Comparative Genomics and Phylogenetic Relationships of Two Endemic and Endangered Species (Handeliodendron Bodinieri and Eurycorymbus Cavaleriei) of two Monotypic Genera within Sapindales

Jiaxin Yang
Chinese Academy of Sciences
Guoxiong Hu
Guizhou university
Guangwan Hu (guangwanhu@wbgcas.cn)
Chinese Academy of Sciences

Research Article

Keywords: Endangered, Monotypic genus, Endemic, Handeliodendron, Eurycorymbus, Chloroplast genome, Phylogenetic analysis, Sapindaceae

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-822059/v1

Version of Record: A version of this preprint was published at BMC Genomics on January 6th, 2022. See the published version at https://doi.org/10.1186/s12864-021-08259-w.

Abstract

Background

Handeliodendron Rehder and *Eurycorymbus* Hand.-Mazz. are the monotypic genera in the Sapindaceae family. The phylogenetic relationship of these endangered species *Handeliodendron bodinieri* (Lévl.) Rehd. and *Eurycorymbus cavaleriei* (Lévl.) Rehd. et Hand.-Mazz. with other members of Sapindaceae s.l. is not well resolved. A previous study concluded that the genus *Aesculus* might be paraphyletic because *Handeliodendron* was nested within it based on small DNA fragments. Thus, their chloroplast genomic information and comparative genomic analysis with other Sapindaceae species are necessary and crucial to understand the circumscription and plastome evolution of this family.

Results

The chloroplast genome sizes of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* are 151,271 and 158,690 bp, respectively. Results showed that a total of 114 unique genes were annotated in *H. bodinieri* and *E. cavaleriei*, and the *ycf1* gene contained abundant SSRs in both genomes. Comparative analysis revealed that gene content, PCGs, and total GC content were remarkably similar or identical within 13 genera from Sapindaceae, and the chloroplast genome size of four genera was generally smaller within the family, including *Acer*, *Dipteronia*, *Aesculus*, and *Handeliodendron*. IR boundaries of the *H. bodinieri* showed a significant contraction, whereas it presented a notable expansion in *E. cavaleriei* cp genome. *Ycf1*, *ndhC-trnV-UAC*, and *rpl32-trnL-UAG-ccsA* were remarkably divergent regions in the Sapindaceae species. Phylogenetic analysis based on different datasets, including whole chloroplast genome sequences, coding sequences, large single-copy, small single-copy, and inverted repeat regions, consistently demonstrated that *H. bodinieri* was sister to the clade consisted of *Aesculus chinensis* and *A. wangii*, strongly support *Eurycorymbus cavaleriei* as sister to *Dodonaea viscosa*.

Conclusion

This study revealed that the cp genome size of the Hippocastanoideae was generally smaller across Sapindaceae, and three highly divergent regions could be used as the specific DNA barcodes within Sapindaceae. Phylogenetic results strongly support that the subdivision of four subfamilies within Sapindaceae, and *Handeliodendron* is not nested within the genus *Aesculus*.

1. Introduction

Handeliodendron bodinieri (Lévl.) Rehd. and *Eurycorymbus cavaleriei* (Lévl.) Rehd. et Hand.-Mazz. (Sapindaceae) are deciduous woody plants endemic to China, and belong to the two monotypic genera *Handeliodendron* and *Eurycorymbus*, respectively [1]. Previous studies revealed that *H. bodinieri* and *E. cavaleriei* are economically important plants. Their seed kernel have been used as raw materials for producing biodiesel, sources of protein and edible oil and are the high medicinal and nutritional values [2, 3]. *H. bodinieri* species are mainly distributed in the south Guizhou province and the northwest of Guangxi Zhuang

Autonomous Region [1], and a small population found in Yunan province, China [4]. *E. cavaleriei* are distributed in Fujian, Guangdong, Guangxi, Guizhou, Hunan, Jiangxi, Sichuan, Taiwan, and Yunnan provinces [5]. The two species are categorized as endangered according to Information System of Chinese Rare and Endangered Plants (ISCREP) (http://www.iplant.cn/rep/). Significantly, *H. bodinieri* has been listed among the level-I state-protected wild plants, whereas *E. cavaleriei* has been classified as the level-II state-protected wild plants in the list. In addition, *E. cavaleriei* was designated as Near Threatened (NT) by the International Union for Conservation of Nature (IUCN) Red List (https://www.iucnredlist.org/species/31353/9628640).

Historically, there were considerable controversies about the circumscription of Sapindaceae, predominantly the traditionally defined Aceraceae and Hippocastanaceae should be incorporated into Sapindaceae or separated. The concept of Sapindaceae s. str. was first proposed by Jussieu in 1789, which supported that the family is different from Aceraceae. After this, the works of Radlkofer maintained the family distinct from Aceraceae and Hippocastanaceae, and proposed the first worldwide system of classification for Sapindaceae [6, 7]. In this classification system of Radlkofer, Sapindaceae was divided into two subfamilies Eusapindaceae and Dyssapindaceae. The Eusapindaceae consists of nine tribes, including Paullinieae, Thouinieae, Sapindeae, Aphanieae, Lepisantheae, Melicocceae, Nephelieae, and Cupanieae, whereas the Dyssapindaceae contains five tribes (including Koelreuterieae, Cossignieae, Doratoxyleae, and Harpullieae). Muller and Leenhouts [8] proposed the revised classification of Sapindaceae based upon the macromorphology and pollen morphology, in which the tribe Aphanieae was incorporated into Lepisantheae. They considered Aceraceae and Hippocastanaceae as related to the Sapindaceae basing on pollen character, thus suggested that they be retained as tribes within the Sapindaceae. However, most classification systems [9–11] maintained Aceraceae and Hippocastanaceae as distinct families. On the basis of phytochemistry, Umadevi and Daniel [12] proposed a system of classification for Sapindaceae. He divided the family into four subfamilies: Aceroideae (including all members of the former Aceraceae), Sapindoideae (including the former Hippocastanaceae as a tribe), Dodoneoideae (*Dodonaea*), and Koelreuterioideae (all the other tribes of Dodoneoideae of Radlkofer). Judd [13] agreed with the broader concept of Sapindaceae, supported the inclusion of both families within Sapindaceae. The system of Thorne [14] divided the family into five subfamilies: Dodonaeoideae, Koelreuterioideae, Sapindoieae, Hippocastanoideae (*Aesculus* and *Billia*), and Aceroideae (*Acer* and *Dipteronia*). Based on plastid genes *matK* and *rbcL*, Harrington et al. [15] conducted the phylogenetic analysis in Sapindaceae, and supported the recognition of a broadly defined Sapindaceae incorporating Aceraceae and Hippocastanaceae. They further proposed that the subdivision of four subfamilies within Sapindaceae: Xanthoceroideae (including the single genus *Xanthoceras*), Hippocastanoideae, Dodonaeoideae, and Sapindoideae. The work of Harrington et al. [15] has been adopted by Thorne et al. [16]. By increasing taxon sampling and expanding DNA data, Buerki et al. [17] revealed the relationships at subfamilial and tribal levels in Sapindaceae. They supported the merge of Aceraceae and Hippocastanaceae into Sapindaceae, and recognized four subfamilies within Sapindaceae (Xanthoceroideae, Hippocastanoideae, Dodonaeoideae, and Sapindoideae). Nevertheless, based on molecular sequence data, morphology and biogeography, Buerki et al. [18] resurrected the traditional families Aceraceae and Hippocastanaceae, further proposed a new family, Xanthoceraceae (including the single genus *Xanthoceras*). The concept of a broadly defined Sapindaceae that includes Aceraceae, Hippocastanaceae, and *Xanthoceras*, was adopted by the newly published Angiosperm Phylogeny Group (APG) [19].

For the *Handeliodendron bodinieri*, Rehder [20] proposed the new genus *Handeliodendron*, and it resembles the Hippocastanaceae in opposite and digitately 5-foliolate leaves. However, other characters showed that it is a closer affinity with the Sapindaece and is best placed in the tribe Harpullieae [20]. The work of Muller and Leenhouts [8] considered that *Delavaya* and *Handeliodendron* are the intermediate between Harpullieae and Hippocastanaceae. Judd [13] supported that *Handeliodendron* should be classified together with members of the Hippocastanaceae. The molecular phylogenetic analysis of Sapindaceae demonstrated that *Handeliodendron* should be transferred from Harpullieae to tribe Hippocastaneae (subfamily Hippocastanoideae) containing *Aesculus* and *Billia*, and strongly supported *H. bodinieri* was sister to the *Aesculus* plus *Billia* [15]. According to *Flora of China*, Xia et al. [21] maintained the Hippocastanaceae as a distinct family, comprised of *Handeliodendron*, *Aesculus*, and *Billia*. Buerki et al. [17] supported that *Handeliodendron*, *Aesculus*, and *Billia* were members of the *Aesculus* group in the subfamily Hippocastanoideae, but their work lacked samples of *Handeliodendron* and *Billia*. By increasing samples and DNA markers, Buerki et al. [18] concluded the genus *Aesculus* might be paraphyletic because the *Handeliodendron* and *Billia* were nested within it, but the relationship presented weak supported (for *Handeliodendron*: BS = 68; for *Billia*: BS = 76). For the genus *Eurycorymbus*, it was a component of the Harpullieae in the subfamily Dodonaeoideae [6, 8]. Molecular analyses indicated that *Eurycorymbus* belongs to a member of the *Dodoneae* group in the Dodonaeoideae, the results strongly supported *E. cavaleriei* was sister to *Euphorianthus longifolius* [17, 18].

Despite the systematic position of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* tending to show more stability within the Sapindaceae, previous studies were mainly based on the morphology and/or limited DNA regions. At present, the next-generation sequencing (NGS) technologies and bioinformatics tools bring great convenience for acquiring and analyzing genome-scale data, while obtaining genome-scale nuclear data remains significantly difficult in terms of expense. On the contrary, the plastid genome has become extremely easy to obtain because of its highly conserved nature and much smaller size. Meanwhile, an increasing number of the chloroplast genomes have been widely applied to solve phylogenetic relationships at different taxonomical levels within angiosperms [22–29]. Besides, recent studies have demonstrated cp genome can serve as super barcodes for species and taxonomic groups [23, 30–33]. Although some cp genomes in Sapindaceae s.l. species have been reported and deposited in GenBank database, like *Acer* [34], *Dodonaea* [35], *Aesculus* [36–39], *Dimocarpus* [40], most of which were the genus *Acer* [41–46]. We also noticed that previous studies have revealed the complete cp genomes of *H. bodinieri* and *E. cavaleriei* [47–49], but focus mainly on the size and gene contents of the cp genome. Moreover, a comprehensive genomics analysis is still a useful framework for understanding phylogenetic relationships within Sapindaceae. Unfortunately, a detailed comparative analysis was further not conducted. Due to insufficient molecular data, the result of Tian, et al [49] showed *H. bodinieri* is closely related to the genus *Mangifera*. However, according to the work of Harrington et al. [15], *H. bodinieri* was closely related to *Aesculus* plus *Billia*. Therefore, it is necessary to perform a comprehensive cp genomic comparison and phylogenetic analysis in the Sapindaceae.

In the current study, we sequenced and assembled the complete chloroplast genomes of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei*. The main objectives of this study were to 1) compare and analyze the gene organizations of *H. bodinieri* and *E. cavaleriei* cp genomes; 2) reveal the cp genome structural and size variation in Sapindaceae; 3) explore the highly divergent regions of the cp genomes from Sapindaceae

species; 4) reconstruct the phylogeny of *H. bodinieri* and *E. cavaleriei* within Sapindales at the cp genome level.

2. Materials And Methods

2.1 Plant Material and DNA Extraction

Fresh leaves of cultivated plants *Handeliodendron bodinieri* (voucher Number: HXZ-Z-0001) and *Eurycorymbus cavaleriei* (voucher Number: Hu-Z-0001) were collected from Guizhou Botanical Garden and Pingtang county of Guizhou, China, respectively. These sample collections were approved by the Guizhou Academy of Sciences and Guizhou Forestry Bureau, Guizhou province, China. The voucher specimens were deposited in the Herbarium of Nature Museum, Guizhou University (GACP). Total genomic DNA of *E. cavaleriei* was extracted following a modified cetyltrimethylammonium bromide (CTAB) approach [50], while that of *H. bodinieri* could not be obtained by the same method. Total genomic DNA of *H. bodinieri* was eventually extracted CTAB approach as improved by Tong [51] .

2.2 Chloroplast Genome Sequencing, Assembly and Annotation

The genome was sequenced using the Illumina Hiseq 2500 platform at Wuhan BGI Technology Service Co., Ltd. (Wuhan, China). Complete chloroplast genomes of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* were assembled from raw reads using GetOrganelle v1.7.1 with default parameters [52], which will automatically call the SPAdes [53], Bowtie2 [54] and BLAST [55]. The assembly workflow including five key steps: 1). Mapping reads to seed and assembling seed-mapped reads for parameter estimation; 2). Recruiting more target-associated reads through extending iterations; 3). Conducting de novo assembly; 4). Roughly filtering for target-like contigs; 5). Identifying target contigs and exporting all configurations [52]. The final assembly results were checked using Bandage [56].

Complete chloroplast genomes were annotated using the PGA (Plastid Genome Annotator) software [57], with *Litchi chinensis* (NC_035238) as main reference. Subsequently, Geneious 11.0.4 [58] was used to manually adjusted start and stop codons based on multiple cp genomes of Sapindaceae species. The structural features of cp genome maps were drawn online using OGDRAW [59]. The final cp genomes of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* were submitted to GenBank of National Center for Biotechnology Infromation (NCBI), where *H. bodinieri* and *E. cavaleriei* were given MK552107 and MK552106 accession numbers, respectively.

2.3 Genome features Analysis

Length of the whole chloroplast genome, numbers of genes, and categories of genes were analyzed in Geneious 11.0.4 [58]. MEGA [60] was adopted to calculate the guanine-cytosine (GC) content. A total of 25 species from Sapindaceae were compared in this analysis, including *Aesculus wangii*, *Ae. chinensis*, *Acer. buergerianum* subsp. *ningpoense*, *A. tataricum* subsp. *ginnala*; *A. davidii*, *A. truncatum*, *Acer miaotaiense*, *A. griseum*, *A. wilsonii*, *A. sino-oblongum*, *A. morrisonense*, *A. palmatum*, *Dipteronia dyeriana*, *D. sinensis*, *Dimocarpus longan*, *Dodonaea viscosa*, *Sapindus mukorossi*, *Xanthoceras sorbifolium*, *Litchi chinensis*,

Koelreuteria paniculata, *Pometia tomentosa*, and *Nephelium lappaceum*. CodonW software (http://codonw.sourceforge.net/culong.html#CodonW) was used for analyzing the codon preference. In the current study, all sequences downloaded from NCBI as well as their corresponding GenBank accession numbers are presented in the Table S1.

2.4 Dispersed repeats and Simple Sequence Repeats Detection

Dispersed repeats, including forward, reverse, palindromic, and complement repeat types, were identified online using REPuter program [61]. The minimal repeat size was limited to no less than 30 bp with the Hamming distance equal to three, and with the other settings retained as default. Furthermore, IR regions are the most typical palindromic repeat sequences in the cp genome, hence were not included in the analysis. The simple sequence repeats (SSRs) of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* chloroplast genomes were detected using MIcroSAtellite identification tool (MISA) [62]. The parameters were adjusted for the identification of mononucleotide, dinucleotide, tri nucleotide, tetra nucleotide, penta nucleotide and hexa nucleotidemotifs with a minimum of 10, 5, 4, 3, 3, and 3 repeats, respectively.

2.5 Whole cp Genome Sequence Comparisons

The mVISTA program [63] was employed to compare the whole cp genome divergence with related species in Shuffle-LAGAN mode, with the *Eurycorymbus cavaleriei* chloroplast genome as the reference. To explore the highly divergent regions of the cp genomes in the Sapindaceae species, the software DnaSP version 5.1 [64] was used to calculate nucleotide diversity (Pi). The step size and window length were set as 200 bp and 600 bp, respectively. Geneious 11.0.4 [58] was used to detect the contraction/expansion of the inverted repeat regions (IRs), and the final graph of expansions/contractions was visualized using Adobe Illustrator. The genome rearrangement analyses of *H. bodinieri* and *E. cavaleriei* cp genomes were performed using Mauve with default settings [65].

2.6 Phylogenetic analysis

To ascertain the phylogenetic position of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* within the Sapindales, a total of 42 species were analyzed, of which 40 complete cp genome sequences were downloaded from NCBI (Table S1). Among these species, *Euscaphis japonica* in the order Crossosomatales was used as outgroup. Phylogenies were constructed by Maximum Likelihood (ML) and Bayesian Inference (BI) analyses using the complete genome sequences, coding sequence (CDS), LSC, SSC, and IR regions. For the CDS dataset, we extracted 79 CDSs using Geneious 11.0.4 [58]. Each CDS matrix was aligned individually by MAFFT [66] with the codon model. All alignments were eventually concatenated into one supermatrix by PhyloSuite [67]. ModelFinder [68] was used to select the best-fit model for constructing phylogenetic tree based on the BIC standard. ML analysis was conducted using IQ-TREE with 1000 bootstrap replicates [69], whereas BI analysis were carried in MrBayes 3.2.2 [70]. For BI analysis, the two independent Markov Chain Monte Carlo (MCMC) analyses were run for 10,000,000 generations. Trees were sampled every 1,000 generations, and the first 25% of trees generated were discarded as burn-in. Finally, the phylogenetic trees were edited using Figtree 1.4 (https://github.com/rambaut/figtree).

3. Results

3.1 General Characteristics of the cp Genomes

Chloroplast genome structures of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* are conserved and their cp genome sizes were 151,271 and 158,690 bp, respectively (Fig. 1). Both genomes presented the quadripartite structures including a pair of inverted repeats (IRs) of 25,724 bp and 26,910 bp, a large single copy (LSC) region of 85,092 bp and 86,874 bp, and a small single copy (SSC) region of 15,812 bp and 17,966 bp in *H. bodinieri* and *E. cavaleriei* respectively (Table 1). In the whole genome, the total GC content of *H. bodinieri* and *E. cavaleriei* were 37.8% and 37.9%, respectively. Moreover, the GC content was unevenly distributed in the cp genome of *H. bodinieri* and *E. cavaleriei*. The IR region of *H. bodinieri* showed the highest GC contents (43.1%), followed by 36.0% in the LSC region, whereas the SSC region exhibited the lowest GC content of 31.5%. Similarly, the IR region of *E. cavaleriei* showed higher GC contents (42.8%) than that of the LSC region (36.1%) and SSC region (32.3%). In the coding sequences (CDS), the GC content of *H. bodinieri* and *E. cavaleriei* were 38.1% and 38.4%, respectively.

Table 1
Comparison of chloroplast genome feature of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei*

Species	Location	length (bp)	T (U) (%)	C (%)	A (%)	G (%)	G + C (%)
H. bodinieri	LSC	85,092	32.8	18.5	31.1	17.5	36.0
	SSC	18,731	34.2	16.3	34.4	15.1	31.5
	IR	25,724	28.3	20.9	28.6	22.3	43.1
	CDS	78,970	31.4	19.5	30.5	18.6	38.1
	Total	155,271	31.5	19.3	30.6	18.6	37.8
E. cavaleriei	LSC	86,874	32.7	18.6	31.3	17.5	36.1
	SSC	17,996	33.8	16.7	33.9	15.6	32.3
	IR	26,910	28.5	20.8	28.6	22.1	42.8
	CDS	79,251	31.3	19.7	30.3	18.7	38.4
	Total	158,690	31.4	19.3	30.7	18.6	37.9

A total of 114 unique genes were annotated in *H. bodinieri* and *E. cavaleriei*, including 77 protein-coding genes (PCGs) in *H. bodinieri*, whereas 79 PCGs were annotated in *E. cavaleriei*. 31 tRNA genes and four rRNA genes were annotated in the two species (Table 2). Among these, three genes (*infA*, *rpl22*, *rpl2*) and one gene (*infA*) had the stop codon appearing prematurely, thus, were annotated as pseudogenes in *H. bodinieri* and *E. cavaleriei*, respectively. In total, 18 genes were duplicated in the IR regions of *H. bodinieri* cp genome, including seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*), and seven PCGs (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *ycf2*, *ycf15*). For the cp genome of *E. cavaleriei*, seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), four rRNA

genes (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*), and eight PCGs (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps19*, *rps12*, *ycf2*, *ycf15*) were located in the IR regions.

Table 2
Summary of assembled gene functions of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* chloroplast genomes

Gene Family	Gene Names
Subunits of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
Subunits of NADH dehydrogenase	*ndhA*, ndhB*(×2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Subunits of cytochrome	*petA, petB*, petD*, petG, petL, petN*
Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
Subunit of rubisco	*rbcL*
Subunit of Acetyl-CoA-carboxylase	*accD*
c-type cytochrome synthesis gene	*ccsA*
Envelop membrane protein	*cemA*
Protease	*clpP***
Translational initiation	*ψinfA*
Maturase	*matK*
Large subunit of ribosome	*rpl2*(×2), rpl14, rpl16*, rpl20, rpl22^{H}, rpl23(×2), rp32, rpl33, rpl36*
DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*
Small subunit of ribosome	*rps2^{H}, rps3, rps4, rps7(×2), rps8, rps11, rps12**(×2), rps14, rps15, rps16*, rps18, rps19(×E)*
rRNA Genes	*rrn4.5(×2), rrn5(×2), rrn16(×2), rrn23(×2)*

* Genes containing a single intron; ** Genes containing two introns; (×2) Genes are located within the IR regions and therefore are duplicated; (×E) Genes present as two copies in the IR regions of *E. cavaleriei*; ψ indicates a pseudogene; ^{H} Pseudogene in *H. bodinieri* only.

Gene Family	Gene Names
tRNA Genes	*trnA-UGC(×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU(×2), trnI-GAU(×2), trnK-UUU, trnL-CAA(×2), trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU(×2), trnP-UGG, trnQ-UUG, trnR-ACG(×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(×2), trnV-UAC, trnW-CCA, trnY-GUA*
Unknown function	*ycf1, ycf2(×2), ycf3**, ycf4, ycf15(×2)*

* Genes containing a single intron; ** Genes containing two introns; (×2) Genes are located within the IR regions and therefore are duplicated; (×E) Genes present as two copies in the IR regions of *E. cavaleriei*; ψ indicates a pseudogene; ^{H} Pseudogene in *H. bodinieri* only.

Among the annotated genes of *H. bodinieri* and *E. cavaleriei* cp genomes, 18 genes contain introns, including *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl16*, *rpl2*, *rpoC1*, *rps16*, *trnV-UAC*, *trnL-UAA*, *trnK-UUU*, *trnI-GAU*, *trnG-UCC*, *trnA-UGC*, *clpP*, *rps12*, and *ycf3*. Gene *clpP*, *rps12*, and *ycf3* contain two introns, whereas the other 15 genes have only one intron (Table 3). *Rps12* is a trans-spliced gene with the 5' end exon located in the LSC region, but the 3' end in the IR region, as in most other angiosperms. In addition, the longest intron was detected in *trnK-UUU* of both cp genomes, and its length was 2,514 bp and 2,496 bp, respectively. Similar to other cp genomes, the *matK* gene is located in the intron of *trnK-UUU*.

Table 3
Genes with introns in the chloroplast genomes of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* as well as the lengths of the exons and introns.

Gene	Location	*H. bodinieri*					*E. cavaleriei*				
		Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
atpF	LSC	159	756	408			159	750	408		
clpP	LSC	69	836	291	645	228	69	853	291	658	228
ndhA	SSC	552	1143	540			558	1089	540		
ndhB	IR	777	682	756			777	680	756		
petB	LSC	6	789	657			6	794	657		
petD	LSC	9	740	525			9	677	525		
rpl16	LSC	9	1049	402			9	923	402		
rpl2	IR	393	661	435			390	664	435		
rpoC1	LSC	435	674	1620			435	711	1620		
*rps12**	LSC	114	-	232	537	26	114	-	232	533	26
rps16	LSC	39	829	225			39	836	225		
trnV-UAC	LSC	39	589	37			39	590	37		
trnL-UAA	LSC	37	542	49			37	542	49		
trnK-UUU	LSC	37	2514	38			37	2496	38		
trnI-GAU	IR	42	953	35			42	946	35		
trnG-UCC	LSC	23	724	48			23	721	48		
trnA-UGC	IR	38	841	35			38	810	35		
ycf3	LSC	126	731	228	746	153	126	732	228	767	153

We further compared these basic characteristics of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* cp genomes with other genera of Sapindaceae (Table 4). Significantly, found the cp genome size of the Hippocastanoideae was generally smaller across Sapindaceae, their size ranged from 152,688 bp (*Acer tataricum* subsp. *ginnala*) to 157,367 bp (*A. palmatum*). Overall, the full-length of cp genome ranged from 152,688 bp (*Acer tataricum* subsp. *ginnala*) to 163,258 bp (*Koelreuteria paniculata*), but the total GC content was similar among 25 cp genomes of Sapindaceae. By comparing single copy regions, found that *K.*

paniculata possessed the largest LSC region with a length of 90,236 bp, whereas *Sapindus mukorossi* possessed the largest SSC region (18,874 bp). Among these cp genomes, the length of IR regions varied from 25,656 bp (*Aesculus chinensis*) to 30,103 bp (*Litchi chinensis*). Interestingly, we also discovered *L. chinensis* presenting the smallest SSC region in these cp genomes. This analysis indicated that the number of rRNA is identical, whereas the number of tRNA (from 37 to 40) and PCGs (83–89) were remarkably similar.

Table 4
Statistics on the basic features of the chloroplast genomes from Sapindaceae species.

Species	size (bp)	LSC (bp)	IR (bp)	SSC (bp)	GC content (%)	No. rRNA	No. tRNA	No. PCGs
Acer tataricum subsp. *ginnala*	152,688	82,529	26,306	17,795	38.2	8	40	87
Handeliodendron bodinieri	155,271	85,092	25,724	18,731	37.8	8	37	89
Aesculus chinensis	155,528	85,489	25,656	18,727	37.9	8	37	83
Aesculus wangii	155,871	84,882	26,390	18,209	38.0	8	40	84
Acer truncatum	156,262	86,018	26,086	18,072	37.9	8	40	89
Acer miaotaiense	156,595	86,327	26,100	18,068	37.9	8	40	89
Acer griseum	156,857	85,227	26,742	18,146	37.9	8	40	86
Acer buergerianum subsp. *ningpoense*	156,911	85,314	26,752	18,093	37.9	8	40	89
Acer davidii	157,044	85,410	26,761	18,112	37.9	8	40	86
Acer wilsonii	157,067	85,418	26,760	18,129	37.9	8	39	88
Dipteronia dyeriana	157,071	85,529	26,730	18,082	38.0	8	40	87
Dipteronia sinensis	157,080	85,455	26,766	18,093	37.8	8	40	88
Acer sino-oblongum	157,118	85,558	26,722	18,119	37.9	8	39	88
Acer morrisonense	157,197	85,655	26,728	18,086	37.8	8	40	86
Acer palmatum	157,367	85,829	26,689	18,160	37.8	8	40	89
Eurycorymbus cavaleriei	158,690	86,874	26,910	17,996	37.9	8	37	87
Dodonaea viscosa	159,375	87,204	27,100	17,971	37.9	8	37	88
Sapindus mukorossi	160,481	85,649	27,979	18,874	37.7	8	39	88
Pometia tomentosa	160,818	85,666	28,396	18,360	37.9	8	37	88
Dimocarpus longan	160,833	85,708	28,428	18,269	37.8	8	37	87
Xanthoceras sorbifolium	161,231	85,299	28,620	18,692	37.7	8	38	86
Nephelium lappaceum	161,356	86,009	28,597	18,153	37.8	8	37	87
Litchi chinensis	162,524	85,750	30,103	16,568	37.8	8	37	87
Koelreuteria paniculata	163,258	90,236	27,377	18,268	37.3	8	37	85

3.1.1 Chloroplast Repeated Sequences and SSRs

In the current study, a total of 32 and 39 repeat sequences were detected in *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* cp genomes, respectively. In cp genome of *H. bodinieri*, there were 14 forward (F), 1 reverse (R), 16 palindromic (P), and 1 complement (C) repeats (Fig. 2, A). For *E. cavaleriei* cp genome, the number of the F, R, P, and C repeats was 16, 2, 20, and 1, respectively. We found that the length of repeat sequences ranged from 30 to 51 bp in *H. bodinieri*, 30–72 bp in *E. cavaleriei* (Fig. 2, B). In total, the results revealed that the P and F repeats were most abundant in all these repeat sequences, and most of palindromic and forward repeats were with 30–40 bp in length.

Here, we observed the simple sequence repeats of the *H. bodinieri* and *E. cavaleriei* cp genomes. The total number of the SSRs was 98 in *H. bodinieri*, whereas 60 in *E. cavaleriei* (Table S2; Fig. 2, C). In *H. bodinieri*, we detected five categories of SSRs, including mononucleotide, dinucleotide, tri nucleotide, tetra nucleotide, and penta nucleotide repeats. Additionally, none of hexa-nucleotides were detected in *H. bodinieri* cp genome. The number of mononucleotide, dinucleotide, tri nucleotide, tetra nucleotide, and penta nucleotide repeats were 75, 6, 7, 9, and 1, respectively (Fig. 2, C). The finding not only showed that the mononucleotide repeats were the most abundant in the cp genome, but also had an outstanding base preference, mainly consist of A or T. Notably, five SSRs were identified in the *ycf1* gene of *H. bodinieri* cp genome, consisting of mononucleotide repeats that contain four poly (T) and one poly (A). In total, there were four types of SSRs in the *E. cavaleriei* cp genome, including mononucleotide, dinucleotide, tri nucleotide, and tetra nucleotide repeats (Table S3). The number of mononucleotide, dinucleotide, tri nucleotide, tetra nucleotide, and penta nucleotide repeats were 38, 5, 8, and 8, respectively. Among these SSRs, the most dominant of SSRs were A or T mononucleotides. We also observed four SSRs in the gene *ycf1*, consist of mononucleotide repeats. Within *H. bodinieri* and *E. cavaleriei* cp genomes, most SSRs were located in the intergenic spacer regions (IGS) (Fig. 2, D and E).

3.1.2. Relative Synonymous Codon Usage Analysis

The total number of the codons was 26,028 in *Handeliodendron bodinieri*, 26,445 in *Eurycorymbus cavaleriei*. Among the codons, the number of the amino acids less than 1000 was tyrosine (Tyr), glutamine (Gln), histidine (His), methionine (Met), tryptophan (Trp), cysteine (Cys) in both the *H. bodinieri* (Table S4) *and E. cavaleriei* (Table S5). The leucine (Leu) was the most amino acid encoded in the analysis, accounting for 10.5% and 10.6% on average of all amino acids in the *H. bodinieri and E. cavaleriei* cp genomes, respectively. However, the Cys has the lowest number of codons in both the *H. bodinieri and E. cavaleriei* cp genomes excluding the stop codons. The codon usage frequency and relative synonymous codon usage (RSCU) were summarized in Fig. 3. In *H. bodinieri* cp genome, 30 codons had RSCU values more than 1.00, and they all ended with A or U excluding UUG. For the *E. cavaleriei*, there were 31 codons with RSCU values more than 1.00, 29 of which ended with A or U codons, whereas two ended with C and G codons (UCC and UUG). Moreover, discovered that the RSCU values of three codons (AUG, UGG, and UCC) is 1.00 in the *H. bodinieri* cp genome, while only two codons (AUG and UGG) in *E. cavaleriei* cp genome.

3.2 Comparative Chloroplast Genomic Analysis

To identify the sequence divergence of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* cp genomes, the genomic rearrangement was detected, with *Litchi chinensis* as the reference (Fig. 4). In total, 16 cp

genomes of 13 genera from Sapindaceae were used for analysis. Comparative analysis showed that all cp genomes exhibited highly conserved, an indication that inversion and translocation in genes or plastid segments was not detected in the final results.

In addition, we performed multiple sequence alignment of the complete chloroplast genome sequences from different families in Sapindales with *E. cavaleriei* as the reference (Fig. 5). The comparison analyses revealed that coding regions were more conserved than the non-coding regions, and the SSC and LSC regions exhibited more variation than IR regions in all cp genomes. Moreover, there were almost no variation in four rRNA genes within LSC regions, and the genes were highly conserved. A total of five genes, including *matK*, *accD*, *ycf1*, *ndhF*, and *rpl22*, were detected the most divergent in these cp genomes. As shown in Fig. 5, the significant variations were detected in the intergenic regions of the LSC and SSC, including *trnH-psbA*, *trnk-rps16*, *rps16-trnQ*, *psbM-trnY*, *psbZ-trnG*, *trnL-trnF*, *trnF-ndhJ*, and *rpl32-trnL*.

3.2.1 Divergence Hotspot Identification

A total of 13 cp genome sequences were aligned among the Sapindaceae species to calculate nucleotide diversity (Pi), including *Acer davidii*, *A. miaotaiense*, *Aesculus chinensis*, *Ae. wangii*, *Dimocarpus longan*, *Dipteronia dyeriana*, *D. sinensis*, *Dodonaea viscosa*, *Litchi chinensis*, *Nephelium lappaceum*, *Pometia tomentosa*, *Sapindus mukorossi*, *Xanthoceras sorbifolium*, and *Handeliodendron bodinieri*, *Eurycorymbus cavaleriei*. Based on this analysis, we identified three remarkably divergent regions among these complete cp genomes, which included *ndhC-trnV-UAC*, *rpl32-trnL-UAG-ccsA*, and *ycf1* (Fig. 6). Gene *ycf1* is the most divergent region with the highest Pi value (0.127) and is located in the SSC region, whereas *ndhC-trnV-UAC* and *rpl32-ccsA* intergenic spacers were located in LSC regions. These highly divergent regions could be used as potential molecular markers for phylogenetic reconstruction of the family Sapindaceae. Overall, the result of this study revealed that sequence divergence was concentrated in the LSC and SSC regions, whereas IR regions presented less divergence, consistent with the mVISTA results (Fig. 5).

3.2.2. Expansion and Contraction of IRs

Comparison of the single-copy (SC) and inverted repeat (IR) boundary region among six families within the order Sapindales was presnted Fig. 7. A total of nine cp genomes were included, *Koelreuteria paniculata* (Sapindaceae), *Litchi chinensis* (Sapindaceae), *Leitneria floridana* (Simaroubaceae), *Toona ciliata* (Meliaceae), *Anacardium occidentale* (Anacardiaceae), *Boswellia sacra* (Burseraceae), *Citrus limon* (Rutaceae), as well as our newly sequenced species. The SSC/IRa boundary located in the coding region of gene *ycf1* in all cp genomes, with 2,516 bp to 4,602 bp in SSC region. The gene *ycf1* has the largest fragment in SSC region of *H. bodinieri*. In five cp genomes, including *H. bodinieri*, *L. chinensis*, *T. ciliata*, *A. occidentale*, and *C. limon*, the gene *ndhF* spanned IRb/SSC boundary, with 7–36 bp in the IRb region. However, the gene *ndhF* was wholly located in the IRb region of four cp genomes (*E. cavaleriei*, *K. paniculata*, *L. floridana*, and *B. sacra*), which was separated from the IRb/SSC border by a spacer varying from 0 to 84 bp. The gene *ycf1* in the border region between IRb and SSC is treated as a pseudogene because of the incomplete duplication of the normal copy. Similarly, *rpl22* and *rps19* genes in IRa region near the IRa/LSC boundary region, and was annotated as a pseudogene, including *E. cavaleriei*, *K. paniculata*, *T. ciliata*, *B. sacra*, *L. floridana*, and *A. occidentale*. In all cp genomes, there was significant variation in the LSC/IRb boundary regions. The LSC/IRb

boundary was crossed by the gene *rpl22* in five cp genomes, and the length of the *rpl22* fragment located in the LSC region ranged from 185 bp (*L. floridana*) to 449 bp (*E. cavaleriei*). The *rpl22* was entirely located in the IRb region of *C. limon* cp genome. In *H. bodinieri* cp genome, *rps19* and *rpl2* genes were entirely located within the LSC and IRb region near the LSC/IRb boundary, respectively. In *A. occidentale* cp genome, we also found that *rps19* gene was located in the LSC/IRb boundary, and 178 bp extended into the LSC region. In *L. chinensis*, the *rps16* and *rps3* genes in the near LSC/IRb were located in LSC and IRb regions, respectively. In the IRa/LSC boundary regions, gene *trnH* was completely located in LSC region of all cp genomes, which was 0–38 bp away from the IRb/SSC boundary. In a word, IR regions of the *H. bodinieri* showed a significant contraction, whereas it presented a notable expansion in *E. cavaleriei* cp genome.

3.6. Phylogenetic Analysis of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* within Sapindales

In this study, phylogenetic analyses were performed based on different datasets, including complete chloroplast genome sequences, coding sequence (CDS), LSC, SSC, and IR regions, with corresponding results in Fig. 8, Fig. 9, Figure S1, Figure S2, and Figure S3, respectively. Detailed information of best-fit model for ML and BI tree are listed in Table S6. All phylogenetic trees consistently revealed that *Handeliodendron bodinieri* is sister to the clade consisting of *Aesculus chinensis* and *Aesculus wangii* (BS = 100, PP = 1.00), strongly support *Eurycorymbus cavaleriei* being sister with *Dodonaea viscosa* (BS = 100, PP = 1.00).

4. Discussion

Features of Complete Chloroplast Genome and Comparative Analyses

GetOrganelle is a state-of-the-art toolkit to accurately assemble organelle genomes from whole-genome sequencing data [52]. In the current study, GetOrganelle was utilized to assemble the complete chloroplast genome sequences of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* based on the newly sequenced Illumina data. The total size of *H. bodinieri* cp genome is 158,690 bp, which is different from the previous reports of Chen et al. [47] and Du et al. [48]. The full-length of *H. bodinieri* cp genome is 151,271 bp and is also different from that of the previous report [49]. Their genome exhibited a typical quadripartite structure, including a pair of inverted repeats (IRs), a large single copy (LSC) region, and a small single copy (SSC) region, which was the same as that reported for most other angiosperms [71–73]. This study revealed that gene content and gene order in *H. bodinieri* and *E. cavaleriei* cp genomes were quite similar to that of other published Sapindaceae species [74–76]. Additionally, GC content was unevenly exhibited in the cp genome of *H. bodinieri* and *E. cavaleriei* and the IR region showed higher GC contents than that of the LSC region and SSC region, which may be attributed to four rRNA genes with low A/T content. Within the *H. bodinieri* and *E. cavaleriei* cp genomes, the *clpP*, *rps12*, and *ycf3* genes possess one intron, and 15 genes contain two introns. Pseudogenization of the gene *infA* was has been detected in *H. bodinieri* and *E. cavaleriei* cp genomes, the same results were observed in those of other Sapindaceae species [77–79]. Besides, we detected the pseudogenization of *rpl22* and *rpl2* genes in *H. bodinieri* cp genome and the latter was also annotated as pseudogene in that of *Acer takesimense* [80]. A detailed comparative analysis of the complete chloroplast genomes revealed that genomic structure, gene content, PCGs, and total GC content were remarkably similar

or identical within 11 genera from Sapindaceae. Significantly, the cp genome size of the Hippocastanoideae was generally smaller across Sapindaceae, which presented they have a close relationship to some extent.

The study demonstrated that palindromic (P) and forward (F) repeats were the most abundant in all dispersed repeats, and most of them were 30–40 bp in length, similar to previous studies [33, 81–83]. Simple sequence repeats (SSRs) were widely used as a molecular marker for studies of genetic diversity and population structure [84–86]. We observed five and four types of SSRs in *H. bodinieri* and *E. cavaleriei* cp genomes, respectively. Among these SSRs, the most dominant of SSRs were A or T mononucleotides. Furthermore, most of SSRs were in the intergenic spacer regions (IGS), which is consistent with other studies [78, 87]. The finding not only showed that the mononucleotide repeats were the most abundant in the cp genome but also had an A/T base preference.

Within *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* coding sequences, the leucine (Leu) was the frequent amino acid, while the least abundant amino was cysteine (Cys) excluding the stop codons. Moreover, the findings of this study revealed that most codons ended with A or U when RSCU values more than 1.00. As a common phenomenon in cp genomes of plants, similar results have been reported in previous studies [72, 88–90].

The mVISTA results revealed that coding regions were more conserved than the non-coding regions, and the SSC and LSC regions exhibited more variation than IR regions in all cp genomes. The results of this study are consistent with previous finings in other species [29, 73, 74, 91]. In total, we identified five genes present significant variations in these cp genomes, such as *matK*, *accD*, *ycf1*, *ndhF*, and *rpl22*. Additionally, eight intergenic regions, namely *trnH-psbA*, *trnk-rps16*, *rps16-trnQ*, *psbM-trnY*, *psbZ-trnG*, *trnL- trnF*, *trnF-ndhJ*, and *rpl32-trnL* also present significant variations. The nucleotide diversity tests indicate IR regions showed relatively low diversity, but we found *ndhC-trnV-UAC*, *rpl32-trnL-UAG-ccsA*, and *ycf1* were remarkably divergent regions in LSC and SSC regions, which could be used as the specific DNA barcodes for Sapindaceae species. Gene *ycf1* was a high variable genic region in cp genome of plants [23, 71, 88], which has abundant SSRs in our newly sequenced species.

Up to date, cp genome sequences of 13 genera from Sapindaceae are deposited in GenBank database (Table 4). Comparative analysis showed that *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* cp genomes exhibited highly conserved, and the phenomenon of inversion and translocation in genes or plastid segments was not detected in both species. Nevertheless, we found that there is still some significant variation in LSC and IR boundary regions. In the family Sapindaceae, the LSC/IRb boundary was traversed by the gene *rpl22* in four cp genomes, but it is different from those of *H. bodinieri* and *Litchi chinensis*. Interestingly, their boundaries were similar to that of the member of Simaroubaceae (*Leitneria floridana*). We noticed that IRa/LSC boundary regions showed some differences, but gene *trnH* was completely located in LSC region of all cp genomes. In different families, the SSC/IRa and IRb/SSC boundary regions exhibit highly conserved, which resembles most previous studies [92–94]. The pseudogenes, *ycf1* and *rpl22* were present at the IRb/SSC and IRa/LSC boundaries of *E. cavaleriei* cp genome, respectively, because the incomplete duplication of the normal copy. The finding revealed the IR boundary regions of *E. cavaleriei* cp genome are similar to the *K. paniculata*, *T. ciliata*, *B. sacra*, and *L. floridana*, while the IR boundary regions of *H. bodinieri* cp genome resemble that of other reported *Acer* species (*Acer miaotaiense*, *A. sterculiaceum*, *A. amplum*) [46,

78]. IR regions of the *H. bodinieri* and *E. cavaleriei* cp genomes showed noticeable contraction and expansion, respectively.

Phylogenetic Analysis

Phylogenetic analysis based on different datasets consistently demonstrated that *Handeliodendron bodinieri* was sister to the clade that consisted of *Aesculus chinensis* and *A. wangii* (PP = 1.00, BS = 100). On the basis of plastid *matK* and *rbcL* DNA sequences, the work of Harrington et al. [15] strongly supported *H. bodinieri* was sister to *Aesculus* plus *Billia*. Subsequently, based on a combination of nuclear (ITS) and plastid (*matK*, *rpoB*, *trnD-trnT*, *trnK-matK*, *trnL-trnF*, and *trnS-trnG*) markers, Buerki et al. [18] suggested the possible paraphyly of the *Aesculus* because *Handeliodendron* and *Billia* were nested within it. Our results strongly supported that *H. bodinieri* had a close relationship with the *Aesculus*, which was consistent with the works of Harrington et al. [15] and Buerki et al. [18]. However, the present analyses could not demonstrate the paraphyly of the *Aesculus* because of lacking sufficient samples. In this study, we found *Eurycorymbus cavaleriei* as sister to *Dodonaea viscosa* with strong support (PP = 1.00, BS = 100), which similar to previous studies [45, 47, 48, 74, 95]. Buerki et al. [17, 18] conducted molecular phylogenetic studies of Sapindaceae s. lat. based on comprehensive sampling. Their results demonstrated that *Eurycorymbus* belongs to a member of the *Dodoneae* group within Dodonaeoideae, strongly supported *Euphorianthus longifolius* as sister to *E. cavaleriei*, but the *E. longifolius* was not sampled here. Furthermore, for the Sapindaceae, it remains ambiguous in terms of the phylogenetic relationship to other members within Sapindales. In whole cp genome analysis, Sapindaceae formed a sister of the Rutaceae + Simaroubaceae + Meliaceae + Burseraceae + Anacardiaceae clade (BS = 99, PP = 1, Fig. 9), but the clade consist of Rutaceae + Simaroubaceae + Meliaceae + Burseraceae + Anacardiaceae was weakly supported (BS = 51, PP = 0.67). In the CDS analysis, Sapindaceae was sister to Anacardiaceae and Burseraceae with weak support (BS = 51, PP = 1, Fig. 9), the result is consistent with previous studies [93, 96]. In the LSC analysis, Sapindaceae was sister to the Burseraceae + Anacardiaceae + Nitrariaceae clade (BS = 49, PP = 0.99, Figure S1), In the SSC analysis, Sapindaceae was sister to the Rutaceae + Simaroubaceae + Meliaceae clade (BS = 52, PP = 1, Figure S2), which is similar to the previous work [95, 97]. Overall, as the sister group of Sapindaceae, most results exhibited weak support in maximum likelihood analysis. This work will contribute to a comprehensive understanding of plastome evolution in Sapindaceae species and provide valuable chloroplast genomic information for further elucidating the circumscription of Sapindaceae at the cp genome level.

5. Conclusion

In this work, we sequenced and assembled the complete chloroplast genome of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei*. Their gene order, gene content, and molecular structure are similar to that of cp genomes of other Sapindaceae species. Comparative analysis of complete cp genomes revealed that the cp genome size of the Hippocastanoideae was generally smaller across Sapindaceae. We detected three highly divergent regions, which could be used as the specific DNA barcodes within Sapindaceae. Phylogenetic results consistently confirm that *H. bodinieri* has a close relationship with the genus *Aesculus*, strongly support *E. cavaleriei* as sister to *Dodonaea viscosa*. As the national-level protected species, both *H. bodinieri* and *E. cavaleriei* attract scientific attention in many aspects, thus this work will provide valuable chloroplast genomic information, and contribute to facilitating future studies in population genetics and conservation biology.

6. Abbreviations

LSC: large single copy; SSC: small single copy;IR: inverted repeat;BI: Bayesian inference;ML: Maximum Likelihood; PCGs: protein-coding genes;rRNA: Ribosomal RNA; SSR: simple sequence repeats; tRNA: Transfer RNA; chloroplast: cp; CTAB: cetyltrimethylammonium bromide; CDS: coding sequence; RSCU: Relative synonymous codon usage

7. Declarations

Acknowledgements

Not applicable

Authors' contributions

HGX collected these materials and identified species. HGX designed the experiments and organized the manuscript. YJX and HGW performed the analyses. YJX wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study got financial support from grants from National Science & Technology Fundamental Resources Investigation Program of China (Grant No. 2019FY101800), National Natural Science Foundation of China (31970211), and Sino-Africa Joint Research Center, CAS (SAJC202101)

Availability of data and materials

All data generated or analyzed during this study are included in this published article and the complete chloroplast genome sequences of *Handeliodendron bodinieri* and *Eurycorymbus cavaleriei* are deposited in the genbank with ID no: MK552107 and MK552106, respectively. The accession numbers corresponding to the additional datasets used and analyzed in this study can be found in Table S1. These were retrieved from National Center for Biotechnology Information database.

Ethical approval and consent to participate

The authors have complied with the relevant institutional, national and international guidelines in collecting biological materials for the study. Collection permits for materials for sequencing were granted by the Guizhou Academy of Sciences and Guizhou Forestry Bureau, Guizhou province, China. The studycontributes tofacilitating future studies in population geneticsand conservation biology.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

8. References

1. Luo, X.R., Chen, D.Z., Flora Reipublicae Popularis Sinicae, Sapindaceae. Science Press. Vol. 47. 1985, Beijing. 4-72.
2. Cao, L.M., Teng, T., Wu, Y.X., Cao, M. Physicochemical property and fatty acid composition of *Eurycorymbus cavaleriei* (Sapindaceae) seed oil. China Oils and Fats, 2014; 39(8): p. 95-97.
3. Cao, L.M., Wang, Y.H., Zhao, C., Zou, B.B. Analysis on nutrient components of seed kernel of *Eurycorymbus cavaleriei.* Journal of Plant Resources and Environment, 2015; 24(4): p. 114-115.
4. Chen, X.X., Xuan, X.X., Qiu, L.F., Wei, Y.Y., Qin, J.L. *Handeliodendron* Rehder, A Newly Recorded Genus of the Family Sapindaceae in Yunnan Province. Journal of Southwest Forestry University (Natural Sciences), 2017; 30(3): p. 85-87.
5. Xia, N.H., Gadek, P.A., *Eurycorymbus*, in *Flora of China*. 2007, Science Press & Missouri Botanical Garden Press: Beijing & St. Louis. p. 8.
6. Radlkofer, L. Sapindaceae. In: Engler, A. (ed.) Das Pflanzenreich IV, 165 (Heft 98a-h),. Leipzig: Werlag von Wilhelm Engelmann. 1933.
7. Radlkofer, L. Uber die Gliederung der Familie der Sapindaceen. Sitz. -Ber. Akad. Wiss. Munchen. 1890; 20: p. 105-379.
8. Muller, J., Leenhouts, P.W., A general survey of pollen types in Sapindaceae in relation to taxonomy. In: Ferguson, I.K, Müller, J. (Eds.), The evolutionary significance of the exine. Linnean Soc. Symp. Ser. 1, pp. 407-445. London: Academic Press. 1976.
9. Dahlgren, G. An updated angiosperm classification. Bot. J. Linn. Soc., 1989; 100: p. 197-203.
10. Cronquist, A., The Evolution and Classification of Flowering Plants. 1988, New York Botanic Gardens: New York.
11. Takhtajan, A., Systema Magonoliophytorum. 1987, Leningrad: Soviet Sciences Press.
12. Umadevi, I., Daniel, M. Chemosystematics of the Sapindaceae. Feddes Repertorium, 1991; 102(7-8): p. 607-612.DOI: 10.1002/fedr.19911020711.
13. Judd, W.S. Angiosperm family pairs: preliminary phylogenetic analyses. Harvard Pap. Bot., 1994; 5: p. 1-51.
14. Thorne, R.F. The classification and geography of the flowering plants: Dicotyledons of the class angiospermae (Subclasses magnoliidae, ranunculidae, caryophyllidae, dilleniidae, rosidae, asteridae, and lamiidae). The Botanical Review, 2000; 66(4): p. 441-647.DOI: 10.2307/4354381.
15. Harrington, M.G., Edwards, K.J., Johnson, S.A., Chase, M.W., Gadek, P.A. Phylogenetic inference in Sapindaceae sensu lato using plastid matK and rbcL DNA sequences. Syst. Bot., 2005; 30(2): p. 366-382.DOI: 10.1600/0363644054223549.
16. Thorne, R.F., Reveal, J.L. An Updated Classification of the Class Magnoliopsida (“Angiospermae”). The Botanical Review, 2007; 73(2): p. 67-182.DOI: 10.1663/0006-8101(2007)73[67:AUCOTC]2.0.CO;2.
17. Buerki, S., Forest, F., Acevedo-Rodriguez, P., Callmander, M.W., Nylander, J.A., Harrington, M., et al. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). Mol. Phylogenet Evol., 2009; 51(2): p. 238-258.DOI: 10.1016/j.ympev.2009.01.012.

18. Buerki, S., Lowry, P.P., Alvarez, N., Razafimandimbison, S.G., Küpfer, P., Callmander, M.W. Phylogeny and circumscription of Sapindaceae revisited: molecular sequence data, morphology and biogeography support recognition of a new family, Xanthoceraceae. Plant Ecol. Evol., 2010; 143(2): p. 148-159.DOI: 10.5091/plecevo.2010.437.
19. APGIV. 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): p. 1-20.
20. Rehder, A. *Handeliodendron*, a new genus of sapindaceae. J. Arnold Arboretum, 1935; 16(1): p. 65-67.
21. Xia, N.H., Turland, N.J., Gadek, P.A., *Handeliodendron*, in *Flora of China*. 2007, Science Press & Missouri Botanical Garden Press: Beijing & St. Louis. p. 1-2.
22. Zhang, R., Wang, Y.H., Jin, J.J., Stull, G.W., Bruneau, A., Cardoso, D., et al. Exploration of Plastid Phylogenomic Conflict Yields New Insights into the Deep Relationships of Leguminosae. Syst. Biol., 2020; 69(4): p. 613-622.DOI: 10.1093/sysbio/syaa013.
23. Kim, G.B., Lim, C.E., Kim, J.S., Kim, K., Lee, J.H., Yu, H.J., et al. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: insights into evolutionary divergence and phylogenomic implications. BMC Genomics, 2020; 21(1): p. 415.DOI: 10.1186/s12864-020-06812-7.
24. Xie, D.F., Tan, J.B., Yu, Y., Gui, L.J., Su, D.M., Zhou, S.D., et al. Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. Ann. Bot., 2020; 125(7): p. 1039-1055.DOI: 10.1093/aob/mcaa024.
25. Yan, M., Fritsch, P.W., Moore, M.J., Feng, T., Meng, A., Yang, J., et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. Mol. Phylogenet. Evol., 2018; 121: p. 198-211.DOI: 10.1016/j.ympev.2018.01.004.
26. Zhao, F., Chen, Y.P., Salmaki, Y., Drew, B.T., Wilson, T.C., Scheen, A.C., et al. An updated tribal classification of Lamiaceae based on plastome phylogenomics. BMC Biol, 2021; 19(1): p. 2.DOI: 10.1186/s12915-020-00931-z.
27. Wang, Y.H., Qu, X.J., Chen, S.Y., Li, D.Z., Yi, T.S. Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. Tree Genetics & Genomes, 2017; 13(2).DOI: 10.1007/s11295-017-1124-1.
28. Liu, B.B., Liu, G.N., Hong, D.Y., Wen, J. *Eriobotrya* Belongs to *Rhaphiolepis* (Maleae, Rosaceae): Evidence From Chloroplast Genome and Nuclear Ribosomal DNA Data. Front Plant Sci., 2019; 10: p. 1731.DOI: 10.3389/fpls.2019.01731.
29. Wang, J.H., Moore, M.J., Wang, H., Zhu, Z.X., Wang, H.F. Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. Gene, 2021; 765: p. 145103.DOI: 10.1016/j.gene.2020.145103.
30. Cai, C.N., Ma, H., Ci, X.Q., Conran, J.G., Li, J. Comparative phylogenetic analyses of Chinese *Horsfieldia* (Myristicaceae) using complete chloroplast genome sequences. J. Syst. Evol., 2020.DOI: 10.1111/jse.12556.
31. Teshome, G.E., Mekbib, Y., Hu, G.W., Li, Z.Z., Chen, J.M. Comparative analyses of 32 complete plastomes of Tef (*Eragrostis* tef) accessions from Ethiopia: phylogenetic relationships and mutational hotspots. PeerJ, 2020; 8: p. e9314.DOI: 10.7717/peerj.9314.

32. Zhang, W., Sun, Y.Z., Liu, J., Xu, C., Zou, X.H., Chen, X., et al. DNA barcoding of *Oryza*: conventional, specific, and super barcodes. Plant Mol Biol., 2021; 105(3): p. 215-228.DOI: 10.1007/s11103-020-01054-3.
33. Chen, X.L., Zhou, J.G., Cui, Y.X., Wang, Y., Duan, B.Z., Yao, H. Identification of *Ligularia* Herbs Using the Complete Chloroplast Genome as a Super-Barcode. Front Pharmacol., 2018; 9: p. 695.DOI: 10.3389/fphar.2018.00695.
34. Fu, Q.D., Yu, X.D., Xia, X.H., Zheng, Y.Q., Zhang, C.H. Complete chloroplast genome sequence of *Acer nikoense* (Sapindaceae). Mitochondrial DNA B, 2020; 5(3): p. 3118-3119.DOI: 10.1080/23802359.2020.1797574.
35. Saina, J.K., Gichira, A.W., Li, Z.Z., Hu, G.W., Wang, Q.F., Liao, K. The complete chloroplast genome sequence of *Dodonaea viscosa*: comparative and phylogenetic analyses. Genetica, 2018; 146(1): p. 101-113.DOI: 10.1007/s10709-017-0003-x.
36. Zhang, Z.Y., Chen, Y., Jiang, X.B., Zhu, P., Li, L., Zeng, Y.L., et al. The complete chloroplast genome of *Aesculus chinensis*. Mitochondrial DNA B, 2019; 4(1): p. 1955-1956.DOI: 10.1080/23802359.2019.1617056.
37. Liu, Z.G., Zhang, J.J., Zhou, Y.X., Liu, Y.F., Hu, Z.G., Zheng, G.H., et al. The complete chloroplast genome of *Aesculus chinensis* var. *wilsonii*. Mitochondrial DNA B, 2020; 5(3): p. 2547-2549.DOI: 10.1080/23802359.2020.1780972.
38. Zheng, W., Wang, W., Harris, A.J., Xu, X.D. The complete chloroplast genome of vulnerable *Aesculus wangii* (Sapindaceae), a narrowly endemic tree in Yunnan, China. Conserv. Genet. Resour., 2017; 10(3): p. 335-338.DOI: 10.1007/s12686-017-0818-x.
39. Zhang, D.X., Sun, K., Xiang, Q.H., Wang, X.R., Xu, J.H., Wang, Q., et al. The complete chloroplast genome sequence of *Aesculus Chinensis* Bunge, a major street tree. Mitochondrial DNA B, 2019; 4(1): p. 1686-1687.DOI: 10.1080/23802359.2019.1605857.
40. Wang, K.Y., Li, L., Zhao, M.Z., Li, S.K., Sun, H.H., Lv, Y.X., et al. Characterization of the complete chloroplast genome of longan (*Dimocarpus longan* Lour.) using illumina paired-end sequencing. Mitochondrial DNA B, 2017; 2(2): p. 904-906.DOI: 10.1080/23802359.2017.1413310.
41. Dong, P.B., Liu, Y., Gao, Q.Y., Yang, T., Chen, X.Y., Yang, J.Y., et al. Characterization of the complete plastid genome of *Acer tsinglingense*, an endemic tree species in China. Mitochondrial DNA B, 2019; 4(2): p. 4065-4066.DOI: 10.1080/23802359.2019.1689863.
42. Dai, W.T., Li, S.Q., Gao, X.F., Xu, B. The complete chloroplast genome of *Acer pentaphyllum* (Sapindaceae), a critically endangered maple endemic to China. Mitochondrial DNA B, 2020; 5(1): p. 470-471.DOI: 10.1080/23802359.2019.1704647.
43. Shi, Z.W., Sun, B., Pei, N.C., Shi, X. The complete chloroplast genome of *Acer tutcheri* Duthie (Acereae, Sapindaceae): an ornamental tree endemic to China. Mitochondrial DNA B, 2020; 5(3): p. 2686-2687.DOI: 10.1080/23802359.2020.1787899.
44. Ling, L.Z., Zhang, S.D. The complete chloroplast genome of an endangered and endemic species, *Acer yangbiense* (Aceraceae). Mitochondrial DNA B, 2019; 5(1): p. 224-225.DOI: 10.1080/23802359.2019.1699878.

45. Yang, H.X., Zha, X., Cao, S.L., Wang, Y., Gao, F., Zhou, Y.J. Complete chloroplast genome sequence of *Acer ginnala*, an important ornamental tree. Mitochondrial DNA B, 2020; 5(1): p. 609-610.DOI: 10.1080/23802359.2019.1710606.
46. Wang, W.C., Chen, S.Y., Zhang, X.Z. Complete plastomes of 17 species of maples (Sapindaceae: *Acer*): comparative analyses and phylogenomic implications. Plant Syst. Evol., 2020; 306(3).DOI: 10.1007/s00606-020-01690-8.
47. Chen, Z., Qiao, O.M., Liu, B.B., Sun, H.L. Complete chloroplast genome of *Eurycorymbus Cavaleriei* (Sapindaceae), a tertiary relic rare tree. Mitochondrial DNA B, 2019; 4(2): p. 3250-3251.DOI: 10.1080/23802359.2019.1670113.
48. Du, X.M., Xin, G.L., Ren, X.L., Liu, H.D., Hao, N., Jia, G.L., et al. The complete chloroplast genome of *Eurycorymbus cavaleriei* (Sapindaceae), a Tertiary relic species endemic to China. Conserv. Genet. Resour., 2018; 11(3): p. 283-285.DOI: 10.1007/s12686-018-1009-0.
49. Tian, X.M., Li, X.L., Miao, H.Y., Xue, C.L., Wang, B.L., Guo, Y., et al. Complete chloroplast genome of an endangered endemic tree, *handeliodendron bodinieri* (levl.) rehder (sapindaceae) from karst forests of southwest China. Mitochondrial DNA B, 2019; 4(2): p. 3272-3273.DOI: 10.1080/23802359.2019.1671251.
50. Doyle, J.J., Doyle, J.L. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem. Bull., 1987; 19: p. 11-15.
51. Tong, Z.G., Wang, F.R., Zhang, Z., Zhao, J.B., Zhang, K.C., Yan, G.H., et al. A method for DNA extraction from mature leaves of fruit trees. Journal of Fruit Science, 2008; 25(1): p. 122-125.
52. Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol., 2020; 21(1): p. 241.DOI: 10.1186/s13059-020-02154-5.
53. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol., 2012; 19(5): p. 455-477.DOI: 10.1089/cmb.2012.0021.
54. Langmead, B., Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods, 2012; 9(4): p. 357-359.DOI: 10.1038/nmeth.1923.
55. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. BLAST+: architecture and applications. BMC Bioinformatics, 2009; 10(1): p. 421.DOI: 10.1186/1471-2105-10-421.
56. Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics, 2015; 31(20): p. 3350-3352.DOI: 10.1093/bioinformatics/btv383.
57. Qu, X.J., Moore, M.J., Li, D.Z., Yi, T.S. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods, 2019; 15: p. 50.DOI: 10.1186/s13007-019-0435-7.
58. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 2012; 28(12): p. 1647-1649.DOI: 10.1093/bioinformatics/bts199.
59. Lohse, M., Drechsel, O., Bock, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet, 2007; 52(5-6): p. 267-274.DOI: 10.1007/s00294-007-0161-y.

60. Kumar, S., Stecher, G., Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol., 2016; 33(1870-1874).DOI: 10.1093/molbev/msw054.
61. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res., 2001; 29(22): p. 4633-4642.DOI: 10.1093/nar/29.22.4633.
62. Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M. MISA-web: a web server for microsatellite prediction. Bioinformatics, 2017; 33(16): p. 2583-2585.DOI: 10.1093/bioinformatics/btx198.
63. Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., et al. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics, 2000; 16(11): p. 1046-1047.DOI: 10.1093/bioinformatics/16.11.1046.
64. Librado, P., Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009; 25(11): p. 1451-1452.DOI: 10.1093/bioinformatics/btp187.
65. Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res., 2004; 14(7): p. 1394-1403.DOI: 10.1101/gr.2289704.
66. Katoh, K., Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol., 2013; 30(4): p. 772-780.DOI: 10.1093/molbev/mst010.
67. Zhang, D., Gao, F.L., Jakovlic, I., Zou, H., Zhang, J., Li, W.X., et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol. Ecol. Resour., 2020; 20(1): p. 348-355.DOI: 10.1111/1755-0998.13096.
68. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods, 2017; 14(6): p. 587-589.DOI: 10.1038/nmeth.4285.
69. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol., 2015; 32(1): p. 268-74.DOI: 10.1093/molbev/msu300.
70. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol., 2012; 61(3): p. 539-542.DOI: 10.1093/sysbio/sys029.
71. Yik, M.H., Kong, B.L., Siu, T.Y., Lau, D.T., Cao, H., Shaw, P.C. Differentiation of Hedyotis diffusa and Common Adulterants Based on Chloroplast Genome Sequencing and DNA Barcoding Markers. Plants, 2021; 10(1).DOI: 10.3390/plants10010161.
72. Zhu, B., Feng, Q., Yu, J., Yu, Y., Zhu, X.X., Wang, Y., et al. Chloroplast genome features of an important medicinal and edible plant: *Houttuynia cordata* (Saururaceae). PLoS One, 2020; 15(9): p. e0239823.DOI: 10.1371/journal.pone.0239823.
73. Alzahrani, D.A., Yaradua, S.S., Albokhari, E.J., Abba, A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genomics, 2020; 21(1): p. 393.DOI: 10.1186/s12864-020-06798-2.
74. Dong, F., Lin, Z.C., Lin, J., Ming, R., Zhang, W.P. Chloroplast Genome of Rambutan and Comparative Analyses in Sapindaceae. Plants, 2021; 10(2).DOI: 10.3390/plants10020283.

75. Wang, R.K., Fan, J.S., Chang, P., Zhu, L., Zhao, M.R., Li, L.L. Genome Survey Sequencing of *Acer truncatum* Bunge to Identify Genomic Information, Simple Sequence Repeat (SSR) Markers and Complete Chloroplast Genome. Forests, 2019; 10(2).DOI: 10.3390/f10020087.
76. Yu, T., Gao, J., Huang, B.H., Dayananda, B., Ma, W.B., Zhang, Y.Y., et al. Comparative Plastome Analyses and Phylogenetic Applications of the *Acer* Section Platanoidea. Forests, 2020; 11(4).DOI: 10.3390/f11040462.
77. Kim, S.C., Baek, S.H., Hong, K.N., Lee, J.W. Characterization of the complete chloroplast genome of *Koelreuteria paniculata* (Sapindaceae). Conserv. Genet. Resour., 2017; 10(1): p. 69-72.DOI: 10.1007/s12686-017-0767-4.
78. Ma, Q.Y., Wang, Y.A., Zhu, L., Bi, C.W., Li, S.X., Li, S.S., et al. Characterization of the Complete Chloroplast Genome of *Acer truncatum* Bunge (Sapindales: Aceraceae): A New Woody Oil Tree Species Producing Nervonic Acid. Biomed. Res. Int., 2019; 2019: p. 7417239.DOI: 10.1155/2019/7417239.
79. Lin, N., Moore, M.J., Deng, T., Sun, H., Yang, L.S., Sun, Y.X., et al. Complete plastome sequencing from *Toona* (Meliaceae) and phylogenomic analyses within Sapindales. Appl. Plant Sci., 2018; 6(4): p. e1040.DOI: 10.1002/aps3.1040.
80. Kim, H.T., Pak, J.H., Kim, J.S. The complete chloroplast genome sequence of *Acer takesimense* (Sapindaceae), an endemic to Ullenung Island of Korea. Mitochondrial DNA B, 2019; 4(1): p. 1531-1532.DOI: 10.1080/23802359.2019.1601521.
81. Wu, L.W., Nie, L.P., Xu, Z.C., Li, P., Wang, Y., He, C.N., et al. Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of Three *Paeonia* Section Moutan Species (Paeoniaceae). Front. Genet., 2020; 11: p. DOI: 10.3389/fgene.2020.00980.DOI: 10.3389/fgene.2020.00980.
82. Cheon, K.S., Kim, K.A., Kwak, M., Lee, B., Yoo, K.O. The complete chloroplast genome sequences of four *Viola species* (Violaceae) and comparative analyses with its congeneric species. PLoS One, 2019; 14(3): p. e0214162.DOI: 10.1371/journal.pone.0214162.
83. Li, D.M., Zhao, C.Y., Liu, X.F. Complete Chloroplast Genome Sequences of Kaempferia Galanga and Kaempferia Elegans: Molecular Structures and Comparative Analysis. Molecules, 2019; 24(3).DOI: 10.3390/molecules24030474.
84. Singh, R.B., Mahenderakar, M.D., Jugran, A.K., Singh, R.K., Srivastava, R.K. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. Gene, 2020; 753: p. 144800.DOI: 10.1016/j.gene.2020.144800.
85. Ramzan, M., Sarwar, S., Kauser, N., Saba, R., Hussain, I., Shah, A.A., et al. Assessment of Inter simple sequence repeat (ISSR) and simple sequence repeat (SSR) markers to reveal genetic diversity among *Tamarix* ecotypes. J. King Saud Univ. Sci., 2020; 32(8): p. 3437-3446.DOI: 10.1016/j.jksus.2020.10.003.
86. Motahari, B., Shabanian, N., Rahmani, M.S., Mohammad-Hasani, F. Genetic diversity and genetic structure of *Acer monspessulanum* L. across Zagros forests of Iran using molecular markers. Gene, 2021; 769: p. 145245.DOI: 10.1016/j.gene.2020.145245.
87. Zhao, Z.Y., Wang, X., Yu, Y., Yuan, S., Jiang, D., Zhang, Y.J., et al. Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses. PeerJ, 2018; 6: p. e6032.DOI: 10.7717/peerj.6032.

88. Li, W., Zhang, C., Guo, X., Liu, Q., Wang, K. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. PLoS One, 2019; 14(5): p. e0216645.DOI: 10.1371/journal.pone.0216645.
89. Zhang, F.J., Wang, T., Shu, X.C., Wang, N., Zhuang, W.B., Wang, Z. Complete Chloroplast Genomes and Comparative Analyses of *L. chinensis*, *L. anhuiensis*, and *L. aurea* (Amaryllidaceae). Int. J. Mol. Sci., 2020; 21(16).DOI: 10.3390/ijms21165729.
90. Wen, F., Wu, X.Z., Li, T.J., Jia, M.L., Liu, X.S., Liao, L. The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. BMC Genomics, 2021; 22(1): p. 161.DOI: 10.1186/s12864-021-07484-7.
91. Khayi, S., Gaboun, F., Pirro, S., Tatusova, T., El Mousadik, A., Ghazal, H., et al. Complete Chloroplast Genome of *Argania spinosa*: Structural Organization and Phylogenetic Relationships in Sapotaceae. Plants, 2020; 9(10): p. DOI:10.3390/plants9101354.DOI: 10.3390/plants9101354.
92. Meng, D., Xiaomei, Z., Wenzhen, K., Xu, Z. Detecting useful genetic markers and reconstructing the phylogeny of an important medicinal resource plant, *Artemisia selengensis*, based on chloroplast genomics. PLoS One, 2019; 14(2): p. e0211340.DOI: 10.1371/journal.pone.0211340.
93. Khan, A., Asaf, S., Khan, A.L., Al-Harrasi, A., Al-Sudairy, O., AbdulKareem, N.M., et al. First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. PLoS One, 2019; 14(1): p. e0208511.DOI: 10.1371/journal.pone.0208511.
94. Hong, Z., Wu, Z.Q., Zhao, K.K., Yang, Z.J., Zhang, N.N., Guo, J.Y., et al. Comparative Analyses of Five Complete Chloroplast Genomes from the Genus *Pterocarpus* (Fabacaeae). Int. J. Mol. Sci., 2020; 21(11).DOI: 10.3390/ijms21113758.
95. Chi, X.F., Zhang, F.Q., Dong, Q., Chen, S.L. Insights into Comparative Genomics, Codon Usage Bias, and Phylogenetic Relationship of Species from Biebersteiniaceae and Nitrariaceae Based on Complete Chloroplast Genomes. Plants, 2020; 9(11).DOI: 10.3390/plants9111605.
96. Shi, C., Han, K., Li, L., Seim, I., Lee, S.M., Xu, X., et al. Complete Chloroplast Genomes of 14 Mangroves: Phylogenetic and Comparative Genomic Analyses. Biomed Res Int, 2020; 2020: p. 8731857.DOI: 10.1155/2020/8731857.
97. Wang, L., He, N., Li, Y., Fang, Y., Zhang, F. Complete Chloroplast Genome Sequence of Chinese Lacquer Tree (Toxicodendron vernicifluum, Anacardiaceae) and Its Phylogenetic Significance. Biomed Res Int, 2020; 2020: p. 9014873.DOI: 10.1155/2020/9014873.

Figures

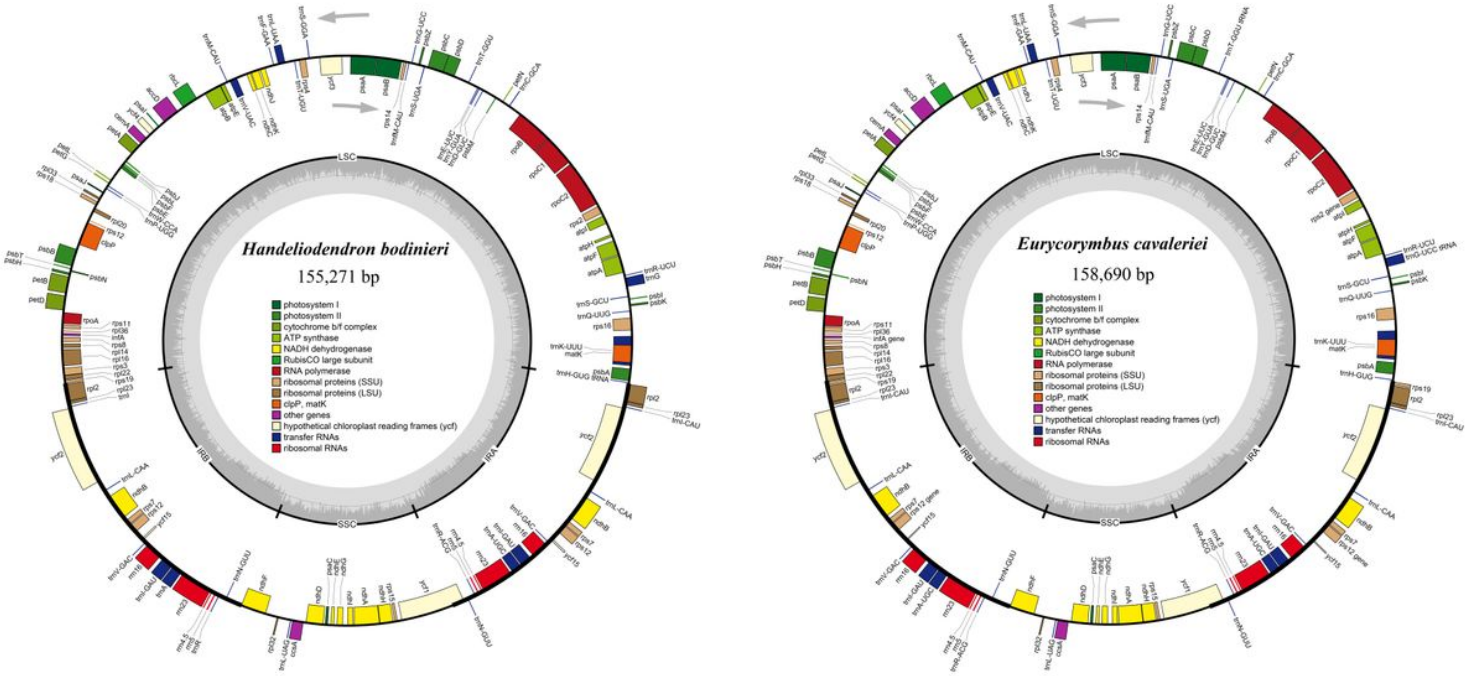

Figure 1

Circular gene map of chloroplast genomes of Handeliodendron bodinieri and Eurycorymbus cavaleriei. The gray arrowheads indicate the direction of the genes. Genes on the outside and inside of the circle are transcribed in clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are color coded. The innermost darker gray corresponds to GC content, whereas the lighter gray corresponds to AT content. IR, LSC, SSC indicate inverted repeat, large single copy region, and small single copy region, respectively.

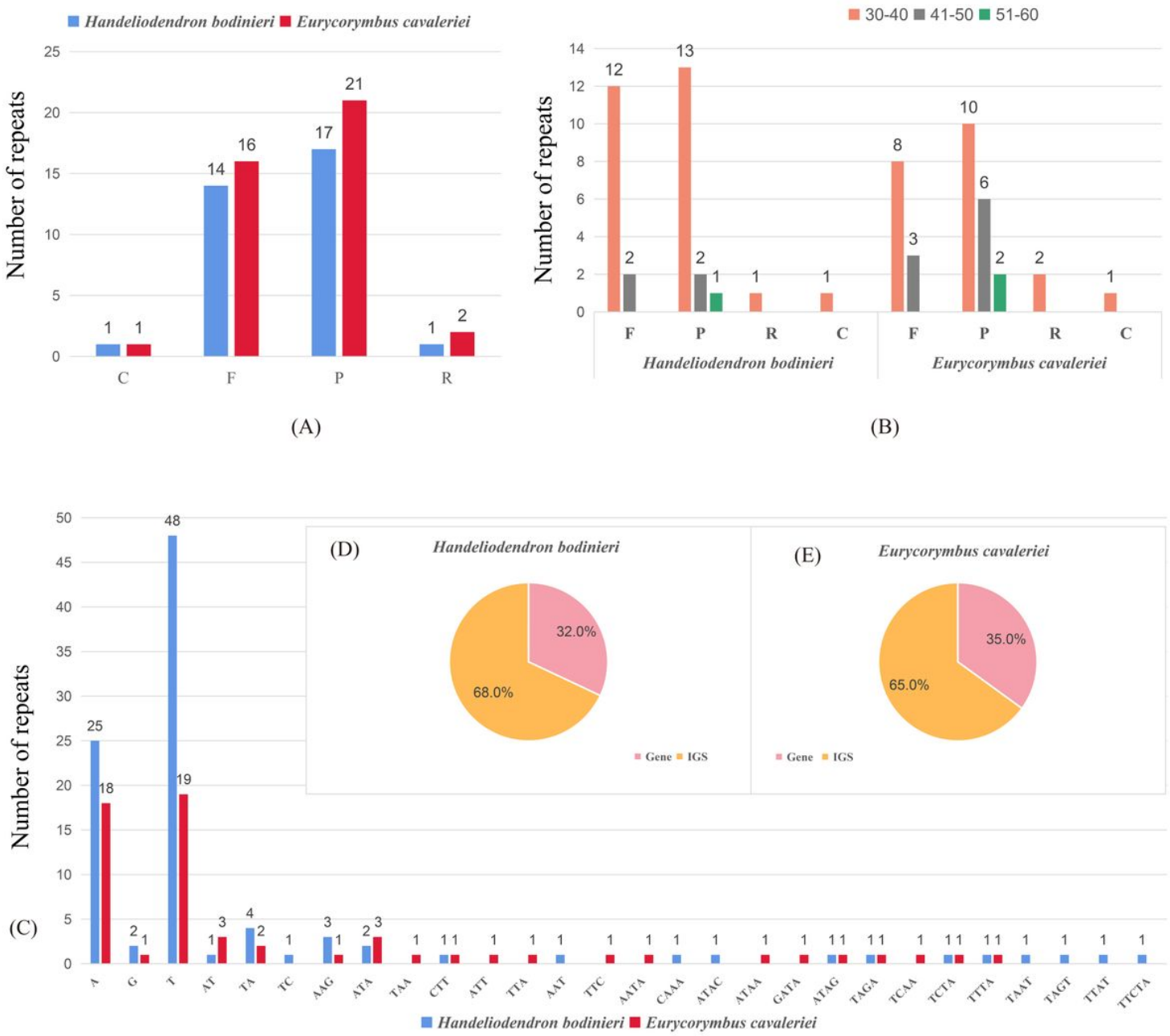

Figure 2

Analyses of repeated sequences in Handeliodendron bodinieri and Eurycorymbus cavaleriei complete chloroplast genomes. (A) Total of dispersed repeats, the F, R, P, C indicate forward, reverse, palindromic, and complement repeats, respectively; (B) Frequency of dispersed repeats by length; (C) Numbers and types of SSR in the chloroplast genomes of H. bodinieri and E. cavaleriei; (D) SSR distribution between Gene and intergenic spacer regions (IGS) of H. bodinieri; (E) SSR distribution between Gene and intergenic spacer regions (IGS) of E. cavaleriei.

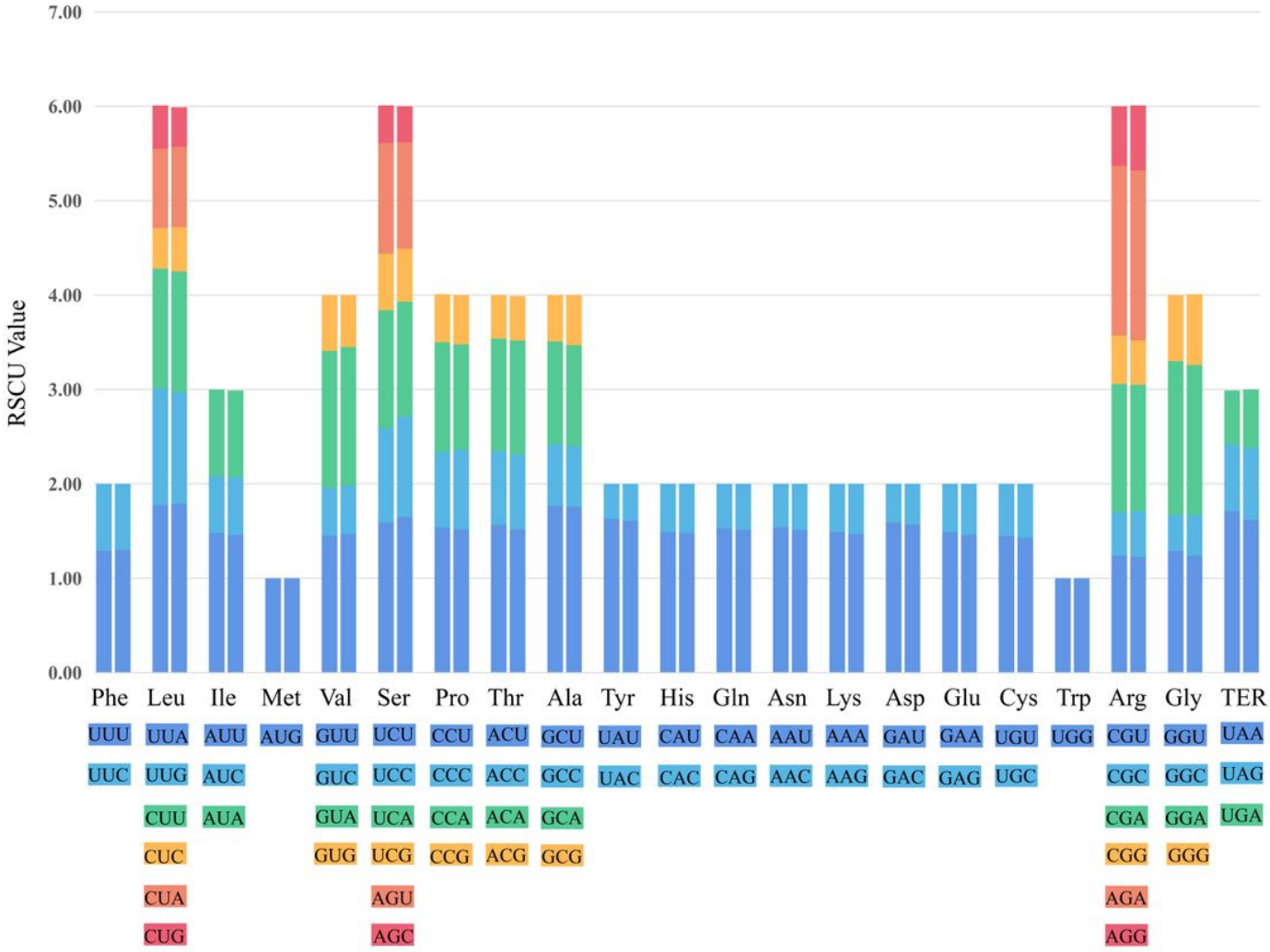

Figure 3

Codon contents of 20 amino acids and stop codons in all coding sequences of the Handeliodendron bodinieri and Eurycorymbus cavaleriei chloroplast genomes. The histogram on the left side presents each amino acid codon usage within H. bodinieri cp genome, and the right side denotes that of within E. cavaleriei cp genome. The colour of the histogram corresponds to the colour of codons.

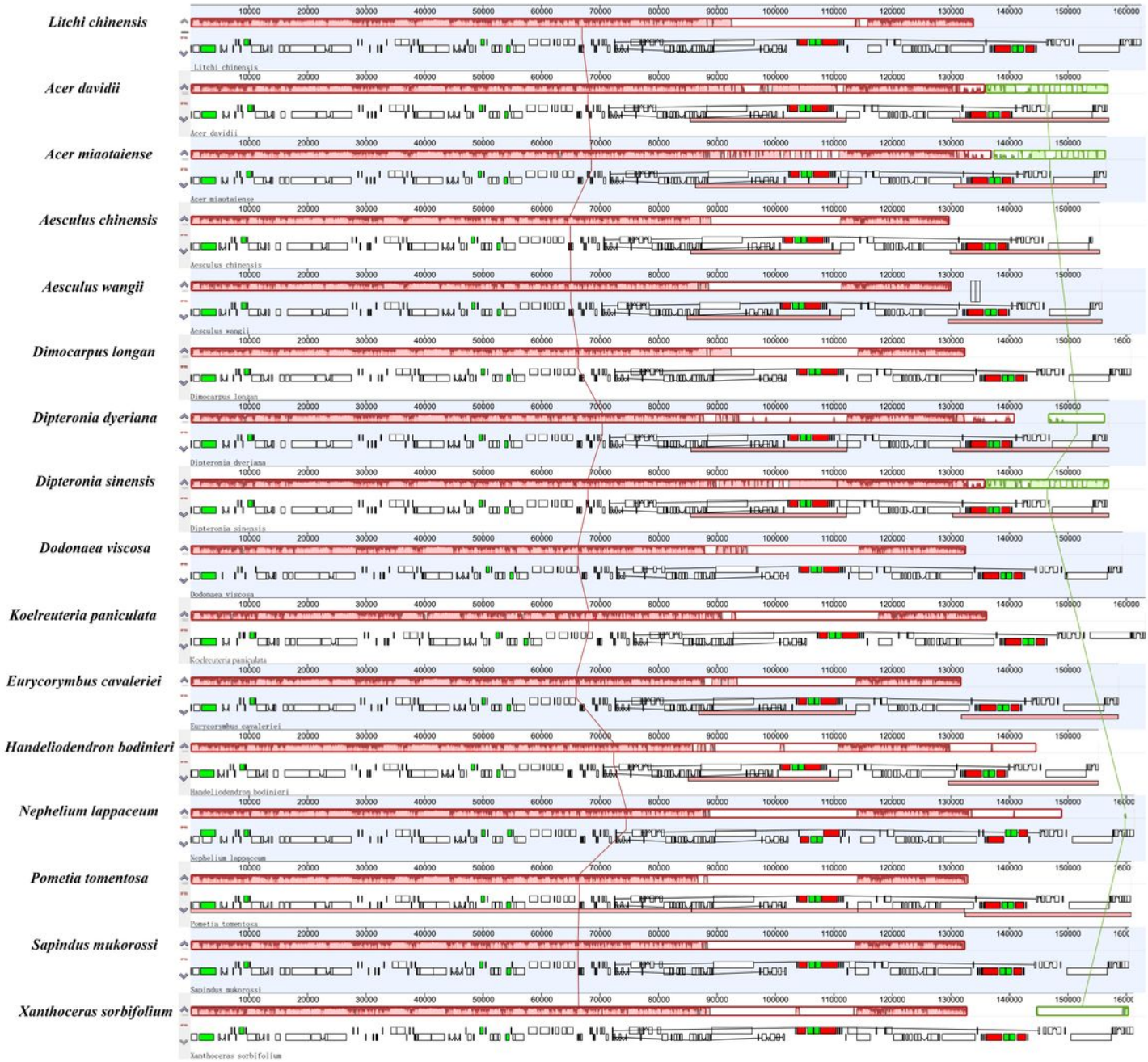

Figure 4

Genomic rearrangement of the 16 Sapindaceae chloroplast genomes, with Litchi chinensis set as a reference genome.

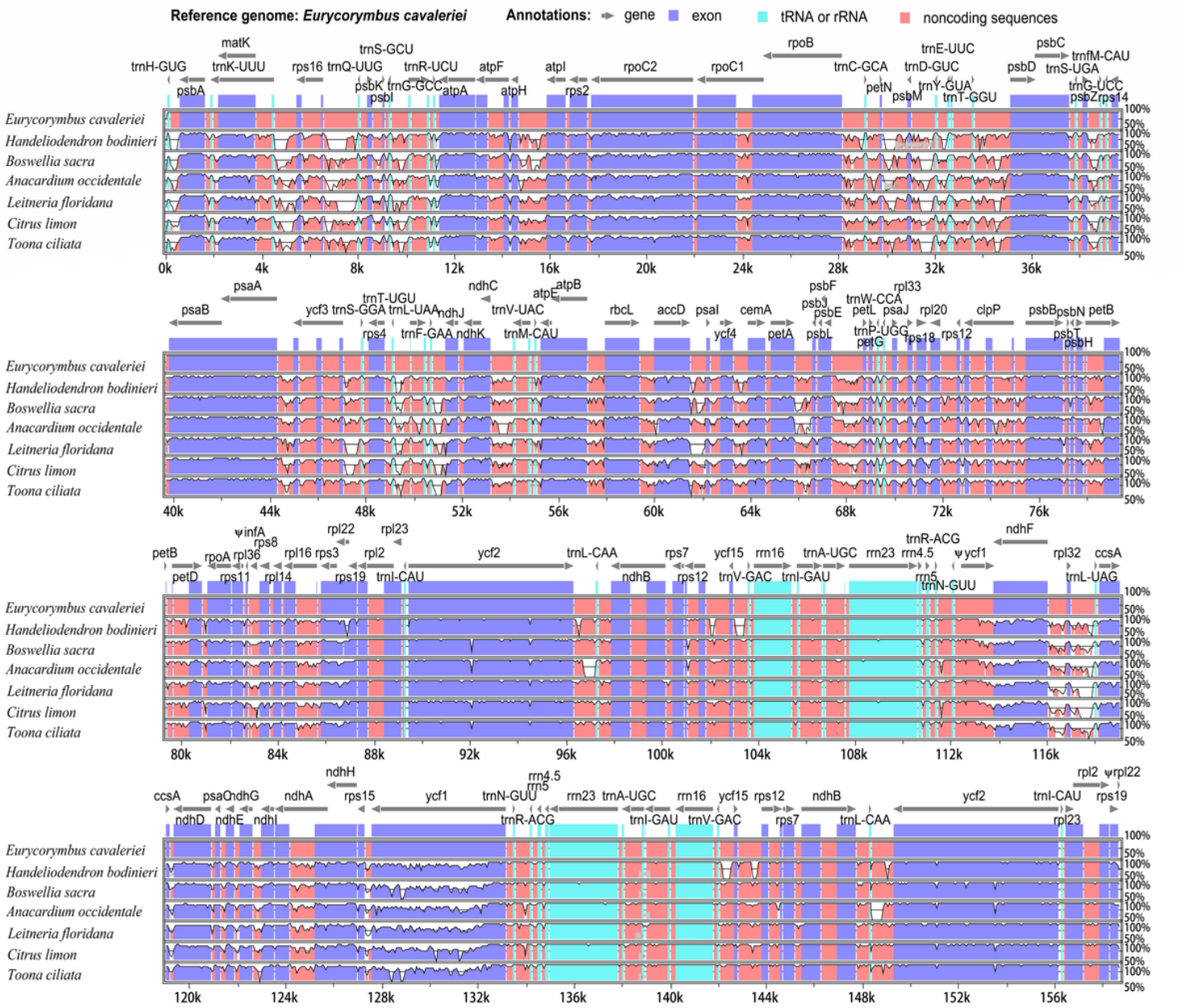

Figure 5

Visualization of genome alignment of the complete genome of seven complete chloroplast genomes from Sapindales. The cp genome of Eurycorymbus cavaleriei is used as the reference. X-axis indicates the sequence coordinates in the whole cp genome. Y-axis represents the similarity of the aligned regions, indicating percent identity to the reference genome (50-100%).

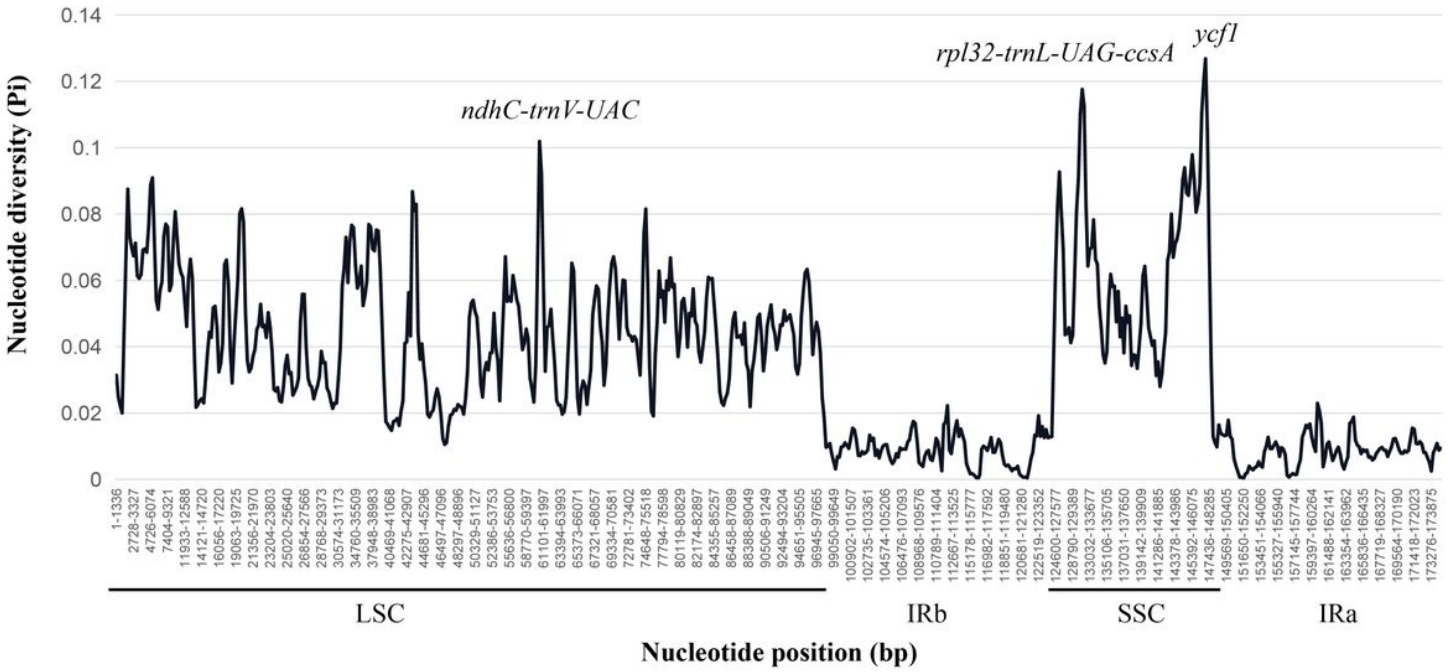

Figure 6

Sliding window analysis based on the 13 cp genome sequences of Sapindaceae. X-axis: position of the Midpoint of a window; Y-axis: Nucleotide diversity of each window (Pi).

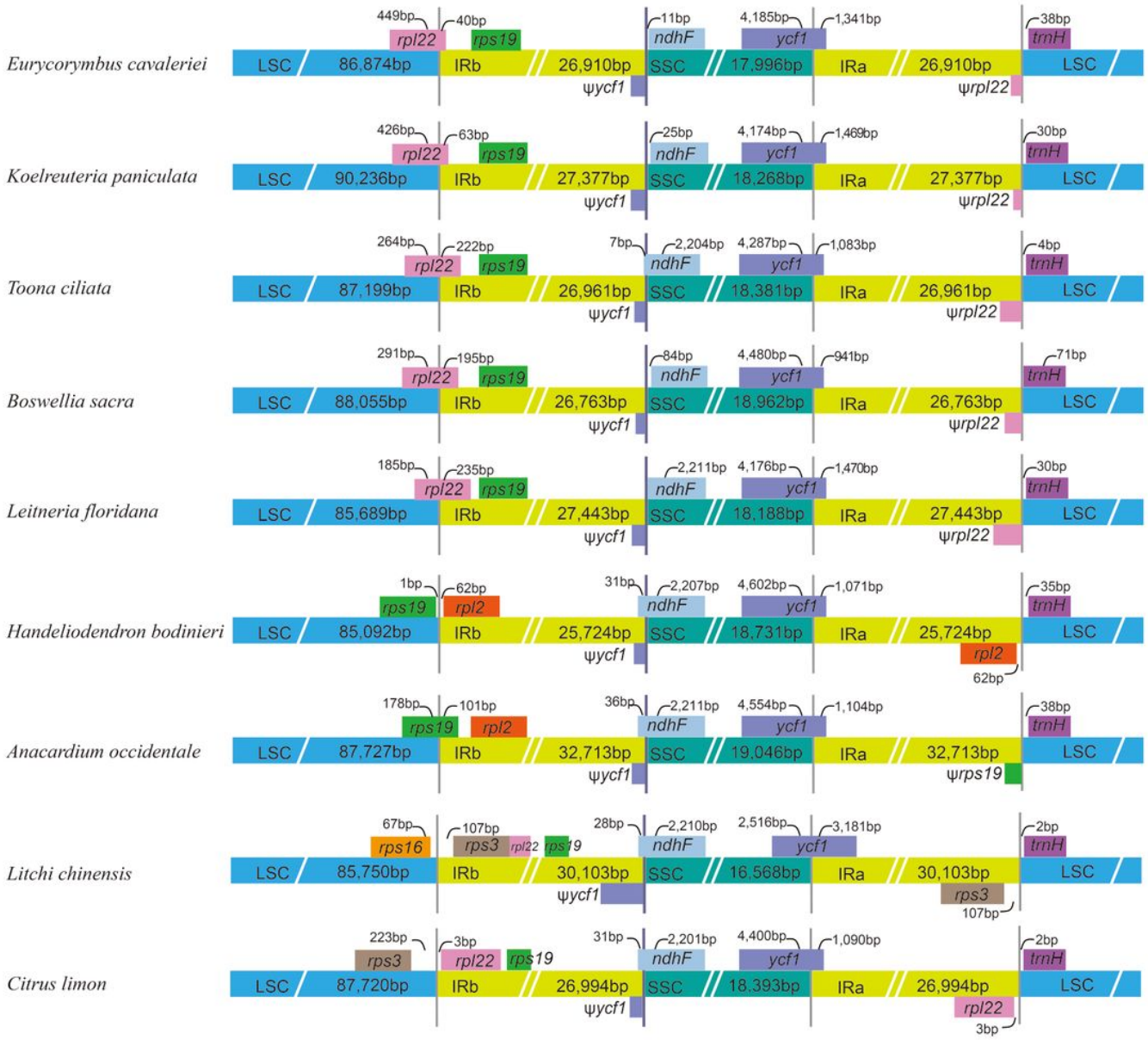

Figure 7

Comparison of the Large Single-Copy (LSC), Small Single-Copy (SSC), and inverted repeat (IR) boundary regions across nine Sapindales chloroplast genomes, ψ indicates a pseudogene.

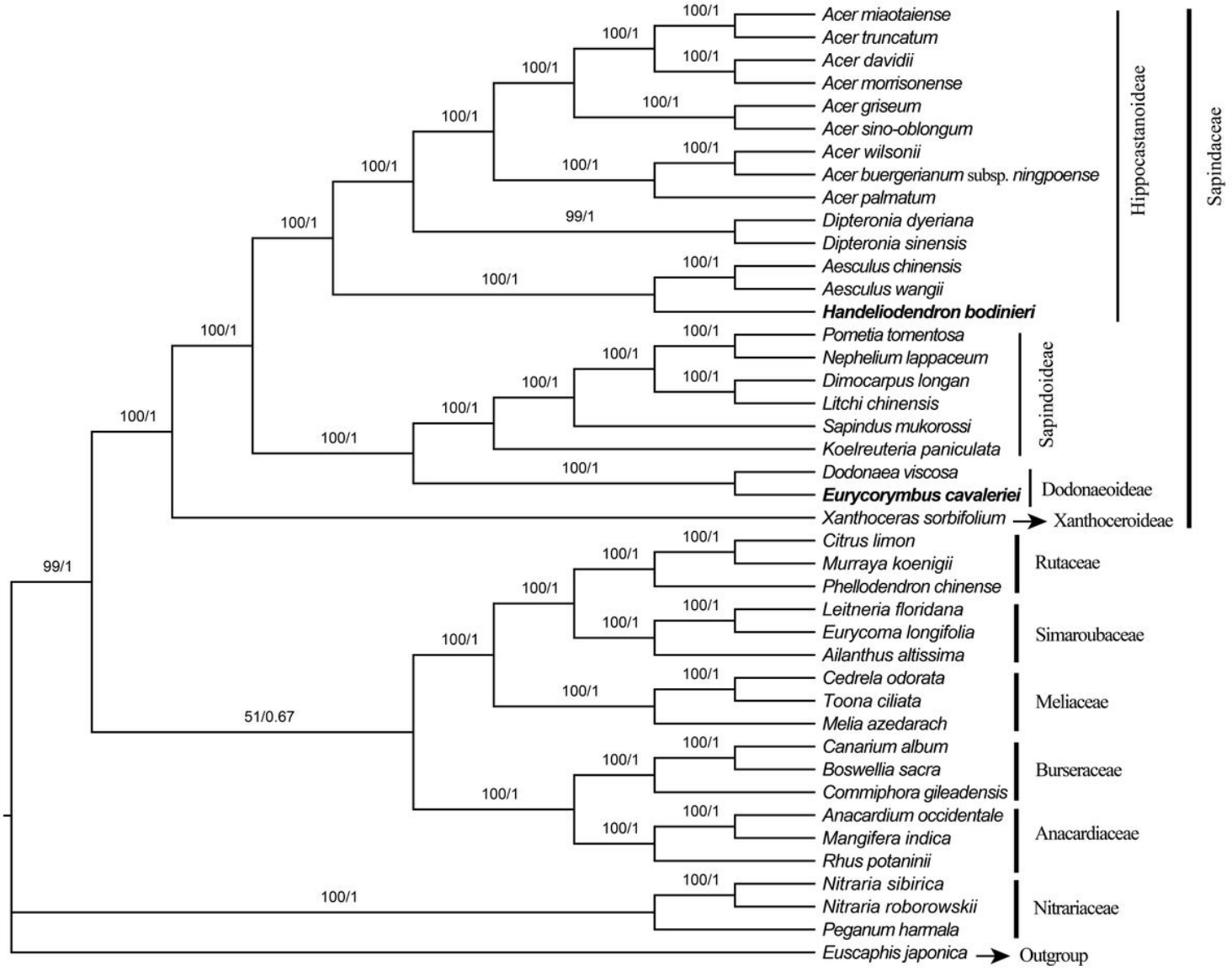

Figure 8

Phylogenetic tree reconstruction of Sapindales using the maximum likelihood (ML) and Bayesian inference (BI) method based on complete genome sequences. Only the ML tree is shown, because its topology is identical to that of the obtained BI tree. ML supports/ BI posterior probabilities values are indicated on the nodes

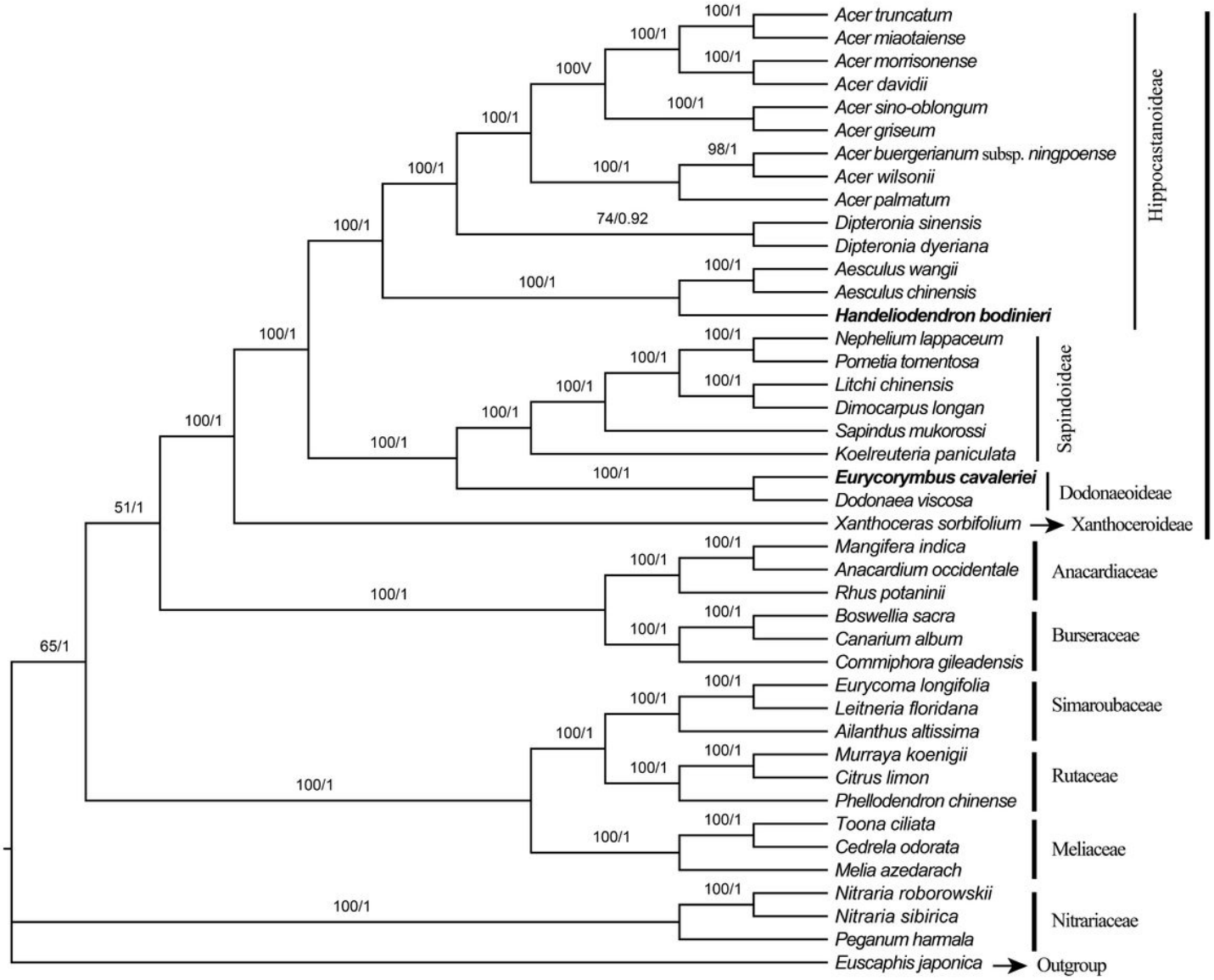

Figure 9

Phylogenetic tree reconstruction of Sapindales using the maximum likelihood (ML) Bayesian inference (BI) method based on 79 coding sequences. Only the ML tree is shown, because its topology is identical to that of the obtained BI tree. ML supports/BI posterior probabilities values are indicated on the node

Supplementary Files

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

- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- TableS1.xlsx

- TableS2.docx
- TableS3.docx
- TableS4.docx
- TableS5.docx
- TableS6.docx
- supplementarylegends.docx