Molecular Phylogenetics and Evolution 134 (2019) 74–86

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Plastid phylogenomic insights into the evolution of Caryophyllales

Gang Yao[a,b,1], Jian-Jun Jin[b,c,1], Hong-Tao Li[a], Jun-Bo Yang[a], Venkata Shiva Mandala[d,e], Matthew Croley[d], Rebecca Mostow[d,2], Norman A. Douglas[f], Mark W. Chase[g,h], Maarten J.M. Christenhusz[g,i], Douglas E. Soltis[f,j], Pamela S. Soltis[j], Stephen A. Smith[k], Samuel F. Brockington[l], Michael J. Moore[d], Ting-Shuang Yi[a,b,*], De-Zhu Li[a,b,*]

[a] Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[b] Key Laboratory for Plant Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[c] Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[d] Department of Biology, Oberlin College, Oberlin, OH 44074, USA
[e] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[f] Department of Biology, University of Florida, Gainesville, FL 32611-8525, USA
[g] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK
[h] School of Biological Sciences, The University of Western Australia, Crawley, WA 6009, Australia
[i] Department of Environment and Agriculture, Curtin University, Bentley, WA 6102, GPO Box U1987, Perth, Western Australia 6845, Australia
[j] Florida Museum of Natural History, University of Florida, Gainesville, FL 32611-7800, USA
[k] Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA
[l] Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

ARTICLE INFO

Keywords:
Caryophyllales
Gene loss
Molecular dating
Phylogenomics
Plastome
Rapid radiation
Substitution rate

ABSTRACT

The Caryophyllales includes 40 families and 12,500 species, representing a large and diverse clade of angiosperms. Collectively, members of the clade grow on all continents and in all terrestrial biomes and often occupy extreme habitats (e.g., xeric, salty). The order is characterized by many taxa with unusual adaptations including carnivory, halophytism, and multiple origins of C_4 photosynthesis. However, deep phylogenetic relationships within the order have long been problematic due to putative rapid divergence. To resolve the deep-level relationships of Caryophyllales, we performed phylogenomic analyses of all 40 families of Caryophyllales. We time-calibrated the molecular phylogeny of this clade, and evaluated putative correlations among plastid structural changes and rates of molecular substitution. We recovered a well-resolved and well-supported phylogeny of the Caryophyllales that was largely congruent with previous estimates of this order. Our results provide improved support for the phylogenetic position of several key families within this clade. The crown age of Caryophyllales was estimated at ca. 114.4 million years ago (Ma), with periods of rapid divergence in the mid-Cretaceous. A strong, positive correlation between nucleotide substitution rate and plastid structural changes was detected. Our study highlights the importance of broad taxon sampling in phylogenomic inference and provides a firm basis for future investigations of molecular, morphological, and ecophysiological evolution in Caryophyllales.

1. Introduction

Over the past two decades, enormous progress has been made in clarifying phylogenetic relationships among angiosperms (Soltis et al., 1999; Soltis et al., 2011; APG, 2016; Zeng et al., 2017). However, the establishment of the phylogenetic position of clades that have experienced rapid radiation has been difficult (Whitfield and Lockhart, 2007; Moore et al., 2010; Xi et al., 2012). Determining the accurate phylogenetic position of clades is often complicated by long branches (Whitfield and Lockhart, 2007; Philippe et al., 2011), where mutational saturation can lead to long-branch attraction (LBA) (Felsenstein, 1978; Bergsten, 2005). This problem is particularly acute when taxon sampling is sparse. In addition, incomplete lineage sorting (ILS) typically leads to the non-coalescence of haplotypes during rapid radiations,

* Corresponding authors at: Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China.
E-mail addresses: tingshuangyi@mail.kib.ac.cn (T.-S. Yi), dzl@mail.kib.ac.cn (D.-Z. Li).
[1] These authors contributed equally to this work.
[2] Current address: Department of Integrative Biology, Oregon State University, Corvallis, OR 97331, USA.

https://doi.org/10.1016/j.ympev.2018.12.023
Received 8 April 2018; Received in revised form 17 December 2018; Accepted 19 December 2018
Available online 05 February 2019
1055-7903/

which may also obscure the underlying phylogenetic pattern (Galtier and Daubin, 2008). Hence, ancient rapid radiations represent some of the most difficult-to-resolve nodes in the Tree of Life.

The Caryophyllales are a large angiosperm clade containing approximately 12,500 species distributed among 749 genera and 40 families (APG, 2016; Walker et al., 2018), whose early branching history is characterized by multiple, nested rapid radiations that have proved resistant to resolution using standard phylogenetic loci (Yang et al., 2015; Walker et al., 2017; Walker et al., 2018). The clade occurs on every continent, with representatives growing in all terrestrial habitats and many aquatic habitats, including a fully aquatic species (*Aldrovanda vesiculosa*, Droseraceae) (Hernández-Ledesma et al., 2015). Many subclades of Caryophyllales are famous for their adaptations to extreme environments, including extreme drought/heat (e.g., Cactaceae), extreme cold (e.g., Montiaceae), and areas of high salt stress (e.g., Amaranthaceae). Specialized ecophysiological adaptations have evolved in concert with these extreme environments, typically multiple times across Caryophyllales. These include various forms of succulence (Hernández-Ledesma et al., 2015), as well as C_4 and CAM (crassulacean acid metabolism) photosynthesis (Stevens, 2001 onwards), with 63.9% of the C_4 plant lineages in eudicots belonging to this order (Sage et al., 2011). In addition, a large and diverse clade of carnivorous plants falls within Caryophyllales, including sundews (*Drosera*), Old World pitcher plants (*Nepenthes*), and the Venus fly-trap (*Dionaea*). Most core Caryophyllales are characterized by the presence of betalain pigments that are unique to Caryophyllales and of increasing economic interest due to their antioxidant properties (Lopez-Nieves et al., 2018). Caryophyllales also include many other economically important species, including crops and culinary herbs (e.g., *Chenopodium quinoa*, *Beta vulgaris*, *Fagopyrum esculentum, Fallopia multiflora*), edible fruits (e.g., *Hylocereus undatus*), and numerous ornamentals (e.g., *Dianthus caryophyllus*). Studying ecophysiological and biochemical evolution within the Caryophyllales, depends on a robust phylogenetic framework for the whole clade.

Our understanding of the composition and relationships among members of Caryophyllales has changed dramatically since the 1960s (e.g., Behnke, 1976; Giannasi, 1992; Yang et al., 2015). Circumscription and deep phylogenetic relationships of this order have been greatly clarified by a series of molecular studies (e.g., Cuénoud et al., 2002; Brockington et al., 2009; Schäferhoff et al., 2009; Crawley and Hilu, 2012a,b; Yang et al., 2015, 2018). The monophyly of Caryophyllales *sensu* APG (2016) has been strongly supported in all previous analyses based on one to many loci. Earlier analyses usually recovered two main sister clades (Brockington et al., 2009; Crawley and Hilu, 2012a,b): (1) the core Caryophyllales, which comprises a strongly supported clade of betalain-producing families and their close anthocyanic relatives (e.g., Caryophyllaceae, Molluginaceae) that corresponds roughly to the former Centrospermae (Caryophyllales s.s.), plus Rhabdodendraceae, Simmondsiaceae, Physenaceae, and Asteropeiaceae; and (2) the non-core Caryophyllales, which includes the carnivorous clade, as well as the FTPP clade (including Plumbaginaceae, Polygonaceae, Tamaricaceae, and Frankeniaceae) (Cuénoud et al., 2002; Brockington et al., 2009; Walker et al., 2018). Within both major clades of Caryophyllales, several other well-supported deep-level clades exist, including the carnivorous, FTPP, globular inclusion, phytolaccoid, Portulacineae, and raphide clades, which have been circumscribed previously (Cuénoud et al., 2002; Brockington et al., 2009; Crawley and Hilu, 2012a; Yang et al., 2015; Table 1). Improved understanding of the relationships among families required the breakup of several traditionally recognized families, particularly Molluginaceae, Phytolaccaceae, and Portulacaceae (Stevens, 2001 onwards; Hernández-Ledesma et al., 2015; Liu et al., 2015; APG, 2016).

Despite our greatly improved understanding of the composition and phylogeny of Caryophyllales, many of the relationships among families of Caryophyllales remain uncertain, especially within the carnivorous, phytolaccoid, and Portulacineae clades (Supplementary Figs. S1–S3). Several deep nodes have been only poorly to moderately supported in both plastome-scale (Brockington et al., 2009; Arakaki et al., 2011; Ruhfel et al., 2014), and transcriptome-based (Yang et al., 2015, 2018; Walker et al., 2018) studies. Although these studies were all character-rich, none included the comprehensive family-level sampling that is essential for resolving rapid radiations.

Here we present a plastid phylogenomic analysis aimed at resolving deep relationships within Caryophyllales and examining the molecular and temporal diversification of the clade. Our dataset includes representatives of all families of Caryophyllales, with a sampling design intended to test deeper relationships throughout the order. We also evaluate the impact of different analytical methods on phylogenetic inference of Caryophyllales, as well as evaluate plastome structural evolution across the clade. We also provide new estimates of diversification times across the clade and evaluate the correlation between rates of substitution and plastome structural changes.

2. Materials and methods

2.1. Taxon sampling

Plastomes of 95 species representing 80 genera within Caryophyllales were newly sequenced for this project (Supplementary Table S1). To these, we added sequences that were already available from earlier studies: (1) sequences from the complete plastomes of 22 additional Caryophyllales (Schmitz-Linneweber et al., 2001; Logacheva et al., 2008; Li et al., 2014; Sloan et al., 2014; Steflova et al., 2014; Cho et al., 2015; Sloan et al., 2014; Kang et al., 2015; Sanderson et al., 2015; Chaney et al., 2016; Dong et al., 2016; Fan et al., 2016; Gurusamy et al., 2016; Ho et al. unpublished; Mardanov et al. unpublished; Park, unpublished; Xu et al. unpublished), (2) sequences of 83 plastid genes from 13 more Caryophyllales species (Arakaki et al., 2011) (Supplementary Table S1), and (3) sequence data for nine plastid genes (*atpB*, *matK*, *ndhF*, *psbB*, *psbT*, *psbN*, *rbcL*, *rpoC2* and *rps4*) and the inverted repeat (IR) region of the 27 Caryophyllales species from Brockington et al. (2009) (Supplementary Table S1). In total, the final matrix included 168 accessions representing all 40 Caryophyllales families, 105 genera, and 141 species (Supplementary Table S1). All Caryophyllales families were represented by multiple individuals in the final matrix, except for 12 families that comprise three or fewer genera (Supplementary Table S1). In larger families, we included representatives of several subfamilies or tribes. For outgroups, we included 18 species of asterids and Berberidopsidales (representing 15 families and 7 orders; Supplementary Table S1) based on the previous phylogenetic results of Moore et al. (2010) and Ruhfel et al. (2014).

2.2. DNA extraction, sample preparation, and sequencing

Total DNA was extracted from fresh leaves following the modified CTAB method of Yang et al (2014). Four different methods were used for sample preparation prior to sequencing, representing contributions from the laboratories of different coauthors at different times (Table S1): (1) unenriched genomic DNA sequencing (Zhang et al., 2017b); (2) plastome amplification through long-range PCR (Yang et al., 2014); (3) targeted enrichment (Stull et al., 2013); (4) transcriptome sequencing (Walker et al., 2018). Paired-end 90-bp or 100-bp sequencing was conducted using an Illumina GAIIx or HiSeq 2000 at the University of Florida ICBR sequencing facility, or using a HiSeq 2000 at BGI-Shenzhen following the manufacturer's protocols. For each library, > 200 Mb (for the long-range PCR method) or > 2 Gb (for genomic DNA sequencing) of DNA sequence data was obtained. The plastomes of *Aldrovanda vesiculosa* (accession Moore 1652), *Antigonon leptopus*, and *Cerastium arvense* were derived from transcriptome data; methods for sequencing are described in Yang et al. (2015).

Table 1
Family composition of deeper-level clades in Caryophyllales.

Clade	Families included	Reference
The carnivorous clade	Ancistrocladaceae, Dioncophyllaceae, Drosophyllaceae, Droseraceae, Nepenthaceae	Crawley and Hilu (2012a,b)
The FTPP clade	Frankeniaceae, Plumbaginaceae, Polygonaceae, Tamaricaceae	Crawley and Hilu (2012a,b)
The globular inclusion clade	Agdestidaceae, Aizoaceae, Anacampserotaceae, Barbeuiaceae, Basellaceae, Cactaceae, Corbichoniaceae, Didiereaceae, Gisekiaceae, Halophytaceae, Kewaceae, Lophiocarpaceae, Molluginaceae, Montiaceae, Nyctaginaceae, Petiveriaceae, Phytolaccaceae, Portulacaceae, Sarcobataceae, Talinaceae	Cuénoud et al. (2002), Brockington et al. (2009)
The phytolaccoid clade	Agdestidaceae, Nyctaginaceae, Petiveriaceae, Phytolaccaceae, Sarcobataceae	Yang et al. (2015)
The Portulacineae clade (or the succulents clade)	Anacampserotaceae, Basellaceae, Cactaceae, Didiereaceae, Halophytaceae, Montiaceae, Portulacaceae, Talinaceae	Crawley and Hilu (2012a,b)
The raphide clade	Aizoaceae, Agdestidaceae, Gisekiaceae, Nyctaginaceae, Petiveriaceae, Phytolaccaceae, Sarcobataceae	Cuénoud et al. (2002), Brockington et al. (2009), Crawley and Hilu (2012a,b)

2.3. Plastome assembly and annotation

Raw sequence reads were filtered using NGS QC analyses within CLC Genomics Workbench (http://www.clcbio.com/), using default parameters. Filtered, high-quality reads were assembled *de novo* within CLC Genomics Workbench using default parameters except that word size was set to 60. For each individual, the correct position of assembled contigs was determined by BLASTing contigs (Local BLAST: ncbi-blast-2.2.30, using default parameters) against the complete plastome sequences of *Rheum palmatum* (NCBI accession NC_027728; Fan et al., 2016), *Silene chalcedonica* (NC_023359; Sloan et al. 2014), and *Spinacia oleracea* (NC_002202; Schmitz-Linneweber et al., 2001). For several individuals, one to several gaps remained after the initial assembly. To close these gaps, SOAPdenovo 1.0.4 (Li et al., 2010) was adopted, with the kmer value set to 81. This approach allowed most plastomes to be completely assembled. We also conducted assembly using a "reference-architecture-free" strategy as implemented in the pipeline GetOrganelle (https://github.com/Kinggerm/GetOrganelle). This strategy enriched the chloroplast reads from raw reads using the "baiting and iterative mapping approach" (Hahn et al., 2013), then assembled the reads into contigs using SPAdes 3.9.0 (Bankevich et al., 2012). The contig graph was then visualized, edited and exported as the complete or nearly complete circular plastome using Bandage 0.7.1 (Wick et al., 2015).

Assembled plastome sequences were annotated using the online program DOGMA (Wyman et al., 2004). Positions of start and stop codons and the exon/intron boundaries were checked manually using Geneious 7.1.4 (Kearse et al., 2012) against the annotated plastome of *Spinacia oleracea*.

2.4. Phylogenetic dataset construction and analyses

Coding regions of all 83 plastid-coding genes (all 79 protein-coding genes and four rRNA genes; Supplementary Table S2; Moore et al., 2010) were extracted from the assembled plastomes and aligned using MAFFT 7.221 (Katoh and Standley, 2013) with default settings, with subsequent manual modifications in MEGA 7.0 (Kumar et al., 2016). In the few cases where genes were absent, they were treated as missing data.

To investigate the phylogenetic effects of character inclusion/exclusion and to minimize systematic error due to poor alignment, seven phylogenetic data matrices were constructed (Supplementary Table S3), as follows: (i) basic combined-complete matrix, including all genes for the 163 accessions that possessed complete or nearly complete plastome data; (ii) combined-complete matrix, constructed based on the basic combined-complete matrix, with the removal of poorly aligned regions in *accD*, *ycf1*, and *ycf2* using Gblocks 0.91b (Castresana, 2000) and exclusion of gaps with > 50% missing data; (iii) combined-incomplete matrix, constructed based on the combined-complete matrix, with the inclusion of 27 additional individuals for which partial plastome data were available (Supplementary Table S4); (iv) all-gapped matrix, constructed based on the basic combined-complete matrix, excluding ambiguously aligned sites and including all gap positions; (v) half-gapped matrix, constructed based on the basic combined-complete matrix, excluding gaps with > 50% missing data; (vi) no-gapped matrix, constructed based on the basic combined-complete matrix, excluding all gaps; and (vii) unsaturated matrix, constructed based on the basic combined-complete matrix, excluding *accD*, *ycf1*, and *ycf2* because DAMBE 5 (Xia, 2013) indicated that these loci were saturated.

For each dataset, Partitionfinder 1.0.1 (Lanfear et al., 2012) was used to evaluate the optimal partitioning strategy, employing the Corrected Akaike Information Criterion (AICc). We reconstructed trees using maximum likelihood (ML) on all matrices using RAxML-HPC2 (8.1.24) (Stamatakis, 2006a) on the CIPRES cluster (Miller et al., 2010), employing the GTR + Γ model (Supplementary Table S3) and the default number of rate categories (C = 25). We conducted a rapid bootstrap analysis using the GTR + Γ model (Stamatakis, 2006b) with 1000 pseudoreplicates. All tree files obtained were visualized and edited using FigTree 1.4.0 (Rambaut, 2012).

2.5. Molecular dating

Recent empirical studies have suggested that missing data may have little impacts on the accuracy of molecular dating analyses (e.g., Zheng and Wiens, 2015; Pozzi, 2016). Considering that the phylogenetic positions of the 27 individuals with greater proportions of missing data were well-resolved with strong support (see Results), we conducted the dating analysis of Caryophyllales on the combined-incomplete matrix, which includes the best taxon sampling. However, to minimize the computational burden especially for the Bayesian analysis, we reduced the combined-incomplete matrix by keeping a single individual per species. Penalized likelihood (PL) dating analysis, which performs well for large datasets (e.g., Xiang et al., 2016; Zhang et al., 2017b), was conducted using treePL (Smith and O'Meara, 2012). Detailed procedures of dating analyses follow Zhang et al. (2017b). We used a maximum age calibration of 125 Million years ago (Ma) for the root, based on age estimates for the split between Berberidopsidales and the asterid + Caryophyllales clade (Wikström et al., 2001; Bell et al., 2010; Zeng et al., 2017), and for crown group superasterids (Magallón et al., 2015; Wikström et al., 2015). Moreover, 125 Ma represents the fossil-based age of eudicots (Doyle and Hotton, 1991), within which Caryophyllales are nested. Ten fossils employed in previous studies were used as minimum-age calibrations, including five from Caryophyllales and five from asterids (Supplementary Table S5). A detailed discussion of these fossil calibrations is provided in Supplementary Methods S1.

Age estimation was also conducted in BEAST 2.4.7 (Bouckaert et al. 2014), using a lognormal relaxed molecular clock and the birth-death model of speciation. To minimize the computational burden and reduce missing data, a combined matrix including only seven plastid genes (*atpB*, *matK*, *ndhF*, *psbB*, *rbcL*, *rpoC2*, *rps4*; total length 14,202 bp) of Brockington et al. (2009) was applied in the analysis. A constraint ML topology derived from the reduced combined-incomplete matrix was used. Lognormal distribution priors were used for all fossil constraints,

with offsets set to the minimum fossil ages, a mean of 1.0 Ma and a standard deviation of 0.5 Ma. We used a normal distribution around the root, with mean of 125.0 Ma and SD of 1.0 Ma. Nucleotide model selection was conducted for the combined matrix using the AICc in jModelTest2 (Posada, 2008), yielding the GTR + I + Γ model. The MCMC analysis ran for 709,465,000 generations, sampling every 1000 generations, until a sufficient effective sample size (ESS) value (> 200) was obtained for all relevant parameters. The first 10% of the trees was discarded as burn-in, as determined using Tracer 1.6 (Rambaut et al., 2014).

2.6. Correlation between substitution rate and plastome structural changes

Phylogenetic generalized least-squares (PGLS) regression (Martins and Hansen, 1997), with the variance-covariance matrix approximating Brownian motion, was used to test for the correlation between substitution and plastome structural change rates. The variance-covariance matrix for the PL-dated tree was calculated using "corBrownian" as implemented in the R package "ape" (Paradis et al., 2004). We fit the regression by maximizing the restricted log-likelihood using "gls" as implemented in the R package "nlme" (Pinheiro et al., 2016). The substitution rate for each species from the combined-complete matrix was characterized as the phylogenetic distance from root to tip along the ML tree. We characterized the structural changes for each species as the number of discrete events (where an event is a protein-coding gene loss, intron loss, or pseudogenization) compared to the outgroup *Cornus capitata*, which possesses the ancestral angiosperm plastome structure (Fu et al., 2017). We mapped relevant structural changes onto the phylogenetic tree using parsimony (Fig. 3). Inverted repeat (IR) region expansion/contraction is commonly observed in angiosperms and in many cases has resulted in multiple genes moving in or out of the IR (Zhu et al., 2016). In Caryophyllales, however, IR expansion/contraction has only been observed in a few clades, and most shifts are small, involving only 1–3 genes (see Results). Hence, changes in IR boundaries were not included in the present analysis but are described and mapped on the tree in Fig. 3.

3. Results

3.1. Plastome and dataset characteristics

Plastomes of 103 species (including outgroups) were newly sequenced for this study; voucher information and GenBank accession numbers are provided in Supplementary Table S1. Complete plastomes were obtained for most species, although one to a few small gaps (< 1000 bp) were observed in 44 plastomes and the highly rearranged plastomes of Droseraceae, *Hypertelis,* and *Pharnaceum* were not completed (Supplementary Table S1). These gaps were mostly located in non-coding regions. Plastome sizes for completed or nearly completed plastomes ranged from 129,601 (*Physena sessiliflora*) to ~170,974 bp (*Afrobrunnichia erecta*). The length and proportion of missing data for all genes included in this matrix are presented in Supplementary Table S2. The proportion of missing data for each of the 27 accessions included only in the combined-incomplete matrix is presented in Supplementary Table S4. The combined-complete matrix had an aligned length of 73,135 bp with only 4.1% missing data, and the combined-incomplete matrix had an aligned length of 73,207 bp with 14.6% missing data. Full statistics for all matrices are presented in Supplementary Table S3. The optimal partitioning scheme for each data matrix is provided in Supplementary Table S6.

3.2. Phylogenetic results

ML analyses of the combined-complete (shown in Fig. 1a and b) and combined-incomplete (Supplementary Fig. S4) matrices yielded identical topologies. Almost all relationships among the major clades and families of Caryophyllales were well resolved and strongly supported in analyses of these two major matrices. However, the first two nodes at the base of the Centrospermae clade had low bootstrap support (BS) (60–65%), and the sister clade of Barbeuiaceae also received low support (BS < 50%). Nearly identical topologies of Caryophyllales were also recovered from ML analyses of the basic combined-complete matrix (Supplementary Fig. S5) and the other modified matrices (Supplementary Figs. S6–S9), except that analyses of the no-gapped matrix (Supplementary Fig. S8) yielded alternative infrafamilial relationships of Amaranthaceae and Cactaceae and of the placements of Barbeuiaceae and Kewaceae within the globular inclusion clade. Highly congruent levels of support were also obtained in ML analyses of the partitioned (Supplementary Figs. S10a–g) and unpartitioned (Fig. 1a and b; Supplementary Figs. S4–S9) matrices, except for two nodes (Supplementary Table S7). Higher support was observed for some nodes with the removal of sites with more gapped characters (Supplementary Table S7).

Because highly congruent trees of Caryophyllales were obtained across all analyses, we focus on describing phylogenetic relationships using the combined-complete matrix tree (Fig. 1a and b). The monophyly of Caryophyllales, core and non-core Caryophyllales, and all major clades within the order (e.g., Centrospermae, the carnivorous clade, the FTPP clade, the globular inclusion clade, the phytolaccoid clade, the Portulacineae clade, and the raphide clade) was strongly supported (all with BS = 100%, except the globular inclusion clade with BS = 92%). Within core Caryophyllales, Asteropeiaceae + Physenaceae, Simmondsiaceae, and Rhabdodendraceae were successively sister to the Centrospermae clade. Within the latter, *Microtea* and *Stegnosperma* were successively sister to the remaining members of the clade, with weak BS. The remainder of the Centrospermae clade was resolved into two sister clades: (1) a clade including Achatocarpaceae, Amaranthaceae, Caryophyllaceae, and Macarthuriaceae (BS = 91%), and (2) a clade including all remaining taxa (BS = 100%). Within the non-core Caryophyllales (Fig. 1b), Frankeniaceae + Tamaricaceae appeared sister to Plumbaginaceae + Polygonaceae, and this FTPP clade was sister to the carnivorous clade, all with maximum support.

3.3. Molecular dating

Results from PL analyses revealed that age estimates for most nodes within Caryophyllales were affected strongly by the presence or absence of the long branch leading to the clade of *Hypertelis* + *Pharnaceum* in Molluginaceae (Supplementary Table S8), which may result from the fact that the PL method uses a global smoothing value for all rate shifts. Excluding this clade resulted in the estimated ages for most nodes in the PL analysis being largely consistent with those in the BEAST analysis, especially for major clades within Caryophyllales (Supplementary Table S8). The ages estimated (with the *Hypertelis* + *Pharnaceum* clade excluded) for the stem and crown of Caryophyllales were ca. 122.4 Ma (age interval: 123.1–122.0 Ma), 114.4 Ma (115.8–113.2 Ma) in the PL analysis and ca. 124.4 Ma [95% highest posterior densities (HPD): 127.2–121.3 Ma], 116.5 Ma (121.4–111.1 Ma) in the BEAST analysis, respectively. The 95% HPD obtained from the BEAST analysis were much broader than the age intervals obtained from the PL analysis (Supplementary Table S8). Because of the problems resulting from the long branch to the *Hypertelis* + *Pharnaceum* clade, we mainly discuss results from the PL analysis with this clade excluded (Fig. 2; Supplementary Table S9).

The earliest divergences within the order were dated to the mid-Cretaceous (especially from the Albian to the Turonian period, 113.0–89.8 Ma), with most modern families estimated to have been present by the end of the Cretaceous (ca. 66.0 Ma), although the earliest divergences in several families of the phytolaccoid and Portulacineae clades were dated to the Paleocene and Eocene. The early divergence of the Centrospermae clade was estimated to have occurred extremely

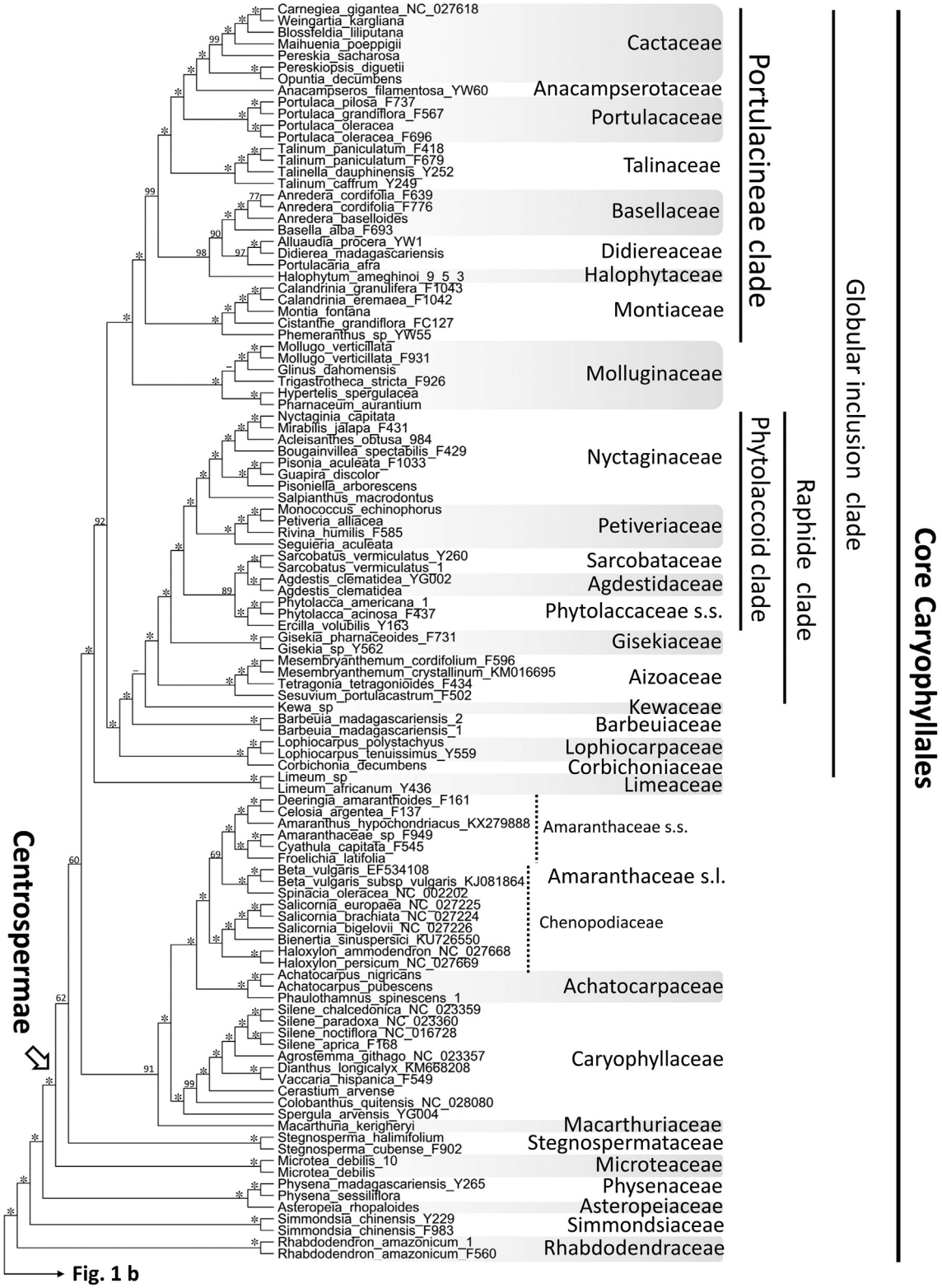

Fig. 1. Maximum likelihood tree of Caryophyllales inferred from the combined-complete matrix. Asterisks and numbers associated with nodes are ML bootstrap support. Asterisks indicate that the support is 100% BS and dashes denote that the support is < 50% BS. The crown node of Centrospermae and Caryophyllales are shown by the hollow and solid arrowhead, respectively. (a) Phylogeny of the core Caryophyllales, (b) phylogeny of the non-core Caryophyllales and outgroups.

rapidly, over a time span as short as 3.5 million years (95.5–92.0 Ma) in the mid-Cretaceous near the Cenomanian-Turonian boundary (Fig. 2; ca. 93.9 Ma). Absolute ages estimated for Caryophyllales are shown in Supplementary Table S9.

3.4. Plastome structure and molecular evolution

No large-scale rearrangements were detected across Caryophyllales, with the exception of eight highly rearranged plastomes: *Aldrovanda vesiculosa*, *Dionaea muscipula*, *Drosera indica* and *D. rotundifolia*

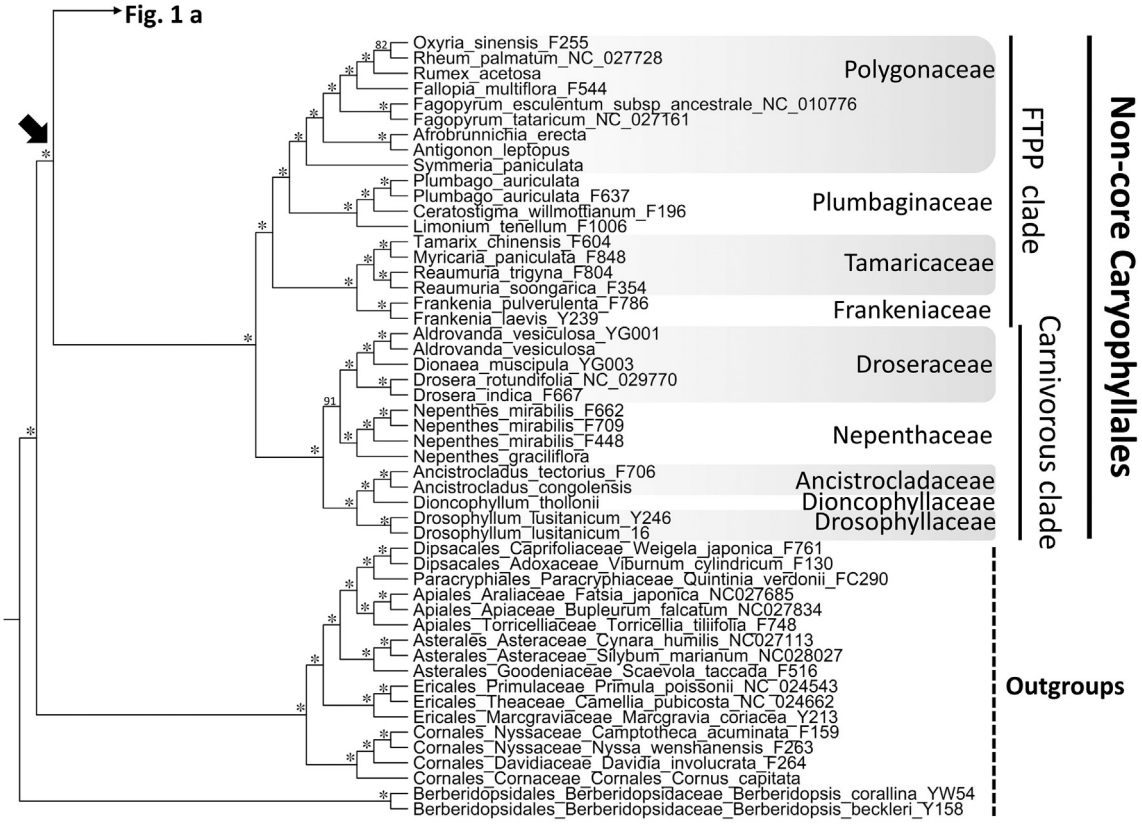

Fig. 1. (*continued*)

(Droseraceae), *Physena madagascariensis* and *P. sessiliflora* (Physenaceae), and *Hypertelis spergulacea* and *Pharnaceum aurantium* (Molluginaceae). A small inversion (represented by pentagon in Fig. 3), involving only *accD*, was detected in *Sarcobatus* (Sarcobataceae). The evolutionary rate was elevated significantly in the *Hypertelis* + *Pharnaceum* clade (Fig. 3), with an average accumulation of molecular changes (calculated in the same way as Smith and Donoghue 2008) approximately seven times higher than that of its sister clade. IR expansion (represented by red rectangle in Fig. 3) was detected in a number of clades (Fig. 3). The IR has expanded to include all of the *ycf1* gene in Polygonaceae and most Plumbaginaceae, except for *Limonium*. The IR has expanded to include all of *ycf1* and *rps15* and also a portion of *ndhH* in *Silene noctiflora* (Caryophyllaceae), and has expanded to include all of *rps19*, *rpl22*, *rps3*, *rpl16*, and *rpl14* in *Afrobrunnichia* (Polygonaceae), yielding a total IR length of 34,631 bp. Other IR expansions were detected in *Anacampseros* (Anacampserotaceae) and *Rhabdodendron* (Rhabdodendraceae; expanded to include 2340 bp of *ycf1*). Small IR contractions (represented by purple rectangle in Fig. 3) were detected in *Drosophyllum* (Drosophyllaceae; contracted to include only 186 bp of *ycf1*) and *Pisoniella* (Nyctaginaceae; contracted to exclude *rps19* and all but the first 47 bp of *rpl2*), and the IR is absent (represented by blue rectangle in Fig. 3) in *Carnegiea gigantea* (Cactaceae; Sanderson et al., 2015).

Various gene and intron losses (represented by circle and square in Fig. 3, respectively) were observed throughout Caryophyllales (Fig. 3). Pseudogenization of the *rpl23* gene was rampant, with a minimum of 11 independent losses. Other individual genes were also found to have been pseudogenized or deleted in various Caryophyllales, as for example *infA* in Caryophyllaceae and *ccsA* in *Simmondsia*. However, the most significant gene alterations were associated with Droseraceae, Physenaceae, and Molluginaceae (*Hypertelis* and *Pharnaceum*) (Fig. 3). The *ndh* complex of 11 plastid genes was also found to be lost/pseudogenized in six different clades: Cactaceae subfam. Cactoideae (represented by *Blossfeldia*, *Carnegiea* and *Weingartia* in the present study), *Anacampseros* (Anacampserotaceae), *Pharnaceum* (Molluginaceae), *Simmondsia* (Simmondsiaceae), Droseraceae, and *Drosophyllum* (Drosophyllaceae). Intron losses displayed less homoplasy, although the *rpl2* intron was lost twice, once at the base of the clade of Centrospermae and Physenaceae + Asteropeiaceae, and once in Droseraceae (Fig. 3). Intron loss was most common in the rearranged plastomes of *Physena*, Droseraceae, *Hypertelis,* and *Pharnaceum* (Fig. 3).

A significant positive relationship between substitution rates and plastome structural change rates was detected by PGLS regression (coefficient: 2.813E-3, p-value $< 1.0E-3$); full results are provided in Supplementary Table S10.

4. Discussion

4.1. Caryophyllales phylogeny

Our results provide yet another example of using phylogenomics to disentangle relationships of many angiosperm clades that have experienced rapid radiations (e.g., Moore et al., 2010; Xi et al., 2012; Ma et al., 2014). In general, our data improved resolution and support throughout Caryophyllales compared with previous studies, with almost all nodes supported with > 90% BS. These improvements include several key nodes that were poorly resolved in past studies, most notably the relationships within the phytolaccoid clade and the Portulacineae clade (positions of Agdestidaceae and Halophytaceae; Fig. 1a). However, while the resolution may be high, some relationships may still be uncertain. It is important to emphasize that at deeper phylogenetic levels the plastome represents a single, non-recombining locus, albeit one with numerous informative characters. Although the vastly increased number of characters in plastid phylogenomic studies reduces some of the stochastic error inherent in phylogenomic studies, it may exacerbate other sources of systematic error (Jeffroy et al., 2006; Philippe et al., 2011) and cannot account for complex reticulate evolution. Hence we must remain cautious in any conclusions based solely

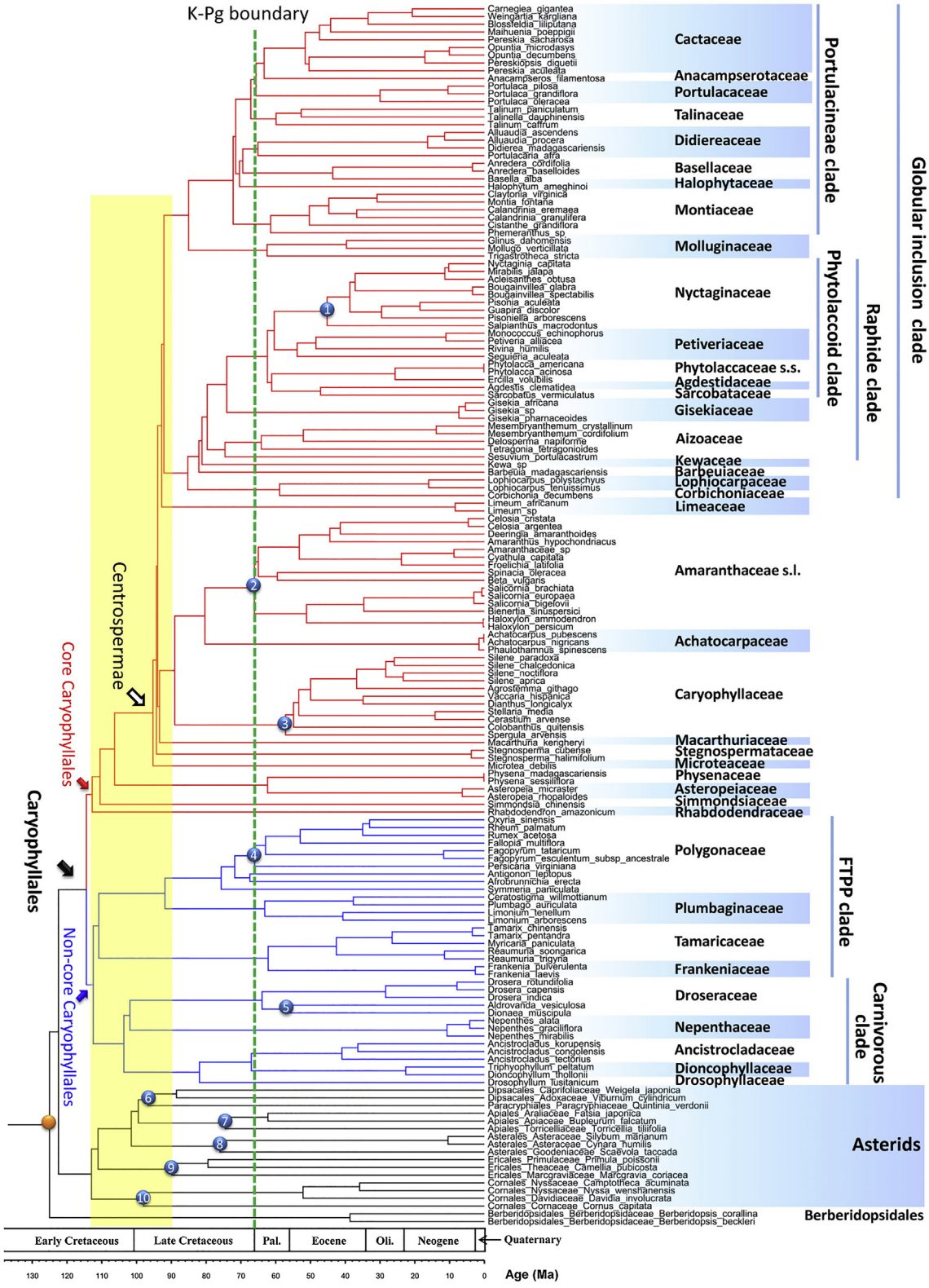

Fig. 2. Chronogram using the modified combined-incomplete matrix (including 159 species) and ten fossil calibrations, from the penalized likelihood analysis. Positions of the fossil calibrations are depicted with numbered circles. The orange circle indicates the root of tree. The yellow region indicates the Albian-Turonian period (113.0–89.8 Ma) in the Mid-Cretaceous. Oli.: Oligocene; Pal.: Paleocene.

on plastome data.

The only other available phylogenomic study of Caryophyllales with comparable taxon sampling to that employed here is the transcriptome-based analyses of Walker et al. (2018). Their nuclear-based topology is largely congruent with the plastid tree presented here, but there are several well-supported differences in regions of short branches that may reflect true incongruence of nuclear vs. plastid data. For example, early divergences in the Centrospermae differed between the current study

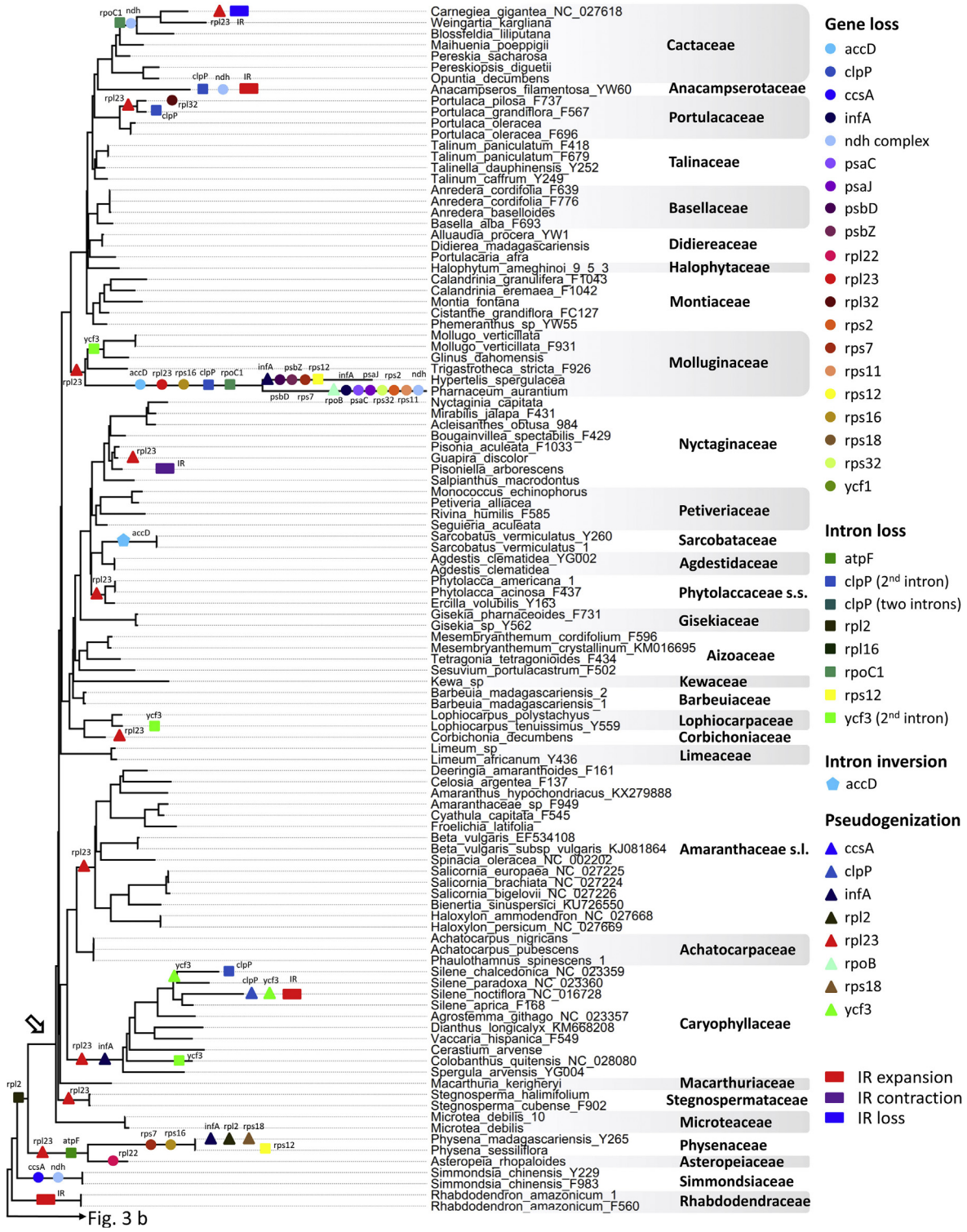

Fig. 3. Maximum likelihood tree of Caryophyllales inferred from the combined-complete matrix, with plastid structural changes (protein-coding gene loss, intron loss, intron inversion, pseudogene formation, IR contraction, expansion and loss) indicated. The crown node of Centrospermae is shown by the hollow arrowhead. (a) Phylogeny of the core Caryophyllales, (b) phylogeny of the non-core Caryophyllales. The crown node of Caryophyllales is shown by the solid arrowhead.

and that of Walker et al. (2018): (1) *Macarthuria* (Macarthuriaceae) was recovered as sister to *Stegnosperma* in Walker et al. (2018) rather than as sister to the clade of Achatocarpaceae, Amaranthaceae, and Caryophyllaceae; (2) Basellaceae were recovered as sister to all Portulacineae families except Montiaceae, rather than as sister to Didiereaceae; (3) Anacampserotaceae were sister to the Portulacaceae + Cactaceae rather than to Cactaceae alone; and (4) Chenopodiaceae in the traditional sense were monophyletic. However, as mentioned in Walker et al. (2018), it is evident that there is quite a bit of conflict within the nuclear data, especially for those nodes conflicted with the plastid topology, such as relationships within the non-core Caryophyllales and also within Chenopodiaceae. There are many possible explanations for the conflict among gene trees, including ILS and ancient hybridization (including paleoallopolyploidy; Yang et al., 2018). At this point, given

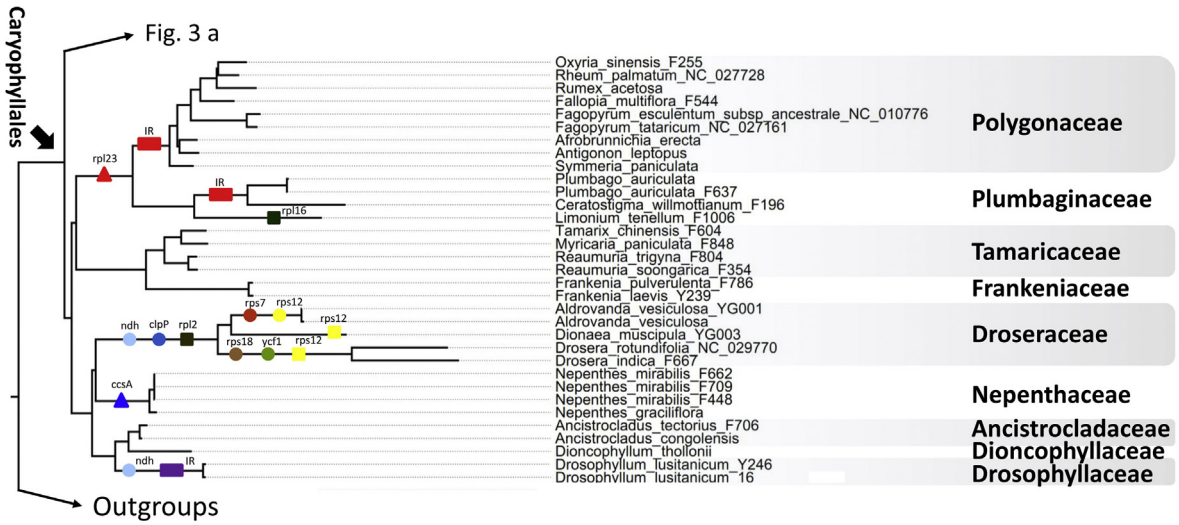

Fig. 3. (*continued*)

current data, it is difficult to sort between ILS and/or hybridization. To more confidently sort between these possibilities will require complete nuclear genomes from many lineages of Caryophyllales, which will allow for analyses of synteny.

Other sources of systematic error in phylogenomic studies may result from insufficient or biased taxon sampling, inaccurate alignment, improper phylogenetic reconstruction methods, and/or incorrect models of sequence evolution (Jeffroy et al., 2006; Rodriguez-Ezpeleta et al., 2007; Philippe et al., 2011). The influence of broad taxon sampling and missing data on the accuracy of phylogenetic inference has been highlighted in previous studies (e.g., Wiens, 2005; Jeffroy et al., 2006; Pick et al., 2010); we made an effort to develop a broad, representative sampling across Caryophyllales to minimize such problems. The nuclear phylotranscriptomic study of Walker et al. (2018) also has good taxonomic coverage, but is missing several small, rare families due to lack of fresh tissue for transcriptome analysis; their inclusion might have altered the best topology recovered for the Caryophyllales backbone. In our own study, we included 27 accessions with large amounts of missing data (Supplementary Table S4), because previous studies have demonstrated that including such taxa may improve support and alter topology (Wiens, 2003; Jiang et al., 2014). Our combined-incomplete matrix resolved the position of these 27 taxa with high support (Supplementary Table S4; Fig. S4), and it appears that including these taxa helped improve support for several key nodes, such as the sister relationship of Droseraceae to Nepenthaceae, which increased from BS = 91% in the ML analysis of the combined-complete matrix (Fig. 1b) to 98% in the combined-incomplete matrix (Supplementary Fig. S4).

Phylogenetic reconstruction is also dependent on accurate sequence alignment (Morrison and Ellis, 1997; Ogden and Rosenberg, 2006). The removal of problematic sequence regions has been suggested as an effective way to reduce alignment artifacts and improve the robustness of phylogenomic reconstruction (Goremykin et al., 2010; Zhong et al. 2011; Parks et al., 2012; Som, 2015), although removal of sites based on level of variability alone has been shown to be misguided, at least when taxon sampling is not limited (Drew et al., 2014). However, results from our study indicate that the topology is largely immune to removal of highly variable sites (including analyses of the all-gapped, half-gapped, no-gapped and unsaturated matrices), although some portions of the tree with short branches were affected (Supplementary Figs. S5–S9, Table S7). These included the sister relationships of Basellaceae and Didiereaceae, Droseraceae and Nepenthaceae, and Amaranthaceae *s.s.* and the *Beta* + *Spinacia* clade, as well as some infrafamilial relationships in Cactaceae.

4.2. Evolutionary and taxonomic implications

The broad congruence of our results with those from previous studies (e.g., Schäferhoff et al., 2009; Brockington et al., 2009), and especially with those of Walker et al. (2018), suggests that the Caryophyllales phylogeny has mostly come into focus. This higher confidence in Caryophyllales relationships will serve as an excellent foundation for evaluating evolutionary questions and revising taxonomy throughout the clade. It also allows for greater confidence in assessing putative morphological synapomorphies. For example, the sister relationship of Droseraceae and Nepenthaceae is supported by pollen morphology; both families have inaperturate pollen grains united in tetrads (Halbritter et al., 2012; Yu et al. 2018). Within the phytolaccoid clade, both our plastome analyses and Walker et al. (2018) recover a monocarpellate Nyctaginaceae + Petiveriaceae clade sister to a multicarpellate (Phytolaccaceae *s.s.*, (Agdestidaceae + Sarcobataceae)) clade with strong support (Fig. 1a). Fruits of the two monogeneric families Agdestidaceae and Sarcobataceae are characterized by winged achenes, in contrast to the berries present in their sister-clade, Phytolaccaceae (Nienaber and Thieret, 2003; Welsh et al., 2003). These morphological differences, as well as the extreme differences in floral characters, habit, and habitat between *Agdestis* and *Sarcobatus*, support the breakup of Phytolaccaceae *s.l.* into Agdestidaceae, Petiveriaceae, and Phytolaccaceae *s.s.* (Hernández-Ledesma et al., 2015; APG, 2016; Walker et al., 2018).

Amaranthaceae and Chenopodiaceae have been traditionally accepted as separate families based on their different bracteole, filament, and sepal morphologies (Kung and Tsien, 1979; Kubitzki et al., 1993). However, molecular phylogenetic studies have typically found that Chenopodiaceae are not monophyletic. Studies based on a few genes have tended to recover a paraphyletic Chenopodiaceae that includes Amaranthaceae *s.s.* (Kadereit et al., 2003; Müller and Borsch, 2005; Hernández-Ledesma et al., 2015). Based on these studies, many authors (Cuénoud et al., 2002; Crawley and Hilu, 2012a), including APG (2016), have advocated a broad Amaranthaceae including Chenopodiaceae. The recent phylotranscriptomic study of Walker et al. (2018) found some support for a monophyletic Chenopodiaceae (including subfam. Polycnemoideae) that was sister to Amaranthaceae *s.s.*, but also found alternative trees with a paraphyletic Chenopodiaceae. In our own plastome tree, Chenopodiaceae were not monophyletic (Fig. 1a; Supplementary Figs. S4–S7, S9). However, no genomic-scale analyses of the Amaranthaceae *s.l.* (neither here nor in Walker et al., 2018) have included full representation of early-diverging lineages. For example, we lack Polycnemoideae, *Bosea,* and *Charpentiera*, and it is possible that

inclusion of these taxa would alter the plastome and/or nuclear topologies.

4.3. Origin and diversification of Caryophyllales

Our study provides the first comprehensive family-level molecular dating analysis of Caryophyllales (Fig. 2; Supplementary Table S9). The stem age (ca. 122.4 Ma) and crown age (ca. 114.4 Ma) of Caryophyllales estimated here are largely congruent with those reported recently (e.g. ca. 121.2 Ma and 107 Ma, respectively, from the PL analysis of Magallón et al., 2015). Furthermore, ages estimated for the stem nodes of some families within the order are also largely consistent with those reported by Magallón et al. (2015), although our estimates for the crown group age of Cactaceae (ca. 53 Ma) are somewhat older.

Our results imply rapid diversification within the Caryophyllales near the Aptian –Albian boundary during the mid-Cretaceous, giving rise to much of the family-level diversity in the order, with continued diversification before and after the Cretaceous-Paleogene (K-Pg) boundary (ca. 66.0 Ma). Similar radiations have been documented among the asterid (Wikström et al., 2015) and rosid (Wang et al., 2009) clades, and similar K-Pg boundary diversifications have been suggested in Menispermaceae (Wang et al., 2012), Orchidaceae (Givnish et al., 2015), and Meliaceae (Koenen et al., 2015). Evidence from the fossil record also indicates a significant increase in generic-level richness in angiosperms during the Albian–Turonian period (113.0–89.8 Ma) in the mid-Cretaceous, and once again near the K-Pg boundary (Lidgard and Crane, 1988).

Some have suggested that a decrease in atmospheric CO_2 concentration may have triggered the ecological radiation and phylogenetic diversification of angiosperms in the mid-Cretaceous (McElwain et al., 2005). Brodribb and Feild (2010) have suggested that changes in angiosperm vein architecture that occurred during the mid-Cretaceous may provide a mechanistic link between climate change and angiosperm radiation during this period. Furthermore, Feild et al. (2011) noted a striking escalation of angiosperm vein densities in mid-Cretaceous fossils and again in fossils near the K-Pg boundary. Angiosperm pollen characters also evolved rapidly during the mid-Cretaceous (e.g., Luo et al., 2015; Zhang et al., 2017a). The mass extinction that occurred near the K-Pg boundary (Schulte et al., 2010) may have also produced vacant niches for angiosperm diversification (Wang et al., 2012). Our molecular dating results for Caryophyllales are consistent with these fossil studies in suggesting that the mid- to late Cretaceous and early Paleogene were important periods for angiosperm diversification.

4.4. Plastome structural evolution

Our broad taxon sampling provides unprecedented insights into plastome structural evolution throughout Caryophyllales (Fig. 3), including several putative structural synapomorphies (Supplementary Table S11). For example, previous studies have suggested that the *rpl2* intron might be absent throughout the Centrospermae (Palmer et al., 1988), which our study strongly supports. Furthermore, we detected *rpl2* intron loss in the Asteropeiaceae + Physenaceae clade, which is sister to the Centrospermae, providing an important structural synapomorphy to support the sequence-based phylogenetic results. Previous studies have also detected the pseudogenization of *rpl23* in Caryophyllales, including in members of Amaranthaceae (Schmitz-Linneweber et al., 2001), Caryophyllaceae (Raman and Park, 2015), and Polygonaceae (Logacheva et al., 2008), suggesting the possibility that *rpl23* may be a pseudogene throughout Caryophyllales. Somewhat unexpectedly, we found that many taxa outside of these three economically important families have full-length, apparently intact *rpl23* genes (Fig. 3). Our results instead imply that *rpl23* has become a pseudogene repeatedly (at least 11 times) throughout the order. Functional gene transfer to the nucleus has been found in most cases of plastid gene alteration in autotrophic plants (the *ndh* complex excepted; Ruhlman and Jansen, 2014; Daniell et al., 2016). The propensity for pseudogene formation across Caryophyllales may suggest that a functional copy of *rpl23* was transferred to the nucleus early on in Caryophyllales evolution, perhaps even in the common ancestor of all extant Caryophyllales, which would allow for the subsequent loss of the plastid copy throughout the clade in a pattern similar to that observed here.

We also detected at least six independent losses of the *ndh* complex (Fig. 3). The 11 plastid *ndh* genes encode subunits of the plastid NDH (NADH dehydrogenase-like) complex, which plays an important role in mediating cyclic electron flow around photosystem I and facilitates chlororespiration (Martín and Sabater, 2010). However, the loss of the *ndh* genes has been widely reported throughout plants and is often associated with unusual trophic status. Examples include heterotrophic and autotrophic orchids (Kim et al., 2015), mycoheterotrophic Ericaceae (Braukmann and Stefanović, 2012), and many parasitic plants such as *Aneura* (Aneuraceae; Wickett et al., 2008), *Cuscuta* (Convolvulaceae; Funk et al., 2007), and Orobanchaceae (Wicke et al., 2013). The gene family has also been lost in carnivorous Lentibulariaceae (Wicke et al., 2014), conifers (Braukmann et al., 2009), *Erodium* (Geraniaceae; Blazier et al., 2011), *Najas* (Hydrocharitaceae; Peredo et al., 2013), and *Carnegiea gigantea* (Cactaceae; Sanderson et al., 2015) — the last representing the first report of *ndh* family loss in Caryophyllales. We detected *ndh* gene family loss in two additional cacti (*Blossfeldia* and *Weingartia*; Fig. 3), suggesting that the gene family may be absent throughout subfamily Cactoideae.

Earlier studies have hypothesized that the absent/truncated plastid *ndh* genes may have been transferred to the nuclear genome, retaining their function there (Sanderson et al., 2015). Alternatively, the *ndh* function may have been assumed by the AA-sensitive CEF pathway (Hertle et al., 2013). Other authors have suggested that the NDH complex may be dispensable under mild, non-stressing environmental conditions (Martín and Sabater, 2010; Barrett et al., 2014). Two of the six *ndh* family losses that we found in Caryophyllales correspond to the carnivorous Droseraceae and Drosophyllaceae, but not the carnivorous Nepenthaceae (Fig. 3). Previous studies suggested that carnivorous plants may have much lower maximal photosynthetic rates than those of non-carnivorous species (Méndez and Karlsson, 1999; Ellison and Gotelli, 2001) because to some extent they may make use of organic carbon obtained from their prey (Wicke et al., 2014). In a study of Droseraceae and Lentibulariaceae, Méndez and Karlsson (1999) suggested that the differences of photosynthetic rate between carnivorous and non-carnivorous plants may be connected with the anatomy of their leaves. Nepenthaceae species have much larger, often greener leaves than many other carnivorous plants, and are similar in size to many other non-carnivorous plants. This may indicate that Nepenthaceae have a higher photosynthetic rate compared with other carnivorous plants, which may explain the retention of the *ndh* genes.

Both structural rearrangement and gene/intron loss (including pseudogenization) were strongly correlated with increases in the rate of molecular evolution (Fig. 3; Supplementary Table S10). The greatest increases in evolutionary rate occurred in Droseraceae (especially *Drosera*) and the clade of *Hypertelis* + *Pharnaceum* (Molluginaceae), associated with highly rearranged plastomes. The case of Molluginaceae is most instructive in demonstrating the correlation between structural rearrangements and evolutionary rates. All members of Molluginaceae have similar habits (typically herbs) and grow in similar habitats (typically open, sunny, semiarid regions) (Thulin et al., 2016). However, the clade of Molluginaceae that includes *Hypertelis* and *Pharnaceum,* as well as *Adenogramma, Coelanthum, Polpoda, Psammotropha*, and *Suessenguthiella*, possesses a rate of plastid molecular evolution 4–8 times faster than remaining Molluginaceae (Thulin et al., 2016; Smith et al., 2018), suggesting that all members of this rapidly evolving clade, and not just *Hypertelis spergulacea* and *Pharnaceum aurantium*, may possess highly rearranged plastomes. Importantly, phylotranscriptomic studies of Molluginaceae (Yang et al., 2015;

Walker et al., 2018) have not detected a similarly highly elevated rate of sequence evolution in nuclear transcriptomes of *Hypertelis, Pharnaceum,* or *Suessenguthiella*, suggesting in this instance little connection in molecular rates of evolution among genomic compartments as a whole. Similar correlations between extensive plastome rearrangement and elevated mutation rates have also been observed elsewhere (Jansen et al., 2007), including in Campanulaceae (Cosner et al., 2004), Geraniaceae (Chumley et al., 2006), and Fabaceae (Cai et al., 2008). The causes of the correlation between structural changes and increased substitution rate in plastomes are unclear, although recent work in Geraniaceae is consistent with a hypothesis that changes in plastid-targeted DNA replication, recombination, and repair mechanisms play an important role (Zhang et al., 2016).

5. Moving forward

Our results improve resolution and support for deep-level relationships in Caryophyllales through comprehensive family-level sampling, and hence provide a solid basis for future evolutionary studies within the order. As with many large phylogenomic datasets, our analyses resolve many of the outstanding relationships within Caryophyllales. However, significant conflict may underlie these relationships, and further understanding of Caryophyllales evolution will require additional sampling and analyses of key clades using genomic-scale data, including plastomes, transcriptomes, and whole genomes. The Caryophyllales represents an important system for understanding plastome evolution, providing important new insights into how gene loss and structural rearrangement influence and are influenced by interactions with other genomic compartments. This is especially true given the burgeoning availability of transcriptomic (e.g., Walker et al., 2018) and genomic data across Caryophyllales (Dohm et al., 2014; Jarvis et al., 2017). Together, these datasets will allow for further investigation of the levels of gene transfer from the plastome to the nucleus, and will also facilitate an understanding of the causes of phylogenetic incongruence between plastid and nuclear loci.

Acknowledgements

This study was supported by grant awards from the National Key Basic Research Program of China (2014CB954100), the Chinese Academy of Sciences (XDPB0201; 2017-LSF-GBOWS-02), the National Natural Science Foundation of China (31500180), the Applied Fundamental Research Foundation of Yunnan Province (2014GA003), and the US National Science Foundation (DEB 1352907 and DEB 1354048 to MJM, SAS, and SFB, and DEB 1054539 to MJM). The authors thank Lucia Lohmann and two anonymous reviewers for improvements to the manuscript. The authors thank the Shanghai Chenshan Botanical Garden, Kunming Botanical Garden, Missouri Botanical Garden, Royal Botanic Gardens Kew, Royal Botanic Garden Edinburgh, San Francisco Botanical Garden, and Wuhan Botanical Garden for providing samples, and Dr. Chunlei Xiang, Dr. Greg Stull, and Mary Merello for help with sampling. This study was facilitated by the molecular laboratory of the Germplasm Bank of Wild Species, KIB/CAS.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2018.12.023.

References

Angiosperm Phylogeny Group (APG), 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 181, 1–20.
Arakaki, M., Christin, P.A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R.M., Spriggs, E., Moore, M.J., Edwards, E.J., 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. PNAS 108, 8379–8384.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Computat. Biol. 19, 455.
Barrett, C.F., Freudenstein, J.V., Li, J., Mayfield-Jones, D.R., Perez, L., Pires, J.C., Santos, C., 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Mol. Biol. Evol. 31, 3095–3112.
Behnke, H.D., 1976. Ultrastructure of sieve-element plastids in Caryophyllales (Centrosperma), evidence for the delimitation and classification of the order. Plant Syst. Evol. 126, 31–54.
Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-revisited. Am. J. Bot. 97, 1296–1303.
Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.
Blazier, J.C., Guisinger-Bellian, M.M., Jansen, R.K., 2011. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol. Biol. 76, 263–272.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computat. Biol. 10, e1003537.
Braukmann, T.W.A., Kuzmina, M., Stefanovic, S., 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr. Genet. 55, 323–337.
Braukmann, T., Stefanović, S., 2012. Plastid genome evolution in mycoheterotrophic Ericaceae. Plant Mol. Biol. 79, 5–20.
Brockington, S.F., Alexandre, R., Ramdial, J., Moore, M.J., Crawley, S., Dhingra, A., Hilu, K., Soltis, D.E., Soltis, P.S., 2009. Phylogeny of the Caryophyllales sensu lato: Revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. Int. J. Plant Sci. 170, 627–643.
Brodribb, T.J., Feild, T.S., 2010. Leaf hydraulic evolution led a surge in leaf photosynthetic capacity during early angiosperm diversification. Ecol. Lett. 13, 175–183.
Cai, Z., Guisinger, M., Kim, H.G., Ruck, E., Blazier, J.C., McMurtry, V., Kuehl, J.V., Boore, J., Jansen, R.K., 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J. Mol. Evol. 67, 696–704.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Chaney, L., Mangelson, R., Ramaraj, T., Jellen, E.N., Maughan, P.J., 2016. The complete chloroplast genome sequences for four *Amaranthus* species (Amaranthaceae). Appl. Plant Sci. 4, 1600063.
Cho, K.S., Yun, B.K., Yoon, Y.H., Hong, S.Y., Mekapogu, M., Kim, K.H., Yang, T.J., 2015. Complete Chloroplast Genome Sequence of Tartary Buckwheat (*Fagopyrum tataricum*) and Comparative Analysis with Common Buckwheat (*F. esculentum*). PLoS ONE 10, e0125332.
Chumley, T.W., Palmer, J.D., Mower, J.P., Fourcade, H.M., Calie, P.J., Boore, J.L., Jansen, R.K., 2006. The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. 23, 2175–2190.
Cosner, M.E., Raubeson, L.A., Jansen, R.K., 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol. Biol. 4, 27.
Crawley, S.S., Hilu, K.W., 2012a. Caryophyllales: Evaluating phylogenetic signal in *trnK* intron versus *matK*. J. Syst. Evol. 50, 387–410.
Crawley, S.S., Hilu, K.W., 2012b. Impact of missing data, gene choice, and taxon sampling on phylogenetic reconstruction: the Caryophyllales (angiosperms). Plant Syst. Evol. 298, 297–312.
Cuénoud, P., Savolainen, V., Chatrou, L.W., Powell, M., Grayer, R.J., Chase, M.W., 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. Am. J. Bot. 89, 132–144.
Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17, 134.
Drew, B.T., Ruhfel, B.R., Smith, S.A., Moore, M.J., Briggs, B.G., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., 2014. Another look at the root of the angiosperms reveals a familiar tale. Syst. Biol. 63, 368–382.
Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörenson, T.R., Stracke, R., Reinhardt, R., Goesmann, A., Kraft, T., Schulz, B., Stadler, P.F., Schmidt, T., Gabaldón, T., Lehrach, H., Weisshaar, B., Himmelbauer, H., 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505, 546–549.
Dong, W., Xu, C., Li, D., Jin, X., Li, R., Lu, Q., Suo, Z., 2016. Comparative analysis of the complete chloroplast genome sequences in psammophytic *Haloxylon* species (Amaranthaceae). PeerJ 4, e2699.
Doyle, J., Hotton, C., 1991. Diversification of early angiosperm pollen in a cladistics context. In: Blackmore, S., Barnes, S. (Eds.), Pollen and Spores: Pattern of Diversification. Clarendon Press, Oxford, UK, pp. 169–195.
Ellison, A.M., Gotelli, N.J., 2001. Evolutionary ecology of carnivorous plants. Trends Ecol. Evol. 16, 623–629.
Fan, K., Sun, X.J., Huang, M., Wang, X.M., 2016. The complete chloroplast genome sequence of the medicinal plant *Rheum palmatum* L. (Polygonaceae). Mitochondrial DNA A 27, 2935–2936.
Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.
Feild, T.S., Brodribb, T.J., Iglesias, A., Chatelet, D.S., Baresch, A., Upchurch, G.R., Gomez, B., Mohr, B.A.R., Coiffard, C., Kvacek, J., Jaramillo, C., 2011. Fossil evidence for Cretaceous escalation in angiosperm leaf vein evolution. PNAS 108, 8363–8366.

Fu, C.N., Li, H.T., Milne, R., Zhang, T., Ma, P.F., Yang, J., Li, D.Z., Gao, L.M., 2017. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. BMC Genom. 18, 956.
Funk, H.T., Berg, S., Krupinska, K., Maier, U.G., Krause, K., 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7, 45.
Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Philos. Trans. Roy. Soc. Lond. B Biol. Sci. 363, 4023–4029.
Giannasi, D.E., 1992. Ecolutionary relationships of the Caryophyllidae based on comparative *rbcL* sequences. Syst. Biol. 17, 1–15.
Givnish, T.J., Spalink, D., Ames, M., Lyon, S.P., Hunter, S.J., Zuluaga, A., Iles, W.J.D., Clements, M.A., Arroyo, M.T.K., Leebens-Mack, J., Endara, L., Kriebel, R., Neubig, K.M., Whitten, W.M., Williams, N.H., Cameron, K.M., 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc. Roy. Soc. B. Biol. Sci. 282, 2108–2111.
Goremykin, V.V., Nikiforova, S.V., Bininda-Emonds, O.R., 2010. Automated removal of noisy data in phylogenomic analyses. J. Mol. Evol. 71, 319–331.
Gurusamy, R., Lee, D.-H., Park, S., 2016. The complete chloroplast genome sequence of *Dianthus superbus* var. *longicalycinus*. Mitochondrial DNA 27, 2015–2017.
Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. Nucl. Aci. Res. 41 e129 e129.
Halbritter, H., Hesse, M., Weber, M., 2012. The unique design of pollen tetrads in *Dionaea* and *Drosera*. Grana 51, 148–157.
Hernández-Ledesma, P., Berendsohn, W.G., Borsch, T., Mering, S.V., Akhani, H., Arias, S., Castañeda-Noa, I., Eggli, U., Eriksson, R., Flores-Olvera, H., Fuentes-Bazán, S., Kadereit, G., Klak, C., Korotkova, N., Nyffeler, R., Ocampo, G., Ochoterena, H., Oxelman, B., Sanchez, R.K.R.A., Schlumpberger, B.O., Uotila, P., 2015. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. Willdenowia 45, 281–383.
Hertle, A.P., Blunder, T., Wunder, T., Pesaresi, P., Pribil, M., Armbruster, U., Leister, D., 2013. PGRL1 is the elusive ferredoxin-plastoquinone reductase in phptosynthetic cyclic electron flow. Mol. Cell 49, 511–523.
Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. PNAS 104, 19369–19374.
Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmockel, S.M., Li, B., Boem, T.J., Borm, T.J.A., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N., Kharbatia, N.M., Rupper, R.R., Sharp, A.R., Dally, N., Boughton, B.A., Woo, Y.H., Gao, G., Schijlen, E.G.W.M., Guo, X., Momin, A.A., Negrão, S., Al-Babili, S., Gehring, C., Roessner, U., Jung, C., Murphy, K., Arold, S.T., Gojobori, T., van der Linden, C.G., van Loo, E.N., Jellen, E.N., Maughan, P.J., Tester, M., 2017. The genome of *Chenopodium quinoa*. Nature 542, 307–312.
Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.
Jiang, W., Chen, S.-Y., Wang, H., Li, D.-Z., Wiens, J.J., 2014. Should genes with missing data be excluded from phylogenetic analyses? Mol. Phylogenet. Evol. 80, 308–318.
Kadereit, G., Borsch, T., Weising, K., Freitag, H., 2003. Phylogeny of amaranthaceae and chenopodiaceae and the evolution of C_4 photosynthesis. Int. J. Plant Sci. 164, 959–986.
Kang, Y., Lee, H., Kim, M.K., Shin, S.C., Park, H., Lee, J., 2015. The complete chloroplast genome of antarctic pearlwort, *Colobanthus quitensis* (kunth) bartl. (caryophyllaceae). DNA Seq. 27, 4677–4678.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kim, H.T., Kim, J.S., Moore, M.J., Neubig, K.M., Williams, N.H., Whitten, W.M., Kim, J.H., 2015. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. PLoS ONE 10, e0142215.
Koenen, E.J.M., Clarkson, J.J., Pennington, T.D., Chatrou, L.W., 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. New Phytol. 207, 327–339.
Kubitzki, K., Rohwer, J.G., Bittrich, V., 1993. The Families and Genera of Vascular Plants II. Springer, Berlin, Heidelberg & New York.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Kung, H., Tsien, C., 1979. Flora Reipublicae Popularis Sinicae 25(2). Science Press, Beijing.
Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Li, H., Cao, H., Cai, Y.F., Wang, J.H., Qu, S.P., Huang, X.Q., 2014. The complete chloroplast genome sequence of sugar beet (*Beta vulgaris* ssp. *vulgaris*). Mitochondrial DNA 25, 209–211.
Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J., 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265–272.
Lidgard, S., Crane, P.R., 1988. Quantitative analyses of the early angiosperm radiation. Nature 331, 344–346.
Liu, B., Ye, J., Liu, S., Wang, Y., Yang, Y., Lai, Y., Zeng, G., Lin, Q., 2015. Families and genera of Chinese angiosperms: a synoptic classification. Biod. Sci. 23, 225–231.
Logacheva, M.D., Samigullin, T.H., Dhingra, A., Penin, A.A., 2008. Comparative chloroplast genomics and phylogenetics of *Fagopyrum esculentum* ssp. *ancestrale*–A wild ancestor of cultivated buckwheat. BMC Plant Biol. 8, 59.
Lopez-Nieves, S., Yang, Y., Timoneda, A., Wang, M., Feng, T., Smith, S.A., Brockington, S.F., Maeda, H.A., 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytol. 217, 896–9008.
Luo, Y., Lu, L., Wortley, A.H., Li, D.Z., Wang, H., Blackmore, S., 2015. Evolution of angiosperm pollen. 3. Monocots. Ann. Mo. Bot. Gard. 101, 406–455.
Ma, P.F., Zhang, Y.X., Zeng, C.X., Guo, Z.H., Li, D.Z., 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intracTable bamboo tribe Arundinarieae (Poaceae). Syst. Biol. 63, 933–950.
Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., Hernández-Hernández, T., 2015. A metacalibrated time-tree cocuments the early rise of flowering plant phylogenetic diversity. New Phytol. 207, 437–453.
Martín, M., Sabater, B., 2010. Plastid *ndh* genes in plant evolution. Plant Physiol. Biochem. 48, 636–645.
Martins, E.P., Hansen, T.F., 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149, 646–667.
McElwain, J.C., Willis, K.J., Lupia, R., 2005. Cretaceous CO_2 decline and the radiation and diversification of Angiosperms. Ecol. Stud. 177, 133–165.
Méndez, M., Karlsson, P.S., 1999. Costs and benefits of carnivory in plants: insights from the photosynthetic performance of four carnivorous plants in a subarctic environment. Oikos 86, 105–112.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetics trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA, pp. 1–8.
Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, G., Solti, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. PNAS 107, 4623–4628.
Morrison, D.A., Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. Mol. Biol. Evol. 14, 428–441.
Müller, K., Borsch, T., 2005. Phylogenetics of Amaranthaceae based on *matK/trnK* sequence data – evidence from parsimony, likelihood, and Bayesian analyses. Ann. Mol. Bot. Gard. 92, 66–102.
Nienaber, M.A., Thieret, J.W., 2003. Phytolaccaceae. In: Flora of North America Editorial CommitteeFlora of North America Editorial Committee (Ed.), Flora of North America North of Mexico 4. Oxford University Press, New York, pp. 3–11.
Ogden, T.H., Rosenberg, M.S., 2006. Multiple sequence alignment accuracy and phylogenetic inference. Syst. Biol. 55, 314–328.
Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W., Manhart, J.R., 1988. Chloroplast and variation and plant phylogeny. Ann. Mol. Bot. Gard. 75, 1180–1206.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.
Parks, M., Cronn, R., Liston, A., 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic dataset from *Pinus* L. (Pinaceae). BMC Evol. Biol. 12, 100.
Peredo, E.L., King, U.M., Les, D.H., 2013. The plastid genome of *Najas flexilis*: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. PLoS ONE 8, e68591.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Pick, K.S., Philippe, H., Schreiber, F., Erpenbeck, D., Jackson, D.J., Wrede, P., Wiens, M., Alié, A., Morgenstern, B., Manuel, M., et al., 2010. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Mol. Biol. Evol. 27, 1983–1987.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team., 2016. _nlme: Linear and nonlinear mixed effects models_.R package version 3.1-128, < http://CRAN.R-project.org/package=nlme≥.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
Pozzi, L., 2016. The role of forest expansion and contraction in species diversification among galagos (primates: Galagidae). J. Biogeogr. 43, 1930–1941.
Raman, G., Park, S.J., 2015. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. PLoS ONE 10, e0141329.
Rambaut, A., 2012. FigTree version 1.4.0. < http://tree.bio.ed.ac.uk/software/Figtee/≥.
Rambaut, A., Suchard, M.A., Drummond, A.J., 2014. Tracer v1.6. < http://beast.bio.ed.ac.uk/Tracer≥.
Rodriguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56, 389–399.
Ruhfel, B.R., Gitzendammer, M.A., Soltis, P.S., Soltis, D.E., Burleigh, J.G., 2014. From algae to angiosperms – inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14, 23.
Ruhlman, T.A., Jansen, R.K., 2014. The plastid genomes of flowering plants. In: Maliga, P. (Ed.), Chloroplast Biotechnology: Methods and Protocols, Methods in Molecular Biology. Humana Press, New York, pp. 3–38.
Sage, R.F., Christin, P.A., Edwards, E.J., 2011. The C_4 plant lineages of planet Earth. J. Exp. Bot. 62, 3155–3169.
Sanderson, M.J., Copetti, D., Búrquez, A., Bustamante, E., Charboneau, J.L.M., Eguiarte, L.E., Kumar, S., Lee, H.O., Lee, H.O., Lee, J., McMahon, M., Steele, K., Wing, R., Yang, T.J., Zwickl, D., Wojciechowski, M.F., 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted

repeat. Am. J. Bot. 102, 1115–1127.
Schäferhoff, B., Müller, K.F., Borsch, T., 2009. Caryophyllales phylogenetics: disentangling Phytolaccaceae and Molluginaceae and description of Microteaceae as a new isolated family. Willdenowia 39, 209–228.
Schmitz-Linneweber, C., Maier, R.M., Alcaraz, J.P., Cottet, A., Herrmann, R.G., Mache, R., 2001. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol. Biol. 45, 307–315.
Schulte, P., Alegret, L., Arenillas, I., Arz, J.A., Barton, P.J., Bown, P.R., Bralower, T.J., Christeson, G.L., Claeys, P., Cockell, C.S., Collins, G.S., Deutsch, A., Goldin, T.J., Goto, K., Grajales-Nishimura, J.M., Grieve, R.A.F., Gulick, S.P.S., Johnson, K.R., Kiessling, W., Koeberl, Kring, C.D.A., MacLeod, K.G., Matsui, T., Melosh, J., Montanari, A., Morgan, J.V., Neal, C.R., Nichols, D.J., Norris, R.D., Pierazzo, E., Ravizza, G., Rebolledo-Vieyra, M., Reimold, W.U., Robin, E., Salge, T., Speijer, R.P., Sweet, A.R., Urrutia-Fucugauchi, J., Vajda, V., Whalen, M.T., Willumsen, P.S., 2010. The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. Science 327, pp. 1214–1218.
Sloan, D.B., Triant, D.A., Forrester, N.J., Bergner, L.M., Wu, M., Taylor, D.R., 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol. Phylogenet. Evol. 72, 82–89.
Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322, 86–89.
Smith, S.A., O'Meara, B.C., 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28, 2689–2690.
Smith, S.A., Brown, J.W., Yang, Y., Bruenn, R., Drummond, C.P., Brockington, S.F., Walker, J.F., Last, N., Douglas, N.A., Moore, M.J., 2018. Disparity, diversity, and duplications in the Caryophyllales. New Phytol. 217, 836–854.
Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlsward, B.S., Bell, C.D., Latvis, M., Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth, C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.-L., Hilu, K.W., Davis, C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue, M.J., Soltis, S., 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98, 704–730.
Soltis, P.S., Soltis, D.E., Chase, M.W., 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402, 402–404.
Som, A., 2015. Causes, consequences and solutions of phylogenetic incongruence. Brief. Bioinform. 16, 536–548.
Stamatakis, A., 2006a. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stamatakis, A., 2006b. Phylogenetic models of rate heterogeneity: A high performance computing perspective. In: Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS2006). IEEE Computer Society Press, Washington, pp. 278–286.
Steflova, P., Hobza, R., Vyskot, B., Kejnovsky, E., 2014. Strong accumulation of chloroplast DNA in the Y chromosomes of *Rumex acetosa* and *Silene latifolia*. Cytogenet. Genome Res. 142, 59–65.
Stevens, P.F., 2001 onwards. Angiosperm Phylogeny Website. Version 12, July 2012 [and more or less continuously updated since]. < http://www.mobot.org/MOBOT/research/APweb/ ≥.
Stull, G.W., Moore, M.J., Mandala, V.S., Douglas, N.A., Kates, H.R., Qi, X., Brockington, S.F., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., 2013. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Appl. Plant Sci. 1, 1200497.
Thulin, M., Moore, A.J., El-Seedi, H., Larsson, A., Christin, P.A., Edwards, E.J., 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family Corbichoniaceae. Taxon 65, 775–793.
Walker, J.F., Yang, Y., Moore, M.J., Mikenas, J., Timoneda, A., Brockington, S.F., Smith, S.A., 2017. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. Am. J. Bot. 104, 858–867.
Walker, J.F., Yang, Y., Feng, T., Timoneda, A., Mikenas, J., Hutchinson, V., Edwards, C., Wang, N., Ahluwalia, S., Olivieri, J., Walker-Hale, N., Majure, L.C., Puente, R., Kadereit, G., Lauterbach, M., Eggli, U., Flores-Olvera, H., Ochoterena, H., Brockington, S.F., Moore, M.J., Smith, S.A., 2018. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight to the evolution of Caryophyllales. Am. J. Bot. 105, 446–462.
Wang, H., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., Davis, C.C., Latvis, M., Manchester, S.R., Soltis, D.E., 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. PNAS 106, 3853–3858.
Wang, W., Ortiz, R.D.C., Jacques, F.M.B., Xiang, X.G., Li, H.-L., Lin, L., Li, R.-Q., Liu, Y., Soltis, P.S., Soltis, D.E., Chen, Z.-D., 2012. Menispermaceae and the diversification of tropical rainforests near the Cretaceous-Paleogene boundary. New Phytol. 195, 470–478.
Welsh, S.L., Cromptom, C.W., Clemants, S.E., 2003. Chenopodiaceae. In: Flora of North America Editorial CommitteeFlora of North America Editorial Committee (Ed.), Flora of North America North of Mexico 4. Oxford University Press, New York, pp. 258–404.
Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265.
Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31, 3350–3352.
Wicke, S., Müller, K.F., de Pamphilis, C.W., Quandt, D., Wickett, N.J., Zhang, Y., Renner, S.S., Schneeweiss, G.M., 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25, 3711–3725.
Wicke, S., Schäferhoff, B., de Pamphilis, C.W., Müller, K.F., 2014. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. Mol. Biol. Evol. 31, 529–545.
Wickett, N.J., Zhang, Y., Hansen, S.K., Roper, J.M., Kuehl, J.V., Plock, S.A., Wolf, P.G., dePamphilis, C.W., Boore, J.L., Goffinet, B., 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol. Biol. Evol. 25, 393–401.
Wiens, J.J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. 52, 528–538.
Wiens, J.J., 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Syst. Biol. 54, 731–742.
Wikström, N., Kainulainen, K., Razafimandimbison, S.G., Smedmark, J.E., Bremer, B., 2015. A revised time of the asterids: establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). PLoS ONE 10, e0126690.
Wikström, N., Savolainen, V., Chase, M.W., 2001. Evolution of the angiosperms: calibrating the family tree. Proc. Roy. Soc. Lond. B Biol. Sci. 268, 2211–2220.
Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252–3255.
Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J., Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. PNAS 109, 17519–17524.
Xia, X., 2013. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30, 1720–1728.
Xiang, Y., Huang, C.H., Hu, Y., Wen, J., Li, S., Yi, T., Chen, H., Xiang, J., Ma, H., 2016. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol. Biol. Evol. 34, 262–281.
Yang, J.B., Li, D.Z., Li, H.T., 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Res. 14, 1024–1031.
Yang, Y., Moore, M.J., Brockington, S.F., Soltis, D.E., Wong, G.K.S., Carpenter, E.J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., Sage, R.F., Covshoff, S., Hibberd, J.M., Nelson, M.N., Smith, S.A., 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Mol. Biol. Evol. 32, 2001–2014.
Yang, Y., Moore, M.J., Brockington, S.F., Mikenas, J., Olivieri, J., Walker, J.F., Smith, S.A., 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy event in Caryophyllales, including two allopolyploidy event. New Phytol. 217, 855–870.
Yu, Y., Wortley, A.H., Lu, L., Li, D.Z., Wang, H., Blackmore, S., 2018. Evolution of angiosperm pollen. 5. Basal Superasteridae (Berberidopsidales, Caryophyllales, Cornales, Ericales and Santalales) plus Dilleniales. Ann. Mo. Bot. Gard. 103, 106–161.
Zeng, L., Zhang, N., Zhang, Q., Endress, P.K., Huang, J., Ma, H., 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 214, 1338–1354.
Zhang, J., Ruhlman, T.A., Sabir, J.S.M., Blazier, J.C., Weng, M.L., Park, S., Jansen, R.K., 2016. Coevolution between Nuclear-Encoded DNA Replication, Recombination, and Repair Genes and Plastid Genome Complexity. Genome Biol. Evol. 8, 622–634.
Zhang, M.Y., Lu, L., Wortley, A.H., Wang, H., Li, D.Z., Blackmore, S., 2017a. Evolution of angiosperm pollen: 4. Basal eudicots. Ann. Mol. Bot. Gard. 102, 141–182.
Zhang, S.D., Jin, J.J., Chen, S.Y., Chase, M.W., Soltis, D.E., Li, H.T., Yang, J.B., Li, D.Z., Yi, T.S., 2017b. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytol. 214, 1355–1367.
Zheng, Y., Wiens, J.J., 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST? Mol. Phylogenet. Evol. 85, 41–49.
Zhong, B., Deusch, O., Goremykin, V.V., Penny, D., Biggs, P.J., Atherton, R.A., Nikiforova, S.V., Lockhart, P.J., 2011. Systematic error in seed plant phylogenomics. Genome Biol. Evol. 3, 1340–1348.
Zhu, A., Guo, W., Gupta, S., Fan, W., Mower, J.P., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209, 1747–1756.