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.

Plastome Structure and Phylogenetic Relationships of Styracaceae (Ericales)

Xiu-lian Cai
Hainan University
Jacob B. Landis
Cornell University
Hong-Xin Wang
Hainan University
Jian-Hua Wang
Hainan University
Zhi-Xin Zhu
Hainan University
Huafeng Wang (wanghuafeng2012@foxmail.com)
Hainan University https://orcid.org/0000-0002-0218-1728

Review

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

Posted Date: January 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-55283/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

---

Abstract

Background: The Styracaceae are a woody, dicotyledonous family containing 12 genera and an estimated 160 species. Recent studies have shown that *Styrax* and *Sinojackia* are monophyletic, *Alniphyllum* and *Bruinsmia* cluster into a clade with an approximately 20-kb inversion in the Large Single-Copy (LSC) region. *Halesia* and *Pterostyrax* are not supported as monophyletic, while *Melliodendron* and *Changiostyrax* always form sister clades . *Perkinsiodendron* and *Changiostyrax* were newly established genera of Styracaceae. However, the phylogenetic relationship of Styracaceae at the genera level needs further research.

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. In our analyses, all genera were strongly supported as monophyletic. *Styrax* was sister to the remaining genera. *Alniphyllum* and *Bruinsmia* form a clade. *Halesia diptera* does not cluster with *Perkinsiodendron*, while *Perkinsiodendron* and *Rehderodendron* form a clade. *Changiostyrax* is sister to a clade of *Pterostyrax* and *Sinojackia*,

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

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, usually 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, analyzing 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 (one species), *Halesia* J. Ellis ex L (two species) , *Melliodendron* Hand.-Mazz (one species), *Parastyrax* Siebold & Zucc.(two species), *Perkinsiodendron* P. W. Fritsch (one species), *Pterostyrax* W.W. Sm.(four species), *Rehderodendron* Hu (one species), and *Sinojackia* Hu (seven species). On the other hand, the ovary is superior and a persistent hypanthium forms only at the base of the fruit at maturity [*Alniphyllum* Matsum (three species), *Bruinsmia* Boerl. & Koord (two species) *Styrax* L (130 species)]. Moreover, the ovary of *Huodendron* Rehder (four species) 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 and the genera within have been unstable since the establishment of the family by Dumoritor in 1829 [77]. 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 these families have some original characteristics and some new evolutionary characters, which may have arisen via parallel evolution. Based on embryological and anatomical studies, Herbert [10] suggested that the Styracaceae and Theaceae may have originated from a common ancestor, having many common characteristics. 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 placed in the Styracaceae by various authors [12]. *Symplocos*, *Diclidanthera* and *Lissocarpa* were excluded from Styracaceae by Perkins [13]. *Symplocos* was treated as an independent family (Symplocaceae Desf) [14]. *Diclidanthera* was placed in Polygalaceae[7,14], and *Lissocarpa* was placed in 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: (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 and the phylogenetic relationships between genera remain ambiguous. The phylogeny of Ericales based on the chloroplast gene *rbcL* [22] 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 the chloroplast gene *ndhF*, including *Styrax* and *Halesia*, which formed a strongly supported sister-group relationship. Albach et al. [24] came to the same conclusion based on the DNA gene sequences *atpB*, *ndhF*, *rbcL* and 18S [23] 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]. Additionally, this analysis 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 *Alniphyllum* and *Bruinsmia* clustered in a clade with an approximate 20-kb inversion in the Large Single-Copy (LSC) region. The tree species of *Pterostyrax* were not supported as monophyletic, with *Halesia carolina* L and *Pterostyrax hispidus* Siebold & Zucc forming a clade.

The chloroplast genomes of most angiosperms are maternally inherited. The rate of evolution of genes in the chloroplast is relatively slow overall, but differences have been observed across different regions of the plastome,

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 directly or indirectly 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 the phylogeny of Styracaceae 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 some genera, 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 greater resolution and support. The objectives of this study are: 1) infer the plastome structural evolution of Styracaceae, 2) resolve the phylogenetic relationships of Styracaceae, 3) use selective pressure analysis to test for the presence of adaptive evolution in all genes.

Methods

Plant Samples, DNA Extraction, Sequencing and assembly

We collected 28 plastomes of Styracaceae, including 12 newly sequenced 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.Our field collections were permitted by the government following local ethics and laws. Collected plant leaves were put directly into silica gel to dry. 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). Samples yielding at least 0.8 µg DNA were selected for subsequent library construction and *de novo* assembly. Genomic DNA of selected samples were used to build paired-end libraries with insert sizes of 200–400bp. Sequencing of 12 accessions was completed using BGISEQ-500 2x100 at BGI (Shenzhen, China), according to the manufacturer's instructions [78]. This yielded approximately 8 Gb of high-quality data per sample of 100 bp paired-end reads. Raw reads were trimmed using SOAPfilter_v2.2 (BGI-Shenzhen, China) with the following criteria: reads with more than 10 percent base of N, reads with more than 40 percent of low quality (value less than 10), and reads contaminated by adaptors and PCR duplication. Approximately 6Gb of clean data (high-quality reads>Q35) were obtained for each sample.

For all samples, plastomes were assembled using MITObim v1.8[38] with default parameters and using plastomes of related species as templates for assembly (Table 2). The assembly was ordered using BLAST and aligned (> 90% similarity and query coverage) according to the reference chloroplast genome (Table 2).

Genome annotation

Plastomes were annotated using Geneious R11.0.4 [79] using the same reference plastomes used for assembly. Start/stop codons and intron/exon boundaries were further corrected using Dual Organellar GenoMa Annotator (DOGMA) [39]. In addition, tRNAscan-SE1.21 was used to further verify all tRNA genes. We also re-annotated the downloaded assembled plastomes from previous studies 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 Styracaceae plastomes were drawn using OrganellarGenome DRAW (ORDRAW) [40], with subsequent manual editing. Genome comparisons across the 26 Styracaceae species (selecting one sequence per 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 evolutionary histories 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 aligning with 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. 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*(nucleotide diversity) < 0.001] were filtered out from selective pressure analysis.

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 phylogenetic hypotheses, topologies were constructed by both maximum likelihood (ML) and Bayesian inference (BI) methods using not only the complete genome sequences, but by using seven additional data sets (i.e. LSC, SSC, IR, coding, non-coding, combination of LSC+SSC, and concatenation of LSC+SSC+one IR). The best-fitting models of nucleotide substitutions were determined by the Akaike Information Criterion (AIC) in Modeltest 3.7 [45] (Table 4). 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 (79 genes×3 codon positions). Branch lengths for all partitions were used for the ML analyses.

Maximum likelihood analyses were conducted using RAXML-HPC v8.2.8 [47] with 1000 bootstrap replicates on the CIPRES Science Gateway website [49]. Bayesian inference (BI) analyses were performed in MrBayes v3.2[48] on the CIPRES Science Gateway portal[49] with the following conditions used for the protein-coding dataset: starting from random trees, Markov chain Monte Carlo (MCMC) simulations were ran for 900,000,000 generations with four incrementally heated chains, sampling every 1,000 generations. BI analyses were set up identically for the remaining data sets, except that 50,000,000 generations were simulated. Convergence of the MCMC chains was determined by examining the average standard deviation of the split frequencies (< 0.01). The first 25% of the trees were discarded as burn-in. The effective sample size (ESS > 200) was determined by using Tracer v 1.7 [80].

Result

Plastome Structure of Styracaceae

In this study, the plastomes of Styracaceae and outgroups displayed a typical quadripartite structure and similar lengths. 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 plastomes were composed of a large single-copy (LSC) region (ranging from 83,200 bp to 88,258 bp), a small single-copy (SSC) region (ranging from 17,556 bp to 19, 235 bp), and two inverted repeat IR regions (IRa and IRb) (ranging from 24,243 bp to 26,761 bp)(Tab. 4). 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.1).

Comparative genomic analysis and divergence hotspot regions

To investigate the levels of sequence divergence, the 26 Styracaceae plastomes were plotted using mVISTA, with *Pterostyrax hispidus* as the reference (Fig.2). 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 sequence divergence of *Huodendron* in *clpP* intron lower than 50%.

Nucleotide diversity analyses showed that the proportion of variable sites in noncoding region were higher than that in coding region, and the greatest diversity change was in the intergenic spacer region (Fig.3). Among all 209 loci (79 coding genes and 130 non-coding regions), nucleotide diversity (*pi*) values of coding genes ranged from 0.001(*rpl23*) to 0.156 (*atpH*), four loci were greater than 0.1 (*psbK*, *psbI*, *rpoC2*, *atpH*). Nucleotide diversity of non-coding genes ranged from 0 (*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.3).

Selective pressures in plastome evolution of Styracaceae

The results showed that the 79 protein coding genes mainly possessed synonymous substitutions (Fig.4). In addition, *rps12* (0.8874), *rps19* (0.5076) and *rps11* (0.4466) had the highest synonymous substitution rate. The locus with the highest rate of non-synonymous substitution was *ycf1* (1.016). 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 were seven genes with ω value greater than 1: *rps4* (1.087), *rpl23* (1.126), *accD* (1.839), *rpoC1* (1.990), *psaA* (2.175), *rpoA* (1.578) and *ndhH* (3.459). (Fig.5)

Phylogenetic analyses

The optimal partitioning scheme identified under the Akaike information criterion with correction (AICc) using relaxed clustering analysis in PartitionFinder (lnL= -189247.90; AICc=379952.05) contained 64 partitions (Table S1). BI analyses and ML analyses using the unpartitioned and partitioned schemes produced identical topologies (Fig6). The genera within Styracaceae were all recovered as monophyletic with strong support (BS/PP=100/1). All species of *Styrax* form a clade sister to the rest of the family (BS/PP=100/1). The second branch is *Huodendron*, followed by two genera with unique plastome reversals, *Alniphyllum* and *Bruinsmia*. *Halesiadiptera* did not cluster with *Perkinsiodendron* but was sister to the remaining genera (BS/PP=100/1), while *Perkinsiodendron* and *Rehderodendron* form a clade (BS/PP=100/1). *Changiostyrax* is sister to a clade composed of *Pterostyrax* and *Sinojackia* (BS/PP=65/0.67). *Pterostyrax* and *Sinojackia* were the last to diverge from each other and show strong support (BS/PP=85/1). To test for conflicting signals across different data, we use six data sets for analyses (S1-S6). The ML and BI analyses produced similar topologies over all data sets except for the different positions of *Sinojackia sarcocarpa* (L.) Q. Luo, *Changiostyrax dolichocarpus* (C. J. Qi) Tao Chen and *Pterostyrax hispidus* in the IR regions (Fig S1). In tree topology of IR regions, *Sinojackia* and *Pterostyrax* were not monophyletic. Characteristics of all data sets are shown in Table 2.

Discussion

Plastome structure comparisons and sequence divergence hotspots

This study included 31 plastomes, 28 representative taxa from 11 genera of Styracaceae, and three outgroups. Plastomes 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. The plastome 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), with a typical tetragonal structure [50]. The plastome of *Alniphyllum fortunei*, which was first reported in this study, detected a 20-kb inversion which includes 14 coding genes from *trnQ-UUG* to *rpoB*. This inversion was verified by PCR and Sanger sequencing by Yan et al [51]. The 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 has the same gene composition and relative position as the normal plastome structure 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 gene inversion has also found in *Lotus japonicas*, *Arabidopsis thaliana* [53] and members of Oleaceae [54]. Because of their scarcity, plastid inversions are of great value to the study of genome evolution [55, 56]. Previous

studies have suggested that gene inversions are closely related to repetitive sequences, 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* being 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 rate of variation in 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 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 exhibit a weak correlation with the degree of variation in the genome [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*). The chloroplast NADH dehydrogenase (NDH) complex participates in the circular electron transport and chlorine respiration around the light system [69]. However, due to NDH complex existing in low abundance and being of a fragile nature, it is difficult to analyze its function. The plants of Styracaceae are mainly distributed in the tropics and subtropics, which are subjected to growing conditions of high light and high temperature. 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, as well 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 over 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].

Phylogeneticanalyses

We constructed seven data matrices by different sequence segmentation, and analyzed the phylogeny of the different matrices to maximize the resolution phylogenetic relationships and to test if there are conflicting signals. Overall, the phylogenetic relationships constructed by the different data matrices do consistent topologies with moderate support. The phylogenetic based on the complete plastome is consistent with the results of six of the data sets except the IR region. According to Fritsch et al.'s [1] analysis of morphology and three DNA sequence data sets, *Styrax* is monophyletic, forming a clade with *Huodendron*. However, our analyses show that *Styrax* remained monophyletic with high support (BS/PP=100/1) and is 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 rates of substitution in the two genera..

Fritsch et al. [1] and Yao et al. [25], consistently showed that *Melliodendron* formed a clade with *Changiostyrax*, whereas in all our data sets, except in the LSC data set, *Melliodendron* and *Changiostyrax* do not form a clade. *Changiostyrax* is strongly supported as sister to a clade composed of *Pterostyrax* and *Sinojackia*. *Halesia* and *Pterostyrax* have not been previously fully resolved [1, 25, 26]. Here, we collected four accessions of *Pterostyrax* to analyze and *Pterostyrax* was recovered as monophyletic in all analyses except for *P. hispidus* was observed as being excluded from the other two species with a relatively low support value (BS/PP=56/1) in the IR data set. The conflicting signal from different partitions of the chloroplast are more likely to be 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 sets, with *Perkinsiodendron* being a new genus established from *Halesia macgregorii* Chun based on molecular data and morphological characters [21]. Furthermore, our study strongly supports the monophyly of *Sinojackia* based on plastid data, as has been detected in previous studies [25], except in IR data set where *Sinojackia sarcocarpa* is separated from the other species. 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 taxonomic sampling and biological factors such as hybridization/introgression, incomplete lineage sorting, gene duplication and/or loss, and horizontal gene transfer [74, 75, 76]. However, most of these reasons do not explain differences observed between different partitions of complete plastome sequences.

Conclusions

Our results presented here utilize a phylogenomic data set to investigate phylogenetic relationships among the genera of Styracaceae. Based on 28 complete plastomes, our results show that the plastome structure of Styracaceae have small differences except for *Alniphyllum* and *Bruinsmia*, which have an approximately 20-kb inversion. Our results clearly indicate that all genera of Styracaceae are monophyletic, and the establishment of *Perkinsiodendron* and *Changiostyrax* are supported. Nevertheless, the lack of *Parastyrax* species in the sequence data, necessitates that our results may need to be further verified by increasing taxon sampling and using nuclear genes. 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

Declarations

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).

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, College of Tropical Crops, 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. 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. The Genera of Flowering Plant.Oxford, 1967:2:34–39
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. New Phytol. 2017:215, 1235-1248.
33. Allen, J.F., de Paula, W. B., Puthiyaveetil, S., Nield, J. A structural phylogenetic map for chloroplast photosynthesis. Trends Plant Sci. 2011:16, 645-655.
34. Carbonell-Caballero, J., Alonso, R., Iba, V.E., Terol, J., Talon, M., Dopazo, J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. Mol. Biol. Evol.2015: 32, 2015-2035.
35. Hu, S.L., Sablok, G., Wang, B., Qu, D., Barbaro, E., Viola, R., Li, M.A., Varotto, C.Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. BMC Genom. 2015:16, 306.
36. APG IV. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. 2016:181, 1–20. http://dx.doi. org/10.1111/boj.12385.
37. Doyle J.J., Doyle J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 1987:19: 11-15.
38. Hahn, C., Bachmann, L., Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 2013:41, e129.
39. Wyman S.K., Jansen R.K., Boore J.L. Automatic annotation of organellar genomes with DOGMA. Bioinformatics, 2004:20: 3252-3255.
40. Lohse, M., Drechsel, O., Kahlau, S., Bock, R. Organellar GenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research 2013:41, W575.
41. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I.Vista: computational tools for comparative genomics. Nucleic Acids Research 2004:32, 273–279.
42. Katoh K., Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 2013:30: 772-780.

43. Librado P., Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009:25: 1451-1452.
44. Yang Z.H. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution, 2007:24: 1586-1591.
45. Posada, D., Crandall,A. Modeltest: testing the model of DNA substitution. Bioinformatics 1998:14 (9), 817-818.
46. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 2016:34, 772–773.
47. Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
48. Ronquist, F., Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003:19, 1572–1574.
49. Miller, M.A., Pfeiffer, W., Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE). 2010:14, 1–8.
50. Jansen R.K., Ruhlman T.A. Plastomes of seed plants. Sharkey T.D. (Eds.). Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration, Vol. 35. 2012:103-126.
51. Yan, M.H., Moore, M.J., Meng, A.P., Yao, X.H., Wang, H.C. The first complete plastome sequence of the basal asterid family Styracaceae (Ericales) reveals a large inversion. Plant Syst. Evol. 2017:303, 61–70. http://dx.doi.org/10.1007/s00606-016-1352-0.
52. Yi‐ing Liao Liao, Liu Y , Liu X , et al. The complete chloroplast genome of Myriophyllum spicatum reveals a 4-kb inversion and new insights regarding plastome evolution in Haloragaceae. Ecology and Evolution, 2020:10(6).
53. Tomohiko K , Takakazu K , Shusei S , et al. Complete Structure of the Chloroplast Genome of a Legume, Lotus japonicus. Dna Research. 2000:(6):6.
54. Lee HL, Jansen RK, Chumley TW, Kim KJ. Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. Molec Biol Evol. 2007:24:1161–1180.
55. Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. Complete plastid genome sequence of the chickpea (Cicer arietinum) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). Molec Phylogen Evol 2008:48:1204–1217. doi:10.1016/j.ympev.
56. Johansson JT. There large inversions in the chloroplast genomes and one loss of the chloroplast gene rps16 suggest an early evolutionary split in the genus Adonis (Ranunculaceae). Pl Syst Evol. 1999:218:133–143.
57. Palmer JD. Plastid chromosomes: structure and evolution. In: Vasil IK, Bogorad L, editors. Cell culture and somatic cell genetics in plants. Vol. 7A. The molecular biology of plastids. San Diego: Academic Press. p. 1991:5–53.
58. Pombert JF, Lemieux C, Turmel M. The complete chloroplast DNA sequence of the green alga Oltmannsiellopsis viridis reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol. 2006:4:3.
59. Raubeson LA, Jansen RK. Chloroplast genomes of plants. In: Henry R, editor. Diversity and evolution of plants—genotypic and phenotypic variation in higher plants. Oxfordshire: CABI Publishing. p.2005:45–68.

60. Huang H., Shi C., Liu Y., et al. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology. 2014:14: 151.
61. Jansen R.K., Ruhlman T.A. Plastomes of seed plants. Sharkey T.D. (Eds.). Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration, Vol. 35. Dordrecht, Netherlands: Springer: 2012:103-126.
62. Perry A.S., Wolfe K.H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. Journal of Molecular Evolution, 2002:55: 501-508.
63. Joey Shaw, Edgar B. Lickey, Edward E. Schilling, and Randall L. Small. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany. 2007:94(3):p.275-288.
64. , Shaw, R., L., & Small. Chloroplast dna phylogeny and phylogeography of the north american plums (prunus subgenus prunus section prunocerasus, rosaceae). American Journal of Botany. 2005:92: 2011–2030.
65. Curtis, SE, Clegg MT. Molecular evolution of chloroplastDNA sequences. Molecular Biology and Evolution 1984:1: 291–301.
66. G AUT , B. S. Molecular clocks and nucleotide substitution rates in higher plants. Evolutionary biology, vol.1998: 30, 93–120.
67. Wolfe, K. H.Li, W. H.Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences, USA 1987:84: 9054–9058.
68. Weng ML, Blazier JC, Govindu M, et al. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Molecular Biology and Evolution, 2014:31(3): 645-659.
69. Sazanov L A, Burrows P A , Nixon P J. The plastid ndh genes code for an NADH-specific dehydrogenase: Isolation of a complex I analogue from pea thylakoid membranes. Proceedings of the National Academy of Sciences, 1998:95(3):1319-1324.
70. Fan X, Zhang J, Li W, et al. The NdhV subunit is required to stabilize the chloroplast NADH dehydrogenase-like complex in Arabidopsis. The Plant Journal, 2015:82(2):221-231.
71. Liqiang, Wang, Hui. Complete plastome sequence of Iodes cirrhosa Turcz. the first in the Icacinaceae, comparative genomic analyses and possible split of Idoes species in response to climate changes. Peerj.2019
72. Bisson, G.P. et al. Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis. J. Bacteriol. 2012:194, 6441–6452.
73. Wolfe, K. H. Protein-coding genes in chloroplast DNA:compilation of nucleotide sequences, data base entries, and rates of molecular evolution. In L. Bogorad and I. K. Vasil [eds.], Cell culture and somatic cell genetics of plants, vol. 1991:7B, 467–482.
74. Naciri, Y., Linder, P. Species delimitation and relationships: The dance of the seven veils. Taxon 2015:64, 3-16.
75. Nicola, M.V., Johnson, L.A., Pozner, R. Unraveling patterns and processes of diversification in the South Andean-Patagonian Nassauvia subgenus Strongyloma (Asteraceae, Nassauvieae). Mol. Phylogenet. Evol.

2019:136, 164-182.

76. Lin, H.Y., Hao, Y.J., Li, J.H., Fu, C.X., Pamela, S.S., Douglas, E.S., Zhao, Y.P. Phylogenomic conflict resulting from ancient introgression following 24 species diversification in Stewartia s.l. (Theaceae). Mol. Phylogenet. Evol. 2019:135, 1-11.
77. Dumortier B-C. Analyse des familles des plantes :avec l'indication des principaux genres quis'y rattachent (French Edition) [M]. Tournay Impr. de J. Casterman, Ainé, 1829.
78. Gao S, Mu F, Yang Z, Liu X, Jiang H, Liao S, et al. BGISEQ-500 WGS library Protocols.io 2018; http://dx.doi.org/10.17504/protocols.io.ps5dng6
79. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
80. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67, 901–904.

Tables

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

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

Table 3 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

Table 4 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

Figures

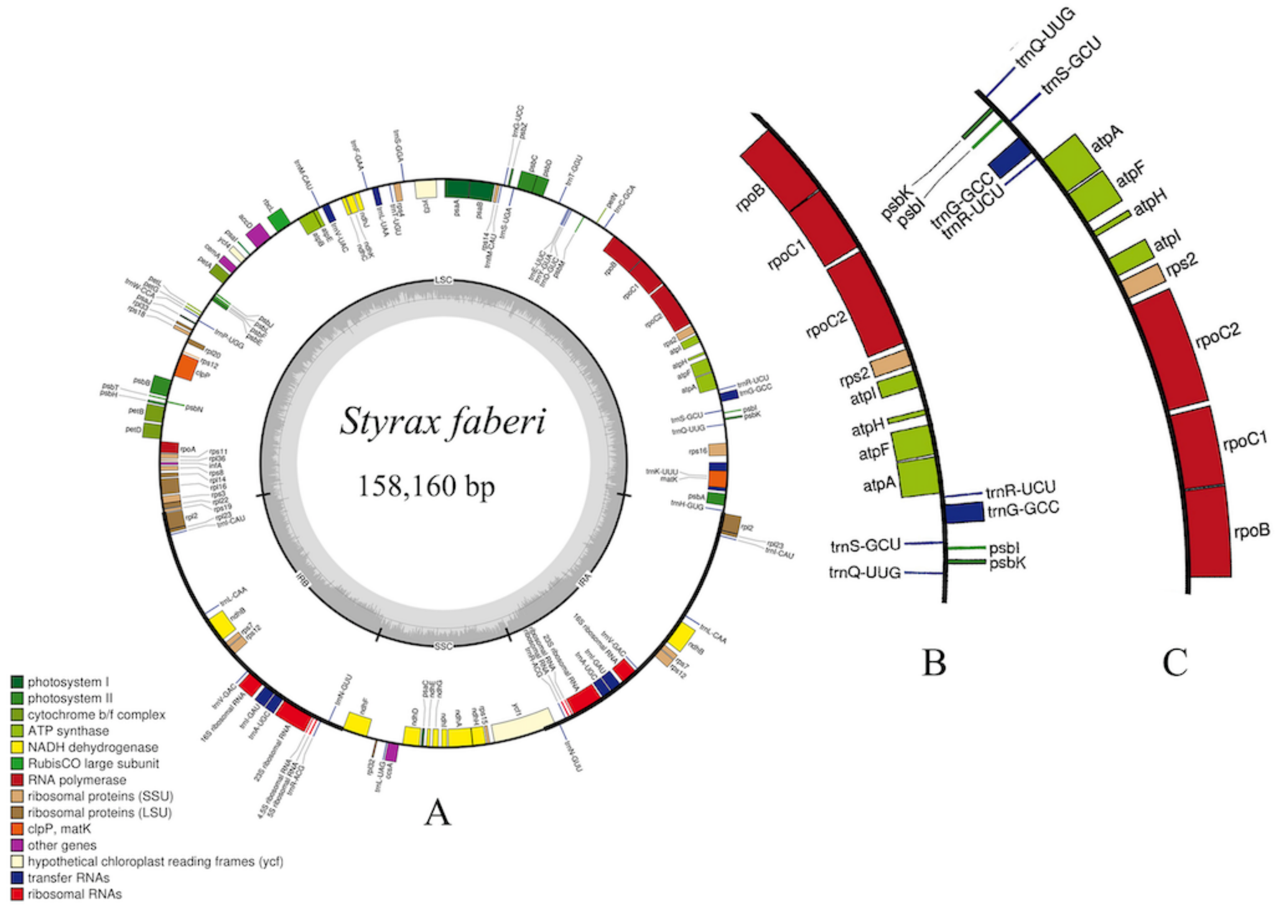
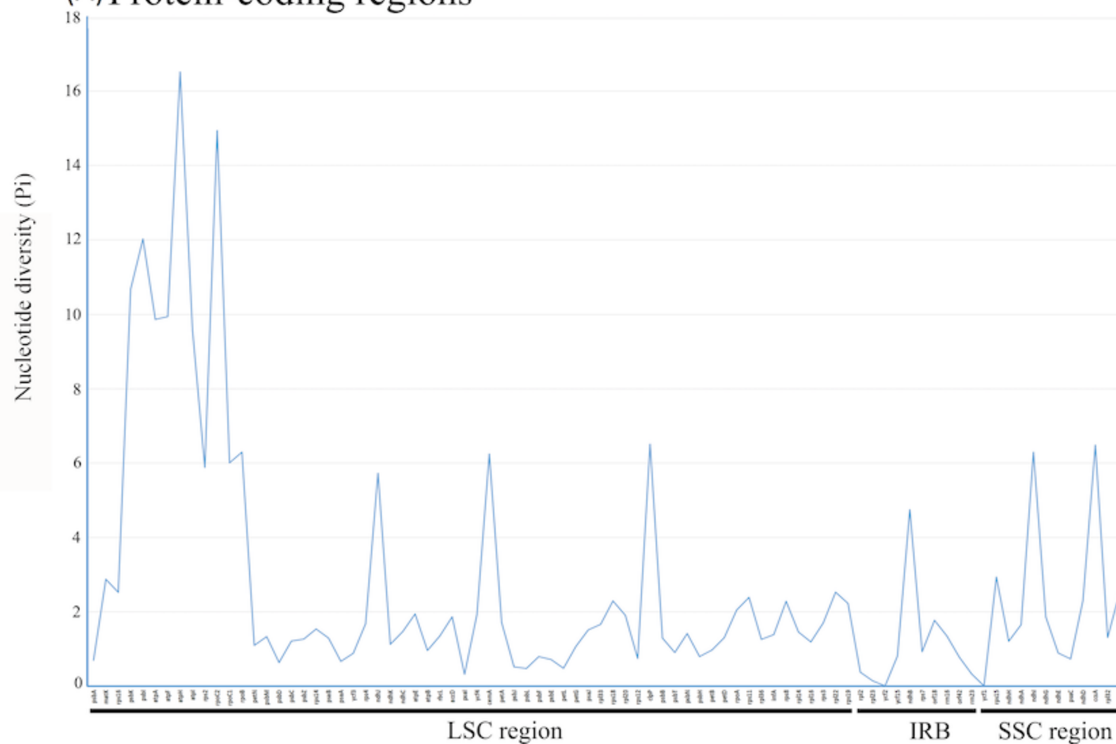


Figure 1

Gene map of the Styrax faberi. (A) The inverted order of genes in Alniphyllum fortunei; (B) The corresponding region of Styrax faberi.

Figure 2

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.



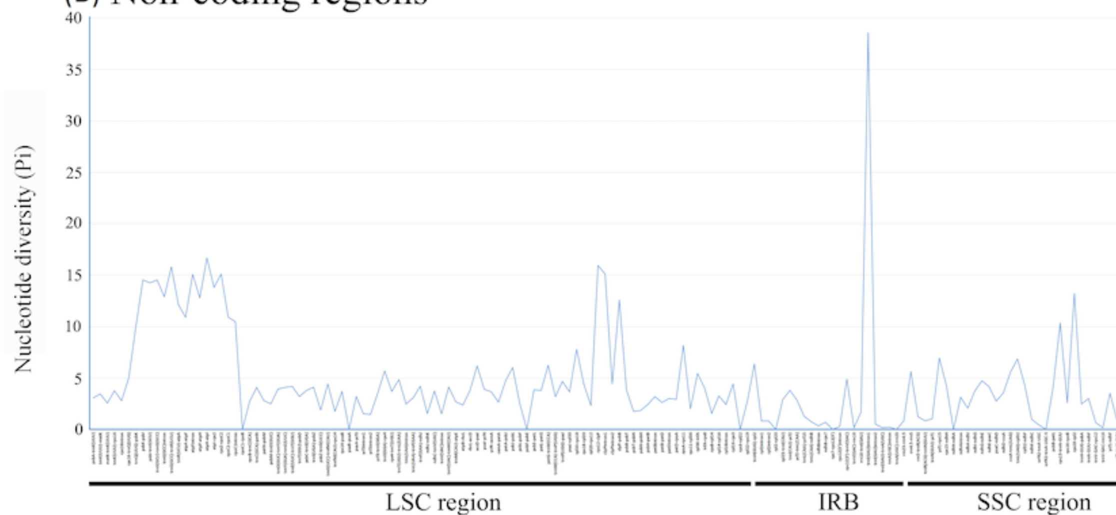


Figure 3

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

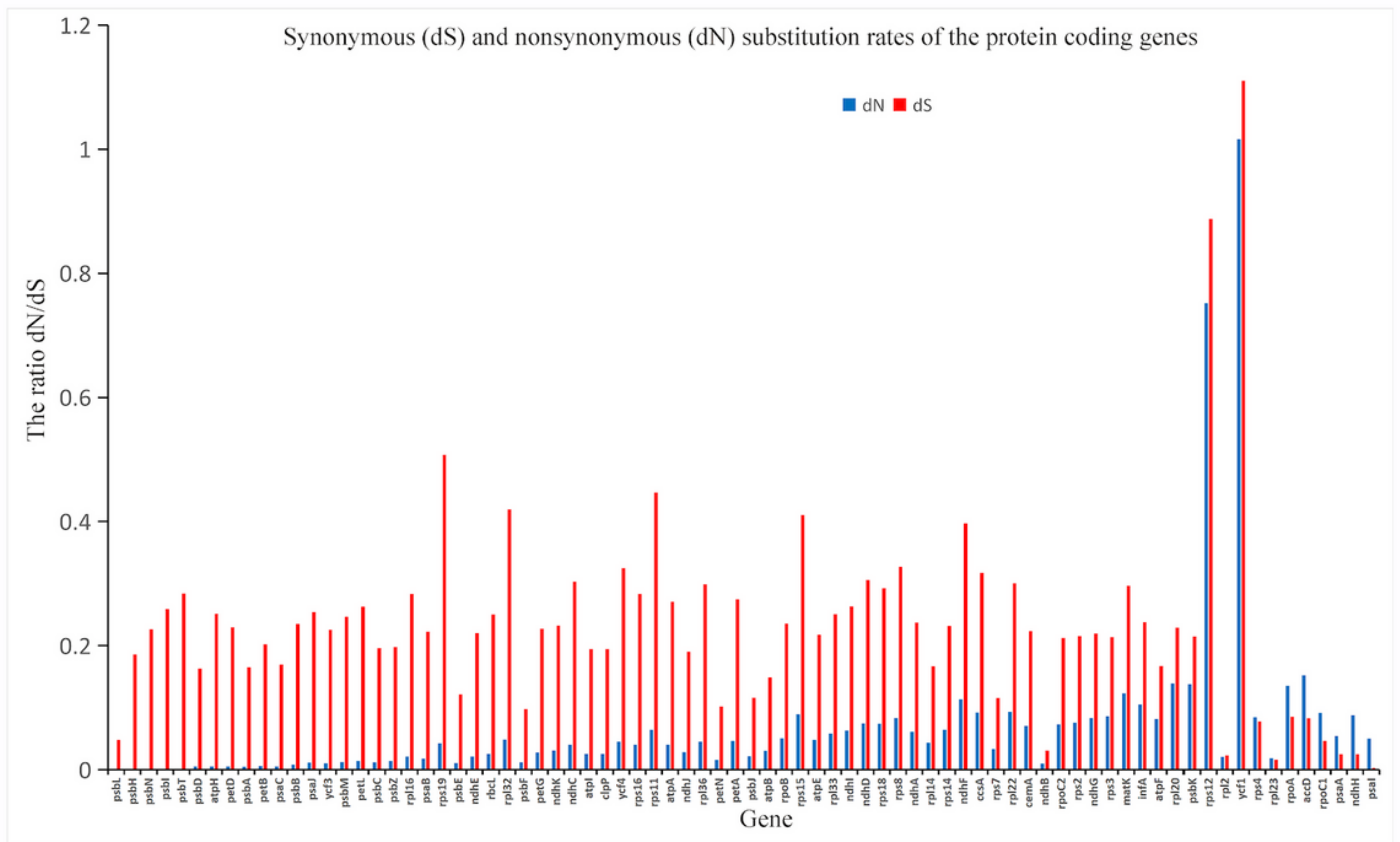


Figure 4

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

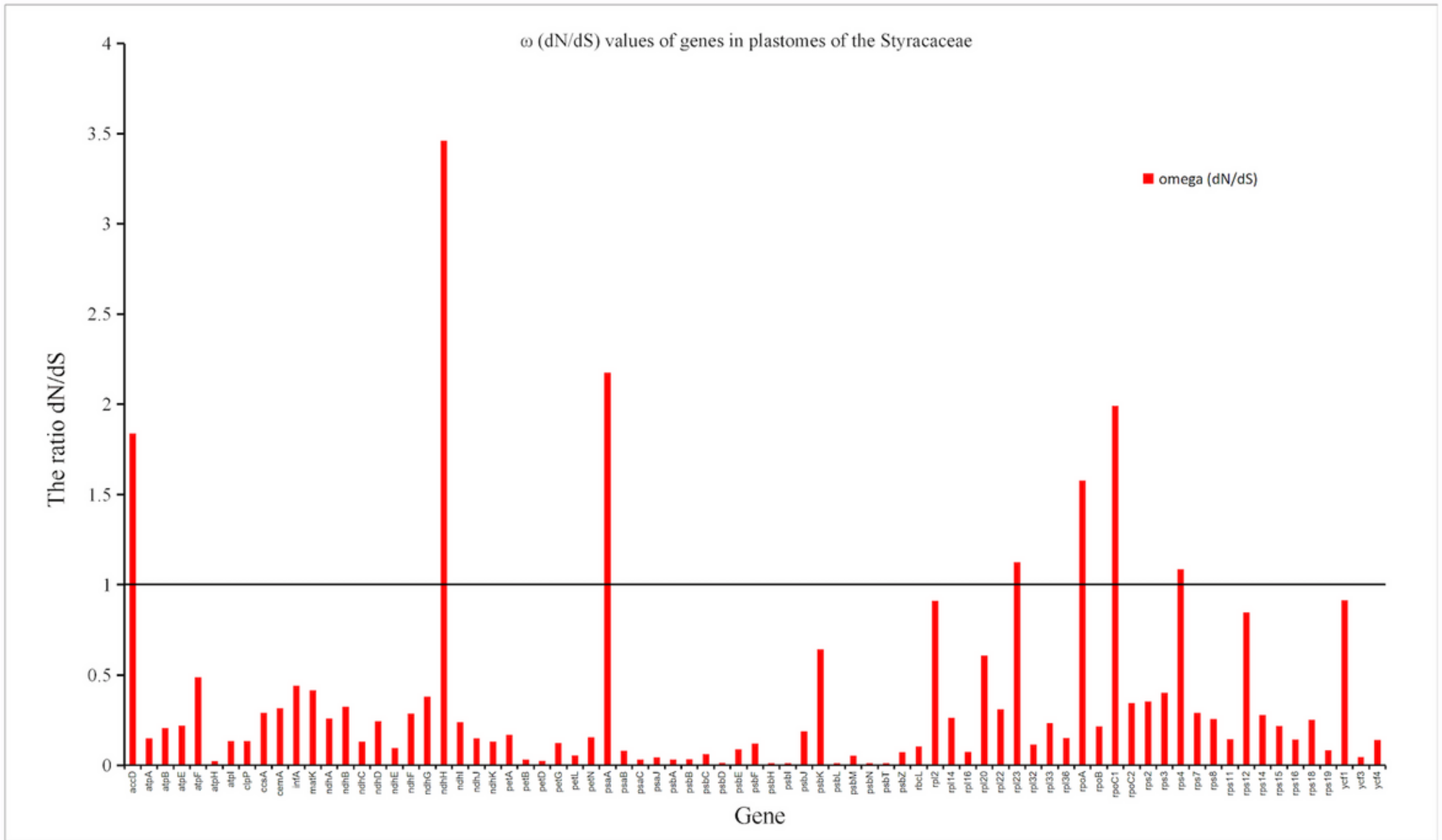


Figure 5

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.

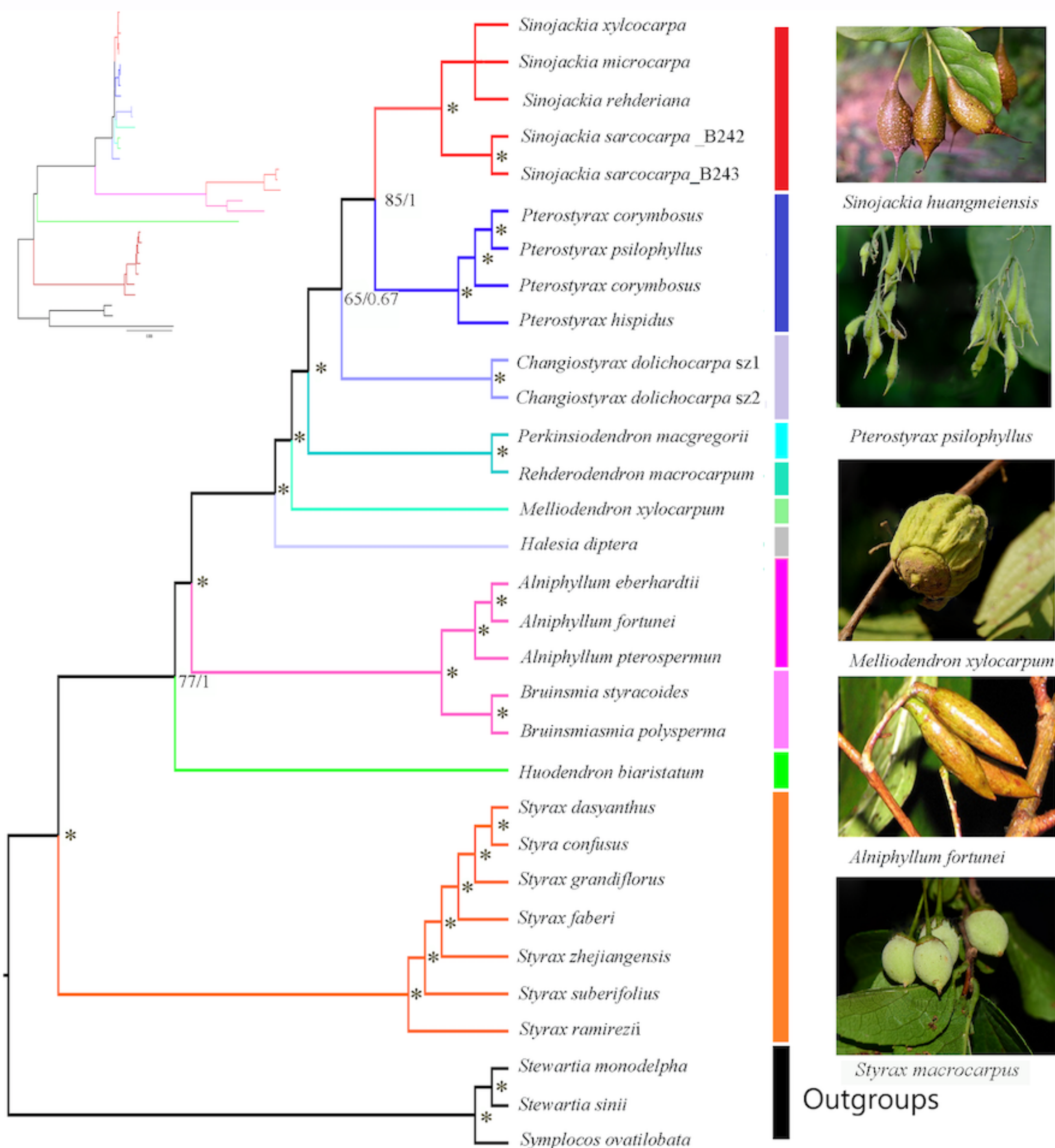


Figure 6

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 next to the nodes are maximum likelihood bootstrap support/Bayesian posterior probability; asterisks indicate 100%/1.0 support values. The genera of Styracaceae are indicated by different branch colors. The inset shows the same tree as a phylogram.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigS1.jpg
- FigS2.jpg
- FigS3.jpg
- FigS4.jpg
- FigS5.jpg
- FigS6.jpg
- TableS1.pdf