Comparative analyses of Diospyros (Ebenaceae) plastomes: Insights into genomic features, mutational hotspots, and adaptive evolution

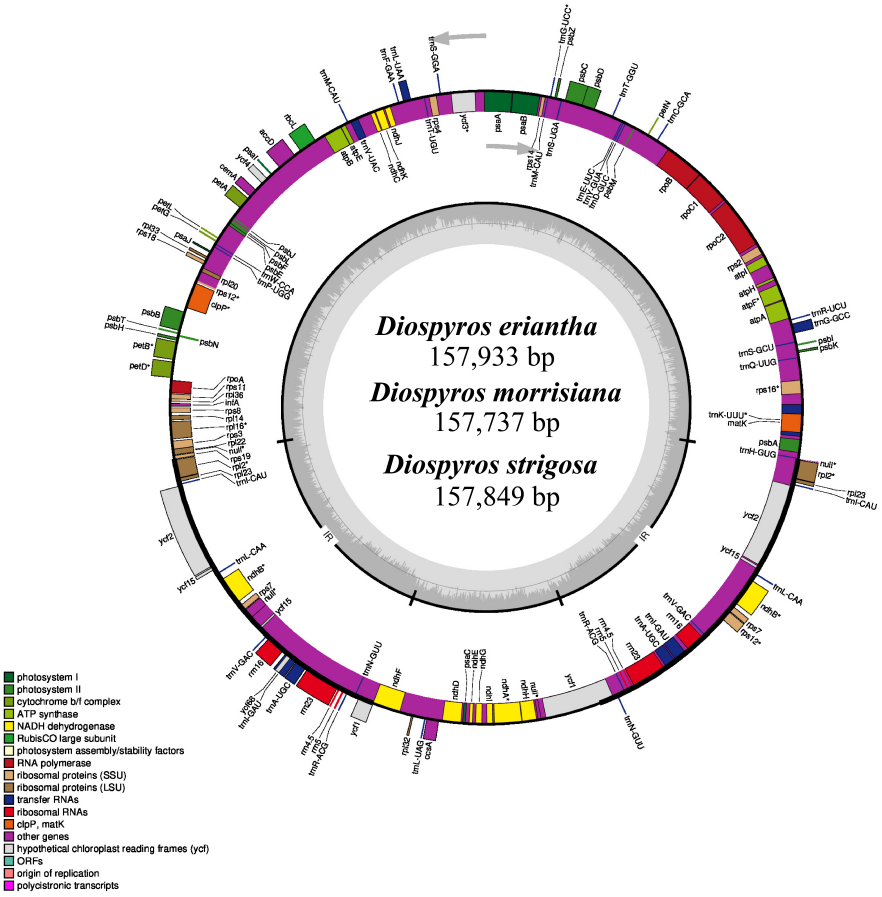
Jing Sun[1], Qing Ma[2], Yue Huang[1], Chan-Juan Lai[1], Pan Li[3], Xin-Jie Jin[1], and Yong-Hua Zhang[1]

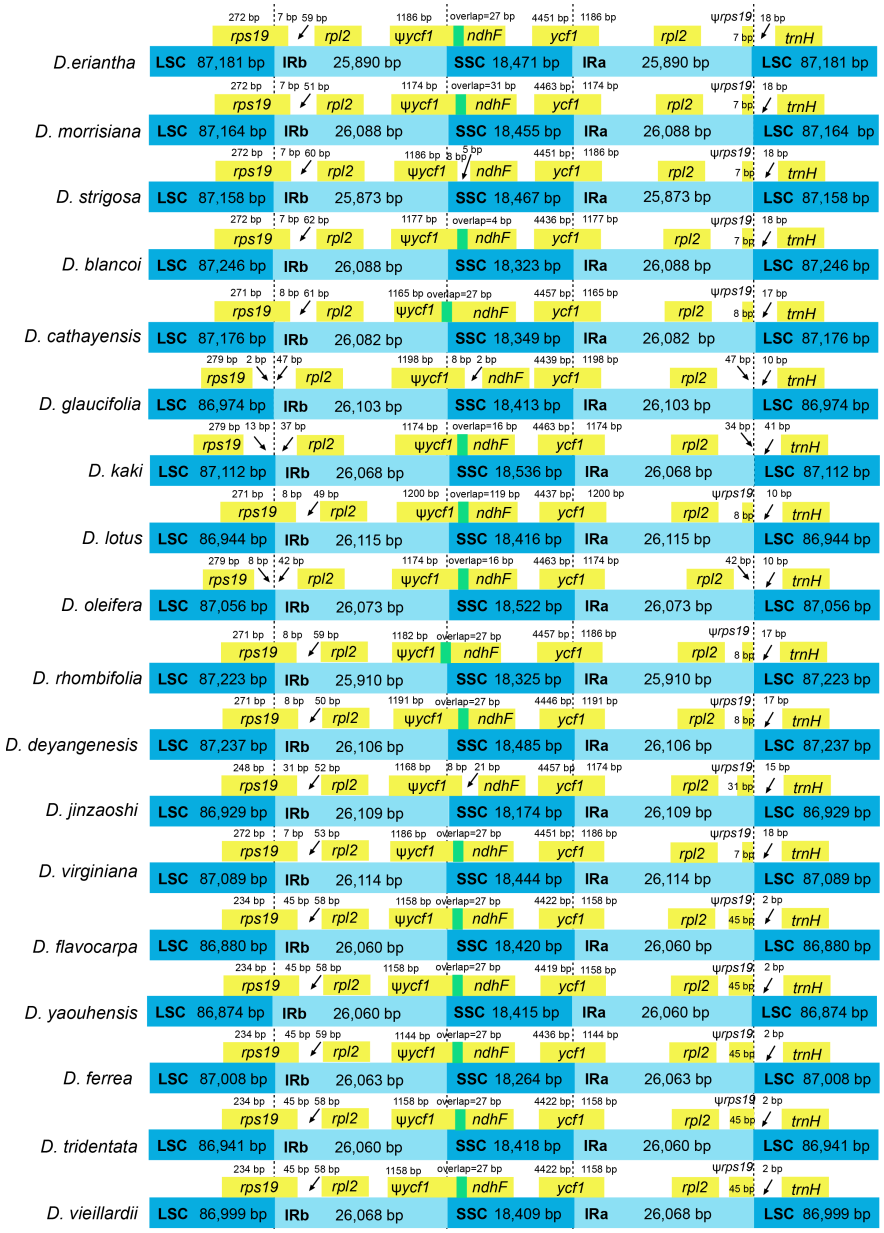
[1]Wenzhou University
[2]Zhejiang Shuren University
[3]Zhejiang University

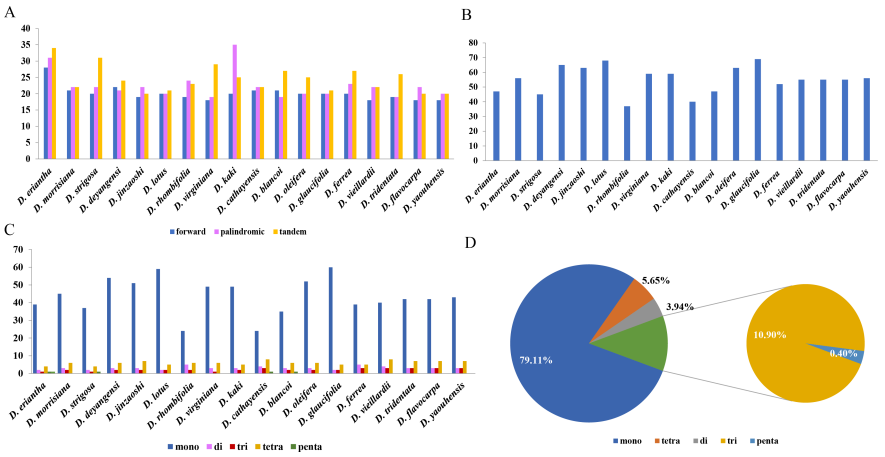
October 25, 2022

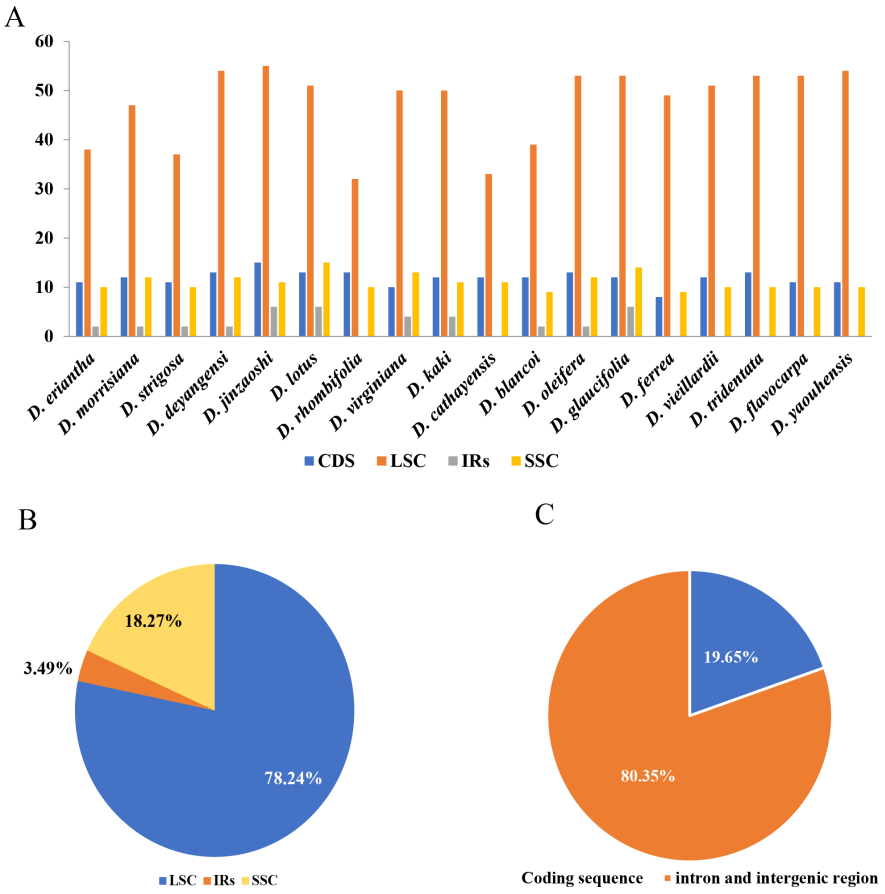
Abstract

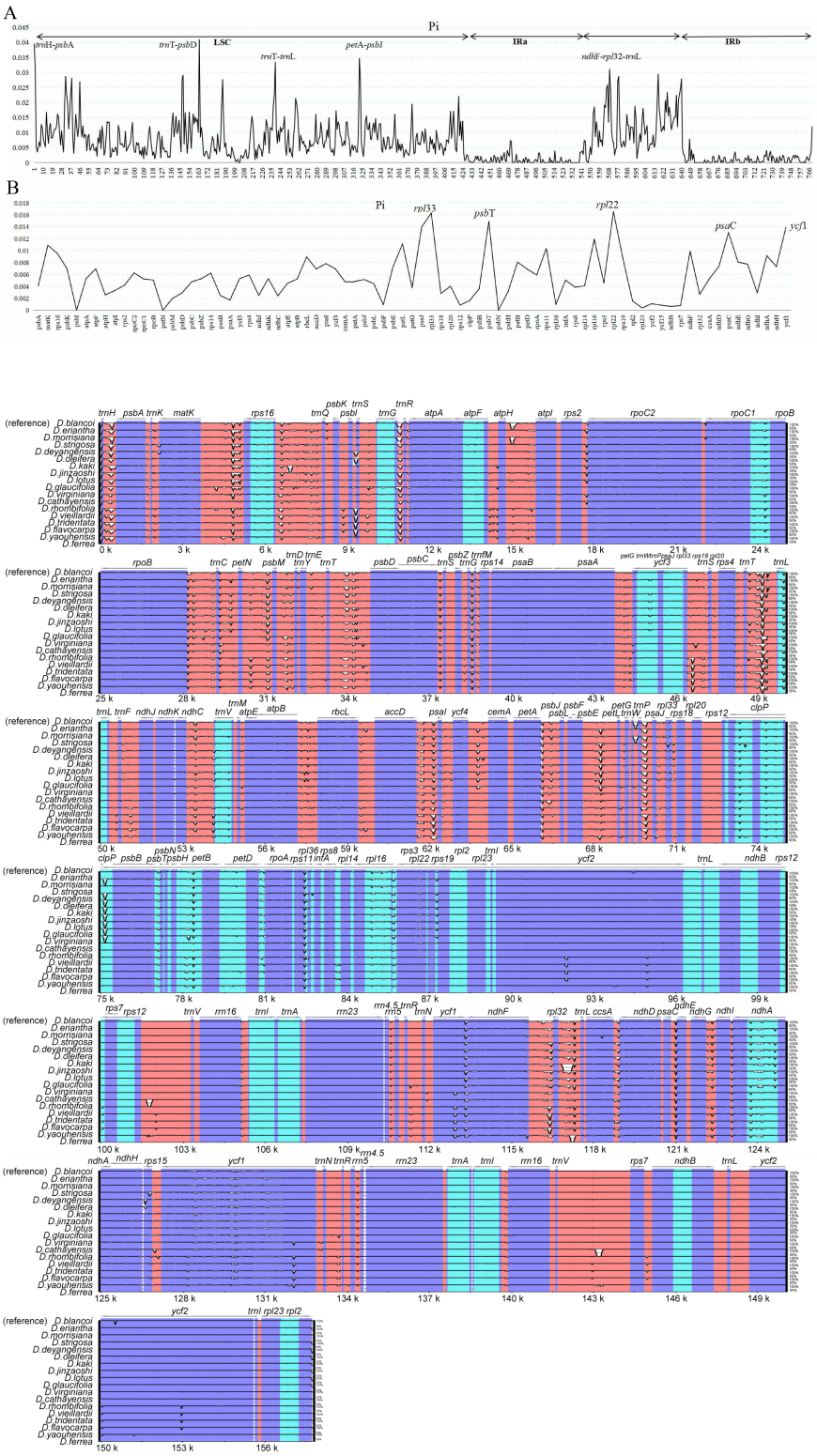
Diospyros (Ebenaceae) is a widely distributed genus of trees and shrubs native to tropical and subtropical regions, with numerous species valued for their fruits (persimmons), timber, and medicinal values. However, information regarding their plastomes and chloroplast evolution is scarce. The present study performed comparative genomic and evolutionary analyses on plastomes of 18 accepted Diospyros species, including three newly sequenced ones. Our study showed a highly conserved genomic structure across the species, with plastome size ranging from 157,321 bp (D. jinzaoshi) to 157,934 bp (D. deyangensis). These plastomes encoded 134–138 genes, including 89–91 protein-coding genes, 1–2 pseudogenes (Ψycf1 for all, Ψrps19 for a few), 37 tRNA genes, and 8 rRNA genes. Comparative analysis of Diospyros identified the intergenic regions (trnH-psbA, rps16-trnQ, trnT-psbD, petA-psbJ, trnL-trnF-ndhJ) as the mutational hotspots in these species. Phylogenomic analyses identified three main groups within the genus designated as the evergreen, deciduous, and island groups. The codon usage analysis identified 30 codons with relative synonymous codon usage (RSCU) values greater than 1 and 29 codons ending with A and U bases. A total of three codons (UUA, GCU, and AGA) with highest (RSCU) values were identified as the optimal codons. ENC-plot indicated the significant role of mutational pressure in shaping codon usage, while most protein-coding genes in Diospyros experienced relaxed purifying selection (Ka/Ks < 1). Additionally, the ndhG, rpoC1, and ycf3 genes showed positive selection (Ka/Ks $>$ 1) in the island, deciduous, and both deciduous and evergreen species, respectively. Thus, the results provide a foundation for elaborating Diospyros's genetic architecture and taxonomy, conserving genetic diversity and enriching genetic resources.











Hosted file

Table S1.docx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S2.docx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S3.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S4.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S5.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S6.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S7.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Hosted file

Table S8.xlsx available at https://authorea.com/users/517336/articles/591922-comparative-analyses-of-diospyros-ebenaceae-plastomes-insights-into-genomic-features-mutational-hotspots-and-adaptive-evolution

Comparative analyses of *Diospyros* (Ebenaceae) plastomes: Insights into genomic features, mutational hotspots, and adaptive evolution

Jing Sun[1,4], Ma Qing[2,4], Huang Yue[1], Chan-Juan Lai[1], Pan Li[3], Xin-Jie Jin[1*] & Yong-Hua Zhang[1*]

[1]College of Life and Environmental Science, Wenzhou University, Wenzhou, 325035, China.

[2]College of Biology and Environmental Engineering, Zhejiang Shuren University, Hangzhou 310015, China.

[3]Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058, China.

[4]These authors contributed equally to this work.

*Corresponding authors: Yong-Hua Zhang (zhangyhua@wzu.edu.cn) & Xin-Jie Jin (xinjie_jin@yeah.net)

ABSTRACT

Diospyros (Ebenaceae) is a widely distributed genus of trees and shrubs native to tropical and subtropical regions, with numerous species valued for their fruits (persimmons), timber, and medicinal values. However, information regarding their plastomes and chloroplast evolution is scarce. The present study performed comparative genomic and evolutionary analyses on plastomes of 18 accepted *Diospyros* species, including three newly sequenced ones. Our study showed a

highly conserved genomic structure across the species, with plastome size ranging from 157,321 bp (*D. jinzaoshi*) to 157,934 bp (*D. deyangensis*). These plastomes encoded 134–138 genes, including 89–91 protein-coding genes, 1–2 pseudogenes (Ψ *ycf*1 for all, Ψ *rps*19 for a few), 37 tRNA genes, and 8 rRNA genes. Comparative analysis of *Diospyros* identified the intergenic regions (*trn*H-*psb*A, *rps*16-*trn*Q, *trn*T-*psb*D, *pet*A-*psb*J, *trn*L-*trn*F-*ndh*J) as the mutational hotspots in these species. Phylogenomic analyses identified three main groups within the genus designated as the evergreen, deciduous, and island groups. The codon usage analysis identified 30 codons with relative synonymous codon usage (RSCU) values greater than 1 and 29 codons ending with A and U bases. A total of three codons (UUA, GCU, and AGA) with highest (RSCU) values were identified as the optimal codons. ENC-plot indicated the significant role of mutational pressure in shaping codon usage, while most protein-coding genes in *Diospyros* experienced relaxed purifying selection (Ka/Ks < 1). Additionally, the *ndh*G, *rpo*C1, and *ycf*3 genes showed positive selection (Ka/Ks > 1) in the island, deciduous, and both deciduous and evergreen species, respectively. Thus, the results provide a foundation for elaborating *Diospyros*'s genetic architecture and taxonomy, conserving genetic diversity and enriching genetic resources.

KEYWORDS: *Diospyros*, Plastome, Hyper-variable region, Genetic diversity

INTRODUCTION

44 *Diospyros* (Ebenaceae) is a genus well-known for hardwood and delicious fruits. It
45 is also used for medicines in tropical and temperate regions (Lee et al., 1996;
46 Wallnöfer, 2001; Luo et al., 2021; White, 1956, Lin et al., 2020). *Diospyros* is the
47 largest genus of the Ebenaceae family, with about 500 evergreen or deciduous shrub
48 and tree species distributed worldwide (Lee et al., 1996; The plant list, 2002). But
49 only a few members in the genus are economically important, so it is important to
50 distinguish the species for conservation and utilization of wild relatives. The genus is
51 characterized by male cymose inflorescence, solitary female flowers, fleshy berries
52 with enlarged persistent calyx at the base, and a dioecious breeding system (Lee et
53 al., 1996). However, the morphological similarities make it difficult to distinguish
54 the species, hindering research and economic development.

55 Previous infrafamilial classification based on a phylogenetic approach
56 (multilocus) proposed that Ebenaceae consists of two subfamilies, Lissocarpoideae
57 and Ebenoideae, and four genera, *Lissocarpa*, *Euclea*, *Royena*, and *Diospyros*
58 (Duangjai et al. 2006). Previous studies found that *Diospyros* belongs to the
59 Ebenoideae subfamily (Ebenaceae) and is closely associated with *Euclea* Murray
60 and *Royena* L. (Duangjai et al. 2006; Duangjai et al. 2009; Linan et al. 2019; Li et al.
61 2018; Fu et al. 2016; Samuel et al. 2019). Within the genus, about 11 (or 12) clades
62 were supported by molecular phylogenetic studies based on multilocus or genomes
63 (Duangjai et al. 2006; Duangjai et al. 2009; Linan et al. 2019). However, there is
64 little study of *Diospyros* about phylogeny-based evolution analysis. Some *Diospyros*
65 spp. have adapted to high latitudes towards a deciduous habit but the species in low

66 latitudes towards a evergreen habit (Lee et al., 1996; Duangjai et al. 2009), while
67 few taxa are endemic to island environments (Turner et al., 2016). Therefore, to
68 understand the strategies to adapt to different environmental conditions, the research
69 for leaf habits of *Diospyros* has great significance (Tomlinson, et al. 2013; Yao, et
70 al., 2020). The high-latitude or high-elevation species, such as *D. kaki* Thunb. and *D.*
71 *lotus* L., are deciduous, while low-latitude or low-elevation species, such as *D.*
72 *cathayensis* Steward and *D. ferrea* (Willd.) Bakh., are evergreen (Lee et al., 1996).
73 Research has established that the plants on islands have been shaped by ancestral
74 bottlenecks, rapid and recent radiations in phenotypic characters, and repeated and
75 convergent evolution of potentially adaptive traits during the diversification
76 (Fernández-Mazuecos et al., 2020). *Diospyros* taxa of the islands (New Caledonia)
77 also experienced similar evolutionary pressure (Turner et al., 2016). Adaptive
78 evolution of *Diospyros* spp. driven by natural or mutation selection is the basis of
79 biodiversity and a significant driving force of speciation (Morgan, 1925). However,
80 the relationship between environmental adaptation (leaf habits) and genetic diversity
81 in *Diospyros* species has rarely been discussed (See Samuel et al. 2019). Therefore,
82 on the basis of previous molecular phylogenetic researches, it is of great significance
83 to study the adaptive evolution of *Diospyros*, which have obvious leaf habits, by
84 using new molecular markers such as plastomes.

85 The structurally stable and maternally inherited plastomes with low
86 recombinant levels play a pivotal role in phylogenetic and evolutionary studies
87 (Jansen et al., 2007; Wicke et al. 2011; Xia et al. 2022a; Xia et al. 2022b). The genes

88 in plastomes primarily encode proteins related to photosynthesis and other
89 biochemical pathways, including starch storage, nitrogen and sulfate metabolism,
90 and chlorophyll, carotenoid, or fatty acid synthesis (Wicke et al. 2011; Mohanta et al.
91 2020). Moreover, plastomes are considered conserved in terms of genomic structures
92 and substitution rates among most Angiosperms, which make plastomes into a
93 widely used molecular marker. Additionally, several studies have detected positive
94 selection signals in plastid genes during evolution. For example, accelerated
95 evolutionary rates of *mat*K (Maturase K) in the low-altitude and recently derived
96 lineages of *Dysosma* have been related to the adaptation of the genus to high-altitude
97 environments (Ye et al. 2018). Furthermore, analysis of the Ka/Ks ratios of
98 Cardamineae suggested positive selection on the *ycf*2 (hypothetical chloroplast RF21)
99 gene in watercress, possibly allowing the species to adapt to specific living
100 environments (Yan et al. 2019). Most plastid genes are under selection pressure due
101 to their significant roles in maintaining essential cellular functions and, therefore,
102 often retain the adaptive characteristics during evolution (Wicke et al. 2011). The
103 codon usage bias in plastomes serves as a suitable strategy for identifying the
104 principal evolutionary driving forces (Kapralov et al. 2007; Jiang et al. 2014; Gao et
105 al. 2022). For example, the effective number of codons (ENC)-plot showing
106 deviations from the expected curve for a few genes suggested that apart from natural
107 selection, mutational pressure also played a major role in shaping codon usage in
108 *Helianthus annuus* (Gao et al. 2022). These findings have demonstrated that the

genetic diversity in plastomes provides useful information about plants' adaptive evolution.

Therefore, the present study mainly aimed to study the adaptive evolution of *Diospyros* using plastomes. We included plastomes of 18 accepted *Diospyros* species with two leaf habits: deciduous (clade IX in Duangjai et al. 2009, subtropical to temperate regions of the Northern Hemisphere) and evergreen (clade III & XI in Duangjai et al. 2009; island specialized taxa from New Caledonia and general evergreen taxa from Asia). The specific objectives of the study were to (1) evaluate the plastome variations in *Diospyros* among the 18 species; (2) develop new and efficient plastid DNA (ptDNA) markers for DNA barcoding and perform the phylogenetic analyses for *Diospyros* species identification; and (3) analyze the Ka/Ks ratios and the codon usage bias of plastid genes to explore the value differences in each leaf habits and (or) the island taxa which are associated with environmental pressure.

MATERIALS AND METHODS

DNA extraction

The plastomes of three *Diospyros* species, *D. strigosa* Hemsl., *D. morrisiana* Hance, and *D. eriantha* Champ. ex Benth., were sequenced for the first time in this study collected from South China Botanical Garden and Guangdong Province (Table 1). The specimens have been deposited in the Herbarium of Wenzhou University (Table 1). Genomic DNA was extracted from approximately 20 mg of silica-dried leaves

131 using DNA Plantzol Reagent (Hangzhou Lifefeng Biotechnology Co., Ltd,
132 Hangzhou, China). The quality and quantity of the extracted DNA samples were
133 assessed using agarose gel electrophoresis and ultraviolet-microspectrophotometry.
134

135 **Genome sequencing, assembly, and annotation**

136 Approximately 1 μg of the extracted DNA with a concentration higher than 12.5
137 ng/μL was used for plastome sequencing at the Beijing Genomics Institute (BGI,
138 Wuhan, China). Before sequencing, total DNA was sheared into fragments shorter
139 than 800 bp. The DNA fragments' quality was evaluated using Agilent Bioanalyzer
140 2100 (Agilent Technologies), and the pooled library was sequenced on an Illumina
141 HiSeq X10 platform to obtain 150 bp long raw reads.
142 The raw reads were filtered by removing the sequences with a Phred score lower
143 than 30, and the remaining ones were used for genome assembly using GetOrganelle
144 toolkit (Jin et al., 2020). The command lines used for the assembly were as follows:
145 get_organelle_reads.py -1 forward.fq -2 reverse.fq -o plastome_output -R 15 -k
146 21,45,65,85,105 -F plant_cp. The newly sequenced plastomes of *Diospyros* species
147 were annotated with Geneious Prime 2021 (Biomatters, Auckland, New Zealand),
148 using the plastome sequence of *D. virginiana* L. (GenBank accession No. MF288577)
149 as the reference. The CPGAVAS2 web server
150 (http://www.herbalgenomics.org/cpgavas) predicted the types and structures of all
151 the protein-coding and noncoding genes in the plastome. The location of the start
152 and stop codons, exon-intron boundaries, and the tRNA gene length and types were

153 confirmed by comparing the annotation results from CPGAVAS2 and Geneious.
154 Finally, the plastome maps for the newly sequenced species were drawn using the
155 online tool OrganellarGenomeDRAW (Lohse et al., 2007). Plastomes of 15 other
156 *Diospyros* species and two outgroups (*Primula malacoides* and *Impatiens balsamina*)
157 (Table 2, Fig. 6) were downloaded from NCBI GenBank repository and re-annotated
158 using the earlier method. According to the leaf habits of *Diospyros* species, it can be
159 divided into evergreen (five species), deciduous (eight species), and island groups
160 (five species) (Table 2).
161

162 **Plastome comparison**

163 The GenBank accession numbers of the plastomes of the 18 *Diospyros* species used
164 for comparative analyses are shown in Table 2. The plastome sequences of these 18
165 *Diospyros* species were aligned using the LAGAN model implemented in the
166 mVISTA software to evaluate the degree of variation (Frazer et al., 2004), using
167 default parameters and *Diospyros blancoi* as the reference. The rearrangement in the
168 sequences was detected using the whole genome alignment tool Mauve implemented
169 in Geneious (Darling et al., 2004).
170

171 **Detection of repeated sequences**

172 Repeated sequences are essential components of the gene regulatory network; they
173 are identical or complementary nucleotide fragments distributed throughout the
174 genome. Two large families of repeated sequences, the dispersed repeated sequence

175 (DRS, including forward, reverse, complement, and palindromic sequences) and the
176 tandem repeated sequences (TRS, known as satellite DNA), can be readily
177 recognized based on their distribution pattern in the genome (Sperling & Li, 2013).
178 The satellite DNA refers to the repetitions of short sequences of the DNA and is of
179 three types: macrosatellites, minisatellites, and microsatellites (simple sequence
180 repeats or SSRs) (Hoy, 2013). The DRS in the plastomes of 18 *Diospyros* species
181 were predicted with REPuter (Kurtz et al., 2001), and the forward, reverse,
182 palindromic, and complementary repeat sequences were identified using the
183 following parameters: length of repeat unit ≥ 30 bp, sequence consistency ≥ 90%
184 (Hamming distance = 3). Meanwhile, the Tandem Repeats Finder (TRF) web server
185 (https://tandem.bu.edu/trf/trf.html) was used to search for TRS in the plastomes
186 using default settings (Benson, 1999), and the MISA software to identify SSRs
187 (Beier et al., 2017), with the minimum length of SSR fragment set to 10 bp and the
188 minimum repetition threshold values for mono-, di-, tri-, tetra-, penta-, and
189 hexanucleotide set to 10, 5, 4, 3, 3, and 3, respectively. Finally, all the detected
190 repeat sequences were manually checked and corrected to remove the redundant
191 ones.
192
193 **Analysis of codon usage**
194 Codon usage bias refers to the unequal usage of synonymous codons in genetic
195 material (Hershberg & Petrov, 2008; Guo et al., 2017; Plotkin & Kudla, 2011). For
196 codon usage analysis, protein-coding sequences longer than 300 bp with ATG as the

197 start codon were isolated from each plastome. CodonW
198 (http://codonw.sourceforge.net) analyzed the number and types of codons encoding
199 the proteins and calculated the effective number of codons (ENC), the relative
200 synonymous codon usage (RSCU), and the GC3 (Guanine and cytosine content at the
201 third codon position) values. Further, the effect of base composition on codon usage
202 bias was evaluated by ENC plotting, with ENC and GC3 values along the y-axis and
203 x-axis. The observed ENC value was compared with the expected ENC value using
204 the following equation (Wright, 1990):
205 $ENC = 2 + GC3s + 29/[GC3s^2 + (1 - GC3s)^2]$.
206 The effects of gene mutation and natural selection on codon usage bias were evaluated
207 by PR2 plotting with [A3/(A3 + T3)] and [G3/(G3 + C3)] along the y-xis and x-axis;
208 this plot reflects the potential biased usage of A/T and G/C in the third codon position.
209

210 **Analysis of genetic diversity and selective pressure**

211 The plastomes were aligned using the MUSCLE alignment software implemented in
212 Geneious to screen for the highly divergent regions among the 18 *Diospyros* species
213 (Edgar, 2004). The protein-coding genes, noncoding genes, and the intergenic
214 regions were extracted from the plastomes to analyze the nucleotide diversity (Pi)
215 among the *Diospyros* species using DnaSP (v5.0) (Librado & Rozas, 2009) based on
216 the number of overall mutation and the average nucleotide variation. Then, to
217 evaluate the effect of environmental pressure on the evolution of *Diospyros* species,
218 the Ka/Ks ratios of all the annotated protein-coding gene sequences in the plastomes

219 were calculated in Microsoft Excel. In general, the ratio of Ka/Ks < 1 (especially less
220 than 0.5) indicates purifying selection; Ka/Ks > 1 indicates probable positive
221 selection whereas Ka/Ks values close to 1 indicate neutral evolution, or relaxed
222 selection (Kimura, 1983).
223

224 **Phylogenomic inferences**

225 The plastomes of the 18 *Diospyros* species were further used for phylogenomic
226 analysis, with *Impatiens* (Balsaminaceae) and *Primula* (Primulaceae, the sister
227 family of Ebenaceae) as outgroups, to explore the evolutionary relationship among
228 the species. Maximum Likelihood (ML) and Bayesian Inference (BI) methods were
229 employed for the phylogenomic reconstruction of *Diospyros*. The best-fit nucleotide
230 substitution model for ML and BI analyses was determined by ModelTest (v3.7)
231 (Drummond et al., 2002), and the GTR + I + G model was finally selected for
232 phylogenomic analysis. ML and BI analyses were performed using the
233 RAxML-HPC (v8.1.11) (Stamatakis, 2014) and MrBayes (v3.2.3) (Ronquist, 2013)
234 online tools available from the CIPRES Science Gateway. The ML analysis was
235 conducted with 1000 bootstrap replicates using default settings. For BI analysis, four
236 parallel Markov Chains were run simultaneously to iterate 1,000,000 generations,
237 with the first 25% of samples discarded as burn-in. The phylogenetic trees were
238 sampled every 1000 generations to construct the final consensus tree.
239

240 **RESULTS**

241 **Genome structure and nucleotide variation**

242 The three newly generated *Diospyros* plastome sequences have been deposited in the
243 GenBank (OP480008, OP480009, OP485441) (Table 1). Similar to most angiosperm,
244 these three *Diospyros* species have plastomes with a classic tetrad structure, with two
245 inverted repeats (IR) separated by a large single copy (LSC) region and a small
246 single copy (SSC) region (Fig. 1). The plastome sequences of the *Diospyros* species
247 ranged from 157,321 bp to 157,934 bp, including IRs ranging from 25,873 bp to
248 26,120 bp, SSC from 18,174 bp to 18,560 bp, and LSC from 86,874 bp to 87,246 bp
249 (Table 2). A total of 134–138 genes, including 89–91 protein-coding genes, 1–2
250 pseudogenes, 37 tRNA genes, and 8 rRNA genes were identified in these species,
251 among which 10 protein-coding genes, 7 tRNA genes, and 4 rRNA genes were
252 repeated in the two IRs (Table 2, Table S1). Among the protein-coding genes, the
253 *ycf*15 had only two copies in the IR in *D. eriantha* and *D. strigosa* and four in the
254 other *Diospyros* species. The *ycf*1 in the IRb of all *Diospyros* species (a short Ψ*ycf*1)
255 and the *rps*19 in the IRa region in most *Diospyros* species (a short Ψ*rps*19) were
256 identified as pseudogenes (Table 2, Table S1). Six tRNAs and nine kinds of
257 protein-coding genes had one intron, while the *clp*P, *ycf*3, and *rps*12 genes had two
258 (Table S1). The *mat*K gene was found embedded in the intronic region of *trn*K-UUU,
259 consistent with various other plant taxa. Meanwhile, the trans-spliced *rps*12 gene,
260 with the 5' and 3' ends located in the LSC and IR, had two independent transcription
261 units.

262 The overall GC content of *Diospyros* species was 37.4%, while that of the

263 coding sequences (CDS) was 37.7% (Table 2). For all the species, the GC content of
264 IR (43.0%–43.1%) was higher than those of the LSC (35.3%–35.4%) and SSC
265 (30.7%–30.9%) regions.

266 Multiple plastome comparisons among the *Diospyros* species using mVISTA
267 and Mauve alignment showed a high degree of collinearity. The gene organization
268 and distribution patterns in the plastome were highly consistent among the
269 *Diospyros* species (Fig. S1). No rearrangement of DNA fragments, including
270 inversion or translocation, was detected among *Diospyros* plastomes sequences (Fig.
271 S2). However, slight differences were observed in different regions throughout the
272 plastome sequence. The sequence similarity among *Diospyros* plastomes sequences
273 was much higher in the two IRs, especially the rRNA coding regions. By contrast,
274 the nucleotide mutation rate was high in the noncoding regions, especially the
275 intergenic spacer (IGS) regions (Figs. S1–2).

276 Contraction and expansion of IR indicate plastome evolution and are correlated
277 with plastome size. The present study found conserved plastome structure in terms
278 of the length of IRs and gene location at the IR/SSC/LSC boundaries among the 18
279 *Diospyros* species (Fig. 2). In all the species, the *rpl*2 and *trn*H genes were located
280 on different sides of the IRa/LSC boundary. The *ycf*1 gene spanned the SSC/IRa
281 boundary with a part of the gene extended to the IRa, forming a pseudogene (Ψ*ycf*1)
282 at the corresponding position near the IRb/SSC boundary. Extension of the short
283 Ψ*ycf*1 fragment into the SSC region was observed in all *Diospyros* species, and an
284 extension of a short portion of *ndh*F into the IRb was observed in *D. cathayensis* and

285 *D. rhombifolia*. The analysis also detected Ψ*ycf*1 and *ndh*F overlap in all species
286 except *D. glaucifolia*, *D. strigosa*, and *D. jinzaoshi*. The *rps*19 gene spanned the
287 LSC/IRb region in all the species except *D. glaucifolia*, *D. kaki*, and *D. oleifera*, in
288 which the gene was found 2, 13, and 8 bp away from the LSC/IRb junction. In
289 addition, *rps*19 formed a pseudogene (Ψ*rps*19) in all the species except *D.*
290 *glaucifolia*, *D. kaki*, and *D. oleifera*, where the gene was at the IRa/LSC boundary
291 (Fig. 2).
292
293 **Repetitive sequences in plastomes**
294 REPuter identified 1204 repeated sequences, including 18–28 forward repeats,
295 19–35 palindromic repeats, and 20–34 tandem repeats, in the 18 *Diospyros* species
296 (Table S3–4, Fig. 3). However, no reverse complementary sequences were detected
297 in the *Diospyros* plastomes. Among the species, *D. eriantha* had the maximum (93)
298 forward, palindromic, and tandem repeats. Tandem repeats were more prevalent and
299 accounted for 36.46% of all the repeat types. On the contrary, forward repeats were
300 relatively rare and accounted for only 30.07% of the repeat types (Table S4). The
301 length of the dispersed repeats, including forward and palindromic repeats, varied
302 from 30 bp to 90 bp, while more than half of the tandem repeats were 18 bp to 30 bp
303 long (Table S3). The longest tandem repeats were detected in *D. kaki* (43 bp) and *D.*
304 *blancoi* (58 bp) and were located in the IGS of *ndh*H and *rps*15, respectively (Table
305 S3).

306 Additionally, 991 SSR loci were detected from the 18 *Diospyros* plastomes. The
307 number of SSR loci in each species varied from 37 (*D. rhombifolia*) to 69 (*D.*
308 *glaucifolia*) (Table S4, Fig. 3). Most identified SSRs were mononucleotide repeats
309 (79.11%), followed by tetra- (10.90%), di- (5.65%), and trinucleotide (3.94%)
310 repeats (Table S4, Fig. 3). Four pentanucleotide repeats were detected in 4 (*D.*
311 *blancoi*, *D. cathayensis*, *D. eriantha*, and *D. strigosa*) of the 18 species, while no
312 hexanucleotide repeats were detected in the genus. Most SSRs (78.24%) were found
313 in the LSC region of the plastome, and only 18.27% and 3.49% were found in the
314 SSC and IR regions, respectively (Table S3–4, Fig. 4). In addition, 19.65% of the
315 SSRs were found in the CDS, while the other 80.35% were found in the introns and
316 IGS (Table S3–S4, Fig. 4).
317

318 **Nucleotide diversity of plastomes**

319 The alignment of the plastomes discovered five hypervariable regions with a Pi
320 higher than 0.03 (*trn*H-*psb*A, *rps*16-*trn*Q, *trn*T-*psb*D, *pet*A-*psb*J, *trn*L-*trn*F-*ndh*J)
321 among the 18 *Diospyros* species (Table S5, Fig. 5). Analysis of the CDS and their
322 nucleotide polymorphisms among the plastomes of the 18 species identified *rpl*33,
323 *psb*T, *rpl*22, *psb*C, and *ycf*1 as the genes with the highest nucleotide polymorphism
324 (Pi > 0.012, Fig. 5). Meanwhile, most nucleotide mutations were detected in the LSC
325 and SSC regions. The nucleotide diversity values (Pi) of the LSC and SSC regions
326 were 0–0.04 and 0–0.03, respectively, while that of the IR was 0–0.01 (Table S5, Fig.
327 5).

328 Further analysis revealed high variability in the gene spacer, with a Pi value
329 significantly higher than that of the gene-coding region (CDS) (Fig. 5). These
330 findings suggest that hypervariable DNA fragments between the different *Diospyros*
331 species could be used as ptDNA barcodes for taxonomic classification, species
332 discrimination, and phylogenetic reconstruction and inference.
333

334 **Phylogenetic inference**

335 Phylogenetic analysis based on complete plastome sequences revealed a close
336 relationship between *D. eriantha* and *D. strigose*. Meanwhile, *D. morrisiana* was
337 found clustered with *D. glaucifolia* and *D. lotus* (Fig. 6). *Diospyros kaki*, *D. oleifera*,
338 and the two cultivated species *D. deyangensis* and *D. jinzaoshi* formed a clade.
339 Notably, *Diospyros* species living in similar habitats clustered together in the
340 phylogenetic tree, and the five island species formed a clade at the base of the genus.
341 All the deciduous species formed a sister clade to the clade of four evergreen species.
342 However, the evergreen species, *D. blancoi*, was relatively isolated and created a
343 single lineage; it was identified as a sister to all other deciduous and evergreen
344 species (Fig. 6).
345

346 **Selective pressure in CDS genes**

347 Then, to evaluate the evolutionary forces acting on the protein-coding homologous
348 genes in the 18 *Diospyros* species, the Ka/Ks values of CDS were calculated (Table
349 S6). Our results showed a Ka/Ks value of less than 1 for most genes, indicating that

350 most homologous genes were under purifying selection. However, the Ka/Ks values
351 of *rps*16 and *ycf*3 in all species were more than 1, suggesting that these genes were
352 under positive selection in the *Diospyros* species. Additionally, *ndh*G in island
353 species, *rpo*C1 in deciduous species, and *ycf*3 in deciduous and evergreen species
354 were also under positive selection (Fig. 7A, Table S6a). Furthermore, to examine the
355 selective pressure on plastid genes with different functions, the CDS were classified
356 into photosynthesis-related, self-replication-related, and other functional genes
357 (Table S6). For species in the evergreen, deciduous, and island groups, the Ka/Ks
358 values of photosynthesis-related and self-replication-related genes were significantly
359 lower than the other genes (Fig. 7B, Table S6b). The Ka/Ks values of
360 photosynthesis-related and self-replication-related genes were extremely low in
361 species from the island group, suggesting strong purifying selection (Fig. 7B, Table
362 S6b). Meanwhile, the Ka/Ks values of both photosynthesis-related and
363 self-replication-related genes in the evergreen species were significantly higher than
364 their homologs in deciduous and island species (Fig. 7C, Table S6c).
365

366 **Codon usage bias**

367 The comparison of the occurrence frequencies of different codons in the 18
368 *Diospyros* plastomes identified leucine (Leu) as the most used amino acid (10.35%),
369 and its encoding codon UUA with a maximum RSCU value of 1.94 accounted for
370 3.35% of all the codons (Table S7). On the contrary, cysteine (Cys) was the least
371 used amino acid (1.05%), but serine (Ser) encoding codon AGC had a minimum

372 RSCU value of 0.33 (Table S5). In addition, AUG and UGG encoding methionine
373 (Met) and tryptophan (Trp) had an RSCU value of 1, indicating no bias in the codon
374 usage for these two amino acids (Table S7). Moreover, 30 codons had an RSCU >1,
375 of which 16 had U in its third position, 12 had A, and one had G, which indicates
376 that the codons ending with U or A are preferred in the *Diospyros* plastomes (Table
377 S7).

378 Further, the ENC-GC3 plot was obtained by taking the ENC value of each gene
379 as the ordinate and the GC3 value as the abscissa to explore the kind of suffered
380 stress (mutation pressure or natural selection) (Fig. 8). The ENC value ranged from
381 32.36 to 59.25 and the GC3 value from 0.143 to 0.346 (Table S8). Figure 8A shows
382 that most genes are close to the standard curve, and a few are far below it, indicating
383 the influence of mutation pressure and natural selection on the codon usage bias of
384 *Diospyros* genes. Then, to accurately evaluate the difference between the observed
385 value (ENC_{obs}) and the expected value (ENC_{exp}) of ENC, the
386 (ENC_{exp}-ENC_{obs})/ENC_{exp} ratio was calculated (Table S6). The ENC frequency
387 ranging from -0.1 to 0.1 indicated a slight difference between ENC_{exp} and ENC_{obs}
388 values of most genes. The difference values in the codon usage bias of *Diospyros*
389 genes was related to the difference in GC3, indicating a significant influence of
390 mutation pressure on codon usage bias.

391 Detailed analysis showed considerable deviation in the observed ENC values
392 from the standard curve for eight genes (*rps*18, *rps*14, *psb*A, *rpl*16, *rps*8, *psb*D, *ycf*3,
393 and *clp*P) of all the species (Fig. 8A). Then, to explore the potential differences in

394 the main driving force of codon usage bias in *Diospyros* species with different leaf
395 habits and living habitats, all the 18 *Diospyros* species were divided into three
396 groups: evergreen, deciduous, and island species. Genes from these three groups are
397 presented using different colors in the ENC and PR2 (parity rule 2) plots. Among all
398 the genes, *ycf*3 from the island group showed the highest ENC value, while *rps*18
399 from the deciduous and evergreen groups had the lowest (Table S8; Fig. 8). PR2 plot
400 showed slight disequilibrium in A/T and G/C usage in the third codon position of
401 CDS of the 18 *Diospyros* plastomes (Fig. 8C). More genes were distributed in the
402 quadrant IV (at the right bottom of the Fig. 8C) than the other three quadrants,
403 indicating frequent use of G and T in the third codon position. This observation
404 suggests that the existing codon usage pattern may be due to the combined action of
405 natural selection and mutation.
406

407 **DISCUSSION**

408 **Phylogenetic relationships of *Diospyros* species**

409 Recently, researchers have discussed using plastomes as super-barcodes
410 for plant species identification (Hernandez-Leon et al., 2013). The
411 phylogenetic analysis of this study showed that the plastomes are helpful
412 as a super-barcode for *Diospyros* species identification (Fig. 6). Breeding,
413 intensive management, and germplasm conservation in *Diospyros* demand
414 an understanding of the genetic relationship of the taxa. The present study
415 found a topology of *Diospyros* consistent with earlier research which also

416 reported based on plastome itself (Li et al., 2018). We carried out the
417 phylogenetic analysis using more samples and thus revealed reliable
418 results with greater precision. Notably, species clustering was based on
419 leaf habits (Fig. 6). The island species formed a monophyletic clade at the
420 basal portion of the tree and was a sister to the monophyletic clade of the
421 deciduous and evergreen species. Except for the evergreen species *D.*
422 *blancoi*, eight the deciduous species and four of the evergreen species
423 formed two sister clades. The plastome-based evidence obtained in this
424 study for the deciduous clade supports the previous phylogenetic analysis
425 demanding the upgradation of *D. deyangensis* and *D. jinzaoshi* to species
426 rank based on morphological, molecular, and chromosomal features
427 (number). In the plastome-based tree, *D. kaki*, the dioecious *D.*
428 *deyangensis*, and the polygamous *D. oleifera* shared a common furcation.
429 Meanwhile, *D. glaucifolia* and *D. lotus* were genetically close to *D.*
430 *morrisiana*, identical to the classification based on phenotypic
431 characteristics (Lee et al., 1996), which is similar to Tang et al. (2014). In
432 addition to the similar phylogenetic relationships among the three species,
433 *Diospyros morrisiana* has relatively smaller leaves and fruits than *D.*
434 *glaucifolia* and *D. lotus* (Lee et al., 1996). Meanwhile, *Diospyros*
435 *virginiana* was identified as the basal taxa of the deciduous clade. The
436 fruits of *D. virginiana* are an important food for wildlife, native people,
437 and Euro-American colonists. These fruits have never been

438 commercialized, despite the selection of superior clones over the years
439 (Boufford, 2022). Therefore, *D. virginiana*, as the base group of
440 deciduous group and its wild existence, can be used as a species for
441 cultivation and breeding. In the evergreen clade, *D. blancoi* appeared
442 relatively isolated and formed a paraphyletic group with the remaining evergreen
443 species. *Diospyros blancoi* is located at the base of the whole deciduous and
444 evergreen groups and has extensive application value (e.g. strong heartwood and
445 fruit as medicine), which is of research significance (Howlader et al., 2012;
446 Krisdianto, 2005). Meanwhile, *Diospyros eriantha* and *D. strigosa*
447 clustered together based on plastomes sequences, consistent with the
448 similarities in the morphological characteristics. *Diospyros rhombifolia*
449 and *D. cathayensis* clustered together and formed sister to the
450 monophyletic clade of *D. eriantha* and *D. strigosa*. For the island clade
451 included the *D. ferrea* complex, which has trimerous flowers with a
452 trilocular ovary (biovulate) and is found throughout the Old World tropics
453 (Lee et al., 1996). Elucidating the boundaries between the different
454 *Diospyros* species would improve our understanding of the cultivated
455 species' origin, phylogeny, and taxonomy and help decide the breeding
456 strategy. The phylogenetic results of this study are generally consistent
457 with previous studies. This study further found that *Diospyros* species are
458 clustered into the three groups (evergreen, deciduous, and island groups).
459

460 **Adaptive evolution of *Diospyros* plastomes**

461 We found that the Ka/Ks values of 79 common genes among the species
462 were less than 1. We also found that the Ka/Ks values of photosynthesis-related
463 and self-replication-related genes were significantly lower than other genes in the
464 evergreen, deciduous, and island groups (Fig. 7). This observation indicated that
465 most important photosynthesis-related and self-replication-related genes are
466 undergoing strong purifying selection. Purifying selection usually reduces
467 genetic diversity and maintain gene homozygosity via the selective
468 removal of deleterious alleles (Cvijović et al., 2018). In addition, the
469 functional importance of a protein determines its evolutionary rate (Wang
470 et al., 2011). Our study found that the Ka/Ks values of photosynthesis-related
471 and self-replication-related genes were extremely low in species from the island
472 group, indicating these species suffered more strong purifying selection than those
473 in other leaf habits. This indicated that the purifying selection of these two type
474 genes of island species is more intense than evergreen and deciduous species.
475 Meanwhile, evergreen species, primarily distributed in the tropics, have
476 undergone less purification. In addition, Ka/Ks pairwise calculation
477 detected a positive gene selection signal based on the values of *ndh*G in
478 island species, *rpo*C1 in deciduous species, and *ycf*3 in both deciduous and
479 evergreen species. These results indicate that the plastid genes are likely to
480 be involved in the adaptation to latitude or precipitation. However, A
481 small portion of total DNA represented by organelle genomes, such as

482 plastomes, cannot fully display a large number of selected sites. Therefore,
483 a nuclear, genome-wide transcriptome approach is necessary to confirm
484 the selection pressure on *Diospyros* species for future research.

485 Typically, the usage pattern of the third base of the codon is closely
486 related to codon usage bias (Gao et al., 2022). The GC composition drives
487 codon and amino acid usage, and the GC content of the third base of a
488 codon (GC3) reflects codon usage patterns (Chen et al., 2013). Previous
489 studies have shown that dicots and monocots use A/U and C/G as ending
490 codons, respectively (Yao et al., 2008; Liu et al., 2020). Our study found
491 that the average GC content and GC3 values of *Diospyros* codons were
492 37.6%–37.7% and 14.3%–34.6%, respectively, indicating that the
493 *Diospyros* codons also preferred A/T(U) in the third position, consistent
494 with the RSCU values of *Diospyros* genes.

495 Mutation pressure and natural selection are the major factors
496 influencing codon usage bias in any organism (Sharp et al., 2010; Rao et al.,
497 2011). However, the main factors affecting codon usage bias vary
498 significantly among species. According to the parity rule 2 analysis, the
499 GT content at the third position of a codon is higher than AC content.
500 However, A and T were used more frequently than G and C in the third
501 position of the codons of *Diospyros* genes, which suggested natural
502 selection as one of the main reasons for *Diospyros* codon usage bias.
503 Further ENC-plot analysis showed that the ENC value of most genes was

504 close to the expected value, suggesting that the codon usage bias of these
505 genes was related to GC3, and mutation was the main factor influencing.
506 Additionally, a few genes in the plot (*rps*18 and *rps*14) were well below the
507 expected curve, indicating the influence of natural selection on the codon
508 deviations of these genes. Integrated analysis of the ENC-plot and PR2
509 plot revealed that mutation and natural selection jointly affected the
510 codon usage bias of *Diospyros* genes, and mutation pressure played a
511 significant role, consistent with the reports on CDS in *Oncidium* (Xu et al.,
512 2011) and the findings in Rosaceae (Liu et al., 2021). Moreover, studies
513 in *Drynaria* also indicated mutation pressure as the driving force of codon
514 usage bias (Shen et al., 2021). However, Li *et al*. (2022) reported natural
515 selection as the main factor influencing codon usage bias of *Pinus densata*
516 plastome genes. These results suggest that various pressures influence
517 plastomes, and codon usage preferences of plastome genes vary among the
518 dicotyledon taxa.
519

520 **Potential ptDNA barcodes of *Diospyros***

521 Taxonomic classification is challenging in *Diospyros* (Lee et al., 1996). Moreover,
522 the worldwide distribution and phenotypic plasticity make it difficult to identify the
523 wild *Diospyros* species (Ebenaceae) (Lin et al., 2020). Generally, in such cases
524 barcodes are used. However, only a limited number of DNA barcodes (e.g.,
525 *rbc*L, *mat*K, and *trn*H-*psb*A) are available to resolve the phylogenetic

526 relationships among the groups (Duangjai et al. 2009; Linan et al., 2019).
527 Therefore, comparing more plastomes for developing variable DNA
528 barcodes is important for *Diospyros* species. Generally, the mutational
529 hotspots have the potential to resolve taxonomic issues. They provide
530 adequate genetic information for species identification and, therefore, can
531 be used to develop novel DNA barcodes. The five potential mutational
532 hotspots (*trn*H-*psb*A, *rps*16-*trn*Q, *trn*T-*psb*D, *pet*A-*psb*J, *trn*L-*trn*F-*ndh*J)
533 identified in this study could be suitable barcodes for *Diospyros*
534 classification. In addition, five other potential mutational hotspots (*rpl*33,
535 *psb*T, *rpl*22, *psb*C, and *ycf*1) were identified with high nucleotide
536 polymorphisms in CDS. By comparison, in a previous study on *Diospyros*,
537 eight potential mutational hotspots (*trn*H-*psb*A, *rps*16-*trn*Q, *rpo*B-*trn*C,
538 *rps*4-*trn*T-*trn*L, *ndh*F, *ndh*F-*rpl*32-*trn*L, *ycf*1a, and *ycf*1b) showed high
539 divergence in plastomes and were recommended as core DNA barcodes
540 (Li et al., 2018). Of these, *ycf*1 has been widely applied in plant
541 phylogeny and DNA barcoding studies (Parks et al., 2011; Yang et al.,
542 2017; Dastpak et al., 2018). *Trn*H-*psb*A, *trn*L-*trn*F-*ndh*J, *pet*A-*psb*J and
543 *rps*16-*trn*Q have also been used for phylogenetic studies (Shaw et al.,
544 2005; Shaw et al., 2007). Meanwhile, *Trn*T-*psb*D, *rpl*33, *psb*T, *rpl*22, and *psb*C
545 are novel hotspots identified as potential barcodes in this study.

546

547 **CONCLUSION**

548 The present study analyzed the plastome sequences of 18 *Diospyros*
549 species and performed phylogenetic analysis to provide valuable genetic
550 information. The findings based on this analysis partially supported the
551 previous classifications based on morphological features. In addition, the
552 study offers new insights into the phylogenetic relationships between the
553 species of the three groups (evergreen, deciduous, and island groups).
554 Comparative plastome analysis revealed conserved genome structures and
555 low nucleotide polymorphism. The study also identified mutational
556 hotspots as phylogenetically informative markers that will contribute to
557 future studies on *Diospyros* systematics and species identification. The
558 study also assessed the adaptive evolution of the three groups (major
559 lineages) in *Diospyros* for the first time using Ka/Ks, ENC-plot, and PR2
560 plot. This integrated analysis revealed natural selection and mutation
561 pressure as the driving forces of *Diospyros*' evolution. In this study,
562 plastomes of Diospyros provided adequate genetic information for
563 understanding adaptive evolution. Thus, our results provide a framework
564 for further studies on the systematics and ecology of *Diospyros*, including
565 a formal, subgeneric classification. However, we should focus on a
566 comprehensive molecular sampling of all species in future research.
567

568 ACKNOWLEDGMENT

The authors thank Dr. Zu-Lin Ning from Horticultural Center, South China National Botanical Garden in Guangzhou, for the help in the material collection.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant No.31800309) to Y.H. Zhang, the Research Funds for Zhejiang Provincial Public Welfare Technology and Application Research Project (Grant No. LGN21C020007) to Q. Ma and the Scientific Research Project of Baishanzu National Park (Grant No. 2021KFLY06) to Y.H. Zhang.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Y. H. Zhang and X. J. Jin conceived and designed the study; J. Sun, Y. Huang, and C. J. Lai performed the experiments and data analysis; Y. H. Zhang contributed to material collection; Q. Ma, X. J. Jin, and J. Sun wrote the manuscript; P. Li, Q. Ma, and Y. H. Zhang edited the manuscript. All authors have approved the final manuscript.

DATA AVAILABILITY STATEMENT

The *Diospyros* plastomes generated in this study are available in the NCBI GenBank repository (details in Table 2).

REFERENCES

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics, 33(16): 2583–2585.

https://doi.org/10.1093/bioinformatics/xxxxx

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research, 27(2): 573–580.

Boufford, D. E. F. Ebenaceae In: Flora of North America North of Mexico [Online], Flora of North America Editorial Committee, Ed., New York and Oxford, Vol. 8, http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=242416451. Accessed [August 23, 2022].

CBOL Plant Working Group. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America, 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., Wu, X. H., Huang, X., Gu, X. B., Wang, S. X., Peng, X. R., & Yang, G. Y. (2013). Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. Biochem. Biophys. Biochemical and Biophysical Research Communications, 430: 1344–1348. https://doi.org/10.1016/j.bbrc.2012.12.078

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. Genetics, 209(4): 1235–1278. https://doi.org/10.1534/genetics.118.301058

Darling, C. E. A., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research, 14: 1394–1403. https://doi.org/10.1101/gr.2289704

Dastpak, A., Osaloo, S. K., Maassoumi, A. A., & Safar, K. N. (2018). Molecular phylogeny of *Astragalus* sect. *Ammodendron* (Fabaceae) inferred from chloroplast *ycf*1 gene. Annales Botanici Fennici, 55: 75–82. https://www.jstor.org/stable/26789423

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics, 161(3): 1307–1320. https://doi.org/10.1093/genetics/161.3.1307

Duangjai, S., Wallnöfer, B., Samuel, R., Munzinger, J., & Chase, M. W. (2006). Generic delimitation and relationships in Ebenaceae sensu Lato: evidence from six plastid DNA regions. American Journal of Botany, 93(12): 1808–1827. https://doi.org/10.3732/ajb.93.12.1808

Duangjai, S., Samuel, R., Munzinger, J., Forest, F., Wallnöfer, B., Barfuss, M. H., & Chase, M. W. (2009). A multi-locus plastid phylogenetic analysis of the pantropical genus *Diospyros* (Ebenaceae), with an emphasis on the radiation and biogeographic origins of the New Caledonian endemic species. Molecular Phylogenetics and Evolution, 52(3): 602-620. https://doi.org/10.1016/j.ympev.2009.04.021

Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fernández-Mazuecos, M., Vargas, P., McCauley, R. A., Monjas, D., Otero, A., Chaves, J. A., ... & Rivas-Torres, G. (2020). The radiation of Darwin's giant daisies in the Galápagos Islands. Current Biology, 30(24): 4989-4998. https://doi.org/10.1016/j.cub.2020.09.019

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

Fu, J. M., Liu, H. M., Hu, J. J., Liang, Y. Q., Liang, J. J., Wuyun, T., & Tan, X. F.

642 (2016). Five complete chloroplast genome sequences from *Diospyros*: genome
643 organization and comparative analysis. PLoS One, 11(7):
644 https://doi.org/e0159566. 10.1371/journal.pone.0159566
645 Gao, Y., Lu, Y., Song, Y., & Jing, L. (2022). Analysis of codon usage bias of WRKY
646 transcription factors in *Helianthus annuus*. BMC Genomic Data, 23: 46.
647 https://doi.org/10.1186/s12863-022-01064-8
648 Guo, Y., Liu, J., Zhang, J., Liu, S., & Du, J. (2017). Selective modes determine
649 evolutionary rates, gene compactness and expression patterns in Brassica. Plant
650 Journal, 91, 34–44. https://doi.org/10.1111/tpj.13541
651 Hernandez-Leon, S., Gernandt, D. S., Perez de la Rosa, J. A., & Jardon-Barbolla, L.
652 (2013). Phylogenetic relationships and species delimitation in *Pinus* section
653 Trifoliae inferred from plastid DNA. PLoS One, 8: e70501.
654 https://doi.org/10.1371/journal.pone.0070501
655 Hershberg, R., & Petrov, D. A. (2008). Selection on codon bias. Annual Review of
656 Genetics, 42, 287–299. https://doi.org/10.1146/annurev.genet.42.110807.091442
657 Hoy, M. A. (2013). *Chapter 12 - Molecular Systematics and the Evolution of*
658 *Arthropods*. In M. A. Hoy (Eds), *Insect Molecular Genetics (Third Edition).*
659 Pittsburgh, Academic Press, pp. 521–589.
660 Howlader, M. S. I., Rahman, M. M., Kalhipa, A. B. R., Ahmed, F., & Rahman, M. M.
661 (2012). Antioxidant and Antidiarrhoeal Potentiality of *Diospyros blancoi*.
662 International Journal of Pharmacology, 8(5): 403–409.
663 https://doi.org/10.3923/ijp.2012.403.409
664 Jansen, R. K., Cai, Z. Q., Raubeson, L. A., Danielle, H., dePamphilis, C. W.,
665 Leebens-Mack, J., Müller, K. F., Guisinger-Bellian, M., Haberle, R. C., Hansen,
666 A. K., Chumley, T. W., Lee, S. B., Peery, R., McNeal, J. R., Kuehl, J. V., &

Boore J. L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America, 104(49): 19369–19374. https://doi.org/10.1073/pnas.07091211

Jiang, B., Gao, L., Li, J., Zhou, Y., Su, Y. J., & Wang, T. (2014). Adaptive evolution of the chloroplast genome in AA-genome *Oryza* species. Science China Press, 59(20): 1975–1983. https://doi.org/10.1360/N972014-00127

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology, 21: 241. https://doi.org/10.1186/s13059-020-02154-5

Kapralov, M. V., & Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology, 7: 73. https://doi.org/10.1186/1471-2148-7-73

Kimura, M. (1983). The neutral theory of molecular evolution and the world view of the neutralists. Genome, 31, 24–31. https://doi.org/10.1266/jjg.66.367

Krisdianto A. (2005). Anatomical and physical properties of Bisbul wood (*Diospyros blancoi* A.DC.). Indonesian Journal of Forestry Research, 2(1): 57–67. https://doi.org/10.20886/ijfr.2005.2.1.57-67

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research, 29(22): 4633–4642. https://doi.org/10.1093/nar/29.22.4633

Lee, S., Michael, G. G., & Frank, W. (1996). *Diospyros.* In Z. Y. Wu, P. H. Raven & D. Y. Hong (Eds.), *Flora of China* (Vol. 15). Beijing/Missouri/St. Louis, Science

Press/Botanical Garden Press, pp. 215–234.

Li, W. Q., Liu, Y. L., Yang, Y., Xie, X. M., Lu, Y. Z., Yang, Z. R., Jin, X. B., Dong, W. P., & Suo, Z. L. (2018). Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*. BMC Plant Biology, 18: 210.

Li, J. F., Li, Y. Q., Tang, J. R., Chen, S., Chen, L., Cai, N. H., Xu, Y. L., & Tang, H. Y. (2022). Comparison of codon preference patterns in the chloroplast genome of *Pinus densata*. J. Biol., Online. https://doi.org/10.7717/peerj.12173

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25(11): 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Linan, A. G., Schatz, G. E., Lowry II, P. P., Miller, A., & Edwards, C. E. (2019). Ebony and the Mascarenes: the evolutionary relationships and biogeography of *Diospyros* (Ebenaceae) in the western Indian Ocean. Botanical Journal of the Linnean Society, 190: 359–373. https://doi.org/10.1093/botlinnean/boz034

Lin, J., Wang, S. Y., Gong, X., & Zhao, Y. J. (2020). Research progress in identification of persimmon germplasm resources and phylogenetic relationship of *Diospyros* plants. Chinese Journal of Tropical Agriculture, 40(10), 83–89. https://doi.org/10.12008/j.issn.1009-2196.2020.10.014

Liu, H. B., Lu, Y. Z., Lan, B. L., & Xu, J. C. (2020). Codon usage by chloroplast gene is bias in *Hemiptelea davidii*. Journal of Genetics, 99: 8. https://doi.org/10.1007/s11033-009-9521-7

Liu, X. Y., He Z. J., & Qiu Y. M. (2021). Codon bias in the chloroplast genome of four Rosaceae fruit trees. Molecular Plant Breeding, Online. https://kns.cnki.net/kcms/detail/46.1068.s.20210204.1406.012.html

Lohse, M., Drechsel, O., & Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW):

717 a tool for the easy generation of high-quality custom graphical maps of plastid
718 and mitochondrial genomes. Current Genetics, 52: 267–274.
719 https://doi.org/10.1007/s00294-007-0161-y
720 Luo, Z. R., Zhang, Q. L., Xu, L. Q., & Guo, D. Y. (2021). Advances in genetic
721 diversity and breeding of persimmon. Deciduous Fruits, 53(3): 01–05.
722 https://doi.org/10.13855/j.cnki.lygs.2021.03.001
723 Mohanta, T. K., Mishra, A. K., Khan, A., Hashem, A., Abd Allah, E. F., & Al-Harrasi,
724 A. (2020). Gene loss and evolution of the plastome. Genes, 11(10): 1133.
725 https://doi.org/10.1101/676304
726 Morgan, T. H. (1925). *Evolution and Genetics (Second Edition).* Princeton, Princeton
727 University Press, pp. 330.
728 Parks, M., Liston, A., & Cronn, R. (2011). Newly developed primers for complete
729 *ycf*1 amplification in *Pinus* (Pinaceae) chloroplasts with possible family-wide
730 utility. American Journal of Botany, 98: e185–188.
731 https://doi.org/10.3732/ajb.1100088
732 Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and
733 consequences of codon bias. Nature Reviews Genetics, 12, 32–42.
734 https://doi.org/10.1038/nrg2899
735 Rao, Y., Wu, G., Wang, Z., Chai, X., Nie, Q., & Zhang, X. (2011). Mutation bias is the
736 driving force of codon usage in the Gallus gallus genome. DNA Research, 18(6):
737 499–512. https://doi.org/10.1093/dnares/dsr035
738 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S.,
739 Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2013). MrBayes 3.2:
740 efficient Bayesian phylogenetic inference and model choice across a large model
741 space. Systematic Biology, 61(3): 539–542.

https://doi.org/10.1093/sysbio/sys029

Samuel, R., Turner, B., Duangjai, S., Munzinger, J., Paun, O., Barfuss, M. H. J., & Chase, M. W. (2019). Systematics and evolution of the Old World Ebenaceae, a review with emphasis on the large genus *Diospyros* and its radiation in New Caledonia. Botanical Journal of the Linnean Society, 189: 99–114. https://doi.org/10.1093/botlinnean/boy081

Sharp, P. M., Emery, L. R., & Zeng, K. (2010). Forces that infuence the evolution of codon bias. Philosophical Transactions of The Royal Society B Biological Sciences, 365: 1203–1212. https://doi.org/10.1098/rstb.2009.0305

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E., & Small, R. L. (2005). The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany, 92: 142–166. https://doi.org/10.3732/ajb.92.1.142

Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany, 94: 275–288. https://doi.org/10.3732/ajb.94.3.275

Shen, Z. F., Lu, T. Q., Zhang, Z. R., Cai, C. T., & Tian, B. (2021). Condon preference of chloroplast genome of *Drynaria*. Guihaia, 41(2): 266–273. https://doi.org/10.11931/guihaia.gxzw201904013

Sperling, A. K., & Li, R. W. (2013). *Repetitive Sequences.* In S. Maloy, & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)*, (Vol. 6). Pittsburgh, Academic Press, pp. 150–154.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and

post-analysis of large phylogenies. Bioinformatics, 30(9): 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Tang, D. L., Hu, Y., Zhang, Q. L., Yang, Y., & Luo, Z. R. (2014). Discriminant analysis of "Jinzaoshi" from persimmon (Diospyros kaki Thunb.; Ebenaceae): A comparative study conducted based on morphological as well as ITS and matK sequence analyses. Scientia Horticulturae, 168: 168–174. http://dx.doi.org/10.1016/j.scienta.2014.01.033

The plant list (TPL). Specialized Information Services: Enter a Genus (*Diospyros*) in search. http://www.theplantlist.org/tpl1.1/search?q=*Diospyros* (Accessed August 23, 2022).

Turner, B., Paun, O., Munzinger, J., Chase, M. W., & Samuel, R. (2016). Sequencing of whole plastid genomes and nuclear ribosomal DNA of *Diospyros* species (Ebenaceae) endemic to New Caledonia: many species, little divergence. Annals of Botany, 117(7): 1175-1185. https://doi.org/10.1093/aob/mcw060

Tomlinson, K. W., Poorter, L., Sterck, F. J., Borghetti, F., Ward, D., de Bie, S., & van Langevelde, F. (2013). Leaf adaptations of evergreen and deciduous trees of semi-arid and humid savannas on three continents. Journal of Ecology, 101: 430–440. https://doi.org/10.1111/1365-2745.12056

Wallnöfer, B. The biology and systematics of Ebenaceae: a review. (2001). Annalen des Naturhistorischen Museums in Wien, 103B: 485–512.

Wang, D. P., Liu, F., Wang, L., Huang, S., & Yu, J. (2011). Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. Biology Direct, 6: 13. https://doi.org/10.1186/1745-6150-6-13

White, F. (1956). Distribution of the African species of *Diospyros*. Webbia: Journal of

Plant Taxonomy and Geography, 11(1): 525–540. https://doi.org/10.1080/00837792.1956.10669649

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Molecular Biology, 76: 273–297. https://doi.org/10.1007/s11103-011-9762-4

Wright, F. (1990). The 'effective number of codons' used in a gene. Gene, 87(1): 23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Xia, M. Q., Liao, R. Y., Zhou, J. T., Lin, H. Y., Li, J. H., Li, P., Fu, C. X., & Qiu, Y. X. (2022a). Phylogenomics and biogeography of Wisteria: Implications on plastome evolution among inverted repeat-lacking clade (IRLC) legumes. Journal of Systematics and Evolution, 60(2): 253–265. https://doi.org/10.1111/jse.12733

Xia, M. Q., Liu, Y., Liu, J. J., Chen, D. H., Shi, Y., Chen, Z. X., Chen, D. R., Jin, R. F., Chen, H. L., Zhu, S. S., Li, P., Si, J. P., & Qiu, Y. X. (2022b). Out of the Himalaya-Hengduan Mountains: Phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere. Molecular Phylogenetics and Evolution, 169(2022): 107431. https://doi.org/10.1016/j.ympev.2022.107431

Xu, C., Cai, X. N., Chen, Q. Z., Zhou, H. X., Cai, Y., & Ben, A. I. (2011). Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium* gower ramsey. Evolutionary Bioinformatics, 7: 271–278. https://doi.org/10.4137/EBO.S8092

Yan, C., Du, J. C., Gao, L., Li, Y., & Hou, X. L. (2019). The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in Cardamineae.

Gene, 699: 24–36. https://doi.org/10.1016/j.gene.2019.02.075

Yang, J., Vazquez, L., Chen, X., Li, H., Zhang, H., Liu, Z., & Zhao, G. (2017). Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes. Frontiers in Plant Science, 8: 816. https://doi.org/10.3389/fpls.2017.00816

Yao, L. J., Ding, Y., Yao, L., Ai, X. R., & Zang, R. G. (2020). Trait gradient analysis for evergreen and deciduous species in a subtropical forest. Forests, 11: 364. https://doi.org/10.3390/f11040364

Yao, Z., Hanmei, L., & Yong, G. (2008). Analysis of characteristic of codon usage in waxy gene of *Zea mays*. Journal of Maize Sciences, 16(2): 16–21. https://doi.org/10.13597/j.cnki.maize.science.2008.02.036

Ye, W. Q., Yap, Z. Y., Li, P., Comes, H. P., & Qiu, Y. X. (2018). Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Molecular Phylogenetics and Evolution, 127: 978–987. https://doi.org/10.1016/j.ympev.2018.07.001

Yu, X. Q., Maki, M., Drew, B. T., Paton, A. J., Li, H. W., Zhao, J. L., ... & Li, J. (2014). Phylogeny and historical biogeography of *Isodon* (Lamiaceae): rapid radiation in south-west China and Miocene overland dispersal into Africa. Molecular phylogenetics and evolution, 77, 183-194. https://doi.org/10.1016/j.ympev.2014.04.017

842

843 Table 1. Geographic information and specimen voucher number of the *Diospyros*

844 species sequenced in this study.

Species	Voucher no.	Plastome	Locality
Diospyros strigosa	ZYH18080301	OP480009	South China Botanical Garden
Diospyros morrisiana	ZYH18072101	OP485441	Heishiding, Zhaoqing, China (N 23°27’09", E 111°53’11")
Diospyros eriantha	ZYH18080302	OP480008	South China Botanical Garden

845

Table 2 Plastome features of 18 *Diospyros* species. The newly sequenced data is shown in bold.

Species	GenBank	Habit	Total (bp)	LSC (bp)	SSC (bp)	IR (bp)	CDS (bp)	Gene	CDS	Pseudo	tRNA	rRNA
D. eriantha	**OP480008**	Evergreen	157432	87181	18471	25890	80379	136	89	2	37	8
D. strigosa	**OP480009**	Evergreen	157371	87158	18467	25873	80416	134	89	2	37	8
D. blancoi	KX426216	Evergreen	157745	87246	18323	26088	80700	138	91	2	37	8
D. cathayensis	MF288576	Evergreen	157689	87176	18349	26082	80817	138	91	2	37	8
D. rhombifolia	MF288578	Evergreen	157368	87223	18325	25910	80859	138	91	2	37	8
D. morrisiana	**OP485441**	Deciduous	157737	87164	18455	26088	80838	138	91	2	37	8
D. glaucifolia	KM504956	Deciduous	157593	86974	18413	26103	80817	137	91	1	37	8
D. kaki	KT223565	Deciduous	157784	87112	18536	26068	80823	137	91	1	37	8
D. lotus	KM522849	Deciduous	157590	86944	18416	26115	80940	138	91	2	37	8
D. oleifera	KM522850	Deciduous	157724	87056	18522	26073	80817	137	91	1	37	8

Species	GenBank	Habit	Total (bp)	LSC (bp)	SSC (bp)	IR (bp)	CDS (bp)	Gene	CDS	Pseudo	tRNA	rRNA
D. deyangensis	MF288575	Deciduous	157934	87237	18485	26106	80826	138	91	2	37	8
D. jinzaoshi	KM522848	Deciduous	157321	86929	18174	26109	80781	138	91	2	37	8
D. virginiana	MF288577	Deciduous	157761	87089	18444	26114	80958	138	91	2	37	8
D. flavocarpa	MG049699	Island	157420	86880	18420	26060	80685	138	91	2	37	8
D. yaouhensis	MG049731	Island	157409	86874	18415	26060	80682	138	91	2	37	8
D. ferrea	MG049698	Island	157398	87008	18264	26063	80706	138	91	2	37	8
D. tridentata	MG049723	Island	157479	86941	18418	26060	80673	138	91	2	37	8
D. vieillardii	MG049728	Island	157544	86999	18409	26068	80680	138	91	2	37	8