

## Research article

# Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Diospyros oleifera*, the first representative from the family Ebenaceae



Yang Xu<sup>a</sup>, Yi Dong<sup>a</sup>, Wenqiang Cheng<sup>a</sup>, Kaiyun Wu<sup>a</sup>, Haidong Gao<sup>b</sup>, Lei Liu<sup>b</sup>, Lei Xu<sup>b</sup>,  
Bangchu Gong<sup>a,\*</sup>

<sup>a</sup> Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, 311400, China

<sup>b</sup> Geneioneer Biotechnologies Co. Ltd, Nanjing, 210023, China

## ARTICLE INFO

## Keywords:

*Diospyros oleifera*  
Mitochondrial genome  
Phylogenetic analysis

## ABSTRACT

Plant mitochondrial genomes are a valuable source of genetic information for a better understanding of phylogenetic relationships. However, no mitochondrial genome of any species in Ebenaceae has been reported. In this study, we reported the first mitochondrial genome of an Ebenaceae model plant *Diospyros oleifera*. The mitogenome was 493,958 bp in length, contained 39 protein-coding genes, 27 transfer RNA genes, and 3 ribosomal RNA genes. The rps2 and rps11 genes were missing in the *D. oleifera* mt genome, while the rps10 gene was identified. The length of the repetitive sequence in the *D. oleifera* mt genome was 31 kb, accounting for 6.33%. A clear bias in RNA-editing sites were found in the *D. oleifera* mt genome. We also detected 28 chloroplast-derived fragments significantly associated with *D. oleifera* mt genes, indicating intracellular tRNA genes transferred frequently from chloroplasts to mitochondria in *D. oleifera*. Phylogenetic analysis based on the mt genomes of *D. oleifera* and 27 other taxa reflected the exact evolutionary and taxonomic status of *D. oleifera*. Ka/Ks analysis revealed that 95.16% of the protein-coding genes in the *D. oleifera* mt genome had undergone negative selections. But, the rearrangement of mitochondrial genes has been widely occur among *D. oleifera* and these observed species. These results will lay the foundation for identifying further evolutionary relationships within Ebenaceae.

## 1. Introduction

Mitochondria are the main organelle involved in energy metabolism in plants [1, 2]. They supply ATP via oxidative phosphorylation for metabolism, cell differentiation, apoptosis, cell growth, and cell division and are abundant in energy-consuming tissues involved in essential biological functions [1, 2, 3, 4]. Therefore, mitochondria play an important role in plant productivity and development [2, 5, 6]. According to endosymbiotic theory, plant mitochondria are believed to have descended from free-living bacteria-independent microorganisms, which explains the presence of their genomes [5, 7].

During evolution, the plant mitochondrial (mt) genome underwent dramatic changes in, for example, the gene order, genome structure, and migration of sequences from other organelles [5, 7, 8, 9]. Thus, plants have about 100–10,000 times larger and more structurally complex mitochondrial (mt) genomes than animals [10, 11, 12]. The mt genomes of plants demonstrate significant genome size variation, from 66 kb [13]

to 11.3 Mb [14]; the number of protein-coding genes varies from 14 to 67 [15]; and the number of tRNA genes varies from 3 to 27 [9]. There are variations in mitochondrial genomes not only between plant species but also within the same species [9, 12, 16, 17], in stark contrast to the conserved structure of plant chloroplast genomes [16, 17, 18]. Thus, mt genomes have been used as a valuable source of genetic information and for investigation of essential cellular processes in many phylogenetic studies [18, 19, 20, 21].

While, these characteristics of plant mt genomes (bigger size, more structural complexity, and low conservation across species) make plant mitochondrial genome assembly difficult [1, 8, 10]. To date, more than 5000 plant chloroplast genomes have been sequenced, but only about 400 mt genome sequences are available ([www.ncbi.nlm.nih.gov/genome/organelle/](http://www.ncbi.nlm.nih.gov/genome/organelle/), 11/11/2021). In addition, sequenced plants largely differ in their classification, and only three complete mitochondrial genomes of species from the order Ericales have been identified.

\* Corresponding author.

E-mail address: [gongbc@126.com](mailto:gongbc@126.com) (B. Gong).

<https://doi.org/10.1016/j.heliyon.2022.e09870>

Received 22 February 2022; Received in revised form 18 April 2022; Accepted 30 June 2022

2405-8440/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Diospyros* L., from the Ebenaceae family, is a plant genus that includes over 500 species widely distributed across tropical and subtropical regions [22] and that is one of the largest angiosperm genera [23]. Among these species, *Diospyros oleifera* and *Diospyros kaki* have been cultivated as an important fruit crop in China, Korea, Japan for centuries, due to its edible fruit is rich in vitamins, sugars, nutrients and antioxidants vital for optimum health with various medicinal and chemical [24, 25]. Morphological, molecular, and genomic studies have shown that *D. oleifera* can be used as a model plant [24, 26]. Chloroplast genome sequencing has been performed in 15 species of *Diospyros* [26, 27], and nuclear genome sequencing has been performed in *D. oleifera* [23, 28] and *Diospyros lotus* [29, 30]. However, to date, no mt genome of any species in Ebenaceae has been reported.

Fortunately, advancements in long read sequencing, such as PacBio and Oxford Nanopore, have made organelle genome sequencing easier and faster. Therefore, in this study, we constructed the complete mt genome of *D. oleifera* based on PacBio and Illumina data, performed a phylogenetic analysis, and compared the complete mt genomes of *D. oleifera* and related genera. These results will help better understand the features of the *D. oleifera* mitochondrial genome and lay the foundation for identifying further evolutionary relationships within Ebenaceae.

## 2. Materials and methods

### 2.1. Samples and mitogenome sequencing

Due to the advancement of sequence technology, long reads, used for *de novo* assembly of organelle genomes without the need for organelle

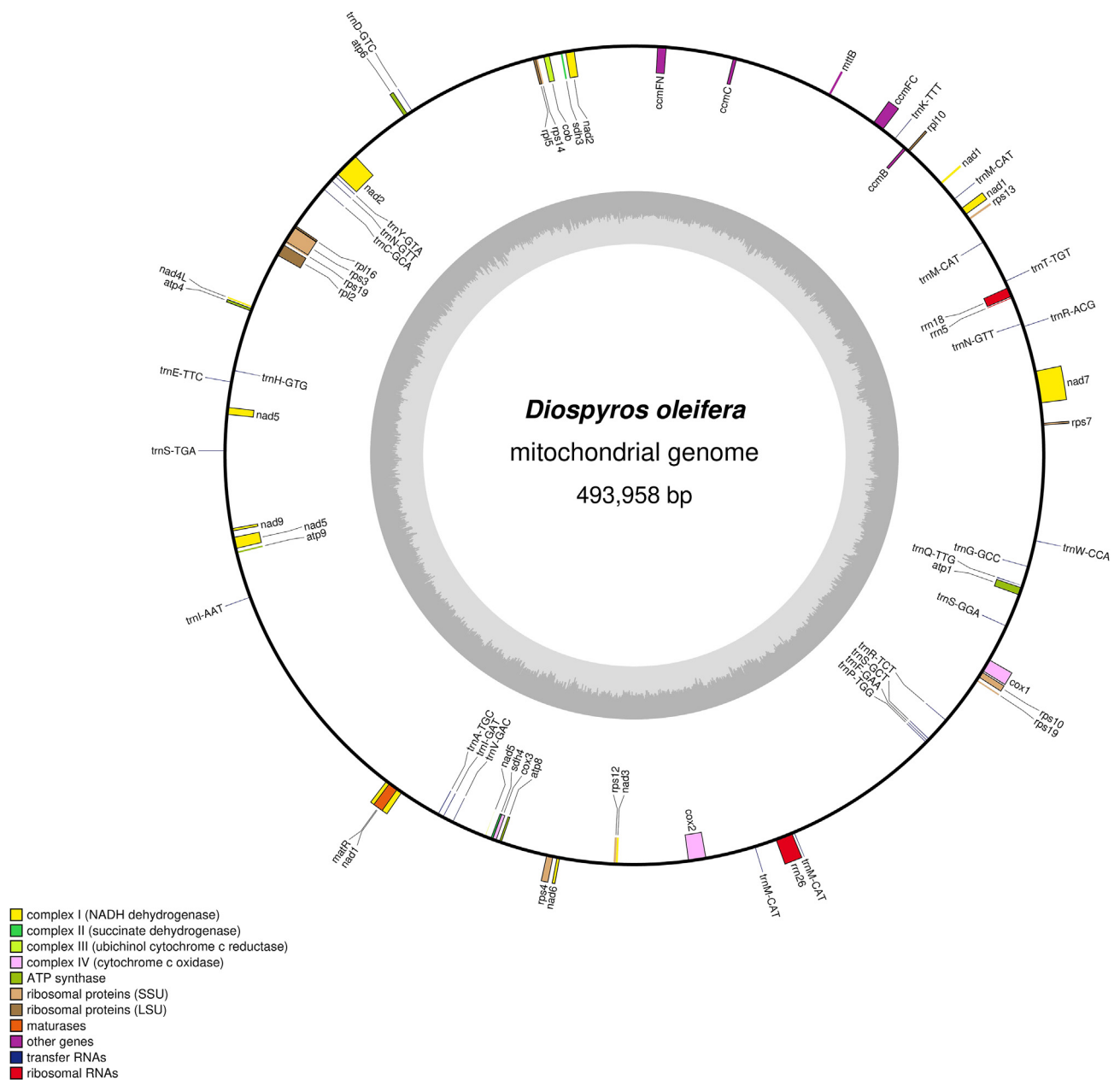
DNA isolation, could be easily generated from high throughput sequencing. The well-established methodology is quite efficient and well accepted in the scientific community [1, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42].

In this study, Mature leaves of *D. oleifera* (at latitude 34.27569 and longitude 107.75079) were used to isolate total DNA following the protocol for the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) and the SMRTbell Libraries protocol for PacBio data (Pacific Biosciences, Menlo Park, CA, USA). All these whole-genome Illumina HiSeq and PacBio sequencing data were deposited in the NCBI GenBank (accession no. PRJNA562043) and the Persimmon Genome Website ([http://www.kak.iwi.zju.edu.cn/cgi-bin/persimmon/about\\_genome.cgi](http://www.kak.iwi.zju.edu.cn/cgi-bin/persimmon/about_genome.cgi)). Sequencing reads of the mitochondria were filtered and extracted from these WGD sequencing data of *D. oleifera*. Raw data of second-generation sequencing were filtered using fastp version 0.20.0 software (<https://github.com/OpenGene/fastp>) [43]. The three-generation sequencing data of mitochondrial reads were error-corrected, trimmed, and *de-novo*-assembled using a Canu assembler (version 1.5) with default parameters [44]. Then, the contig sequence was obtained. The gene databases of plant mitochondria (the mitochondrial gene sequences of species published on the NCBI) were compared using blast v2.6 (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), and contigs that matched with the mitochondrial gene as the seed sequence were selected. The original data were used to extend and circularize the contigs to obtain the ring-dominant structure (or secondary ring), and then, the assembly was polished using NextPolish 1.3.1 (<https://github.com/Nextomics/NextPolish>) [45]. The assembly results were calibrated using second- and third-generation data, and the parameters were set as *rerun* = 3 and *-max\_depth* = 100. Then, the final assembly results were obtained.

**Table 1.** GenBank accession numbers of mitochondrial genomes for species sampled in this study.

Classification Status	Order	Familiy	Species	Length (bp)	Accession number		
Ingroup	Asterids	Apiales	Apiaceae	<i>Daucus carota</i> #	281,132	NC_017855	
	Asterids	Aquifoliales	Aquifoliaceae	<i>Ilex pubescens</i>	517,520	NC_045078	
	Asterids	Asterales	Asteraceae	<i>Chrysanthemum boreale</i>	211,002	NC_039757	
	Asterids	Asterales	Asteraceae	<i>Helianthus annuus</i>	300,945	NC_023337	
	Asterids	Asterales	Asteraceae	<i>Lactuca sativa</i>	363,324	NC_042756	
	Asterids	Asterales	Asteraceae	<i>Lactuca serriola</i>	363,328	NC_042378	
	Asterids	Asterales	Campanulaceae	<i>Codonopsis lanceolata</i>	403,704	NC_037949	
	Asterids	Asterales	Campanulaceae	<i>Platycodon grandiflorus</i>	1,249,593	NC_035958	
	Asterids	Ericales	Ebenaceae	<i>Diospyros oleifera</i> *#	493,958	MW970112	
	Asterids	Ericales	Ericaceae	<i>Rhododendron simsii</i> #	802,707	WJXA01000014	
	Asterids	Ericales	Ericaceae	<i>Vaccinium macrocarpon</i> #	459,678	NC_023338	
	Asterids	Gentianales	Rubiaceae	<i>Scyphiphora hydrophyllacea</i>	354,155	MT610041	
	Asterids	Lamiales	Lamiaceae	<i>Salvia miltiorrhiza</i>	499,236	NC_023209	
	Asterids	Lamiales	Lentibulariaceae	<i>Utricularia reniformis</i>	857,234	NC_034982	
	Asterids	Lamiales	Oleaceae	<i>Olea europaea</i>	710,808	MW262896	
	Asterids	Lamiales	Phrymaceae	<i>Mimulus guttatus</i>	525,671	NC_018041	
	Asterids	Solanales	Convolvulaceae	<i>Ipomoea nil</i>	265,768	NC_031158	
	Asterids	Solanales	Solanaceae	<i>Capsicum annuum</i>	511,530	KJ865410	
	Asterids	Solanales	Solanaceae	<i>Nicotiana tabacum</i>	430,597	NC_006581	
	Asterids	Solanales	Solanaceae	<i>Solanum lycopersicum</i> #	446,257	NC_035963	
Commelinids	Poales	Poaceae	<i>Oryza sativa</i>	637,692	JF281153		
	Poales	Poaceae	<i>Zea mays</i>	680,603	DQ645539.1		
	Rosids	Brassicales	Brassicaceae	<i>Arabidopsis thaliana</i>	367,808	NC_037304	
	Rosids	Fabales	Fabaceae	<i>Glycine max</i>	402,558	NC_020455	
	Rosids	Rosales	Rosaceae	<i>Malus domestica</i> #	396,947	NC_018554	
	Rosids	Vitales	Vitaceae	<i>Vitis vinifera</i>	773,279	NC_012119	
	Outgroup	Conifers	Ginkgoales	Ginkgoaceae	<i>Ginkgo biloba</i>	346,544	KM672373
		Conifers	Pinales	Pinaceae	<i>Pinus taeda</i>	1,191,054	MF991879

\*Represents the new mitogenome in this study. #Represents these species were used for mitogenome synteny and rearrangements through Mauve software.



**Figure 1.** The circular map of *D. oleifera* mt genome. Gene map showing 69 annotated genes of different functional groups.

## 2.2. Genome annotation

The assembled *D. oleifera* mt genome was annotated using the GeSeq tool [46]. To confirm the annotated results, the assembled *D. oleifera* mt genome was also BLAST-searched against protein-coding genes and ribosomal RNA (rRNA) genes of available plant mt genomes at the NCBI. Then, the sequence coordinates of the identified protein-coding genes (PCGs) were manually verified for start and stop codons. The annotations of transfer RNA (tRNA) genes were also confirmed by tRNAscan-SE [47]. ViennarNA-2.4.14 [48] was used to visualize the secondary structure of tRNA. The physical circular map was drawn using the Organellar Genome DRAW (OGDraw) v1.2 program [49]. The final annotated mt genome sequences of *D. oleifera* have been deposited in the NCBI GenBank (accession no. MW970112).

Strand asymmetry was calculated according to the formulas: AT-skew =  $[A - T]/[A + T]$  and GCskew =  $[G - C]/[G + C]$  [50]. The possible RNA-editing sites in the PCGs of *D. oleifera* were predicted using the

online predictive RNA editor for plant mitochondrial genes (PREP-Mt) [51] suite of servers (<http://prep.unl.edu/>). The codon frequencies were calculated using the Codon Usage tool in the Sequence Manipulation Suite ([bioinformatics.org/sms2/codon\\_usage.html](http://bioinformatics.org/sms2/codon_usage.html)) [52]. The relative synonymous codon usage (RSCU [53]) was calculated using the CAI Python package of Lee [54].

## 2.3. Analysis of repeated sequences

Three kinds of repeats (simple sequence, tandem, and dispersed) were detected in the *D. oleifera* mitochondrial genome. The MicroSatellite (MISA) identification tool Perl script was used to detect simple sequence repeats [55]. The repeats of mono-, di-, tri-, tetra-, penta-, and hexanucleotide bases with 12, 6, 4, 3, 3, and 3 repeat numbers, respectively, were identified. Tandem repeats (>6 bp repeat units) were detected using Tandem Repeats Finder v4.09 software (<http://tandem.bu.edu/trf/trf.submit.options.html>) [56] with default parameters (matching

**Table 2.** Gene profile and organization of the *D. oleifera* mt genome.

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
ATP synthase	<i>atp1</i>	1530	ATG	TGA	510
	<i>atp4</i>	579	ATG	TAA	193
	<i>atp6</i>	807	ATG	TGA	269
	<i>atp8</i>	480	ATG	TAA	160
	<i>atp9</i>	285	ATG	TAG	95
Cytochrome c biogenesis	<i>ccmB</i>	621	ATG	TGA	207
	<i>ccmC</i>	753	ATG	TGA	251
	<i>ccmFC*</i>	1353	ATG	TAA	451
	<i>ccmFN</i>	1755	ATG	TGA	585
Ubichinol cytochrome c reductase	<i>cob</i>	1182	ATG	TGA	394
Cytochrome c oxidase	<i>cox1*</i>	1584	ACG (ATG)	TAA	528
	<i>cox2**</i>	780	ATG	TAA	260
	<i>cox3</i>	798	ATG	TGA	266
Maturases	<i>matR</i>	1968	ATG	TAG	656
Transport membrane protein	<i>mttB</i>	375	ATG	TAG	125
NADH dehydrogenase	<i>nad1****</i>	978	ATG	TAA	326
	<i>nad2****</i>	1467	ATG	TAA	489
	<i>nad3</i>	357	ATG	TAA	119
	<i>nad4***</i>	1488	ATG	TGA	496
	<i>nad4L</i>	273	ATG	TAA	91
	<i>nad5****</i>	2013	ATG	TAA	671
	<i>nad6</i>	618	ATG	TAA	206
	<i>nad7****</i>	1185	ATG	TAG	395
	<i>nad9</i>	588	ATG	TAG	196
Ribosomal proteins (LSU)	<i>rpl10</i>	489	ATG	TAA	163
	<i>rpl16</i>	435	ND	TAA	145
	<i>rpl2*</i>	1005	ATG	TAA	335
	<i>rpl5</i>	564	ATG	TAA	188
	Ribosomal proteins (SSU)	<i>rps10*</i>	333	ACG (ATG)	CGA
	<i>rps12</i>	378	ATG	TGA	126
	<i>rps13</i>	351	ATG	TGA	117
	<i>rps14</i>	303	ATG	TAG	101
	<i>rps19(2)</i>	(231,231)	ATG	TAA	77
	<i>rps3*</i>	1752	ATG	TAG	584
	<i>rps4</i>	1326	ATG	TAA	442
	<i>rps7</i>	447	ATG	TAA	149
Succinate dehydrogenase	<i>sdh3</i>	306	ATG	TGA	102
	<i>sdh4</i>	432	ATG	TGA	144
Ribosomal RNAs	<i>rrn18</i>	1904			
	<i>rrn26</i>	3373			
	<i>rrn5</i>	119			
Transfer RNAs	<i>trnA-TGC*</i>	67			
	<i>trnC-GCA</i>	71			
	<i>trnD-GTC</i>	74			
	<i>trnE-TTC</i>	72			
	<i>trnE-GAA</i>	74			
	<i>trnG-GCC</i>	72			
	<i>trnH-GTG</i>	74			
	<i>trnI-AAT</i>	69			
	<i>trnI-GAT*</i>	74			
	<i>trnK-TTT</i>	73			
	<i>trnM-CAT(4)</i>	(73,74,74,77)			
	<i>trnN-GTT(2)</i>	(72,72)			

**Table 2 (continued)**

Group of genes	Gene name	Length	Start codon	Stop codon	Amino acid
	<i>trnP-TGG</i>	75			
	<i>trnQ-TTG</i>	72			
	<i>trnR-ACG</i>	74			
	<i>trnR-TCT*</i>	72			
	<i>trnS-GCT</i>	88			
	<i>trnS-GGA</i>	87			
	<i>trnS-TGA</i>	87			
	<i>trnT-TGT*</i>	75			
	<i>trnV-GAC</i>	72			
	<i>trnW-CCA</i>	74			
	<i>trnY-GTA</i>	83			

probability of 80 and indel probability of 10). Direct and inverted repeats were detected using the vmatch (v2.3.0) Perl script with the minimal repeat size set to 30 bp.

#### 2.4. Chloroplast-to-mitochondrion-DNA transformation

The *D. oleifera* cp genome (NC\_030787.1) was downloaded from the NCBI Organelle Genome Resources Database. The protein-coding and tRNA genes, which were transferred from chloroplasts to mitochondria, were identified using Blastn software with the following screening criteria: matching rate  $\geq 70\%$ , E-value  $\leq 1e - 10$ , and length  $\geq 30$  bp.

#### 2.5. Phylogenetic tree construction and Ka/Ks analysis

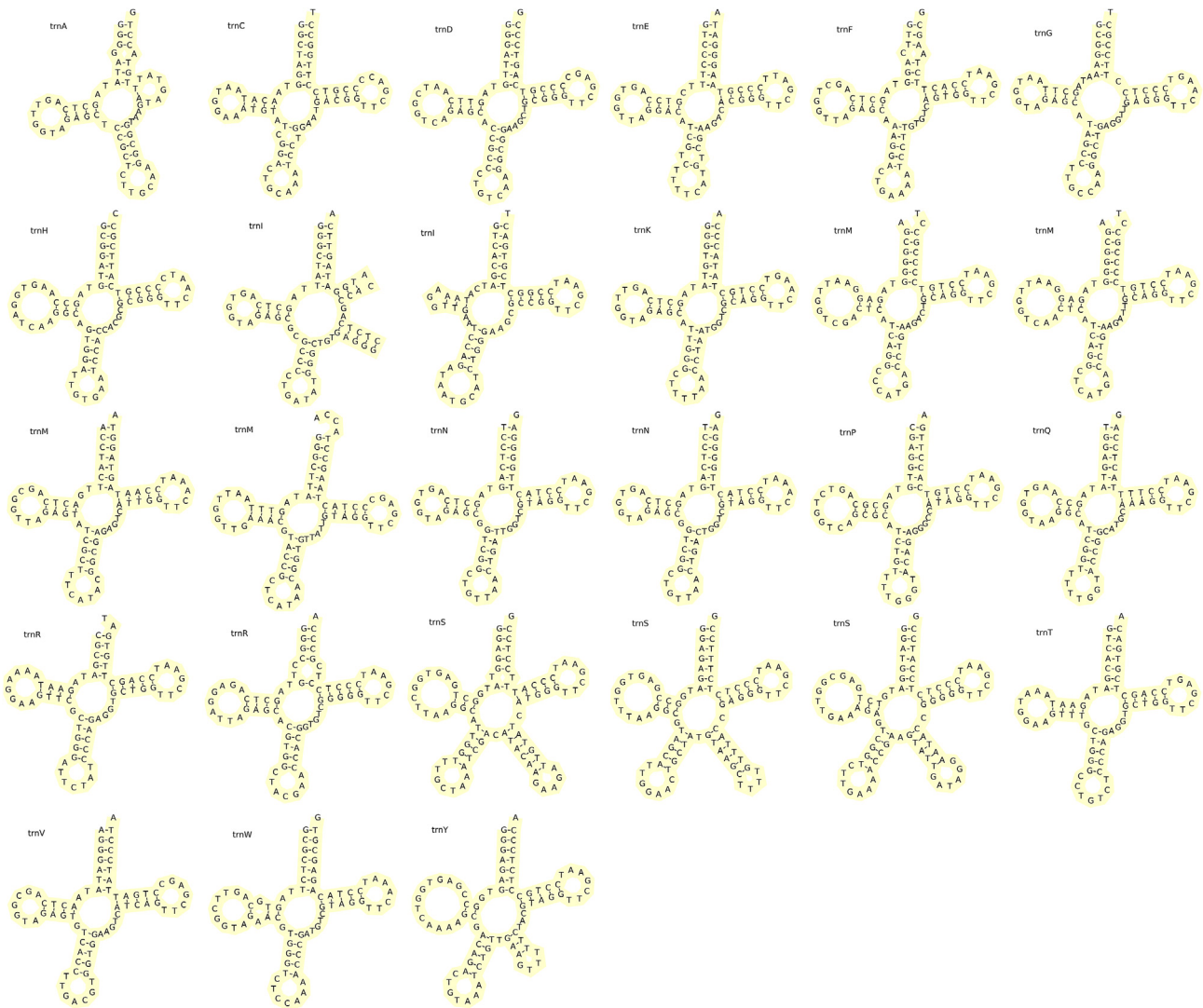
The *D. oleifera* and Twenty-seven other species with complete or nearly complete mitogenomes were used in phylogenetic analyses, representing twenty families. Two species from Conifers were used as outgroup. All species were listed in Table 1. The mt genomes were downloaded from the NCBI Organelle Genome Resources Database, and the conserved protein-coding genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl5*, *rps12*, *rps13*, *rps3*, and *rps4*) were extracted and aligned using MAFFT v7.402 [57] with default parameters. ModelTest-NG v0.1.3 was used to determine the best-fit model, and a maximum likelihood (ML) tree was generated using RAXMLv8.2.12 with the best-fit substitution model (GTRGAMMA) at 1000 bootstrap replicates [58].

The synonymous (Ks) and nonsynonymous (Ka) substitution rates of the protein-coding genes in the *D. oleifera* mt genome were analyzed using the 27 species. In this analysis, KaKs\_Calculator (v2.0) [59] with the MLWL model was used to calculate Ka/Ks. Genome synteny and rearrangements among the using six representative species (Table 1) mitogenomes were analyzed using the progressive Mauve algorithm as implemented in Mauve ver. 2.4.0 software [60].

### 3. Results and discussion

#### 3.1. Genomic features of the *D. oleifera* mt genome

The plant mitochondrial genome greatly varies in size, from 66 kb in *Viscum scurruloideum* [13] to 11.3 Mb in *Silene conica* [14]. We assembled the complete mt genome of *D. oleifera* in a single circular contig of 493,958 bp (GenBank accession number MW970112). The relatively medium size of the *D. oleifera* mt genome is similar to that of *Vaccinium macrocarpon* (459,678 bp) [3] and some asterids, such as *Solanum lycopersicum* (446,257 bp) [61], *Salvia miltiorrhiza* (499,236 bp) [62], and *Capsicum annuum* (511,530 bp) [63]; smaller than that of *Rhododendron simsii* (802,707 bp) [64] and *Olea europaea* (710,808 bp) [65]; and larger than that of *Daucus carota* (281,132 bp) [66] and *Malus domestica* (396,947 bp) [67].



**Figure 2.** Secondary structures of tRNAs of *D. oleifera*. Each region of tRNA is named as follows [69, 70]: Amino acid accepting stem, AAS (upper arm); dihydrouridine stem and loop, DSL (left arm); anticodon stem and loop, ASL (lower arm); thymidine stem and loop, TSL (right arm); variable stem and loop, VSL (between ASL and TSL).

**Table 3.** Composition and skewness of the *D. oleifera* mt genome.

<i>D.oleifera</i>	Size (bp)	A%	T%	G%	C%	A + T%	G + C%	AT-skew	GC-skew
Mitogenome	493958	27.27	27.03	22.8	22.9	54.3	45.7	0.004	-0.002
PCGs	32400	26.59	30.3	21.9	21.21	56.89	43.11	-0.065	0.016
tRNAs	2021	22.46	26.03	28.95	22.56	48.49	51.51	-0.073	0.124
rRNAs	5396	25.8	22.07	29.23	22.91	47.87	52.13	0.078	0.121

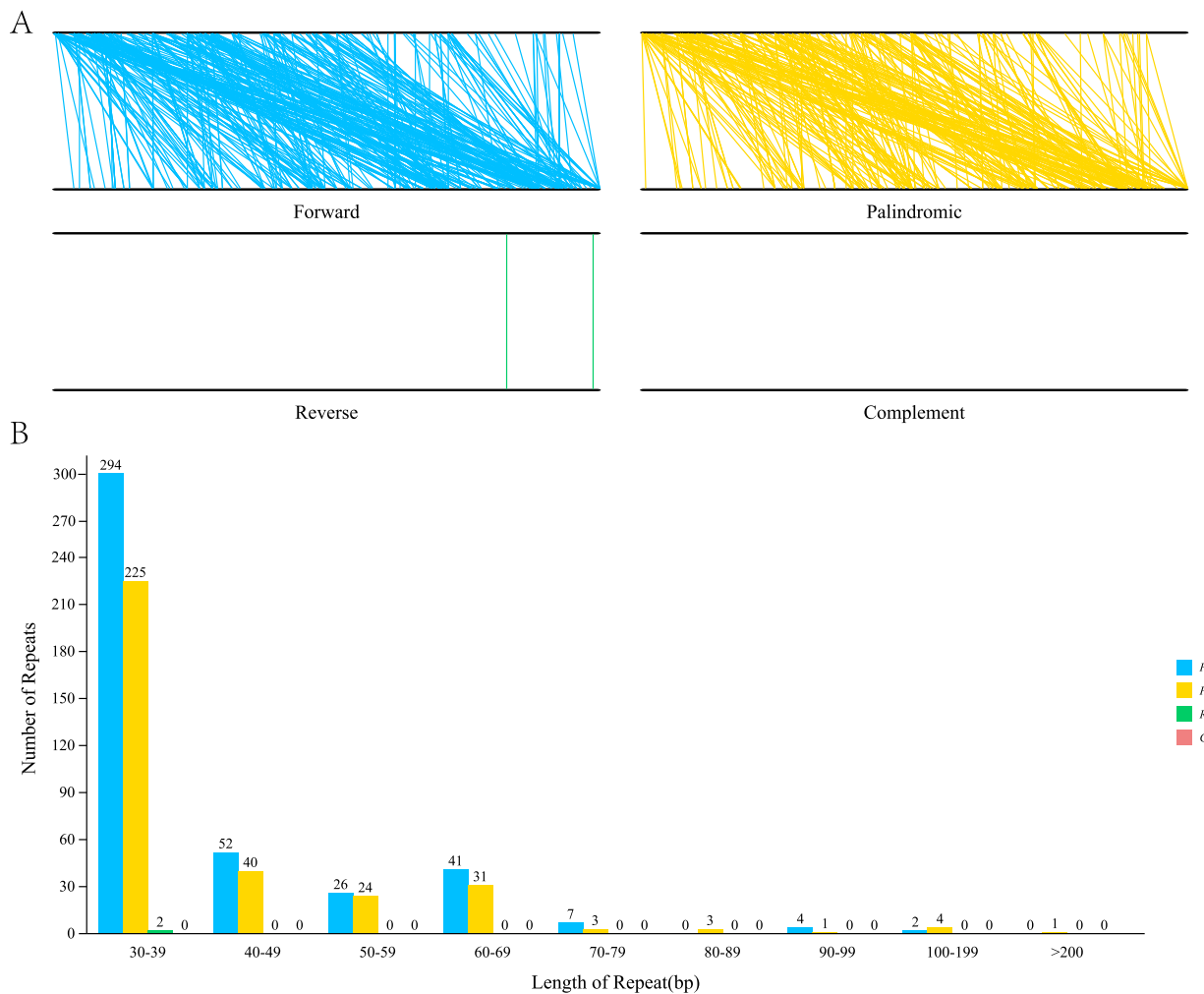
In the *D. oleifera* mt genome, 69 genes (39 protein-coding genes, 27 tRNA genes, and 3 rRNA genes (*rrn5*, *rrn18*, and *rrn26*)) were annotated. The functional categorization and physical locations of the annotated genes are shown in Figure 1. The 38 different proteins (*rps19* has two copies) could be divided into 10 classes (Table 2): ATP synthase (five genes), cytochrome C biogenesis (four genes), ubiquinol cytochrome c reductase (one gene), cytochrome C oxidase (three genes), maturases (one gene), transport membrane protein (one gene), NADH dehydrogenase (nine genes), ribosomal proteins (LSU; four genes), ribosomal proteins (SSU; nine genes), and succinate dehydrogenase (two genes). ATG was used as the starting codon by almost all the protein-coding genes, and the four stop codons (TAA, TGA, TAG, and CGA) had utilization rates of 48.71%, 30.77%, 17.95%, and 2.57%, respectively.

Previous studies have shown that *rps10* is missing in the mt genomes of most plants, such as *Arabidopsis thaliana*, *Brassica napus*, and *Beta vulgaris*, and that its function is replaced by the nuclear gene [9]. However, the *rps10* gene was found in the *D. oleifera* mt genome. The absence of *rps2* and *rps11* genes in the *D. oleifera* mt genome, consistent with *R. simsii* [64] and *V. macrocarpon* [3], supports Adams' speculation that *rps2* and *rps11* genes were lost in the early evolution of eukaryotic plants [3]. Similar to *Nicotiana tabacum* [68] and *M. luteus* [4], the *D. oleifera* mt genome has no *rps1* gene, whereas *rps1* is present in the *V. macrocarpon* mt genome [3] and two copies of *rps1* are present in the *R. simsii* mt genome [64].

The persimmon mitochondria have 27 tRNAs (23 typical tRNA genes, one more *trnN-GTT* and three more *trnM-CAT*). The average length of

**Table 4.** Distribution of perfect tandem repeats in the *D. oleifera* mt genome.

NO.	Size	Copy	Repeat sequence	Percent Matches	Start	End
1	30	1.9	TACTACAATCCGTACGATAACTAGAATCCG	82	123393	123450
2	18	2.2	GCTTGATTCGGTGTAAC	90	143948	143987
3	20	2	TTTGATTTTCATCTTCATATAC	90	176075	176115
4	14	2.8	GGAGCTGACACCCT	84	210479	210515
5	15	2.4	AAATAAAAAATAAA	90	273479	273514
6	19	2.1	AACAACCTATCTTGCACGA	90	308468	308506
7	15	6.7	ACAACCTATTATGCG	70	308469	308572
8	18	2.1	AATACTAATAGAATAGAA	90	335217	335254
9	18	2.4	CATAGTCGCGAGCTGTTT	81	400200	400242
10	6	4.2	AAAGAA	100	409196	409220
11	18	5.2	TATTGATGATAGTGACGA	92	456597	456686
12	9	6.8	ATTGATGAT	73	456613	456673



**Figure 3.** The repeats in the *D. oleifera* mt genome. A: The synteny between the mt genome and its copy showing the direct repeats. B: The length distribution of reverse and inverted repeats in the *D. oleifera* mt genome. The number on the histograms represents the repeat number of designated lengths shown on the horizontal axis.

these tRNAs is 67–88 bp, with a total length of 1479 bp. The number of tRNAs in the *D. oleifera* mt genome is more than that in other asterids, such as *V. macrocarpon* (18) [3], *R. simsii* (23) [64], *M. luteus* (24) [4], and *N. tabacum* (21) [68]. This may be because some tRNAs in the *D. oleifera* mt genome have multiple copies; for example, *trnN-GTT* has two copies and *trnM-CAT* has four copies. The secondary structures are shown in

**Figure 2.** Following terms for Agris et al. [69, 70], secondary structures of most tRNAs were recovered as ordinal cloverleaf structures, which includes amino acid accepting stem (AAS), dihydrouridine stem and loop (DSL), anticodon stem and loop (ASL), thymidine stem and loop (TSL), furthermore, *trnI-GAT*, *trnS-GCT*, *trnS-GGA*, *trnS-TGA*, *trnY-GTA* were with an addition variable stem and loop (VSL). And, consist with many

**Table 5.** Prediction of RNA editing sites.

Type	Effect	Number	Percentage (%)		
Hydrophilic	CGT (R) => TGT (C)	28	13.40		
	CGC (R) => TGC (C)	13			
	CAT (H) => TAT (Y)	20			
	CAC (H) => TAC (Y)	8			
Hydrophobic	GCT (A) => GTT (V)	3	30.29		
	GCG (A) => GTG (V)	7			
	GCC (A) => GTC (V)	2			
	CTT (L) => TTT (F)	13			
	CTC (L) => TTC (F)	5			
	CCT (P) => CTT (L)	19			
	CCG (P) => CTG (L)	35			
	CCC (P) => TTC (F)	6			
	CCC (P) => CTC (L)	7			
	CCA (P) => CTA (L)	45			
	CCT (P) => TTT (F)	14			
	Hydrophilic-hydrophobic	TCT (S) => TTT (F)		44	47.57
		TCG (S) => TTG (L)		49	
		TCC (S) => TTC (F)		29	
TCA (S) => TTA (L)		78			
CGG (R) => TGG (W)		30			
ACT (T) => ATT (I)		4			
ACG (T) => ATG (M)		6			
ACA (T) => ATA (I)		5			
Hydrophilic-stop		CGA (R) => TGA (X)	3	0.77	
		CAA (Q) => TAA (X)	1		
Hydrophobic-hydrophilic	CCT (P) => TCT (S)	21	7.77		
	CCG (P) => TCC (S)	9			
	CCA (P) => TCA (S)	6			
	CCG (P) => TCG (S)	4			

report [19, 21, 38, 39, 42], G-T (U) matches were also found in mostly tRNA secondary structures in the *D. oleifera* mt genome.

The total gene length added up to 8% of the total mt genome length, with protein-coding regions comprising only 6.5% (32 kb) of the genome length. The gene content of *D. oleifera* is similar to that of the published mt genomes of asterids, especially *Mimulus guttatus* (7.4%) [4] and *Helianthus annuus* (8.5%) [71]. We found 54 genes with no introns,

accounting for 78.26% of the total, consistent with the result conclusion that 63.2%–100% of mitochondrial genes in most plants have no introns [8, 9]. In addition, 30 introns were found in the other 15 *D. oleifera* mt genes; *nad1*, *nad2*, *nad5*, and *nad7* had 4 introns; *nad4*, 3 introns; and *cox2*, 2 introns.

The nucleotide composition of the whole mt genome (Table 3) was found to be A (27.27%), T (27.03%), C (22.90%), and G (22.80%). The overall GC content was 45.7%, consistent with that of other asterids (*V. macrocarpon* 45.33% [3], *D. carota* 45.41% [66], *Ilex pubescens* 45.55% [35], *Camellia sinensis* 45.70% [33], and *R. simsii* 45.86% [64]). The GC skew was positive in CDS regions and negative in the mitochondrial genome. Strikingly, the GC content of the PCGs (43.11%) was lower than that of other CDS regions (tRNAs and rRNAs).

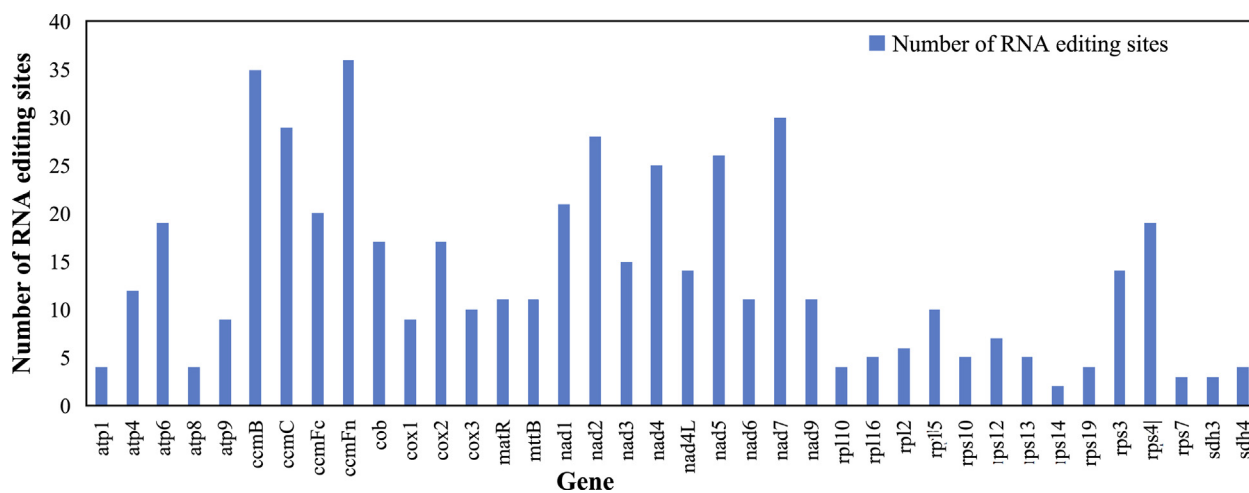
### 3.2. Repeat sequences analysis

Simple sequence repeats (SSRs, or microsatellites) are DNA stretches consisting of short, tandem units of sequence repetitions 1–6 base pairs in length [72]. We identified 87 SSRs in the *D. oleifera* mt genome. The proportions of different repeat units are shown in Table 1. Tetranucleotide repeats were the most abundant SSR type, constituting 68.97% of all identified SSRs, and there were 7 SSRs in di-, tri-, and pentanucleotide repeats, accounting for 8.05% of all identified SSRs. There were only three mono- and hexanucleotide repeats in the *D. oleifera* mt genome. AAAG/CTTT motifs (16) were most recurrent motifs, representing 18.39% of all identified SSRs (Table S1).

Tandem repeats (satellite DNA) are core repeating units of 1–200 bases repeated several times in tandem [73]. As shown in Table 4, 12 tandem repeats 6 to 30 bp long were observed in the *D. oleifera* mt genome.

In addition, 760 non-tandem repeats, with 30 bp or more in length, were detected in the *D. oleifera* mt genome. Of the 760 non-tandem repeats, 426 were direct, 332 were palindromic, and 2 were reverse. The longest direct-type repeat was 115 bp long, while the longest inverted repeat was 331 bp long (Table S2). As shown in Figure 3, the 30–39 bp repeats were most abundant for both repeat types.

The repetitive sequence in the *D. oleifera* mt genome was 31 kb, accounting for 6.33% of the total mitochondria. This is considered a medium proportion of repeats, higher than that in *Boea hygrometrica*, (1.5%) and *V. macrocarpon* (3%) and lower than that in *N. tabacum* (13%) [68] and *D. carota* (16%) [66]. The different proportions of repeats may be because the mitochondria of *B. hygrometrica*, *V. macrocarpon*, and



**Figure 4.** The distribution of RNA-editing sites in the *D. oleifera* mt protein-coding genes. The blue bars represent the number of RNA-editing sites of each gene.

*D. oleifera* are mainly short repeating units, whereas those of tobacco and carrots are mainly longer repeating units [66].

### 3.3. The prediction of RNA editing

The number of RNA-editing sites varies in different species and is usually frequent in angiosperm and gymnosperm mitochondria. We predicted 515 RNA-editing sites within all the 38 protein-coding genes (Table 5) in the *D. oleifera* mt genome, which is more than those in *A. thaliana* (441) [5], *Suaeda glauca* (261) [73], *Eucalyptus grandis* (470) [74], and *Citrullus lanatus* (463) [75] and less than those in gymnosperms with larger mt genomes, such as *Taxus cuspidata* (974), *Pinus taeda* (1179), *Cycas revoluta* (1206), and *Ginkgo biloba* (1306) [32]. However, whether the number of RNA-editing sites is positively correlated with the size of the mt genome requires further research.

The selection of mitochondrial RNA-editing sites in *D. oleifera* shows a high degree of compositional bias. All RNA-editing sites are the C-T editing type, which is consistent with the fact that C-T is the most common editing type found in plant mt genomes [76, 77, 78]. In previous studies, almost half of the mitochondrial RNA editing occurred at the second codon position [73, 77]. The proportion of RNA-editing sites at the second codon position in the *D. oleifera* mt genome is also about 45.72% (235), slightly less than that at the first codon position (259; 50.39%). However, no editing site was found at the third position of triplet codons, consistent with the fact that RNA-editing sites are rare in plant mt genomes [73, 78].

Due to mitochondrial RNA editing, the *D. oleifera* mt genome has more RNA-editing sites but fewer editing types (Table 5). There were only 29 codon transfer types, corresponding to 14 amino acid transfer types, among the 515 RNA-editing sites. The types of transfer are comparable to those of most gymnosperms (30–40 codons; around 20 amino acids) [32, 76] but less than those of monocotyledonous and dicotyledonous plants (50–60 codons; around 30 amino acids) [74, 75,

78]. Among the 29 codon transfer types, TCA => TTA was the most common, with 78 sites. A leucine tendency after RNA editing, supported by the fact that 45.24% (233 sites) of the edits are converted to leucine, was found in the amino acids of predicted editing codons. After RNA editing, 43.59% of the amino acids remained hydrophobic. However, 47.57% of the amino acids were predicted to change from hydrophilic to hydrophobic, while 7.77% were predicted to change from hydrophobic to hydrophilic.

The number and type of RNA-editing sites differed among the mt genomes of *D. oleifera* and other species. Like with most angiosperms [73, 76], ribosomal proteins (except *rps4*) and ATPase subunits (except *atp6*) had a relatively small number of RNA-editing-derived substitutions (1–12 sites), while the transcripts of NADH dehydrogenase subunits and cytochrome c biogenesis genes were significantly edited (11–36 sites; Figure 4), and *ccmFn* and *ccmB* had the most RNA-editing sites predicted (36, 35).

In *D. oleifera*, 10,611 amino acids were encoded. The most frequently used amino acids were Leu (10.25%), Ser (9.23%), and Arg (6.86%), and the least common amino acids were Trp (1.52%) and Met (2.65%) (Figure 4). The relative synonymous codon usage (RSCU) value for *D. oleifera* for the third codon position is shown in Figure 5. Consistent with most of the currently studied mitochondrial genomes [1, 73, 76], the use of both two- and four-fold degenerate codons was biased toward the use of codons abundant in A or T.

### 3.4. Chloroplast-derived mitogenomic sequences

The transfer of DNA sequences among chloroplast and mt genomes has been frequently observed in the mt genomes of plants [79]. In many cases, the chloroplast DNA content in the mt genomes of most plants is 3%–6%, sometimes reaching up to about 10% [80]. The *D. oleifera* mt genome contained 28 chloroplast insertions, ranging in length from 32 to 5703 bp (Figure 6, Table 6), with a total length of 32.83 kb, accounting for 6.65% of the total length of the genome, which is greater than the

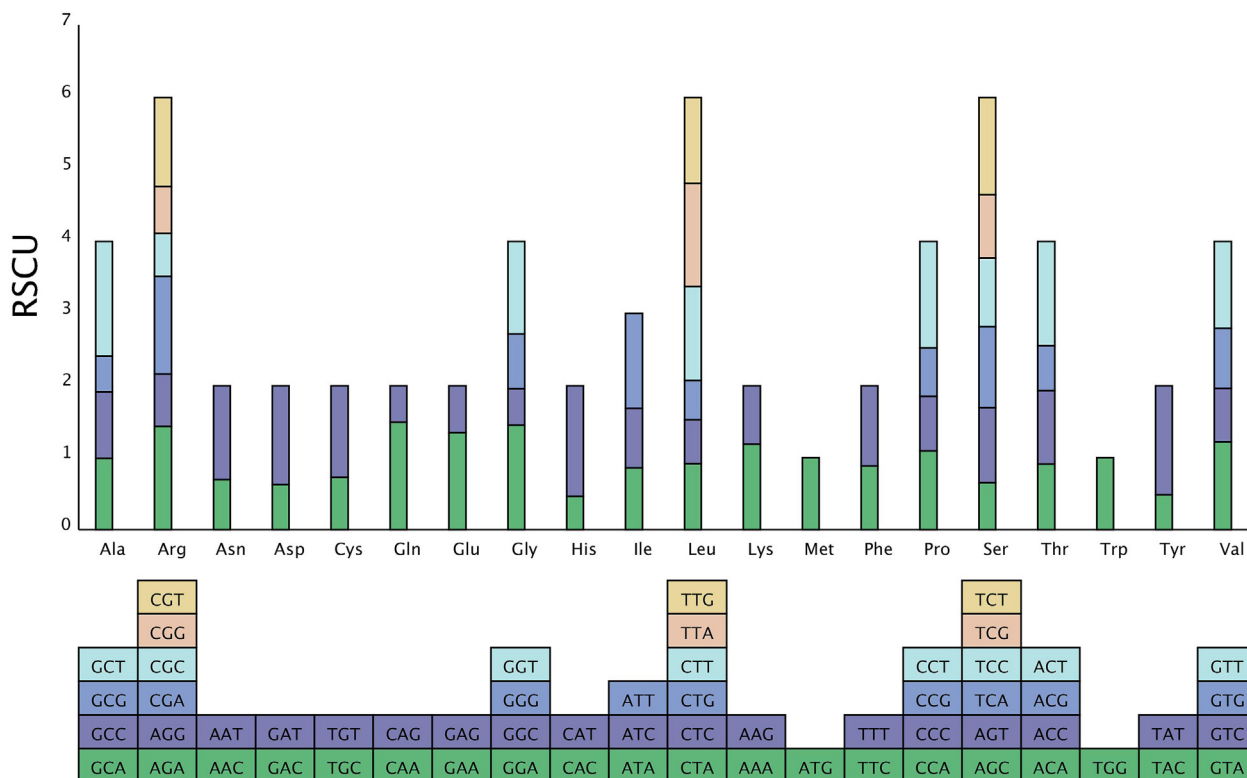
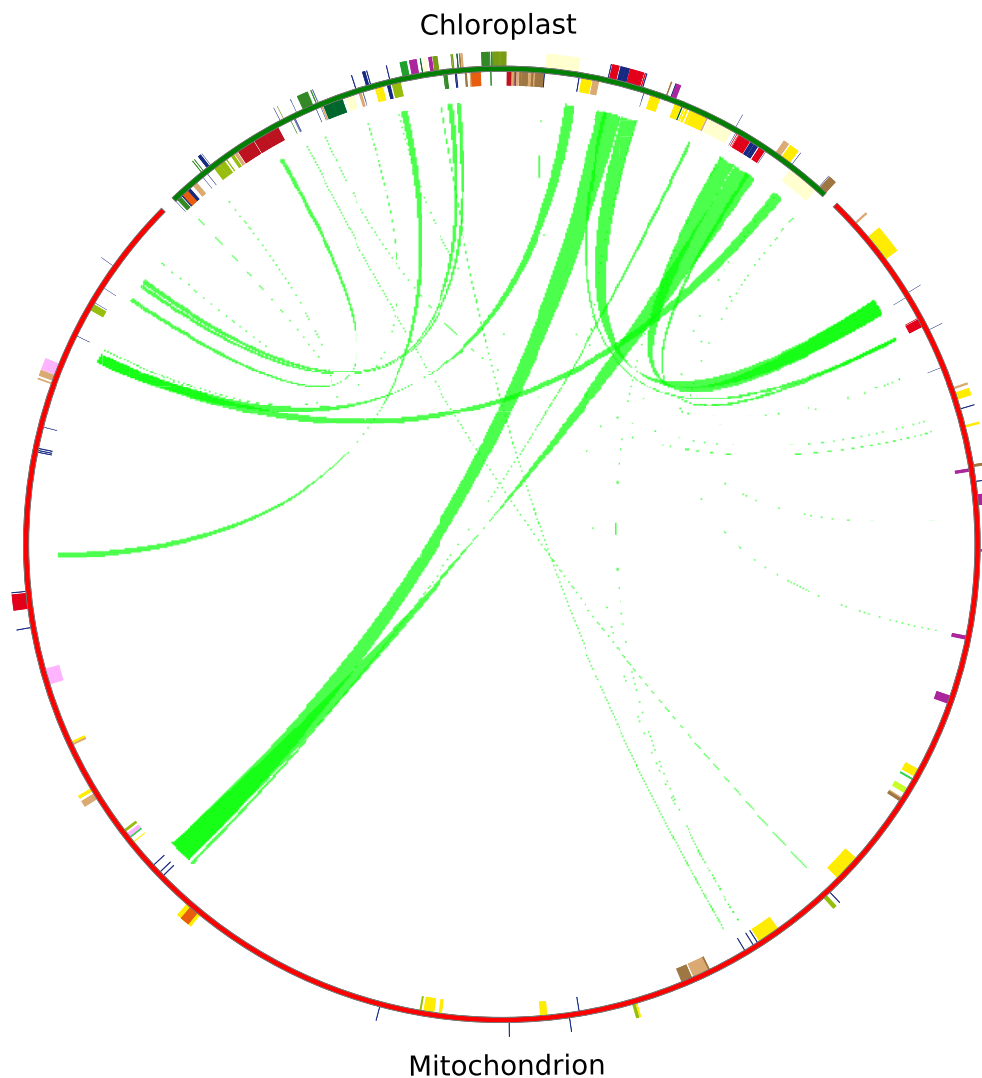


Figure 5. Relative synonymous codon usage in the *D. oleifera* mt genome.





**Figure 6.** DNA and gene transfer between Chloroplast and Mitochondrial genomes in *D. oleifera*. The track shows complete genomes of cp and mt in green and red respectively.

mitochondrial genome lengths of *Liriodendron tulipifera* (3%) [31], and *N. tabacum* (2.5%) [68]; comparable to those of *C. lanatus* (6%) [75], *E. grandis* (6%) [74], and *Oryza sativa* (6.3%) [81]; and less than those of *Vitis vinifera* (8.8%) [36] and *Cucurbita pepo* (11.5%) [75].

Among the transfer DNA sequences, some chloroplast protein-coding genes, such as *atpB*, *atpE*, *rps12*, *rpoB*, *petA*, *psaA*, and *psbC*, lost their integrity while migrating from the cp to the mitochondria, and only partial sequences of those cp-derived PCGs could be found in the *D. oleifera* mt genome (Table 6). In the *D. oleifera* mt genome, 11 chloroplast-derived tRNAs with a complete sequence were identified: *trnA-UGC*, *trnD-GUC*, *trnI-GAU*, *trnM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnR-ACG*, *trnS-GCU*, *trnS-GGA*, *trnV-GAC*, and *trnW-CCA*. The different completeness levels of the transferred PCGs and tRNA genes showed that tRNA genes are much more conserved in the mt genome than PCGs, indicating that tRNA genes play an indispensable role in mitochondria. The transfer of these tRNAs can be traced back to the retention of an earlier horizontal gene transfer event. In accordance with the present results, cp-derived *trnM-CAU* first appeared in gymnosperms [82]; cp-derived *trnD-GUC* mainly appeared in dicotyledons, not in monocotyledons [76]; and cp-derived *trnM-CAU* and *trnD-GUC* were both found in the *D. oleifera* mt genome. However, the absence of cp-derived *trnH-GTG*, which is commonly found in angiosperms [3, 74, 76, 82], and the presence of cp-derived *trnA-UGC*, lost during early evolution of

terrestrial plants [80, 83], indicate that special evolutionary events may be occurring during *D. oleifera* formation.

### 3.5. Phylogenetic, *Ka/Ks* and gene arrangement analysis

To detect the evolutionary status of the *D. oleifera* mt genome, a phylogenetic analysis was performed on *D. oleifera*, together with 27 other species: 23 eudicots (19 asterids and 4 rosids), 2 monocotyledons, and 2 gymnosperms (designated as outgroups). Phylogenetic relationships were analyzed using the concatenated dataset (28 PCGs: *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl5*, *rps12*, *rps13*, *rps3* and *rps4*) through ML phylogenetic analysis. The abbreviations and accession numbers of the mt genomes investigated in this study are listed in Table S1. As outgroups, the two gymnosperms were distinct from the other angiosperms. The phylogenetic tree (Figure 7) strongly supported the separation of asterids from rosids and the separation of eudicots from monocots. Moreover, the taxa from 20 families (Apiaceae, Aquifoliaceae, Asteraceae, Brassicaceae, Campanulaceae, Convolvulaceae, Ericaceae, Ebenaceae, Fabaceae, Ginkgoaceae, Lamiaceae, Lentibulariaceae, Oleaceae, Phrymaceae, Pinaceae, Poaceae, Rosaceae, Rubiaceae, Solanaceae, and Vitaceae) were well clustered. In addition, the monophyly of *D. oleifera*, which belongs to the single genus of *Diospyros* in the

**Table 6.** Chloroplast insertions in the mitochondrial genome of *D. oleifera*.

Chloroplast insertion	Start	End	Length	Chloroplast genes carried	Mitochondrial gene
1	102039	107741	5703	<i>rps12-rrn16-rrn23-trnA-UGC-trnI-GAU-trnV-GAC</i>	<i>nad5/trnA-TGC/trnI-GAT/trnV-GAC/</i>
2	137054	142756	5703	<i>rps12-rrn16-rrn23-trnA-UGC-trnI-GAU-trnV-GAC</i>	<i>nad5/trnA-TGC/trnI-GAT/trnV-GAC</i>
3	132935	136943	4009	<i>rps12-rrn23-rrn4.5-rrn5-trnN-GUU-trnR-ACG</i>	<i>trnN-GTT/trnR-ACG/</i>
4	107852	111860	4009	<i>rps12rrn23-rrn4.5-rrn5-trnN-GUU-trnR-ACG</i>	<i>trnN-GTT/trnR-ACG</i>
5	148183	150621	2439	<i>ycf15-ycf2</i>	ORF
6	94174	96612	2439	<i>rps12-ycf15-ycf2</i>	ORF
7	55481	56973	1493	<i>atpB-atpE</i>	ORF
8	66698	67774	1127	<i>psbE-psbF-psbJ-psbL</i>	ORF
9	24713	25651	939	<i>rpoB-rpoC1</i>	ORF
10	68737	69666	939	<i>petG-petL-trnPUGG-trnWCCA</i>	<i>trnW-CCA</i>
11	103614	104477	888	<i>rps12-rrn16</i>	<i>rrn18</i>
12	140318	141181	888	<i>rps12-rrn16</i>	<i>rrn18</i>
13	124702	125419	719	<i>rps12-ndhA-ndhH</i>	<i>nad5</i>
14	65137	65375	245	<i>petA</i>	<i>nad1</i>
15	31892	32085	197	<i>trnDGUC</i>	<i>nad1/trnD-GTC</i>
16	47055	47223	171	<i>trnSGGA</i>	<i>trnS-GGA</i>
17	36672	36818	147	<i>psbC</i>	<i>nad1</i>
18	1096	1190	96	<i>psbA</i>	ORF
19	9282	9370	92	<i>trnSGCU</i>	<i>trnS-GGA</i>
20	133140	133220	82	<i>rps12-trnN-GUU</i>	<i>nad1/trnN-GTT</i>
21	111575	111655	82	<i>rps12-trnN-GUU</i>	<i>nad1/trnN-GTT</i>
22	54808	54886	79	<i>TrnM-CAU</i>	<i>nad1/trnM-CAT</i>
23	155661	155735	77	<i>TrnI-CAU</i>	<i>nad1/ccmC/orf</i>
24	89060	89134	77	<i>rps12-trnI-CAU</i>	<i>nad1/ccmC/orf</i>
25	95837	95897	61	<i>rps12-ycf2</i>	<i>nad1/orf</i>
26	148898	148958	61	<i>ycf2</i>	<i>nad1/orf</i>
27	155661	155692	32	<i>trnI-CAU</i>	ORF
28	89103	89134	32	<i>rps12-trnI-CAU</i>	ORF

Ebenaceae family, was well supported based on mt genomes (Figure 7). Consistent with previous comparative genome studies [23, 28, 29], this study also found that the clade united *V. macrocarpon* and *R. simsii* and then formed a sister cluster with the Ebenaceae family with high confidence (bootstrap value of 100%). In general, the phylogenetic tree topology was in line with the evolutionary relationships among those species, indicating the consistency of traditional taxonomy with the molecular classification.

To evaluate selective pressures during the evolutionary dynamics of protein-coding genes among closely related species, the nonsynonymous (Ka) and synonymous (Ks) substitution ratio (Ka/Ks) was calculated. For the Ka/Ks calculation, 28 PCGs from the *D. oleifera* mt genome were compared with the mt genomes of 27 species.

As shown in Figure 8, for the gene-specific substitution rates, Ka/Ks ranged from 0.031 at the *cox1* gene in *V. macrocarpon* to 4.321 at the *atp4* gene in *D. carota*. In 58 cases (except *Glycine max*, *O. sativa*, *Platycodon grandiflorus*, *Scyphiphora hydrophyllacea* and *Z. mays*), the Ka/Ks values of *D. oleifera* gene-specific substitution rates were higher than 1, compared with 22 other species, suggesting positive selection during evolution. Among the 22 species, nine substitution genes with higher Ka/Ks values were found between the *D. oleifera* and *V. vinifera* mt genomes and six genes between the *D. oleifera* and *V. macrocarpon* mt genomes. The *atp4* and *atp8* genes exhibited the highest average rate (1.348 and 0.751) and 15 and 5 Ka/Ks values above 1, respectively, suggested to be the result of positive or relaxed selection [2]. However, most genes had undergone negative selection pressures during evolution, supported by the fact that the Ka/Ks values of 654 proteins, accounting for 91.59% of the proteins

in *D. oleifera*, were less than 1 compared to the other plant species. The *atp1* and *cox1* genes have the smallest average Ka/Ks values (0.212 and 0.272), indicating strong purifying selection [34, 84]. These results show that mt genes are highly conserved during the evolutionary process in green plants.

Because of no mt genome of any species in Ebenaceae has been reported, Synteny of entire mitochondrial genomes was only compared among four Asterids (including three Ericales, one Apiales, and one Solanales) and one Rosids species in this study to assess the degree of structural rearrangement between different lineages. Figure 9 and Figs.1 showed that the rearrangement of mitochondrial genes has been widely occur among these six species, which is accords with many mitogenome observations [20, 37, 38, 39, 41, 42]. When using *D. oleifera* as a reference genome, The dot-plot analyses showed sequences or synteny were seldom shared, and only short stretches of synteny among species (Figs.2). These Large rearrangement events have indicated differentiation within these six species mitogenome. Understandably, species that have close evolutionary relationships share more clusters [20, 41, 42], for example, In general, longer synteny sequences with higher similarity were found between *D. oleifera* and *V. macrocarpon* than that between *D. oleifera* and *M. domestica*.

Among genus *Diospyros*, The most well-known species is *D. kaki*, which has been cultivated as an important fruit crop, due to its edible fruit [25]. However, *D. kaki* are hexaploid ( $2n = 6 \times = 90$ ) or nonaploid ( $2n = 9 \times = 135$ ) and their origin, and polyploidization mechanisms are unclear, which has hampered genome sequencing and molecular breeding [23, 28]. Phylogenetic analyses based on the nuclear [23, 28]

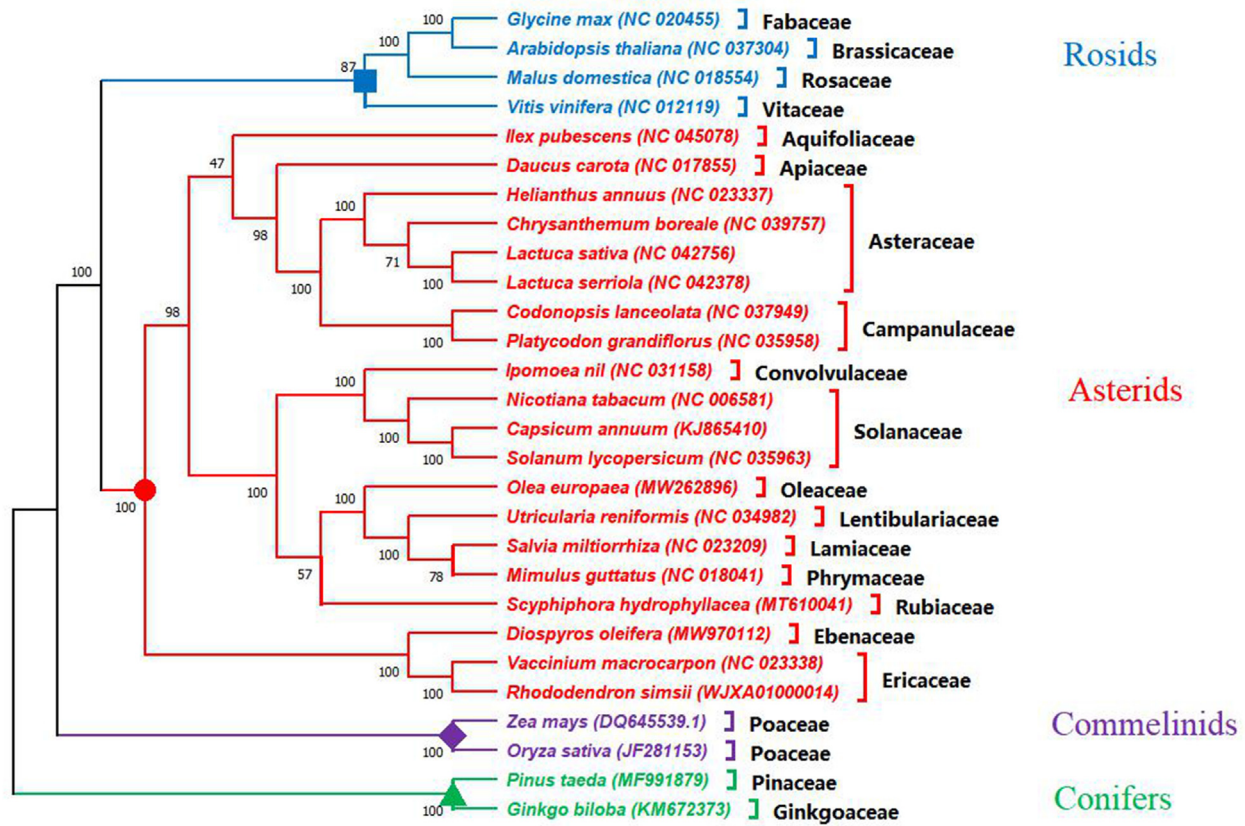


Figure 7. The phylogenetic relationships of *D. oleifera* with other 27 plant species using the maximum likelihood (ML) analysis. The bootstrapping values are listed in each node. The number after the species name is the GenBank accession number. Colors indicate the groups that the specific species belongs.

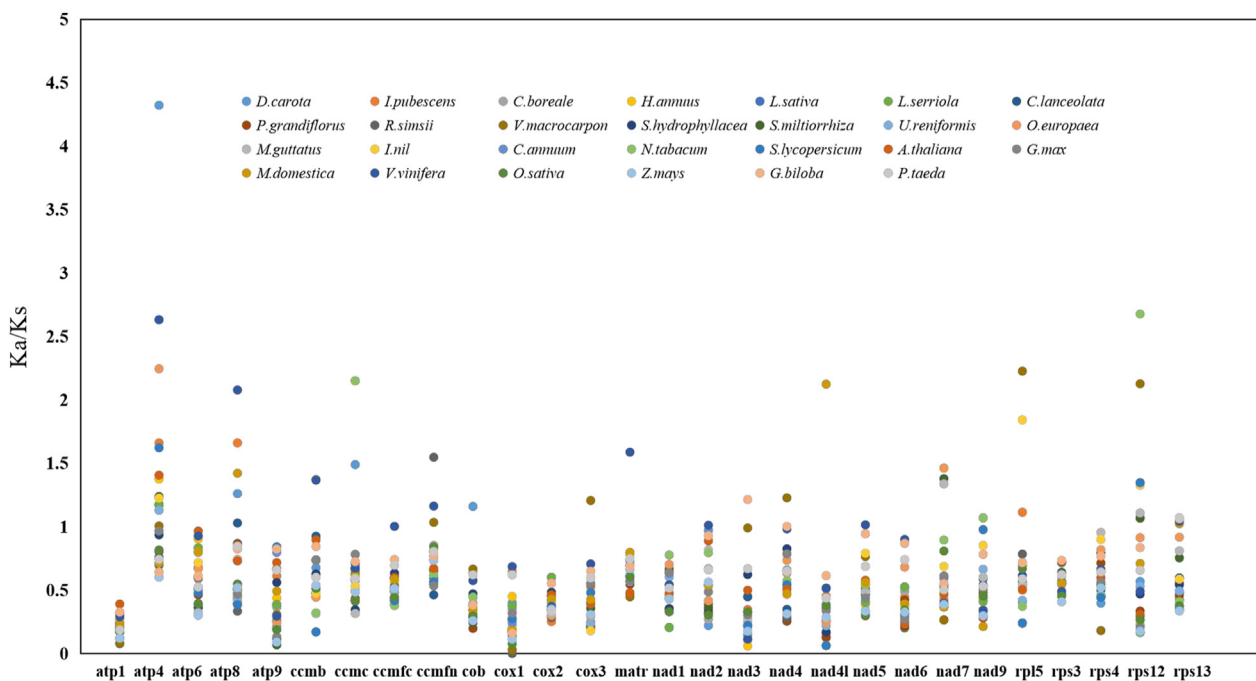
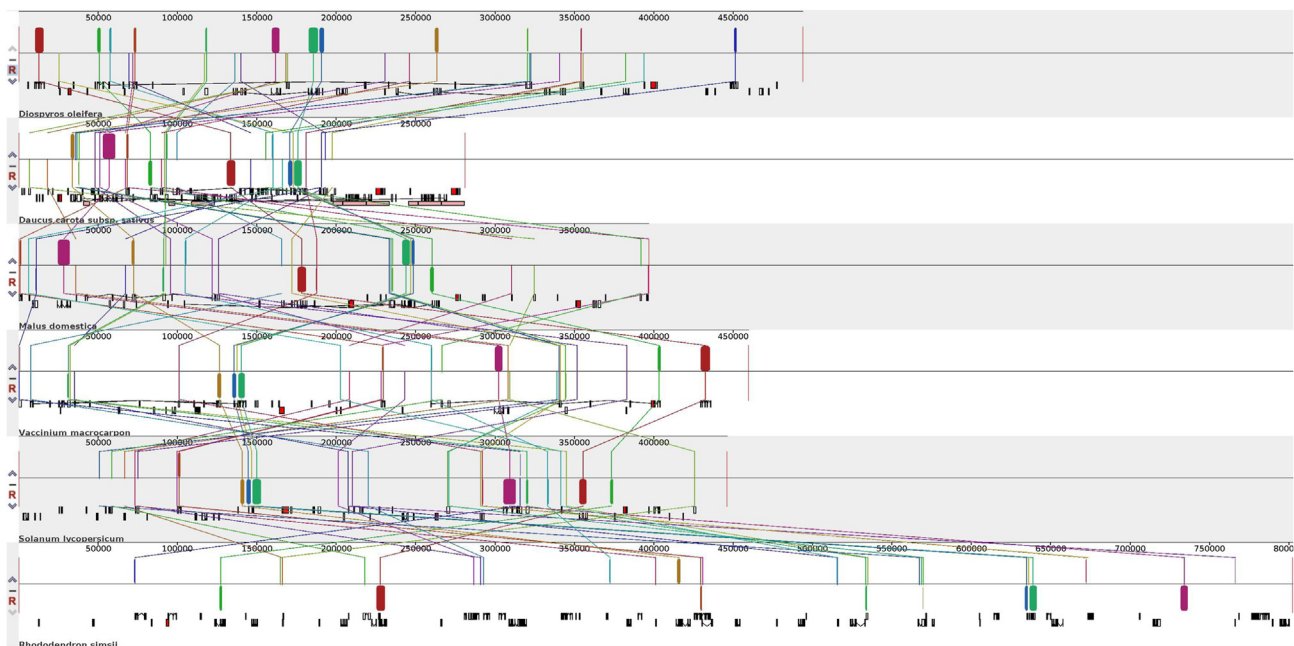


Figure 8. The Ka/Ks values of 28 protein-coding genes of *D. oleifera* versus 27 species.

and chloroplast [26, 27] genome and mtDNA non-coding fragments [85] have indicated that *D. oleifera* is more closely related to *D. kaki* [24] and could be used as a model plant for studies of *Diospyros* [24, 26]. So, as the nuclear and mt genome of hexaploid cultivated persimmon both remains

unpublished, the availability of the *D. oleifera* mt genome provides more alternative comparable reference for *D. kaki* than *D. lotus* does. In addition, our results will lay the foundation for identifying further evolutionary relationships within Ebenaceae. However, due to the lack of



**Figure 9.** Synteny analysis of *D. oleifera* and other five species mitogenomes as generated with Mauve. The sizes and relative positions of the homologous fragments varied across mitogenomes.

adequate representative mitogenomes, more *Ebenaceae* mitogenomes are needed to be sequenced to better resolve the phylogeny and evolutionary biology of Ebenaceae.

#### 4. Conclusions

Here, we presented the first mitochondrial genome assembly and annotation of an Ebenaceae model plant *Diospyros oleifera* as well as the mitochondrial genome in the family Ebenaceae. The mitogenome was 493,958 bp in length, contained 39 protein-coding genes, 27 transfer RNA genes, and 3 ribosomal RNA genes. Comparative analysis of gene structure, codon usage, repeat regions and RNA-editing sites shows that *rps2* and *rps11* genes are missing, and a clear bias of RNA-editing sites is existing in the *D. oleifera* mt genome. In addition, the phenomenon that intracellular tRNA genes transferred frequently from chloroplasts to mitochondria was also observed in *D. oleifera*. Moreover, Phylogenetic analysis based on the mt genomes of *D. oleifera* and 27 other taxa indicates consistency in molecular and taxonomic classification. Furthermore, The  $K_a/K_s$  analysis based on code substitution revealed that most of the coding genes had undergone negative selections, indicating the conservation of mt genes during the evolution. These results will help in better understanding the features of the *D. oleifera* mitochondrial genome and lay the foundation for identifying further evolutionary relationships within Ebenaceae. However, due to the lack of adequate representative mitogenomes, more *Ebenaceae* mitogenomes are needed to be sequenced to better resolve the phylogeny and evolutionary biology of Ebenaceae.

#### Declarations

##### Author contribution statement

Yang Xu: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Yi Dong: Performed the experiments.

Wenqiang Cheng; Haidong Gao, Lei Liu and Lei Xu: Analyzed and interpreted the data.

Kaiyun Wu: Contributed reagents, materials, analysis tools or data.

Bangchu Gong: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

##### Funding statement

Ph.D. Yang Xu was supported by National Key R & D Program of China [2018YFD1000606].

Ph.D. Yang Xu was supported by National Key R & D Program of China [2019YFD1000600].

Bangchu Gong was supported by Key Project for New Agricultural Cultivar Breeding in Zhejiang Province, China [2021C02066-10].

##### Data availability statement

Data associated with this study [The final annotated mt genome sequences of *D. oleifera*] has been deposited at NCBI GenBank under the accession number MW970112.

##### Declaration of interests statement

The authors declare no conflict of interest.

##### Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2022.e09870>.

#### References

- [1] C. Bi, N. Lu, Y. Xu, C. He, Z. Lu, Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches, *Int. J. Mol. Sci.* 21 (11) (2020).
- [2] A. Kozik, B.A. Rowan, D. Lavelle, L. Berke, M.E. Schranz, R.W. Michelmore, A.C. Christensen, The alternative reality of plant mitochondrial DNA: one ring does not rule them all, *PLoS Genet.* 15 (8) (2019), e1008373.
- [3] D. Fajardo, B. Schlautman, S. Steffan, J. Polashock, N. Vorsa, J. Zalapa, The American cranberry mitochondrial genome reveals the presence of selenocysteine (tRNA-Sec and SECIS) insertion machinery in land plants, *Gene* 536 (2) (2014) 336–343.
- [4] M. Vallejo-Marin, A.M. Cooley, M.Y. Lee, M. Folmer, M.R. McKain, J.R. Puzey, Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor, *Am. J. Bot.* 103 (7) (2016) 1272–1288.
- [5] S. Greiner, R. Bock, Tuning a menage a trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants, *Bioessays* 35 (4) (2013) 354–365.
- [6] Y. Ogihara, Y. Yamazaki, K. Murai, A. Kanno, T. Terachi, T. Shiina, N. Miyashita, S. Nasuda, C. Nakamura, N. Mori, S. Takumi, M. Murata, S. Futo, K. Tsunewaki,

- Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome, *Nucleic Acids Res.* 33 (19) (2005) 6235–6250.
- [7] J.N. Timmis, M.A. Aylliffe, C.Y. Huang, W. Martin, Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes, *Nat. Rev. Genet.* 5 (2) (2004) 123–135.
- [8] N. Chevigny, D. Schatz-Daas, F. Lotfi, J.M. Gualberto, DNA repair and the stability of the plant mitochondrial genome, *Int. J. Mol. Sci.* 21 (1) (2020).
- [9] T. Kubo, K.J. Newton, Angiosperm mitochondrial genomes and mutations, *Mitochondrion* 8 (1) (2008) 5–14.
- [10] C. Best, R. Mizrahi, O. Osterseker-Biran, Why so complex? The intricacy of genome structure and gene expression, associated with angiosperm mitochondria, may relate to the regulation of embryo quiescence or dormancy-intrinsic blocks to early plant life, *Plants* 9 (5) (2020).
- [11] A.C. Christensen, Plant mitochondrial genome evolution can be explained by DNA repair mechanisms, *Genome Biol. Evol.* 5 (6) (2013) 1079–1086.
- [12] F. Palumbo, N. Vitulo, A. Vannozzi, G. Magon, G. Barcaccia, The mitochondrial genome assembly of fennel (*Foeniculum vulgare*) reveals two different atp6 gene sequences in cytoplasmic male sterile accessions, *Int. J. Mol. Sci.* 21 (13) (2020).
- [13] E. Skippington, T.J. Barkman, D.W. Rice, J.D. Palmer, Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes, *Proc. Natl. Acad. Sci. U.S.A.* 112 (27) (2015) E3515–E3524.
- [14] D.B. Sloan, A.J. Alverson, J.P. Chuckalovcak, M. Wu, D.E. McCauley, J.D. Palmer, D.R. Taylor, Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates, *PLoS Biol.* 10 (1) (2012), e1001241.
- [15] G. Petersen, A. Cuenca, I.M. Moller, O. Seberg, Massive gene loss in mistletoe (*Viscum, Viscaceae*) mitochondria, *Sci. Rep.* 5 (2015), 17588.
- [16] V. Knoop, The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective, *Curr. Genet.* 46 (3) (2004) 123–139.
- [17] S. O'Conner, L. Li, Mitochondrial fostering: the mitochondrial genome may play a role in plant orphan gene evolution, *Front. Plant Sci.* 11 (2020), 600117.
- [18] L. Yan, W. Xu, D. Zhang, J. Li, Comparative analysis of the mitochondrial genomes of flesh flies and their evolutionary implication, *Int. J. Biol. Macromol.* 174 (2021) 385–391.
- [19] R. Zhang, J. Li, S. Geng, J. Yang, X. Zhang, Y. An, C. Li, H. Cui, X. Li, Y. Wang, The first mitochondrial genome for Phaidoptera (Lepidoptera) with phylogenetic analyses of Zygaenoidea, *Int. J. Biol. Macromol.* 149 (2020) 951–961.
- [20] K. Tyagi, V. Kumar, N. Poddar, P. Prasad, I. Tyagi, S. Kundu, K. Chandra, The gene arrangement and phylogeny using mitochondrial genomes in spiders (Arachnida: araneae), *Int. J. Biol. Macromol.* 146 (2020) 488–496.
- [21] A. Sharma, C. Siva, S. Ali, P.K. Sahoo, R. Nath, M.A. Laskar, D. Sarma, The complete mitochondrial genome of the medicinal fish, *Cyprinus semiplotum*: insight into its structural features and phylogenetic implications, *Int. J. Biol. Macromol.* 164 (2020) 939–948.
- [22] S. Duangjai, B. Wallnofer, R. Samuel, J. Munzinger, M.W. Chase, Generic delimitation and relationships in *Ebenaceae sensu lato*: evidence from six plastid DNA regions, *Am. J. Bot.* 93 (12) (2006) 1808–1827.
- [23] Q.G. Zhu, Y. Xu, Y. Yang, C.F. Guan, Q.Y. Zhang, J.W. Huang, D. Grierson, K.S. Chen, B.C. Gong, X.R. Yin, The persimmon (*Diospyros oleifera* Cheng) genome provides new insights into the inheritance of astringency and ancestral evolution, *Hortic. Res.* 6 (2019) 138.
- [24] S. Kanzaki, The origin and cultivar development of Japanese Persimmon (*Diospyros kaki* Thunb.), *J. Jpn. Soc. Food Sci. Technol.* 63 (2016) 328–330.
- [25] R.Z. Wang, Y. Yang, G.C. Li, Chinese persimmon germplasm resources, *Acta Hort.* 20 (3) (1997) 43–50.
- [26] J. Fu, H. Liu, J. Hu, Y. Liang, J. Liang, T. Wuyun, X. Tan, Five complete chloroplast genome sequences from *Diospyros*: genome organization and comparative analysis, *PLoS One* 11 (7) (2016), e0159566.
- [27] W. Li, Y. Liu, Y. Yang, X. Xie, Y. Lu, Z. Yang, X. Jin, W. Dong, Z. Suo, Interspecific chloroplast genome sequence diversity and genomic resources in *Diospyros*, *BMC Plant Biol.* 18 (1) (2018) 210.
- [28] Y. Suo, P. Sun, H. Cheng, W. Han, S. Diao, H. Li, Y. Mai, X. Zhao, F. Li, J. Fu, A high-quality chromosomal genome assembly of *Diospyros oleifera* Cheng, *GigaScience* 9 (1) (2020).
- [29] T. Akagi, K. Shirasawa, H. Nagasaki, H. Hirakawa, R. Tao, L. Comai, I.M. Henry, The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants, *PLoS Genet.* 16 (2) (2020), e1008566.
- [30] W. Mao, G. Yao, S. Wang, L. Zhou, G. Hu, Chromosome-level genomes of seeded and seedless date plum based on third-generation DNA sequencing and hi-c analysis, *Forest. Res.* 1 (2021) 9.
- [31] S. Dong, L. Chen, Y. Liu, Y. Wang, S. Zhang, L. Yang, X. Lang, S. Zhang, The draft mitochondrial genome of *Magnolia biondii* and mitochondrial phylogenomics of angiosperms, *PLoS One* 15 (4) (2020), e0231020.
- [32] S.L. Kan, T.T. Shen, P. Gong, J.H. Ran, X.Q. Wang, The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome, *BMC Evol. Biol.* 20 (1) (2020) 10.
- [33] H.C. Rawal, P.M. Kumar, B. Bera, N.K. Singh, T.K. Mondal, Decoding and analysis of organelle genomes of Indian tea (*Camellia assamica*) for phylogenetic confirmation, *Genomics* 112 (1) (2020) 659–668.
- [34] X. Wang, F. Cheng, D. Rohlsen, C. Bi, C. Wang, Y. Xu, S. Wei, Q. Ye, T. Yin, N. Ye, Organellar genome assembly methods and comparative analysis of horticultural plants, *Hortic. Res.* 5 (2018) 3.
- [35] Z. Xu, Y. Hao, Y. Xu, Characterization of the complete mitochondrial genome of *lex pubescens*, *Mitochondrial. DNA. Resour. B.* 4 (1) (2019) 2003–2004.
- [36] X. Yin, Y. Gao, S. Song, D. Hassani, J. Lu, Identification, characterization and functional analysis of grape (*Vitis vinifera* L.) mitochondrial transcription termination factor (mTERF) genes in responding to biotic stress and exogenous phytohormone, *BMC Genom.* 22 (1) (2021) 136.
- [37] L. Ren, X. Zhang, Y. Li, Y. Shang, S. Chen, S. Wang, Y. Qu, J. Cai, Y. Guo, Comparative analysis of mitochondrial genomes among the subfamily Sarcophaginae (Diptera: sarcophagidae) and phylogenetic implications, *Int. J. Biol. Macromol.* 161 (2020) 214–222.
- [38] X.-D. Xu, J.-Y. Guan, Z.-Y. Zhang, Y.-R. Cao, K.B. Storey, D.-N. Yu, J.-Y. Zhang, Novel tRNA gene rearrangements in the mitochondrial genomes of praying mantises (Mantodea: mantidae): translocation, duplication and pseudogenization, *Int. J. Biol. Macromol.* 185 (2021) 403–411.
- [39] Z. Chen, Y. Liu, Y. Wu, F. Song, W. Cai, H. Li, Novel tRNA gene rearrangements in the mitochondrial genome of *Camarochiloides weiwii* (Hemiptera: pachynomidae), *Int. J. Biol. Macromol.* 165 (2020) 1738–1744.
- [40] Z. Feng, Y. Wu, C. Yang, X. Gu, J.J. Wilson, H. Li, W. Cai, H. Yang, F. Song, Evolution of tRNA gene rearrangement in the mitochondrial genome of *ichneumonoid wasps* (Hymenoptera: ichneumonoidea), *Int. J. Biol. Macromol.* 164 (2020) 540–547.
- [41] Y. Wu, H. Yang, Z. Feng, B. Li, W. Zhou, F. Song, H. Li, L. Zhang, W. Cai, Novel gene rearrangement in the mitochondrial genome of *Pachyneuron aphidis* (Hymenoptera: pteromalidae), *Int. J. Biol. Macromol.* 149 (2020) 1207–1212.
- [42] M. Monnens, S. Thijs, A.G. Briscoe, M. Clark, E.J. Frost, D.T.J. Littlewood, M. Sewell, K. Smeets, T. Artois, M.P.M. Vanhove, The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of atp8 and genome plasticity in flatworms, *Int. J. Biol. Macromol.* 162 (2020) 454–469.
- [43] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics* 34 (17) (2018) 1884–1890.
- [44] T.K. Mondal, H.C. Rawal, S. Chowrasia, D. Varshney, A.K. Panda, A. Mazumdar, H. Kaur, K. Gaikwad, T.R. Sharma, N.K. Singh, Draft genome sequence of first monocot-halophytic species *Oryza coarctata* reveals stress-specific genes, *Sci. Rep.* 8 (1) (2018), 13698.
- [45] S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, N.H. Bergman, A.M. Phillippy, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation, *Genome Res.* 27 (5) (2017) 722–736.
- [46] J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: a fast and efficient genome polishing tool for long-read assembly, *Bioinformatics* 36 (7) (2020) 2253–2255.
- [47] M. Tillich, P. Lehwar, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq - versatile and accurate annotation of organelle genomes, *Nucleic Acids Res.* 45 (W1) (2017) W6–W11.
- [48] P.P. Chan, T.M. Lowe, tRNAscan-SE: searching for tRNA genes in genomic sequences, *Methods Mol. Biol.* 1962 (2019) 1–14.
- [49] M. Lohse, O. Drechsel, R. Bock, OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes, *Curr. Genet.* 52 (5-6) (2007) 267–274.
- [50] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, *J. Mol. Evol.* 41 (1995) 353–358.
- [51] R. Lorenz, S.H. Bernhart, C. Honer Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA package 2.0, *Algorithm Mol. Biol.* 6 (2011) 26.
- [52] J.P. Mower, PREP-Mt: predictive RNA editor for plant mitochondrial genes, *BMC Bioinf.* 6 (2005) 96.
- [53] P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences, *Biotechniques* 28 (6) (2000) 1102–1104.
- [54] Lee, Python implementation of codon adaptation index, *J. Open. Source. Softw.* 3 (2018) 905.
- [55] T. Thiel, W. Michalek, R.K. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* 106 (3) (2003) 411–422.
- [56] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, *Nucleic Acids Res.* 27 (2) (1999) 573–580.
- [57] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Mol. Biol. Evol.* 30 (4) (2013) 772–780.
- [58] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, *Bioinformatics* 14 (9) (1998) 817–818.
- [59] D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, *Dev. Reprod. Biol.* 8 (1) (2010) 77–80.
- [60] A.C. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, *Genome Res.* 14 (2004) 1394–1403.
- [61] M. Causse, J. Giovannoni, M. Bouzayen, M. Zouine, The chloroplast and mitochondrial genomes of tomato, in: M. Causse, J. Giovannoni, M. Bouzayen, M. Zouine (Eds.), *Compendium of Plant Genomes*, Springer, Berlin, Germany, 2016, pp. 111–137. Chapter 7.
- [62] H. Chen, C. Liu, The chloroplast and mitochondrial genomes of *Salvia miltiorrhiza*, in: S. Lu (Ed.), *Compendium of Plant Genomes*, Springer, Berlin, Germany, 2019, pp. 55–68 (2019).
- [63] Y.D. Jo, Y. Choi, D.H. Kim, B.D. Kim, B.C. Kang, Extensive structural variations between mitochondrial genomes of CMS and normal peppers (*Capsicum annuum* L.) revealed by complete nucleotide sequencing, *BMC Genom.* 15 (2014) 561.
- [64] J. Xu, H. Luo, S. Nie, R.G. Zhang, J.F. Mao, The complete mitochondrial and plastid genomes of *Rhododendron simsii*, an important parent of widely cultivated azaleas, *Mitochondrial DNA B Resour.* 6 (3) (2021) 1197–1199.

- [65] F. Nardi, A. Carapelli, J.L. Boore, G.K. Roderick, R. Dallai, F. Frati, Domestication of olive fly through a multi-regional host shift to cultivated olives: comparative dating using complete mitochondrial genomes, *Mol. Phylogenet. Evol.* 57 (2) (2010) 678–686.
- [66] M. Iorizzo, D. Senalik, M. Szklarczyk, D. Grzebelus, D. Spooner, P. Simon, De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome, *BMC Plant Biol.* 12 (2012) 61.
- [67] V.V. Goremykin, P.J. Lockhart, R. Viola, R. Velasco, The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants, *Plant J.* 71 (4) (2012) 615–626.
- [68] Y. Sugiyama, Y. Watase, M. Nagase, N. Makita, S. Yagura, A. Hirai, M. Sugiura, The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants, *Mol. Genet. Genom.* 272 (6) (2005) 603–615.
- [69] P.F. Agris, E.R. Erusyal, A. Narendran, V.Y.P. Väre, S. Vangaveti, S.V. Ranganathan, Celebrating wobble decoding: half a century and still much is new, *RNA Biol.* 15 (2018) 537–553.
- [70] P.F. Agris, F.A.P. Vendeix, W.D. Graham, tRNA's wobble decoding of the genome: 40 years of modification, *J. Mol. Biol.* 366 (2007) 1–13.
- [71] C.J. Grassa, D.P. Ebert, N.C. Kane, L.H. Rieseberg, Complete mitochondrial genome sequence of sunflower (*Helianthus annuus* L.), *Genome Announc.* 4 (5) (2016).
- [72] Y. Xu, W. Cheng, C. Xiong, X. Jiang, K. Wu, B. Gong, Genetic diversity and association analysis among germplasms of *Diospyros kaki* in Zhejiang Province based on SSR markers, *Forest* 12 (4) (2021) 422.
- [73] Y. Cheng, X. He, S. Priyadarshani, Y. Wang, L. Ye, C. Shi, K. Ye, Q. Zhou, Z. Luo, F. Deng, L. Cao, P. Zheng, M. Aslam, Y. Qin, Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*, *BMC Genom.* 22 (1) (2021) 167.
- [74] D. Pinard, A.A. Myburg, E. Mizrahi, The plastid and mitochondrial genomes of *Eucalyptus grandis*, *BMC Genom.* 20 (1) (2019) 132.
- [75] A.J. Alverson, X. Wei, D.W. Rice, D.B. Stern, K. Barry, J.D. Palmer, Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae), *Mol. Biol. Evol.* 27 (6) (2010) 1436–1448.
- [76] A.A. Edera, M.V. Sanchez-Puerta, Computational detection of plant RNA editing events, *Methods Mol. Biol.* 2181 (2021) 13–34.
- [77] P. Robles, V. Quesada, Organelle genetics in plants, *Int. J. Mol. Sci.* 22 (4) (2021).
- [78] L. Verhage, Targeted editing of the Arabidopsis mitochondrial genome, *Plant J.* 104 (6) (2020) 1457–1458.
- [79] S.C. Straub, R.C. Cronn, C. Edwards, M. Fishbein, A. Liston, Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae), *Genome Biol. Evol.* 5 (10) (2013) 1872–1885.
- [80] K.L. Adams, Y.L. Qiu, M. Stoutemyer, J.D. Palmer, Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution, *Proc. Natl. Acad. Sci. U.S.A.* 99 (15) (2002) 9905–9912.
- [81] S. Asaf, A.L. Khan, A.R. Khan, M. Waqas, S.M. Kang, M.A. Khan, R. Shahzad, C.W. Seo, J.H. Shin, L.J. Lee, Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species, *PLoS One* 11 (4) (2016), e0152937.
- [82] A.O. Richardson, J.D. Palmer, Horizontal gene transfer in plants, *J. Exp. Bot.* 58 (1) (2007) 1–9.
- [83] U. Bergthorsson, K.L. Adams, B. Thomason, J.D. Palmer, Widespread horizontal transfer of mitochondrial genes in flowering plants, *Nature* 424 (6945) (2003) 197–201.
- [84] J.C. Fay, C.I. Wu, Sequence divergence, functional constraint, and selection in protein evolution, *Annu. Rev. Genom. Hum. Genet.* 4 (2003) 213–235.
- [85] D.C. Hu, Z.R. Luo, Polymorphisms of amplified mitochondrial DNA non-coding regions in *Diospyros* spp, *Sci. Hortic.* 109 (3) (2006) 275–281.