Plastome Structure and Phylogenetic Relationships of Styracaceae (Ericales)

Xiu-lian Cai[a], Jacob B. Landis[b], Hong-Xin Wang [a], Jian-Hua Wang [a], Zhi-Xin Zhu [a], Hua-Feng Wang [a,*]

[a]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, School of Life and Pharmaceutical Sciences, Hainan University, Haikou 570228, China

[b]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14850, USA

***Authors for correspondence**:

Hua-Feng Wang, E-mail: wanghuafeng2012@foxmail.com;

Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

ABSTRACT

Background: The Styracaceae are a woody, dicotyledonous family containing 12 genera and an estimated 160 species. Recent studies have shown that *Styrax* is monophyletic, *Alniphyllum* and *Bruinsmia* cluster into a clade with an approximately 20-kb inversion in the LSC. *Halesia* and *Pterostyrax* are not supported as monophyletic, while *Melliodendron* and *Changiostyrax* always from a clade sister to the rest of the family. However, the phylogenetic relationship of Styracaceae at the level of genera remains ambiguous.

Results: We collected 28 complete plastomes of Styracaceae, including 12 sequences newly reported here and 16 publicly available complete plastome sequences, comprising 11 of the 12 genera of Styracaceae. All species possessed the typical quadripartite structure of angiosperm plastomes, and the sequence difference is small, except for the large 20-kb (14 genes) inversion region found in *Alniphyllum* and *Bruinsmia*. Seven coding sequences (*rps4*, *rpl23*, *accD*, *rpoC1*, *psaA*, *rpoA* and *ndhH*) were identified to possess positively selected sites. Phylogenetic reconstructions based on seven data sets (i.e., LSC, SSC, IR, Coding, Non-coding, combination of LSC+SSC and concatenation of LSC+SSC+one IR) produced similar topologies and most relationships are consistent with previous findings. In our study, *Pterostyrax* was strongly supported as monophyletic; *Melliodendron* and *Changiostyrax* as successively sister to the rest of the family.

Conclusion: Our results clearly indicate that *Pterostyrax* is monophyletic, and the establishment of *Perkinsiodendron* and *Changiostyrax* are supported. A 20-kb reverse sequence was found in the newly published sequence of *Alniphyllum fortunei*, which confirmed the existence of large inversion sequence in *Alniphyllum* and *Bruinsmia*.

Keywords: Styracaceae; Plastome; Genome structure; Phylogeny; positive selection

Background

The Styracaceae DC. & spreng (Ericales) comprise an angiosperm clade of 12 genera and over 160 species, mainly distributed in regions of Asia, as well as tropical and temperate America, and the Mediterranean [1]. The family consists of shrubs or trees with stellate pubescent or epidermal scales, simple leaves, inflorescence of raceme, cyme or panicle, and actinomorphic flowers with varying degrees of synsepaly and sympetaly [2]. The fruit of Styracaceae is a drupe or capsule, with persistent calyx, surrounding or united with the fruit. The Styracaceae have been included in a number of morphological studies, including leaf anatomy [3], wood anatomy [4], pollen morphology [5] and floral morphology and anatomy [2], but distinguishing between genera in the family primarily involves variation in fruit morphological characters (e.g. hypanthium at maturity). On one hand the ovary is inferior with a persistent hypanthium combined with the fruit at maturity (i.e., *Changiostyrax* C.T. Chen, *Halesia* J. Ellis ex L , *Melliodendron* Hand.-Mazz, *Parastyrax* Siebold & Zucc., *Perkinsiodendron* P. W. Fritsch, *Pterostyrax* W.W. Sm., *Rehderodendron* Hu, and *Sinojackia* Hu), while on the other hand the ovary is superior and a persistent hypanthium forms only at the base of the fruit at maturity (*Alniphyllum* Matsum., *Bruinsmia* Boerl. & Koord, *Styrax* L.,)(Fig.1). Moreover, the ovary of *Huodendron* Rehder is semisuperior with a persistent hypanthium extending from the base to about two-thirds of the fruit length [1, 2], a feature considered to be transitional.

The systematic position of Styracaceae has been revised numerous times. Early researchers thought Styracaceae was positioned in the order Ebenales, along with the well-known Sapotaceae, Ebenaceae, and Symplocaceae, and the small family Lissocarpaceae [6, 7, 8, 9]. However, Cronquist [6] showed that there were some original characteristics and some new evolutionary characters in each family of persimmons, which may have arisen via parallel evolution. Based on embryological and anatomical studies, Herbert [10] suggested that the family Styracaceae and Theaceae, which have many common characteristics, may have originated from the

ancestors of Theaceae. According to molecular systematic studies, Styracaceae has been recognized as part of the order Ericales *sensu lato* [11]. Within the family, phylogenetic resolution generally remains poor. At most 17 genera have been included in Styracaceae, with *Symplocos* L, *Diclidanthera* Mart*, Afrostyrax* Perk et Gil, *Foveolaria* Ruiz et pav., *Pamphilia* Mart. ex A. DC, *Huapierre* et De Wil, and *Lissocarpa* Benth having been placed in the Styracaceae by various authors [12]. *Symplocos*, *Diclidanthera* and *Lissocarpa* were excluded from Styracaceae by Perkins [13], after which they were treated as independent families (*Symplocos*) or combined into Polygalaceae (*Diclidanthera*) and Ebenaceae (*Lissocarpa*), *Diclidanthera* is placed in Polygalaceae [6, 7, 14], and *Lissocarpa* is the closest relative of Ebenaceae [15]. *Afrostyrax* was once included in the genus *Styrax* [16], but was later reclassified into Huaceae [6, 7, 14,17]. According to taxonomic revisions, *Pamphilia* was classified into *Styrax* [18]. Fritsch [19] combined *Foveolaria* into *Styrax* by implementing morphological phylogenetic analyses. In addition, two new genera have been established in the family Styracaceae: (1) Chen [20] segregated *Sinojackia dolichocarpa* as a new monotypic genus *Changiostyrax*, and (2) according to morphological and DNA sequences, *Halesia macgregorii* was removed from *Halesia* to become a new genus, *Perkinsiodendron* P.W. Fritsch [21].

Although the phylogenetic placement of the family has been resolved, there are few phylogenetic studies above the genus level, with the phylogenetic relationships between genera remaining ambiguous. In the phylogeny of Ericales based on the single chloroplast gene *rbcL* [22], the results showed that *Styrax* and *Clethra* Gronov.ex L. (Clethraceae) were clustered in a clade, while *Halesia*, *Rehderodendron*, and *Sinojackia* formed a clade that was sister to *Diapensia* L. and *Galax* Rafin. (Diapensiaceae). Therefore, Styracaceae was considered to be polyphyletic. However, this conclusion does not always hold true. Olmstead et al. [23] inferred the phylogeny of Asteridae based on multiple genes, including *Styrax* and *Halesia*, which formed a strongly supported sister-group relationship. Albach et al. [24] came to the same conclusion as Olmstead et al. [23] when analyzing molecular evolution of four genes within the Asterids. In addition, the phylogeny of Styracaceae based on morphology

plus three DNA sequences (chloroplast *trnL intron*/*trnL-trnF* spacer and *rbcL* with the nuclear ribosomal DNA region ITS) recovered a monophyletic relationship of Styracaceae [1]. Early molecular phylogenetic studies of Styracaceae based on morphological characters and three DNA loci showed that *Pterostyrax* and *Halesia* were not supported as monophyletic, *Styrax* and *Huodendron* formed a clade that was sister to a clade of *Alniphyllum* and *Bruinsmia*, and a sister relationship was found between *Halesia macgregorii* and *Rehderodendron macrocarpum* [1]. Based on ITS, the plastid *psbA-trnH* intergenic spacer, and microsatellite data, Yao et al. [25] recovered *Sinojackia* as monophyletic and reported a similar topology as Fritsch et al. [1] with weak support for six genera within Styracaceae. Yan et al. [26] conducted phylogenetic analyses of the Styracaceae based on 19 chloroplast genomes. The results showed that *Styrax* was monophyletic, while *Huodendron*, *Alniphyllum* and *Bruinsmia* clustered in a clade with an approximate 20-kb inversion in the Large Single-Copy (LSC) region. The tree species *Pterostyrax* were not supported as monophyletic, however, *Halesia carolina* L and *Pterostyrax hispidus* Siebold & Zucc formed a clade that was sister to the remainder of the family. Additionally, the systematic positions of *Halesia* and *Pterostyrax* have not yet been fully resolved.

The chloroplast genomes of most angiosperms are of matrilineal inheritance. The rate of evolution of genes in the chloroplast is relatively slow, but the rates of evolution across different regions of the genome have observed differences, which can be applied to phylogenentic studies of various taxonomic scales. Due to a conserved structure, small effective population size, and lack of recombination, chloroplast genomes have been extensively used to infer phylogenetic relationships and histories [27, 28, 29]. With the advent of next-generation sequencing (NGS) technologies, whole-plastome sequencing has become cheaper and faster than ever before. As a result, whole-plastome sequence data have recently been employed to generate highly resolved phylogenies or to efficiently barcode and identify plant species, especially in taxonomically complex groups [30, 31, 32]. Moreover, previous studies have uncovered signatures of natural (purifying or positive/adaptive) selection in some plastome gene regions (e.g. *psbA, matK, rbcL*) which encode proteins

involved in photosynthesis [33, 34, 35].

Despite progress in understanding the Styracaceae phylogeny, most advances have been based on relatively limited molecular and/or morphological data. Only one study has examined Styracaceae phylogeny using plastome-scale data [26], but this study employed only 19 taxa and included only one or two accessions per genus. Here, we increased samples for each genus, especially *Sinojackia* (five accessions) and *Styrax* (seven accessions). We analyzed 28 complete plastomes for resolving the broader phylogeny of Styracaceae. Compared with phylogenetic studies limited to a few complete plastomes or a few plastid loci, plastome phylogenomic studies provide potentially much greater resolution and support. The objectives of this study are: 1) infer the plastome structural evolution of Styracaceae, 2) elucidate Styracaceae plastome diversity and evolution, 3) use selective pressure analysis to test for the presence of adaptive evolution in all genes located in the two single-copy regions and one of the two Inverted Repeat (IR) region regions.

Methods

Plant Samples, DNA Extraction and Sequencing

We collected 28 plastomes of Styracaceae, including 12 new Styracaceae plastomes, and 16 previously sequenced plastomes of Styracaceae (Table 1), with representatives from 11 of the 12 genera described by APG IV [36]. We used *Symplocos ovatilobata* Noot (Symplocaceae), *Stewartia monadelpha* Siebold et Zucc, and *Stewartia sinii* (Y. C. Wu) Sealy (Theaceae) as outgroups. A total of 31 sequences were analyzed in total. Our field collection followed the ethics and legality of the local government and was permitted by the government. The formal identification of the plant material was undertaken by Guowen Xie, and voucher herbarium specimens were deposited at the Institute of Tropical Agriculture and Forestry (HUTB), Hainan University, Haikou, China.

Total genomic DNA was extracted from dried leaf tissue using cetyltrimethyl ammonium bromide (CTAB) protocol of Doyle and Doyle [37]. The genomic DNA of each sample was quantified and analyzed with an Agilent BioAnalyzer 2100 (Agilent Technologies). After extracting genomic DNA, approximately 0.8 μg of DNA for was used for library construction with BGI's kit (Embro-seq PGS Kit) developed by the Beijing Genomics Institute. Libraries were sequenced using the BGISEQ-500

platform at BGI Shenzhen, China and produced approximately eight Gb of high-quality reads per sample with 100 bp paired-end reads. Raw reads were trimmed using SOAPfilter-v2.2 with the following criteria (1) reads with >10 percent base of N; (2) reads with > 40 percent of low quality (value <= 10); (3) reads contaminated by adaptor and produced by PCR duplication.

Genome assembly and annotation

Sequenced reads were used to assemble plastomes with MITObim-v1.8 [38]. Plastomes of related species were used as templates for plastome assembly (Table 2). Assembled plastomes were annotated using Geneious-R11.0.5 (Biomatters Ltd., Auckland, New Zealand) and Dual Organellar GenoMa Annotator (DOGMA) [39], with further manual corrections for the start/stop codons and intron/exon boundaries. The assembly was ordered using BLAST and aligned (> 90% similarity and query coverage) according to the reference chloroplast genome (Table 2). In addition, tRNAscan-SE1.21 was used to further verify all of the tRNA genes. We also re-annotated the sequences downloaded from previously assembled plastomes before using them in our analyses. The 12 newly generated complete plastome sequences were deposited in GenBank (accession numbers in Tables 1 and 2)

Genome comparative and structural analyses

Graphical maps of Dipsacales plastomes were drawn using OrganellarGenome DRAW (ORDRAW) [40], with subsequent manual editing. Genome comparisons across the 26 (removing two homologous sequences) Styracaceae species were performed in Shuffle-LAGAN mode on the mVISTA program [41], using the annotation of *Pterostyrax hispidus* Siebold & Zucc as a reference. To evaluate whether different chloroplast genome regions underwent different evolution patterns and to explore highly variable regions for future population genetic and species identification studies, we sequentially extracted both coding regions and noncoding regions (including intergenic spacers and introns) after alignment using MAFFT v7 [42] under the criteria that the aligned length was >200 bp and at least one mutation

site was present. Finally, nucleotide variability of these regions was evaluated with DNASP V5.10 [43].

Selective pressure analysis

The analyses of selective pressures were conducted along the phylogenetic tree of Styracaceae (see below) for each plastid gene located in the Large Single-Copy (LSC) region, Inverted Repeat (IR) region and Small Single-Copy (SSC) region regions. Non-synonymous (dN) and synonymous (dS) substitution rates of each plastid gene were calculated using the yn00 program in PAML v4.9 [44]. In addition, we used the CODEML program in PAML to detect signatures of natural selection among specific lineages. Genes were considered to be under positive/negative selection at a certain clade when its ω value from the two-ratio model was higher/lower than 1 (neutral selection). To avoid potential convergence biases, genes with too few mutations (Pi < 0.001) were filtered out from selective pressure analysis. To determine if the genes are under selection, we conducted the selection analysis of the exons of each protein-coding gene by using CODEML program in PAML.

Phylogenetic analyses

Phylogenetic analyses were conducted on the 31 plastomes, using *Symplocos ovatilobata, Stewartia sinii,* and *S. monadelpha* as outgroups. Chloroplast sequences were aligned using MAFFT v7.037 [42]. In order to evaluate possible alternative hypotheses of phylogeny, topologies were constructed by both maximum likelihood (ML) and Bayesian inference (BI) methods using not only the complete genome sequences, but also included seven data sets (i.e. LSC, SSC, IR, coding, non-coding, combination of LSC+SSC, and concatenation of LSC+SSC+one IR) for analysis. The best-fitting models of nucleotide substitutions were determined by the Akaike Information Criterion (AIC) in Modeltest 3.7 [45], so then we can determine the best-fitting model for these data sets (Table 4). However, for the coding data set, Partitionfinder-2.1.1 [46] was used to select the best-fit partitioning scheme of all 79 possible gene-by-codon position partitions under three models (GTR, GTR + G, GTR

+ I + G); GTR + I + G was recovered as having lowest AICc score. The best partition scheme from PartitionFinder was used in MrBayes and BEAST, with each partition having its own GTR + I + G model (GTR + G model in RAxML).

Maximum likelihood analyses were conducted using RAXML-HPC v8.2.8 [47] with 1000 bootstrap replicates on the CIPRES Science Gateway website (Miller et al., 2010) with the GTR+I + G substitution model. Bayesian inference (BI) analyses were performed in MrBayes v3.2[48]. Based on the calculated model, BI analyses were set up identically for the all data sets at the CIPRES Science Gateway website [49], except that 50,000,000 generations were used.

Result

Plastome Structure of Styracaceae

In this study, the plastomes of Styracaceae species and two other members of the family (*Symplocos* and *Stewartia*) displayed a typical quadripartite structure and similar lengths, containing a pair of inverted repeat IR regions (IRa and IRb), one large single-copy (LSC) region, and one small single-copy (SSC) region (Table 3). Plastome sizes ranged from 155,185 bp (*Alniphyllum pterospermu* Matsum) to 158,879 bp (*Pterostyrax hispidus*) with a maximum read depth of at least 40× for each plastome. The genomes were composed of an LSC region (ranging from 83,200 bp to 88,258 bp), SSC region (ranging from 17,556 bp to 19, 235 bp), and two IR copies (ranging from 24,243 bp to 26,761 bp)(Tab. 3). Their overall GC content was nearly identical (36.70-37.40%). In all species, the GC content of the LSC and SSC regions (about 35% and 30%) were lower than those of the IR regions (about 43%). The 31 plastomes encoded 113 genes, including 79 protein-coding genes, 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes.

Comparison of the genome structures among Styracaceae, revealed an inversion of a large segment spanning *trnQ-UUG* to *rpoB* (20-kb) in the LSC region of *Alniphyllum fortunei* (Hemsl.) Makino (Fig 2). This inversion was also shown in *A. eberhardtii* Guill, *A. pterospermum* Matsum*, Bruinsmia polysperma* (C. B. Clarke)

Steenis and *B. styracoides* Boerl. & Koord, suggesting that the inversion is common to *Bruinsmia* and *Alniphyllum*. The large 20-kb inversion is the same as the normal genome in gene composition and relative position.

Comparative genomic analysis and divergence hotspot regions

To investigate the levels of sequence divergence, the 28 Styracaceae plastomes were plotted using mVISTA, with *Pterostyrax hispidus* as the reference (Fig.3). The sequence divergence was low among all plastomes. Notably, the proportion of variability in coding regions and inverted repeats (IRs) showed higher conservation than non-coding and small single-copy (SSC) regions. The mutation rate of *ycf1* was the highest observed. The variation rates of *Styrax* and *Huodendron* in the large and small single copy regions were higher than other species, and the consistency of *Huodendron* in *clpP* intron was lower than 50%.

Nucleotide diversity analyses showed that the proportion of variable sites in noncoding region was higher than that in coding region, and the region with the greatest diversity change was located in the intergenic spacer region (Fig. 4). Among all 209 loci (79 coding genes and 130 non-coding regions), nucleotide diversity values of coding genes ranged from 0.001(*rpl23*) to 0.156 (*atpH*), 4 loci were greater than 0.1 (*psbK*, *psbI*, *rpoC2*, *atpH*). Nucleotide diversity of non-coding gene ranged from 0 (*ycf1*, *ycf2*, *rpoC1-rpoB*, *psaB-psaA*, *psbF-psbE*, *rps3-rpl22*, *rpl2-rpl23*, *rps7-rps12*, *trnA* (UGC)-*rrn23*, *ndhH-ndhA*, *orf42-trnA-UGC*, *ycf2-ycf15)* to 0.385 (*trnI* intron1). Seven of these loci possessed values >0.15: e.g. *atpF* intron (0.151), *clpP* intron1 (0.151), *rps*2-*rpoC*2 (0.151), *trnG*(*GCC*)-*trnR*(*UCU*) (0.158), *rps*12-*clpP* (0.159), *atpH-atpI* (0.166), *trnI*(*GAU*) intron1 (0.385) (Fig.4).

Selective pressures in plastome evolution of Styracaceae

The results showed that the 79 protein coding genes mainly possessed synonymous substitutions (Fig.5). In addition, *rps12* (0.8874), *rps19* (0.5076) and *rps11* (0.4466) had the highest synonymous substitution rate. The highest rate of non-synonymous substitution is *ycf1* (1.016), followed by *rps12* (0.751). The rate of

non-synonymous substitutions in other genes was low, in which the rate of non-synonymous substitution of *psb* was the lowest, and the non-synonymous substitution of *psbL*, *psbH*, *psbN*, *psbI* and *psbT* was zero. Among the 79 protein coding genes of Styracaceae, there are 7 genes with ω value greater than 1, which are *rps4* (1.087), *rpl23* (1.126), *accD* (1.839), *rpoC1* (1.990), *psaA* (2.175), *rpoA* (1.578) and *ndhH* (3.459). These seven genes are under positive selection, including the NADH dehydrogenase gene (*ndhh*), ribosome protein coding gene (*rps4* & *rpl23*), RNA polymerase gene (*rpoC1* & *rpoA*), photosynthetic gene (*psa*) and an additional protein gene (*accD*) (Fig.6)

Phylogenetic analyses

The ML and BI analyses produced similar topologies over all gene data set. Thus, the data were combined into a single overall analysis (Fig. 7). Characteristics of all data sets are given in Table 2. Both ML and BI analyses of the 31 plastomes generated almost identical topologies except for the different position of *Sinojackia sarcocarpa* L. Q. Luo and *Changiostyrax dolichocarpus* (C. J. Qi) Tao Chen in the IR regions (Fig. 7). In all data sets, the Styracaceae were strongly supported as monophyletic (BS=100%, PP=1). All species of *Styrax* form a clade sister to the rest of the family. The second branch is *Huodendron*, followed by two genera with unique plastome reversal structures, *Alniphyllum* and *Bruinsmia*, which has the longest branches in the unequal branch evolutionary tree. *Halesia diptera* did not cluster with *Perkinsiodendron* but formed a sister relationship with all the remaining species with strong support, while *Perkinsiodendron* and *Rehderodendron* form a clade. In the Large Single-Copy (LSC) region, *Melliodendron* and *Changiostyrax* form a sister relationship with weak support values (BS=25%, PP=0.55). However, *Changiostyrax* appeared as an independent branch with weak support values (BS=25/65/63/69100%, PP=0.67/0.71/1) in the rest of data sets. *Pterostyrax* and *Sinojackia* appear successively sister to the rest of the family with strong support except for the position of *Pterostyrax hispidus*, which was poorly supported in the IR regions (BS=56%, PP=1).

Discussion

Plastome structure comparisons and sequence divergence hotspots

This study included 31 plastomes, 28 representative taxa from 11 genera of Styracaceae, and 3 sequences as outgroup. The plastid size of Styracaceae is within the normal range of angiosperms (120-190kb), and the size, structure, gene sequence and content of the whole family are highly conserved (155,185bp-158,879 bp), which is a typical tetragonal structure [50]. The chloroplast genome of *Alniphyllum fortunei*, which was first reported in this study, found about 20-kb reverse sequence in 14 coding genes from *trnQ-UUG* to *rpoB*. The same inverse sequence has been shown to exist in other members of the genus (*Alniphyllum eberhardtii*), and is not due solely to the gene assembly [51]. Plastid structure is usually conserved in most angiosperms, but large inversions have been detected in many taxa. For example, a 4-kb inverted fragment in the LSC between *rpoB-trnT* was found in *Myriophyllum spicatum* [52], and a large fragment gene inversion was also found in *Lotus japonicas*, *Arabidopsis thaliana* [53] and Oleaceae [54]. Because of the scarcity of inversions, plastid inversions are of great value to the study of genome evolution [55, 56]. Previous studies suggested that gene inversions are closely related to the repetitive sequence, and dispersed repetitive sequences promote inversions through intermolecular recombination [57, 58, 59]. In the comparative analysis of the plastome structure of Styracaceae, we found that the degree of variation of *Styrax* and *Huodendron* is the same, which is consistent with the phylogenetic results of *Styrax* and *Huodendron* which are close relatives [1, 26].

In the sequence divergence analysis, the variation in loci of the noncoding region is higher than those of the coding region, which is similar to previous results of most angiosperms [60, 61, 62]. The results also show that the degree of evolution in the noncoding region is greater than that of coding region, and highly variable noncoding regions are of great value for the study of plant phylogenetics [63, 64]. In addition, the variation rate of the IR region was lower than the two single copy regions. Previous studies have shown that the accumulation of point mutations in the

inverted repeat region is slower than the single copy region [65, 66, 67].

Positive Selection Analysis

In the selection pressure analysis, Styracaceae is dominated by non-synonymous substitutions. A previous study indicated that the rate of non-synonymous substitutions is positively correlated with the degree of variation in the genome, while the rate of synonymous substitution exhibits a weak correlation [68]. There are seven coding genes under positive selection, including five gene types: NADH dehydrogenase gene (*ndhH*), ribosomal protein coding gene (*rps4* & *rpl23*), RNA polymerase gene (*rpoC*1 & *rpoA*), a photosynthetic gene (*psaA*) and one additional protein gene (*accD*). Chloroplast NADH dehydrogenase (NDH) complex participates in the circular electron transport and chlorine respiration around the light system [69], NDH subunits participate in the stability of NDH complexes, especially under high light conditions [70]. However, due to NDH complex existing in low abundance and afragile nature, it is difficult to analyze its function. The plants of Styracaceae are mainly distributed in the tropics and subtropics, which are subjected to high light and high temperature conditions. Ribosomal proteins are a part of the ribosomal complex, which is a translation mechanism, and is essential for the correct production of proteins required for normal cell function. The selection of ribosomal proteins may increase the stability of ribosomal complexes under high light conditions, such as high temperature, which is similar to the selection of *ndh* proteins under high light conditions[71]. However, whether these ribosomal proteins have increased stability than those of the original proteins under strong light or related conditions has not been determined, and further experimental verification is still needed. The gene *rpoC* is in the same operon as *rpoA*, which encodes the β subunit of RNA polymerase. Increasing the *rpoA* & *rpoC* mutations may lead to alterations in cell wall metabolism, possibly as a result of altered transcription [72].

Phylogenetic analyses

Phylogenetic analyses of all data sets strongly support a monophyletic

classification of Styracaceae (BS=100%, PP=1). Our results are consistent with those of the most comprehensive previous phylogenetic studies of the family [1, 23, 24, 26]. According to Fritsch et al. [1] analysis of morphology and three DNA sequence data sets, *Styrax* is a monophyletic, forming a clade with *Huodendron*. However, in our analysis of seven data sets, both using ML and BI, *Styrax* remained monophyletic with high support rate (BS=100%, PP=1) and sister to the remainder of the family, which is consistent with the conclusions of Yan et al [26]. *Alniphyllum* and *Bruinsmia* formed a clade that has the longest branches in the unequal branch evolutionary tree which may be due to the sequence difference caused by the inversion of 20-kb.

In the analyses of Fritsch et al. [1] and Yao et al. [25], *Melliodendron* consistently formed a clade with *Changiostyrax*, whereas in all our data set, except in the LSC data set, *Melliodendron* and *Changiostyrax* are strongly supported as successively sister to all remaining Styracaceae. *Halesia* and *Pterostyrax* have not been previously fully resolved [1, 25, 26]. Here, we collected four accessions of *Pterostyrax* to analyze, with the species *Pterostyrax hispidus* observed as being excluded from the other two species with a relatively low support value (BS=56%, PP=1) in the LSC data set. The four sequences of *Pterostyrax* form a clade with strong support (BS=100%, PP=1) in all other data sets. In previous morphological analyses, the species of *Pterostyrax* form a monophyletic group, however conflicting data from the chloroplast maybe have been caused by homoplasy rather than hybridization [1]. Our study only included one species of *Halesia*, and its systematic relationship needs to be further studied by increasing the sample size or combining with nuclear gene analysis. *Perkinsiodendron* and *Rehderodendron* form a clade in our all data set, with *Perkinsiodendron* being a new genus established from *Halesia macgregorii* Chun based on molecular data and morphological characters [21]. Our results strongly support the establishment of *Halesia macgregori* as *Perkinsiodendron macgregori*. Our study also strongly supports the monophyly of *Sinojackia* in plastid data, as has been detected in previous studies [25], except *Sinojackia sarcocarpa* which appears to be separate from the other species in IR data set. The different topological structure of IR data set may be the result of a slower mutation and evolution rate in the reverse

repeat region compared to that of the single copy region [65, 66, 67, 73]. There are many possible reasons for differences between different data sets in inferring phylogenetic trees, including differences in taxon sampling and biological factors such as hybridization/introgression, incomplete lineage sorting, gene duplication and/or loss, and horizontal gene transfer [74, 75, 76]. In addition, the results of the molecular phylogenies were consistent with those of fruit morphology.

Conclusions

With 28 plastomes of Styracaceae, our study provides new insights to the phylogenetics and plastome evolution of the family. Our results clearly indicate that *Pterostyrax* is monophyletic, and the establishment of *Perkinsiodendron* and *Changiostyrax* are supported. A 20-kb reverse sequence was found in the newly published sequence of *Alniphyllum fortunei*, which confirmed the existence of large inversion sequence in *Alniphyllum* and *Bruinsmia*. Nevertheless, the lack of *Parastyrax* species in the sequence data, necessitates that our results may need to be further verified by increasing taxon sampling or using nuclear genes. There is the possibility that the inclusion of additional genera may alter the topology and/or support values.

Abbreviations

BI: Bayesian Inference; CTAB: Cetyltrimethylammonium bromide; dN: Nonsynonymous; DnaSP: DNA Sequences Polymorphism; dS: synonymous; IR: Inverted repeat; LSC: Large single copy; GTR: General time reversible ML: Maximum Likelihood; PI: Phylogenetic informativeness; rRNA: Ribosomal RNA; SSC: Small single copy; tRNA: Transfer RNA

Author's Contributions

HW, XC, and JL designed this study; XC,HW and JH designed experiments, sequenced chloroplast genomes; ZZ analyzed the data; HW, XC, and JL drafted the manuscript; All authors have read and approved the final

manuscript.

Funding

This research was funded by National Natural Scientific Foundation of China (31660055 and 31660074) and by a start-up fund from Hainan University (kyqd1633).

Declarations

Availability of data and materials

All sequences used in this study are available from the National Center for Biotechnology Information (NCBI) (accession numbers: MT700470- MT700481; see Additional Table 2).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version (Supporting information).

Author details

[a]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, School of Life and Pharmaceutical Sciences, Hainan University, Haikou 570228, China

[b]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14850, USA

References

1. Fritsch, P.W., Morton, C.M., Chen, T., Meldrum, C. Phylogeny and biogeography of the Styracaceae. Int. J. Plant Sci. 2001:162, S95-S116.
2. Dickison, W.C. Floral anatomy of the Styracaceae, including observations on intraovarian trichomes. Bot. J. Linn. Soc. 1993:112, 223–255.
3. Dickison, W.C. A note on the wood anatomy of Dillenia (Dilleniaceae). IA WA Bull. 1979:2 & 3:57-60.
4. Dickison W C, Phend K D . Wood Anatomy of the Styracaceae: Evolutionary and Ecological Considerations. IAWA Journal.1985:6(1):3-22.
5. Morton, C. M. & Dickison, W. C. Comparative pollen morphology of the Styracaceae. Grana 1992:31: 1-15. 19n. Odensc. ISSN 0017-3134
6. Cronquist, A. An Integratcd System of classification of Flowering Plants. NewYork: CDlumbia Univ.Pesrs 1981:492- 506.
7. Takhtajan A. Diversity and classification of flowering plants. New York: Columbia University Press.1997.
8. Thorne RF. Classification and geography of the flowering plants [J]. Botanical Review, 1992:58(3): 225-348.
9. Voss EG. An integrated system of classification of flowering plants-Cronquist, A [J].Economic Botany, 1983:37(4): 498-498.
10. Copeland H F. The Kingdom of Organisms. Quarterly Review of Biology, 1938:13:383.
11. APG. An ordinal classification for the families of flowering plants. Ann. Mo. Bot. Gard. 1998:85 (4), 531–553. http://dx.doi.org/10.2307/2992015.
12. Bentham G, JD Hooker. Genera plantarum. Vol 2.LovellReeve, London. 1279 pp. 1873.
13. Perkins J. Styracaceae. In A Engler, ed. Pflanzenreich IV, 241 (Heft 30). Engelmann, Leipzig.1907.
14. Thorne, R.F. The classification and geography of the flowering plants: dicotyledons of the class angiospermae (subclasses Magnoliidae, Ranunculidae,

Caryophylliidae, Dilleniidae, Rosidae, Asteridae, and Lamiidae). Bot. Rev. 2000:66, 441–647.
15. Franceschi, D. de. Phylogenie des Ebenales: analyse de l'ordre et origine biogeographique des especes indiennes. Pub. Dept. Ecol. Institut Franais de Pondicherry1993:33: 1-153.
16. HutehinsonJ. The Genera of Flowering Plant.Oxford, 1967:2:34 一 39
17. Baas P. Anatomical contributions to plant taxonomy. 2. The affinities of Hua Pierre and Afrostyrax Perkins et Gilg. Blumea 1972:20:161–192.
18. Wallnöfer B. A revision of Styrax L. section Pamphilia (Mart. ex A.DC.) B.Walln.(Styracaceae). Annalen des Naturhistorischen Museums in Wien. Serie B für Botanikund Zoologie, 1997:99B: 681-720.
19. Fritsch PW. Phylogeny of Styrax based on morphological characters, with implications for biogeography and infrageneric classification. Systematic Botany, 1999:24(3):356-378.
20. Chen CT. Changiostyrax, a new genus of Styracaceae from China. Guihaia, 1995:15:289-292.
21. Fritsch, P.W., Yao, X., Simison, W.B., Cruz, B.C., Chen, T. Perkinsiodendron, a new genus in the Styracaceae based on morphology and DNA sequences. J. Bot. Res. I.Tex. 2016:10, 109–117.
22. Morton, C. M. , Chase, M. W. , & Swensen, K. A. K. M. A molecular evaluation of the monophyly of the order ebenales based upon rbcl sequence data. Systematic Botany, 1996:21(4), 567-586.
23. Olmstead, R.G., Kim, K.J., Jansen, R.K., Wagstaff, S.J. The phylogeny of the Asteridae sensu lato based on chloroplast ndhF gene sequences. Mol. Phylogenet. Evol. 2000:16, 96–112.
24. Albach, D.C., Soltis, P.S., Soltis, D.E., Olmstead, R.G. Phylogenetic analysis of Asterids based on sequences of four genes. Ann. Mo. Bot. Gard. 2001:88, 163–212. http://dx.doi.org/10.2307/2666224.
25. Yao, X.H., Ye, Q., Fritsch, P.W., Cruz, B.C., Huang, H. Phylogeny of Sinojackia (Styracaceae) based on DNA sequence and microsatellite data: implications for

taxonomy and conservation. Ann. Bot. 2008:101, 651–659.
26. Yan M , Fritsch P W , Moore M J , et al.Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. Molecular Phylogenetics & Evolution, 2018:121:198–211
27. Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., Morris, A.B. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am. J. Bot. 2014:101, 1987-2004.
28. Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots Proc. Natl. Acad. Sci. USA. 2010:107, 4623-4628.
29. Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebensmack, J.,Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.B., Peery, R., McNeal, J.R., Kuehl, J.V., Boore, J.L. Analysis of 81 genes from 64 plastomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. USA. 2007:104, 19369-19374.
30. Barrett, C.F., Specht, C.D., Leebens-Mack, J., Stevenson, D.W., Zomlefer, W.B., Davis, J.I. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? Ann. Bot. 2014:113, 119-133.
31. Malé, P.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Res. 2014:14, 966-975.
32. Yu, X.Q., Gao, L.M., Soltis, D.E., Soltis, P.S., Yang, J.B., Fang, L., Yang, S.X., Li, D.Z. Insights into the historical assembly of East Asian subtropical evergreen broadleaved forests revealed by the temporal history of the tea family.

555 New Phytol. 2017:215, 1235-1248.
556 33. Allen, J.F., de Paula, W. B., Puthiyaveetil, S., Nield, J. A structural phylogenetic
557 map for chloroplast photosynthesis. Trends Plant Sci. 2011:16, 645-655.
558 34. Carbonell-Caballero, J., Alonso, R., Iba, V.E., Terol, J., Talon, M., Dopazo, J. A
559 phylogenetic analysis of 34 chloroplast genomes elucidates the relationships
560 between wild and domestic species within the genus Citrus. Mol. Biol. Evol.2015:
561 32, 2015-2035.
562 35. Hu, S.L., Sablok, G., Wang, B., Qu, D., Barbaro, E., Viola, R., Li, M.A., Varotto,
563 C.Plastome organization and evolution of chloroplast genes in Cardamine
564 species adapted to contrasting habitats. BMC Genom. 2015:16, 306.
565 36. APG IV. An update of the Angiosperm Phylogeny Group classification for the
566 orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 2016:181,
567 1–20. http://dx.doi. org/10.1111/boj.12385.
568 37. Doyle J.J., Doyle J.L. A rapid DNA isolation procedure for small quantities of
569 fresh leaf tissue. Phytochemical Bulletin, 1987:19: 11-15.
570 38. Hahn, C., Bachmann, L., Chevreux, B. Reconstructing mitochondrial
571 genomes directly from genomic next-generation sequencing reads—a baiting
572 and iterative mapping approach. Nucleic Acids Research 2013:41, e129.
573 39. Wyman S.K., Jansen R.K., Boore J.L. Automatic annotation of organellar
574 genomes with DOGMA. Bioinformatics, 2004:20: 3252-3255.
575 40. Lohse, M., Drechsel, O., Kahlau, S., Bock, R. Organellar GenomeDRAW-a suite
576 of tools for generating physical maps of plastid and mitochondrial genomes and
577 visualizing expression data sets. Nucleic Acids Research 2013:41, W575.
578 41. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I.Vista:
579 computational tools for comparative genomics. Nucleic Acids Research
580 2004:32, 273–279.
581 42. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7:
582 improvements in performance and usability. Molecular Biology and Evolution,
583 2013:30: 772-780.

584 43. Librado P., Rozas J. DnaSP v5: a software for comprehensive analysis of DNA
585 polymorphism data. Bioinformatics. 2009:25: 1451-1452.
586 44. Yang Z.H. PAML 4: phylogenetic analysis by maximum likelihood. Molecular
587 Biology and Evolution, 2007:24: 1586-1591.
588 45. Posada, D., Crandall, K.A. Modeltest: testing the model of DNA substitution.
589 Bioinformatics 1998:14 (9), 817-818.
590 46. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B.
591 PartitionFinder 2: new methods for selecting partitioned models of evolution
592 for molecular and morphological phylogenetic analyses. Molecular Biology
593 and Evolution 2016:34, 772–773.
594 47. Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and
595 post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
596 48. Ronquist, F., Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference
597 under mixed models. Bioinformatics 2003:19, 1572–1574.
598 49. Miller, M.A., Pfeiffer, W., Schwartz, T. Creating the CIPRES Science
599 Gateway for inference of large phylogenetic trees. Gateway Computing
600 Environments Workshop (GCE). 2010:14, 1–8.
601 50. Jansen R.K., Ruhlman T.A. Plastomes of seed plants. Sharkey T.D. (Eds.).
602 Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and
603 Respiration, Vol. 35. 2012:103-126.
604 51. Yan, M.H., Moore, M.J., Meng, A.P., Yao, X.H., Wang, H.C. The first complete
605 plastome sequence of the basal asterid family Styracaceae (Ericales) reveals a
606 large inversion. Plant Syst. Evol. 2017:303, 61–70.
607 http://dx.doi.org/10.1007/s00606-016-1352-0.
608 52. Yi TEL ing Liao Liao, Liu Y , Liu X , et al. The complete chloroplast genome of
609 Myriophyllum spicatum reveals a 4‐kb inversion and new insights regarding
610 plastome evolution in Haloragaceae. Ecology and Evolution, 2020:10(6).
611 53. Tomohiko K , Takakazu K , Shusei S , et al. Complete Structure of the Chloroplast
612 Genome of a Legume, Lotus japonicus. Dna Research. 2000:(6):6.
613 54. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast

614 genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping
615 inversions. Molec Biol Evol. 2007:24:1161–1180.
616 55. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid
617 genome sequence of the chickpea (Cicer arietinum) and the phylogenetic
618 distribution of rps12 and clpP intron losses among legumes (Leguminosae). Molec
619 Phylogen Evol 2008:48:1204–1217. doi:10.1016/j.ympev.
620 56. Johansson JT. There large inversions in the chloroplast genomes and one loss of
621 the chloroplast gene rps16 suggest an early evolutionary split in the genus Adonis
622 (Ranunculaceae). Pl Syst Evol. 1999:218:133–143.
623 57. Palmer JD. Plastid chromosomes: structure and evolution. In: Vasil IK, Bogorad L,
624 editors. Cell culture and somatic cell genetics in plants. Vol. 7A. The molecular
625 biology of plastids. San Diego: Academic Press. p. 1991:5–53.
626 58. Pombert JF, Lemieux C, Turmel M. The complete chloroplast DNA sequence of
627 the green alga Oltmannsiellopsis viridis reveals a distinctive quadripartite
628 architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol.
629 2006:4:3.
630 59. Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry R, editor.
631 Diversity and evolution of plants—genotypic and phenotypic variation in higher
632 plants. Oxfordshire: CABI Publishing. p.2005:45–68.
633 60. Huang H., Shi C., Liu Y., et al. Thirteen Camellia chloroplast genome sequences
634 determined by high-throughput sequencing: genome structure and phylogenetic
635 relationships. BMC Evolutionary Biology. 2014:14: 151.
636 61. Jansen R.K., Ruhlman T.A. Plastomes of seed plants. Sharkey T.D. (Eds.).
637 Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and
638 Respiration, Vol. 35. Dordrecht, Netherlands: Springer: 2012:103-126.
639 62. Perry A.S., Wolfe K.H. Nucleotide substitution rates in legume chloroplast DNA
640 depend on the presence of the inverted repeat. Journal of Molecular Evolution,
641 2002:55: 501-508.
642 63. Joey Shaw, Edgar B. Lickey, Edward E. Schilling, and Randall L. Small.
643 Comparison of whole chloroplast genome sequences to choose noncoding regions

644 for phylogenetic studies in angiosperms: the tortoise and the hare III. American
645 Journal of Botany. 2007:94(3):p.275-288.
646 64. J., Shaw, R., L., & Small. Chloroplast dna phylogeny and phylogeography of the
647 north american plums (prunus subgenus prunus section prunocerasus, rosaceae).
648 American Journal of Botany. 2005:92: 2011–2030.
649 65. Curtis, SE, Clegg MT. Molecular evolution of chloroplastDNA sequences.
650 Molecular Biology and Evolution 1984:1: 291–301.
651 66. G AUT , B. S. Molecular clocks and nucleotide substitution rates in higher
652 plants. Evolutionary biology, vol.1998: 30, 93–120.
653 67. Wolfe, K. H.，Li, W. H.，Sharp, P. M. Rates of nucleotide substitution vary greatly
654 among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the
655 National Academy of Sciences, USA 1987:84: 9054–9058.
656 68. Weng ML, Blazier JC, Govindu M, et al. Reconstruction of the ancestral plastid
657 genome in Geraniaceae reveals a correlation between genome rearrangements,
658 repeats, and nucleotide substitution rates. Molecular Biology and Evolution,
659 2014:31(3): 645-659.
660 69. Sazanov L A, Burrows P A , Nixon P J. The plastid ndh genes code for an
661 NADH-specific dehydrogenase: Isolation of a complex I analogue from pea
662 thylakoid membranes. Proceedings of the National Academy of Sciences,
663 1998:95(3):1319-1324.
664 70. Fan X, Zhang J, Li W, et al. The NdhV subunit is required to stabilize the
665 chloroplast NADH dehydrogenase-like complex in Arabidopsis. The Plant Journal,
666 2015:82(2):221-231.
667 71. Liqiang, Wang, Hui. Complete plastome sequence of Iodes cirrhosa Turcz. the
668 first in the Icacinaceae, comparative genomic analyses and possible split of Idoes
669 species in response to climate changes. Peerj.2019
67072. Bisson, G.P. et al. Upregulation of the phthiocerol dimycocerosate biosynthetic
671 pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis. J. Bacteriol.
672 2012:194, 6441–6452.
67373. Wolfe, K. H. Protein-coding genes in chloroplast DNA:compilation of nucleotide

674 sequences, data base entries, and rates of molecular evolution. In L. Bogorad and I. K.
675 Vasil [eds.], Cell culture and somatic cell genetics of plants, vol. 1991:7B, 467–482.
67674. Naciri, Y., Linder, P. Species delimitation and relationships: The dance of the seven
677 veils. Taxon 2015:64, 3-16.
67875. Nicola, M.V., Johnson, L.A., Pozner, R. Unraveling patterns and processes of
679 diversification in the South Andean-Patagonian Nassauvia subgenus
680 Strongyloma (Asteraceae, Nassauvieae). Mol. Phylogenet. Evol. 2019:136,
681 164-182.
682 76. Lin, H.Y., Hao, Y.J., Li, J.H., Fu, C.X., Pamela, S.S., Douglas, E.S., Zhao, Y.P.
683 Phylogenomic conflict resulting from ancient introgression following 24 species
684 diversification in Stewartia s.l. (Theaceae). Mol. Phylogenet. Evol. 2019:135,
685 1-11.
686
687

1. Figure legends

Fig.1 Fruit morphology of Styracaceae. A, *Halesia diptera* B, *Perkinsiodendron macgregorii* C, *Rehderodendron kwangtungense* D, *Pterostyrax psilophyllus* E, *Melliodendron xylocarpum* F, *Alniphyllum fortunei* G, *Changiostyrax dolichocarpa* H, *Sinojackia xylocarpa* I, *Bruinsmia polysperma* J, *Huodendron tibeticum* K,*Styrax macrocarpus*.

Fig. 2 Gene map of the Styracaceae plastome. (A) The inverted order of genes in *Alniphyllum fortunei*; (B) The corresponding region of *Styrax confusus*.

Fig. 3. Visualization of the alignment of 26 Styracaceae plastome sequences. The plastome of *Pterostyrax hispidus* was used as the reference. The Y-axis depicts percent identity to the reference genome (50-100%) and the X-axis depicts sequence coordinates within the plastome. Genome regions were color-coded according to coding and non-coding regions.

Fig. 4. Comparison of the nucleotide diversity (Pi) values across 28 Styracaceae plastomes. (A) Protein-coding regions. (B) Non-coding regions.

Fig. 5 Synonymous (dS) and nonsynonymous (dN) substitution rates of the protein coding genes.

Fig. 6 ω (dN/dS) values of genes in plastomes of the Styracaceae. The red line represents neutral selection, while values above one represents positive/adaptative selection, and values below one represents negative/purifying selection.

Fig. 7 Optimal phylogenetic tree resulting from analyses of 79 protein-coding genes using Maximum Likelihood (ML). Bayesian inference (BI) topology is the same as ML. Support values above the branches are maximum likelihood bootstrap support/Bayesian posterior probability; asterisks indicate 100%/1.0 support values. The genera of Styracaceae are indicated by different colors, which correspond to branch colors. The inset shows the same tree as a phylogram.

716 Fig. S1 Bayesian inference (BI) and Maximum likelihood (ML) phylogram of
717 Styracaceae based on LSC regions, with ambiguous sites excluded from analysis.
718 Numbers associated with branches are ML bootstrap values and Bayesian posterior
719 probabilities, respectively. Asterisks indicate 100% bootstrap support or 1.0 posterior
720 probability whereas hyphens indicate the bootstrap support or posterior probability
721 <50% or 0.5. Clade designations are described in the text.
722
723 Fig. S2 Bayesian inference (BI) and Maximum likelihood (ML) phylogram of
724 Styracaceae based on SSC regions, with ambiguous sites excluded from analysis.
725 Numbers associated with branches are ML bootstrap values and Bayesian posterior
726 probabilities, respectively. Asterisks indicate 100% bootstrap support or 1.0 posterior
727 probability whereas hyphens indicate the bootstrap support or posterior probability
728 <50% or 0.5. Clade designations are described in the text.
729
730 Fig. S3 Bayesian inference (BI) and Maximum like- lihood (ML) phylogram of
731 Styracaceae based on IR regions, with ambiguous sites excluded from analysis.
732 Numbers associated with branches are ML bootstrap values and Bayesian posterior
733 probabilities, respectively. Asterisks indicate 100% bootstrap support or 1.0 posterior
734 probability whereas hyphens indicate the bootstrap support or posterior probability
735 <50% or 0.5. Clade designations are described in the text.
736
737 Fig. S4 Bayesian inference (BI) and Maximum like-lihood (ML) phylogram of
738 Styracaceae based on complete plastome sequences, with ambiguous sites excluded
739 from analysis. Numbers associated with branches are ML boot- strap values and
740 Bayesian posterior probabilities, respec- tively. Asterisks indicate 100% bootstrap
741 support or 1.0 posterior probability whereas hyphens indicate the bootstrap support or
742 posterior probability <50% or 0.5. Clade designations are described in the text.
743
744 Fig. S5 Bayesian inference (BI) and Maximum likelihood (ML) phylogram of
745 Styracaceae based on plastome LSC+SSC regions, with ambiguous sites excluded

from analysis. Numbers associated with branches are ML bootstrap values and Bayesian posterior probabilities, respectively. Asterisks indicate 100% bootstrap support or 1.0 posterior probability whereas hyphens indicate the bootstrap support or posterior probability <50% or 0.5. Clade designations are described in the text.

Fig. S6 Bayesian inference (BI) and Maximum likelihood (ML) phylogram of Styracaceae based on plastome noncoding regions, with ambiguous sites excluded from analysis. Numbers associated with branches are ML bootstrap values and Bayesian posterior probabilities, respectively. Asterisks indicate 100% bootstrap support or 1.0 posterior probability whereas hyphens indicate the bootstrap support or posterior probability <50% or 0.5. Clade designations are described in the text.

759 **Table 1** Plant collection information and GenBank accession numbers for plastomes of Styracaceae and outgroups included in this
760 study

Family	Species name	Specimen collection and voucher specimen	Locality	Accession number
Styracaceae	*Alniphyllum eberhardtii*	Yan M.H. 201,401 (HIB)	Kunming Institute of Botany,China	NC_031892_1
Styracaceae	*Alniphyllum fortunei*	HUTB LC	Lushan Mountain, Jiujiang, Jiangxi	MT700470
Styracaceae	*Styrax grandiflorus*	NA	Yunnan, China	NC_030539_1
Styracaceae	*Alniphyllum pterospermum*	NA	Wuhan,Hubei,China	NC_041126_1
Styracaceae	*Bruinsmia polysperma*	Wang Hong 9805 (HIB)	Pu'er, Jinggu County, Yunnan, China	NC_030180_1
Styracaceae	*Bruinsmia styracoides*	P.W. Fritsch 1886 (CAS)	Sabah, Malaysia	NC_041137_1
Styracaceae	*Changiostyrax dolichocarpa*	HUTB SZ1	Hupingshan,Hunan,China	MT700471
Styracaceae	*Changiostyrax dolichocarpa*	HUTB SZ2	Hupingshan,Hunan,China	MT700472
Styracaceae	*Halesia diptera*	P.W. Fritsch 1975 (CAS)	University of California Botanical Garden, California,	NC_041128_1
Styracaceae	*Halesia_carolina*	P.W. Fritsch 1974 (CAS)	University of California Botanical Garden, California,	NC_041127_1
Styracaceae	*Huodendron biaristatum*	Yan M.H. 201,403 (HIB)	Wuhan Botanical Garden, Hubei, China	NC_041132_1
Styracaceae	*Melliodendron xylocarpum*	YXQ138	NA	MF179500_1
Styracaceae	*Perkinsiodendron macgregorii*	Zhao C.X. 201,401 (HIB)	Nanyue Arboretum, Hunan, China	MG719841_1
Styracaceae	*Pterostyrax corymbosus*	Yan M.H. 201,405 (HIB)	Wuhan Botanical Garden, Hubei, China	NC_041134_1
Styracaceae	*Pterostyrax hispidus*	P.W. Fritsch 1970 (CAS)	Quarryhill Botanical Garden, California, U.S.A.	NC_041135_1
Sstyracaceae	*Pterostyrax psilophyllus*	Yan M.H. 201,406 (HIB)	Wuhan Botanical Garden, Hubei, China	NC_041133_1
Styracaceae	*Rehderodendron macrocarpum*	Zhao C.X. 201,402 (HIB)	Nanyue Arboretum, Hunan, China	NC_041139_1
Styracaceae	*Sinojackia microcarpa*	HUTB B274	Jiande,Zhejiang, China	MT700474
Styracaceae	*Sinojackia rehderiana*	HUTB PZ13	Pengze, Jiangxi,China	MT700475
Styracaceae	*Sinojackia sarcocarpa*	HUTB B242	Leshan, Sichuan,China	MT700476
Styracaceae	*Sinojackia sarcocarpa*	HUTB B243	Sichuan Normal University,China	MT700477

Styracaceae	*Sinojackia xylocarpa*	HUTB NJ	Nanjing, Botanical, Garden, Jiangsu,China	MT700481
Theaceae	*Stewartia monadelpha*	S. Sakaguchi s. n	Nara, Kinki, Japan	NC_041468_1
Theaceae	*Stewartia sinii*	H. Y. Lin 16105	Jinxiu Co., Guangxi, China	NC_041470_1
Styracaceae	*Styrax confusus*	HUTB SS	Lushan Mountain, Jiujiang, Jiangxi	MT700478
Styracaceae	*Styrax faberi*	HUTB B197	Lushan Mountain, Jiujiang, Jiangxi	MT700480
Styracaceae	*Styrax ramirezii*	P. W. Fritsch 1472 (CAS)	University of California Botanical Garden, California,U.S.A	NC_041138_1
Styracaceae	*Styrax suberifolius*	Zhao C.X. 201,403 (HIB)	Nanyue Arboretum, Hunan, China	NC_041125_1
Styracaceae	*Styrax zhejiangensis*	NA	NA	NC_038209_1
Styracaceae	*Styrax dasyanthus*	HUTB CH	Lushan Mountain, Jiujiang, Jiangxi	MT700479
Symplocaceae	*Symplocos ovatilobata*	HUTB	Diaoluo Mountain,Hainan, China	NC_036489_1

761

762 **Table 2** GenBank accession numbers, and template plastome for assembly for 12 newly sequenced genomes

Family	**Species name**	**Accession number**	**Locality**	**Template for plastome assembly**
Styracaceae	*Alniphyllum fortunei (Hemsl.) Makino*	MT700470	Lushan Mountain, Jiujiang, Jiangxi	KX765434.1
Styracaceae	*Pterostyrax corymbosus Sieb. et Zucc.*	MT700473	Lushan Mountain, Jiujiang, Jiangxi	KY709672.1
Styracaceae	*Changiostyrax dolichocarpa*	MT700471	Hupingshan,Hunan,China	MF179499.1
Styracaceae	*Changiostyrax dolichocarpa*	MT700472	Hupingshan,Hunan,China	MF179499.1
Styracaceae	*Sinojackia rehderiana Hu*	MT700475	Pengze, Jiangxi,China	MF179499.1
Styracaceae	*Sinojackia xylocarpa Hu*	MT700481	Nanjing Botanical Garden, Jiangsu,China	KY709672.1
Styracaceae	*Sinojackia microcarpa C.T. Chen & G. Y. Li*	MT700474	Jiande,Zhejiang, China	KY626040.1
Styracaceae	*Sinojackia sarcocarpa L. Q. Luo*	MT700476	Sichuan Normal University,China	KY709672.1
Styracaceae	*Sinojackia sarcocarpa L. Q. Luo*	MT700477	Leshan, Sichuan,China	KY709672.1
Styracaceae	*Styrax confusus Hemsl.*	MT700478	Lushan Mountain, Jiujiang, Jiangxi	MF179493.1

Styracaceae	*Styrax dasyanthus Perk*	MT700479	Lushan Mountain, Jiujiang, Jiangxi	MF179493.1
Styracaceae	*Styrax faberi Perkins Wenzhou*	MT700480	Lushan Mountain, Jiujiang, Jiangxi	KX111381.1

763

764

Table 3 Summary of major plastome characteristics in Styracaceae and outgroups.

Latin name	cpDNA size (bp)	LSC size (bp)	SSC size (bp)	IRs size (bp)	Total GC content (%)	LSC (%)	SSC (%)	IR (%)	tRNA	rRNA	Coding gene	Number
Alniphyllum eberhardtii	155384	83710	18153	26761	37.10%	35.20%	30.20%	42.40%	30	4	79	NC_031892_1
Alniphyllum fortunei	155490	83773	18153	26782	37.10%	35.20%	30.20%	42.40%	30	4	79	MT700470
Alniphyllum pterospermum	155185	83200	18583	26701	37.10%	35.20%	30.10%	42.50%	30	4	79	NC_041126_1
Bruinsmia polysperma	157879	86495	18725	26329	36.80%	34.90%	30.30%	42.20%	30	4	79	NC_030180_1
Bruinsmia styracoides	156434	86251	19235	25574	36.70%	34.80%	29.80%	42.60%	30	4	79	NC_041137_1
Changiostyrax dolichocarpa	158881	88086	18609	26091	37.30%	35.30%	30.50%	43.00%	30	4	79	MT700471
Changiostyrax dolichocarpa	158781	88030	18606	26072	37.30%	35.30%	30.50%	43.00%	30	4	79	MT700472
Halesia diptera	158849	88165	18528	26078	37.20%	35.20%	30.50%	43.00%	30	4	79	NC_041128_1
Huodendron biaristatum	158499	87731	18988	25990	36.80%	34.70%	30.30%	42.70%	30	4	79	NC_041132_1
Melliodendron xylocarpum	157131	90159	18486	24243	37.20%	35.30%	30.60%	43.20%	30	4	79	MF179500_1
Perkinsiodendron macgregorii	158602	88189	18293	26060	37.20%	35.20%	30.60%	43.00%	30	4	79	MG719841_1
Pterostyrax corymbosus	158836	88102	18557	26088	37.20%	35.20%	30.50%	43.00%	30	4	79	NC_041134_1
Pterostyrax corymbosus	158890	85662	18561	26106	37.20%	35.30%	30.50%	43.10%	30	4	79	MT700473
Pterostyrax hispidus	158879	88195	18516	26087	37.20%	35.20%	30.50%	43.00%	30	4	79	NC_041135_1
Pterostyrax psilophyllus	158835	88101	17556	26089	37.20%	35.20%	30.50%	43.00%	30	4	79	NC_041133_1
Rehderodendron macrocarpum	157934	87508	18316	25368	37.20%	35.20%	30.60%	43.00%	30	4	79	NC_041139_1
Sinojackia microcarpa	157554	87142	18238	26089	37.30%	35.30%	30.70%	43.00%	30	4	79	MT700474
Sinojackia rehderiana	158872	88077	18516	26091	37.20%	35.20%	30.50%	43.00%	30	4	79	MT700475
Sinojackia sarcocarpa	158901	88168	18556	26090	37.20%	35.20%	30.50%	43.00%	30	4	79	MT700476
Sinojackia sarcocarpa	158834	88092	18881	25931	37.20%	35.20%	30.60%	43.10%	30	4	79	MT700477
Sinojackia xylocarpa	158637	87947	18552	26068	37.20%	35.20%	30.50%	43.00%	30	4	79	MT700481
Stewartia monadelpha	158447	87545	18134	26378	37.30%	35.30%	30.50%	42.80%	30	4	79	NC_041468_1
Stewartia sinii	158478	87531	18962	26363	37.30%	35.30%	30.60%	42.80%	30	4	79	NC_041470_1
Styrax confusus	158261	87837	18299	26064	37.00%	34.80%	30.30%	42.90%	30	4	79	MT700478

Styrax faberi	158160	87785	18225	26073	36.90%	34.80%	30.20%	42.90%	30	4	79	MT700480
Styrax grandiflorus	158052	87648	18310	26047	36.90%	34.80%	30.20%	42.90%	30	4	79	NC_030539_1
Styrax ramirezii	158315	87990	18051	26363	37.00%	34.80%	30.40%	43.00%	30	4	79	NC_041138_1
Styrax suberifolius	158480	87763	18051	26363	37.00%	34.80%	30.30%	42.80%	30	4	79	NC_041125_1
Styrax zhejiangensis	157387	87195	17988	25953	37.00%	34.80%	30.30%	42.80%	30	4	79	NC_038209_1
Styrax dasyanthus	158165	87736	18960	25736	36.90%	34.80%	30.30%	43.00%	30	4	79	MT700479
Symplocos ovatilobata	157417	87447	17792	26089	37.40%	35.40%	30.80%	43.00%	30	4	79	NC_036489_1

Table 4 Data characteristics and models selected in Maximal Likelihood and Bayes Inference analyses for phylogenetic data sets. IR: Inverted repeat; LSC: Large single copy; SSC: Small single copy;

Datasets	**No. of taxa**	**No. of site**	**No. of variable**	**Parsimony informative sites**	**Best Fit Model**	**Model in ML**	**Model in BI**
Whole plastomes	31	180369	31865 (17.66%)	21804 (12.08%)	GTR + I + G	GTR + I + G	TVM + I + G
Coding	31	79755	13242 (16.60%)	9395 (11.78%)	GTR + I + G	GTR + I + G	GTR + I + G
Non-coding	31	131319	21014 (16.00%)	11940 (9.09%)	TVM + I + G	GTR + I + G	TVM + I + G
IRb	31	28419	1900 (6.68%)	938 (3.30%)	TVM + I + G	GTR + I + G	TVM + I + G
LSC	31	104030	23519 (22.60%)	17151 (16.49%)	GTR+I+G	GTR + I + G	GTR + G
SSC	31	22329	5021 (22.49%)	3024 (13.54%)	TVM + I + G	GTR + I + G	GTR+I+G
LSC+SSC	31	126237	28623 (22.67%)	20158 (15.96%)	GTR + I + G	GTR + I + G	GTR + I + G