BSJ

Plastome structure, phylogeny and evolution of plastid genes in *Reevesia* (Helicteroideae, Malvaceae)

Li-Yang Geng[1] · Tian-Yi Jiang[1] · Xin Chen[1] · Qiang Li[1] · Jian-Hui Ma[1] · Wen-Xiang Hou[1] · Chen-Qian Tang[1] ·
Qin Wang[1,2] · Yun-Fei Deng[2,3]

Abstract

Reevesia is an eastern Asian-eastern North American disjunction genus in the family Malvaceae s.l. and comprises approximately 25 species. The relationships within the genus are not well understood. Here, 15 plastomes representing 12 *Reevesia* species were compared, with the aim of better understanding the species circumscription and phylogenetic relationships within the genus and among genera in the family Malvaceae s.l. The 11 newly sequenced plastomes range between 161,532 and 161, 945 bp in length. The genomes contain 114 unique genes, 18 of which are duplicated in the inverted repeats (IRs). Gene content of these plastomes is nearly identical. All the protein-coding genes are under purifying selection in the *Reevesia* plastomes compared. The top ten hypervariable regions, SSRs, and the long repeats identified are potential molecular markers for future population genetic and phylogenetic studies. Phylogenetic analysis based on the whole plastomes confirmed the monophyly of *Reevesia* and a close relationship with *Durio* (traditional Bombacaceae) in subfamily Helicteroideae, but not with the morphologically similar genera *Pterospermum* and *Sterculia* (both of traditional Sterculiaceae). Phylogenetic relationships within *Reevesia* suggested that two species, *R. pubescens* and *R. thyrsoidea*, as newly defined, are not monophyletic. Six taxa, *R. membranacea*, *R. xuefengensis*, *R. botingensis*, *R. lofouensis*, *R. longipetiolata* and *R. pycnantha*, are suggested to be recognized.

Keywords Adaptive evolution · Genome structure · Phylogeny · Plastome · *Reevesia*

Introduction

Malvaceae Juss. is a family constituting the monophyletic core Malvales with Bombacaceae, Sterculiaceae and Tiliaceae (Alverson et al. 1998, 1999; Judd and Manchester 1997). However, recent cladistic analyses of the morphological, anatomical, palynological and chemical

Li-Yang Geng and Tian-Yi Jiang are contributed equally to this work.

✉ Xin Chen
chenxinzhou@njfu.edu.cn

1   Co-Innovation Center for Sustainable Forestry in Southern China, College of Life Sciences, Nanjing Forestry University, Nanjing, Jiangsu 210037, China

2   Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

3   Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou 510650, China

characteristics (Judd and Manchester 1997) and molecular evidences (Alverson et al. 1998, 1999; Bayer et al. 1999a, b; Hernández-Gutiérrez and Magallón 2019; Richardson et al. 2015; Wang et al. 2021) have suggested that Tiliaceae, Sterculiaceae, and Bombacaceae are non-monophyletic and that only the Malvaceae is likely monophyletic. In addition, Tiliaceae, Sterculiaceae, and Bombacaceae were merged into the expanded Malvaceae (Judd and Manchester 1997). With sequence analyses of the plastid genes *atpB* and *rbcL*, the expanded Malvaceae was divided into nine subfamilies, i.e., Grewioideae, Byttnerioideae, Tilioideae, Helicteroideae, Dombeyoideae, Brownlowioideae, Sterculioideae, Malvoideae and Bombacoideae (Bayer et al. 1999a; Byng et al. 2016; Stevens 2001).

Traditionally, the genus *Reevesia* was placed into tribe Helictereae under family Sterculiaceae (Bentham and Hooker 1862; Cronquist 1988; Hsue 1984a; Hutchinson 1967; Takhtajan 1997; Thorne 1992; Wu et al. 2003). When it was established by Lindley (1827), it was assigned to Byttneriaceae (including Sterculiaceae) and presumed that

it is similar to *Pterospermum* Schreb. in having petals and capsules and *Sterculia* L. in having calyx and stamens. However, due to recent research, *Reevesia* was merged into the expanded Malvaceae with Sterculiaceae (Byng et al. 2016; Stevens 2001). Accordingly, the genus *Reevesia* was assigned to tribe Helictereae under subfamily Helicteroideae. *Reevesia* is easily distinguished from other genera by having the large terminal inflorescences, the bisexual flowers with the presence of white corolla and conspicuous long androgynophore, woody capsules and membranous-winged seeds (Bayer and Kubitzki 2003; Tang et al. 2007).

Reevesia Lindl. is a genus representing an example of well-known eastern Asian-eastern North American floristic disjunction (Tiffney 1985; Wen 1999; Wen et al. 2010; Wolf 1980). The genus contains 25 species, of which 23 species occur in East and Southeast Asia and two species in Mexico and Nicaragua (Bayer and Kubitzki 2003; Dorr 2017; Mabberley 2017; Solheim 1991; Xu and Hsue 2001). The genus was considered to be restricted to East Asia from Northeast India to Southeast China and Indochina for a long time (Hsue 1984a; Long et al. 1985; Mabberley 1987; Phengklai et al. 2001; Wu 1991; Wu et al. 2003). The American genus *Veeresia* Monach. & Moldenke was considered to resemble Asia *Reevesia* and was named so by an anagram of the latter's name, but differs by the presence of the staminodes (Monachino 1940). However, further studies revealed that both genera have minute tooth-like staminodes at the summit of the anther column, and therefore, *Veeresia* was treated as the synonym of *Reevesia* (Bayer and Kubitzki 2003; Dorr 2017; Machuca 2017; Solheim 1991; Tang et al. 2007; Terada and Suzuki 1998).

Some species of *Reevesia* are of significant economic value as ornamental, timber, fiber and medicine (Bayer and Kubitzki 2003; Chang et al. 2013; Hsue 1984b). The species of *Reevesia* are well-known as "nine-layer barks", "oil hemp tree", "hard shell fruit tree", etc. (Hsue 1984b). Some species have been cultivated as ornamental garden plants for their large inflorescence and white fragrant flowers. The plants of *Reevesia* are rich in fiber, which can be used as raw material for weaving sacks, rope and paper making (Bayer and Kubitzki 2003; Hsue 1984b).

The species delimitation of *Reevesia* depends on vegetative features including indumentum, petiole length, shape and size of leaf blades (Anthony 1926; Feng et al. 2022; Hsue 1984b; Solheim 1991; Tang et al. 2007). However, this is not accurate. Long (1985) thought that using leaf blade shapes to distinguish species in *Reevesia* with the same flower morphology characteristics is unreasonable. Due to the lack of significant differences in reproductive characteristics, the challenge of species delimitation is often enhanced. For example, *R. siamensis* Craib (1924) was treated as a variety of *R. pubescens* Mast. by Anthony

(1926). However, this treatment was not supported by studies of pollen morphology. *Reevsia siamensis* has pentacolporate pollen grains, while tetracolporate pollen grains are found in *R. pubescens* (Long et al. 1985). However, the phylogenetic relationships within *Reevesia* have not been determined due to the lack of plastome.

The plastid organelle is a remarkable feature in plants, and is considered to have evolved from cynobacterium since tens of millions of years ago (Timmis et al. 2004). Compared with the nuclear genome, the plastid genome (plastome) has a simple structure with a low molecular weight and multiple copies. Previous studies have revealed that the plastome, whose size ranges from 120 to 170 kb and includes 120–130 genes under normal conditions, consists of a characteristic circular quadripartite structure that includes a large single copy (LSC) and small single copy (SSC) separated by a pair of inverted repeats (Aldrich et al. 1985, 1988; Bendich 2004). However, there are some exceptions to the normal structure, such as IRLC (IR-lacking clade) in Fabaceae (Moghaddam et al. 2022), three tandem repeats with the same direction in *Euglena gracilis* and IR regions with direct duplication in *Porphyra purpurea* (Xing and Liu 2008). Due to its low substitution rates of nucleotides and relatively conserved variation in genomic structures, plastomes are extensively utilized for phylogenetic relationship inference, DNA barcoding and molecular markers development (Du et al. 2021; Luo et al. 2021; Ren et al. 2020; Tian et al. 2021). With the development of high-throughput sequencing technologies, it has been demonstrated that the plastomes are remarkable at resolving phylogenetic relationships at different taxonomic levels (Barrett et al. 2016; Ruhsam et al. 2015; Wang et al. 2016; Wu and Chaw 2016; Xie et al. 2019; Zhang et al. 2022b). However, up to date, the plastomes in *Reevesia* have been sequenced for four species, *R. botingensis* Hsue, *R. pycnantha* Y. Ling, *R. rotundifolia* Chun, and *R. thyrsoidea* Lindl., and are available for public usage (Quan et al. 2020; Wang et al. 2021; Zhang et al. 2022b).

In the present study, eleven plastomes of *Reevesia* species were newly sequenced. The aims are to: (1) compare the structural variation, examine variations of simple sequence repeats (SSRs), and calculate the nucleotide diversity for future population genetic and phylogenetic studies; (2) identify the hypervariable regions of these plastomes as DNA barcodes for species delimitation and identification; (3) test for the presence of adaptive evolution in all genes located in the two single copy regions and one of the two IR regions, using selective pressure analysis; and (4) resolve phylogenetic relationships within the genus and among genera within the family Malvaceae s.l.

Materials and methods

Taxon sampling and DNA extraction

Eleven samples representing ten *Reevesia* species were newly sequenced in the present study. Among them, eight species were sequenced for the first time. Samples were collected in their native habitats during several botanical surveys, or obtained from Nanjing Botanical Garden MEM, Sun Yat-sen or Shanghai Jiao Tong University. Leaf samples were preserved with silica gel and the voucher specimens were deposited at the Herbarium of Nanjing Forestry University (NF). A list of all the newly sequenced samples with geographical origin, collection history, herbarium voucher information and GenBank accession numbers were included in Table 1. Total genomic DNA was extracted from silica gel-dried leaf material. DNA was extracted using the CTAB protocol modified from Doyle and Doyle (1987) and Super

Plant Genomic DNA Kit (Polysacchardes & Polyphenoilics-rich; TIANGEN, Beijing).

Plastome sequencing, assembly, and annotation

The DNA concentration was quantified using a Qubit Fluorometer or microplate reader and visualized in a 1% agarose-gel electrophoresis. The genomic DNA was sheared to produce 270 bp fragments with Covaris, and was purified using AxyPrep Mag PCR clean up Kit. The selected fragments were amplified after suffering from end repair and adding polyA tail and adaptor ligation. The processed fragments were heat denatured to single strand after purification. The single strands were circularized, and single strand circular DNA was obtained as the final library. Then, the DNA extractions were sent to BGI (The Beijing Genomics Institute, China) for library preparation and sequencing.

DNA was sequenced with 101 bp paired-end reads using the Illumina HiSeq 4000 platform at BGI. Raw reads

Table 1 Basic information and NCBI GenBank accession numbers of 11 *Reevesia* samples used in this study

Taxa	Voucher	Clean bases (G)	Clean reads	Average coverage(×)	Genbank accession number
R. botingensis (1)	CHINA. Hainan, Xin Chen, Jiabao Li and Yinquan Huo 1708 (NF)	2,437,589,400	8,125,298	229	OP856555
R. glaucophylla	CHINA. Hunan, Xin Chen, Xiangxiang Fu and Chunhui Wang 1789 (NF)	1,984,305,900	6,614,353	248	OP868812
R. lofouensis	CHINA. Guangdong, Xin Chen, Kailing Zhu and Jiabao Li 1521 (NF)	2,081,719,800	13,878,132	187	OL701984.1
R. longipetiolata	CHINA. Hainan, Xin Chen, Jiabao Li 1722 (NF)	2,410,089,600	8,033,632	232	OP868809
R. membranacea (1)	CHINA. Guangxi, Haiying Nong GX-001-1 (NF)	2,315,680,800	15,437,872	101	OP868813
R. membranacea (2)	CHINA. Guangxi, Haiying Nong GX-001-3 (NF)	2,292,920,700	15,286,138	155	OP868814
R. orbicularifolia	CHINA. Shanghai, Dongqin Tang 1661 (NF)	2,166,109,800	14,440,732	104	OL701983.1
R. pubescens	CHINA. Zhejiang, Ting Wang 1663 (NF)	1,921,122,600	12,807,484	194	OL701985.1
R. thyrsoidea (1)	CHINA. Guangdong, Chengxiang Xiu 1729 (NF)	2,609,793,000	17,398,620	104	OP868810
R. tomentosa	CHINA. Guangdong, Zhengyang Niu and Qin Wang 1653 (NF)	2,167,034,400	14,446,896	103	OP868811
R. xuefengensis	CHINA. Hunan, Xin Chen, Kailing Zhu and Jiabao Li 1507 (NF)	2,104,865,400	14,032,436	187	OL701987.1

were first filtered to obtain high-quality data by removing adapter sequences and low-quality reads using the program SOAPnuke (Chen et al. 2018). Bases were discarded in case of quality < Q20 of read ends. The remaining reads were trimmed if average quality in a 5-bp window was < Q20. After trimming, reads were removed if their length fell below 36 bp. The number and quality of clean reads obtained from each *Reevesia* sample were evaluated with FastQC v.0.11.9 (Andrews 2010), and the details were provided in Table 1.

The mapped plastid reads were assembled into contigs using the GetOrganelle v.1.7.5.3 with the suitable parameter setting for *Reevesia* (wordsize = 102, k-mer = 75,85,105,115,127) by mapping the reads to reference genomes obtained from GenBank (*Durio zibethinus* Moon, NC_036829.1, *Gossypium tomentosum* Nutt. ex Seem., HQ325745.1 and *R. thyrsoidea*, NC_041441.1). Bandage v0.8.1 software was used to map all reads to the assembled plastid sequence for visualization processing and obtaining accurate plastomes (Wick et al. 2015). Relative position and direction of contigs were manually adjusted according to the reference genome using Geneious v.9.0.2 (Kearse et al. 2012).

Successfully assembled plastomes were annotated using the PGA program (Qu et al. 2019) with *Durio zibethinus*, *R. rotundifolia* as references. The start/stop codons and the exon/intron boundaries of genes were manually adjusted in Geneious v.9.0.2 (Kearse et al. 2012). The plastomes were visualized using OGDRAW v.1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Greiner et al. 2019). The complete plastid sequences and gene annotation of the 11 newly assembled *Reevesia* Samples were submitted to NCBI database (https://www.ncbi.nlm.nih.gov) under the accession numbers listed in Table 1.

Genome comparison and structural analysis

For each plastome, the LSC, SSC, IRa and IRb regions were plotted with boundary positions being compared using IRscope online software (https://irscope.shinyapps.io/irapp/; Amiryousefi et al. 2018).

All protein-coding genes (CDSs) for each *Reevesia* plastome were extracted using Geneious v.9.0.2. The amount of codon and relative synonymous codon usage (RSCU) ratio was calculated using CodonW v.1.4.2 software (http://codonw.sourceforge.net/) with default parameters.

Repeated sequences analysis

Repeat structure in 15 *Reevesia* plastomes was examined and compared. SSRs were identified using the MISA online software (https://webblast.ipk-gatersleben.de/misa/). The

thresholds of repeats were 12 for mono-, 6 for di-, 4 for tri-, and 3 for tetra-, penta- and hexa-nucleotides.

Sequence divergence analysis

Genomic similarity analysis was performed using the Shuffle-LAGAN mode in mVISTA (Brudno et al. 2003; Frazer et al. 2004), with *R. botingensis* (OP856555) as the reference. The nucleotide diversity (Pi) for all protein-coding and noncoding (intergenic spacer and intron) regions was calculated using sliding window analysis to detect the most divergent regions (i.e., mutation hotspots) in DnaSP v.5.10.01 (Librado and Rozas 2009; Rozasas and Sánchez 2003).

Selective pressure analysis

Selective pressures were analyzed for 80 common CDSs of 15 *Reevesia* plastomes included in the phylogenetic analysis. The values of nonsynonymous (dN) and synonymous nucleotide (dS) substitution rates and the ratio ($\omega = dN/dS$) for each protein-coding gene were calculated using the Codeml program in PAML4.9 with the site-specific model (seqtype = 1, model = 0, NSsites = 0, 1, 2, 3, 7, 8; Yang 2007; Yang and Nielsen 2002). The likelihood ratio tests (LRTs) were used to identify positively selected sites in comparisons of M0 vs. M3, M1 vs. M2, M7 vs. M8 (Yang and Nielsen 2002). The codon frequencies were set by the F3×4 model. The Bayes Empirical Bayes (BEB) method was used to identify codons under positive selection (Yang et al. 2005). BEB values higher than 0.95 indicate sites that are potentially under positive selection. For CDSs that were detected under positive selection from the site-specific model, the branch model was used to detect signatures of positive selection along specific lineages by using three models (one-ratio, free-ratio and two-ratio model). The one-ratio model (m0), assumes the same ω ratio for all branches in the phylogeny. This model was compared to the free-ratio model (m1), which assumes that an independent ω ratio for each branch. The m0 was then compared to the two-ratio model (m2), which assumes that the *Reevesia* clade (set as the foreground branch) has a ω value different from the ratio of the other lineages (set as the background branch). Similarly, the LRT was used to identify positively selected branches in comparisons of m0 vs. m1 and m0 vs. m2 (Whelan and Goldman 1999). In general, genes were considered to be under strong positive selection when $\omega > 1$, under neutral evolution when $\omega = 1$ and under higher pressure of negative selection when $\omega < 1$.

Phylogenetic reconstruction

Outgroups included representatives in eight subfamilies of Malvaceae s.l. and one species of the sister family Dipterocarpaceae. All the samples available for *Pterospermum* (Dombeyoideae, Malvaceae) and *Sterculia* (Sterculioideae, Malvaceae) were included because preliminary morphological investigations suggested the affinities to *Reevesia* (Lindley 1827).

Datasets were separately aligned using MAFFT v.7.037 (Katoh and Standley 2013) as implemented in Geneious v.9.0.2 using default settings and manually adjusted. Phylogenetic relationships were inferred using maximum likelihood (ML) and Bayesian inference (BI). Best-fit nucleotide substitution models were determined using MrModeltest v.2.3 (Nylander 2004) based on the Akaike Information Criterion (AIC). ML analyses were performed using RAxML v.8.2.10 (Stamatakis 2014) with GTR+GAMMA model, and they comprised 100 runs and 1,000 thorough bootstrap (BS) replicates. BI was conducted in MrBayes v.3.2.7 (Ronquist and Huelsenbeck 2003). The Markov chain Monte Carlo (MCMC) was carried out for 6,000,000 generations sampling trees every 1000th generation. The initial 25% of the sampled trees were discarded as burn-in, and the two runs were combined. Both ML analyses and BI analyses were visualized by the FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

Results

Plastome features of *Reevesia*

The number of clean reads in the 11 newly sequenced *Reevesia* plastomes ranged from 6.6 to 17.4 million. Mean read coverage across these accessions ranged from 101 to 248x (Table 1). Assembled plastomes ranged in size from 161,532 (*R. botingensis*; OP856555) to 161,945 bp (*R. glaucophylla*). All plastomes show the typical quadripartite structure of angiosperms, which consists of an LSC with a length between 90,225 and 90,665 bp, an SSC with a length between 20,289 and 20,317 bp; and a pair of IRs with a length between 25,465 and 25,495 bp (Table S1, Fig. 1). The total GC content is nearly similar, 36.8% for 10 samples and 36.9% for *R. botingensis* (OP856555; Table S1). The GC content in the IR regions (42.9–43.0%) is noticeably above that of the LSC (34.6–34.7%) and SSC (31.4–31.5%) regions in each plastome (Table S1).

Each of the sequences encoded 80 protein-coding genes, 30 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes. Seven protein-coding genes (*rps12, rpl2, rpl23, ycf2, ycf15, ndhB*, and *rps7*), seven tRNAs (*trnI-CAU*,

trnI-GAU, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*, and *trnA-UGC*), and all four rRNAs (*rrn4.5, rrn5, rrn16* and *rrn23*) had two copies located at the IR regions (Table S2). Eighteen genes contained one (*atpF, ndhA, ndhB, petB, petD, rpl2, rpl16, rpoC1, rps12, rps16, trnA-UGC, trnG-GCC, trnI-GAU, trnK-UUU, trnL-UAA* and *trnV-UAC*) or two (*clpP* and *ycf3*) introns, and six of these were the tRNA genes (Table S2, Fig. 1). The genomes consist of 56.2–56.3% coding regions (48.9–49.0% protein coding genes and 7.3% RNA genes) and 43.7–43.8% non-coding regions, including both intergenic spacers and introns (Table S1).

The boundaries between the IR and LSC/SSC regions of the 12 *Reevesia* species (15 individuals) and the three species in other genera in Malvaceae s.l. were compared (Fig. 2). At the IRb/LSC boundary, the IR ends within the *rps19* gene so that 6 bp of the 5' end of the gene is duplicated in 14 *Reevesia* plastomes. In *R. botingensis* (MN533972.1), *Pterospermum kingtungense* C.Y. Wu ex H.H. Hsue and *Sterculia lanceolata* Cav., the IRb/LSC boundary shifted into the IGS between the *rps19* and *rp12* genes, and in *Durio zibethinus*, the boundary further shifted into the IGS between *rpl23* and *trnI*. The IRa/LSC boundary is located between the *rpl2* and *trnH* genes in all species except in *Durio zibethinus*, which is located between the *trnI* and *trnH* genes. The IRa/SSC boundary in all 15 *Reevesia* samples, *Durio zibethinus* and *Sterculia lanceolata* are located between *ycf1* and *trnN*, while in *R. botingensis* (MN533972.1) and *Pterospermum kingtungense*, the boundary is located within *ycf1*. The IRb/SSC boundary of all the 17 plastomes are located between *trnN* and *ndhF* genes.

Codon preference analysis

Codon usage patterns of protein coding sequences for the 15 *Reevesia* plastomes were calculated based on the relative synonymous codon usage (RSCU) value. The total sequence sizes of the coding genes for codon analysis were 72,906–79,164 bp and 26,302–26,388 codons were encoded (Table S3). Leucine encoded with the highest number of codons ranging from 2,577 to 2,785, followed by isoleucine, with the number of codons encoded between 2,088 and 2,263. Cysteine is the least abundant (286–305). The RSCU values varied slightly among the 15 *Reevesia* sequences. Thirty codons are used frequently with RSCU > 1 and 32 codons used less frequently with RSCU < 1. UUA showed a preference in all the 15 plastomes. The frequency of use for the start codons AUG and UGG, encoding methionine and tryptophan, showed no bias (RSCU = 1). Codons with A (31.84%) or U (37.92%) in the third position are all 69.76%, thus the codon usage is biased toward A or U at the third codon position.
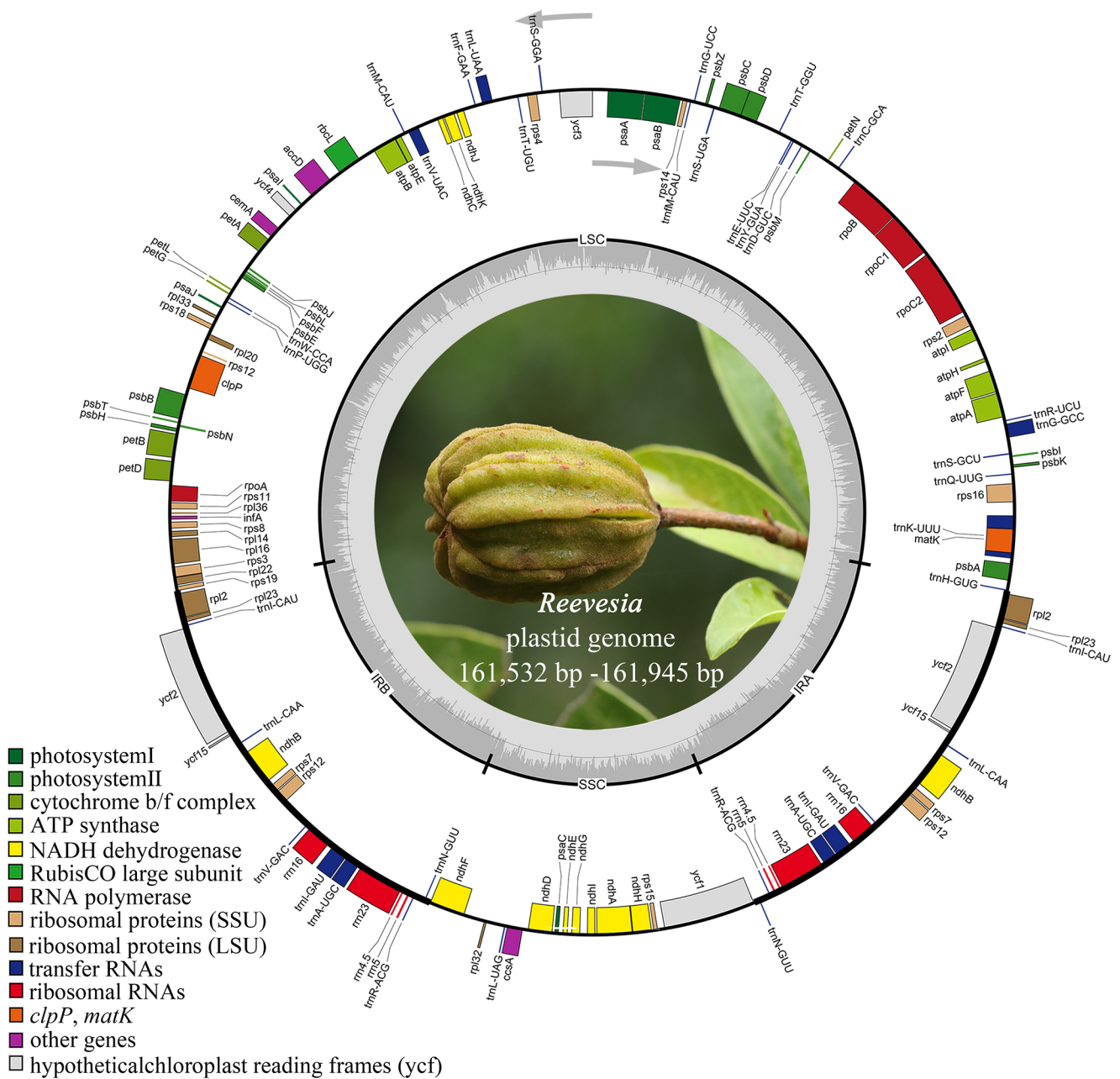
Fig. 1 Gene map of 11 *Reevesia* plastomes. The outer circle shows the genes at each locus, and inverted repeat regions are indicated with thicker lines. Genes on the outside of the outer circle are transcribed in a counterclockwise direction, while genes on the inside of the outer circle are transcribed in a clockwise direction. The inner circle indi-cates the range of the LSC, SSC, and IRs, and also shows a GC content graph of the genome. In the GC content graph, the dark gray lines indicate GC content, while light gray lines indicate the AT content at each locus

Repeated sequences analysis

The total number of SSRs in 15 *Reevesia* plastomes ranged from 35 to 45 (Table S4). Six kinds of SSRs were detected in the fifteen species, namely mono-, di-, tri-, tetra-, penta- and hexa- nucleotides. The most common SSRs are mononucleotides, which range from 14 to 24, followed by tetranucleotides ranging from 7 to 10, and dinucleotides ranging from 4 to 8. The number of trinucleotides and pentanucleotides varied from 2 to 7 and 1 to 3, respectively (Fig. 3a). SSRs are mainly distributed in the intergenic regions (65.0–76.2%), with much lower quantities distributed in the intron regions (15.8–27.5%) and exon regions (6.7–10.8%; Fig. 3b). All mononucleotide and dinucleotide SSRs belong to AT type and the majority of trinucleotide, tetranucleotide, and pentanucleotide SSRs are especially rich in AT type.

Fig. 2 Comparisons of the LSC, SSC and IRs boundaries within *Reevesia* and among three other Malvaceae s.l. plastomes
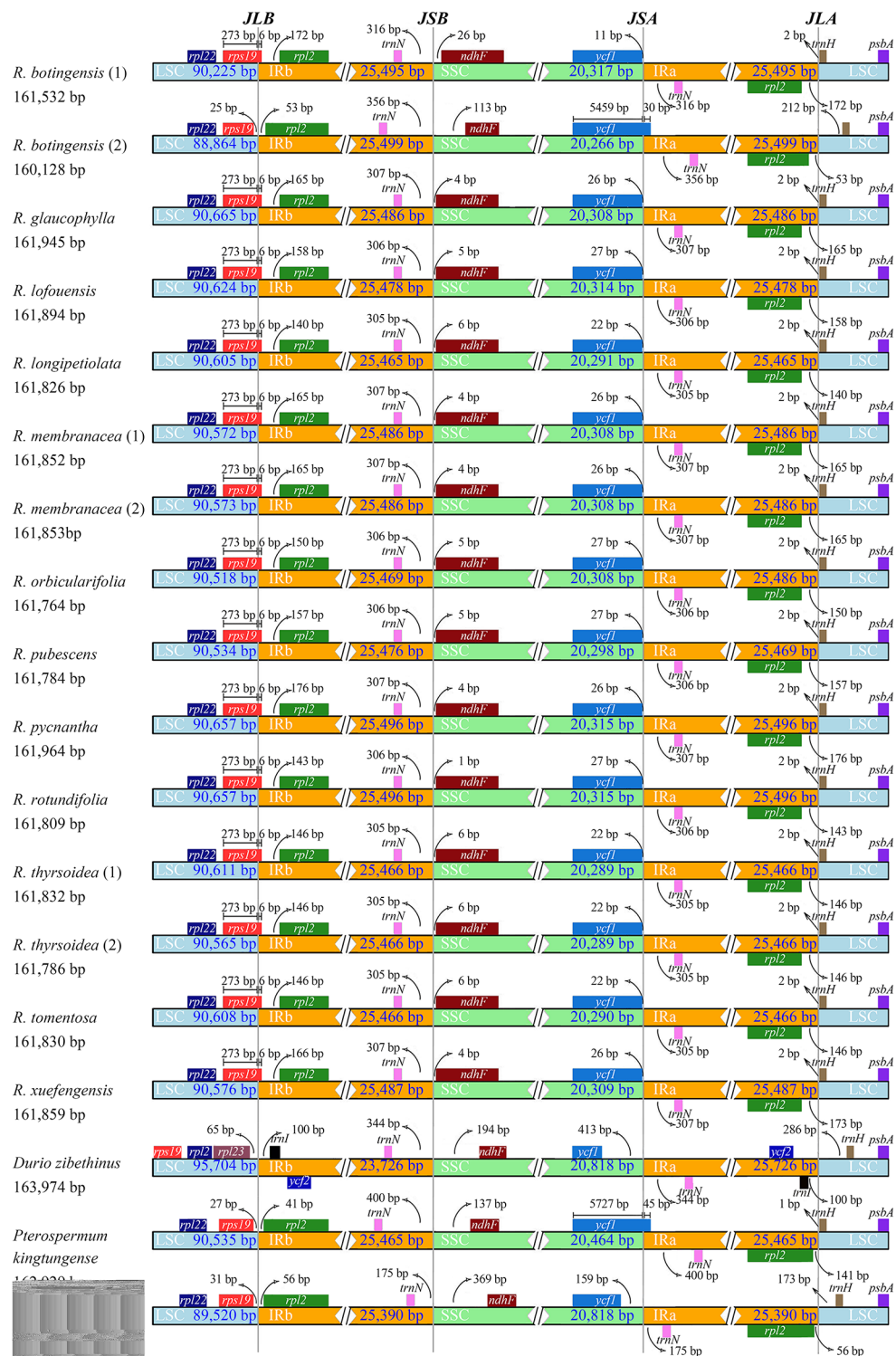
Polymorphic variation and hypervariable regions

Comparative plastome analysis revealed that noncoding regions are generally more divergent than coding regions and LSC/SSC regions are more divergent than IR regions (Fig. 4). Comparison of nucleotide diversity in the LSC, SSC and IR regions indicated that the SSC region shows the highest nucleotide diversity (Pi = 0.00374), followed by LSC (Pi = 0.00269) and the IR (Pi = 0.00171) regions. For the 80 protein-coding regions (CDSs), Pi values for each locus ranged from 0 to 0.00540, with an average of 0.00110, whereas 15 regions (i.e. *infA*, *ycf15*, *rps16*, *psbZ*, *petD*,
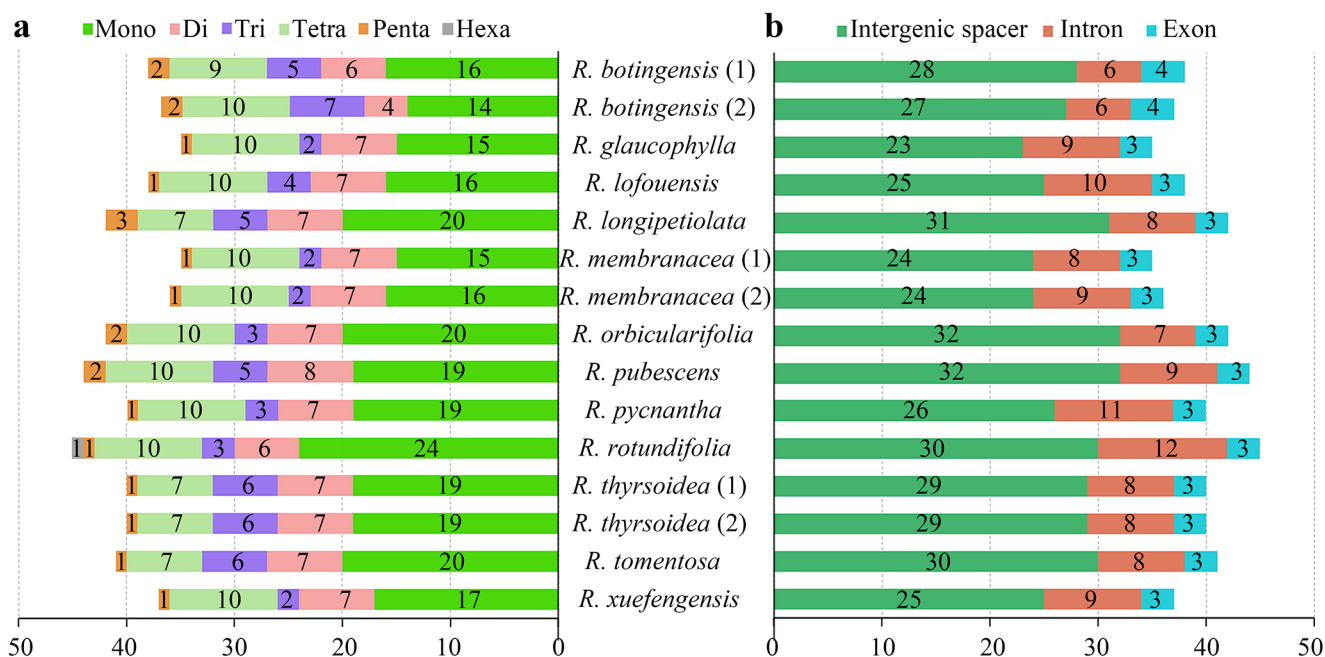
Fig. 3 Distribution of repeats in *Reevesia* plastomes. **a**. Distribution of SSRs types. **b**. Distribution of SSRs among intergenic spacer, intron and exon regions

ndhF, rpl32, ycf1, rps15, petN, ndhH, ccsA, rpl20, psbI and *rpl36*) had relatively high values (Pi > 0.002; see Fig. 5a and Table S5). Across all 133 noncoding regions, Pi values ranged from 0 to 0.04063, with an average of 0.00363. Likewise, thirteen of those regions had relatively high values (Pi > 0.007; see Fig. 5b and Table S5), i.e. *rps19-rpl2, ccsA-ndhD, psbZ-trnG, rps2-rpoC2, infA-rps8, ndhH-rps15, trnH-psbA, petN-psbM, ndhD-psaC, rpl33-rps18, trnG-trnR, trnD-trnY* and *trnM-atpE*.

Selective pressure analysis

Analysis of 80 protein-coding sequences across 15 *Reevesia* plastomes indicated that all CDSs were subjected to purifying selection (ω<1; Table S6). Fifty-five genes are under strong purifying selection with the dN/dS ratio of 0.000. No coding sequences with positively selected sites were identified, even in the seven genes with relatively high dN/dS ratios (ω>0.5): *ndhD* (ω=0.968), *petD* (ω=0.882), *accD* (ω=0.829), *rpoA* (ω=0.721), *ycf4* (ω=0.658), *rpl20* (ω=0.574) and *matK* (ω=0.509).

Phylogenetic analysis

Thirty-nine sequences were used in the phylogenetic analyses. Eleven out of 39 sequences were newly generated (Table S7). The topologies of the trees resulting from ML and BI analyses based on whole plastid genomes are congruent. There are only slight differences in support values

between the phylogenetic trees. Therefore, only the ML topology is shown here with the BI support values added at each node (Fig. 6).

The whole plastid genome analysis strongly supported the monophyly of Malvaceae s.l. and its two major phylogenetic lineages, i.e., Byttneriina and Malvadendrina. Byttneriina includes subfamilies Byttnerioideae and Grewioideae, near the base of the plastome phylogeny, and appears as sister to the remaining subfamilies. Six monophyletic branches that correspond to Malvoideae, Bombacoideae, Dombeyoideae, Helicteroideae, Sterculioideae and Tilioideae, nested in one lineage Malvadendrina.

Reevesia forms a well-supported monophyletic group (ML=100, PP=1.0). The genus is sister to *Durio zibethinus* and nested well within subfamily Helicteroideae (Fig. 6). Within *Reevesia*, three full-supported clades are resolved. Clade A comprises two samples of *R. botingensis*, which were resolved as sisters to the remainder of the genus. Clade B includes the type of the genus, *R. thyrsoidea*, together with *R. longipetiolata* and *R. tomentosa*. Clade C consists of eight taxa, i.e., *R. glaucophylla, R. lofouensis* Chun & Hsue, *R. membranacea* Hsue, *R. orbicularifolia* Hsue, *R. pubescens, R. xuefengensis* (C.J. Qi) C.J. Qi, *R. pycnantha* and *R. rotundifolia*.

Fig. 4 Comparison of plastomes of 15 *Reevesia* samples. Purple blocks indicate conserved genes, while red blocks indicate noncoding sequences (CNS). White blocks represent regions with sequence variation among the 15 plastomes. Gray arrows indicate the direction of gene transcription
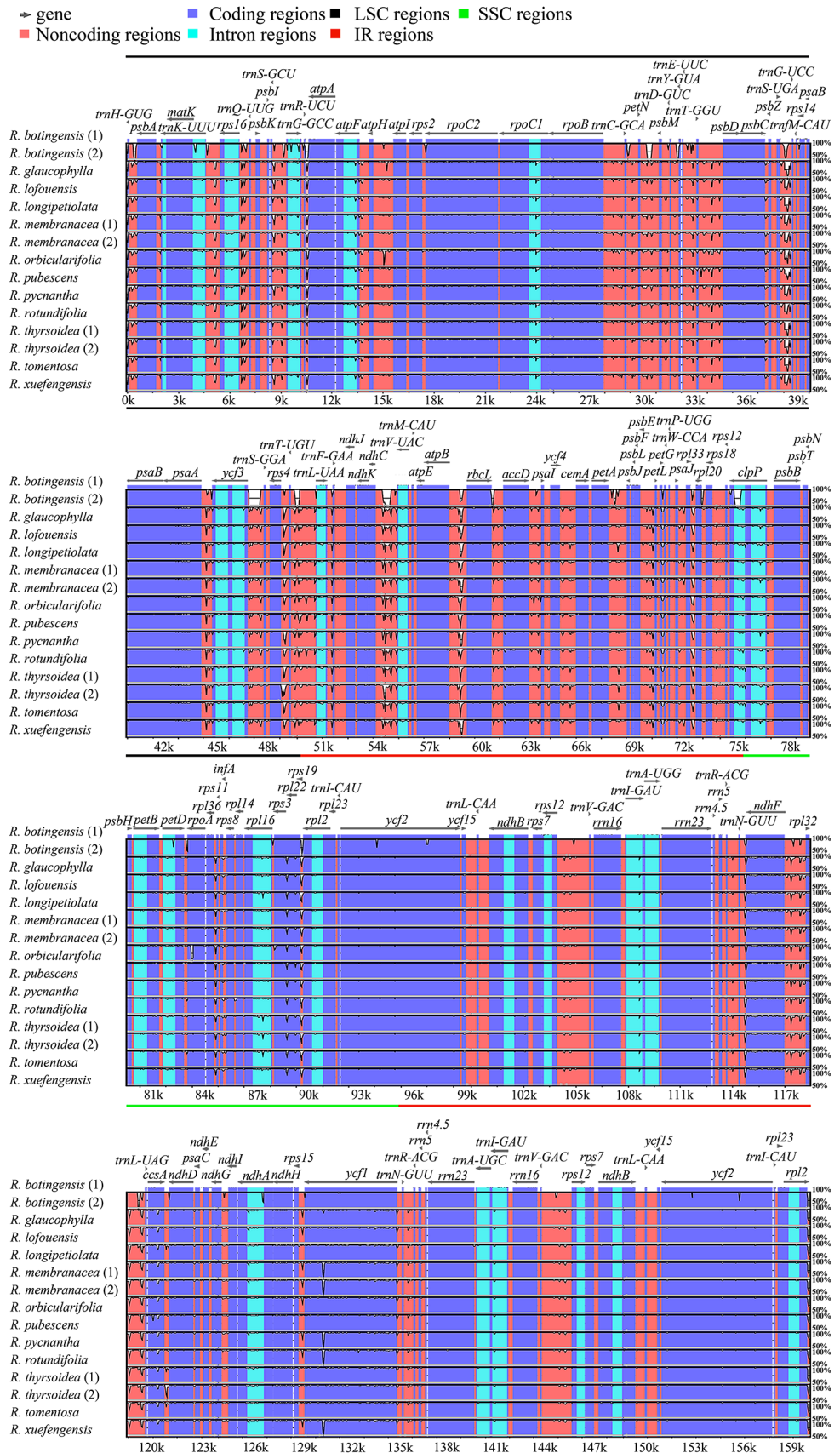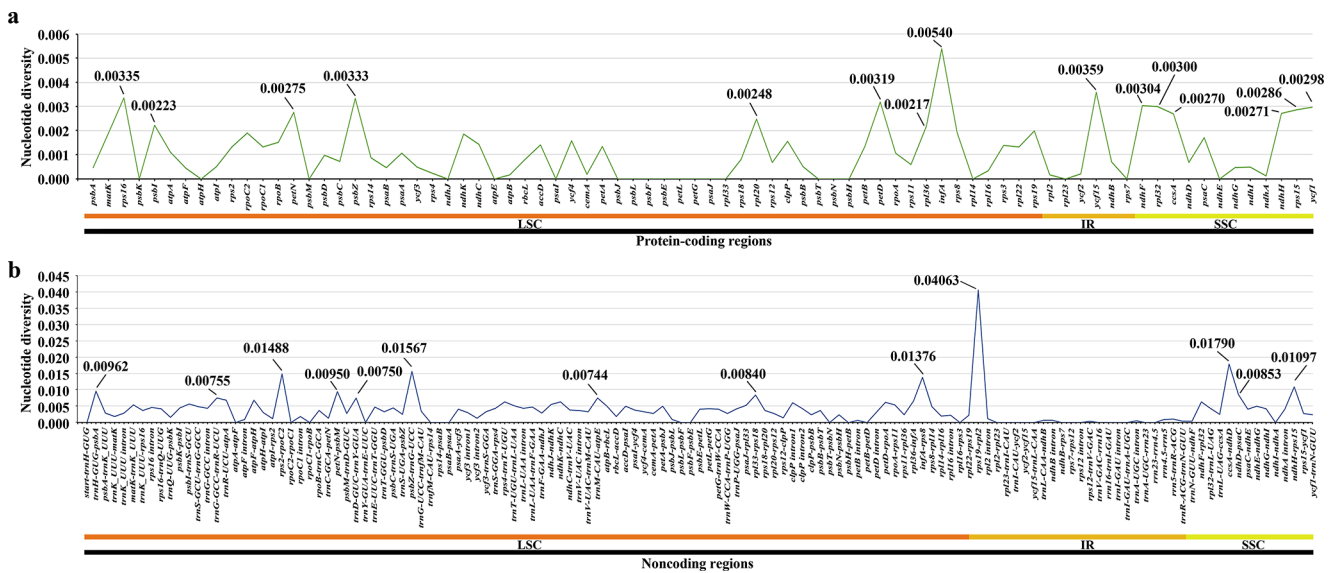
Fig. 5 Comparison of nucleotide variability (Pi) values in *Reevesia* plastomes. **a**. Pi values among protein-coding genes (CDS). **b**. Pi values among intergenic spacer (IGS) regions

Discussion

Plastome structure comparisons

The comparative analysis of plastomes provides new insights into sequence variation and molecular evolutionary patterns. In the present study, 15 *Reevesia* plastomes representing 12 species were compared to 23 plastomes reported in Malvaceae s.l. to clarify phylogenetic relationships and resolve taxonomic uncertainties.

In terms of plastome size, the 11 newly sequenced *Reevesia* samples (161,532–161,945 bp) are slightly larger than most plastomes reported in Malvaceae s.l. (Fig. 1; Cai et al. 2015; Cheon et al. 2017; Heckenhauer et al. 2019; Lee et al. 2006; Wang et al. 2021). All the newly sequenced *Reevesia* plastomes contain 132 genes, which is consistent with plastomes reported previously in the genus (Quan et al. 2020; Wang et al. 2021; Zhang et al. 2022b). Indeed, the gene content and order of the *Reevesia* plastomes are highly conserved. The GC content of the plastomes newly sequenced (36.8–36.9%) are in accordance with that of other reported *Reevesia* plastomes and other Malvales (Heckenhauer et al. 2019; Wang et al. 2021).

The IR regions are resistant to re-combinational loss and therefore help in the stabilization of the plastome (Perry and Wolfe 2002). The IR boundaries are highly conserved among species in the genus *Reevesia* which all possess 18 completely duplicated genes. However, shifts in IR boundaries were detected in different genera of Malvaceae s.l. The IR/LSC of *Reevesia* have expanded relative to the other three genera *Durio*, *Pterospermum* and *Sterculia* to include

a small portion of *rps19*, which is consistent with the results of recent studies (Wang et al. 2021).

Codon usage preference is closely related to gene expression, mutation pressure on DNA sequences and natural selection (Lyu and Liu 2020; Pfitzinger et al. 1987; Zhou et al. 2013). Analysis of the *Reevesia* plastomes revealed that leucine was the most abundant and cysteine was the least abundant amino acid, which has also been frequently reported in the chloroplast genomes of other angiosperms (Ren et al. 2022; Somaratne et al. 2020; Wen et al. 2021). The bias for high representation of A and U at third codon position is observed in this study and is also shown in other land plant plastomes (Moghaddam et al. 2022; Wen et al. 2021).

SSRs are short stretches which have tandemly repeated DNA motifs of 1–6 nucleotides and are often used as genetic markers to identify related species, to explore population relationships and to explore evolutionary history due to their high polymorphism levels (Li et al. 2019; Powell et al. 1995; Zhang et al. 2012). Six types of SSRs were detected in *Reevesia* plastomes, of which A/T mononucleotides were the most common as reported in previous studies (Ren et al. 2022). SSRs in *Reevesia* are mainly restricted to intergenic spacers in LSC regions. Furthermore, the similarity across the twelve species is supported by SSRs alignment.

All the protein-coding genes are under purifying selection in the *Reevesia* plastomes compared. However, eight of them (*ndhD*, *petD*, *accD*, *rpoA*, *ycf4*, *rpl20* and *matK*) evolved rapidly with relatively high dN/dS ratio. Among these genes, positively selected sites were detected in *accD* by Wang et al. (2021) and in *petD* by Wu et al. (2018) in Malvaceae s.l. plastomes. Purifying selection is responsible
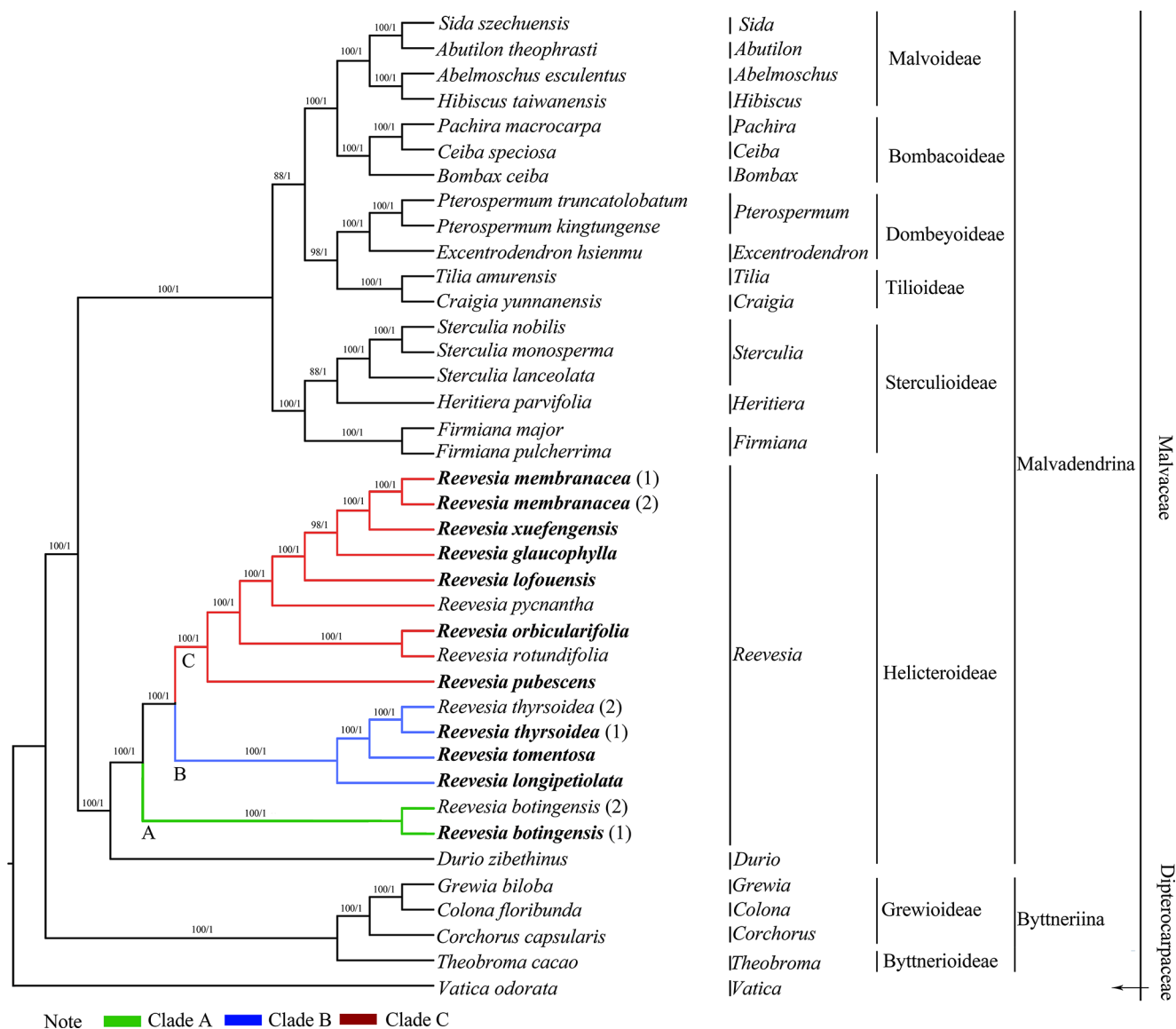
Fig. 6 Phylogenetic tree based on plastomes resulting from the maximum likelihood analysis (ML) with Bayesian inference (BI) value at nodes. Names of taxa newly sequenced in *Reevesia* are in bold

for genomic sequence conservation across long evolutionary timescales (Cvijović et al. 2018) and can be attributed to morphological similarities and habitat homogeneity among *Reevesia* species.

The nucleotide diversity was higher in the non-coding regions than the coding regions, which is generally consistent with most previous studies of the chloroplast genomes of angiosperms (Cao et al. 2022; Ye et al. 2018). At present, data from non-coding regions is the most commonly used tool for phylogeographic and phylogenetic studies at low taxonomic levels of plants (Mapaya and Cron 2021; Shaw et al. 2014). The *rps19-rpl2*, *ccsA-ndhD*, *psbZ-trnG*, *rps2-rpoC2*, *infA-rps8*, *ndhH-rps15*, *trnH-psbA*, *petN-psbM*, *ndhD-psaC*, and *rpl33-rps18* are identified as the top ten

variable regions in the non-coding regions at the species level within *Reevesia* (Figs. 3 and 4). Among them, two, *rps19-rpl2* and *ccsA-ndhD* were identified as the best possible choices for low-level phylogenetic studies of Malvaceae s.l. (Wang et al. 2021). However, to determine whether these variable regions could serve as powerful DNA barcodes for species delimitation and identification in the genus, a broad sampling of well-dispersed species across the genus need to be included for further studies.

The cp. genome sequences of three species with two samples were compared in detail. *R. thyrsoidea* and *R. membranacea* were more similar in the IR/LSC boundary pattern, and *R. botingensis* obviously different between the two samples (Fig. 2). *R. botingensis* and *R. membranacea*

exhibited the distinct SSRs patterns, and these distinctions may be helpful for population genetic analysis in *Reevesia* (Fig. 3). All codons of *R. thyrsoidea* and *R. membranacea* had identical RSCU values, 53 of which were different from those of the two *R. botingensis* samples (Table S3). Discrepancies in annotation protocols may eventually explain all the differences in the two chloroplast genomes of *R. botingensis*.

Phylogenetic analysis

Whole plastid genomes are used widely for inferring phylogenetic relationships at the generic (Saski et al. 2005; Simmonds et al. 2021), supra-generic (Heckenhauer et al. 2019; Wang et al. 2021; Ye et al. 2018), interspecific and intraspecific species levels (Song et al. 2022; Wicke et al. 2011). In this phylogenetic analysis of *Reevesia*, the monophyly of the genus and the relationship with other genera in Malvaceae s.l. were investigated. The relationships among species within the genus and the delimitation of some species were also examined.

In consistent with previous studies based on molecular (Alverson et al. 1999; Nyffeler et al. 2005; Wang et al. 2021) morphological (Judd and Manchester 1997) and biogeographical data (Bayer et al. 1999), the 39 species analyses of whole plastid genomes strongly support the monophyly of the Malvaceae s.l. Within Malvaceae s.l., two major lineages, Byttneriina and Malvadendrina, and the eight subfamilies included in this study were found to be monophyletic.

The genus *Reevesia* is characterized by a number of vegetative and reproductive characteristics, such as longitudinal fissured bark, flowers with conspicuous long androgynophores, 5-angled capsules and winged seeds. The monophyly of *Reevesia* is supported here and in previous molecular studies based on whole plastid genomes (Wang et al. 2021) or a few sequences (Bayer et al. 1999; Nyffeler and Baum 2000).

When the genus *Reevesia* was established by Lindley (1827), it was compared with *Pterospermum* and *Sterculia*. However, in our analysis, *Reevesia* is sister to *Durio zibethinus* in Durioneae, and did not nest in the same lineage with *Pterospermum* and *Sterculia*. It is in accordance with the result of Wang et al. (2021) based on whole plastid genomes data. Previous studies based on a few molecular sequences (Alverson et al. 1999; Bayer et al. 1999; Hernández-Gutiérrez and Magallón 2019; Nyffeler and Baum 2000) have demonstrated that *Reevesia* was sister to the monotypic genus *Ungeria* Schott & Endl. and more closely related to *Helicteres* L., *Mansonia* J.R. Drumm. ex Prain, *Triplochiton* K. Schum. and *Durio* rather than to *Pterospermum* and *Sterculia*. Traditionally, *Pterospermum* was included in

Helictereae (Bentham and Hooker 1862; Hutchinson 1967), or elevated to a separate tribe Pterospermeaeae C.Y. Wu et Y. Tang (Tang 1992), although it is now shown to be distantly related to *Reevesia* and is placed in Dombeyoideae (Alverson et al. 1999; Bayer et al. 1999). *Sterculia*, the type of Sterculiaceae, is also distantly related to *Reevesia* and is included in Sterculioideae.

Within *Reevesia*, three full-supported clades are resolved. Clade I includes *Reevesia botingensis*, which is the first diverging species among the genera. This species is endemic to China, occurring in the Hainan islands, and it can be easily distinguished from other species by having oblanceolate and glabrous leaves. Clade II consists of three species, *R. longipetiolata*, *R. thyrsoidea* and *R. tomentosa* formed a full supported clade B, and the remaining eight taxa were included in clade C. In terms of vegetative and reproductive characters, clade B and C are rather heterogenous and it is difficult to come up with good and consistent synapomorphies. Thus, it remains unclear if this indeed represents the basal split in *Reevesia*.

Three species with two samples, *Reevesia botingensis*, *R. thyrsoidea* and *R. membranacea*, are all supported as monophyletic. However, the recent treatment of two species, *R. pubescens* by Hsuee (1984b) and Tang et al. (2007)d *thyrsoidea* by Feng et al. (2022), are not supported here. Both Hsuee (1984b) and Tang et al. (2007) treated *R. membranacea* as a synonym of *R. pubescens* and *R. xuefengensis* as a variety of *R. pubescens*. In our analysis, both *R. membranacea* and *R. xuefengensis* are not clustered with *R. pubescens*. *Reevesia pubescens* located at the base of clade C, while *R. membranacea* was sister to *R. xuefengensis* and located at the distal end of clade C. When published, *R. membranacea* was considered to be different from other species in the genus in having the membranous leaves and biauriculate protuberances at the middle part of petals (Hsue, 1963). *Reevesia xuefengensis* was first described as a variety of *R. pubescens* by Qi (1984). In 2000, Qi himself elevated it to a species rank. It is obviously different from *R. pubescens* by having the deciduous and lately glabrous leaves. Recently, Feng et al. (2022) merged four species, i.e., *R. botingensis*, *R. lofouensis*, *R. longipetiolata* and *R. pycnantha*, with *R. thyrsoidea*, all of which were recognized by Hsuee (1984b) and Tang et al. (2007) as independent species. Our results supported treating them as independent species. Morphologically, *R. botingensis* differs from *R. thyrsoidea* in having the oblanceolate leaves; *R. lofouensis* differs in the present of densely yellowish stellate pubescent on the branchlets, *R. longipetiolata* differs in having cuneate leaf base and *R. pycnantha* differs in having papery and deciduous leaves. Furthermore, these five species can be distinguished from each other in blooming time: *R. botingensis* flowers from late December to February; *R. longipetiolata* flowers in early March; *R. lofouensis* flowers in mid-May while *R. thyrsoidea* flowers in mid-April in

the evergreen broad-leaved forest of Luofu Mountain; and R. pycnantha flowers from May to July. Therefore, we propose to treat them as distinct species.

Conclusions

In summary, 15 plastomes representing 12 *Reevesia* species were analyzed here. Although plastomes of *Reevesia* show a conservative quadripartite structure, the top ten hypervariable regions identified in noncoding regions, and the long repeated sequenced and SSR screened are informative for species delimitation and identification, phylogenetic and population genetic studies in this genus. A phylogenetic analysis based on 39 plastomes supported the monophyly of *Reevesia*, the eight subfamilies of Malvaceae s.l. included and Malvaceae s.l. Moreover, the analysis indicated that *Reevesia* forms a natural group and is sister to *Durio* within subfamily Helicteroideae. Redefining of *R. pubescens* and *R. thyrsoidea* is proposed for the two species newly delimitated are not monophyletic. Overall, the results of this study provide better understanding of species delimitation and the phylogenetic position of the genus *Reevesia*.

Declarations

References

Aldrich J, Cherney B, Merlin E, Williams C, Mets L (1985) Recombination within the inverted repeat sequences of the Chlamydomonas reinhardii chloroplast genome produces two orientation isomers. Curr Genet 9:233–238. https://doi.org/10.1007/BF00420317

Aldrich J, Cherney B, Williams C, Merlin E (1988) Sequence analysis of the junction of the large single copy region and the large inverted repeat in the petunia chloroplast genome. Curr Genet 14:487–492. https://doi.org/10.1007/BF00521274

Alverson WS, Karol KG, Baum DA, Chase MW, Swensen SM, McCourt R, Sytsma R (1998) Circumscription of the Malvales and relationships to other Rosidae: evidence from *rbcL* sequence data. Am J Bot 85:876. https://doi.org/10.2307/2446423

Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA (1999) Phylogeny of the core Malvales: evidence from *ndhF* sequence data. Am J Bot 86:1474–1486. https://doi.org/10.2307/2656928

Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34:3030–3031. https://doi.org/10.1093/bioinformatics/bty220

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Anthony J (1926) The Genus *Reevesia* Lindl. Edinburgh (ed) The Royal Botanic Garden, Edinburgh, vol 15. Her Majesty's Stationery Office, Edinburgh, pp 121–129. Royal Botanic Garden

Barrett CF, Baker WJ, Comer JR et al (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol 209:855–870. https://doi.org/10.1111/nph.13617

Bayer C, Dorr LJ (1999b) A synopsis of the neotropical genus Pentaplaris, with remarks on its systematic position within core Malvales. Brittonia 51:134–148. https://doi.org/10.2307/2666620

Bayer C, Kubitzki K (2003) Malvaceae. In: Kubitzki K (ed) The families and genera of vascular plants. Flowering plants·dicotyledons malvales, capparales and non-betalain caryophyllales, vol 5. Springer, Berlin, pp 225–311

Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW (1999a) Support for an expanded family concept of Malvaceae with a recircumscribed order Malvales: a combined analysis of plastid atpB and rbcL DNA sequences. Bot J Linn Soc 129:267–303. https://doi.org/10.1111/j.1095-8339.1999.tb00505.x

Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. Plant Cell 16:1661–1666. https://doi.org/10.1105/tpc.160771

Bentham G, Hooker JD (1862) Genera Plantarum ad exemplaria imprimis in herbariis kewensibus servata, vol 2. Reeve & Company, London

Brudno M, Do CB, Cooper GM et al (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13:721–731. https://doi.org/10.1101/gr.926603

Byng JW, Chase MW, Christenhusz MJM, Fay MF, Judd WS, Mabberley DJ, Sidow A, Batzoglou S (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181:1–20. https://doi.org/10.1111/boj.12385

Cai J, Ma P-F, Li H-T, Li D-Z (2015) Complete plastid genome sequencing of four *Tilia* species (Malvaceae): a comparative analysis and phylogenetic implications. PLoS ONE 10:e0142705. https://doi.org/10.1371/journal.pone.0142705

Cao Q, Gao QB, Ma XL, Zhang FQ, Xing R, Chi XF, Cen SL (2022) Plastome structure, phylogenomics and evolution of plastid genes

in *Swertia* (Gentianaceae) in the Qing-Tibetan Plateau. BMC Plant Biol 22:195. https://doi.org/10.1186/s12870-022-03577-x

Chang H-S, Chiang MY, Hsu H-Y, Yang C-W, Lin C-H, Lee S-J, Chen I-S (2013) Cytotoxic cardenolide glycosides from the root of *Reevesia Formosana*. Phytochemistry 87:86–95. https://doi.org/10.1016/j.phytochem.2012.11.024

Chen YX, Chen YS, Shi CM et al (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience 7:1–6. https://doi.org/10.1093/gigascience/gix120

Cheon SH, Jo S, Kim HW, Kim YK, Sohn JY, Kim KJ (2017) The complete plastome sequence of *durian*, *Durio zibethinus* L. (Malvaceae). Mitochondrial DNA Part B Resources 2:763–764. https://doi.org/10.1080/23802359.2017.1398615

Craib WG (1924) Contributions to the flora of Siam. Additamentum XIV. In: Royal Gardens K (ed) Bulletin of miscellaneous information. Royal Gardens, Kew, London, p 90

Cronquist A (1988) The evolution and classification of flowering plants, 2nd edn. New York Botanical Garden, New York

Cvijović I, Good BH, Desai MMV (2018) The effect of strong purifying selection on genetic diversity. Genetics 209:1235–1278. https://doi.org/10.1534/genetics.118.301058

Dorr LJ (2017) Two innovaions in Mexican Malvaceae. J Bot Res Inst Tex 11:25–28

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Du ZY, Lu K, Zhang K, He YM, Wang HT, Chai GQ, Shi JG, Duan YZ (2021) The chloroplast genome of *Amygdalus* L. (Rosaceae) reveals the phylogenetic relationship and divergence time. BMC Genomics 22:645. https://doi.org/10.1186/s12864-021-07968-6

Feng H-Z, Meng K-K, Wang L-Y (2022) Taxonomic studies on the genus *Reevesia* (Malvaceae: Helicteroideae) I: the identities of *R. thyrsoidea*. Phytotaxa 547:001–019. https://doi.org/10.11646/phytotaxa.547.1.1

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:W273–W279. https://doi.org/10.1093/nar/gkh458

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47:W59–W64. https://doi.org/10.1093/nar/gkz238

Heckenhauer J, Paun O, Chase MW, Ashton PS, Kamariah AS, Samue R (2019) Molecular phylogenomics of the tribe *Shoreeae* (Dipterocarpaceae) using whole plastid genomes. Ann Bot 123:857–865. https://doi.org/10.1093/aob/mcy220

Hernández-Gutiérrez R, Magallón S (2019) The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. Mol Phylogenet Evol 140:106606. https://doi.org/10.1016/j.ympev.2019.106606

Hsue HH (1963) Materials for the flora of Hainan (II) *Reevesia* Lindl. Acta Phytotaxonomica Sinica 8:271–275

Hsue HH (1984a) Chinese sterculiaceous plants and their economic significance. Guihaia 4:149–155

Hsue HH (1984b) Sterculiaceae. In: Wu ZH (ed) Flora republicae popularis sinicae, vol 49. Science, Beijing, pp 112–189

Hutchinson J (1967) Key to the families of flowering plants of the world. Clarendon, Oxford

Judd WS, Manchester SR (1997) Circumscription of Malvaceae (Malvales) as determined by a preliminary cladistic analysis of morphological, anatomical, palynological, and chemical characters. Brittonia 49:384–405. https://doi.org/10.2307/2807839

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir R, Wilson A et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Lee SB, Kaittanis C, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) The complete chloroplast genome sequence of Gossypium hirsutum: organization and phylogenetic relationships to other angiosperms. BMC Genomics 7:61. https://doi.org/10.1186/1471-2164-7-61

Li D-M, Zhao C-Y, Liu X-F (2019) Complete chloroplast genome sequences of *Kaempferia Galanga* and *Kaempferia Elegans*: molecular structures and comparative analysis. Molecules 24:474. https://doi.org/10.3390/molecules24030474

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lindley J (1827) An account of a new genus of plants called *Reevesia*. Q J Sci Literature Art 2:109–112

Long H, He LK, Hsue HH (1985) Pollen morphology of Sterculiaceae from China (II) *Reevesia* Lindl. J South China Agric Univ 6:24–34

Luo C, Huang WL, Sun HY, Yer HY, Li XY, Li Y, Yan B, Wang Q, Wen YH, Huang MJ, Huang HQ (2021) Comparative chloroplast genome analysis of Impatiens species (Balsaminaceae) in the karst area of China: insights into genome evolution and phylogenomic implications. BMC Genomics 22:571. https://doi.org/10.1186/s12864-021-07807-8

Lyu XL, Liu Y (2020) Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. Mol Biol Physiol 11:e02605–e02620. https://doi.org/10.1128/mBio.02605-20

Mabberley DJ (1987) The plant-book: a portable dictionary of higher plants. Cambridge University Press, Cambridge

Mabberley DJ (2017) Mabberley's plant-book: a portable dictionary of plants, their classification and uses, 4th edn. Cambridge University Press, Cambridge

Machuca MK (2017) Ayenia Grisea (Malvaceae-Byttnerioideae), a new species for Mexico and validation of *Reevesia Clarkii* (Malvaceae/Helicteroideae). Acta Bot Mexicana 120:113–120. https://doi.org/10.21829/abm120.2017.1187

Mapaya RJ, Cron GV (2021) A phylogeny of Emilia (Senecioneae, Asteraceae) - implications for generic and sectional circumscription. Taxon 70:127–138. https://doi.org/10.1002/TAX.12417

Moghaddam M, Ohta A, Shimizu M, Terauchi R, KazempourOsaloo S (2022) The complete chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): comparative analysis with related IR-lacking clade species. BMC Plant Biol 22:75. https://doi.org/10.1186/s12870-022-03465-4

Monachino J (1940) A new genus and species of Sterculiaceae. Bull Torrey Bot Club 67:621–622. https://doi.org/10.2307/2481584

Nyffeler R, Baum DA (2000) Phylogenetic relationships of the durians (Bombacaceae-Durioneae or /Malvaceae/Helicteroideae/Durioneae) based on chloroplast and nuclear ribosomal DNA sequences. Plant Syst Evol 224:55–82. https://doi.org/10.1007/BF00985266

Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock BA, Chase MW, Baum DA (2005) Phylogenetic analysis of the malvadendrina clade (Malvaceae s.l.) based on plastid DNA sequences. Organisms Divers Evol 5:109–123. https://doi.org/10.1016/j.ode.2004.08.001

Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Uppsala University, Evolutionary Biology Centre, Uppsala. https://github.com/nylander/MrModeltest2

Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted

repeat. J Mol Evol 55:501–508. https://doi.org/10.1007/s00239-002-2333-y

Pfitzinger H, Guillemaut P, Weil JH, Pillay DTN (1987) Adjustment of the tRNA population to the codon usage in chloroplasts. Nucleic Acids Res 15:1377–1386. https://doi.org/10.1093/nar/15.4.1377

Phengklai C (2001) Sterculiaceae. In: Santisuk T, Larsen K (eds) Flora of Thailand, vol 7. The Forest Herbarium, Royal Forest Department, Bangkok, pp 351–654

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763. https://doi.org/10.1073/pnas.92.17.7759

Qi CJ (1984) New Taxa from Hunan. J Univ Chin Acad Sci 22:493–494

Qu X-J, Moore MJ, Li D-Z, Yi T-S (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods 15:50. https://doi.org/10.1186/s13007-019-0435-7

Quan GM, Zou PS, Liu GF, Sun MM, Wang W, Dai SP (2020) The complete chloroplast genome of *Reevesia thyrsoidea* (Malvaceae). Mitochondrial DNA Part B Resour 5:292–293. https://doi.org/10.1080/23802359.2019.1674206

Ren T, Li Z-X, Xie D-F, Gui L-J, Peng C, Wen J, He X-J (2020) Plastomes of eight *Ligusticum* species: characterization, genome evolution, and phylogenetic relationships. BMC Plant Biol 20:519. https://doi.org/10.1186/s12870-020-02696-7

Ren J, Tian J, Jiang H, Zhu X-X, Mutie FM, Wanga VO, Ding S-X, Yang J-X, Dong X, Chen L-L, Cai X-Z, Hu G-W (2022) Comparative and phylogenetic analysis based on the chloroplast genome of *Coleanthus subtilis* (Tratt.) Seidel, a protected rare species of monotypic genus. Front Plant Sci 13:828467. https://doi.org/10.3389/fpls.2022.828467

Richardson JE, Whitlock BA, Meerow AW, Madrinan S (2015) The age of chocolate: a diversification history of Theobroma and Malvaceae. Front Ecol Evol 3:120. https://doi.org/10.3389/fevo.2015.00120

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497. https://doi.org/10.1093/bioinformatics/btg359

Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei WB, Thomas PI, Gardner MF, Ennos RA, Hollingsworth PM (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? Mol Ecol Resour 15:1067–1078. https://doi.org/10.1111/1755-0998.12375

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK (2005) Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Mol Biol 59:309–322. https://doi.org/10.1007/s11103-005-8882-0

Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot 101:1987–2004. https://doi.org/10.3732/ajb.1400398

Simmonds SE, Smith JF, Davidson C, Buerki S (2021) Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, *Piper* and *Peperomia* (Piperaceae). Mol Phylogenet Evol 163:107229. https://doi.org/10.1016/J.YMPEV.2021.107229

Solheim SL (1991) *Reevesia* and *Ungeria* (Sterculiaceae): a taxonomic and biogeographic study. University of Wisconsin, Madison

Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ (2020) The complete chloroplast genomes of two *Lespedeza* species: insights into

codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). Plants 9:51. https://doi.org/10.3390/plants9010051

Song WC, Ji CX, Chen ZM, Cai HH, Wu XM, Shi C, Wang S (2022) Comparative analysis the complete chloroplast genomes of nine *Musa* species: genomic features, comparative analysis, and phylogenetic implications. Front Plant Sci 13:832884. https://doi.org/10.3389/fpls.2022.832884

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stevens PF (2001) Angiosperm Phylogeny Website. Version 14. http://www.mobot.org/MOBOT/research/APweb/

Takhtajan A (1997) Diversity and classification of flowering plants. Columbia University, New York

Tang Y (1992) On the affinities of *Pterospermum Schreb* (Sterculiaceae). Guihaia 12:8–14

Tang Y, Gilbert MG, Dorr LJ (2007) Sterculiaceae. In: Wu ZY, Raven PH, Hong DY (eds) Flora of China, vol 12. Science Press, Beijing, Missouri Botanical Garden, St. Louis, pp 302–330

Terada K, Suzuki M (1998) Revision of the so-called '*Reevesia*' fossil woods from the Tertiary in Japan — a proposal of the new genus Wataria (Sterculiaceae). Rev Palaeobot Palynol 103:235–251. https://doi.org/10.1016/S0034-6667(98)00039-6

Thorne RF (1992) Classification and geography of the flowering plants. Bot Rev 58:225–348. https://doi.org/10.1007/BF02858611

Tian SB, Lu PL, Zhang ZH, Wu JQ, Zhang H, Shen HB (2021) Chloroplast genome sequence of Chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume chloroplast genomes. BMC Genomics 22:194. https://doi.org/10.1186/s12864-021-07467-8

Tiffney BH (1985) Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. J Arnold Arboretum 66:73–94. https://doi.org/10.5962/bhl.part.13179

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135. https://doi.org/10.1038/nrg1271

Wang Y, Zhan D-F, Jia X, Dai H-F, Mei W-L, Chen X-T, Peng S-Q (2016) Complete chloroplast genome sequence of Aquilaria sinensis (Lour.) Gilg and evolution analysis within the Malvales order. Front Plant Sci 7(280). https://doi.org/10.3389/fpls.2016.00280

Wang J-H, Moore MJ, Wang HX, Zhu Z-X, Wang H-F (2021) Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene 765:145103. https://doi.org/10.1016/j.gene.2020.145103

Wen J (1999) Evolution of eastern Asian and eastern north American disjunct distributions in flowering plants. Annu Rev Ecol Syst 30:421–455. https://doi.org/10.1146/annurev.ecolsys.30.1.421

Wen J, Ickert-Bond S, Nie ZL, Li R (2010) Timing and modes of evolution of eastern asian-north American biogeographic disjunctions in seed plants. In: Long M, Gu H, Zhou Z (eds) Darwin's heritage today. Higher Education Press, Beijing, pp 252–269

Wen F, Wu XZ, Li TJ, Jia ML, Liu XS, Liao L (2021) The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. BMC Genomics 22:161. https://doi.org/10.1186/s12864-021-07484-7

Whelan S, Goldman N (1999) Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol Biol Evol 16:1292–1299. https://doi.org/10.1093/oxfordjournals.molbev.a026219

Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants:

Springer

gene content, gene order, gene function. Plant Mol Biology 76:273–297. https://doi.org/10.1007/s11103-011-9762-4

Wolf JA (1980) Tertiary climates and floristic relationships at high latitudes in the Northern Hemisphere. Palaeogeogr Palaeoclimatol Palaeoecol 30:313–323. https://doi.org/10.1016/0031-0182(80)90063-2

Wu ZY (1991) The areal-types of Chinese genera of seed plants. Acta Bot Yunnanica 1:1–139

Wu C-S, Chaw S-M (2016) Large-scale comparative analysis reveals the mechanisms driving plastomic compaction, reduction, and inversions in conifers II (Cupressophytes). Genome Biol Evol 8:3740–3750. https://doi.org/10.1093/gbe/evw278

Wu ZY, Lu AM, Tang CY, Chen ZD, Li DZ (2003) The families and genera of angiosperms in China: a comprehensive analysis. Science, Beijing

Wu Y, Liu F, Yang D-G, Li W, Zhou X-J, Pei X-Y, Liu Y-G, He K-L, Zhang W-S, Ren Z-Y, Zhou K-H, Ma X-F, Li Z-H (2018) Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. Front Plant Sci 9:376. https://doi.org/10.3389/fpls.2018.00376

Xie D-F, Yu H-X, Price M, Xie C, Deng Y-Q, Chen J-P, Yu Y, Zhou S-D, He X-J (2019) Phylogeny of Chinese *Allium* species in section Daghestanica and adaptive evolution of *Allium* (Amaryllidaceae, Allioideae) species revealed by the chloroplast complete genome. Front Plant Sci 10:460. https://doi.org/10.3389/fpls.2019.00460

Xing SC, Liu CJ (2008) Progress in chloroplast genome research. Progress Biochem Biophys 35:21–28

Xu SJ, Hsue HH (2001) Geographical distribution of Sterculiaceae. J Trop Subtropical Bot 9:19–30

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Yang ZH, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specifc lineages. Mol Biol Evol 19:908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148

Yang ZH, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118. https://doi.org/10.1093/molbev/msi097

Ye W-Q, Yap Z-Y, Lia P, Comesb HP, Qiu Y-X (2018) Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Mol Phylogenet Evol 127:978–987. https://doi.org/10.1016/j.ympev.2018.07.001

Zhang Q, Li J, Zhao YB, Korba SS, Han YP (2012) Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. Plant Mol Biol Rep 30:539–546. https://doi.org/10.1007/s11105-011-0366-6

Zhang Y-H, Chen P-D, Chen D-J, Ke X-R, Wang H-F (2022) Complete plastome sequence of *Reevesia pycnantha* Y. Ling (Malvaceae): a rare medicinal tree species in South Asia. Mitochondrial DNA Part B Resour 7:343–345. https://doi.org/10.1080/23802359.2022.2032439

Zhang Y-M, Han L-J, Yang C-W, Yin Z-L, Tian X, Qian Z-G, Li G-D (2022b) Comparative chloroplast genome analysis of medicinally important *Veratrum* (Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Plant Divers 44:70–82. https://doi.org/10.1016/j.pld.2021.05.004

Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature 495:111–115. https://doi.org/10.1038/nature11833