Phylogenomic and structural analyses of 18 complete plastomes
across all families of early-diverging eudicots, including an
angiosperm-wide analysis of IR gene content evolution

Yanxia Sun^a, Michael J. Moore^b, Shoujun Zhang^a, Pamela S. Soltis^c, Douglas E.
Soltis^c, Tingting Zhao^a, Aiping Meng^a, Xiaodong Li^a, Jianqiang Li^{a,*}, Hengchang
Wang^{a,*}

^a*Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan
Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, China*

^b*Department of Biology, Oberlin College, Oberlin, Ohio, USA*

^c*Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA*

* Corresponding author.

E-mail address: lijq@wbgcas.cn (Jianqiang Li); hcwang@wbgcas.cn (Hengchang
Wang).

ABSTRACT

The grade of early-diverging eudicots includes five major lineages: Ranunculales, Trochodendrales, Buxales, Proteales and Sabiaceae. To examine the evolution of plastome structure in early-diverging eudicots, we determined the complete plastome sequences of eight previously unsequenced early-diverging eudicot taxa, *Pachysandra terminalis* (Buxaceae), *Meliosma* aff. *cuneifolia* (Sabiaceae), *Sabia yunnanensis* (Sabiaceae), *Epimedium sagittatum* (Berberidaceae), *Euptelea pleiosperma* (Eupteleaceae), *Akebia trifoliata* (Lardizabalaceae), *Stephania japonica* (Menispermaceae) and *Papaver somniferum* (Papaveraceae), and compared them to previously published plastomes of the early-diverging eudicots *Buxus*, *Tetracentron*, *Trochodendron*, *Nelumbo*, *Platanus*, *Nandina*, *Megaleranthus*, *Ranunculus*, *Mahonia* and *Macadamia*. All of the newly sequenced plastomes share the same 79 protein-coding genes, 4 rRNA genes, and 30 tRNA genes, except for that of *Epimedium*, in which *infA* is pseudogenized and *clpP* is highly divergent and possibly a pseudogene. The boundaries of the plastid Inverted Repeat (IR) were found to vary significantly across early-diverging eudicots; IRs ranged from 24.3-36.4 kb in length and contained from 18 to 33 genes. Based on gene content, the IR was classified into six types, with shifts among types characterized by high levels of homoplasy. Reconstruction of ancestral IR gene content suggested that 18 genes were likely present in the IR region of the ancestor of eudicots. Maximum likelihood phylogenetic analysis of a 79-gene, 97-taxon data set that included all available

40 early-diverging eudicots and representative sampling of remaining angiosperm
41 diversity largely agreed with previous estimates of early-diverging eudicot
42 relationships, but resolved Trochodendrales rather than Buxales as sister to
43 *Gunneridae*, albeit with relatively weak bootstrap support, conflicting with what has
44 been found for these three clades in most previous analyses. In addition, Proteales was
45 resolved as sister to Sabiaceae with the highest support (bootstrap > 90%) yet
46 observed in plastome-scale phylogenetic analyses.

47

1. Introduction

Over the past 30 years, the plastome has served as a valuable resource for phylogenetics, population genetics and species identification. During the past decade, rapid improvements in sequencing technology have led to the routine sequencing of complete plastomes, allowing for increasingly detailed understanding of the molecular and structural evolution of plastomes as well as increased resolution of deep-level relationships (Straub et al., 2012; Ruhfel et al., 2014; Ruhlman and Jansen, 2014). For example, it is now well known that the plastomes of most angiosperms are highly conserved in structural organization, gene arrangement, and gene content (Palmer, 1991; Raubeson and Jansen, 2005; Wicke et al., 2011; Jansen and Ruhlman, 2012; Ruhlman and Jansen, 2014).

The hallmark of the plastome is its quadripartite structure, with two inverted repeat regions (IR_A and IR_B) separating the small and large single-copy regions (SSC and LSC) (Palmer and Stein, 1986). The IRs of angiosperm plastomes can fluctuate greatly in size, and the positions of all four junctions (IR_A/SSC, IR_A/LSC, IR_B/SSC and IR_B/LSC) can vary, even among closely related species (Goulding et al., 1996; Plunkett and Downie, 2000; Downie and Jansen, 2015). This variation in IR size is often the largest contributor to variation in overall plastome size and gene number. For example, among early-diverging eudicots, the IR gene number in most plastomes is 18 or 19 (Moore et al., 2006; Hansen et al., 2007; Raubeson et al., 2007; Kim et al., 2009; Nock et al., 2014). However, the IR of *Mahonia bealei* (Berberidaceae) has

expanded to 33 genes (Ma et al., 2013), while those of *Tetracentron sinense* and *Trochodendron aralioides* (Trochodendraceae) contain 24 genes (Sun et al., 2013).

Although plastome sequences have played an important role in resolving relationships among flowering plants (*e.g.* Jansen et al., 2007; Moore et al., 2007, 2010; Xi et al., 2012), problem areas remain. For example, some of the earliest divergences among *Eudicotyledoneae* [all italicized clade names in the current paper follow Cantino et al. (2007); otherwise, APG III (Angiosperm Phylogeny Group, 2009) ordinal and family names are employed] have not been resolved with strong support, most likely due to the rapid diversification of this grade (Bell et al., 2010; Nock et al., 2014). Early-diverging eudicots consist of five major lineages—Ranunculales, Sabiaceae, Proteales, Trochodendrales and Buxales (*sensu* APG III)—which together contain 12 families (APG III): Berberidaceae, Ranunculaceae, Eupteleaceae, Lardizabalaceae, Menispermaceae, Papaveraceae, Sabiaceae, Platanaceae, Nelumbonaceae, Proteaceae, Trochodendraceae, and Buxaceae. Among the five major lineages, the sister relationship between Sabiaceae and Proteales is only moderately supported in previous studies (*e.g.* Moore et al., 2010; Nock et al., 2014; Ruhfel et al., 2014; Wu et al., 2014), and the relationships among Buxales, Trochodendrales, and *Gunneridae* (*i.e.* all remaining eudicots) have differed in recent analyses (*e.g.* Jansen et al., 2007; Moore et al., 2011; Soltis et al., 2011).

Increased taxon sampling, particularly for taxa representing deep-level divergences, may provide resolution to such problem areas (Hillis et al., 2003; Leebens-Mack et al., 2005). To date, the plastomes of 10 early-diverging eudicot taxa

have been reported (Table 1; Moore et al., 2006; Hansen et al., 2007; Raubeson et al., 2007; Kim et al., 2009; Ma et al., 2013; Sun et al., 2013; Nock et al., 2014; Wu et al., 2014). However, no complete plastomes have been reported from several key families, including Eupteleaceae, Lardizabalaceae, Menispermaceae, Papaveraceae, and Sabiaceae, and for many of the other families, only one complete plastome exists.

To characterize plastome structural evolution and to better understand relationships among early-diverging eudicots, we sequenced the complete plastomes of eight previously unsampled early-diverging eudicot taxa (Table 1) representing seven families, including the first reports of complete plastomes for Eupteleaceae, Lardizabalaceae, Menispermaceae, Papaveraceae, and Sabiaceae. We then compared these to all ten previously published early-diverging eudicot plastomes (Table 1) and included all 18 early-diverging eudicot plastomes in a larger phylogenetic analysis including representatives of all major lineages of angiosperms. Analyses of these plastomes (1) indicate that early-diverging eudicots retain the overall ancestral plastome structure of angiosperms but possess highly variable IR boundaries, and (2) reveal high levels of support for almost all relationships among orders and families of early-diverging eudicots.

2. Materials and methods

2.1. Taxon sampling, chloroplast DNA isolation, sequencing and de novo assembly

The collecting and GenBank accession information for taxa sampled in the present study is listed in Table 1. For all species except *Meliosma*, high-quality chloroplast DNA was obtained following the plastid DNA extraction method of Shi et al. (2012). Approximately 30 g of fresh, young leaf tissue was used for each species. For all species except *Meliosma*, a 500-bp DNA TruSeq Illumina (Illumina Inc., San Diego, CA, USA) sequencing library was constructed for each plastome at the Beijing Genomics Institute (BGI) in Wuhan, Hubei, China, using 2.5–5 ng sonicated chloroplast DNA as input. Libraries were quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and by real-time quantitative PCR. Libraries were multiplexed and paired-end sequenced on an Illumina Hiseq 2000 platform at BGI. Raw sequence reads were subsequently filtered for high quality reads by removing duplicate reads, as well as adapter-contaminated reads, and reads with more than five Ns. Filtered reads were further assembled into contigs with a minimum length of 100 bp using SOAPdenovo with default parameters, except for a kmer value of 57 for *Sabia*, *Stephania* and *Pachysandra*, 49 for *Epimedium*, 73 for *Euptelea*, 71 for *Akebia*, and 45 for *Papaver*. Contigs were ordered by aligning them to previously published plastomes of early-diverging eudicots using BLAST (<http://blast.ncbi.nlm.nih.gov/>). For *Meliosma*, chloroplast isolation, sequencing, and assembly followed the methods of Moore et al. (2006, 2007).

2.2 Genome annotation and alignment

Plastid genomes were annotated using DOGMA (Wyman et al., 2004) and BLAST tools from NCBI (the National Center for Biotechnology Information). Physical maps were drawn using GenomeVx (Conant and Wolfe, 2008), followed by subsequent manual editing with Adobe Illustrator CS5. Boundaries for tRNAs were identified with tRNAscan-SE 1.21 (Schattner et al., 2005) and confirmed by comparison with publicly available early-diverging eudicot plastome sequences. The finished genomes were deposited in GenBank (Table 1).

Whole plastome alignment among 18 early-diverging eudicot species was performed with ProgressiveMauve v 2.4.0 (Darling et al., 2010), including only one copy of the IR (IR_B). In ProgressiveMauve, locally collinear blocks (LCBs) identified during the alignment process are indicated by color to indicate syntenic regions. Sequences of protein-coding genes were extracted from the newly sequenced plastomes and submitted to MultiPipMaker (<http://pipmaker.bx.psu.edu/cgi-bin/multipipmaker>; Schwartz et al., 2000) to create percentage identity plots.

2.3 Phylogenetic analyses

All 79 protein-coding regions were extracted from the plastomes of *Mahonia* (Ma et al., 2013), *Macadamia* (Nock et al., 2014), and the eight newly sequenced early-diverging eudicots. These sequences were then added manually to the 87-taxon alignment of Sun et al. (2013), which already contained all other early-diverging

eudicot plastomes, resulting in a data set with complete coverage of early-diverging eudicot families *sensu* APG III (Angiosperm Phylogeny Group, 2009). GenBank information for all plastomes used for present phylogenetic analyses can be found in Table 1 (in present research) and Table S4 of Moore et al. (2010). Regions of ambiguous alignment and sites with more than 80% missing data were excluded from the alignment.

Maximum likelihood (ML) analyses were conducted using RAxML version 7.4.2, under the general time-reversible (GTR) substitution model and the Γ model of rate heterogeneity (Stamatakis, 2008). We conducted both unpartitioned and partitioned analyses, where the data set was partitioned in various ways by gene and codon position. PartitionFinder version 1.1.1 (Lanfear et al., 2012) was used to select the best-fit partitioning scheme, considering all 237 possible gene-by-codon position partitions. In addition to the best partitioning scheme as selected by PartitionFinder, we also conducted ML analyses of five other partitioning schemes: (1) partitioned with codons 1 and 2 in one partition and codon 3 in a separate partition, considering all genes together (2 partitions); (2) partitioned by codon position, considering all genes together (3 partitions); (3) partitioned by gene, ignoring codon position (79 partitions); (4) partitioned by gene, with codons 1 and 2 in one partition and codon 3 in a separate partition (158 partitions); and (5) partitioned by gene, with each codon position in a separate partition (237 partitions). For all ML analyses, a single set of branch lengths for all partitions was used. For all analyses, 10 independent ML searches were conducted, bootstrap support was estimated with 1000 bootstrap

replicates, and bootstrap proportions were drawn on the tree with highest likelihood score from the 10 independent searches.

MP analyses were conducted in PAUP* 4.0b10 (Swofford, 2000). Shortest trees were obtained by means of heuristic searches with 1000 random taxon addition replicates and tree-bisection-reconnection (TBR) branch swapping and saving all shortest trees per replicate. Bootstrap support (BS) for relationships (Felsenstein, 1985) was estimated from 1000 bootstrap replicates using 10 random taxon additions per replicate, with TBR branch swapping (saving all trees).

2.4 Evolution of IR gene content

To reconstruct the history of gene gain and loss in the IR region in early-diverging eudicots, the presence or absence of genes in the IR region across all taxa in our data set (where this information was publicly available) was mapped to the best ML tree using parsimony. This data matrix is available as supplemental material (Table S2).

3. Results

3.1 Plastome assemblies

For each species, assembly yielded single scaffolds comprising the entire plastome sequence. Junctions between the IR and the SSC and LSC regions were

confirmed with Sanger sequencing. Assembly statistics are presented in Table 2.

3.2 Plastome structural characteristics

All eight newly sequenced plastomes were found to possess the typical quadripartite structure of angiosperms. Seven of the eight plastomes shared identical complements of coding genes, each containing 30 unique tRNA genes, four unique rRNA genes, and 79 unique protein-coding genes (Table S1). As is typical of angiosperms, 18 genes were found to have introns, 16 of which contained a single intron, while two (*ycf3*, *clpP*) possessed two introns (Table S1). In *Epimedium*, *infA* was identified as a severely truncated pseudogene with a length of only 36 bp (Fig. S1), while the *clpP* sequence was highly divergent from the other *clpP* sequences and possessed a 315-bp insertion in exon 1 (Fig. S2). The physical maps of the eight newly sequenced early-diverging eudicot plastomes are shown in Figs. S3-S5.

A comparison of the major structural features of these eight plastid genomes, along with comparable data for the previously published early-diverging eudicot plastomes, is provided in Table 3. *Trochodendron* had the largest plastome, whereas *Papaver* had the smallest. *Platanus*, *Stephania* and *Mahonia* possessed the longest LSC, SSC and IR, respectively, while *Mahonia*, *Epimedium* and *Stephania* possessed the shortest LSC, SSC and IR, respectively. The G/C content of all of these plastomes was approximately 38%, as is typical of angiosperm plastomes.

The IR varied in gene content across early-diverging eudicots, particularly on the

LSC side of the IR (Fig. 1). *Mahonia* had by far the longest IR among early-diverging eudicots, at over 33 kb in length. Based on gene content, the IRs among early-diverging eudicots were divided into six groups (Fig. 1). Subgroup E corresponded to the IR boundaries that are most commonly encountered across eudicots.

The alignments of whole plastomes are provided in Fig. S6. No large-scale rearrangements were identified. Eight blocks (A-H) were shared by all 18 plastomes; Block I, which was located between *rpl32* and *trnL-UAG*, occurred only in the plastomes of Buxales and Trochodendrales (Fig. S6). The MultiPipMaker plot (Fig. S7) showed that most protein-coding genes shared a high degree of conservation, although some genes that are typically highly variable in angiosperms (*e.g. matK*, the 3' end of *ndhF*, *rpl22*, and especially *ycf1*) were also found to be variable across early-diverging eudicots.

3.3 Phylogenetic analyses

The total aligned length of the 79-gene, 97-taxon data set, after character exclusion, was 62,292 characters. The optimal partitioning scheme identified under the Bayesian information criteria (BIC) using relaxed clustering analysis in PartitionFinder ($\ln L = -1103993.09792$; BIC 2214153.94107) contained 34 partitions. Maximum likelihood analyses under the unpartitioned and all partitioned schemes (including the 34-partition, 2-partition, 3-partition, 79-partition, 158-partition and

237-partition schemes) produced identical topologies (Figs. 2, S8-S13).

The support values generated by all ML analyses were highly similar, with most branches possessing 100% support (Figs. 2, S8-S13). The five major early-diverging eudicot lineages (Ranunculales, Sabiaceae, Proteales, Buxales, and Trochodendrales) each formed clades with 100% bootstrap support in all ML analyses (Figs. 2, S8-S13). Relationships among and within these lineages were also resolved with maximal support with the exception of three branches: (1) Trochodendrales were found to be sister to *Gunneridae* with weak/moderate support (ranging from 46% to 86% in the seven ML analyses); (2) Proteales and Sabiaceae were sister with 90%-100% bootstrap supports in the seven ML analyses, and (3) the sister relationship of *Papaver* to the clade comprising of Berberidaceae, Ranunculaceae, Menispermaceae and Lardizabalaceae was supported by 47% to 59% (Figs. 2, S8-S13).

Although the positions of some taxa (e.g., *Ceratophyllum*, *Bulnesia* and *Dillenia*) from the MP analysis were different compared with that from the ML analyses, the relationships within early-diverging eudicots were identical in MP (Fig. S14) and ML analyses (Figs. 2, S8-S13). In the MP analyses, the sister relationship between Trochodendrales and *Gunneridae* was supported by 66%. Sabiaceae was sister to Proteales with 91% support.

3.4 Evolution of IR gene content

Parsimony reconstructions indicated that 18 genes were present in the IR region of the ancestral eudicot plastid genome (Fig. 3). IR gene content is remarkably stable

across most taxa, with major exceptions restricted to taxa with known IR loss (*Medicago*), extensive rearrangements (e.g. *Trachelium*, *Pelargonium*) and/or significant IR boundary changes (e.g. *Illicium*, *Mahonia*, Trochodendrales). Our analyses suggest that the position of *rps19* in or out of the IR has been most labile across angiosperms, with at least 16 individual gains or losses (i.e., not considering cases where *rps19* plus adjacent genes are involved in boundary shifts) across angiosperms, including four gains (including one at the base of Ranunculales) and two losses of this gene across early-diverging eudicots.

4. Discussion

4.1 Genome structure and gene content in early-diverging eudicot plastomes

Although the early-diverging eudicot plastomes analyzed in our study are syntenic and share the typical quadripartite structure of most land plants, they show a remarkable variation in IR length (Fig. 1). This is the main contributing factor to differences in plastome length across early-diverging eudicots (Table 3). The IRs of angiosperm plastomes fluctuate greatly in size, particularly at the LSC/IR_B boundary (J_{LB}) (Downie and Jansen, 2015). In most eudicot plastomes, J_{LB} lies within the *rps19* gene (Shinozaki et al., 1986; Yukawa et al., 2005; Downie and Jansen, 2015), but 13 of the 18 early-diverging eudicot plastomes analyzed here have experienced an expansion at this end of the IR (Fig. 1). In contrast, the SSC/IR boundaries are

comparatively stable across early-diverging eudicots. In most angiosperms, the SSC/IR boundary lies within the 5' end of *ycfI* (Raubeson et al., 2007), and this is the case in all early-diverging eudicots as well.

The ancestral IR gene reconstruction (Fig. 3) reveals a significant amount of homoplasy in IR boundary shifts, which is also evident in the taxonomic distribution of the various IR types in Figs. 1 and 2. Of the three IR types that are shared by more than one taxon, only one (IR type B, which is found only in Trochodendrales and represents the most extreme of the IR boundary shifts that are shared by multiple taxa) shows no evidence of homoplasy. IR types D and E show a particularly large amount of homoplasy, which is perhaps not surprising given that the two types differ only by the shift of J_{LB} from *rps19* to the adjacent *rpl22* (Fig. 3).

In contrast to the seeming instability of the J_{LA} boundary, our analyses reveal that gene content is almost perfectly conserved across early-diverging eudicots. Only three clear cases and one possible additional case of gene loss are apparent. We found that *infA* is a pseudogene in the plastome of *Epimedium* (Fig. S1), as has been found in *Ranunculus* (Raubeson et al., 2007; Fig. S1). This gene has been pseudogenized in a number of other core eudicot plastomes as well, including tobacco (Shinozaki et al., 1986), *Silene* (Sloan et al., 2012) and numerous rosids (Millen et al., 2001) including *Arabidopsis* (Sato et al., 1999) and *Oenothera elata* (Hupfer et al., 2000). In some cases, the pseudogenization of this gene in these plants has been demonstrated to result from plastid-to-nucleus gene transfer (Millen et al., 2001), and such may be the case in *Epimedium* and *Ranunculus* as well.

Another example of gene loss in early-diverging eudicots is the apparent complete loss of *rpoA* from the plastome of *Mahonia bealei* (Ma et al., 2013). The loss of *rpoA*, which produces one of the polypeptides in the plastid-encoded RNA polymerase, is far rarer than the loss of *infA* across land plants. Within angiosperms, *rpoA* is absent from many nonphotosynthetic angiosperms (Krause and Scharff, 2013), including mycoheterotrophic orchids (Logacheva et al., 2011) and holoparasitic plants such as Orobanchaceae (Wolfe et al., 1992) and *Cuscuta* (Krause et al., 2003). This gene is also highly divergent in the massively rearranged plastome of *Pelargonium × hortorum* (Chumley et al., 2006), although no clear evidence of a nuclear copy of *rpoA* has been found (Zhang et al., 2013). The *rpoA* gene is also absent from the plastome of mosses, where it has been transferred to the nucleus (Sugiura et al., 2003).

Finally, the *clpP* sequence from the *Epimedium* plastome is highly divergent from that from other early-diverging eudicot plastomes (Fig. S2), with a large insertion in the first exon, together suggesting the possibility that the gene has been pseudogenized. However, we emphasize that the gene is in frame, and hence we cannot be certain of its status as a pseudogene. Among photosynthetic angiosperms, the loss of *clpP* has been observed in several plastomes, including some legumes (Jansen et al., 2008), *Passiflora*, and *Scaevola* (Wicke et al., 2011).

4.2 Relationships among early-diverging eudicots

The 79-gene, 97-taxon ML trees (Figs. 2, S8-S13) are almost entirely consistent with previous estimates of early-diverging eudicot relationships (*e.g.* Hoot et al., 1999; Soltis *et al.*, 2000; Qiu et al., 2006; Worberg et al., 2007; Wang et al., 2009; Moore et al., 2010, 2011; Soltis et al., 2011; Barniske et al., 2012; Sun et al., 2013; Ruhfel et al., 2014), with maximal bootstrap support at almost all nodes. Among the most important results are the sister relationship of *Euptelea* to remaining Ranunculales, the high support for Sabiaceae as sister to Proteales, and the sister relationship of Trochodendrales to *Gunneridae*. Our work adds to a growing body of evidence that Eupteleaceae are sister to remaining Ranunculales (*e.g.* Kim et al., 2004; Wang et al., 2009, Soltis et al., 2011), although support is only moderate and taxon sampling within most families of Ranunculales remains limited.

The sister relationship of Sabiaceae and Proteales has been recovered in most prior phylogenetic analyses of early-diverging eudicots, but usually with low to moderate support (*e.g.* Worberg et al., 2007; Moore et al., 2010; Soltis et al., 2011; Barniske et al., 2012; Sun et al., 2013; Nock et al., 2014; Ruhfel et al., 2014; Wu et al., 2014). Our work provides perhaps the best support yet obtained for a clade of Sabiaceae and Proteales and strengthens the case for including Sabiaceae in a broadened circumscription of Proteales, as suggested by APG III (Angiosperm Phylogeny Group, 2009).

In contrast to the improved picture of the placement of Sabiaceae, our work complicates the question of which clade is sister to *Gunneridae*. In the majority of prior analyses, including plastome-scale analyses, Buxales have been resolved as

sister to *Gunneridae*, although typically with no more than moderate support (*e.g.* Qiu et al., 2006; Worberg et al., 2007; Barniske et al., 2012; Moore et al., 2010; Sun et al., 2013; Ruhfel et al., 2014; Wu et al., 2014), the lone exception being the 98% support found for Buxales + *Gunneridae* in Soltis et al. (2011). Our study resolves Trochodendrales as sister to *Gunneridae*, albeit with weak support (Fig. 2). This relationship has been found in a handful of other analyses (*e.g.* Hoot et al.; 1999; Soltis et al., 1999; Soltis et al., 2000), most notably in those of Moore et al. (2011) based on complete plastid IR sequences, in which Trochodendrales + *Gunneridae* received 93% support. The current study has much more extensive gene and taxon sampling than any previous plastome-scale analysis of early-diverging eudicots, although it should be emphasized that (1) several key lineages remain unsampled, particularly in Buxales (*e.g.* *Haptanthus*, *Didymeles*), and (2) the plastome is a single linked locus that represents one potential signal from the total genome. Hence, the relationships among Buxales, Trochodendrales, and *Gunneridae* must be considered as yet unsettled. In any case, the fact that the position of Trochodendrales moved in response to additional taxon sampling in the current analysis compared to previous plastome-scale analyses (Moore et al., 2010; Sun et al., 2013; Ruhfel et al., 2014) emphasizes the importance of taxon sampling in future studies of this problem.

Acknowledgments

This work was supported by the Natural Science Foundation of Hubei Province (2015CFB681), National Natural Science Foundation of China (31070191, 31370223)

and 31370242), and the US National Science Foundation (DEB EF-0431266).

References

- Angiosperm Phylogeny Group, 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161, 105–121.
- Barniske, A.M., Borsch, T., Müller, K., Krug, M., Worberg, A., Neinhuis, C., Quandt, D., 2012. Phylogenetics of early branching eudicots: Comparing phylogenetic signal across plastid introns, spacers, and genes. J. Syst. Evol. 50, 85-108.
- Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms revisited. Am. J. Bot. 97, 1296-1303.
- Cantino, P.D., Doyle, J.A., Graham, S.W., Judd, W.S., Olmstead, R.G., Soltis, D.E., Soltis, P.S., Donoghue, M.J., 2007. Towards a phylogenetic nomenclature of *Tracheophyta*. Taxon 56, E1-E44.
- 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 x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. 23, 2175-2190.
- Conant, G.C., Wolfe, K.H., 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24, 861–862.

- 400 Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome
401 alignment with gene gain, loss and rearrangement. PLoS ONE 5, e11147.
- 402 Downie, S.R., Jansen, R.K., 2015. A comparative analysis of whole plastid genomes
403 from the Apiales: expansion and contraction of the inverted repeat,
404 mitochondrial to plastid transfer of DNA, and identification of highly
405 divergent noncoding regions. Syst. Bot. 40, 336–351.
- 406 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the
407 bootstrap. Evolution 39, 783–791.
- 408 Goulding, S.E., Olmstead, R.G., Morden, C.W., Wolfe, K.H., 1996. Ebb and flow of
409 the chloroplast inverted repeat. Mol. Gen. Genet. 252, 195–206.
- 410 Hansen, D.R., Dastidar, S.G., Cai, Z.Q., Penaflor, C., Kuehl, J.V., Boore, J.L., Jansen,
411 R.K., 2007. Phylogenetic and evolutionary implications of complete
412 chloroplast genome sequences of four early-diverging angiosperms: *Buxus*
413 (*Buxaceae*), *Chloranthus* (*Chloranthaceae*), *Dioscorea* (*Dioscoreaceae*), and
414 *Illicium* (*Schisandraceae*). Mol. Phylogenet. Evol. 45, 547–563.
- 415 Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon
416 sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.
- 417 Hupfer, H., Swiatek, M., Hornung, S., Herrmann, R.G., Maier, R.M., Chiu, W.L.,
418 Sears, B., 2000. Complete nucleotide sequence of the *Oenothera elata* plastid
419 chromosome, representing plastome I of the five distinguishable *Euoenothera*
420 plastomes. Mol. Gen. Genet. 263, 581 – 585.

- 421 Hoot, S.B., Magallón, S., Crane, P.R., 1999. Phylogeny of basal eudicots based on
422 three molecular data sets: *atpB*, *rbcL* and 18S nuclear ribosomal DNA
423 sequences. Ann. Mo. Bot. Gard. 86, 1-32.
- 424 Jansen, R.K., Cai Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack,
425 J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley,
426 T.W., Lee, S., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis
427 of 81 genes from 64 plastid genomes resolves relationships in angiosperms
428 and identifies genome-scale evolutionary patterns. P. Natl. Acad. Sci. USA.
429 104,19369-19374.
- 430 Jansen, R.K., Ruhlman, T.A., 2012. Plastid genomes of seed plants, in: Bock, R.
431 Knoop, V. (Eds.), Genomics of chloroplasts and mitochondria, advances in
432 photosynthesis and respiration. Springer, Dordrecht, pp. 103–126
- 433 Jansen, R.K., Wojciechowski, M.F., Sanniyasi, E., Lee, S.B., Daniell, H., 2008.
434 Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the
435 phylogenetic distribution of *rps12* and *clpP* intron losses among legumes
436 (Leguminosae). Mol. Phylogenet. Evol. 48, 1204-1217.
- 437 Kim, S., Soltis, D.E., Soltis, P.S., Zanis, M.J., Suh, Y., 2004. Phylogenetic
438 relationships among early-diverging eudicots based on four genes: were the
439 eudicots ancestrally woody? Mol. Phylogenet. Evol. 31, 16-30.
- 440 Kim, Y.K., Park, C.W., Kim, K.J., 2009. Complete chloroplast DNA sequence from a
441 Korean endemic genus, *Megaleranthis saniculifolia*, and its evolutionary
442 implications. Mol. Cells 27, 365–381.

- 443 Krause, K., Berg., S, Krupinska, K., 2003. Plastid transcription in the holoparasitic
444 plant genus *Cuscuta*: parallel loss of the *rrn16* PEP-promoter and of the *rpoA*
445 and *rpoB* genes coding for the plastid-encoded RNA polymerase. *Planta* 5,
446 815-823.
- 447 Krause, K., Scharff, L.B., 2013. Reduced genomes from parasitic plant plastids:
448 templates for minimal plastomes? *Prog. Bot.* 75, 97-115.
- 449 Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined
450 selection of partitioning schemes and substitution models for phylogenetic
451 analyses. *Mol. Biol. Evol.* 29, 1695-1701.
- 452 Leebens-Mack, J., Raubeson, L.A., Cui, L., Kuehl, J.V., Fourcade, M.H., Chumley,
453 T.W., Boore, J.L., Jansen, R.K., dePamphilis, C.W., 2005. Identifying the
454 early-diverging angiosperm node in chloroplast genome phylogenies:
455 sampling one's way out of the Felsenstein Zone. *Mol. Biol. Evol.* 22,
456 1948-1963.
- 457 Logacheva, M.D., Schelkunov, M.I., Penin ,A.A., 2011. Sequencing and analysis of
458 plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biol.*
459 *Evol.* 3, 1296-1303.
- 460 Ma, J., Yang, B.X., Zhu, W., Sun, L.L., Tian, J.K., Wang, X., 2013. The complete
461 chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a
462 significant expansion of the inverted repeat and phylogenetic relationship with
463 other angiosperms. *Gene* 528, 120-131.

- 464 Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L.,
465 Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W., Calie, P.J., Jermini,
466 L.S., Wolfe, K.H., 2001. Many parallel losses of *infA* from chloroplast DNA
467 during angiosperm evolution with multiple independent transfers to the
468 nucleus. *Plant Cell* 13, 645-658.
- 469 Moore, M.J., Dhingra, A., Soltis, P.S., Shaw, R., Farmerie, W.G., Folt, K.M., Soltis,
470 D.E., 2006. Rapid and accurate pyrosequencing of angiosperm plastid
471 genomes. *BMC Plant Biol.* 6, 17.
- 472 Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale
473 data to resolve enigmatic relationships among early-diverging angiosperms. *P.*
474 *Natl. Acad. Sci. USA.* 104, 19363–19368.
- 475 Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic
476 analysis of 83 plastid genes further resolves the early diversification of
477 eudicots. *P. Natl. Acad. Sci. USA.* 107, 4623–4628.
- 478 Moore, M.J., Hassan, N., Gitzendanner, M.A., Bruenn, R.A., Croley, M., Vandeventer,
479 A., Horn, J.W., Dhingra, A., Brockington, S.F., Latvis, M., Ramdial, J.,
480 Alexandre, R., Piedrahita, A., Xi, Z., Davis, C.C., Soltis, P.S., Soltis, D.E.,
481 2011. Phylogenetic analysis of the plastid inverted repeat for 244 species:
482 insights into deeper-level angiosperm relationships from a long, slowly
483 evolving sequence region. *Int. J. Plant Sci.* 172, 541-558.

- 484 Nock, C.J., Baten, A., King, G.J., 2014. Complete chloroplast genome of *Macadamia*
485 *integrifolia* confirms the position of the Gondwanan early-diverging eudicot
486 family Proteaceae. BMC Genomics 15, S13.
- 487 Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovskaya,
488 O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor,
489 D.W., Testa, C.M., Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey,
490 W., Quandt, D., Davis, C.C., 2006. The deepest divergences in land plants
491 inferred from phylogenomic evidence. P. Natl. Acad. Sci. USA. 103,
492 15511–15516.
- 493 Palmer, J.D., 1991. Plastid chromosomes: structure and evolution, in: Bogorad L.,
494 Vasil, I.K., (Eds.), Cell culture and somatic cell genetics of plants. Academic
495 Press, San Diego, pp. 5–53.
- 496 Palmer, J.D., Stein, D.B., 1986. Conservation of chloroplast genome structure among
497 vascular plants. Curr. Genet. 10, 823-833.
- 498 Plunkett, G.M., Downie, S.R., 2000. Expansion and contraction of the chloroplast
499 inverted repeat in Apiaceae subfamily Apioideae. Syst. Bot. 25, 648–667.
- 500 Raubeson, L.A., Jansen, R.K., 2005. Chloroplast genomes of plants, in: Henry, R.J.,
501 (Eds.), Plant diversity and evolution: Genotypic and phenotypic variation in
502 higher plants. CAB International, London, pp. 45–68.
- 503 Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L.,
504 Jansen, R.K., 2007. Comparative chloroplast genomics: analyses including

- new sequences from the angiosperms *Nuphar advena* and *Ranunculus*
Macranthus. BMC Genomics 8, 174.
- Ruhfel, B.R., Gitzendanner, 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. (Eds.), Chloroplast Biotechnology: Methods and Protocols.
 Springer, New York, pp. 3–38.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Tabata, S., 1999. Complete structure
 of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 6, 283–290.
- Schattner, P., Brooks ,A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and
 snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids
 Res. 33, W686–W689
- Schwartz, S., Zhang, Z, Frazer, K.A. Smit, A., Riemer, C., Bouck, J., Gibbs, R.,
 Hardison, R., Miller, W., 2000. PipMaker - A web server for aligning two
 genomic DNA sequences. Genome Res. 10, 577–586.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayshida, N., Matsubayashi, T.,
 Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi- Shinozaki, K., Ohto, C.,
 Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K.,
 Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M., 1986.
 The complete nucleotide sequence of the tobacco chloroplast genome: its gene
 organization and expression. Embo J. 5, 2043–2049.

- 527 Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y-J., Gao L-Z., 2012. An improved
528 chloroplast DNA extraction procedure for whole plastid genome sequencing.
529 PLoS ONE 7, e31468.
- 530 Sloan, D.B., Alverson, A.J., Wu, M., Palmer, J.D., Taylor, D.R., 2012. Recent
531 acceleration of plastid sequence and structural evolution coincides with
532 extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome
533 Biol. Evol. 4, 294-306.
- 534 Soltis DE, Soltis PS, Chase MW, Mort M, Albach D, Zanis, M., Savolainen, V., Hahn,
535 W., Hoot, S., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J,
536 Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from a
537 combined data set of 18S rDNA, *rbcL* and *atpB* sequences. Bot. J. Linn. Soc.
538 133, 381-461.
- 539 Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington,
540 S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlswald, B.S., Bell,
541 C.D., Latvis, M.,Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth,
542 C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.L., Hilu, K.W., Davis,
543 C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue,
544 M.J., Soltis, P.S., 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot.
545 98, 704 – 730.
- 546 Soltis, P.S., Soltis, D.E., Chase, M.W. 1999. Angiosperm phylogeny inferred from
547 multiple genes as a tool for comparative biology. Nature 402, 402-404.

- 548 Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the
549 RAxML Web-Servers. *Syst. Biol.* 75, 758-771.
- 550 Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012.
551 Navigating the tip of the genomic iceberg: next-generation sequencing for
552 plant systematics. *Am. J. Bot.* 99, 349-364.
- 553 Sun, Y.X., Moore, M.J., Meng, A.P., Soltis, P.S., Soltis, D.E., Li, J.Q., Wang, H.C.,
554 2013. Complete plastid genome sequencing of Trochodendraceae reveals a
555 significant expansion of the inverted repeat and suggests a Paleogene
556 divergence between the two extant species. *PloS ONE* 8: e60429.
- 557 Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., Sugita, M., 2003. Complete
558 chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for
559 the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic*
560 *Acids Res.* 31, 5324-31.
- 561 Swofford, D.L., 2000. PAUP*: phylogenetic analysis using parsimony (*and other
562 methods), version 4. Sinauer, Sunderland, MA.
- 563 Wang, W., Lu, A.M., Ren, Y., Endress, M.E., Chen, Z.D., 2009. Phylogeny and
564 classification of Ranunculales: Evidence from four molecular loci and
565 morphological data. *Perspect. Plant Ecol.* 11, 81 - 110.
- 566 Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D., 2011.
567 The evolution of the plastid chromosome in land plants: gene content, gene
568 order, gene function. *Plant Mol. Biol.* 76, 273-297.

- 569 Wolfe, K.H., Morden, C.W., Palmer, J.D., 1992. Function and evolution of a minimal
570 plastid genome from a nonphotosynthetic parasitic plant. P. Natl. Acad. Sci.
571 USA. 89, 10648-10652.
- 572 Worberga, A., Quandtb, D., Barniskeb, A., Löhnea, C., Hiluc, K.W., Borscha, T., 2007.
573 Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving
574 DNA. Org. Divers. Evol 7, 55 – 77.
- 575 Wu, Z.H., Gui, S.T., Quan, Z.W., Pan, L., Wang, S.Z., Ke, W.D., Liang, D.Q., Ding, Y.,
576 2014. A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae)
577 evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing
578 platforms: insight into the plastid evolution of early-diverging eudicots. BMC
579 Plant Biol. 2014, 14, 289.
- 580 Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar
581 genomes with DOGMA. Bioinformatics, 20, 3252–3255.
- 582 Xi, Z., Ruhfel, B.R., Schaefer, H., Amorim, A.M., Sugumaran, M., Wurdack, K.J.,
583 Endress, P.K., Matthews, M.L., Stevens, P.F., Mathews, S., Davis, C.C., 2012.
584 Phylogenomics and a posteriori data partitioning resolve the Cretaceous
585 angiosperm radiation Malpighiales. P. Natl. Acad. Sci. USA. 109,
586 17519-17524.
- 587 Yukawa, M., Tsudzuki, T., Sugiura, M., 2005. The 2005 version of the chloroplast
588 DNA sequence from tobacco (*Nicotiana tabacum*). Plant Mol. Biol.Rep. 23,
589 359–365.
- 590 Zhang, J., Ruhlman, T.A., Mower, J.P., Jansen, R.K., 2013. Comparative analyses of

591 two Geraniaceae transcriptomes using next-generation sequencing. BMC Plant
592 Biol. 13, 228.
593

594 **Table 1**

595 Species of early-diverging eudicots included in the present study. Collection locality and voucher information are provided for newly sequenced plastomes. WBG =
 596 Wuhan Botanical Garden, Wuhan, China; FLAS = Herbarium of Florida Museum of Natural History; HIB = Herbarium of Wuhan Botanical Garden, Chinese
 597 Academy of Sciences.

APG III Order	Family	Species	Collection locality	Voucher information	GenBank accession number
Buxales	Buxaceae	<i>Buxus microphylla</i> Siebold & Zucc.	—	—	NC_009599
Buxales	Buxaceae	<i>Pachysandra terminalis</i> Siebold & Zucc.	WBG (cultivated)	Y.X. Sun 1401 (HIB)	KU204904
Proteales	Proteaceae	<i>Macadamia integrifolia</i> Maiden &	—	—	NC_025288

		Betche			
Proteales	Nelumbonaceae	<i>Nelumbo nucifera</i> Gaertn.	—	—	NC_025339
Proteales	Platanaceae	<i>Platanus occidentalis</i> L.	—	—	NC_008335
Ranunculales	Berberidaceae	<i>Epimedium sagittatum</i> (Sieb. & Zucc.) Maxim.	WBG (cultivated)	Y.X. Sun 1402 (HIB)	KU204899
Ranunculales	Berberidaceae	<i>Mahonia bealei</i> (Fortune) Pynaert	—	—	NC_022457
Ranunculales	Berberidaceae	<i>Nandina domestica</i> Thunb.	—	—	NC_008336
Ranunculales	Eupteleaceae	<i>Euptelea pleiosperma</i> Hook.f. & Thomson	WBG (cultivated)	Y.X. Sun 1403 (HIB)	KU204900
Ranunculales	Lardizabalaceae	<i>Akebia trifoliata</i> (Thunb.) Koidz.	WBG (cultivated)	Y.X. Sun 1404 (HIB)	KU204898

Ranunculales	Menispermaceae	<i>Stephania japonica</i> (Thunb.) Miers	Ma'anshan Forest Park, Wuhan, China	Y.X. Sun 1405 (HIB)	KU204903
Ranunculales	Papaveraceae	<i>Papaver somniferum</i> L.	WBG (cultivated)	Y.X. Sun 1406 (HIB)	KU204905
Ranunculales	Ranunculaceae	<i>Megaleranthis saniculifolia</i> Ohwi	—	—	NC_012615
Ranunculales	Ranunculaceae	<i>Ranunculus macranthus</i> Scheele	—	—	NC_008796
Trochodendrales	Trochodendraceae	<i>Tetracentron sinense</i> Oliv.	—	—	NC_021425
Trochodendrales	Trochodendraceae	<i>Trochodendron aralioides</i> Siebold & Zucc.	—	—	NC_021426
unplaced	Sabiaceae	<i>Meliosma</i> aff. <i>cuneifolia</i> Franch.	Heronswood Nursery, Kingston, Washington, USA (Heronswood accession DJHC 872)	M. J. Moore 333 (FLAS)	KU204901

unplaced	Sabiaceae	<i>Sabia yunnanensis</i> Franch.	Xiaoheishan Mountain, Yunnan, China	H.C. Wang 1201	KU204902
(HIB)					

598

599 **Table 2**

600 Comparison of assemblies of eight early-diverging eudicot plastomes.

Species	Read length (bp)	No. of plastid reads	Average depth of coverage
<i>Akebia trifoliata</i>	90	246,168	139.9×
<i>Epimedium sagittatum</i>	90	233,867	133.0×
<i>Euptelea pleiosperma</i>	90	213,822	118.9×
<i>Meliosma</i> aff. <i>cuneifolia</i>	115 (average)	44,941	32.2×
<i>Pachysandra terminalis</i>	90	260,291	145.8×
<i>Papaver somniferum</i>	90	176,465	103.8×

<i>Sabia yunnanensis</i>	90	166,637	92.5×
<i>Stephania japonica</i>	90	138,232	78.9×

601

602 **Table 3**

603 Comparison of basic characteristics among 18 early-diverging eudicot plastomes. All lengths are in bp.

Taxon			Total genome length	LSC length	SSC length	IR length	Number of genes within IR	Overall G/C content
Buxales	Buxaceae	<i>Pachysandra</i>	160,623	86,303	18,820	27,750	22	38.3%
		<i>Buxus</i>	159,010	88,143	17,747	26,558	18	38.1%
Proteales	Platanaceae	<i>Platanus</i>	161,791	92,150	19,509	25,066	19	38.0%
	Nelumbonaceae	<i>Nelumbo</i>	163,330	91,910	19,360	26,030	18	38.0%
	Proteaceae	<i>Macadamia</i>	159,714	88,093	18,813	26,404	18	38.1%

Ranunculales	Berberidaceae	<i>Nandina</i>	156,599	85,473	19,002	26,062	19	38.3%
		<i>Epimedium</i>	158,273	87,078	17,059	27,068	18	38.9%
		<i>Mahonia</i>	164,792	73,213	18,631	36,474	33	38.1%
	Ranunculaceae	<i>Megaleranthis</i>	159,924	88,326	18,382	26,608	19	38.0%
		<i>Ranunculus</i>	155,129	84,638	18,909	25,791	19	37.9%
	Eupteleaceae	<i>Euptelea</i>	161,834	90,449	19,311	26,037	19	38.6%
	Lardizabalaceae	<i>Akebia</i>	158,339	87,057	19,024	26,129	19	38.7%
	Menispermaceae	<i>Stephania</i>	157,719	88,693	20,346	24,340	19	38.2%
Trochodendrales	Trochodendraceae	<i>Papaver</i>	152,931	83,029	17,920	25,991	19	38.9%
		<i>Tetracentron</i>	164,467	84,466	19,539	30,231	24	38.1%
		<i>Trochodendron</i>	165,945	85,483	18,974	30,744	24	38.0%
unplaced	Sabiaceae	<i>Meliosma</i>	160,357	89,479	18,590	26,144	18	38.4%

<i>Sabia</i>	162,080	89,998	18,892	26,595	19	38.6%
--------------	---------	--------	--------	--------	----	-------

604

605

- 1 **Fig. 1.** Comparison of gene content in IR among 18 early-diverging eudicot plastomes.
- 2 **Fig. 2.** Phylogram of the best tree determined by RAxML for the 79-gene, 97-taxon data set using 34-partition scheme ($\ln L = -1108895.169560$).
- 3 Numbers associated with branches are ML bootstrap support values; asterisks indicate ML BS=100%. The range of ML bootstrap values for all
- 4 other schemes (including single partition, 2-partition, 3-partition, 79-partition, 158-partition and 237 partition) are presented in parentheses; the
- 5 presence of a single number indicates all analyses shared the same support value. The letters A-F associated with early-diverging eudicots
- 6 correspond to the six IR types from Fig. 1.
- 7 **Fig. 3.** Proposed evolutionary history of IR gene content found in angiosperm plastid genomes. Open text on a branch denotes a gain of that gene,
- 8 whereas strikethrough indicates the loss of that gene. Taxa in red lack information concerning IR gene content. The vertical string of numbers at the
- 9 base of the tree indicates the ancestral gene complement of angiosperms.

Figure 1

Among the five IR types, type 'E' is the typical IR in eudicots.

Figure 2

1=*ndhE*,
2=*ndhG*,
3=*ndhA*,
4=*ndhH*,
5=*rps15*,
6=*rrn23* (1),
7=*ycf1*,
8=*rpl32*,
9=*ndhF*,
10=*ycf3*,
11=*clpP*,
12=*psbB* (1),
13=*trnN*-GUU,
14=*trnR*-ACG,
15=*rrn5*,
16=*rrn4.5*,
17=*rrn23* (2),
18=*trnA*-UGC,
19=*trnI*-GAU,
20=*rrn16*,
21=*trnV*-GAC,
22=*rps12-3'*end,
23=*rps7*,
24=*ndhB*,
25=*trnL*-CAA,
26=*ycf2*,
27=*trnI*-CAU,
28=*rpl23*,
29=*rpl2* (exon1),
30=*rpl2* (exon2),
31=*trnH*-GUG,
32=*rps19*,
33=*rpl22*,
34=*rps3*,
35=*rpl16* (exon1),
36=*rpl16* (exon2),
37=*rpl14*,
38=*rps8*,
39=*infA*,
40=*rpl36*,
41=*rps11*,
42=*rpoA*,
43=*petD*,
44=*petB*,
45=*psbH*,
46=*psbN*,
47=*psbT*,
48=*psbB* (2)

1=ndhE,
2=ndhG,
3=ndhA,
4=ndhH,
5=rps15,
6=rrn23 (1),
7=ycf1,
8=rpl32,
9=ndhF,
10=ycf3,
11=clpP,
12=psbB (1),
13=trnN-GUU,
14=trnR-ACG,
15=rrn5,
16=rrn4.5,
17=rrn23 (2),
18=trnA-UGC,
19=trnI-GAU,
20=rrn16,
21=trnV-GAC,
22=rps12-3'end ,
23=rps7,
24=ndhB,
25=trnL-CAA,
26=ycf2,
27=trnI-CAU,
28=rpl23,
29=rpl2 (exon1),
30=rpl2 (exon2),
31=trnH-GUG,
32=rps19,
33=rpl22,
34=rps3,
35=rpl16 (exon1),
36=rpl16 (exon2),
37=rpl14,
38=rps8,
39=infA,
40=rpl36,
41=rps11,
42=rpoA,
43=petD,
44=petB,
45=psbH,
46=psbN,
47=psbT,
48=psbB (2)

Pentapetalae

Early-diverging eudicots

Monocyledoneae

Magnoliidae

Early-diverging angiosperms

1 2 3 4 5 6 10 11 12 21 22 23 24 26 27 28 29 30

7 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

3 4 5 8 9 33 34 35 36 37 38 39 40 41 42 43

25 26 27 28 29 30

Solanum
Atropa
Nicotiana
Cuscuta
Ipomoea
Antirrhinum
Epifagus
Jasminum
Coffea
Nerium
Ehretia
Aucuba
Helianthus
Lactuca
Scaevola
Trachelium
Anethum
Daucus
Panax
Loniceria
Ilex
Rhododendron
Cornus
Spinacia
Plumbago
Berberidopsis
Ximenia
Phoradendron
Lotus
Medicago
Phaseolus
Glycine
Morus
Ficus
Cucumis
Quercus
Passiflora
Populus
Manihot
Euonymus
Oxalis
Bulnesia
Brassica
Arabidopsis
Gossypium
Citrus
Staphylea
Eucalyptus
Oenothera
Pelargonium
Heuchera
Liquidambar
Vitis
Dillenia
Gunnera
Tetracentron
Trochodendron
Buxus
Pachysandra
Platanus
Macadamia
Nelumbo
Meliosma
Sabia
Epimedium
Mahonia
Nandina
Ranunculus
Megaleranthis
Stephania
Akebia
Papaver
Euptelea
Ceratophyllum
Triticum
Oryza
Zea
Typha
Musa
Elaeis
Yucca
Phalaenopsis
Dioscorea
Lemna
Acorus
Piper
Drimys
Calycanthus
Liriodendron
Chloranthus
Illicium
Nymphaea
Nuphar
Amborella

13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30