among major clades of mesangiosperms with medium to high support along the backbone of the phylogenetic tree, largely supporting the existing phylogenetic framework of monocot orders with minor inconsistencies relative to previous studies [3, 6, 7, 22, 23, 61–63] (Additional file 2: Fig. S18). In addition, we obtained well-supported trees of some eudicot orders (Fig. 1; Additional file 2: Figs. S2, S6). Discrepancies of support and phylograms between 24- and 41-MTCDS may be due to the substitution rate of mitochondrial genes and the number of genes used in these analyses (Fig. 2). These 24 mitochondrial genes are often considered as conservative, core mitochondrial protein coding genes [64–66]. Many factors may lead to phylogenetic inconsistencies between PPA tree and 41-MTCDS concatenation tree, including horizontal gene transfer, hybridization, substitution rate, plastid genome capture, inheritance patterns and rapid radiation of angiosperms [9, 29, 53–55].

Three clades, including Ceratophyllales, Chloranthales, and [Chloranthales + Ceratophyllales] + eudicots, were strongly supported in this study (Fig. 1; Additional file 2: Figs. S2, S6). The magnoliids have often been proposed as the sister of the rest of the mesangiosperms, and Ceratophyllales have been more closely associated with the eudicots [3, 9, 20, 22, 26, 28, 31]. Our results place magnoliids sister to monocots with low support, agreeing with Qiu et al. and Endress et al. [16, 67]. Chloranthales and Ceratophyllales were sister groups, agreeing with previous studies [6, 20, 21, 28] and a morphological study of Chloranthales and Ceratophyllales [67]. However, Ceratophyllales are sister to the remaining of angiosperm (excluding ANA grade) and Chloranthales sister to magnoliids in a large-scale nuclear gene analysis [25]. This would appear to be a case of incongruence between nuclear and mitochondrial genes.

Our mtDNA trees largely support the existing phylogenetic framework of monocot orders (such as [9, 29]), although there are minor inconsistencies relative to previous studies. Three hypotheses have been proposed for the positions of Asparagales and Liliales: sister groups [7, 8, 61, 68], Asparagales diverging before Liliales, and the reverse [9, 22, 63]. Our results indicated that Asparagales and Liliales diverged successively along the monocot backbone with low to high support. Among commelinids, Arecales and Poales diverged successively, which agrees with plastome studies [9, 22, 63] but contrasts with nuclear-based studies with the opposite order [6, 7, 23, 61] (Additional file 2: Fig. S18). It is noteworthy that differences in the sampling in Petrosaviales (mycoheterotrophic *Petrosavia* [3] vs. autotrophic *Japonolirion* [9]) in the plastome-based results have led to their conflicting phylogenetic positions. Petrosaviales were monophyletic here, and the phylogenetic position of Petrosaviales was consistent with most previous studies [6–9, 22, 63]. Heterotrophic species often suffer from extremely divergent plastid and nuclear genes, the former perhaps associated with loss of photosynthesis, whereas mitochondrial genes have resolved phylogenetic positions of mycoheterotrophic taxa [56, 69]. In this study, the phylogenetic positions of Corsiaceae, Mitrastemonaceae, and Rafflesiaceae were well supported. This implies that

mitochondrial phylogenomics can play an important role in resolving phylogenetic placements for heterotrophic groups, although, like the plastome, uniparental inheritance of mitogenomes makes them liable to conflicts generated by hybridization and incomplete lineage sorting (ILS; also referred to as hemiplasy, deep coalescence, retention of ancestral polymorphisms, or trans-species polymorphisms).

Although differing from most previous analyses (see [9, 22, 25, 29]), the position of Berberidopsidales as sister to the rest of the core eudicots (Fig. 1; Additional file 2: Fig. S2; with moderate support) would be supported by the floral development of *Berberidopsis corallina* (Berberidopsidales), which is transitional from basal to the core eudicots [70]. The alternative position, as a member of the superasterids (sensu APG IV, 2016), would also be compatible with its floral features. Although relationships among the orders of asterids were resolved with low support, Columelliaceae were closely related to Escalloniales rather than Bruniales [3, 9, 22] with high support. Phylogenetic relationships among rosids also generally had low support, with a few exceptions such as the nitrogen-fixing and COM clades (Fig. 1; Additional file 2: Figs. S2-S6). The systematic position of Huaceae has been variably estimated in Malvales, Malpighiales, or Oxalidales [12, 18, 22, 71, 72]. Huaceae is characterized by its basal placentation, glandular leaves and anthers opening only apically [72]. Li et al. [9] recognized Huales, but this is premature, given that the position of Huaceae has been either weakly supported or unresolved. The well-supported topology in the COM clade here, Malpighiales-[Huaceae-[Oxalidales-Celastrales]] is consistent with the results of Ruhfel et al. [42]. Relationships among four orders in the COM clade have previously been inconsistent among plastid, mitochondrial and nuclear studies, possibly due to ILS and/or ancient hybridization [73]. Moreover, most clades of rosids had short internodes (Additional file 2: Figs. S2-S4), suggesting rapid radiation [73, 74] approximately 14 Myr as estimated in our divergence time study (Fig. 1). Short divergence time of these clades and potential ILS complicates phylogenetic inference [3, 28]. In addition, gene duplications and losses, recombination, hybridization, horizontal gene transfer, and ancient introgressive hybridization could have rendered inferences of relationships among these clades challenging in previous studies [28].

Recently, there have been numerous publications on the topic of the age of angiosperms [3, 25, 75–79]. It seems that key questions on this topic remain to be resolved, with estimates ranging from approximately 140 million years ago (Ma) to 270 Ma [3, 25, 75–79]. The scenario of rapid early diversification of angiosperms found here is consistent with previous studies [25, 44, 45, 77, 80] but provided some insights. The estimates of crown group node age of major clades, including monocots (226.85 Ma), magnoliids(147.77 Ma), eudicots (205 Ma), rosids (158 Ma), commelinids (175.27 Ma), and asterids (150 Ma) (Fig. 1), were usually older than those in PPA [3] but younger than "old tree crown node age" in Zuntini et al. [25]. The similarity of our results with previous studies may be related to methods of fossil calibrations of molecular evolution [79].

The early rapid radiation and development of the main extant angiosperm clades over a period of ~ 30 Myr during the Early and Middle Triassic are supported by some recent studies [3, 6, 7, 81]. Several rapid radiations have been recorded, including the early eudicots (ca.30 Myr) [6, 76] and rosids (ca. 14 Myr). Overall, angiosperm divergence time estimates found here suggest that the major angiosperm clades diverged in the Triassic to Middle Jurassic, and the subsequent divergence of orders from stem groups occurred roughly in the late Jurassic to Early Cretaceous, with the order-to-family divergences occurring in the Cretaceous. The origin and early diversification of angiosperms proposed in this study are corroborated by other recent independent studies [7, 50, 82–85]: presence of entomophilous traits in early angiosperms [75] and Diptera and Lepidoptera in the Triassic [86], both with implications for angiosperm origin and coevolution with insects. The Permian–Triassic extinction event [87–90] may have provided open ecological niches for radiation of angiosperms [91–93]. In addition, the diversification of insects, most of which, especially phytophagous and pollinating insects, provided opportunities for angiosperm expansion and diversification [86, 94], and intrinsic innovations offered powerful adaptive mechanisms. A series of events acted together to promote the dominance of angiosperms in terrestrial ecosystems.

Conclusions

Mitochondrial gene data for 481 angiosperms representing 335 families and 63 orders were included in this study. Three clades, including Ceratophyllales, Chloranthales, and [Chloranthales + Ceratophyllales] + eudicots, were strongly supported in this study. Among the core eudicots, Berberidopsidales and Dilleniales + Gunnerales were the first and second successive branches. Our results indicated that mitochondrial genomic data were effective at resolving deep node relationships of angiosperm phylogeny and thus represent an important resource for phylogenetics and evolutionary studies of angiosperm.

Methods

Taxon sampling

Sampling included 481 species representing 335 families and 63 orders, accounting for approximately 98.4% of the orders and 80% of the families of angiosperms recognized by APG IV [23]. The sampled species represented four families from three orders of the ANA group, 13 families from four orders of magnoliids, one family each in Ceratophyllaceae and Chloranthaceae, 60 families from eleven orders of monocots, and 256 families from 42 eudicots orders. One order (Vahliales) and 61 families (including Autrobaileyaceae, Hydatellaceae, and Trimeniaceae) were not included because of the lack of plant material. Five gymnosperm species, each representing a major clade of gymnosperms, were used as outgroup, and MTCDSs were downloaded from the National Center for Biotechnology Information (NCBI) (Additional file 1:Table S2).

Mitochondrial genome sequencing

Fresh young leaves were collected, dried using silica gel, and stored at − 20 °C at the Institute of Botany, Chinese Academy of Sciences. Total genomic DNA was extracted using the TIANGEN Plant Genomic DNA Kit, according to the manufacturer's instructions. The quality-checked DNA was sequenced on an Illumina platform by Majorbio Company (Beijing, China) to generate paired-end 150 bp reads. Quality control of raw data was performed using FastQC v0.11.8 [95]. Sequence artifacts, including reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotides (Ns), were removed with Trimmomatic [96]. In addition, 203 raw sequence reads and 192 mitochondrial genome sequences were downloaded from NCBI (Additional file 1: Table S2).

Assembly and annotation of MTCDSs

High-quality Illumina reads were assembled into mitochondrial contigs using BWA [97], SAMtools [98], Geneious Prime (Biomatters Ltd, Auckland, New Zealand), SPAdes v3.13.0, and GetOrganelle v1.7.5.3. [99]. The K-mer value was calculated using KmerGenie v1.7051 (parameters: -l 15 -k 105 -s 10) and used as a parameter in SPAdes. Sequencing depth was examined to avoid contigs with low depth that might have entrained from the nuclear genome [28]. Contigs were further extended using NOVOplasty v4.3.1 [100] and Geneious Prime [101]. Annotation of mtCDSs were annotated using the blastn program for NCBI-BLAST + (2.9.0–2) [102], Geneious Prime (Biomatters Ltd, Auckland, New Zealand) with the option "annotation" with a similarity threshold of 70% and an online tool MITOFY [103] (https://vcru.wisc.edu/cgi-bin/mitofy/mitofy.cgi). The annotation results were checked and manually adjusted in Geneious Prime, and 41 MTCDSs were extracted using a custom Perl script [104].

Phylogenetic analysis

Two datasets, one containing 41 mtCDSs and another containing 24 core mtCDSs, were used for phylogenetic analysis. MTCDSs were individually aligned using MAFFT v7.505 [105] and MUSCLE v5.1.l [106], with default parameter settings, and the poorly aligned regions were removed using Gblocks 0.91b with the least stringent settings. The 41 single-gene nucleotide (41-MTCDSs) alignments were combined in a supermatrix, with a total length of 61,572 bp. Pseudogenes and missing genes were treated as gaps in the datasets. Some poorly aligned regions were removed using Gblocks 0.91b with the least stringent settings [105]. ModelFinder [107] was used to select the best-fit partition model (Edge-unlinked) for RAxML v8.2.12 [108] using the AICc criterion, and GTRGAMMA model was selected, as well as a more fine-grained set of models (Additional file 1: Table S4) for IQ-TREE v1.6.9 [109] was chosen. Maximum-likelihood (ML) analysis used RAxML v8.2.12 [108] and IQ-TREE v1.6.9 [109] under the GTRGAMMA model and set of models aforesaid respectively, starting with 1000 random trees to search for the best tree and obtaining a bootstrap percentage based on 1000 non-parametric iterations. A subset of data, which included 24 MTCDSs present in all samples, often considered as core protein genes [64–66], was filtered from matrix of 41 MTCDSs. The nucleotide sequences of these 24 MTCDSs were used to produce an ML tree with RAxML v8.2.12 [108] as described previously.

Additionally, a coalescent-based species tree was inferred from the single gene tree using ASTRAL v5.7.8. For each single gene of all sampled species, multiple sequence alignments were constructed using MAFFT v7.505. As unrooted input files of ASTRAL, the best ML tree and 100 pseudo-replicate bootstrap trees for all genes were inferred using RAxML v8.2.12. Finally, an optimal species tree, with bootstrap support percentages, was produced.

Assessing phylogenetic conflict

To investigate the phylogenetic conflict between the plastid phylogenomic angiosperm (PPA) and 41-MTCDS concatenation tree, R package, phytools [110], was utilized to create a cophylogeny plot and visualize the incongruences between them. 41-MTCDS concatenation tree were first pruned as a family-level tree to include only families present in PPA trees, and any duplicates were

removed (after random selection per family); we then compared the conflicting phylogenetic relationships of angiosperms inferred in this study with the plastid PPA inferred by Li et al. [3]. In total, 288 angiosperm families were sampled in both PPA tree and 41-MTCDS concatenation tree and reserved for phylogenetic conflict assessment. Incongruences were quantified at family level using the R package TreeDist v.2.9.1 based on the information-based generalized Robinson-Foulds distances [111]. 100,000 random pairs of 288-leaf trees were generated, and the expected distances between them were calculated as 0.52 according to Allen et al. [112], and the calculated tree distances can be provided by comparison with the distance expected for a pair of random trees.

To investigate the phylogenetic conflict between species tree and single-gene tree inferred by MTCDSs, the conflicting and concordant bipartitions were computed using Phyparts while furnishing the summary statistics of internode certainty (ICA) scores. The single gene trees were utilized as input files, with gymnosperms as the outgroup for rerooting purposes. Nodes displaying bootstrap support values (BS) $\leq$ 50% were excluded. Subsequently, the input files were individually aligned to the ASTRAL tree and the 41-MTCDS tree using Phyparts. The single gene tree corresponding to the 24-MTCDS dataset was selected and mapped to the 24-MTCDS tree, enabling the computation of concordant/conflicting bipartitions as well as ICA values for each tree separately.

The quartet sampling (QS) method, a measure used to dissect phylogenetic inconsistencies and discriminate lack of support, was adopted to distinguish the lack of support from conflicting support in the concatenated species tree. QS was conducted in Linux with 100 replicates, utilizing both the concatenated 41-MTCDS tree and the concatenated supermatrix encompassing all mtCDSs. The outcomes were visualized using R scripts. Quartet concordance (QC), quartet differential (QD), quartet informativeness (QI), and quartet fidelity (QF) scores were assessed using the QS method.

Divergence time estimates

Fossil data were searched in fossil and paleontology databases, including PBDB (https://www.paleobiodb.org/), Timetree (http://www.timetree.org/), INTERNATIONAL FOSSIL PLANT NAMES INDEX (http://www.ifpni.org/index.htm), and Fossil Calibration Database Reliable [113]. We selected reliable fossil data as calibration points, confirmed by previous studies. In subsequent molecular dating analyses, we used 46 calibration points (Additional file 1: Table S5; Additional file 2: Fig. S17). Divergence time estimates were determined using TreePL [114]. Forty smoothing parameter values ranging from 10^{20} to 10^{-20} were tested using TreePL for cross-validation, resulting in an optimal smoothing parameter of 0.0001. Penalized likelihood dating analysis was performed in TreePL using the best ML tree and 1000 bootstrap replicates. All bootstrap replicates were analyzed using RAxML to determine branch lengths and then dated. Confidence intervals for the date estimates were calculated from the 1000 bootstrap replicates using TreeAnnotator [115] as implemented in BEAST.

Abbreviations

CDS	Coding sequence
mtCDSs	Mitochondrial protein-coding sequences
41-MTCDS	*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl10*, *rpl16*, *rpl2*, *rpl5*, *rps1*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*, *sdh3*, *sdh4*
24-MTCDS	*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*
Eud	Eudicots
Mon	Monocots
Mag	Magnoliids
Chl	Chloranthales
Cer	Ceratophyllales
Ma	Million Years Ago
Myr	Million Years
HPD	Highest Posterior Density
APG	Angiosperm Phylogeny Group
ANA grade	Amborellales, Nymphaeales and Austrobaileyales
NCBI	National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/)
BP	Bootstrap Percentages
PP	Posterior Probability
COM	Celastrales–Oxalidales–Malpighiales clade
QS	Quartet Sampling
QC	Quartet concordance
QD	Quartet Differential
QI	Quartet Informativeness
QF	Quartet Fidelity
ML	Maximum-likelihood
ICA	Internode Certainty
IQR	Interquartile Range
PPA	Plastid Phylogenomic Angiosperm
ILS	Incomplete Lineage Sorting
DNA	Deoxyribonucleic Acid
AICc	Second-order Akaike information criterion

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02135-9.

Additional file 1: Table S1. Topologies of five mesangiosperm lineages recovered by previous studies. Table s2 Accession number and reference of mitochondrial data. Table s3 List of newly sequenced species and the specimen information in this study. Table s4 A fine-grained set of models for IQ-TREE selected by ModelFinder. Table s5 Details of the 46 fossils used as calibration point prior settings in divergence time analyses of angiosperms.

Additional file 2: Fig. S1 Sixteen recently published phylograms of core eudicots. A–P. Phylograms of core eudicots published by Worberg et al. (2007) and APG (2016) (A), Gitzendanner et al. (2018) (B), Leebens-Mack et al. (2019) (C), Moore et al. (2011) (D), Barniske et al. (2012) (E), Smith et al. (2010) (F, G), Zeng et al. (2017) (H), Yang et al. (2020b) (I), Li et al. (2019, 2021a) (J), Wikstrom et al. (2001) (K), Bremer et al. (2009) (L), Soltis

et al. (2011) (M), Moore et al. (2011) (N), Worberg et al. (2007) (O), and Barniske et al. (2012) (P). Fig. S2 Phylogram of angiosperms based on 24 MTCDSs from 486 sampling. A. Tree of angiosperms orders inferred by RAxML. B. Inlet indicating the relationships among the main lineages of angiosperms. Numbers above the branches represent BP values, and taxa that separated from the APG IV orders are indicated in red. Fig. S3 Phylogram based on the maximum-likelihood (ML) partitioning analysis of the 41-MTCDS concatenation matrix and 486 sampling concatenation matrix. Numbers above the branches represent the ML bootstrap percentages (BP), and taxa that separated from the APG IV orders are indicated in red. Fig. S4 Phylogram of angiosperms based on coalescence methods inferred by ASTRAL. Numbers associated with the nodes represent posterior probability (PP) support values, and taxa that separated from the APG IV orders are indicated in red. Branch length is indicated on the right. Fig. S5 Phylogram of angiosperms based on the 24-MTCDS concatenation matrix of 486 sampling. The branches on the left indicate the topological relationship, a and different colors indicate the different branches of the clade. Numbers above the branches represent BP values, and taxa that separated from the APG IV orders are indicated in red. Fig. S6 Phylogram of angiosperms based on the ASTRAL analysis of 41 single gene trees from 486 sampling. A. The tree of angiosperm orders inferred using ASTRAL. The detailed phylogeny of 486 species is presented in Supplemental Fig. 7. B. An inlet showing the relationship of the main lineages of angiosperms. Numbers above the branches represent PP values, and taxa that separated from the APG IV orders are indicated in red. Fig. S7 Phylogram of angiosperms based on the IQ-TREE analysis of 41-MTCDS concatenation matrix from 486 sampling. A. The tree of angiosperm orders inferred using IQ-TREE, and taxa that separated from the APG IV orders are indicated in red. B. An inlet showing the relationship of the main lineages of angiosperms. Numbers above the branches represent BP values. Fig. S8 Phylogram of the main lineages of mesangiosperms based on 41 mitochondrial protein-coding sequences (MTCDSs) and recently published datasets. A. Phylogram constructed in the present study, based on 41 MTCDSs. B–G. Phylograms published by Moore et al. (2010), Soltis et al. (2011), APG (2016), and Gitzendanner et al. (2018) (B) and Li et al. (2019, 2021) (C), based on plastid gene datasets; published by Yang et al. (2020) (D) and Leebens-Mack et al. (2019) (E), based on nuclear gene datasets; and published by Xue et al. (2022) (F) and Qiu et al. (2010) (G), based on mitochondrial gene datasets. Fig. S9 Tree showing the internode certainty (ICA) scores based on 41 MTCDSs. Numbers associated with nodes represent ICA values, and taxa that separated from the APG IV orders are indicated in red. Fig. S10 Summary of conflicting and concordant for each mitochondrial gene tree on the phylogram based on 41 MTCDSs. The pie charts at each node present the proportion of single gene that support that clade (blue), and two alternative (main and remaining) relationship (green and red), and missing information for nodes (less than 50% bootstrap support, gray). The taxa that separated from the APG IV orders are indicated in red. Fig. S11 Tree showing the ICA scores based on 24 MTCDSs. Numbers associated with nodes represent ICA values, and taxa name that separated from the APG IV orders are indicated in red. Fig. S12 Phylogram of 41 MTCDSs with Quartet Concordance (QC) scores. Full phylogeny with heat map coloration of nodes, based on Quartet Concordance (QC) scores for internal branches: dark green ($QC > 0.2$), light green ($0.2 \geq QC > 0$), light orange ($0 \geq QC \geq -0.05$), or dark orange ($QC < -0.05$). QC/Quartet Differential (QD)/Quartet Informativeness (QI) scores are shown with numbers in the middle of the branch. Fig. S13 Detailed 24-MTCDS tree. Summary of the conflicting and concordant MTCDSs of phylogram based on 24 MTCDSs. The pie charts at each node present the proportion of single gene that support that clade (blue), and two alternative (main and remaining) relationship (green and red), and missing information for nodes (less than 50% bootstrap support, gray). The taxa that separated from the APG IV orders are indicated in red. Fig. S14 ICA scores computed for the ASTRAL tree. Numbers associated with nodes represent ICA values; different colors represent the branches of a clade; and taxa that separated from the APG IV orders are indicated in red. Fig. S15 Summary of the conflicting and concordant for each MTCDSs of the ASTRAL phylogram. The pie charts at each node present the proportion of single gene that support that clade (blue), and two alternative (main and remaining) relationship (green and red), and missing information for nodes (less than 50% bootstrap support, gray). The taxa that separated from the APG IV orders are indicated in red. Fig. S16 Chronogram of angiosperms inferred using treePL, based on the ML partitioned analysis of 41 MTCDs from 486 sampling. Numbers next to the nodes represent the node ages (million years ago [Ma]), and horizontal bars represent 95% confidence intervals of the node ages. Fig. S17 Chronogram of angiosperms based on the analysis of 41 MTCDSs using the penalized likelihood (PL) method. Branch colors indicate divergence time. Black dots indicate the fossil calibration points selected for this study. Fig. S18 Phylograms of monocots based on different datasets. Green lines represent the phylogram inferred from 41-MTCDS concatenation matrix; red lines represent phylograms inferred from nuclear datasets (Zeng et al., 2014; Leebens-Mack et al., 2019; Yang et al., 2020; Baker et al., 2022; and blue lines represent phylograms based on plastid datasets (APG, 2016; Givnish et al., 2018; Li et al., 2021).

Acknowledgements
We thank Prof. De-zhu Li, Prof. Jin-Hua Ran and Prof. Li-Min Lu for their constructive suggestions on our manuscript.

Authors' contributions
X.J. designed the project; X.J. supervised this work and generated funding for this research; X.J., J.L., T.L., W.H., X.Z., C.X., and D.L. collected the plant material; D.L., Z.G., and Z.L. assembled into mitochondrial contigs; D.L., X.J., B.S., Z.G., and M.C. analyzed data and prepared the results; D.L., X.J., and M.C. drafted the manuscript. All authors revised the manuscript and approved the final manuscript text.

Funding
This research was supported by the National Natural Science Foundation of China (32270214, 31870195 to XJ) and the Project of Orchid Biodiversity Survey of China, National Forestry and Grassland Administration (2019070714, 2020100707, 2022070702).

Data availability
Concatenation matrix, tree files and alignment have been deposited at figshare (https://figshare.com/s/717da1044ac73ed6e7f8, https://doi.org/10.6084/m9.figshare.28130231) and are publicly available.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 5 June 2024 Accepted: 17 January 2025
Published: 14 February 2025

References

1. Friis EM, Crane PR, Pedersen KR. Early Flowers and Angiosperm Evolution. Cambridge: Cambridge University Press; 2011.
2. Benton MJ. The origins of modern biodiversity on land. Philos T R Soc B. 2010;365(1558):3667–79.
3. Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants. 2019;5(5):461–70.
4. Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR, et al. A functional phylogenomic view of the seed plants. PLoS Genet. 2011;7(12):e1002411.

5. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A. 2014;111(45):E4859–68.
6. Zeng LP, Zhang Q, Sun RR, Kong HZ, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat Commun. 2014;5:4956.
7. Yang LX, Su DY, Chang X, Foster CSP, Sun LH, Huang CH, Zhou XF, Zeng LP, Ma H, Zhong BJ. Phylogenomic insights into deep phylogeny of Angiosperms based on broad nuclear gene sampling. Plant Commun. 2020;1(2):100027.
8. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019;574(7780):679-+.
9. Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, et al. Plastid phylogenomic insights into relationships of all flowering plant families. BMC Biol. 2021;19(1):232.
10. Cronquist A. The evolution and classification of flowering plants. Bronx, New York: The New York Botanical Garden; 1989.
11. Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA, Kuzoff RK, Kron KA, Chase MW, et al. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann Mo Bot Gard. 1997;84(1):1–49.
12. APG. An ordinal classification for the families of flowering plants. Ann Mo Bot Gard. 1998;85(4):531–553.
13. Qiu YL, Lee JH, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen ZD, Savolainen V, Chase MW. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature. 1999;402(6760):404–7.
14. Qiu YL, Palmer JD. Phylogeny of early land plants: insights from genes and genomes. Trends Plant Sci. 1999;4(1):26–30.
15. Beckert S, Muhle H, Pruchner D, Knoop V. The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol Phylogen Evol. 2001;18(1):117–26.
16. Qiu YL, Li LB, Wang B, Chen ZD, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, et al. The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci U S A. 2006;103(42):15511–6.
17. Zhu XY, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD. Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. BMC Evol Biol. 2007;7:217.
18. APG. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc. 2009;161(2):105–121.
19. Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. Am J Bot. 2010;97(8):1296–303.
20. Qiu YL, Li LB, Wang B, Xue JY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. J Syst Evol. 2010;48(6):391–425.
21. Zhang N, Zeng LP, Shan HY, Ma H. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol. 2012;195(4):923–37.
22. APG. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
23. Transcriptomes OTP. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019;574(7780):679.
24. Guo C, Luo Y, Gao LM, Yi TS, Li HT, Yang JB, Li DZ. Phylogenomics and the flowering plant tree of life. J Integr Plant Biol. 2023;65(2):299–323.
25. Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C, et al. Phylogenomics and the rise of the angiosperms. Nature. 2024;629:843–50.
26. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS, et al. Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot. 2011;98(4):704–30.
27. Qin LY, Hu YH, Wang JP, Wang XL, Zhao R, Shan HY, Li KP, Xu P, Wu HY, Yan XQ, et al. Insights into angiosperm evolution, floral development and chemical biosynthesis from the Aristolochia fimbriata genome. Nat Plants. 2021;7(9):1239.
28. Xue JY, Dong SS, Wang MQ, Song TQ, Zhou GC, Li Z, Van de Peer Y, Shao ZQ, Wang W, Chen M, et al. Mitochondrial genes from 18 angiosperms fill sampling gaps for phylogenomic inferences of the early diversification of flowering plants. J Syst Evol. 2022;60(4):773–88.
29. Hu HY, Sun PC, Yang YZ, Ma JX, Liu JQ. Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. J Integr Plant Biol. 2023;65(6):1479–89.
30. Graham SW, Iles WJD. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. Am J Bot. 2009;96(1):216–27.
31. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A. 2010;107(10):4623–8.
32. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U S A. 2007;104(49):19363–8.
33. Mathews S, Donoghue MJ. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science. 1999;286(5441):947–50.
34. Soltis PS, Soltis DE, Chase MW. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature. 1999;402(6760):402–4.
35. Leebens-Mack J, Raubeson LA, Cui LY, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the felsenstein zone. Mol Biol Evol. 2005;22(10):1948–63.
36. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A. 2007;104(49):19369–74.
37. Finet C, Timme RE, Delwiche CF, Marletaz F. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr Biol. 2012;22(15):1456–7.
38. Goremykin VV, Nikiforova SV, Biggs PJ, Zhong BJ, Delange P, Martin W, Woetzel S, Atherton RA, Mclenachan PA, Lockhart PJ. The evolutionary root of flowering plants. Syst Biol. 2013;62(1):50–61.
39. Xi ZX, Liu L, Rest JS, Davis CC. Coalescent versus concatenation methods and the placement of amborella as sister to water lilies. Syst Biol. 2014;63(6):919–32.
40. Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, Wolfe AD, dePamphilis CW. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc Natl Acad Sci U S A. 2000;97(24):13166–71.
41. Zhang LS, Chen F, Zhang XT, Li Z, Zhao YY, Lohaus R, Chang XJ, Dong W, Ho SYW, Liu X, et al. The water lily genome and the early evolution of flowering plants. Nature. 2020;577(7788):79-+.
42. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol. 2014;14:23. https://doi.org/10.1186/1471-2148-14-23.
43. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot. 2018;105(3):291–301.
44. Friis EM, Pedersen KR, Crane PR. Diversity in obscurity: fossil flowers and the early history of angiosperms. Philos T R Soc B. 2010;365(1539):369–82.
45. Herendeen PS, Friis EM, Pedersen KR, Crane PR. Palaeobotanical redux: revisiting the age of the angiosperms. Nat Plants. 2017;3(3):1.
46. Bonney T. More Letters of Charles Darwin A Record of his Work in a Series of hitherto Unpublished Letters. Nature. 1903;67:529–30.
47. Friedman WE. The meaning of Darwin's "abominable mystery." Am J Bot. 2009;96(1):5–21.
48. Smith SA, Beaulieu JM, Donoghue MJ. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc Natl Acad Sci U S A. 2010;107(13):5897–902.
49. Clarke JT, Warnock RCM, Donoghue PCJ. Establishing a time-scale for plant evolution. New Phytol. 2011;192(1):266–301.
50. Foster CSP, Sauquet H, Van der Merwe M, McPherson H, Rossetto M, Ho SYW. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. Syst Biol. 2017;66(3):338–51.

51. Magallon S, Hilu KW, 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. 2013;100(3):556–73.
52. Magallon S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 2015;207(2):437–53.
53. Palmer JD, Herbon LA. Plant mitochondrial-DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28(1–2):87–97.
54. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogen Evol. 2008;49(3):827–31.
55. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84(24):9054–8.
56. Lin QS, Braukmann TWA, Gomez MS, Mayer JLS, Pinheiro F, Merckx VSFT, Stefanovic S, Graham SW. Mitochondrial genomic data are effective at placing mycoheterotrophic lineages in plant phylogeny. New Phytol. 2022;236(5):1908–21.
57. Qiu YL, Li LB, Hendry TA, Li RQ, Taylor DW, Issa MJ, Ronen AJ, Vekaria ML, White AM. Reconstructing the basal angiosperm phylogeny: evaluating information content of mitochondrial genes. Taxon. 2006;55(4):837–56.
58. Beckert S, Steinhauser S, Muhle H, Knoop V. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial nad5 gene. Plant Syst Evol. 1999;218(3–4):179–92.
59. Sousa F, Civán P, Brazao J, Foster PG, Cox CJ. The mitochondrial phylogeny of land plants shows support for Setaphyta under composition-heterogeneous substitution models. Peerj. 2020;8:e8995.
60. Ran JH, Shen TT, Wang MM, Wang XQ. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. P Roy Soc B Biol Sci. 2018;285(1881):20181012.
61. Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigue LR, Brewer G, Carruthers T, Clarkson JJ, et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Syst Biol. 2022;71(2):301–19.
62. Lughadha EN, Govaerts R, Belyaeva I, Black N, Lindon H, Alkin R, Magil RE, Nicolson N. Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. Phytotaxa. 2016;272(1):82–8.
63. Givnish TJ, Zuluaga A, Spalink D, Gomez MS, Lam VKY, Saarela JM, Sass C, Iles WJD, de Sousa DJL, Leebens-Mack J, et al. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. Am J Bot. 2018;105(11):1888–910.
64. Hall ND, Zhang H, Mower JP, McElroy JS, Goertzen LR. The mitochondrial genome of and characterization of gene content within Poaceae. Genome Biol Evol. 2020;12(1):3684–97.
65. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U SA. 2002;99(15):9905–12.
66. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29(3):380–95.
67. Endress PK, Doyle JA. Reconstructing the ancestral angiosperm flower and its initial specializations. Am J Bot. 2009;96(1):22–66.
68. Zeng LP, Zhang Q, Sun RR, Kong HZ, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nature Communications. 2014;5:4956.
69. Jost M, Samain MS, Marques I, Graham SW, Wanke S. Discordant phylogenomic placement of Hydnoraceae and Lactoridaceae within Piperales using data from all three genomes. Front Plant Sci. 2021;12:642598.
70. De Craene LPR. Floral development of *Berberidopsis corallina*: a crucial link in the evolution of flowers in the core Eudicots. Ann Bot. 2004;94(5):741–51.
71. Wurdack KJ, Davis CC. Malpighiales phylogenetics: gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. Am J Bot. 2009;96(8):1551–70.
72. Bayer C. Huaceae. In: Kubitzki K, editor. Flowering Plants · Eudicots. The Families and Genera of Vascular Plants, vol 9. Berlin, Heidelberg: Springer; 2007. https://doi.org/10.1007/978-3-540-32219-1_23.
73. Sun M, Soltis DE, Soltis PS, Zhu XY, Burleigh JG, Chen ZD. Deep phylogenetic incongruence in the angiosperm clade Rosidae. Mol Phylogen Evol. 2015;83:156–66.
74. Wang HC, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. Rosid radiation and the rapid rise of angiosperm-dominated forests. Proc Natl Acad Sci U S A. 2009;106(10):3853–8.
75. Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, de Morais EB, Bull-Hereñu K, Carrive L, Chartier M, et al. The ancestral flower of angiosperms and its early diversification. Nat Commun. 2017;8:8.
76. Barba-Montoya J, dos Reis M, Schneider H, Donoghue PCJ, Yang ZH. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous terrestrial revolution. New Phytol. 2018;218(2):819–34.
77. Ramírez-Barahona S, Sauquet H, Magallón S. The delayed and geographically heterogeneous diversification of flowering plant families. Nat Ecol Evol. 2020;4(9):1232-+.
78. Benton MJ, Wilf P, Sauquet H. The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. New Phytol. 2022;233(5):2017–35.
79. Sauquet H, Ramírez-Barahona S, Magallón S. What is the age of flowering plants? J Exp Bot. 2022;73(12):3840–53.
80. Crane PR, Lidgard S. Angiosperm diversification and paleolatitudinal gradients in Cretaceous floristic diversity. Science. 1989;246(4930):675–8.
81. Zhang CF, Zhang TK, Luebert F, Xiang YZ, Huang CH, Hu Y, Rees M, Frohlich MW, Qi J, Weigend M, et al. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. Mol Biol Evol. 2020;37(11):3188–210.
82. Magallon S, Castillo A. Angiosperm diversification through time. Am J Bot. 2009;96(1):349–65.
83. Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, et al. Three keys to the radiation of angiosperms into freezing environments (vol 506, pg 89, 2014). Nature. 2014;514(7522):394–394.
84. Beaulieu JM, O'Meara BC, Crane P, Donoghue MJ. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. Syst Biol. 2015;64(5):869–78.
85. Zhang LS, Chen F, Zhang XT, Li Z, Zhao YY, Lohaus R, Chang XJ, Dong W, Ho SYW, Liu X, et al. The water lily genome and the early evolution of flowering plants. Nature. 2020;577(7788):79.
86. Misof B, Liu SL, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. Phylogenomics resolves the timing and pattern of insect evolution. Science. 2014;346(6210):763–7.
87. McElwain JC, Beerling DJ, Woodward FI. Fossil plants and global warming at the Triassic-Jurassic boundary. Science. 1999;285(5432):1386–90.
88. Huynh TT, Poulsen CJ. Rising atmospheric CO_2 as a possible trigger for the end-Triassic mass extinction. Paleogeogr Paleoclimatol Paleoecol. 2005;217(3–4):223–42.
89. Deenen MHL, Ruhl M, Bonis NR, Krijgsman W, Kuerschner WM, Reitsma M, van Bergen MJ. A new chronology for the end-Triassic mass extinction. Earth Planet Sc Lett. 2010;291(1–4):113–25.
90. Chaboureau AC, Sepulchre P, Donnadieu Y, Franc A. Tectonic-driven climate change and the diversification of angiosperms. P Natl Acad Sci USA. 2014;111(39):14066–70.
91. Wang CJ, Visscher H. A molecular biomarker for end-Permian plant extinction in South China. Geology. 2021;49(12):1489–94.
92. Shen J, Zhang YG, Yang H, Xie S, Pearson A. Early and late phases of the Permian-Triassic mass extinction marked by different atmospheric CO_2 regimes. Nat Geosci. 2022;15:839–44.
93. Huang Y, Chen Z-Q, Roopnarine PD, Benton MJ, Zhao L, Feng X, Li Z. The stability and collapse of marine ecosystems during the Permian-Triassic mass extinction. Curr Biol. 2023;33(6):1059–70.
94. Pontarotti P. Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life. Switzerland: Springer International Publishing; 2014.
95. Brown J, Pirrung M, McCue LA. FQC dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics. 2017;33(19):3137–9.

96. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
97. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;arXiv:1303.3997v2. https://doi.org/10.48550/arXiv.1303.3997.
98. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
99. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):1.
100. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18.
101. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
102. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST plus : architecture and applications. BMC Bioinformatics. 2009;10:421.
103. Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.
104. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
105. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564–77.
106. Edgar RC. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun. 2022;13(1):6968.
107. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587.
108. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
109. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
110. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3(2):217–23.
111. Smith MR. Information theoretic generalized Robinson-Foulds metrics for comparing phylogenetic trees. Bioinformatics. 2020;36(20):5007–13.
112. Allen BL, Steel M. Subtree transfer operations and their induced metrics on evolutionary trees. Ann Comb. 2001;5(1):1–15.
113. Ksepka DT, Parham JF, Allman JF, Benton MJ, Carrano MT, Cranston KA, Donoghue PCJ, Head JJ, Hermsen EJ, Irmis RB, et al. The fossil calibration database-a new resource for divergence dating. Syst Biol. 2015;64(5):853–9.
114. Smith SA, O'Meara BC. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 2012;28(20):2689–90.
115. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29(8):1969–73.

Publisher's Note