Mitochondrial genes from 18 angiosperms fill sampling gaps for phylogenomic inferences of the early diversification of flowering plants

Jia-Yu Xue[1,2,3], Shan-Shan Dong[4], Ming-Qiang Wang[1], Tian-Qiang Song[5], Guang-Can Zhou[1,6], Zhen Li[3], Yves Van de Peer [2,3,7], Zhu-Qing Shao[5], Wei Wang[8], Min Chen[1], Yan-Mei Zhang[1], Xiao-Qin Sun[1], Hong-Feng Chen[9], Yong-Xia Zhang[10], Shou-Zhou Zhang[4], Fei Chen[11], Liang-Sheng Zhang[11], Cymon Cox[12], Yang Liu[4]*,Qiang Wang[5]*, and Yue-Yu Hang[1]*

[1]Center for Plant Diversity and Systematics, Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China

[2]College of Horticulture, Nanjing Agricultural University, Nanjing 210093, China

[3]Department of Plant Biotechnology and Bioinformatics, VIB-UGent Center for Plant Systems Biology, Ghent University, Ghent 9052, Belgium

[4]Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, Guangdong, China

[5]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210046, China

[6]College of Agricultural and Biological Engineering (College of Tree Peony), Heze University, Heze 274015, Shandong, China

[7]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0028, South Africa

[8]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, Beijing 100093, China

[9]South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[10]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, Guangdong, China

[11]State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Provincial Key Laboratory of Haixia Applied Plant SystemsBiology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[12]Centro de Ciências do Mar, Universidade do Algarve, Faro 8005-319, Portugal

*Authors for correspondence. Yang Liu. E-mail: yang.liu0508@gmail.com; Qiang Wang. E-mail: wangq@nju.edu.cn; Yue-Yu Hang. E-mail: hangyueyu@cnbg.net

Abstract

The early diversification of angiosperms is thought to have been a rapid process, which may complicate phylogenetic analyses of early angiosperm relationships. Plastid and nuclear phylogenomic studies have raised several conflicting hypotheses regarding overall angiosperm phylogeny, but mitochondrial genomes have been largely ignored as a relevant source of information. Here we sequenced mitochondrial genomes from 18 angiosperms to fill taxon-sampling gaps in Austrobaileyales, magnoliids, Chloranthales, Ceratophyllales, and major lineages of eudicots and monocots. We assembled a data matrix of 38 mitochondrial genes from 107 taxa to assess how well mitochondrial genomic data address current

uncertainties in angiosperm relationships. Although we recovered conflicting phylogenies based on different data sets and analytical methods, we also observed congruence regarding deep relationships of several major angiosperm lineages: Chloranthales were always inferred to be the sister group of Ceratophyllales, Austrobaileyales to mesangiosperms, and the unplaced Dilleniales was consistently resolved as the sister to superasterids. Substitutional saturation, GC compositional heterogeneity, and codon-usage bias are possible reasons for the noise/conflict that may impact phylogenetic reconstruction; and angiosperm mitochondrial genes may not be substantially affected by these factors. The third codon positions of the mitochondrial genes appear to contain more parsimony-informative sites than the first and second codon positions, and therefore produced better resolved phylogenetic relationships with generally strong support. The relationships among these major lineages remain incompletely resolved, perhaps as a result of the rapidity of early radiations. Nevertheless, data from mitochondrial genomes provide additional evidence and alternative hypotheses for exploring the early evolution and diversification of the angiosperms.

Keywords: early angiosperm radiation, incongruence, mitochondrial genome, phylogenomics, systematic error.

1 Introduction

As a result of the recent developments in high-throughput sequencing technologies, genomic data have accumulated rapidly, opening novel avenues for phylogenetics (e.g., Liu et al., 2014; Wickett et al., 2014; Li et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). Before the era of phylogenomics, only a limited number of molecular markers was used for phylogenetic analyses, and the accuracy of inferences was, therefore, inevitably affected by stochastic error (e.g., Rodrigo et al., 1993; Cao et al., 1998; Phillips et al., 2004). As phylogenomics incorporates much more data, it should be able to compensate for this (Wickett et al., 2014; One Thousand Plant Transcriptomes Initiative, 2019). As large amounts of genomic data may provide adequate phylogenetic signals in analyses, phylogenomics stands as an effective and popular strategy to resolve widespread phylogenetic ambiguities (e.g., Edger et al., 2018; Li et al., 2019; Liu et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019).

Angiosperms, which appear to have originated in the Mesozoic era (Zeng et al., 2014; Magallon et al., 2015; Morris et al., 2018; Li et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019; Yang et al., 2020), have evolved into the most diverse land-plant clade on the Earth, with over 350 000 extant species (APG IV, 2016). They exhibit incredible diversity in morphology, physiology, and reproductive forms, as well as adaptations to many different environments. The early diversification of angiosperms involved periods of rapid radiation, occurring within very short periods of time with few transitional fossil records (Herendeen et al., 2017; Coiro et al., 2019). Indeed, the sudden burst of angiosperms was referred to by Darwin as an “abominable mystery” (Friedman, 2009). As a result of their rapid radiation, early angiosperm lineages accumulated only a limited number of substitutions per gene during divergence, consistent with generally poor phylogenetic branch support especially when limited molecular markers are sampled per taxon (e.g., Hilu et al., 2003; Qiu et al., 2006b, 2010). The phylogenetic relationships inferred among early angiosperms have long been in conflict, mainly concerning two ambiguities “below” and “above” Austrobaileyales: (i) below; which lineage is sister to the rest of the extant angiosperms (i.e., Amborellales solely or a clade comprising Amborellales and Nymphaeales) (e.g., Qiu et al., 1999; Leebens-Mack et al., 2005; Qiu et al., 2010; Drew et al., 2014; Zhong & Betancur-R, 2017) and (ii) above; what are the phylogenetic relationships among the five lineages of mesangiosperms (i.e., monocots, eudicots, magnoliids, Chloranthales, and Ceratophyllales) (e.g., Qiu et al., 2006b;

Moore et al., 2007; Wickett et al., 2014; Zeng et al., 2014; One Thousand Plant Transcriptomes Initiative, 2019; Yang et al., 2020). The eudicots and monocots are the two largest lineages of angiosperms, accounting for 75% and 20% of the ~350 000 angiosperm species, respectively (APG IV, 2016). Magnoliids include ~9000 species and were once considered to be the "most primitive" angiosperm lineage (Cronquist, 1981). Chloranthales and Ceratophyllales are two small clades, including only 77 and 6 species, respectively (APG IV, 2016), but are both considered to be ancient angiosperm lineages based on fossil records and their unusual morphological characters (Cronquist, 1981; Eklund et al., 2004; Friis et al., 2006; Dilcher & Wang, 2009; Friis et al., 2010).

In the last decade, large numbers of phylogenomic studies have tried to address these phylogenetic conundrums. On the one hand, relationships among the major angiosperm lineages have become relatively clear or are fully consistent with earlier phylogenetic studies: relevant studies overwhelmingly find Amborellales as the sister group of all other angiosperms (Jansen et al., 2007; Moore et al., 2010; Ruhfel et al., 2014; Wickett et al., 2014; Zeng et al., 2014; Simmons, 2017; Zhong & Betancur-R, 2017; Gitzendanner et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019). On the other hand, ambiguities remain regarding the relationships of the five mesangiosperm lineages. In the last 25 years, 18 different topologies (Table S1) (Mathews & Donoghue, 1999; Qiu & Palmer, 1999; Soltis et al., 1999; Barkman et al., 2000; Graham & Olmstead, 2000; Zanis et al., 2002; Hilu et al., 2003; Leebens-Mack et al., 2005; Qiu et al., 2006a, 2006b; Jansen et al., 2007; Moore et al., 2007; Endress & Doyle, 2009; Finet et al., 2010; Moore et al., 2010; Qiu et al., 2010; Lee et al., 2011; Moore et al., 2011; Soltis et al., 2011; Zhang et al., 2012; Goremykin et al., 2013; Xi et al., 2014; Zeng et al., 2014; Gitzendanner et al., 2018; Li et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019; Yang et al., 2020) have been proposed, but more recent phylogenomic studies seem to converge on several alternative topologies. The first is mainly obtained through the use of mainly plastid genes and one nuclear gene (18S rDNA) and was considered by the Angiosperm Phylogeny Group (APG) as part of their angiosperm classification (Fig. 1A) (APG III, 2009). These studies recover monocots as the sister group of a clade comprising Ceratophyllales and eudicots, with Chloranthales sister to magnoliids (Graham et al. 2006; Moore et al., 2007, 2010; Graham & Iles, 2009; Finet et al., 2010; Ruhfel et al., 2014; Gitzendanner et al., 2018). The second arrangement is obtained using low copy nuclear genes, which resolves successive sister relationships of monocots, magnoliids, eudicots to Chloranthales and Ceratophyllales (Fig. 1B) (Lee et al., 2011; Wickett et al., 2014; Zeng et al., 2014). As such, incongruences have persisted, the recently updated APG IV classification included a summary tree of phylogenetic results with a polytomy comprising Chloranthales, magnoliids and a clade comprising the other mesangiosperm lineages (APG IV, 2016). In 2019, two phylogenomic studies adopting the most comprehensive taxa sampling to date—2881 plastid genomes (Li et al., 2019) (Fig. 1C) and 1124 transcriptomes (One Thousand Plant Transcriptomes Initiative, 2019) (Fig. 1D)—were published, but even so, the incongruence remains.

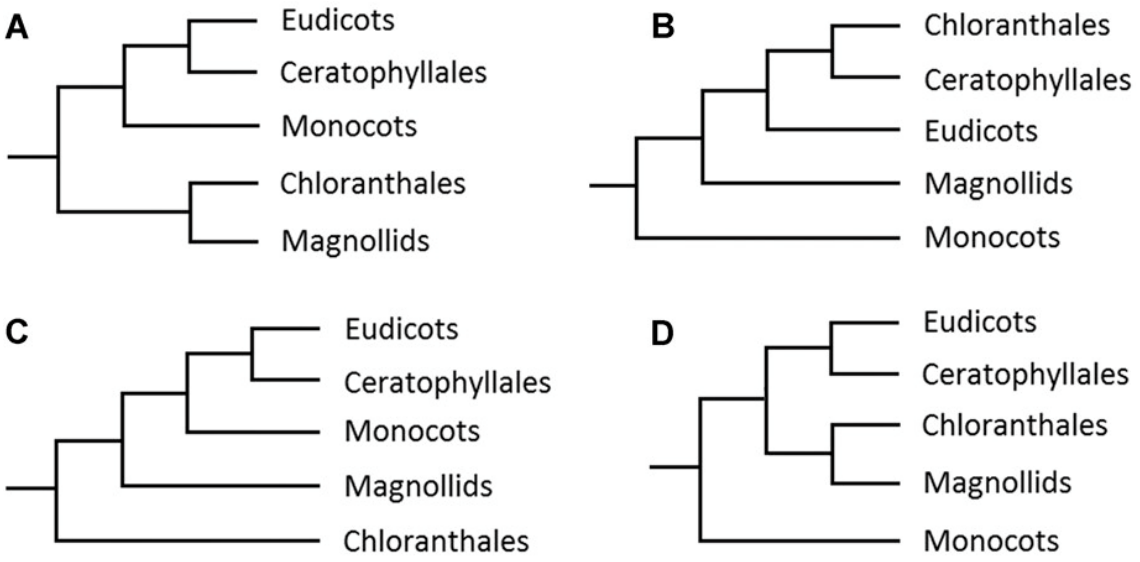

Figure 1. Abbreviated phylogenetic trees of the five mesangiosperm lineages from recent phylogenomic studies using different data sets. **A**, Pt genes (accepted by APG III) (Moore et al., 2007; Graham & Iles, 2009; Finet et al., 2010; Moore et al., 2010; Ruhfel et al., 2014; Gitzendanner et al., 2018). **B**, Nu genes (Lee et al., 2011; Wickett et al., 2014; Zeng et al., 2014). **C**, Pt genes (Li et al., 2019). **D**, Nu genes (One Thousand Plant Transcriptomes Initiative, 2019).

It is noteworthy that all of the inferred relationships discussed above are based on either plastid genes or nuclear sequences, but none was based on large-scale analysis of mitochondrial (mt) sequences. As one of the three genomes in plants, mt genomes have been considered potentially important for phylogenetic research, but only a handful of mt genes (e.g., *matR*, *atp1*, mtSSU, mtLSU, *nad5*, *rps3*) have been employed in phylogenetic studies in plants (e.g., Beckert et al., 1999; Qiu & Palmer, 1999; Barkman et al., 2000; Bowe et al., 2000; Forrest et al., 2006; Qiu et al., 2006b, 2010; Zhu et al., 2007; Mennes et al., 2013), and only a few recent plant studies have included phylogenomic analysis of mitochondrial data (e.g., Bell et al., 2020; Gomez et al., 2020). Mt genomes have been long neglected in favor of plastid genomes or nuclear genes. The reasons for this may include their very slow rate of evolution (Palmer & Herbon, 1988; Knoop, 2004), which makes them largely unsuitable for phylogenetic inference when used individually (e.g., Hilu et al., 2014), and the highly variable structure of mt genomes (Gualberto & Newton, 2017), which complicates the sequencing and subsequent assembly and annotation of mitochondrial sequences (Sloan, 2013). However, mt genes may be well suited to unveiling deep-level phylogenetic relationships. First of all, they comprise approximately 40 genes and may provide abundant phylogenetic signal when used together. Second, the substitution rate of mt genes is, on average, much lower than that of plastid (<3×) or nuclear (<10×) genes (Wolfe et al., 1987; Palmer & Herbon, 1988; Drouin et al., 2008), which should result in lower levels of saturation, and hence, less homoplasy (Nickrent et al., 2000; Qiu et al., 2010), although saturation and homoplasy are not necessarily problematic if the correct model is used and taxon sampling is sufficient (Zwickl & Hillis, 2002). Additionally, mt genes have clear orthologous relationships due to the fact that gene duplication is rare, which can otherwise complicate inferring the evolutionary history of species using nuclear genes where duplication, gene loss, and hybridization are common (Wang et al., 2005; Cui et al., 2006; Freeling, 2009; Van de Peer et al., 2009; Proost et al., 2011; Buggs et al., 2012; Vanneste et al., 2014; Tasdighian et al., 2017), although gene and genome duplication events, to some extent, could be used to infer species trees (e.g., Ness et al., 2011). For example, a phylogenomic study employing 60 streptophyte mt genomes successfully resolved liverworts as the sister group of other land-plant lineage, and successive divergences of mosses, hornworts, and vascular plants, respectively (Liu et al., 2014), but see Bell et al. (2020) for conflicting findings using 139 mitochondrial gene sets.

Since 2000, only one study has employed solely mt genes to reconstruct broad angiosperm phylogeny (Qiu et al., 2010). However, the gene numbers in that study comprised only four loci (*atp1*, *matR*, *nad5*, and *rps3*). Complete mt genome data have not yet been used for phylogenetic analyses in angiosperms; even today, a mere 80 angiosperms have had their entire mt genome sequenced, mostly eudicots or monocots. In contrast, for the magnoliids, a lineage comprising four orders, only one sequenced mt genome has been created, while one lineage of the ANA order—Austrobaileyales and two lineages in mesangiosperms (Chloranthales and Ceratophyllales)—lack mt genome sequences. Here, we used a modified approach for the isolation of mitochondria from plant tissues and a simplified process of mt DNA extraction to obtain high purity and quality DNA for mt genome sequencing (Ahmed & Fu, 2015). Using this technique, we generated mt genomes from 18 additional angiosperm species to fill the gaps in early angiosperm evolution. We then assembled an mt gene matrix comprising 38 mt genes from 107 seed plants and inferred the phylogenetic relationships of major angiosperm clades to address their phylogenetic relationships.

2 Material and Methods

2.1 Taxon sampling

Mitochondrial protein-coding genes from 18 angiosperms, including 1 from Austrobaileyales, 11 from magnoliids, 2 from Chloranthales, 1 from Ceratophyllales, 2 from eudicots, and 1 from monocots, were newly sequenced here. Together with all other available angiosperm mt genomes from the NCBI database, our taxon sampling covered all major lineages of angiosperms (i.e., the three orders of the ANA grade and all five lineages of the Mesangiospermae) (APG IV, 2016). The previously sequenced mt genomes (three gymnosperms and 86 angiosperms) were downloaded from NCBI (http://www.ncbi.nlm.nih.gov). Voucher information and GenBank accession numbers for analyzed samples are provided in Table S2. The final data matrix was composed of 107 seed plants, including 104 angiosperms and three gymnosperm outgroups.

2.2 Mitochondrial DNA extraction, sequencing, assembly, and annotation

Approximately 10–20 g fresh tissues (leaves, young buds, or pericarps) were collected from each plant. The isolation of mitochondria and mt DNA followed a modified method (Ahmed & Fu, 2015), skipping the gradient centrifugation steps. The extracted mt DNA was assessed using a Qubit fluorometer (Invitrogen, San Diego, CA, USA) for quantity and a NanoDrop spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) for quality, and then sequenced using the Illumina HiSeq2000 platform at NovoGene (Nanjing, China). The generated raw data were assembled *de novo* using CLC Genomics Workbench v6.5 (CLC Bio, Aarhus, Denmark) with default parameters. To identify the mt contigs, the assembled contigs were BLASTed against the mt genomes of *Arabidopsis thaliana* (Unseld et al., 1997), *Liriodendron tulipifera* (Richardson et al., 2013), *Nelumbo nucifera* (Gui et al., 2016), and *Spirodela polyrhiza* (Wang et al., 2012). We checked the sequence depth to avoid contigs with low depth (i.e., 10 × lower than the average depth of other mt contigs) that may be derived from nuclear genomes. The mitochondrial protein-coding genes of the above angiosperms were compiled and used to annotate the newly generated mt contigs in Geneious v6.0.3 (Biomatters Ltd, Auckland, New Zealand) with the option “annotation” with a similarity threshold of 60%. The mitochondrial protein-coding sequences were then extracted using a custom Perl script described previously (Liu et al., 2014), https://datadryad.org/stash/dataset/doi:10.5061/dryad.7b470). The assembled contigs and annotation files are available from the Dryad database (https://datadryad.org/resource/doi:10.5061/dryad.fj6q573rq) and related information can be found from Table S8.

2.3 Sequence alignment and concatenation of 38 mt protein-coding genes

Mitochondrial protein-coding genes were aligned individually using MAFFT (Katoh et al., 2005) to build amino acid (aa) alignments. Poorly aligned regions were trimmed by GBLOCKS with the least stringent settings (Talavera & Castresana, 2007), such that nucleotide (nt) alignments were produced based on the corresponding amino acid alignments after removal of ambiguous positions. The above processes were automatically conducted using the program TranslatorX (Abascal et al., 2010). After removing stop codons, the 38 single-gene nucleotide alignments and corresponding amino acid alignments were concatenated into final alignments and converted into appropriate formats using Geneious v6.0.3. Both the nucleotide and amino acid data were used for subsequent phylogenetic analyses.

2.4 Phylogenetic analyses

Collectively, 38 mt protein-coding genes were used for the phylogenetic analyses, and pseudogenes and missing genes were treated as gaps in the data sets. The concatenated nucleotide data set, and the corresponding translated amino acid data set, were analyzed using maximum parsimony (MP), maximum-likelihood (ML), and Bayesian inference (BI).

For the nucleotide analyses, the MP analyses were carried out in PAUP* v.4.0b10 (Swofford, 2003). Heuristic searches were conducted with 1000 random addition replicates, with one tree held at each step during stepwise addition, and using tree-bisection-reconnection (TBR) branch swapping. MulTrees was in effect, and steepest descent was off. Bootstrapping was conducted with 1000 replicates with 10 random addition replicates per bootstrap replicate, and other using the same heuristic research options. ML analyses were performed using the parallel version of RAxML v7.2.3 (Stamatakis, 2006) under the GTR+G model. Non-parametric bootstrap (BS) analyses (Felsenstein, 1985) were implemented using GTR+CAT approximation for 100 pseudoreplicates. Bayesian analyses were performed with MrBayes (Ronquist et al., 2012) using the GTR+G model, with two runs of four chains. The Markov chain Monte Carlo (MCMC) was run for five million generations in each analysis. Burn-in and convergence were assessed using the likelihood of the runs plotted against generations using Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/), ESS values for all parameters are above 200. Posterior probabilities (PPs) of clade support were estimated by sampling trees from the posterior distribution after removal of the first 25% samples as burn-in. PartitionFinder (Lanfear et al., 2012) was used for selecting the optimal data partition scheme and the associated nucleotide substitution models, with an initial partitioning strategy by both locus and codon positions for the nucleotide data set, resulting in 32 partitions (Table S3).

For the amino acid data set, the partition analyses yielded eight partitions after an initial partitioning strategy by genes. The optimal partitioning schemes were then used in the following relevant phylogenetic analyses. For the amino acid data set, ML and BI analyses were performed. The ML analysis was carried out for searching the best tree with the mitochondrial-specific amino acid substitution model stmtREV (Liu et al., 2014), and bootstrap support was estimated based on 100 pseudoreplicates using a GTR–CAT approximation. Using RAxML, and the Bayesian MCMC analyses were implemented in PhyloBayes MPI v1.2d using the CAT+GTR+Γ model (Lartillot et al. 2009). PhyloBayes analyses were performed for five million generations. To quantify the genealogical concordance in mitochondrial phylogenomic data sets, IQ-TREE 2 (Minh et al., 2020b) was used to infer a concatenation-based species tree with 1000 ultrafast bootstrap and an edge-linked partition model, as well as 38 locus trees. Then the gene concordance factor (gCF) and the site concordance factor (sCF) (Minh et al., 2020a) for each branch of the species tree as the fraction of decisive gene trees concordant with this branch were calculated in IQ-TREE 2 with default settings. Three

gymnosperms, namely *Ginkgo biloba*, *Cycas taitungensis*, and *Welwitschia mirabilis* (Chaw et al., 2008; Guo et al., 2016), were used as outgroups in all phylogenetic analyses.

2.5 Evaluation of systematic errors

Systematic errors (evolutionary noise) in each case may conflict with authentic phylogenetic signal and lead to artifactual inference (incorrect topologies), among which substitutional saturation, GC heterogeneity, and codon-usage bias are common factors affecting phylogenetic inferences. To estimate the degree of substitutional saturation, we plotted the uncorrected P-distances against the inferred distances using the method described by Philippe and Forterre (Forterre & Philippe, 1999). The level of saturation is estimated by computing the slope of the regression line in the plot; the shallower the slope, the greater the degree of saturation. We estimated saturation for two subsets of the concatenated data: the combined first and second codon positions, and the third codon positions. The Tamura-Nei model (Tamura & Nei, 1993) was used when calculating inferred distances using the ML method. Base composition of each gene and codon position was obtained from the output of the alignment generated by TranslatorX. To visualize compositional heterogeneity, the GC percentage score of each species was plotted using Excel. The degree of codon-usage bias, synonymous codon-usage order (SCUO), was calculated by CodonO (Angellotti et al., 2007), which is based on Shannon informational theory and enables a measurement of synonymous codon-usage bias within and across genomes, as well as any correlation to GC compositional content.

3 Results

3.1 Data sets

Collectively, we used 107 taxa and 38 mt protein-coding genes (Tables S1, S2) in our phylogenetic analyses. Of the 107 taxa, three were gymnosperms, 12 were magnoliids, 23 were monocots, 63 were eudicots, and two were from Chloranthales; and Amborellales, Nymphaeales, Austrobaileyales, and Ceratophyllales each had one representative, covering 30 angiosperm orders. Eighteen newly sequenced species were included to fill the taxonomic gaps in three lineages, Austrobaileyales, Chloranthales, and Ceratophyllales, and to improve the sampling density of the three orders of magnoliids and major clades of monocots and eudicots. After removal of ambiguously aligned positions, the concatenated nucleotide data set comprised 29 277 characters, with 13 130 variable (44.8%) and 8214 parsimony-informative sites (28.1%), while the corresponding translated amino acid data set contained 9759 characters, with 5395 variable (55.3%) and 3639 parsimony-informative sites (37.3%). The proportions of variable sites for each mitochondrial gene ranged from 28.9% to 67.9%, generally with the highest percentages in the genes encoding ATP synthase subunits (*atp*), and the lowest in subunits encoding the respiratory chain complex I (*nad*). For the informative sites, a very similar pattern was observed, ranging from 15.9% to 45.2% (Fig. S1).

Among all the mt genes, those encoding subunits of the four respiratory chain complexes (complexes I–IV), ATP synthase, and cytochrome *c* biogenesis were almost all (23/25) present in at least 100 species, except for *atp9* (99 species) and *sdh4* (72 species). The other ribosomal protein genes seem to have undergone frequent independent losses, with six out of 13 genes being lost from over 30% of 107 species (Tables S2, S4). Most species (101/107 species) possessed at least three-quarters of the 38 protein-coding genes, except for *Welwitschia*, which has only 13 genes. As *Welwitschia* shows problematic long branches and erroneous placings in single-gene trees, the suspected genes were excluded. other than *Welwitschia*, problematic long branched were observed for *Vitis vinifera*, *Geranium maderense*, *Viscum album*, and our newly sequenced *Acorus gramineus* in both this study and others (Goremykin et al., 2009; Qiu et al., 2010; Park et al., 2015; Petersen et al., 2015),

therefore these taxa were entirely excluded from the analyses after inspection by a preliminary phylogenetic analysis. Eleven of the 18 newly sequenced species all possessed 38 genes, and the eudicot *Clematis terniflora* possessed the fewest genes at 33 (Table S2), although at this time whether the absent genes were truly lost in these species cannot be determined, because the mitochondrial genomes were not fully assembled.

3.2 Phylogenetic results using different data sets and analytical methods

Of all the resulting phylogenies, every lineage with more than one species (Chloranthales, magnollids, eudicots and monocots) was, respectively, recovered as a monophyletic group (Figs. 2, S2–S13), without exception. However, based on the different data sets and methods, the relationships among these major lineages differed, leading to three candidate lineages for the sister group of all other angiosperms and six topologies representing for major relationships in the mesangiosperms. Of all the analyses, the nt and combined first and second codon positions yielded similar results, and the aa and third codon positions yielded similar results. For the same data sets, the results inferred from ML and Bayesian analyses were largely congruent, whereas MP analyses produced more incongruent results (Figs. S2–S13).

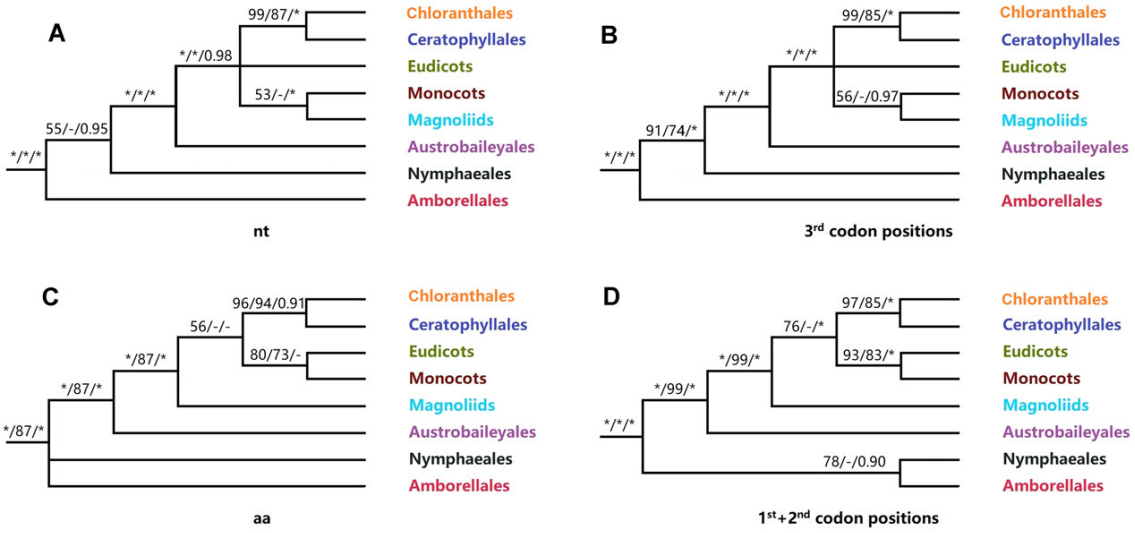

Figure 2. ML phylogenetic trees showing the relationships of angiosperms inferred from **A**, nt data, **B**, third codon positions, **C**, aa data, and **D**, combined first and second codon positions of 38 mt genes of 107 taxa. ML and MP BS support values and Bayesian PP are labeled above nodes, respectively. Asterisks indicate either BS of 100% or PP of 1.00. Nodes with ML BS support values <50% are collapsed. MP BS support values < 50%, Bayesian posterior probabilities < 0.5, or incongruent topologies among ML, MP, and Bayesian inferences are indicated as "−." For details, see Figs. S2–S13. ML, maximum-likelihood; MP, maximum parsimony; BS, bootstrap.

The results of the nt and the third codon positions all recovered Amborellales solely as the first split of angiosperms (ML_{nt}-BS = 55%; BI_{nt}-PP = 0.95; ML_3-BS = 91%; MP_3-BS = 74%; BI_3-PP = 1.0), followed by Nymphaeales (Figs. 2A, S2, S5, S6, S9, S13), with the exception that the MP analysis of the nt data recovered Nymphaeales as sister to all other angiosperms (MP_3-BS = 75%; Fig. S10). Other results that supported Nymphaeales as the first split were exclusively from the MP analyses, with poor support from the aa data (<50%; (Fig. S11) and strong support from the first and second codon positions (92%; Fig. S12). ML and Bayesian analyses of first and second codon positions data both recovered Amborellales sister to Nymphaeales with moderate support (ML_{nt}-BS = 78%; BI_{nt}-PP = 0.90; Figs. 2D, S2, S5, S6, S9, S13), and this monophyletic group was recovered as sister to all other angiosperms; the aa data however, had very poor support for this topology (Figs. 2C, S3, S7). Austrobaileyales,

the other order of the ANA grade, was constantly resolved as the sister to all mesangiosperms based on all data sets and analytical methods, with very strong support (Figs. 2, S2–S13).

The mesangiosperms were resolved as a monophyletic group by all analyses; however, five major mesangiosperm lineages had different relationships in the different analyses, the majority lacking strong support. Despite these different interlineage relationships, one result was consistent, namely, a sister-group relationship between Chloranthales and Ceratophyllales (Figs. 2, S2–S13). This relationship was consistently recovered by all data sets and analytical methods with strong support. As for the other lineages, results derived from the nt and the third codon positions favored magnoliids being sister to monocots (ML_{nt}-BS = 53%; BI_{nt}-PP = 1.0; ML_3-BS = 56%; BI_3-PP = 0.97), and eudicots either as sister to the Chloranthales-Ceratophyllales lineage (ML_{nt}-BS = 42%; BI_{nt}-PP = 0.78) or to the rest of all other mesangiosperms (ML_3-BS = 45%; BI_3-PP = 0.68) with weak support (Figs. 2A, 2B, S2, S5, S6, S9); the analyses involving the aa data and the combined first and second codon positions preferred the grouping of eudicots and monocots and placed magnoliids as the sister group of the rest of the mesangiosperms (Figs. 2C, 2D, S3, S4, S7, S8).

In summary, despite the diverse incongruent phylogenies by the different methods and data sets, the positions and relationships of a few lineages were recovered stably: Austrobaileyales as sister to all mesangiosperms and Chloranthales as sister to Ceratophyllales. In half of the resulted phylogenies, Amborellales, Nymphaeales and Austrobaileyales were successive sisters to mesangiosperms with strong support. When using the same data sets, ML and Bayesian methods (which both take account of DNA substitutional models), generated congruent or similar phylogenies, compared with the MP method, which did not. The different algorithms of the three methods may account for this discrepancy.

3.3 Detection of systematic errors within the nt data set

The different topologies resulting from the different data sets suggested that conflict between the combined first and second codon positions and the third codon positions influenced phylogenetic inferences based on these two data sets. We examined the possible causes of the systematic errors here.

First, the degree of substitutional saturation was estimated for the data set comprised of the combined first and second codon positions, and for the one comprising the third codon positions, (the shallower the slope, the greater the degree of saturation). Both the third codon positions (slope = 0.80; Fig. 3B) and the combined first and second codon positions (slope = 0.87; Fig. 3A) are not highly saturated, and appear to be at similar levels to each other.

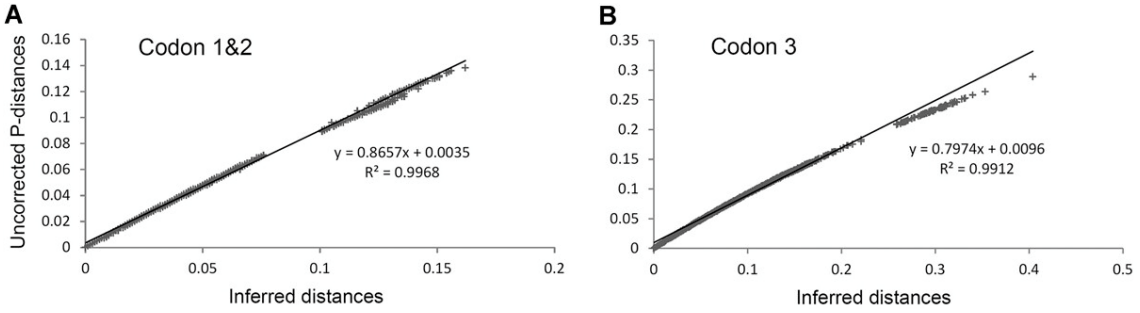

Figure 3. Saturation level in the first and second codon positions, and the third codon positions. The uncorrected P-distances were plotted against the inferred distances. The *x*-axis corresponds to the distance inferred by the maximum-likelihood method using the Tamura-Nei model (Tamura &

Nei, 1993), and the *y*-axis corresponds to the uncorrected distance observed for the same taxa pair. **A**, **B**, include pairwise comparisons between all exemplars across the whole tree.

Second, we calculated GC composition for each major angiosperm lineage and the outgroups, either across all coding regions or at each individual codon position (combined across genes) (Fig. 4; Table S5). Generally, no drastic discrepancy was observed among these lineages. The GC content of all taxa fell within a limited range of 41%–46% and over 90% (98/107) fell in a smaller range of 42%–45%. The GC content of the different codon positions also varied within limited ranges, with the first codons between 47% and 50%, second codons between 41% and 45%, and the third codons between 36% and 39%. In this case, very little GC compositional heterogeneity was detected among lineages.

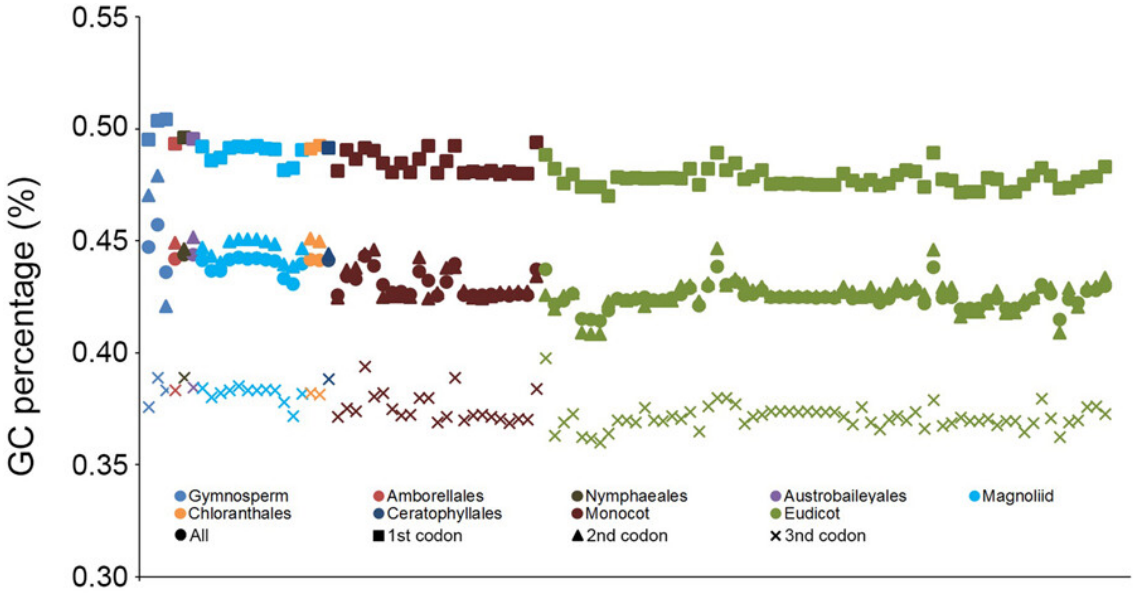

Figure 4. Nucleotide GC content of each species for the entire set of genes, and first, second and third codon positions.

Codon usage is usually correlated with GC content (Stenoien, 2005), because high GC composition results in/reflects GC-rich codons, and low GC composition reversely leads to/reflects AT-rich codons (Sharp et al., 2010). As no GC compositional heterogeneity was detected among these angiosperm lineages, biased codon-usage is very unlikely to be detected as well. Despite this, we still evaluated the codon-usage bias among the taxa sampled in this study. The resulting SCUO values generated by CodonO (Angellotti et al., 2007) showed that all of the taxa indeed had a similar pattern of codon usage, and the degree of biased usage was very low, ranging from 0.04 to 0.06 (Table S4). The dot plots of calculated codon-usage values and their correlation with GC content showed a linear pattern. Different lineages were not separated from each other in this pattern (Fig. 5).

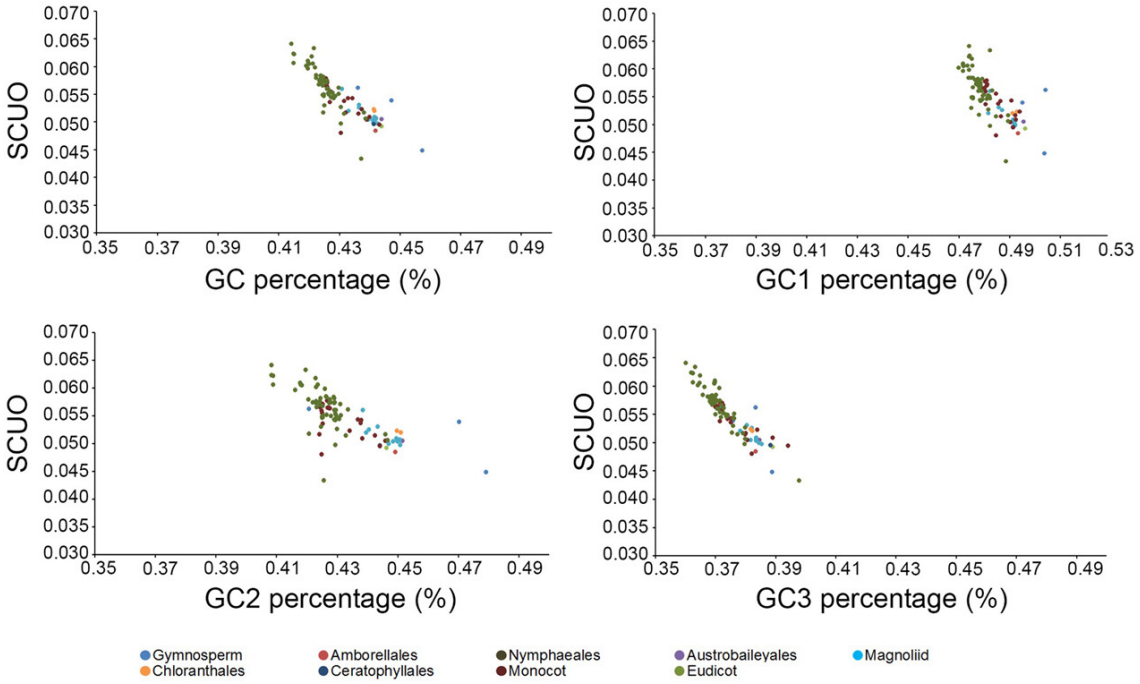

Figure 5. Codon-usage bias of 38 protein-coding genes of 107 taxa and the correlation with GC content. GC, all codon positions; GC1, first codon positions; GC2, second codon positions; GC3, third codon positions; SCUO, synonymous codon usage order.

4 Discussion

4.1 Mitochondrial genomes as an alternative source of information for plant phylogenetic studies

To date, very few studies have employed mitochondrial genes solely to address plant phylogenetic questions, but have rather integrated mt data into combined data sets with plastid or nuclear data (Qiu et al., 1999, 2005; Chaw et al., 2000). Nevertheless, mt genomic data sets comprising 40 concatenated protein-coding genes could provide additional evidence for existing hypotheses (congruent results with pt or nu data) or inform new hypotheses. Liu et al. (2019) recently used data from mt, pt, and nu genomes to infer the phylogeny of moss lineages, and comparison among the three genomic sources produced a stable framework. Other studies used mt genomes to explore the relationships among liverworts, mosses, and hornworts, and obtained largely congruent results with studies that have used pt data alone (Liu et al., 2014; Bell et al., 2020). The angiosperms had an ancient origin, but went through a relatively rapid radiation to generate the major crown lineages (e.g., Soltis et al., 2008; Amborella Genome Project, 2013; Li et al., 2019; Zhang et al., 2020), which may mean that there is poor phylogenetic signal to settle relationships among some of the major extant lineages (Townsend, 2007). This may be the reason why two recent phylogenomic studies that used the reasonably large taxon-sampling strategies (Li et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019) did not reach absolute congruence regarding the relationships of early diverging angiosperms. Mitochondrial phylogenomic studies could serve as new evidence for either of the two hypotheses regarding the relationships of five mesangiosperm lineages (Figs. 1C, 1D). Our study yielded largely congruent results with these two phylogenomic studies. Additionally, our evaluation of angiosperm mt genomes also detected low substitutional rates, homogeneous GC content, and unbiased codon-usage patterns among these early lineages, suggesting that our mt data may be minimally influenced by these possible sources of systematic error. Although horizontal gene transfer (HGT) events are often reported in plant mt genomes (e.g., Bergthorsson et al., 2003; Keeling & Palmer, 2008; Rice

et al., 2013; Davis & Xi, 2015), it is straightforward to distinguish potential HGT genes that have transferred to mt genomes by BLASTing, also these transferred genes are nearly always non-functional and degraded/pseudogenized. In our study, we have excluded the potential HGT genes in *Amborella*, and the species of *Vitis vinifera*, *Geranium maderense*, and *Viscum album* which have high rates of "rampant HGT" or highly elevated substitution rate in mt genes (Goremykin et al., 2009; Rice et al., 2013; Petersen et al., 2015). Additionally, our data set for phylogenetic analyses went through a screening procedure. Species/genes with very high rates or potential for HGT potentials were excluded, including our newly sequenced *Acorus gramineus* (Table S7). Therefore, mt genomes may be well suited for solving the phylogenetic ambiguities of deep nodes of plant phylogeny, providing additional evidence for previous studies and alternative hypotheses of the early evolution and diversification of angiosperms.

4.2 Congruence and incongruence of inferred phylogenies of early angiosperms from different studies

As systematic errors were hardly detected in our data sets, all codon positions may provide the most informative phylogenetic information compared with the other data sets, and therefore should represent the most reliable phylogeny. The nt data of all codon positions yielded a congruent phylogeny using the ML and Bayesian methods (Fig. 6): Amborellales, Nymphaeales, and Austrobaileyales were inferred as successive sister groups to the rest of angiosperms; and the mesangiosperms were composed of two major lineages: a monocots-magnoliids lineage and a lineage comprising eudicots-(Chloranthales-Ceratophyllales). Although our phylogeny of the five mesangiosperm lineages was unlike that of APG IV (Fig. S14), the major relationships inside each lineage were largely congruent, and only three orders of eudicots (Malpighiales, Rosales, and Lamiales) were recovered to different positions (Fig. S15).

A large body of literature is available that addresses the phylogenetic relationships between the major lineages of the angiosperms, but controversy still remains (e.g., Mathews & Donoghue, 1999; Qiu & Palmer, 1999; Soltis et al., 1999; Barkman et al., 2000; Graham & Olmstead, 2000; Zanis et al., 2002; Hilu et al., 2003; Leebens-Mack et al., 2005; Qiu et al., 2006a, 2006b; Jansen et al., 2007; Moore et al., 2007, 2011; Soltis et al., 2008; Endress & Doyle, 2009; Finet et al., 2010; Moore et al., 2010; Qiu et al., 2010; Lee et al., 2011; Zhang et al., 2012; Goremykin et al., 2013; Drew et al., 2014; Wickett et al., 2014; Xi et al., 2014; Zeng et al., 2014; Gitzendanner et al., 2018; Li et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019; Yang et al., 2020). With our new data based on mt genomes, several hypotheses gain new support.

Since the notion of ANA (or ANITA) was proposed, the three orders Amborellales, Nymphaeales and Austrobaileyales have been placed as successive sister groups of the rest of the angiosperms (mesangiosperms) in most studies. Our mt phylogenomic results largely agree with this framework. With respect to the order of divergence at the base of angiosperm phylogeny, our phylogenies based on the nt and the third codon positions recover Amborellales solely as the sister to all of the other angiosperms. However, the combined first and second codon positions data supported the Amborellales-Nymphaeales root for angiosperms, which is consistent with a previous result based on a smaller number (four) of mt genes from 380 angiosperm species (Qiu et al., 2010) and the ML analysis based on combined first and second codon positions from the pt 61-gene data set (Moore et al., 2007). These analyses (including ours) that support the Amborellales-Nymphaeales root used comparatively less data (only the first and second codon positions or very limited number of genes), and so this result may reflect inadequate phylogenetic signal. Here, the third codon positions alone contained 3805 parsimony-informative sites; almost equal to the combined first and second codon positions (4409 sites). As the quantity of sampled data and taxa increased, the Amborellales-root hypothesis has remained popular (Li et al., 2019; One

Thousand Plant Transcriptomes Initiative, 2019). Nevertheless, there is another explanation for the Amborellales-Nymphaeales topology, possibly attributed to the long branch connecting angiosperms to their seed-plant outgroups (Graham & Iles, 2009).

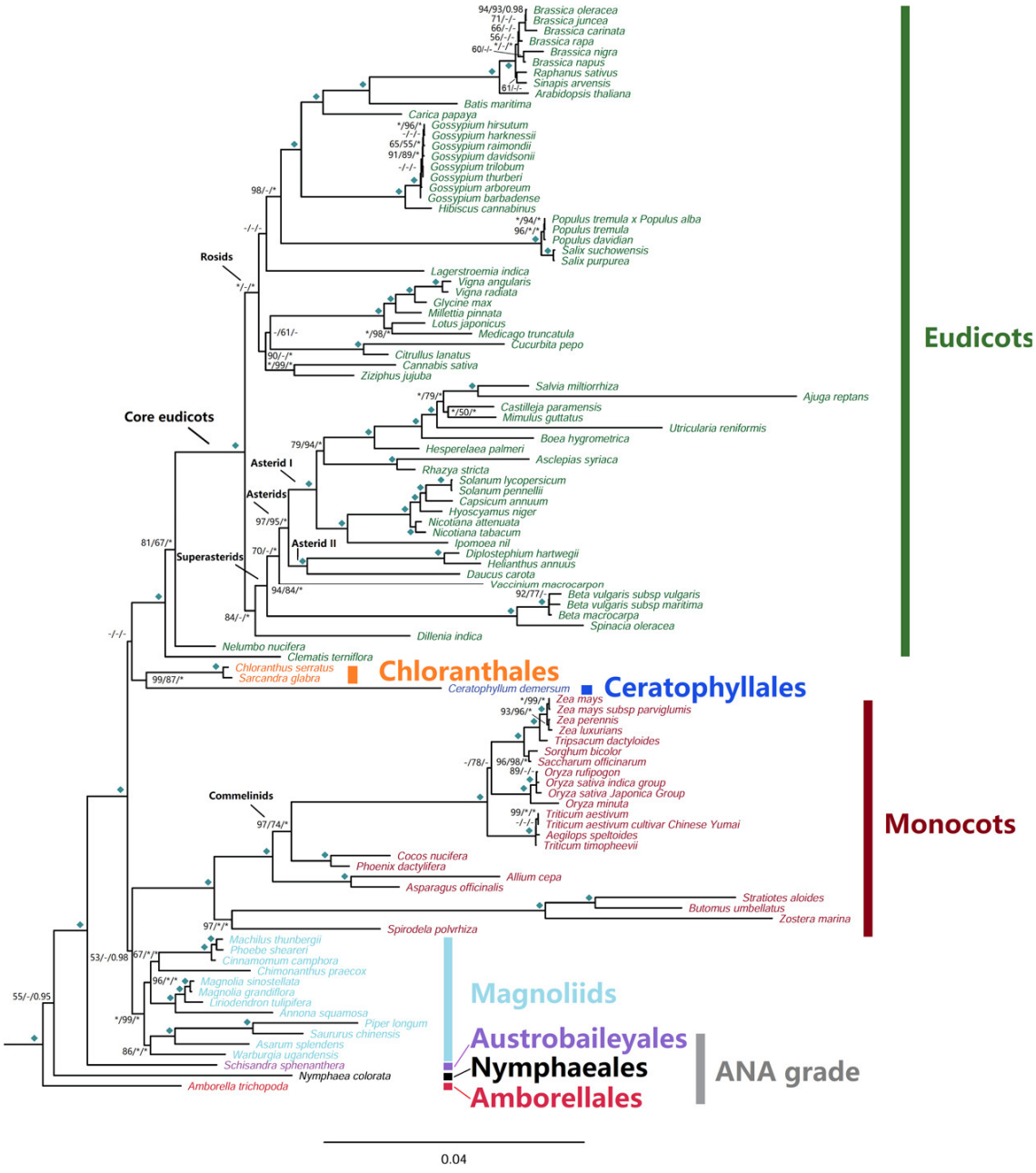

Figure 6. A detailed ML phylogram inferred from the nt data of 38 mt genes of 107 taxa. Asterisks indicate either BS of 100% or PP of 1.00. ML and MP BS support values and Bayesian PP are labeled above nodes, respectively. Diamonds indicate BS of 100% and PP of 1.00 from all ML, MP, and Bayesian analyses. Asterisks indicate either BS of 100% or PP of 1.00. BS support values <50%, Bayesian PP < 0.5, or incongruent topologies among ML, MP, and Bayesian inferences are indicated as "−." Newly sequenced taxa are highlighted in gray. ML, maximum-likelihood; MP, maximum parsimony; BS, bootstrap.

Although our mt genomic results are not in absolute accordance with any of the mentioned topologies, there is some congruence regarding the relationships of specific clades:

i. The sister-group relationship of Chloranthales to Ceratophyllales. This relationship is consistently resolved by all analyses in this study with strong support and should be taken into serious consideration at the next update of the APG system, because studies based on a diversity of data and methods all favor this proposition. A morphological study found evidence based on floral evolution to group Chloranthales and Ceratophyllales (Endress & Doyle, 2009). This relationship also has support from nu phylogenomic analyses (Zeng et al., 2014), and now has additional support from our mt genomic data.
ii. Eudicots as the sister group of the Chloranthales-Ceratophyllales lineage. This topology has poor support from our nt data, but is in accordance with the result of an earlier study based on nu genomic data, which has high resolution (Zeng et al., 2014). Pt genomic studies constantly recover eudicots as sister to Ceratophyllales with moderate support but have difficulty in stably placing Chloranthales (Graham et al., 2006; Moore et al., 2007, 2010; Ruhfel et al., 2014; Givnish et al., 2018; Li et al., 2019), but the slowly evolving inverted repeats supported our topology (Moore et al., 2011). Two recent studies on nu data also supported a sister-group relationship of eudicots and Ceratophyllales, one based on a dense taxon-sampling strategy (One Thousand Plant Transcriptomes Initiative, 2019) and the other using an advanced phylogenetic signal selection strategy (Yang et al., 2020); nevertheless, the results of these two studies differ with regards to the other three lineages. Regardless, the argument concerning the sister-group placement of eudicots seems likely to continue.
iii. The sister-group relationship between magnoliids and monocots. Our results (nt and third codon positions) suggested that magnoliids might be related to monocots, but this topology has rarely been recovered in previous studies (Qiu et al., 2005; Endress & Doyle, 2009). Other studies mainly prefer the successive divergences of the two lineages: nu data favor a prior divergence of monocots, followed by magnoliids (Wickett et al., 2014; Zeng et al., 2014; Yang et al., 2020); whereas pt phylogenomic data incline toward the opposite (Moore et al., 2011; Li et al., 2019). Earlier studies using pt genomes, however, all nested magnoliids and monocots in different clades (Jansen et al., 2007; Moore et al., 2007, 2010). While our mt data support the hypothesis by clustering magnoliids with monocots (Qiu et al., 2005; Endress & Doyle, 2009), this is undoubtedly another unresolved ambiguity worthy of continued exploration.

4.3 Possible causes for the incongruence of early angiosperm phylogenies from different studies

In all of the phylogenetic studies of mesangiosperms, without exception, extremely short branches leading to the five major groups have been observed (e.g., Qiu et al., 2006b; Jansen et al., 2007; Qiu et al., 2010; Moore et al., 2011; Wickett et al., 2014; Zeng et al., 2014; Li et al., 2019; Yang et al., 2020), reflecting the rapid radiation process during the early stage of evolution in mesangiosperms. Phylogenomics should be an effective strategy to solve the ambiguities of rapid radiations (e.g., Xi et al., 2012; Dong et al., 2013; Longo et al., 2017; Pouchon et al., 2018), because abundant data should greatly increase the number of informative sites in analyses and reduce sampling error. However, systematic errors such as substitutional saturation can also accumulate along with the increase of informative sites (Roger & Hug, 2006). Although our assessment of systematic errors discovered angiosperm mt genes may not be substantially affected by such factors, systematic errors might be one notable reason for the inconsistent results obtained by other studies using pt or nu data (e.g., Jansen et al., 2007; Moore et al., 2007; Graham & Iles, 2009; Moore et al., 2010; Wickett et al., 2014; Zeng et al., 2014). Therefore, we assessed some major sources of systematic error

among pt, mt, and nu genes using 14 taxa with data available for all of three genomic compartments.

Although the most conserved low copy genes in the nucleus (59 genes from Zeng et al. (2014)) are selected for phylogenetic analysis, it is noteworthy that they are still much more substitutionally saturated than their pt (79 genes) or mt (38 genes) counterparts (Fig. 7). Pt genes are most subject to codon-usage bias probably because they possess the lowest GC contents among the three genomes. The among-lineage GC compositional heterogeneity could not be evaluated here because too few qualified taxa were involved, but as it is associated with codon-usage bias, pt genes are likely to have the highest among-lineage GC compositional heterogeneity (Fig. 8). Nu and pt genes may be more severely affected by these factors as they on average evolve faster than mt genes and therefore may be likely to generate artifactual results when taxon sampling is not dense enough to overcome potential substitutional saturation. Therefore, although the advancement of sequencing technology facilitates phylogenetic studies with broadened sampling of molecular markers and taxa, we should pay greater attention to the proper utilization of genomic-scale data. Phylogenetic noise due to homoplasy should be taken into account when reconstructing phylogenies. The evaluation and removal of data susceptible to systematic error may be a useful procedure to find out reasons for incongruent results by different methods, and help to obtain more reliable phylogenies. Conversely, although mt genes are likely to suffer the least from systematic errors among the three genomic compartments, they have the fewest and generally most slowly evolving genes for analyses, and therefore, the corresponding analyses may be more prone to sampling error. Our incongruent results from different data sets might be a case in point. Additionally, the choice of methods matters. In this study, the two model-based methods may be better able to deal with problems related to substitutional saturation, substitutional differences, GC composition heterogeneity on phylogenetic analyses.

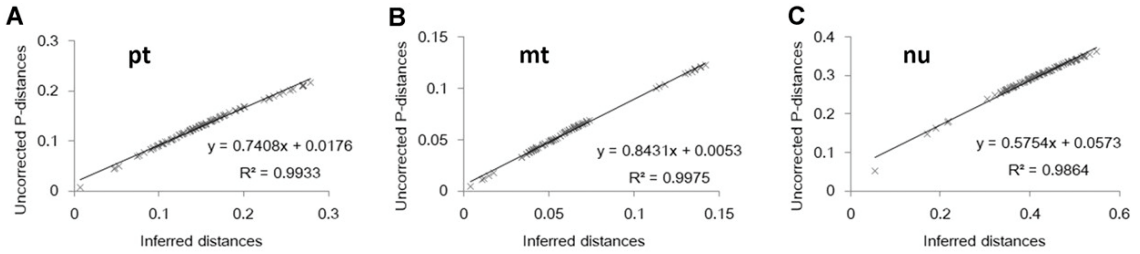

Figure 7. Saturation level of all nucleotide data of **A**, pt, **B**, mt, and **C**, nu genes from 14 taxa. For taxa names, see Table S5. The uncorrected P-distances were plotted against the inferred distances. The *x*-axis corresponds to the distance inferred by the ML method using the Tamura-Nei model (Tamura & Nei, 1993), and the *y*-axis corresponds to the uncorrected distance observed for the same taxa pair. ML, maximum-likelihood.

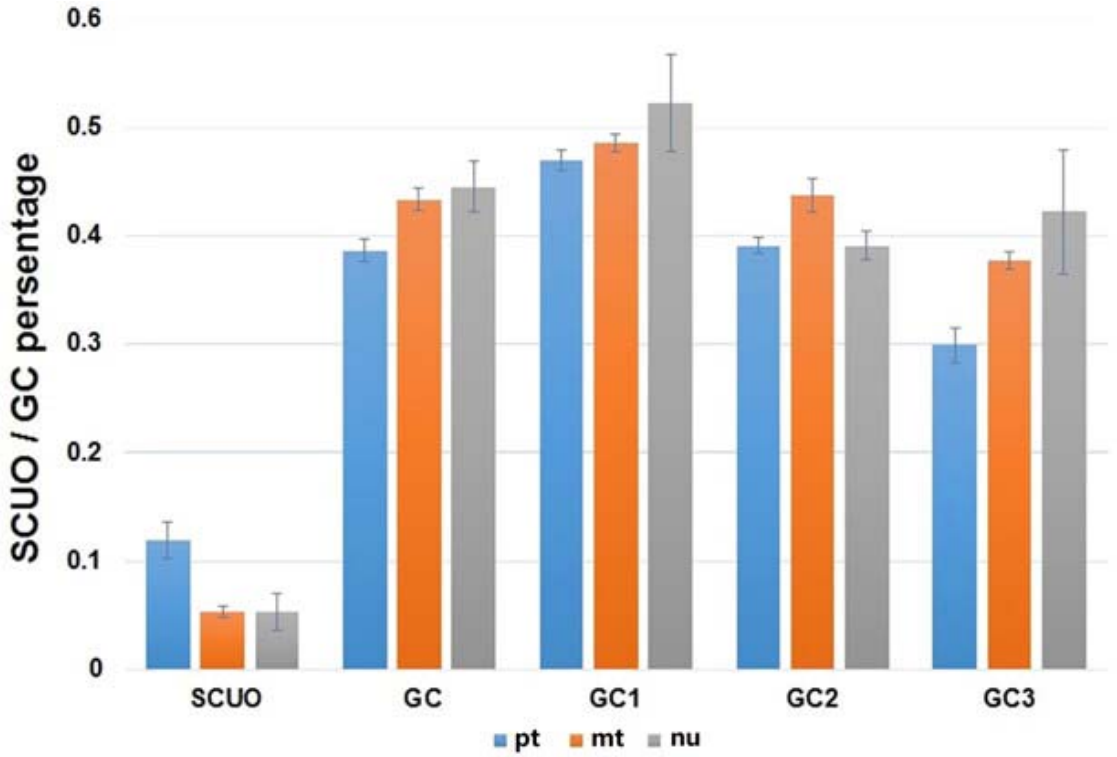

Figure 8. Codon-usage bias and nucleotide GC content among plastid, mitochondrial, and nuclear genes of 14 taxa. For taxa names, see Table S6.

Hybridization events are often responsible for incongruent results inferred from nu and organellar genes. The discrepancy of inheritance modes between nuclear and organellar genomes (biparental vs. uniparental) could lead to such incongruence (e.g., Wendel et al., 1991; Palacios et al., 2000; Zhang et al., 2015; Li et al., 2016; Morales-Briones et al., 2018; Tang et al., 2018). Incomplete lineage sorting (ILS) renders phylogenetic inference more complex, especially for lineages that arose through rapid radiations (e.g., Alexander et al., 2017; Alda et al., 2019; Smith & Hahn, 2020). As mtDNA is inherited uniparentally as one molecule, any among-gene conflict cannot be explained by hybridization or ILS, but may reflect the nature of evolution among mt genes. For those branches with high bootstrap support but low gene/site concordance values (Fig. S16), we may infer that conflicts in phylogenetic inference resulted from the low support of mt genes that resulted from the low substitution rates of plant mtDNAs and short divergence time of these lineages. As mt genes are inherited as a whole and are not subjected to recombination among diploid copies as with nuclear loci, and potential horizontally transferred genes (or very rapidly evolving genes) were identified and excluded before analysis from the data set, the low gCF values at many nodes might therefore suggest weak phylogenetic signal in individual mt loci rather than discordance among individual gene trees, which is consistent with the low substitution rate of the mt genes. For instance, the low gCF value is especially significant for some nodes with very short branches, not only between closely related taxa (e.g., within the genus *Gossypium* and *Brassica*, respectively), but also distantly related lineages (e.g., the node (18.9%) leading to Piperales and Canellales, the node (27.9%) leading to Lauraceae and Calycanthaceae, and the node (25.7%) leading to eudicots and Chloranthales-Ceratophyllales lineage), which might suggest node instability due to a small number of highly influential sites that strongly impact the reference ML topology (whereas these sites are weighed equally in MP analyses). Hence the low sCF values may indicate strong deviation of MP analyses from the reference topology. These unstable nodes should be investigated further using additional lines of evidence.

Acknowledgements

We are grateful to the editor and two anonymous reviewers for their constructive comments to improve our manuscript. We thank Ping Wu for help with statistical analysis, Xianxian Yu for sharing the data of *Schisandra sphenanthera*, and Xiwu Qi for his assistance in extracting protein-coding sequences by PERL scripts. This work was supported by grants from the Jiangsu Key Laboratory for the Research and Uti1ization of Plant Resources (Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, KSPKLB201835 to Jia-Yu Xue), the Shenzhen Key Laboratory of Southern Subtropical Plant Diversity (SLPD-2018-3 to Jia-Yu Xue), the National Natural Science Foundation of China (31470314 to Yang Liu, 31471200 to Qiang Wang and 31470448 to Xiao-Qin Sun), the Public Welfare Forestry Industry Project of State Forestry Administration, China (201504322 to Shou-Zhou Zhang), a postdoctoral fellowship from the Special Research Fund of Ghent University (BOFPDO2018001701 to Zhen Li), the Strategic Priority Research Program of Chinese Academy of Sciences (XDA13020603 to Hongfeng Chen), and the Shenzhen Fairy Lake Botanical Garden Science Foundation (FLSF-2019-02 to Yang Liu). The computational resources (Stevin Supercomputer Infrastructure) and services used for PhyloBayes-MPI were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government–department EWI.

Accession Numbers

Raw sequence reads of 14 species have been deposited in the European Nucleotide Archive under BioProject ID PRJEB33250, three species have been deposited in National Center for Biotechnology Information under BioProject ID PRJNA553292, and the complete mitochondrial genome of *Schisandra sphenanthera* has been deposited in National Center for Biotechnology Information under accession number MH748549.

Author Contributions

JYX, YL, QW and YYH conceived the study. JYX, GCZ, MQW, MC, YMZ and SZZ collected samples and conducted experiments. QW, YL and SD assembled the mitochondrial genomes. JYX, ZQS, YL, WW and ZL analyzed data. JYX, YL, ZL and YVP wrote the manuscript. ZQS, XQS, WW, FC, LZ and HFC participated in the revision of the manuscript. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* 38: W7–W13.
- Ahmed Z, Fu YB. 2015. An improved method with a wider applicability to isolate plant mitochondria for mtDNA extraction. *Plant Methods* 11: 56.
- Alda F, Tagliacollo VA, Bernt MJ, Waltz BT, Ludt WB, Faircloth BC, Alfaro ME, Albert JS, Chakrabarty P. 2019. Resolving deep nodes in an ancient radiation of neotropical fishes in the presence of conflicting signals from incomplete lineage sorting. *Systematic Biology* 68: 573–593.
- Alexander AM, Su YC, Oliveros CH, Olson KV, Travers SL, Brown RM. 2017. Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting

to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution* 71: 475–488.
- Amborella Genome Project. 2013. The Amborella genome and the evolution of flowering plants. *Science* 342: 1241089.
- Angellotti MC, Bhuiyan SB, Chen G, Wan XF. 2007. CodonO: Codon usage bias analysis within and across genomes. *Nucleic Acids Research* 35: W132–W136.
- APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.
- APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.
- Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, Wolfe AD, dePamphilis CW. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* 97: 13166–13171.
- Beckert S, Steinhauser S, Muhle H, Knoop V. 1999. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. *Plant Systematics and Evolution* 218: 179–192.
- Bell D, Lin Q, Gerelle WK, Joya S, Chang Y, Taylor ZN, Rothfels CJ, Larsson A, Villarreal JC, Li FW, Pokorny L, Szovenyi P, Crandall-Stotler B, DeGironimo L, Floyd SK, Beerling DJ, Deyholos MK, von Konrat M, Ellis S, Shaw AJ, Chen T, Wong GK, Stevenson DW, Palmer JD, Graham SW. 2020. Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution. *American Journal of Botany* 107: 91–115.
- Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424: 197–201.
- Bowe LM, Coat G, dePamphilis CW. 2000. Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proceedings of the National Academy of Sciences of the United States of America* 97: 4092–4097.
- Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current Biology* 22: 248–252.
- Cao Y, Janke A, Waddell PJ, Westerman M, Takenaka O, Murata S, Okada N, Paabo S, Hasegawa M. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of Eutherian orders. *Journal of Molecular Evolution* 47: 307–322.
- Chaw SM, Parkinson CL, Cheng YC, Vincent TM, Palmer JD. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proceedings of the National Academy of Sciences of the United States of America* 97: 4086–4091.
- Chaw SM, Shih ACC, Wang D, Wu YW, Liu SM, Chou TY. 2008. The mitochondrial genome of the gymnosperm Cycas taitungensis contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Molecular Biology and Evolution* 25: 603–615.
- Coiro M, Doyle JA, Hilton J. 2019. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist* 223: 83–99.
- Cronquist A. 1981. An integrated system of classification of flowering plants. New York: Columbia University Press.
- Cui LY, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

- Davis CC, Xi Z. 2015. Horizontal gene transfer in parasitic plants. *Current Opinion in Plant Biology* 26: 14–19.
- Dilcher DL, Wang HS. 2009. An early Cretaceous fruit with affinities to Ceratophyllaceae. *American Journal of Botany* 96: 2256–2269.
- Dong WP, Xu C, Cheng T, Lin K, Zhou SL. 2013. Sequencing angiosperm plastid genomes made easy: A complete set of universal primers and a case study on the phylogeny of saxifragales. *Genome Biology and Evolution* 5: 989–997.
- Drew BT, Ruhfel BR, Smith SA, Moore MJ, Briggs BG, Gitzendanner MA, Soltis PS, Soltis DE. 2014. Another look at the root of the angiosperms reveals a familiar tale. *Systematic Biology* 63: 368–382.
- Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* 49: 827–831.
- Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, Sytsma KJ, Koch MA, Al-Shehbaz IA, Pires JC. 2018. Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *American Journal of Botany* 105: 463–469.
- Eklund H, Doyle JA, Herendeen PS. 2004. Morphological phylogenetic analysis of living and fossil Chloranthaceae. *International Journal of Plant Sciences* 165: 107–151.
- Endress PK, Doyle JA. 2009. Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany* 96: 22–66.
- Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.
- Finet C, Timme RE, Delwiche Cf, Marleta F. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology* 20: 2217–2222.
- Forrest LL, Davis EC, Long DG, Crandall-Stotler BJ, Clark A, Hollingsworth ML. 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): Multiple taxa, genomes and analyses. *Bryologist* 109: 303–334.
- Forterre P, Philippe H. 1999. Where is the root or the universal tree of life? *BioEssays* 21: 871–879.
- Freeling M. 2009. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60: 433–453.
- Friedman WE. 2009. The meaning of Darwin's 'abominable mystery'. *American Journal of Botany* 96: 5–21.
- Friis EM, Pedersen KR, Crane PR. 2006. Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. *Palaeogeography, Palaeoclimatology, Palaeoecology* 232: 251–293.
- Friis EM, Pedersen KR, Crane PR. 2010. Diversity in obscurity: Fossil flowers and the early history of angiosperms. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 369–382.
- Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* 105: 291–301.
- Givnish TJ, Zuluaga A, Spalink D, Soto Gomez M, Lam VKY, Saarela JM, Sass C, Iles WJD, de Sousa DJL, Leebens-Mack J, Chris Pires J, Zomlefer WB, Gandolfo MA, Davis JI, Stevenson DW, dePamphilis C, Specht CD, Graham SW, Barrett Cf, Ane C. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105: 1888–1910.
- Gomez MS, Lin Q, da Silva Leal E, Gallaher TJ, Scherberich D, Mennes CB, Smith SY, Graham SW. 2020. A bi-organellar phylogenomic study of Pandanales: inference

of higher-order relationships and unusual rate-variation patterns. *Cladistics* 36: 481–504.
- Goremykin VV, Nikiforova SV, Biggs PJ, Zhong BJ, Delange P, Martin W, Woetzel S, Atherton RA, Mclenachan PA, Lockhart PJ. 2013. The evolutionary root of flowering plants. *Systematic Biology* 62: 50–61.
- Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26: 99–110.
- Graham SW, Iles WJ. 2009. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. *American Journal of Botany* 96: 216–227.
- Graham SW, Olmstead RG. 2000. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Current Genetics* 37: 183–188.
- Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, Saarela JM, Horne EFC, Smith SY, Young WA, O'Brien HE, Brown VL, Pires JC, Olmstead RG, Chase MW, Rai HS. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso: A Journal of Systematic and Evolutionary Botany* 22: 3–21.
- Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annual Review of Plant Biology* 68: 225–252.
- Gui ST, Wu ZH, Zhang HY, Zheng YZ, Zhu ZX, Liang DQ, Ding Y. 2016. The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. *Scientific Reports* 6: 30158.
- Guo WH, Grewe F, Fan WS, Young GJ, Knoop V, Palmer JD, Mower JP. 2016. Ginkgo and Welwitschia mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Molecular Biology and Evolution* 33: 1448–1460.
- Herendeen PS, Friis EM, Pedersen KR, Crane PR. 2017. Palaeobotanical redux: Revisiting the age of the angiosperms. *Nature Plants* 3: 17015.
- Hilu KW, Black CM, Oza D. 2014. Impact of gene molecular evolution on phylogenetic reconstruction: A case study in the rosids (Superorder Rosanae, Angiosperms). *PLoS One* 9: e99725.
- Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW. 2003. Angiosperm phylogeny based on matK sequence information. *American Journal of Botany* 90: 1758–1776.
- Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19369–19374.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* 33: 511–518.
- Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9: 605–618.
- Knoop V. 2004. The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. *Current Genetics* 46: 123–139.
- Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.
- Lartillot, N, Lepage, T, Blanquart, S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25: 2286–2288.

- Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR, Martienssen RA, Coruzzi G, DeSalle R. 2011. A functional phylogenomic view of the seed plants. *PLoS Genetics* 7: e1002411.
- Leebens-Mack J, Raubeson LA, Cui LY, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22: 1948–1963.
- Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research* 26: 1–11.
- Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, Wang H, van der Bank M, Zhang SD, Wang QF, Wang J, Zhang ZR, Fu CN, Yang J, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li DZ. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.
- Liu Y, Cox CJ, Wang W, Goffinet B. 2014. Mitochondrial phylogenomics of early land plants: Mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* 63: 862–878.
- Liu Y, Johnson MG, Cox CJ, Medina R, Devos N, Vanderpoorten A, Hedenas L, Bell NE, Shevock JR, Aguero B, Quandt D, Wickett NJ, Shaw AJ, Goffinet B. 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* 10: 1485.
- Longo SJ, Faircloth BC, Meyer A, Westneat MW, Alfaro ME, Wainwright PC. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. *Molecular Phylogenetics and Evolution* 113: 33–48.
- Magallon S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437–453.
- Mathews S, Donoghue MJ. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.
- Mennes CB, Smets EF, Moses SN, Merckx VS. 2013. New insights in the long-debated evolutionary history of Triuridaceae (Pandanales). *Molecular Phylogenetics and Evolution* 69: 994–1004.
- Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* 37: 2727–2733.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.
- Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19363–19368.
- Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, Vandeventer A, Horn JW, Dhingra A, Brockington SF, Latvis M, Ramdial J, Alexandre R, Piedrahita A, Xi ZX, Davis CC, Soltis PS, Soltis DE. 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species: Insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences* 172: 541–558.
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 107: 4623–4628.
- Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218: 1668–1684.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* 115: E2274–E2283.

- Ness RW, Graham SW, Barrett SCH. 2011. Reconciling gene and genome duplication events: Using multiple nuclear gene families to infer the phylogeny of the aquatic plant family Pontederiaceae. *Molecular Biology and Evolution* 28: 3009–3018.
- Nickrent DL, Parkinson CL, Palmer JD, Duff RJ. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* 17: 1885–1895.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.
- Palacios C, Rossello JA, Gonzalez-Candelas F. 2000. Study of the evolutionary relationships among *Limonium* species (Plumbaginaceae) using nuclear and cytoplasmic molecular markers. *Molecular Phylogenetics and Evolution* 14: 232–249.
- Palmer JD, Herbon LA. 1988. Plant mitochondrial-DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28: 87–97.
- Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, Jansen RK. 2015. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New Phytologist* 208: 570–583.
- Petersen G, Cuenca A, Moller IM, Seberg O. 2015. Massive gene loss in mistletoe (*Viscum*, Viscaceae) mitochondria. *Scientific Reports* 5: 17588.
- Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution* 21: 1455–1458.
- Pouchon C, Fernandez A, Nassar JM, Boyer F, Aubert S, Lavergne S, Mavarez J. 2018. Phylogenomic analysis of the explosive adaptive radiation of the espeletia complex (Asteraceae) in the tropical Andes. *Systematic Biology* 67: 1041–1060.
- Proost S, Pattyn P, Gerats T, Van de Peer Y. 2011. Journey through the past: 150 million years of plant genome evolution. *Plant Journal* 66: 58–65.
- Qiu YL, Dombrovska O, Lee J, Li LB, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW, Renner SS, Soltis DE, Soltis PS, Zanis MJ, Cannone JJ, Gutell RR, Powell M, Savolainen V, Chatrou LW, Chase MW. 2005. Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 166: 815–842.
- Qiu YL, Lee JH, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen ZD, Savolainen V, Chase MW. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402: 404–407.
- Qiu YL, Li LB, Hendry TA, Li RQ, Taylor DW, Issa MJ, Ronen AJ, Vekaria ML, White AM. 2006a. Reconstructing the basal angiosperm phylogeny: evaluating information content of mitochondrial genes. *Taxon* 55: 837–856.
- Qiu YL, Li LB, Wang B, Chen ZD, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC. 2006b. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America* 103: 15511–15516.
- Qiu YL, Li LB, Wang B, Xue JY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391–425.
- Qiu YL, Palmer JD. 1999. Phylogeny of early land plants: Insights from genes and genomes. *Trends in Plant Science* 4: 26–30.
- Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. *Science* 342: 1468–1473.
- Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: Ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* 11: 29.

- Rodrigo AG, Kelly-Borges M, Bergquist PR, Bergquist PL. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *New Zealand Journal of Botany* 31: 257–268.
- Roger AJ, Hug LA. 2006. The origin and diversification of eukaryotes: Problems with molecular phylogenetics and molecular clock estimation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361: 1039–1054.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.
- Sharp, PM, Emery, LR, Zeng, K. 2010. Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1203–1212.
- Simmons MP. 2017. Mutually exclusive phylogenomic inferences at the root of the angiosperms: Amborella is supported as sister and observed variability is biased. *Cladistics* 33: 488–512.
- Sloan DB. 2013. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytologist* 200: 978–985.
- Smith ML, Hahn MW. 2020. New approaches for inferring phylogenies in the presence of paralogs. *Trends in Genetics* . doi: 10.1016/j.tig.2020.08.012.
- Soltis DE, Bell CD, Kim S, Soltis PS. 2008. Origin and early evolution of angiosperms. *Annals of the New York Academy of Sciences* 1133: 3–25.
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Rushworth CA, Gitzendanner MA, Sytsma KJ, Qiu YL, Hilu KW, Davis CC, Sanderson MJ, Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.
- Soltis PS, Soltis DE, Chase MW. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402–404.
- Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
- Stenøien HK. 2005. Adaptive basis of codon usage in the haploid moss *Physcomitrella patens*. *Heredity* 94: 87–93.
- Swofford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), v4.0b10. Sunderland, MA: Sinauer Associates.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56: 564–577.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.
- Tang Q, Edwards SV, Rheindt FE. 2018. Rapid diversification and hybridization have shaped the dynamic history of the genus *Elaenia*. *Molecular Phylogenetics and Evolution* 127: 522–533.
- Tasdighian S, Van Bel M, Li Z, Van de Peer Y, Carretero-Paulet L, Maere S. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *The Plant Cell* 29: 2766–2785.
- Townsend JP. 2007. Profiling phylogenetic informativeness. *Systematic Biology* 56: 222–231.

- Unseld M, Marienfeld JR, Brandt P, Brennicke A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15: 57–61.
- Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K. 2009. The flowering world: a tale of duplications. *Trends in Plant Science* 14: 680–688.
- Vanneste K, Baele G, Maere S, Van de Peer Y. 2014. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Research* 24: 1334–1347.
- Wang WQ, Wu YR, Messing J. 2012. The mitochondrial genome of an aquatic plant, *Spirodela polyrhiza*. *PLoS One* 7: 46747.
- Wang XY, Shi XL, Hao BL, Ge S, Luo JC. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytologist* 165: 937–946.
- Wendel JF, Stewart JM, Rettig JH. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of Gossypium. *Evolution* 45: 694–711.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfel BR, Wafula E, Der JP, Graham SW, Mathews S, Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pokorny L, Shaw AJ, DeGironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis CW, Chen T, Deyholos MK, Baucom RS, Kutchan TM, Augustin MM, Wang J, Zhang Y, Tian ZJ, Yan ZX, Wu XL, Sun X, Wong GKS, Leebens-Mack J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* 111: E4859–E4868.
- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* 84: 9054–9058.
- Xi ZX, Liu L, Rest JS, Davis CC. 2014. Coalescent versus concatenation methods and the placement of Amborella as sister to water lilies. *Systematic Biology* 63: 919–932.
- Xi ZX, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, Mathews S, Davis CC. 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America* 109: 17519–17524.
- Yang L, Su D, Chang X, Foster CSP, Sun L, Huang CH, Zhou X, Zeng L, Ma H, Zhong B. 2020. Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant Communications* 1: 100027.
- Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences of the United States of America* 99: 6848–6853.
- Zeng LP, Zhang Q, Sun RR, Kong HZ, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* 5: 4956.
- Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X, Song A, Chen J, Guo W, Wang Z, Zhuang Y, Wang H, Chen X, Hu J, Liu Y, Qin Y, Wang K, Dong S, Liu Y, Zhang S, Yu X, Wu Q, Wang L, Yan X, Jiao Y, Kong H, Zhou X, Yu C, Chen Y, Li F, Wang J, Chen W, Chen X, Jia Q, Zhang C, Jiang Y, Zhang W, Liu G, Fu J, Chen F, Ma H, Van de Peer Y, Tang H. 2020. The water lily genome and the early evolution of flowering plants. *Nature* 577: 79–84.
- Zhang N, Zeng LP, Shan HY, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* 195: 923–937.
- Zhang Q, Feild TS, Antonelli A. 2015. Assessing the impact of phylogenetic incongruence on taxonomy, floral evolution, biogeographical history, and phylgenetic diversity. *American Journal of Botany* 102: 566–580.

- Zhong BJ, Betancur-R R. 2017. Expanded taxonomic sampling coupled with gene genealogy interrogation provides unambiguous resolution for the evolutionary root of angiosperms. *Genome Biology and Evolution* 9: 3154–3161.
- Zhu XY, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD. 2007. Mitochondrial matR sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology* 7: 217.
- Zwickl DJ, Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.