

RESEARCH ARTICLE

Open Access



Plastid phylogenomic insights into relationships of all flowering plant families

Hong-Tao Li^{1,2†}, Yang Luo^{3†}, Lu Gan^{1,2†}, Peng-Fei Ma^{1,2†}, Lian-Ming Gao^{2,3,4†}, Jun-Bo Yang^{1,2†}, Jie Cai^{1†}, Matthew A. Gitzendanner^{5,6}, Peter W. Fritsch⁷, Ting Zhang¹, Jian-Jun Jin^{1,8}, Chun-Xia Zeng¹, Hong Wang^{2,3}, Wen-Bin Yu⁹, Rong Zhang¹, Michelle van der Bank¹⁰, Richard G. Olmstead¹¹, Peter M. Hollingsworth¹², Mark W. Chase^{13,14}, Douglas E. Soltis^{5,6}, Pamela S. Soltis^{5,6,15}, Ting-Shuang Yi^{1,2*} and De-Zhu Li^{1,2,3*} 

Abstract

Background: Flowering plants (angiosperms) are dominant components of global terrestrial ecosystems, but phylogenetic relationships at the familial level and above remain only partially resolved, greatly impeding our full understanding of their evolution and early diversification. The plastome, typically mapped as a circular genome, has been the most important molecular data source for plant phylogeny reconstruction for decades.

Results: Here, we assembled by far the largest plastid dataset of angiosperms, composed of 80 genes from 4792 plastomes of 4660 species in 2024 genera representing all currently recognized families. Our phylogenetic tree (PPA II) is essentially congruent with those of previous plastid phylogenomic analyses but generally provides greater clade support. In the PPA II tree, 75% of nodes at or above the ordinal level and 78% at or above the familial level were resolved with high bootstrap support (BP \geq 90). We obtained strong support for many interordinal and interfamilial relationships that were poorly resolved previously within the core eudicots, such as Dilleniales, Saxifragales, and Vitales being resolved as successive sisters to the remaining rosids, and Santalales, Berberidopsidales, and Caryophyllales as successive sisters to the asterids. However, the placement of magnoliids, although resolved as sister to all other *Mesangiospermae*, is not well supported and disagrees with topologies inferred from nuclear data. Relationships among the five major clades of *Mesangiospermae* remain intractable despite increased sampling, probably due to an ancient rapid radiation.

Conclusions: We provide the most comprehensive dataset of plastomes to date and a well-resolved phylogenetic tree, which together provide a strong foundation for future evolutionary studies of flowering plants.

Keywords: *Mesangiospermae*, Tree of life, Interfamilial relationships, Plastome, PPA II

Background

Angiosperms, or flowering plants, are by far the largest, most diverse, and most species-rich clade of green plants, with estimates of the number of species ranging from ~295,000 [1] to ~370,000 [2]. Traditionally,

angiosperms were divided into two fundamental groups on the basis of cotyledon numbers, i.e., monocotyledons (monocots or Monocotyledoneae) and dicotyledons (dicots or Dicotyledoneae). Toward the end of the twentieth century, in several morphologically based cladistic analyses (e.g., [3, 4]), the monocots remained as a well-defined group with uniaiperturate or uniaiperturate-derived pollen, but the traditionally defined dicots were recovered as non-monophyletic. The majority of “dicots” formed a well-supported clade termed the tricolpates or

* Correspondence: tingshuangyi@mail.kib.ac.cn; dzl@mail.kib.ac.cn

[†]Hong-Tao Li, Yang Luo, Lu Gan, Peng-Fei Ma, Lian-Ming Gao, Jun-Bo Yang and Jie Cai contributed equally to this work.

¹Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

eudicots [5] based on their triaperturate or triaperturate-derived pollen. These findings were corroborated in subsequent DNA-based phylogenetic studies, and the composition and placement of many of the remaining highly heterogeneous non-eudicots were clarified [6–11]. Among extant angiosperms, three small clades, Amborellales (1 species), Nymphaeales (88 species), and Austrobaileyales (94 species), collectively referred to as the ANA grade, represent the first-branching clades [8, 9]. The remainder belongs to a highly supported clade referred to as core angiosperms or *Mesangiospermae* ([12], comprising over 99.9% of extant angiosperm species), which was resolved into five clades (e.g., [13–17]): eudicots (~210,600 species), monocots (~74,300 species), magnoliids (*Magnoliidae* of [12]; ~10,800 species), Chloranthales (77 species), and Ceratophyllales (four species) [1, 12, 13]. These findings provide a firm understanding of the major clades of angiosperms, reflected in the widely accepted classification of the Angiosperm Phylogeny Group (APG; most recently, APG IV [18]).

During the past three decades, many molecular phylogenetic studies have achieved great progress in clarifying the backbone relationships of angiosperms [7, 9–11, 13, 14, 19–27]. However, the phylogenetic relationships among the eight major clades have remained controversial, hindering our understanding of the origin and early diversification of angiosperms. The debate on whether Amborellales alone or Amborellales + Nymphaeales are sisters to all other extant angiosperms is resolved, with all recent studies supporting Amborellales alone as a sister (e.g., [13, 28–30]). In contrast, the relationships among the five clades of *Mesangiospermae* are far more uncertain, and many contrasting topologies have been recovered from different datasets (nuclear, plastid, or mitochondrial), analytical methods (e.g., concatenation vs. coalescent), and taxon sampling [15, 21, 27, 31–33]. Furthermore, recent phylogenetic analyses with broad taxon sampling but genes from different genomes (2351 angiosperm plastomes in [13]; 682 angiosperm transcriptomes in [28] and 3099 angiosperm samples with target sequence capture data in [34]) yield conflicting topologies. Analyses based on recently sequenced genomes from key major clades of Ceratophyllales, magnoliids, and Nymphaeales [29, 35–39], with their limited taxon sampling, highlight this phylogenetic complexity with major highly conflicting signal.

Through a series of phylogenetic studies that applied broad taxon sampling with a small number of genes [9, 19, 24, 40], more limited taxon sampling with a large number of genes [15, 20–22], or both extensive taxon sampling and many genes [13, 28, 34, 41], great progress has been realized in resolving relationships among the eudicot clades, long recognized taxonomically as angiosperm orders and families (APG IV, which we will use

here for discussion). Hereafter, these clades are also referred to as orders and families following APG IV for clarity and simplicity. However, these analyses either could not resolve or did not produce congruent results for certain parts of the angiosperm tree: (1) the placements of Dilleniales, Saxifragales, Vitales, Santalales, Berberidopsidales, and Caryophyllales in the core eudicots; (2) the interordinal relationships within asterids; and (3) the phylogenetic position and inter ordinal relationships of the Celastrales-Oxalidales-Malpighiales (COM) clade. Moreover, some interfamilial relationships within orders such as Malpighiales, Saxifragales, Commelinales, and Rosales were also not fully resolved.

Phylogenetic analyses based on plastid genes, and more recently complete or nearly complete plastomes, have led the way in reconstructing the phylogenetic backbone for angiosperms over the past three decades [6, 19, 23–25, 27, 42]. Plastomes, usually mapped as circular genomes, have numerous advantages for phylogenetic reconstruction, including mostly uniparental inheritance and a relatively conserved rate of evolution [41]. Recent advances in sequencing technology have made the acquisition of complete plastomes both practical and cost-effective, and an explosion of plastid phylogenomic studies has provided critical insights into historically difficult relationships of the major angiosperm subclades [22, 26, 43–45]. Our previous work [13], the then-largest plastid phylogenomic angiosperm (PPA) tree comprising 2351 angiosperm species representing 353 families and all 64 then-recognized orders, provided a significant advance towards a robust familial-level tree for angiosperms. However, 63 angiosperm families recognized by APG IV [18] and other 10 of 17 newly recognized families recorded by the Angiosperm Phylogeny Website (hereafter abbreviated as APW, last accessed May 23, 2019, [46]) but not recognized by APG IV were missing from the PPA tree, the remaining seven newly recognized families by APW were previously sampled in the PPA tree as genera of other families. These 73 omissions have precluded a full assessment of phylogenetic relationships among all angiosperm families.

In this study, we aim to better resolve evolutionary relationships of angiosperms at the familial level and above by analyzing the largest plastome dataset ever assembled for this purpose. Compared to our previous PPA project [13], the number of angiosperm plastomes has greatly increased from 2694 (1390 genera) to 4627 (2024 genera), a 66.3% increase in samples and a 45.6% increase in generic coverage, and all 433 recognized angiosperm families in APW [46], which provides narrower family circumscriptions than those of the APG system based on recent publications, were sampled accordingly. Our goals are to consolidate plastome-based phylogenetic relationships of the major clades recognized as families, orders,

or more inclusive clades, provide additional perspectives on the early evolutionary history of angiosperms, and provide a robust plastome-based topology for comparison with studies based on the nuclear genome.

Results

Characteristics of the dataset

Our dataset comprised 4792 samples for initial analysis, including 4627 samples representing 4498 angiosperm species from all currently recognized families and orders of angiosperms and 165 samples representing 162 gymnosperm species as the outgroup (Additional file 1: Table S1). The taxonomic circumscription within seed plants followed APW [46]. Using our 86 newly sequenced plastomes representing 57 angiosperm families along with the recently issued plastomes from GenBank, we completed the representatives of 73 families absent from previous work [13] in the current analysis. The alignment of 80 genes from 4792 taxa had < 10% gaps/missing data. To our knowledge, this is the first phylogenomic study to include all currently recognized angiosperm families in APW [46] with plastome data. Overall, plastid phylogenomic analyses resulted in a tree referred to herein as the “PPA II tree” (Figs. 1 and 2; Additional files 2, 3, 4, 5, 6: Figs. S1–S5) with 75% of angiosperm nodes at or above the ordinal level and 78% at or above the familial level receiving bootstrap percentages (BP) \geq 90.

The impact of heterotrophic taxa on phylogenetic inferences

Five heterotrophic families lacked clear phylogenetic positions in our analyses (Additional files 6, 7, 8: Figs. S5–S7). One of these, Rafflesiaceae, was nested within its host family Vitaceae of Vitales with moderate support (BP = 83); similar relationships were also recovered by Molina et al. [47], which suggests that these plastid gene sequences are from the host plant. Thus, Rafflesiaceae were excluded from subsequent analyses. Four other heterotrophic families, Apodanthaceae, Balanophoraceae, Mitrastemonaceae, and Thismiaceae, with long branches, formed a strongly supported “clade” (BP = 100) within Saxifragales, as sister to another holoparasitic family, Cynomoriaceae, with moderate support (BP = 73) (Additional file 8: Fig. S7). Upon removal of Cynomoriaceae, this “clade” was sister to fully mycoheterotrophic *Epipogium* (Orchidaceae), again with a long branch (Additional file 9: Fig. S8a). However, when both Cynomoriaceae and *Epipogium* were removed (Additional file 9: Fig. S8b), these four families formed a “clade” with long-branched *Sarracenia* (Sarraceniaceae) and the long branch persisted upon the successive deletion of its sister in one earlier analysis (Additional file 9: Figs. S8c to S8i). The extremely long branch lengths

involving these taxa suggest a typical case of long-branch attraction, which has been used to explain unusual phylogenetic positions of some heterotrophic plants [48]. Phylogenetic analysis excluding the other four families (Apodanthaceae, Balanophoraceae, Mitrastemonaceae, and Thismiaceae) produced trees that were largely congruent with previous analyses. Moreover, removing these four families plus Cynomoriaceae significantly increased support for many nodes, especially deeper nodes in both monocots and asterids (Figs. 1 and 2 and Additional files 2, 3, 4, 5: Figs. S1–S4).

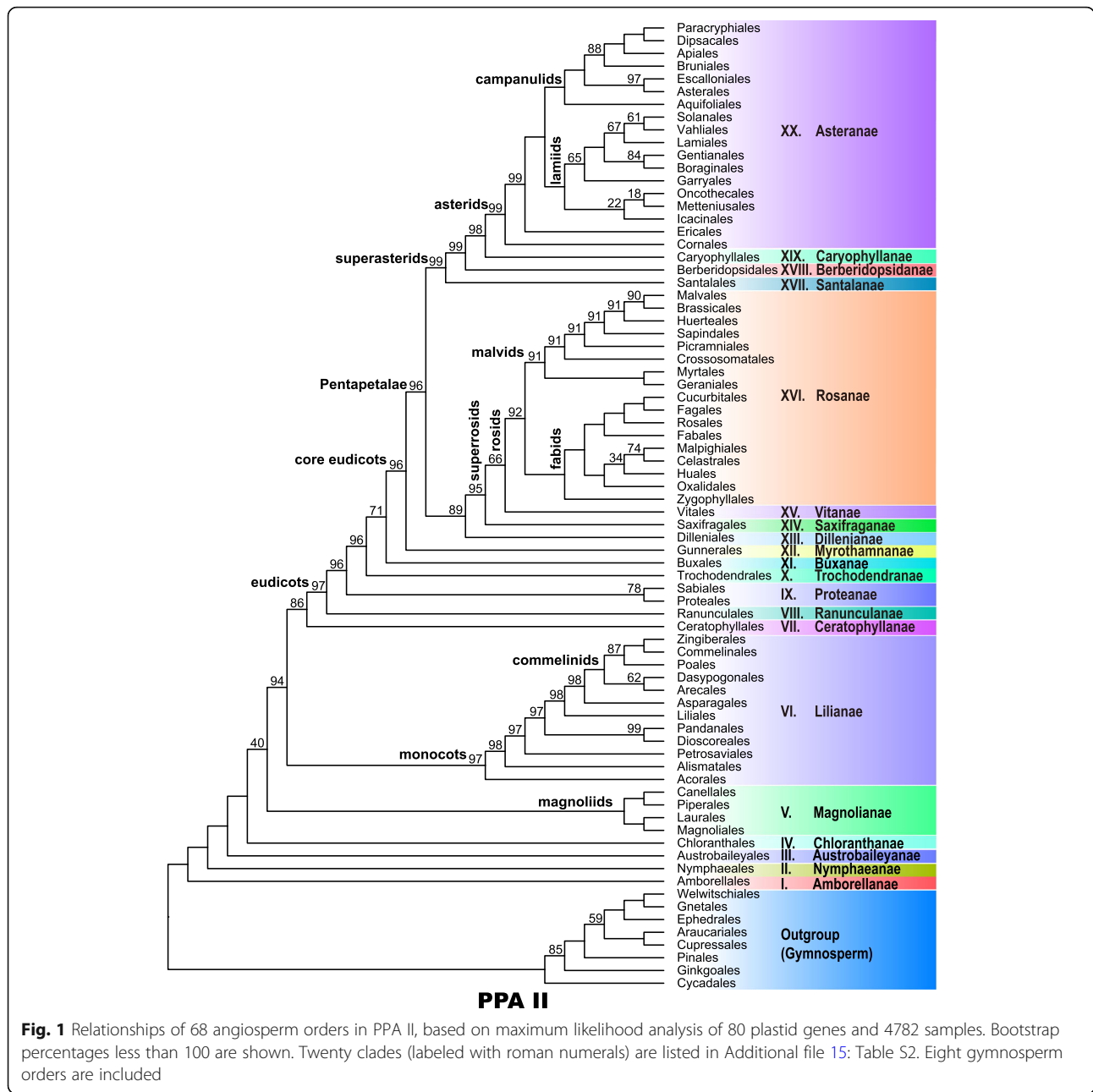
Other fully heterotrophic families seem to have consistent phylogenetic positions as resolved in previous studies. For example, Triuridaceae were supported as a member of Pandanales, Corsiaceae, and Campynemataceae formed a strongly supported (BP = 100) clade sister to all other Liliales, and Cytinaceae and Muntingiaceae formed a clade in Malvales. Phylogenetic positions of partially heterotrophic families such as Burmanniaceae (with both partially and fully mycoheterotrophic plants) and Krameriaceae (hemiparasites) that have retained a larger number of putatively functional plastid genes were resolved with high support.

Phylogenetic relationships at the ordinal level and above

In PPA II, the angiosperm clade received 100 bootstrap support (Figs. 1 and 2, Additional files 2, 3, 4, 5: Figs. S1–S4). Amborellales, Nymphaeales, and Austrobaileyales were supported as successive sisters to *Mesangiospermae* (BP = 100 for all). Although *Mesangiospermae* were strongly supported (BP = 100), relationships among its five major clades (Chloranthales, magnoliids, monocots, Ceratophyllales, and eudicots) were not fully resolved. Chloranthales, magnoliids, monocots, and Ceratophyllales were successive sisters of eudicots with BP of 100, 40, 94, and 86, respectively.

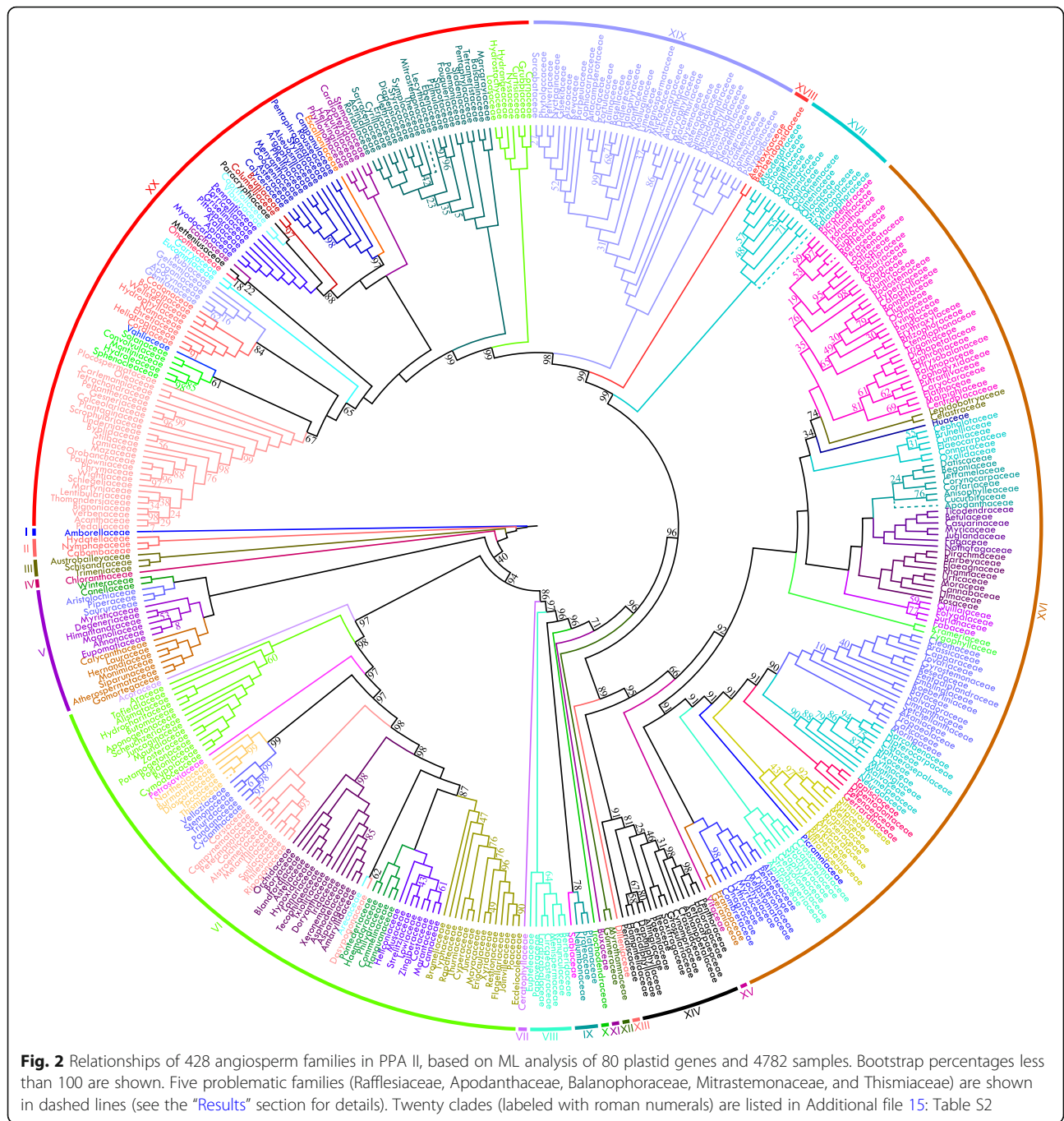
Our results provided strong support (BP = 100) for the monophyly of magnoliids and their four orders, which were further resolved into strongly supported Canellales + Piperales and Laurales + Magnoliales (both pairs BP = 100). However, two interfamilial relationships within Magnoliales were only weakly supported.

Acorales, followed by Alismatales, Petrosaviales, Dioscoreales + Pandanales, Liliales, and Asparagales were strongly supported as successive sisters to the commelinid clade (support at each node; BP = 97, 98, 97, 97, 98, 98, respectively). Within the commelinid clade (BP = 100), the weakly supported (BP = 62) clade of Dasygongonaceae + Arecales was sister (BP = 98) to a strongly supported (BP = 100) clade, within which a clade (BP = 87) comprising Poales was sister to a strongly supported (BP = 100) clade comprising Commelinales and Zingiberales. However, some interfamilial relationships within Zingiberales and Poales received low support.



The monophyly of eudicots received strong support (BP = 97), with Ranunculales sister to all other eudicots, followed by Proteales + Sabiaceae, Trochodendrales, and Buxales with strong to moderate support as successive sisters to the core eudicots (support at each node; BP = 96, 96, 71, respectively). Core eudicots were strongly supported (BP = 96), among which Gunnerales were sister to a highly supported (BP = 96) Pentapetalae, which comprised a moderately supported (BP = 89) Dilleniales + superrosids clade and strongly supported (BP = 99) superasterids.

Within Dilleniales + superrosids, Dilleniales, Saxifragales, and Vitales were weakly to strongly supported as successive sisters to the remaining rosids (support at each node; BP = 89, 95, 66, respectively). The strongly supported (BP = 92) rosids, excluding Vitales, were further divided into malvids (BP = 91) and fabids (BP = 100). Within malvids, Geraniales + Myrtales were supported as sister to the rest (BP = 91), and then Crossosomatales, Picramniales, Sapindales, and Huerteales were strongly supported (support at each node; BP = 91, 91, 91, 91, respectively) as successive sisters to Malvales +



Brassicales (BS = 90). Zygophyllales were sister to the remaining fabid clade, which was further divided into a strongly supported (BP = 100) nitrogen-fixing clade and a strongly supported (BP = 100) COM clade. Within the nitrogen-fixing clade, Fabales, Rosales, and Cucurbitales were successive sisters to Fagales (all BP = 100). Interordinal relationships of the COM clade were poorly resolved, with Huales falling in an isolated position away from Oxalidales.

Within the superasterids, Santalales were sister to the rest, and Berberidopsidales and Caryophyllales were strongly supported (support at each node; BP = 98, 99, respectively) as successive sisters of asterids, within which Cornales were sister (BP = 99) to Ericales + remaining asterids (BP = 99). The remaining asterids (BP = 100) were resolved into two strongly supported clades, campanulids and lamiids (each BP = 100). Within campanulids, Aquifoliales, Escalloniales + Asterales,

Bruniales, Apiales, and Dipsacales were successive sisters to Paracryphiales, and all interordinal campanulid relationships were well supported (BP > 85), whereas most interordinal lamiid relationships were weakly supported.

Major phylogenetic relationships at the familial level

All families with more than one sample included except Aristolochiaceae were resolved as monophyletic, and all families except Hamamelidaceae (BP = 67) were strongly supported (BP ≥ 98). To compare relationships at the interfamilial level from the current PPA II to those of previous studies, we refer to APW [46], which represents the most comprehensive current overview of interfamilial relationships based on previous studies. Our tree was largely consistent with the tree summarized in APW [46], but some incongruence was present (see Additional file 9: Fig. S8 and discussion in Additional file 14: Additional Text). Our analyses clarified some previously unresolved polytomies noted in APW (Additional file 4: Fig. S3), such as relationships among Rhamnaceae, Elaeagnaceae, Barbeyaceae, and Dirachmaceae in Rosales (Fig. 3a) [49–51], relationships among Pentadiplandraceae, Resedaceae + Gyrostemonaceae, Tovariaceae, and [Capparaceae [Cleomaceae + Brassicaceae]] in Brassicales (Fig. 3b) [50, 52–54], relationships among Meliaceae, Simaroubaceae, and Rutaceae in Sapindales [27, 50, 55, 56], relationships among Campynemataceae, Corsiaceae, and Melanthiaceae in Liliales [57–60], as well as some relationships within Malvales [50, 61, 62] and others in Cornales [63–65] (see Additional file 9: Fig. S8 and discussion in Additional file 14: Additional Text).

Our study also greatly improved support for the positions of many families, with all interfamilial relationships of 27 orders (over half of the 49 non-monofamilial orders of extant angiosperms), such as Asparagales,

Asterales, Commelinales, Crossosomatales, Fagales, Myrtales, and Rosales, being strongly supported (BP ≥ 85; Fig. 3c, d for examples, also see Fig. 2, Additional files 3, 4: Figs. S2, S3). Additionally, phylogenetic relationships of the 73 families unsampled in [13] were generally clarified (see Additional file 4: Fig. S3 for details), usually with strong support, such as Corsiaceae sister to Campynemataceae (Liliales, BP = 100), Ixioliriaceae sister to Tecophilaeaceae (Asparagales, BP = 100), Circaeasteraceae sister to Lardizabalaceae (Ranunculales, BP = 100), Anisophylleaceae sister to Cucurbitaceae (Cucurbitales, BP = 76), Stachyuraceae sister to Guamatelaceae + Crossosomataceae (Crossosomatales, BP = 100), Petenaeaceae sister to Tapisciaceae + Dipentodontaceae (Huerteales, BP = 100), Pentadiplandraceae sister to Gyrostemonaceae + Resedaceae (Brassicales, BP = 100), Tovariaceae sister to Capparaceae plus Cleomaceae + Brassicaceae (Brassicales, BP = 100), Macarthuraceae sister to Caryophyllaceae + Achatocarpaceae + Amaranthaceae (Caryophyllales, BP = 86), Loasaceae sister to Hydrostachyaceae (Cornales, BP = 100), and Namaceae sister to [Ehretiaceae [Cordiaceae + Heliotropiaceae]] (Boraginales, BP = 100). However, intractable interfamilial relationships remained in Poales [45, 66], Saxifragales [67, 68], Cucurbitales [49, 50], Oxalidales [27, 50], Malpighiales [69, 70], Santalales [71, 72], Ericales [73, 74], and Lamiales [75, 76] (see Fig. 2, Additional file 3: Fig. S2 and discussion for details in Additional file 14: Additional Text).

Phylogenetic evaluation and comparison of angiosperm family trees

Maximum likelihood (ML) and ASTRAL trees of 428 families (i.e., with five heterotrophic families removed) included in the subdataset generally showed consistent

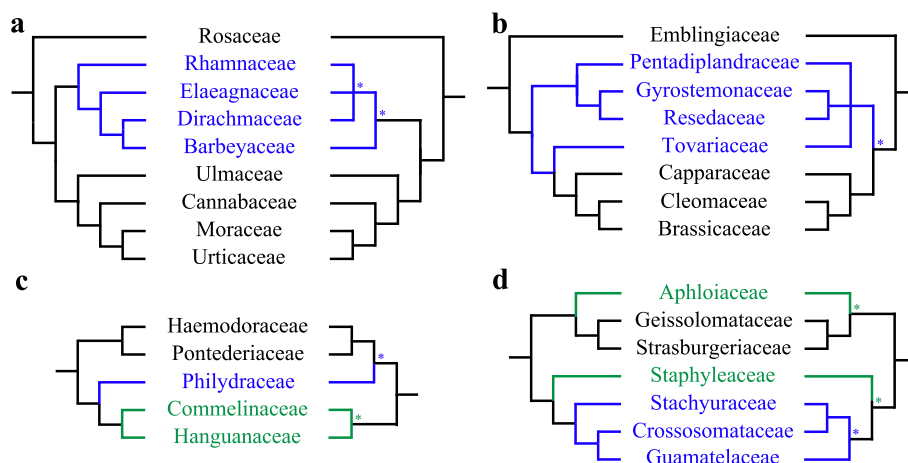


Fig. 3 Familial phylogenetic relationships in PPA II (left) versus APW (right) of Rosales (a), partial Brassicales (b), Commelinales (c), and Crossosomatales (d). All nodes in PPA II have 100 bootstrap percentages. Asterisks (*) represent the nodes with low BP in APW. The blue lines show different phylogenetic positions between PPA II and APW, and the green lines show increased support in PPA II

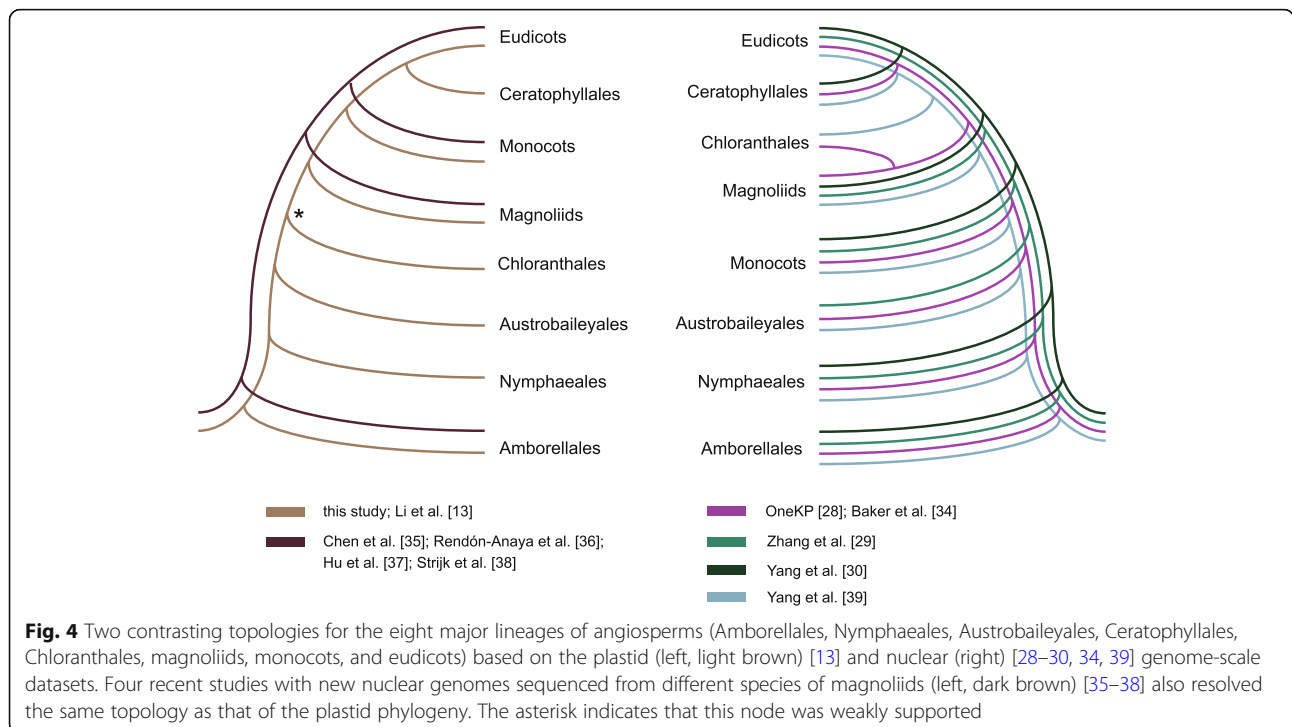
relationships with strong support, only slightly different at some nodes with weak or moderate support (Additional file 10: Fig. S9). Under Quartet Sampling (QS) evaluation, analyses of a pruned plastome dataset indicated strong support for monophyly of the majority of orders (Additional file 11: Fig. S10), but with some alternative relationships among some orders or families. Our results showed that bootstrap values and concordance factors could provide some different information about each branch in the tree, but they tended to display a similar pattern (Additional file 12: Fig. S11). Meanwhile, estimates of gene and site concordance factors (gCF and sCF) were generally correlated across the ML tree of angiosperms, but we note that both measures fell well below standard measures of bootstrap support (Additional file 13: Fig. S12).

Discussion

A plastid phylogenomic analysis including all recognized families provides an unparalleled opportunity to address interfamilial relationships of angiosperms and their associated patterns of phenotypic evolution. Our results are largely congruent with previous analyses [27] but provide higher support for many relationships among major clades, including those recognized as orders and families, and a complete phylogenetic framework of angiosperms at the familial level. Overall, our study represents the first phylogenetic analysis using complete plastomes and a large sampling of all recognized angiosperm families (except Rafflesiaceae and four other heterotrophic

families due to the complete or large number of gene losses in their plastomes), from which the phylogenetic relationships among angiosperm families, orders, and high-level clades could be addressed in a single phylogenetic tree. The higher support for many nodes may be attributed to the much better sampling of representative clades. The monophyly of the angiosperms and their division into eight major clades was supported. Amborellales, Nymphaeales, and Austrobaileyales were resolved as successive sisters to the remaining angiosperms, consistent with current understanding [13, 28–30]. The monophyly of *Mesangiospermae* received 100 BP, and a topology of [Chloranthales [magnoliids [monocots [Ceratophyllales + eudicots]]]] was well supported except for the weakly supported position of magnoliids.

This backbone plastid topology reviewed above has been consistently recovered in previous plastid phylogenomic studies [13, 21, 22]. Recent nuclear phylogenetic analyses have produced multiple topologies [13, 28–30, 34–39]. Notably, for the three clades with the highest species diversity, monocots are more closely related to eudicots than to magnoliids in the plastid tree, whereas magnoliids and eudicots are more closely related in recent nuclear trees (Fig. 4). A recent study [32] using 38 mitochondrial genes of 91 angiosperm taxa representing seven of eight major angiosperm clades (except Ceratophyllales) found that relationships among these major clades were congruent with those of the plastid tree. Nuclear-organellar discordance regarding relationships among the five major *Mesangiospermae* clades,



particularly those among monocots, magnoliids, and eudicots, may imply both rapid radiation as well as reticulate evolution in the early history of angiosperms [13, 28, 39]. More genomic data, particularly those of Chloranthales and Austrobaileyales, should be explored to address this question.

Most angiosperm interordinal relationships have been clarified on the basis of plastome analyses. For the long-controversial phylogenetic positions of a few early-diverging orders in Pentapetalae, our study and most recent plastid phylogenomic studies [13, 28] have supported Dilleniales, Saxifragales, or Vitales as successive sisters of the remaining rosids, and Santalales, Berberidopsidales, and Caryophyllales as successive sisters to the asterids. However, phylogenetic analyses of nuclear data showed substantial discordance regarding the phylogenetic positions of these orders [28, 30, 34, 77]. Dilleniales have been supported as sister to superrosids, superasterids, the remaining Pentapetalae, Gunnerales, or Caryophyllales in recent studies using nuclear gene sequence data [26, 28, 30, 78]. The uncertain position of Dilleniales hampers an accurate understanding of the origin of key trait innovations, such as pentamerous flowers and the distinction between sepals and petals in eudicots. The rapid diversification of core eudicots following two rounds of whole-genome duplication (WGD) currently hinders the confident resolution of relationships [28, 30, 79, 80].

Our study and most recent plastid phylogenomic analyses [13, 45] support the placement of the COM clade (Celastrales, Huales, Oxalidales, Malpighiales) within the fabids, but other analyses based on mitochondrial and nuclear data [15, 28, 31, 33, 81, 82] supported the COM clade within the malvids. Incomplete lineage sorting and/or ancient introgressive hybridization may be the cause of the conflicting positions for this clade [83]. All three topologies among the three large orders (Celastrales, Oxalidales, Malpighiales) within the COM clade were reported in previous studies [83], and our study also failed to resolve relationships among these three orders relative to unplaced Huales (consisting only of Huaceae). Our study provided good support for the phylogenetic positions of Escalloniales, Asterales, Boraginales, Gentianales, Vahliales, Solanales, and Lamiales within asterids. Nevertheless, our analysis did not confidently resolve some interordinal relationships, especially those within lamiids.

Our study did clarify some long-controversial interfamilial relationships within Poales, Saxifragales, Brassicales, Caryophyllales, etc. (please refer to Additional file 14: additional text for more detailed discussion). However, some previously unresolved interfamilial relationships within Saxifragales, Malpighiales, Ericales, and Lamiales [50, 68, 70, 84, 85] remain

unresolved in the current study. Families of these orders may have experienced rapid radiations, which may not be resolved by plastome data. Whereas plastome data have generally been considered to represent uniparental phylogenetic history [86, 87], more complex plastome evolution has been found in Fabaceae [86]. Previous empirical and simulated analyses have suggested that reliable inference of species trees requires the use of large numbers of nuclear loci [87–89]. Increased sampling with hundreds of single-copy nuclear genes may be needed to fully resolve these recalcitrant familial relationships.

Huaceae were placed as sister to the remaining members of Oxalidales in several previous studies, sometimes with relatively high support (BP > 80) [69, 81, 88], so that APG III [89] tentatively included Huaceae in Oxalidales. However, both our previous work [13] and current study strongly supported (BP = 100) the monophyly of Oxalidales (excluding Huaceae), and Huaceae were placed as sister to Celastrales + Malpighiales with weak support in this study (BP = 34) here. In APG IV [18], Dasygogonaceae, Sabiaceae, and Oncothecaceae were placed in Arecales, Proteales, and Icaciniales, respectively, according to the plastid phylogenomic studies of Barrett et al. [90], Sun et al. [44], and Stull et al. [43]. Nevertheless, in recent studies [45] and our study with denser taxon sampling, support for the monophyly of Arecales and Proteales was relatively low (BP < 80). In addition, a poor resolution was also apparent in the weakly supported assemblage of Icaciniales, Oncothecaceae, and Metteniusales (BP < 25). These residual issues in angiosperm phylogeny need to be settled. Thus, we suggest separating Dasygogonales from Arecales, Sabiales from Proteales, Huales from Oxalidales, and Oncothecales from Icaciniales, as the monophyly of all other orders in our tree received strong support (BP ≥ 90).

All recognized families in our study received > 95 BP support, with the exception of Aristolochiaceae and Hamamelidaceae. Aristolochiaceae were found to be paraphyletic in the current study with *Aristolochia* sister to [Saururaceae + Piperaceae] and [*Saruma* + *Asarum*] sister to that clade (Additional file 5: Fig. S4). However, we did not sample Hydnoraceae and Lactoridaceae, both recognized previously by APG III [89] but not APG IV [18]. The monophyly of Hamamelidaceae was weakly supported (BP = 67). These two cases should be the focus of further studies.

Conclusions

Our plastid phylogenomic analysis, which included representatives of all recognized angiosperm families [46], greatly clarified many deep phylogenetic relationships, particularly those at and above the familial level. The robust phylogenetic backbone presented here will provide

a firm basis for future evolutionary studies of flowering plants. Our analyses further indicate that recalcitrant relationships among the five major clades of *Mesangiospermae* and interfamilial relationships such as those of Malpighiales and a few other orders could not be resolved exclusively through increased taxonomic sampling and greater amounts of plastid data but must include the analyses of large numbers of single-copy nuclear genes.

Methods

Taxon sampling

To reconstruct the phylogenetic relationships of angiosperms at the family level, 4627 samples representing 4498 species, 2024 genera, 416 families, and 64 orders recognized by APG IV [18], and 17 additional families recognized by APW [46], were included in the analyses. In addition, 165 samples from 162 species, 77 genera, 12 families, and eight orders of gymnosperms comprised the outgroup. The dataset consisted of 86 newly sequenced plastomes with Illumina HiSeq2500, 2425 samples from our previous work [13, 91], and an additional 2281 plastomes from GenBank (released from January 1, 2017, to April 30, 2019) (Additional file 1: Table S1). The final sampling of 4792 taxa includes representatives of all 72 orders and 445 families of seed plants (Additional file 1: Tables S1 and Additional file 15: Table S2). Order and family circumscriptions of seed plants are as in APW [46].

Molecular techniques

Total genomic DNA was extracted using a modified CTAB protocol [92] from leaf tissue of herbarium specimens and silica-dried materials. The DNA samples were sheared into fragments and used to construct short-insert (500 bp) libraries in accordance with the manufacturer's manual (Illumina, San Diego, CA, USA). Paired-end sequencing of 150 bp was conducted on an Illumina HiSeq 2500. High-quality Illumina sequencing reads were assembled using the GetOrganelle toolkit [93]. The assembled plastomes were annotated using PGA [94] and manually adjusted in Geneious v9.1.8 [95]. Data from complete plastid genomes in GenBank as of April 30, 2019, were downloaded and re-annotated using PGA. For some incomplete plastomes, we used scripts to obtain assembled sequences by mapping contigs to a reference and then extracting the annotated gene fragments.

Phylogenetic inference

All alignments of protein-coding exons and rRNA genes were performed using PASTA [96] before being further locally re-aligned in Geneious v9.1.8 using MAFFT v7.394 [97] and MUSCLE v3.8.425 [98]. Three genes,

infA, *ycf1*, and *ycf2*, were difficult to align and were thus excluded from the phylogenetic analysis. We conducted analyses with and without the inclusion of five heterotrophic families, i.e., Apodanthaceae, Balanophoraceae, Mitrastemonaceae, Rafflesiaceae, and Thismiaceae, given that their plastome sequences are highly reduced and that the retained sequences have unusually high substitution rates that strongly hamper proper alignment and may cause long-branch attraction artifacts in many focal clades. However, for the completeness of the PPA tree, these families were included in the figures following their placement in APW. All aligned genes were concatenated into a supermatrix with a length of 89,357 bp. Maximum likelihood (ML) analyses were performed with RAxML v8.2.12 [99] under the GTRGAMMA model for a partitioned supermatrix. Searches for the best trees were conducted by starting from random trees, and bootstrap percentages were obtained with 1000 non-parametric bootstrap replicates.

To further evaluate the phylogenetic relationships of the backbone tree of angiosperm families, we generated a subdataset of 431 species representing 428 angiosperm families and two outgroup taxa using the Python package ete3 v3.1.2 [100] and pxrms from the phyx package [101]. Maximum likelihood analyses were conducted with RAxML v8.1.2 [99] including 500 rapid bootstraps and a search for the best-scoring tree, employing the GTRGAMMA model. We evaluated clade/branch support under various metrics of branch support including Quartet Sampling [102] with 1000 replicates, gene concordance factors (gCF) [103] and site concordance factors (sCF) [104] from IQtree v2.0 [105], and internode certainty all (ICA) [106] from RAxML v8.1.2 [99]. We compared the angiosperm phylogeny estimated with the concatenated approach and that resulted from the multispecies coalescent-based approach [107, 108] based on 80 single-gene trees from RAxML with local posterior probabilities (LPP) [109] to assess clade/branch support. Two multispecies coalescent-based analyses were executed in which all bipartitions were included and bipartitions with <10 bootstrap support were collapsed prior to the analyses.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01166-2>.

Additional file 1: Table S1. Species sampled in this study. The 4792 individuals sampled including 86 newly sequenced plastomes, involving 4498 angiosperm species and 162 gymnosperm species.

Additional file 2: Figure S1. Phylogenetic tree of 4782 plastomes of 68 orders of angiosperms.

Additional file 3: Figure S2. Phylogenetic tree of 4782 plastomes of 445 families (including 12 gymnosperm families) of seed plants. Five problematic families (Rafflesiaceae, Apodanthaceae, Balanophoraceae,

Mitrastemonaceae, and Thismiaceae) were added manually (see Results for details).

Additional file 4: Figure S3. Angiosperm family-level phylogenetic relationships in PPA II versus APW. Red: different phylogenetic positions between PPA II and APW; green: resolved nodes in PPA II relative to APW. Different phylogenetic positions between PPA II and APW with bootstrap values < 50 in PPA are not shown.

Additional file 5: Figure S4. Phylogenetic tree of 4782 plastomes (with ten plastomes of five problematic families excluded) of 4650 species of seed plants. Bootstrap values are shown.

Additional file 6: Figure S5. Phylogenetic tree of 4792 plastomes of 76 orders (including eight gymnosperm orders) of seed plants.

Additional file 7: Figure S6. Phylogenetic tree of 4792 plastomes of 445 families (including 12 gymnosperm families) of seed plants.

Additional file 8: Figure S7. Phylogenetic tree of 4792 plastomes of 4660 species of seed plants. All bootstrap values are shown.

Additional file 9: Figure S8. Phylogenetic tree of 4792 plastomes with successive removal of the long branch forming a sister relationship with a 'clade' of Mitrastemonaceae, Thismiaceae, Apodanthaceae, and Balanophoraceae.

Additional file 10: Figure S9. Topologies of a pruned Maximum Likelihood phylogeny of 431 representative species ("ML431_pruned") and ASTRAL analysis ("astral431_BS10") of a 43-species subdataset of angiosperms using 80 plastid genes. The ML bootstrap percentages and ASTRAL local posterior probabilities are shown, respectively.

Additional file 11: Figure S10. Family relationships within the pruned angiosperm phylogeny: nodes by Quartet Concordance (QC) scores for internal branches: green (QC > 0.2), blue (0.2 ≥ QC > 0), orange (0 ≥ QC ≥ -0.05, or red (QC < -0.05). QC/Quartet Differential (QD)/Quartet Informativeness (QI) scores are shown for all internal branches.

Additional file 12: Figure S11. Plot showing the relationship between gene and site concordance factors (gCF and sCF) relative to bootstrap support from the pruned angiosperm phylogeny.

Additional file 13: Figure S12. Family relationships within the pruned angiosperm subdataset. In this tree, bootstrap/gCF/sCF scores are shown for each branch.

Additional file 14: Additional Text. Overview of angiosperm phylogeny at the familial level.

Additional file 15: Table S2. Summary of all recognized 433 families, 68 orders, and more inclusive clades for flowering plants, with numbers of known genera and species.

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 Missouri Botanical Garden (BG), RBG Kew, Arnold Arboretum, New York BG, the Beijing BG, Blue Mountains BG, Brisbane BG, Kunming BG, Wuhan BG, RBG Edinburgh, 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 RBG Kew), J. R. Shevock (California), Y.-M. Shui (Kunming), and N. Zamora (Costa Rica) for sampling; B.T. Wursten for *Vahlia capensis* and *Hua gabonica* pictures; J. T. Johansson for *Paracryphia alticola*, *Oncotheaca balansae*, and *Berzelia squarrosa* pictures; and the iFlora High Performance Computing Center of Germplasm Bank of Wild Species (iFlora HPC Center of GBOWS, KIB, CAS) for computing.

Authors' contributions

D.-Z.L., T.-S.Y., H.-T.L., L.-M. G., J.-B.Y., D.E.S., and P.S.S. conceived the project and designed the research. T.-S.Y., J.C., T.Z., L.-M.G., W.-B.Y., M.v.d.B., P.M.H., collected the samples. J.-B.Y. and C.-X.Z. performed the DNA laboratory work. H.-T.L., L.G., J.-J.J., and R.Z. assembled and annotated the plastomic sequences. M.A.G., D.E.S., and P.S.S. prepared the OneKP dataset. H.-T.L. designed and coordinated computational analyses. H.-T.L., Y.L., M.-P. M., H.W., and L.G. prepared the figures and tables. Y.L., T.-S.Y., and L.G. wrote the supplementary information. Y.L., T.-S.Y., D.-Z.L., H.-T.L., P.-F.M., and L.G. wrote the first

manuscript draft with subsequent input from all co-authors, particularly P.W.F., R.G.O., M.W.C., D.E.S., and P.S.S. The author(s) read and approved the final manuscript.

Funding

This work was funded by the Strategic Priority Research Program of Chinese Academy of Sciences (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.), the National Natural Science Foundation of China with a key international (regional) cooperative research project (No.31720103903 to T.-S.Y. and D.E.S.), the Science and Technology Basic Resources Investigation Program of China (2019FY100900 to H.-T. L. and T.-S. Y.), KIB's iFlora initiative (grant No. 2014-4-11 to D.-Z.L.), the open research project of the Germplasm Bank of Wild Species, Kunming Institute of Botany, CAS (grant No. E16O8411D1 to H.-T.L.), the National Natural Science Foundation of China (grant No. 31570333 to H.-T.L.), the Yunling International High-end Experts Program of Yunnan Province, China (grant No. YNQR-GDWG-2017-002 to P.S.S. and YNQR-GDWG-2018-012 to D.E.S.), and the CAS' Youth Innovation Promotion Association (grant No. 2015321 to P.-F.M.).

Availability of data and materials

Sequence alignments underlying analyses and phylogenetic trees are available from figshare (<https://doi.org/10.6084/m9.figshare.16573115>) [110]. Raw reads of 86 new genome skims used in this study are available at the NCBI SRA database as Bioproject PRJNA767934 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA767934>) [111].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China. ²Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650201, Yunnan, China. ³CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China. ⁴Lijiang Forest Ecosystem National Observation and Research Station, Kunming Institute of Botany, Chinese Academy of Sciences, Lijiang 674100, Yunnan, China. ⁵Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. ⁶Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA. ⁷Botanical Research Institute of Texas, 1700 University Drive, Fort Worth, TX 76017, USA. ⁸Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10025, USA. ⁹Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, Yunnan, China. ¹⁰Department of Botany & Plant Biotechnology, University of Johannesburg, PO Box 524, Auckland Park, Johannesburg, Gauteng 2006, South Africa. ¹¹Department of Biology and Burke Museum, University of Washington, Seattle, WA 98195-5325, USA. ¹²Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, Scotland, UK. ¹³Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, England, UK. ¹⁴Department of Environment and Agriculture, Curtin University, Bentley, Western Australia 6102, Australia. ¹⁵Department of Biology, University of Florida, Gainesville, FL 32611, USA.

Received: 17 June 2021 Accepted: 14 October 2021

Published online: 29 October 2021

References

- Christenhusz MJ, Byng JW. The number of known plants species in the world and its annual increase. *Phytotaxa*. 2016;261(3):201–17. <https://doi.org/10.11646/phytotaxa.261.3.1>.

2. Lughadha EN, Govaerts R, Belyaeva I, Black N, Lindon H, Allkin R, et al. 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. <https://doi.org/10.11646/phytotaxa.272.1.5>.
3. Donoghue MJ, Doyle JA. Phylogenetic studies of seed plants and angiosperms based on morphological characters. In: Fernholm B, Bremer K, Jornvall H, editors. *The hierarchy of life: molecules and morphology in phylogenetic analysis*. Amsterdam: Excerpta Medica; 1989. p. 181–93.
4. Donoghue MJ, Doyle JA. Phylogenetic analysis of angiosperms and the relationships of Hamamelidaceae. In: Crane PR, Blackmore S, editors. *Evolution, systematics and fossil history of the Hamamelidaceae*. Oxford: the Clarendon Press; 1989. p. 17–45.
5. Doyle JA, Hotton CL. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore S, Barnes SH, editors. *Pollen spores: patterns of diversification*. Oxford: the Clarendon Press; 1991. p. 169–95.
6. Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann Mo Bot Gard*. 1993;80(3):528–80. <https://doi.org/10.2307/2399846>.
7. Mathews S, Donoghue MJ. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science*. 1999;286(5441):947–50. <https://doi.org/10.1126/science.286.5441.947>.
8. Qiu YL, Lee JH, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, et al. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature*. 1999;402(6760):404–7. <https://doi.org/10.1038/46536>.
9. Soltis PS, Soltis DE, Chase MW. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*. 1999;402(6760):402–4. <https://doi.org/10.1038/46528>.
10. Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, et al. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc Natl Acad Sci USA*. 2000;97(24):13166–71. <https://doi.org/10.1073/pnas.220427497>.
11. Graham SW, Olmstead RG. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am J Bot*. 2000;87(11):1712–30. <https://doi.org/10.2307/2656749>.
12. Cantino PD, Doyle JA, Graham SW, Judd WS, Olmstead RG, Soltis DE, et al. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon*. 2007;56(3):822–46. <https://doi.org/10.2307/25065865>.
13. Li HT, Yi TS, Gao LM, Ma P-F, Zhang T, Yang J-B, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 2019;5(5):461–70. <https://doi.org/10.1038/s41477-019-0421-0>.
14. Soltis PS, Soltis DE, Zanis MJ, Kim S. Basal lineages of angiosperms: relationships and implications for floral evolution. *Int J Plant Sci*. 2000;161(5):S97–S107. <https://doi.org/10.1086/317581>.
15. Zeng LP, Zhang Q, Sun RR, Kong H, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun*. 2014;5(1):4956. <https://doi.org/10.1038/ncomms5956>.
16. APG II. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc*. 2003;141(4):399–436. <https://doi.org/10.1046/j.1095-8339.2003.t01-1-00158.x>.
17. Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, et al. Phylogeny of basal angiosperms: analyses of five genes from three genomes. *Int J Plant Sci*. 2000;161(5):S3–S27. <https://doi.org/10.1086/317584>.
18. APG IV. 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. <https://doi.org/10.1111/boj.12385>.
19. Hillu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, et al. Angiosperm phylogeny based on *matK* sequence information. *Am J Bot*. 2003;90(12):1758–76. <https://doi.org/10.3732/ajb.90.12.1758>.
20. Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, et al. 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. <https://doi.org/10.1093/molbev/msi191>.
21. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA*. 2007;104(49):19369–74. <https://doi.org/10.1073/pnas.0709121104>.
22. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA*. 2007;104(49):19363–8. <https://doi.org/10.1073/pnas.0708072104>.
23. Soltis DE, Gitzendanner MA, Soltis PS. A 567-taxon data set for angiosperms: the challenges posed by Bayesian analyses of large data sets. *Int J Plant Sci*. 2007;168(2):137–57. <https://doi.org/10.1086/509788>.
24. Burleigh JG, Hillu KW, Soltis DE. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evol Biol*. 2009;9(1):61. <https://doi.org/10.1186/1471-2148-9-61>.
25. Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. *Am J Bot*. 2010;97(8):1296–303. <https://doi.org/10.3732/ajb.0900346>.
26. 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 USA*. 2010;107(10):4623–8. <https://doi.org/10.1073/pnas.0907801107>.
27. Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, et al. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot*. 2011;98(4):704–30. <https://doi.org/10.3732/ajb.1000404>.
28. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019;7780:574:679. <https://doi.org/10.1038/s41586-019-1693-2>.
29. Zhang LS, Chen F, Zhang XT, Li Z, Zhao Y, Lohaus R, et al. The water lily genome and the early evolution of flowering plants. *Nature*. 2020;577(7788):1–6. <https://doi.org/10.1038/s41586-019-1852-5>.
30. Yang LX, Su DY, Chang X, Foster CSP, Sun LH, Huang CH, et al. Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant Commun*. 2020;1(2):100027. <https://doi.org/10.1016/j.xplc.2020.100027>.
31. Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA*. 2014;111(45):E4859–68. <https://doi.org/10.1073/pnas.1323926111>.
32. Dong S, Chen L, Liu Y, Wang Y, Zhang S, Yang L, et al. The draft mitochondrial genome of *Magnolia biondii* and mitochondrial phylogenomics of angiosperms. *PLoS One*. 2020;15(4):e0231020. <https://doi.org/10.1371/journal.pone.0231020>.
33. Qiu YL, Li LB, Wang B, Xue JY, Hendry TA, Li RQ, et al. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *J Syst Evol*. 2010;48(6):391–425. <https://doi.org/10.1111/j.1759-6831.2010.00097.x>.
34. Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, et al. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Syst Biol*. 2021;syab035. <https://doi.org/10.1093/sysbio/syab035>.
35. Chen JH, Hao ZD, Guang XM, Zhao CX, Wang PK, Xue LJ, et al. *Liriodendron* genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nat Plants*. 2019;5(1):18–25. <https://doi.org/10.1038/s41477-018-0323-6>.
36. Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan T, Zheng C, Carretero-Paulet L, et al. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen influenced gene space adaptation. *Proc Natl Acad Sci USA*. 2019;116(34):17081–9. <https://doi.org/10.1073/pnas.1822129116>.
37. Hu LS, Xu ZP, Wang MJ, Fan R, Yuan D, Wu B, et al. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat Commun*. 2019;10(1):1–11. <https://doi.org/10.1038/s41467-019-12607-6>.
38. Strijk JS, Hinsinger DD, Roeder MM, Chatrou LW, Couvreur TLP, Erkens RHJ, et al. Chromosome-level reference genome of the soursop (*Annona muricata*): a new resource for Magnoliid research and tropical pomology. *Mol Ecol Resour*. 2021;21(5):1608–19. <https://doi.org/10.1111/1755-0998.13353>.
39. Yang YZ, Sun PC, Lv LK, Wang D, Ru D, Li Y, et al. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat Plants*. 2020;6(3):215–22. <https://doi.org/10.1038/s41477-020-0594-6>.
40. Davis JJ, Stevenson DW, Petersen G, Seberg O, Campbell LM, Freudenstein JV, et al. A phylogeny of the monocots, as inferred from *rbcl* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Syst Bot*. 2004;29(3):467–510. <https://doi.org/10.1600/0363644041744365>.

41. 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. <https://doi.org/10.1002/ajb.1048>.
42. Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, et al. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences. *Syst Biol.* 2000;49(2):306–62. <https://doi.org/10.1093/sysbio/49.2.306>.
43. Stull GW, Duno de Stefano R, Soltis DE, Soltis PS. Resolving basal lamiid phylogeny and the circumscription of Icacinaeace with a plastome-scale data set. *Am J Bot.* 2015;102(11):1794–813. <https://doi.org/10.3732/ajb.1500298>.
44. Sun YX, Moore MJ, Zhang S, Soltis PS, Soltis DE, Zhao T, et al. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Mol Phylogenet Evol.* 2016;96:93–101. <https://doi.org/10.1016/j.ympev.2015.12.006>.
45. Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VK, Saarela JM, 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. <https://doi.org/10.1002/ajb.21178>.
46. Stevens PF. Angiosperm phylogeny website. Version 14 [more or less continuously updated]. <http://www.mobot.org/MOBOT/research/APweb/> 2001 onwards. Accessed 23 May 2019.
47. Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, et al. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol.* 2014;31(4):793–803. <https://doi.org/10.1093/molbev/msu051>.
48. Nickrent DL, Duff RJ, Colwell AE, Wolfe AD, Young ND, Steiner KE, et al. Molecular phylogenetic and evolutionary studies of parasitic plants. In: Soltis DE, Soltis PS, Doyle JJ, editors. *Molecular systematics of plants II*. Boston: Springer; 1998. p. 211–41. https://doi.org/10.1007/978-1-4615-5419-6_8.
49. Li HL, Wang W, Mortimer PE, Li RQ, Li DZ, Hyde KD, et al. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci Rep.* 2015; 5(1):14023. <https://doi.org/10.1038/srep14023>.
50. Sun M, Naeem RH, Su JX, Cao ZY, Burleigh JG, Soltis PS, et al. Phylogeny of the Rosidae: a dense taxon sampling analysis. *J Syst Evol.* 2016;54(4):363–91. <https://doi.org/10.1111/jse.12211>.
51. Zhang SD, Soltis DE, Yang Y, Li DZ, Yi TS. Multi-gene analysis provides a well-supported phylogeny of Rosales. *Mol Phylogenet Evol.* 2011;60(1):21–8. <https://doi.org/10.1016/j.ympev.2011.04.008>.
52. Su JX, Wang W, Zhang LN, Chen ZD. Phylogenetic placement of two enigmatic genera, *Borthwickia* and *Stixis*, based on molecular and pollen data, and description of a new family of Brassicales. *Taxon.* 2012;61(6):601–11. <https://doi.org/10.1093/aob/mcs197>.
53. Cardinal-McTeague WM, Sytsma KJ, Hall JC. Biogeography and diversification of Brassicales: a 103 million year tale. *Mol Phylogenet Evol.* 2016;99:204–24. <https://doi.org/10.1016/j.ympev.2016.02.021>.
54. Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, et al. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am J Bot.* 2018;105(3):463–9. <https://doi.org/10.1002/ajb.21040>.
55. Muellner-Riehl AN, Weeks A, Clayton JW, Buerki S, Nauheimer L, Chiang Y-C, et al. Molecular phylogenetics and molecular clock dating of Sapindales based on plastid *rbcl*, *atpB* and *trnL-trnF* DNA sequences. *Taxon.* 2016;65(5): 1019–36. <https://doi.org/10.12705/655.5>.
56. Koenen EJM, Clarkson JJ, Pennington TD, Chatrou LW. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytol.* 2015; 207(2):327–39. <https://doi.org/10.1111/nph.13490>.
57. Kim JS, Hong J-K, Chase MW, Fay MF, Kim J-H. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: *matK*, *rbcl*, *atpB* and *atpF-H*. *Bot J Linn Soc.* 2013;172(1):5–21. <https://doi.org/10.1111/boj.12039>.
58. Givnish TJ, Zuluaga A, Marques I, Lam VKY, Gomez MS, Iles WJD, et al. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. *Cladistics.* 2016;2016(6):1–25. <https://doi.org/10.1111/cla.12153>.
59. Lam VK, Merckx VS, Graham SW. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *Am J Bot.* 2016;103(4):692–708. <https://doi.org/10.3732/ajb.1500412>.
60. Lam VKY, Darby H, Merckx V, Lim G, Yukawa T, Neubig KM, et al. Phylogenomic inference in extremis: a case study with mycoheterotrophic plastomes. *Am J Bot.* 2018;105(3):480–94. <https://doi.org/10.1002/ajb.21070>.
61. Le Péchon T, Gigord LD. On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l., and Dombeyoideae. In: Walker JM, editor. *Methods in molecular biology*, vol. 1115; 2014. p. 337–63. https://doi.org/10.1007/978-1-62703-767-9_17.
62. Hernandez-Gutierrez R, Magallon S. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Mol Phylogenet Evol.* 2019;140:106606. <https://doi.org/10.1016/j.ympev.2019.106606>.
63. Fu CN, Mo ZQ, Yang JB, Ge XJ, Li DZ, Xiang QJ, et al. Plastid phylogenomics and biogeographic analysis support a trans-Tethyan origin and rapid early radiation of Cornales in the Mid-Cretaceous. *Mol Phylogenet Evol.* 2019;140: 106601. <https://doi.org/10.1016/j.ympev.2019.106601>.
64. Schenk JJ, Hufford L. Effects of substitution models on divergence time estimates: simulations and an empirical study of model uncertainty using Cornales. *Syst Bot.* 2010;35(3):578–92. <https://doi.org/10.1600/036364410792495809>.
65. Xiang QY, Thomas DT, Xiang QP. Resolving and dating the phylogeny of Cornales: effects of taxon sampling, data partitions, and fossil calibrations. *Mol Phylogenet Evol.* 2011;59(1):123–38. <https://doi.org/10.1016/j.ympev.2011.01.016>.
66. Bouchenak-Khelladi Y, Muasya AM, Linder HP. A revised evolutionary history of Poales: origins and diversification. *Bot J Linn Soc.* 2014;175(1):4–16. <https://doi.org/10.1111/boj.12160>.
67. Chen ZD, Yang T, Lin L, Lu L-M, Li H-L, Sun M, et al. Tree of life for the genera of Chinese vascular plants. *J Syst Evol.* 2016;54(4):273–6. <https://doi.org/10.1111/jse.12219>.
68. Dong WP, Xu C, Wu P, Cheng T, Yu J, Zhou S, et al. Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales. *Mol Phylogenet Evol.* 2018;126:321–30. <https://doi.org/10.1016/j.ympev.2018.04.033>.
69. 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. <https://doi.org/10.3732/ajb.0800207>.
70. Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, et al. Angiospermatics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc Natl Acad Sci USA.* 2012;109(43): 17519–24. <https://doi.org/10.1073/pnas.1205818109>.
71. Nickrent DL, Malécot VR, Vidal-Russell R, Der JP. A revised classification of Santalales. *Taxon.* 2010;59(2):538–58. <https://doi.org/10.1002/tax.592019>.
72. Su HJ, Hu JM, Anderson FE, Der JP, Nickrent DL. Phylogenetic relationships of Santalales with insights into the origins of holoparasitic Balanophoraceae. *Taxon.* 2015;64(3):491–506. <https://doi.org/10.12705/643.2>.
73. Larson DA, Walker JF, Vargas OM, Smith SA. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *Am J Bot.* 2020;107(5):773–89. <https://doi.org/10.1002/ajb.21469>.
74. Yan M, Fritsch PW, Moore MJ, Feng T, Meng A, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. *Mol Phylogenet Evol.* 2018;121:198–211. <https://doi.org/10.1016/j.ympev.2018.01.004>.
75. Refulio-Rodriguez NF, Olmstead RG. Phylogeny of lamiidae. *Am J Bot.* 2014; 101(2):287–99. <https://doi.org/10.3732/ajb.1300394>.
76. Xu W-Q, Losh J, Chen C, Li P, Wang R-H, Zhao Y-P, et al. Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. *J Syst Evol.* 2018;57(1):55–65. <https://doi.org/10.1111/jse.12421>.
77. Zeng LP, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytol.* 2017;214(3):1338–54. <https://doi.org/10.1111/nph.14503>.
78. 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(1):23. <https://doi.org/10.1186/1471-2148-14-23>.
79. Jaillon O, Aury J-M, Noel B, Polcristi A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major

- angiosperm phyla. *Nature*. 2007;449(7161):463–7. <https://doi.org/10.1038/nature06148>.
80. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 2012;13(1):R3. <https://doi.org/10.1186/gb-2012-13-1-r3>.
 81. Zhu XY, Chase MW, Qiu YL, Kong H-Z, Dilcher DL, Li J-H, et al. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evol Biol*. 2007;7(1):217. <https://doi.org/10.1186/1471-2148-7-217>.
 82. Burleigh JG, Bansal MS, Eulenstein O, Hartmann S, Wehe A, Vision TJ. Genome-scale phylogenetics: inferring the plant tree of life from 18,896 gene trees. *Syst Biol*. 2010;60(2):117–25. <https://doi.org/10.1093/sysbio/syq072>.
 83. Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Mol Phylogenet Evol*. 2015; 83:156–66. <https://doi.org/10.1016/j.ympev.2014.11.003>.
 84. Luna JA, Richardson JE, Nishii K, Clark JL, Möller M. The family placement of *Cyrtandromoea*. *Syst Bot*. 2019;44(3):616–30. <https://doi.org/10.1600/036364419x15620113920653>.
 85. Rose JP, Kleist TJ, Lofstrand SD, Drew BT, Schoenenberger J, Sysma KJ. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol Phylogenet Evol*. 2018;122:59–79. <https://doi.org/10.1016/j.ympev.2018.01.014>.
 86. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. *Genetics*. 1992;132(2):583–9. <https://doi.org/10.1093/genetics/132.2.583>.
 87. Maddison WP. Gene trees in species trees. *Syst Biol*. 1997;46(3):523–36. <https://doi.org/10.2307/2413694>.
 88. Zhang LB, Simmons MP. Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Syst Bot*. 2006;31(1):122–37. <https://doi.org/10.1600/036364406775971778>.
 89. APG III. 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):105–21. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>.
 90. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytol*. 2016;209(2):855–70. <https://doi.org/10.1111/nph.13617>.
 91. Li HT, Yi TS, Gao LM, Ma P-F, Zhang T, Yang J-B, et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Dryad Dataset*. 2019; <https://doi.org/10.5061/dryad.bq091cg>.
 92. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour*. 2014;14(5):1024–31. <https://doi.org/10.1111/1755-0998.12251>.
 93. Jin JJ, Yu WB, Yang JB, Song Y, de Pamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
 94. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15(1):50. <https://doi.org/10.1186/s13007-019-0435-7>.
 95. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, 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. <https://doi.org/10.1093/bioinformatics/bts199>.
 96. Mirarab S, Nguyen N, Warnow T. PASTA: ultra-large multiple sequence alignment. In: *International conference on research in computational molecular biology*. Pittsburgh: Springer; 2014. p. 177–91. https://doi.org/10.1007/978-3-319-05269-4_15.
 97. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
 98. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
 99. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
 100. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 2016;33(6):1635–8. <https://doi.org/10.1093/molbev/msw046>.
 101. Brown JW, Walker JF, Smith SA. Phyx: phylogenetic tools for Unix. *Bioinformatics*. 2017;33(12):1886–8. <https://doi.org/10.1093/bioinformatics/btx063>.
 102. Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am J Bot*. 2018;105(3):385–403. <https://doi.org/10.1002/ajb2.1016>.
 103. Ané C, Larget B, Baum DA, Smith SD, Rokas A. Bayesian estimation of concordance among gene trees. *Mol Biol Evol*. 2007;24(2):412–26. <https://doi.org/10.1093/molbev/msl170>.
 104. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol*. 2020a;37(9):2727–33. <https://doi.org/10.1093/molbev/msaa106>.
 105. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020b;37(5):1530–4. <https://doi.org/10.1101/849372>.
 106. Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol Biol Evol*. 2014;31(5):1261–71. <https://doi.org/10.1093/molbev/msu061>.
 107. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time. species tree reconstruction from partially resolved gene trees. *BMC Bioinform*. 2018; 19(S6):153. <https://doi.org/10.1186/s12859-018-2129-y>.
 108. Doyle JJ. Defining coalescent genes: theory meets practice in organelle phylogenomics. *Syst Biol*. 2021:syb053 <https://doi.org/10.1093/sysbio/syab053>.
 109. Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol*. 2016;33(7):1654–68. <https://doi.org/10.1093/molbev/msw079>.
 110. Li HT, Luo Y, Gan L, Ma P-F, Gao LM, Yang J-B, et al. Plastid phylogenomic insights into relationships of all flowering plant families. *Figshare Dataset*. 2021; <https://doi.org/10.6084/m9.figshare.16573115>.
 111. Li HT, Luo Y, Gan L, Ma P-F, Gao LM, Yang J-B, et al. Plastid phylogenomic insights into relationships of all flowering plant families: GenBank; 2021. <https://www.ncbi.nlm.nih.gov/sra/PRJNA767934>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

