



OPEN ACCESS

EDITED BY

Xuming Li,
Hugo Biotechnologies Co., Ltd., China

REVIEWED BY

Micai Zhong,
Chinese Academy of Sciences, China
Yue Jing,
South China Agricultural University, China
Fuyan Liu,
Beijing Genomics Institute (BGI), China

*CORRESPONDENCE

Junqin Zhou

✉ zhoujunqin@csuft.edu.cn

Xiaofeng Tan

✉ T19781103@csuft.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 06 March 2024

ACCEPTED 31 July 2024

PUBLISHED 03 September 2024

CITATION

Gu Y, Yang L, Zhou J, Xiao Z, Lu M, Zeng Y and Tan X (2024) Mitochondrial genome study of *Camellia oleifera* revealed the tandem conserved gene cluster of *nad5–nads* in evolution. *Front. Plant Sci.* 15:1396635. doi: 10.3389/fpls.2024.1396635

COPYRIGHT

© 2024 Gu, Yang, Zhou, Xiao, Lu, Zeng and Tan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mitochondrial genome study of *Camellia oleifera* revealed the tandem conserved gene cluster of *nad5–nads* in evolution

Yiyang Gu^{1,2†}, Liying Yang^{1,2,3†}, Junqin Zhou^{1,2,4*}, Zhun Xiao^{1,2,5}, Mengqi Lu^{1,2}, Yanling Zeng^{1,2} and Xiaofeng Tan^{1,2*}

¹Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees, Ministry of Education, Central South University of Forestry and Technology, Changsha, China, ²Academy of Camellia Oil Tree, Central South University of Forestry and Technology, Changsha, China, ³Hunan Horticulture Research Institute, Hunan Academy of Agricultural Sciences, Changsha, China, ⁴College of Landscape Architecture, Central South University of Forestry and Technology, Changsha, China, ⁵School of Foreign Languages, Changsha Social Work College, Changsha, China

Camellia oleifera is a kind of high-quality oil supply species. Its seeds contain rich unsaturated fatty acids and antioxidant active ingredients, which is a kind of high-quality edible oil. In this study, we used bioinformatics methods to decipher a hexaploid *Camellia* oil tree's mitochondrial (mt) genome based on second-generation sequencing data. A 709,596 bp circular map of *C. oleifera* mt genome was found for the first time. And 74 genes were annotated in the whole genome. Mt genomes of *C. oleifera* and three Theaceae species had regions with high similarity, including gene composition and gene sequence. At the same time, five conserved gene pairs were found in 20 species. In all of the mt genomes, most of *nad* genes existed in tandem pairs. In addition, the species classification result, which, according to the gene differences in tandem with *nad5* genes, was consistent with the phylogenetic tree. These initial results provide a valuable basis for the further researches of *Camellia oleifera* and a reference for the systematic evolution of plant mt genomes.

KEYWORDS

Camellia oleifera, mitochondrial genome, *Camellia*, gene cluster, *nad5–nads* tandem genes

Abbreviations: mt, mitochondrial; ORF, open reading frame; tRNA, transfer RNA; rRNA, ribosomal RNA; SSRs, Simple repeats.

1 Introduction

Camellia oleifera Abel., originating in China, is a non-wood forest that blooms and harvests in winter. *C. oleifera* is a kind of high-quality woody edible oil tree in Theaceae, which is famous for its rich unsaturated fatty acids in fruit (Tan, 2013). *Camellia* oil tree not only contains unsaturated fatty acids up to 90% but also is rich in antioxidant active ingredients such as squalene, sterol, vitamin E and polyphenols (He and He, 2002). *C. oleifera* with *Elaeis guineensis* Jacq., *Olea europaea* L. and *Cocos nucifera* L. are known as the world's four famous oil species. In recent years, in China, with the support of national policies, such as sustainable development, green development and targeted poverty alleviation, the development of *C. oleifera* has been widely supported due to its high economic value, wide use, comprehensive development, and utilization potential. *C. oleifera* is called "iron crop" and "green oil reservoir" by farmers (Wu et al., 2018). Studies on the development characteristics of *C. oleifera* fruits (Zhang et al., 2021), self-incompatibility of flowers (Zhou et al., 2020; Lu et al., 2021), and other economic potential exploitation have enabled researchers to have a further understanding of *C. oleifera*. With continuous progress of various technologies, plant genome research has become one of the most popular research topics. For researchers, the genome is a reference book for understanding a species or an organelle. At present, the chloroplast genome of *C. oleifera* has been reported (Wu et al., 2020), and the nuclear genome is also being carried out gradually. *C. oleifera* mitochondrial (mt) genome will also be the focus of the next step.

Mitochondria are important organelles in eukaryotic cells, which are closely related to energy generation, fatty acid synthesis, and active protein synthesis (Niu et al., 2022). Mitochondria are semi-autonomous organelles that encode their own functional genes and are regulated by nuclear genes. The mt genome of plants is more complex than that of animals and single-celled eukaryotes. The mt genomes of plants have different structures, gene contents, and DNA mutation rate strategies to meet their specific needs for physical, photosynthetic, or physiological and biochemical functions (Cui et al., 2009). Concurrently, cytoplasmic male sterility and species evolution in plants are closely related to the mitochondria (Wang et al., 2020). More than 200 plant mt genomes have been published, and most of them are crop species. However, there are only two articles on mt genome of Theaceae plants (Zhang et al., 2019; Rawal et al., 2020). The mt genome sequence is not as conserved as the chloroplast genome sequence, but some gene families are also conserved in the mt genome. There may even be concatenation of two genes or a group of genes. In Niu's study, two gene clusters (*rps12-nad3* and *rps3-rpl16*) were present in eight mitochondrial genomes (Niu et al., 2022). This is actually one of the dependencies on which all genomes can be assembled after sequencing.

With the development of sequencing technology, longer and more accurate reads can be obtained from the sequencing result, which enables many complex plant genomes to be gradually broken down. The sequencing results of one species include not only nuclear genome sequences but also mt genome sequences. Based on this, strategies and software extraction and assembly of mt

genome sequences from whole-genome data are developing (King et al., 2014; Al-Nakeeb et al., 2017). MIA software (Green et al., 2008), for example, is an early mt genome assembly software. It takes mt genome sequences from the ancient human whole-genome sequence data and assembles them together. After then, software, such as MITObim (Hahn et al., 2013), ARC (Hunter et al., 2015), and NOVOPlasty (Dierckxsens et al., 2017), have been developed for every species other than people. Burns et al. (2015) also used MITObim to assemble the mt genome of *Cymbomonas tetramitiformis* when assembling its nuclear genome. Therefore, the genomes of complex species assembled from sequencing data can be used as a kind of data reference for research.

Therefore, the mt genome of *C. oleifera* was selected as the research object in this study. After obtaining all related reads from the sequencing data, the mt genome of *C. oleifera* was assembled by the MITObim software (v4.0.2) (Hahn et al., 2013). According to the sequence information of the genome, its structural characteristics and the situation of coding genes were annotated. The successful analysis of mt genome enables researchers to have a further understanding of *C. oleifera* and focus on the study of biological issues with more time and more energy.

2 Materials and methods

2.1 Materials and data background

The plant material *Camellia oleifera* cv. Huashuo was planted in the experimental field in Wangcheng district, Changsha, Hunan province (28°05'N, 113°2'E). Young leaves of adult *C. oleifera* trees were collected, preserved with liquid nitrogen, and sent to BGI Genomics for sequencing. The BGISEQ 500 Platform was used to build the second-generation sequencing database. The original data used in the experiment was extracted from the next-generation sequencing datasets of *Camellia oleifera*. All reads were trimmed and error corrected and then were provided for the assembly of mt genome of *C. oleifera* as data source.

2.2 Genome assembly

MITObim v4.0.2 (Hahn et al., 2013) was used for *C. oleifera* mt genome assembly. It can directly assemble the mt genomes of non-model species from DNA sequencing data embedded with MIRA and IMAGE modules. The mt genomes of *Camellia sinensis* var. *assamica* cv. Yunkang10 (YK10) (Zhang et al., 2019) and *Camellia sinensis* var. *assamica* cv. TV-1 (TV-1) (Rawal et al., 2020) were used as reference. All the paired-end sequencing reads of *C. oleifera* were mapped to the two mitochondrial reference sequences by BWA v0.7.12 (bwa index -p ref reference.fa) (Li and Durbin, 2009). According to the similarity of sequences, reads needed for mitochondrial assembly were captured preliminarily. Then, Samtools v1.9 was used to extract reads that were paired aligned to the reference sequence (bwa mem -t 16 ref.fa F.fq.gz R.fq.gz | samtools view -bF 12 | samtools sort -@ 16 -m 1G -o output.bam) (samtools view -h output.bam | tail -n +4 | cut -f 1 > map_reads.txt).

All reads were integrated to a new sequence file. This file was used as the input file of MITObim to assemble the mt genome of *C. oleifera*. Concurrently, *Camellia sinensis* mitochondrial sequences (Zhang et al., 2019; Rawal et al., 2020) were used as the main reference sequence in the assembly process.

2.3 Genome annotation and visualization

The GeSeq (Tillich et al., 2017) tool was used to annotate the mt genome of *C. oleifera*. Protein-coding genes, transfer RNA (tRNA), and ribosomal RNA (rRNA) genes were annotated by BLAST with the existing plant mt genome data in NCBI database (<https://www.blast.ncbi.nlm.nih.gov>), including *Camellia sinensis* var. *assamica* cv. Yunkang10 (MK574876.1) (Zhang et al., 2019), *Camellia sinensis* var. *assamica* cv. TV-1 (NC_043914.1) (Rawal et al., 2020), *Vitis vinifera* (NC_012119.1) (Goremykin et al., 2009), *Triticum aestivum* (NC_036024.1) (Cui et al., 2009), *Oryza sativa* subsp. *indica* (NC_007886.1) (Tian et al., 2006), *Zea mays* subsp. *mays* (NC_007982.1) (Clifton et al., 2004), *Glycine max* (NC_020455.1) (Chang et al., 2013), *Gossypium arboreum* (NC_035073.1) (Chen et al., 2017), *Ziziphus jujuba* (NC_029809.1), *Bupleurum falcatum* (NC_035962.1), *Boechera stricta* (NC_042143.1) (Li et al., 2018), and others. Concurrently, tRNA genes in the mt genome were annotated again with tRNA scan-SE tool (Lowe and Eddy, 1997). The annotated genes with coverage and match less than 60% were manually eliminated, and the repeated annotation results were compared. The final annotation results were drawn by Draw Organelle Genome Maps (OGDRAW v1.3.1) tool (Greiner et al., 2019) with a circular map and a linear map.

2.4 Repeat analysis and RNA-editing site prediction

The repeat sequence detection in the mt genome of *C. oleifera* was carried out by MISA v2.1 (Beier et al., 2017) and REPuter tools (Kurtz et al., 2001). Simple repeats (SSR) were verified by MISA, with the minimum number of nucleotide repeats setting as 8, 4, 4, 3, 3, and 3 for monomer, dimer, trimer, tetramer, pentamer, and hexamer, respectively (Zhang et al., 2019). At the same time, incomplete repeats of SSR interrupted by a few bases (spacing less than 100 or equal to 100) were screened, identified, and located. Forward and palindromic repeats are confirmed by REPuter with a minimum length of 50 nt and a minimum fault tolerance of 8 nt. The RNA-editing site was predicted by the PREP-Mt web tool (Mower, 2005). By reading the location information of genes in the annotation file, the base sequences of protein-coding genes were obtained from the genome. The gene sequence file was adjusted according to the format required by the tool, and the threshold was set at 0.2.

2.5 Phylogenetic analysis

The phylogenetic tree was constructed by IQ-Tree software (Nguyen et al., 2015), carried by TBtools (Chen et al., 2020).

Published mt genome data of 19 species were selected from the NCBI database, including 16 dicotyledons as follows: *Arabidopsis thaliana* (Ath, NC_037304.1) (Unsel et al., 1997), *Brassica napus* (Bna, NC_008285.1) (Handa, 2003), *Bupleurum falcatum* (Bfa, NC_035962.1), *Camellia gigantocarpa* (Cgi, OP270590) (Lu et al., 2022), *Camellia sinensis* (Csi, MK574876.1; NC_043914.1) (Zhang et al., 2019; Rawal et al., 2020), *Capsicum annuum* (Can, NC_024624.1), *Carica papaya* (Cpa, NC_012116.1) (Magee et al., 2010), *Glycine max* (Gma, NC_020455.1) (Chang et al., 2013), *Gossypium arboreum* (Gar, NC_035073.1) (Chen et al., 2017), *Gossypium barbadense* (Gba, NC_028254.1) (Tang et al., 2015), *Malus domestica* (Mdo, NC_018554.1) (Goremykin et al., 2012), *Nicotiana tabacum* (Nta, NC_006581.1) (Sugiyama et al., 2005), *Rhazya stricta* (Rst, NC_024293.1) (Park et al., 2014), *Spinacia oleracea* (Sol, NC_035618.1) (Cai et al., 2017), and *Vitis vinifera* (Vvi, NC_012119.1) (Goremykin et al., 2009), two monocots: *Cocos nucifera* (Cnu, NC_031696.1) and *Triticum aestivum* (Tae, NC_036024.1) (Cui et al., 2009), and one gymnosperm: *Ginkgo biloba* (Gbi, NC_027976.1). These include two published species of the genus *Camellia*. The annotation information of all species was compared manually, including the *Camellia oleifera*. The sequences of 15 conserved genes were extracted, respectively, and aligned using Muscle v5 (Edgar, 2021) with default parameters. After integrating the comparison results, the portable IQ-tree software carried by TBtools (Chen et al., 2020) was used to build the phylogenetic tree, setting the model parameter at Auto and Bootstrap value at 1,000 (iqtree -s TBtools5888937064767651616.tmpIn -pre supergene.fa -bb 1000 -bnni -m MFP -nt AUTO).

2.6 Collinearity analysis

MCSanX (Wang et al., 2012) was used to analyze collinearity among the mt genomes of four Theaceae species. The sequences and GFF (general feature format) annotation files of two *Camellia sinensis* mt genomes were downloaded from the NCBI database (MK574876.1; NC_043914.1) and the data number of *Camellia gigantocarpa* is OP270590. The sequence files and annotation files were modified to meet the input file format of One Step MCSanX software in TBtools (Chen et al., 2020). Running the software with the parameters (-CPU for BlastP 2 -E-value 1e-10 -Num of BlastHits 5) obtained the collinearity analysis results of *C. oleifera* and the other three, respectively. Then, the resulting data files were used as input files of Dual Systemy Plot software to draw the colinear map of mt genomes.

3 Results

3.1 Assembly and annotation

The next-generation sequencing yielded 525 G data files, of which 34,720,712 reads were compared to the reference genome. A 709,596-bp circle mitochondrial map of *C. oleifera* was obtained, and GC content reached 45.33% (Figure 1). A total of 42 protein-coding genes (including *orf*), 29 tRNA genes, and 2 rRNA genes (*rrnL* and *rrnS*)

TABLE 1 Gene content of mt genome with three *Camellia* species.

Category		<i>Camellia oleifera</i> cv. Huashuo	<i>Camellia sinensis</i> var. <i>assamica</i>		<i>Camellia gigantocarpa</i>	
			TV-1	YK10		
Genome size (bp)		709,309	707,441	701,719 + 177,329	970,410	
GC content (%)		45.33	45.75	45.63/45.81	45.00	
Total genes number		74	66	73	66	
Protein-coding genes number	Complex I (NADH dehydrogenase)	35	<i>nad1, nad2*</i> , <i>nad3, nad4*</i> , <i>nad4l, nad5*</i> , <i>nad6</i>	31	44	44
	Complex II (succinate dehydrogenase)		<i>sdh3*(3x)</i>			
	Complex III (ubiquinol cytochrome c reductase)		<i>cob</i>			
	Complex IV (cytochrome c oxidase)		<i>cox2</i>			
	Complex V (ATP synthase)		<i>atp1, atp4, atp6, atp9 (2x)</i>			
	Cytochrome c biogenesis		<i>ccmB, ccmC, ccmFc*</i> , <i>ccmFn</i>			
	Ribosomal proteins (SSU)		<i>rps1, rps4(2x), rps7, rps12, rps13(2x), rps14, rps19(2x)</i>			
	Ribosomal proteins (LSU)		<i>rpl5, rpl10</i>			
	Maturases		<i>matR</i>			
	Transport membrane proteins		<i>mttB</i>			
Transfer RNAs		29	<i>trnA-UGC*</i> , <i>trnC-GCA, trnD-GUC (2x), trnF-AAA*</i> , <i>trnF-GAA, trnK-UUU, trnM-CAU (5x), trnN-AUU, trnN-GUU (2x), trnP-UGG, trnS-CGA, trnS-GCU, trnS-UGA* (3x), trnV-GAC, trnY-GUA, trnI-GAU* (2x), trnnull-NNN (4x)</i>	29	24	19
Other genes		8	<i>orf100, orf101a, orf102 (2x), orf104, orf104a, orf106a, orf145</i>	4	2	/
Ribosomal RNAs		2	<i>rnlL, rns</i>	2	3	3
Total SSRs		530	529	665	746	
RNA-editing sites		413/35	429/31	478/44	483/44	

*Gene with introns. Numbers of copies are shown in parenthesis for genes with multiple copies.

least (2) (Supplementary Table 3). By analyzing the relationship between the gene length and the number of RNA-editing sites, it was found that the longer the coding sequence, the more RNA-editing sites. However, there was no absolute linear relationship between them (Figure 2). All the RNA-editing were “C” to “U,” and the number of editing sites at the second base of codon was the highest. There were predicted 267 sites (64.65%) and 20 sites

(4.84%) at second base and first and second base of codon, respectively. Furthermore, no site at third base of codon was predicted (Supplementary Table 4). Among all editing sites, 105 sites (25.42%) enabled serine to convert into leucine, and 89 sites (21.55%) enabled proline to convert into leucine. They account for almost half of the total. However, the conversion of two glutamine and two arginine to terminators was predicted (Table 4).

TABLE 2 Statistics of SSR motifs in the *Camellia oleifera* cv. Huashuo mt genomes.

SSR-motif		Repeats/number	SSR number	SSR %
Monomer	A	8/33, 9/30, 10/4, 11/2, 12/2, 13/1 72)	176	33.2
	C	8/5, 9/3 (8)		
	G	8/14, 9/1 (15)		
	T	8/41, 9/26, 10/8,11/2, 12/2, 15/1, 16/1 (81)		
Dimer	AC	4/5 (5)	237	44.7
	AG	4/41, 5/9, 6/1, 7/2 (53)		
	AT	4/13, 5/4, 6/3, 7/3, 8/1, 9/1 (25)		
	CA	4/1 (1)		
	CG	4/3 (3)		
	CT	4/32, 5/11, 6/2 (45)		
	GA	4/23, 5/9, 6/1 (33)		
	GC	4/2 (2)		
	GT	4/5, 5/1 (6)		
	TA	4/17, 5/6, 6/2 (26)		
	TC	4/26, 5/3, 6/2 (31)		
	TG	4/7 (7)		
Trimer	AAC, AAG, AGC, GCT, TAT, TTG	4/1 (6)	25	4.7
	ACT, CTT, GAA, TTC	4/2 (8)		
	TAA, TTA	4/3 (6)		
	TCT	4/4 (4)		
	GAT	5/1 (1)		
Tetramer	AACA, AACC, AAGC, AAGG, AATA, AATG, ACTA, AGCT, AGTC, ATAA, ATGG, CAAT, CAGA, CCAG, CCGA, CCTT, CGGG, CTAG, CTCC, GAAG, GAGC, GCTT, GGCC, GTGA, TAGA, TAGT, TGAA, TTAC, TTAT, TTCC, TTTG	3/1 (31)	79	14.9
	AAAT, ATAG, ATTC, CAAA, CTTC, CTTG, GAAA, GAAT, GATT, GGAA	3/2 (20)		
	AAGA	3/3 (3)		
	AAAG, AGAA, TCTT	3/4 (12)		
	TTCT	3/5 (5)		
	CTTT	3/6 (6)		
	CCTT, TATC	4/1 (1)		
Pentamer	AATTA, ACTAG, ATAGG, CTAAT, GAAAT, GCCTT, TAAAG, TTAAT, TTATA	3/1 (9)	11	2.1
	TAGAG	3/2 (2)		
Hexamer	CTATCC, TTTCTA	3/1 (2)	2	0.4

3.4 Phylogenetic analysis of mt genomes

A total of 15 conserved genes were found from 20 mt genomes, including the *C.oleifera* mt genomes (*atp1*, *atp6*, *atp9*, *ccmB*, *ccmC*,

ccmFc, *ccmFN*, *cob*, *cox2*, *nad3*, *nad4*, *nad4L*, *rps12*, *rpl5*, and *matR*). The protein sequences of 15 genes were compared and then tandem connected to obtain the super-sequence gene file to construct the phylogenetic tree. The ModelFinder program of IQ-Tree software

TABLE 3 Statistics of long repeat sequences in the *Camellia oleifera* cv. Huashuo mt genomes.

No.	Size (bp)	Palindromic Copy (P)		Size (bp)	Forward Copy (F)	
		Start 1	Start 2		Start 1	Start 2
1	174	17,713	364,406	722	0	87,394
2	162	54,865	684,866	582	727	88,125
3	191	105,254	648,172	433	1,332	88,732
4	281	152,801	709,308	447	1,764	89,165
5	236	153,082	709,073	139	2,244	89,668
6	196	153,452	708,740	487	2,387	89,807
7	322	153,664	708,406	221	2,872	90,293
8	279	154,100	708,013	3,371	3,092	90,514
9	205	154,375	707,811	10,565	6,621	94,050
10	222	154,698	707,473	639	17,179	104,607
11	268	154,920	707,207	4,060	17,884	105,321
12	532	155,184	706,680	406	21,940	109,367
13	202	155,716	706,479	3,366	22,336	109,762
14	414	156,046	705,936	159	25,700	113,113
15	1,108	156,454	704,843	3,883	25,857	113,261
16	585	157,561	704,260	162	54,865	177,382
17	253	158,141	704,013	139	54,957	647,939
18	637	158,392	703,379	171	78,225	409,425
19	2,444	159,024	700,935	302	172,510	681,489
20	2,045	161,607	698,753	151	309,545	309,848
21	2,622	163,649	696,135			
22	310	166,344	695,751			
23	420	166,714	695,271	Long repeat sequences	10,565 * 2 bp	
24	6,180	167,250	688,964			
25	2,080	173,438	686,884	Forward match (F)	20 pairs	
26	1,471	175,651	685,280			
27	2,834	177,197	682,379	Palindromic match (P)	30 pairs	
28	397	330,626	501,796			
29	347	374,491	423,250			
30	302	681,489	689,582			

had tested 546 protein models, and HIVw+F+R3 was selected as the most suitable model. All species were clearly divided into three groups in the phylogenetic tree, including gymnosperm, monocotyledon, and dicotyledon, of which dicotyledon was the main group (Figure 3). *C. oleifera* with two cultivars of *C. sinensis* var. *assamica* (YK10 and TV-1) and *C. gigantocarpa* were grouped together in a branch with a bootstrap value of 100%. Species belonging to the same family were successfully grouped together.

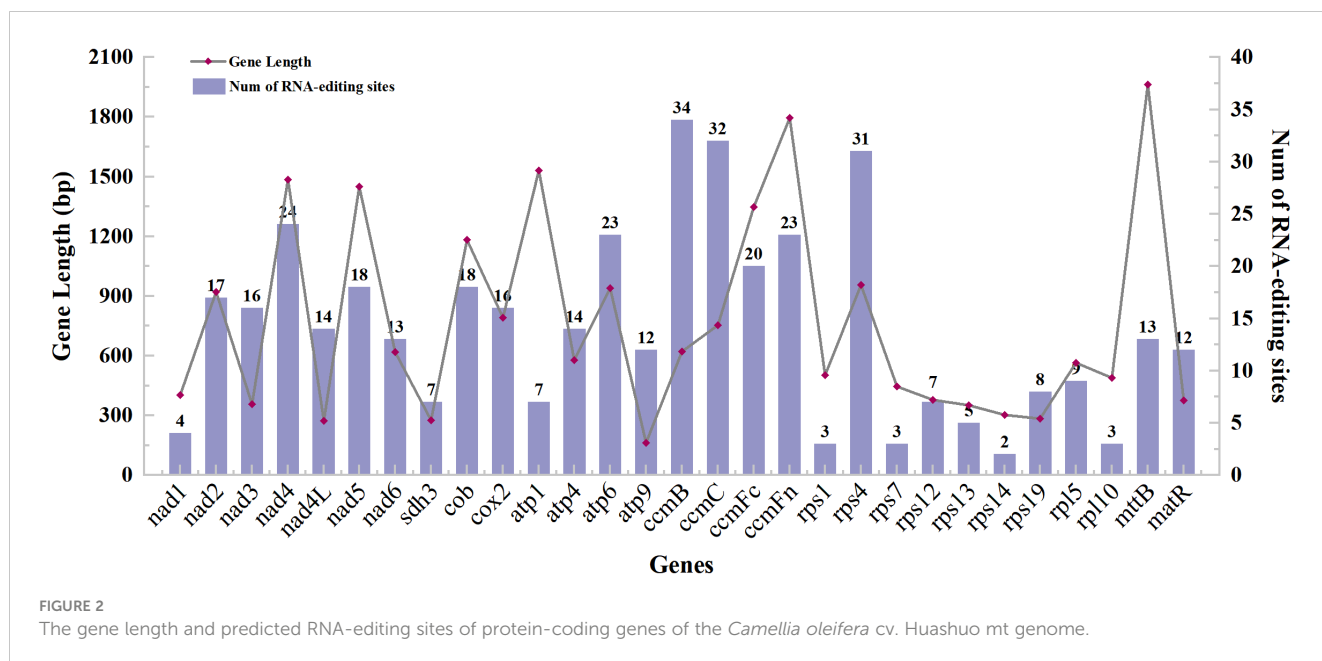
3.5 Comparative analysis with *Camellia* species

At present, only three mt genomes (*C. sinensis* var. *assamica* cv. TV-1, *C. sinensis* cv. Yunkang10, and *C. gigantocarpa*) had been reported in the *Camellia* genus and even the Theaceae family. The mt genome of TV-1 and *C. gigantocarpa* had one single circular map of 707,441 and 970,410 bp, and YK10 had two complete circular maps (701,719 and 177,329 bp). *C. gigantocarpa* has the largest mt genome, and *C. gigantocarpa* and YK10 have the largest number of protein-coding genes (44), followed by *C. oleifera*. It was found that the succinate dehydrogenase group of *C. oleifera* and YK10 had the same gene composition. The NADH dehydrogenase group and ribosomal protein group of *C. oleifera* and TV-1 have the same gene composition. By comparing linear maps of four mt genomes, it was found that the arrangements of some genes in all genomes were similar (Figures 4, 5). The sites selected by the same color boxes in Figure 4 had almost identical gene member, and the sequences of these genes in the four genomes were almost the same. The difference was that the genes were arranged on the sense strand of *C. oleifera*, *C. gigantocarpa*, and TV-1, but on the anti-sense strand of the YK10. For example, the composition of gene clusters *atp1*, *rps7*, *rpl10*, *ccmB*, *mttB*, *atp6* and *nad4* was relatively conserved, and their arrangement was the same on mt genomes of *C. oleifera* and TV-1, but reversed on YK10's. The same was true of the other two gene clusters (*rps13*, *rrnL*, and *cox2*, and *rpl5*, *rps14*, and *cob*).

3.6 Conservation of tandem gene pairs on genomes

To study the conserved gene clusters on the mt genomes of all species, including *Camellia*, we mapped the linear mt genome structure map of 20 species (Figure 6). We found 17 conserved genes on 20 mt genomes, including 15 protein-coding genes and two ribosomal RNAs genes (*rrnL* and *rrnS*). By analyzing the location information and upstream and downstream genes of these 17 genes on the 20 mt genomes, we did not find their arrangement rules on the genome. But back to the location information of all the genes on 20 mt genomes, we found five conserved tandem gene pairs from 20 species, namely, *rps3-rpl6*, *rrn18-rrn5*, *rpl5-cob*, *nad3-rps12*, and *nad1-matR*. Sometimes, *rpl5-rps14-cob* tandem gene pairs existed on the genome instead of *rpl5-cob*.

In all the mt genomes, tandem gene pairs may be located on the positive-sense strand or the antisense strand. The arrangement of some tandem gene pairs was reversed on some mt genomes. For example, on the mt genomes of Bna, Ath, Cpa, and Gba, *nad3-rps2* gene pairs located on the positive-sense strand, but on the anti-sense strand of Vvi, Mdo, and Gma mt genomes. The sequence of *nad1-matR* tandem gene pairs was inversely complementary on the Ath and TV-1 mt genomes. However, some genes in these five tandem gene pairs were missing on the genomes of individual species. For example, *rpl6* is missing on Mdo and Sol mt genomes, and *rrn5* is missing on TV-1 and Col mt genomes. These



differences, such as the number and the type of genes annotated, may be related to the different annotation methods. Therefore, according to the arrangement of these five conserved gene pairs mentioned above, 20 species were divided into six groups. Bna, Ath, Cpa, Gba, and Gar formed group One. Vvi, Mdo, and Gma formed group Two. YK10, TV-1, Cgi, and Col were group Three. Rst, Bfa, Nta, and Can were group Four. Group Five only had Sol. Finally, Tae, Cnu, and Gbi formed group Six. The members of these six groups were consistent with the species branches in the evolutionary tree results. Therefore, these five gene pairs had a certain conservation in the evolutionary process.

3.7 Tandem rule of *nads* on mt genome

When analyzing the number and location information of all gene families on 20 genomes, several *nad-nad* gene pairs were found on each mt genome (Figure 7). For example, on Col, YK10, and TV-1, *nad4* and *nad5* were next to each other and also *nad2* and *nad1*. All *nads* with tandem arrangement in each species were identified, and the rules of *nad-nad* gene tandem were found in the six groups obtained from the previous classification. It was found that *nad1*, *nad5*, *nad6*, and *nad9* genes have the *nad-nad* gene tandem rule in group One, *nad1*, *nad2*, *nad4*, *nad5*, and *nad6* in group Two. In group Four, *nad1*, *nad2*, and *nad5* had the tandem rule and also *nad4L* and *nad5* in group Six. By integrating the

results, it was found that all the mt genomes had the *nad5* gene, and there was also another *nad* gene next to it. So *nad5* always had the tandem rule during the evolution of the *nad* gene family. All the sequences of *nad* family genes were obtained from mt genomes, and the genes with a length less than 200 bp were removed. Finally, 233 *nad* genes were obtained for constructing the phylogenetic tree (Figure 8). Comparing the phylogenetic tree results of all *nad* subfamilies, the genetic distance of *nad5* was different with that of other subfamilies, which was divided into a separate group. At the same time, a species might have multiple *nad5* genes, and the *nad5* genes of all species were located in two branches of the evolutionary tree.

4 Discussion

The common *Camellia* oil tree known to us is a hexaploid plant with huge genomic data and complex structure (Gong et al., 2022). The chloroplast genome and mt genome could help to provide a more comprehensive auxiliary analysis of the genome evolution of *C. oleifera* (Li et al., 2021). With the advancement of sequencing methods, we have obtained more accurate sequences for the genome assembly. At the same time, original and simple methods are used to dig out the information from the genome sequence. So, this paper presents a complete mt genome of *C. oleifera* cv. Huashuo with a length of 709,596 bp, which is larger than most

TABLE 4 The amino acid transformation of RNA-editing sites in the *Camellia oleifera* cv. Huashuo mt genomes.

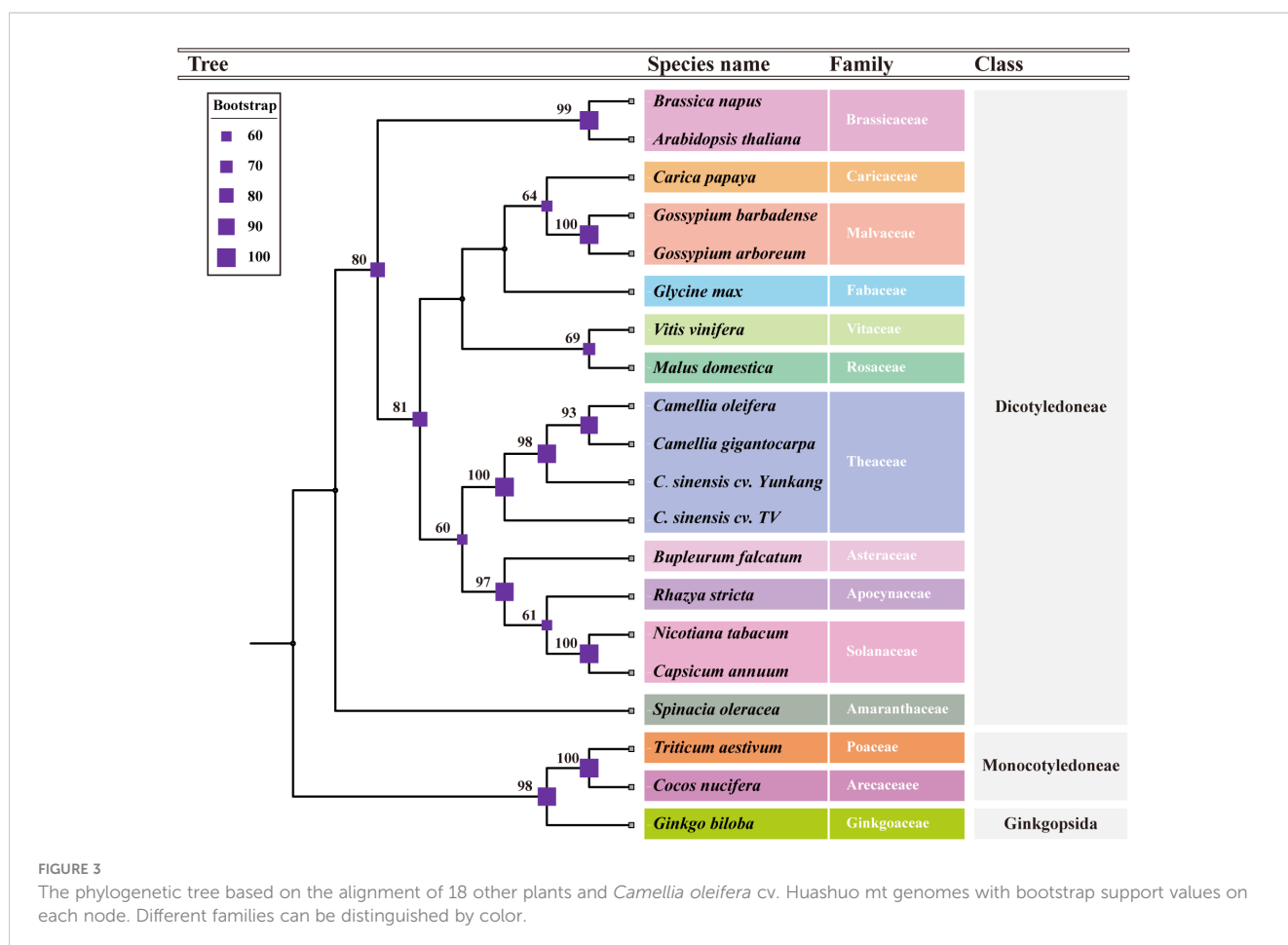
Amino acid	A-V	H-Y	L-F	P-F	P-L	P-S	Q-X	R-C	R-X	R-W	S-F	S-L	T-I	T-M
Count	12	23	13	20	89	33	2	34	2	19	50	105	6	5
Proportion(%)	2.91	5.57	3.15	4.84	21.55	7.99	0.48	8.23	0.48	4.60	12.11	25.42	1.45	1.21
Base position	2	1	1	1/2	2	1	1	1	1	1	2	2	2	2

of the known plant mt genomes. Our lab has reported the chloroplast genome of *C. oleifera* cv. Huashuo (Wu et al., 2020), and the goal of this study is to analyze the mt genome of *C. oleifera* based on the sequencing data.

4.1 Differences in gene quantity and conservation of gene arrangement among *Camellia* species

Compared with mt genomes of *C. sinensis* var. *assamica* (YK10, Zhang et al., 2019 and TV-1 Rawal et al., 2020) and *C. gigantocarpa*, the mt genome size of *C. oleifera* was 709,596 bp, which was similar to the mt genome of TV-1 (707,441 bp) (Table 1). The mt genome of YK10 and *C. gigantocarpa* was longer than theirs. According to all the previous studies, it was found that the evolution of plant mt genome was more complicated than that of chloroplast genome. The genome size, composition, and gene sequence varied greatly among different species, among different individuals within the same species, and even among different cells of the same individual (Li et al., 2011). Moreover, the two cultivated tea plants are from China and India. But even then, like the Bryophyte (Liu et al., 2014), the mitochondria genome remained structurally stable through evolution. Thus, all plants cannot be generalized; they need to have a case-by-case analysis.

A total of 35 protein-coding genes, 29 tRNAs, and 2 rRNAs were annotated in the mt genome of *C. oleifera*. The protein-encoding genes of TV-1 were four fewer than that of *C. oleifera*, and the total number of genes was four fewer than that of *C. oleifera*. The sequenced mt genomes of plants have been annotated to obtain approximately 35 protein-coding genes, such as 35 in *Vitis vinifera* (Goremykin et al., 2009), 36 in *Leucaena Trichandra* (Kovar et al., 2018), and 33 in *Dalbergia odorifera* (Hong et al., 2021). YK10 and Cgi has 9 protein-coding genes more than *C. oleifera*, but 5 and 10 tRNA genes fewer than the other two. Unseld et al. (1997) found that the number of genes was different among different species, which might be related to the number of ribosomal subunit genes and tRNAs (Li et al., 2011). Indeed, *rpl* genes in the *C. oleifera* mt genome (2 genes) and TV-1 (2 genes) are less than those in YK10 and Cgi (4 genes). In addition, the NADH dehydrogenase gene of *C. oleifera* was also lost compared with YK10 and Cgi. *C. oleifera*, Cgi, YK10, and TV-1 all belong to the Theaceae, but there is a great quantity variance in the number of annotated genes. It is possible that the methods used in genome annotation made this situation in closely related species. It is worth paying attention to the occurrence of such cases with comparative genomic analysis in the future. The mt of *C. oleifera* also contained eight ORF genes, which have been found in previous studies to be usually related to the “infertility” problem in plants (Chaumont et al., 1995). For example, the *T2urf13* in maize, *orf79* gene in rice, and *orf224/orf138* in rape



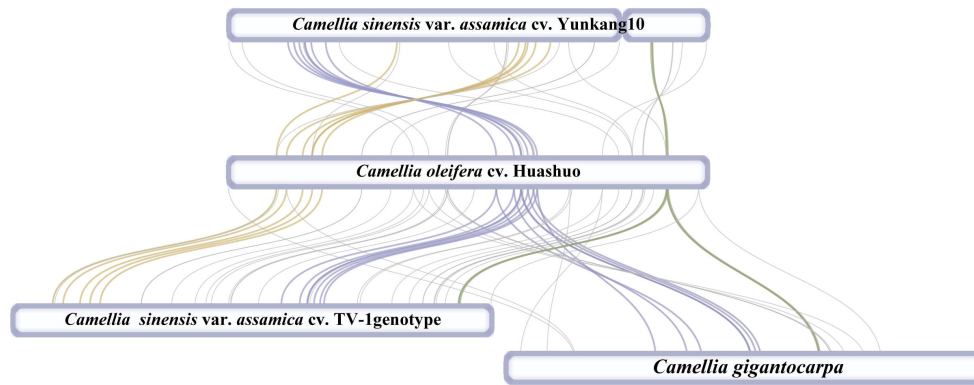


FIGURE 4 Collinearity analysis with mt genomes of four *Camellia* species. The genes indicated by the colors correspond to those in Figure 4.

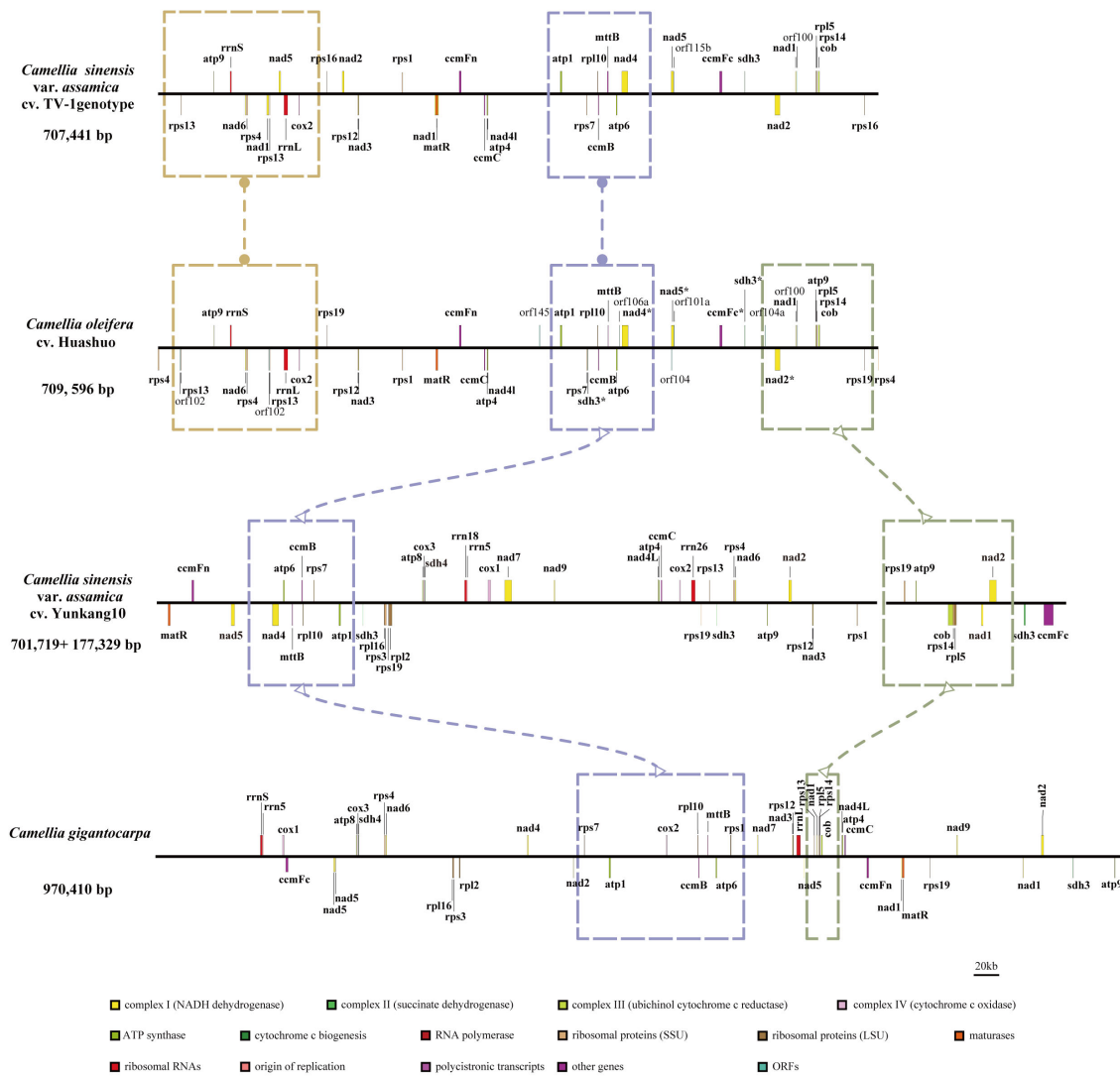


FIGURE 5 Comparison of mt genomes of four *Camellia* species (without transfer RNAs). Parts selected by the same color box have similar structure.

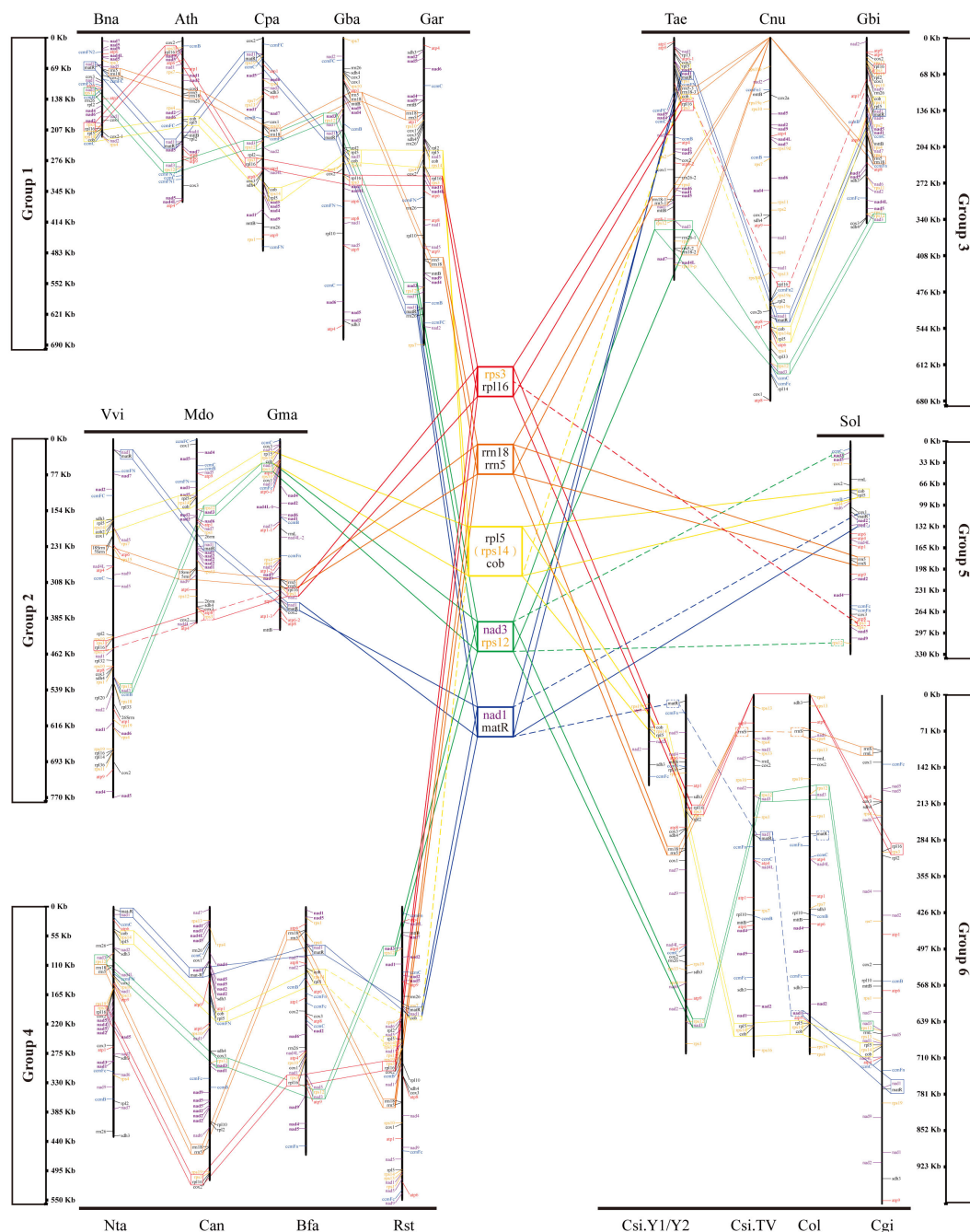


FIGURE 6 Gene localization in mt genomes of all species in the phylogenetic tree. Five gene clusters are in the middle of the figure.

are all associated with cytoplasmic male sterility (Duroc et al., 2005). The function of rich *orf* genes in *C. oleifera* mt can be explored later in combination with plant characteristics and nuclear genes.

RNA-editing sites can affect gene function by changing the protein folding pattern. The genome size and number of protein-coding genes of *C. oleifera* were larger than that of TV-1, but there were fewer RNA-editing sites. The number of protein-coding genes of *C. oleifera* and YKg10 differed by 9, but 65 differences in RNA-

editing sites. These results confirmed that an increase in the number of bases in the mt genome did not lead to an increase in RNA-editing sites. At the same time, the number of editing sites of the same gene in different species is also different. In addition, when studying early terrestrial plants, Zhang found that the number of RNA-editing sites ranged from 0 to 2,152, and different types and numbers of RNA-editing sites contributed to the characteristics of plants (Zhang et al., 2020).

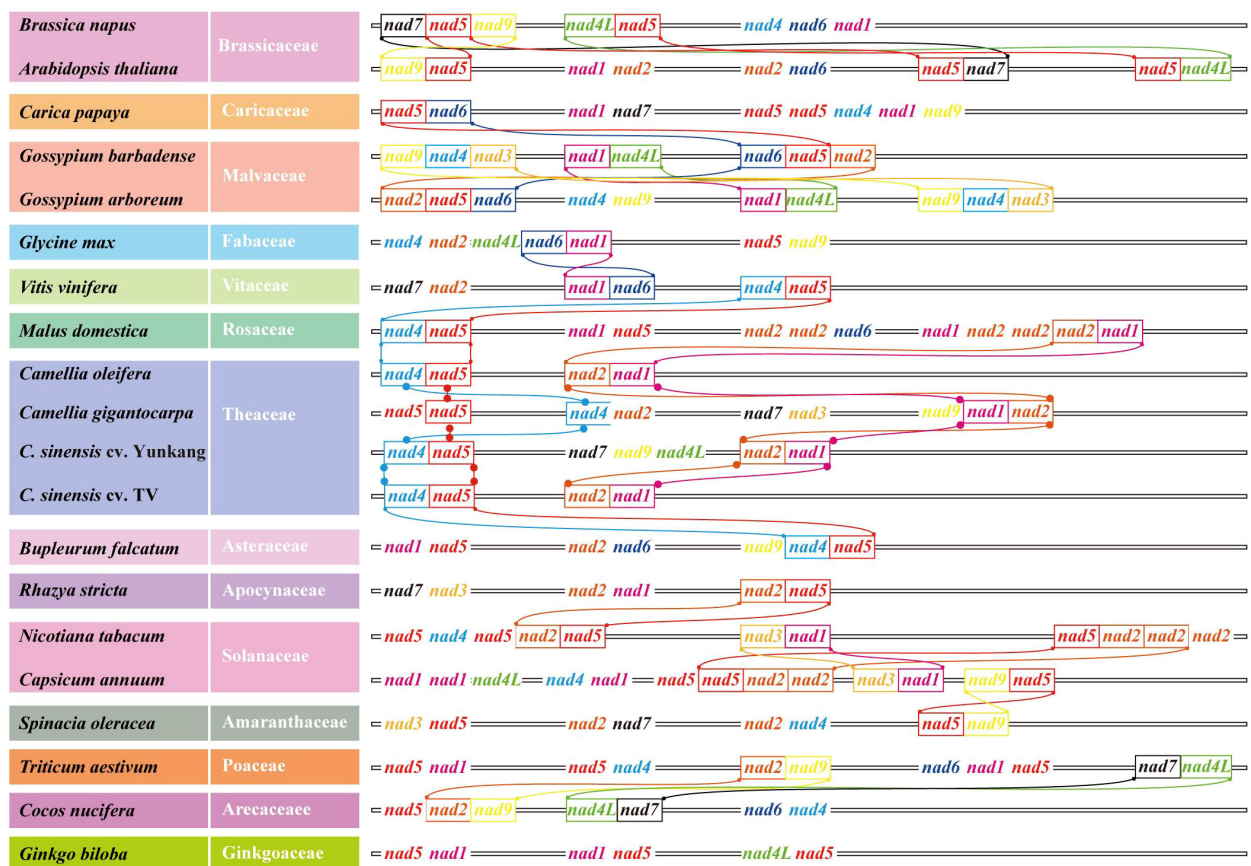


FIGURE 7

Nad-nad gene clusters in mt genomes of all species in the phylogenetic tree. Different *nad* genes were color coded. Identical tandem gene pairs between adjacent species were connected by curves.

4.2 Conserved gene clusters exist during the evolution of species, especially *nad5*

From the mt genome of *C. sinensis* and other previous studies, it is found that the species can be correctly classified by constructing an evolutionary tree through conserved genes. In this study, we screened 15 conserved genes from 20 mt genomes, and finally *C. oleifera*, *C. sinensis*, and *C. gigantocarpa* clustered together in one branch, and other species belonging to the same family also clustered together. We found that the number of conserved genes screened for evolutionary tree construction accounted for about one-third of the total number of genes (excluding tRNA genes). Concurrently, when comparing four mt genomes of *Camellia*, we also found that there were structurally conserved regions in the sequences of *C. oleifera* and other three.

Rearranging the position of genes in the mt genome can occur during evolution as the result of sequence break and sequence recombination. However, some highly conserved genes or gene clusters maintained their original evolutionary patterns throughout the process of evolution (Niu et al., 2022). In this study, five conserved gene clusters were identified across the 20 species: *rps3-rpl6*, *rrn18-rrn5*, *rpl5-cob*, *nad3-rps12*, and *nad1-matR*. But it was observed that some gene of these five gene pairs also were absent in some species. For instance, this study found that *S. oleracea* lacks the *rpl16* gene, which is also missing in *Melastoma*

dodecandrum (Cai et al., 2017; Zhou et al., 2023). Additionally, *C. nucifera* lacks the *rps3*, which was relatively conserved in all species studied. So, this study grouped all species according to the conservativeness of all genes and found that the results of the grouping matched the position on the phylogenetic tree. Therefore, the absence of genes can be used to distinguish different evolutionary groups of species. The conservativeness of gene pairs can also be used to analyze the distance of genome evolution. Niu et al. found two conserved gene clusters (*rps12-nad3* and *rps3-rpl16*) in eight mt genomes (Niu et al., 2022). This result overlaps with the gene pairs found in this study, which shows that the conservativeness of gene clusters is universal in different species. It can be further explored during the study of mt genome.

When analyzing conservative genes, it was found that the *nad* gene has its own genetic rule. The *nad* gene is a part of the large enzyme complex of complex I, and it is active in the mitochondria (Wu et al., 2022). Complex I is one of the enzyme complexes necessary for oxidation phosphorylation. The mitochondrion is the main organelle that produces energy in the cell, which produces adenosine triphosphate (ATP), the main energy source of the cell, through the process of oxidative phosphorylation (Gualberto et al., 2013; Zhao et al., 2023). Therefore, for the whole life of an organism, the *nad* gene is a key role. When analyzing the genetic inheritance rules between different species, it showed that *nad* gene tandem

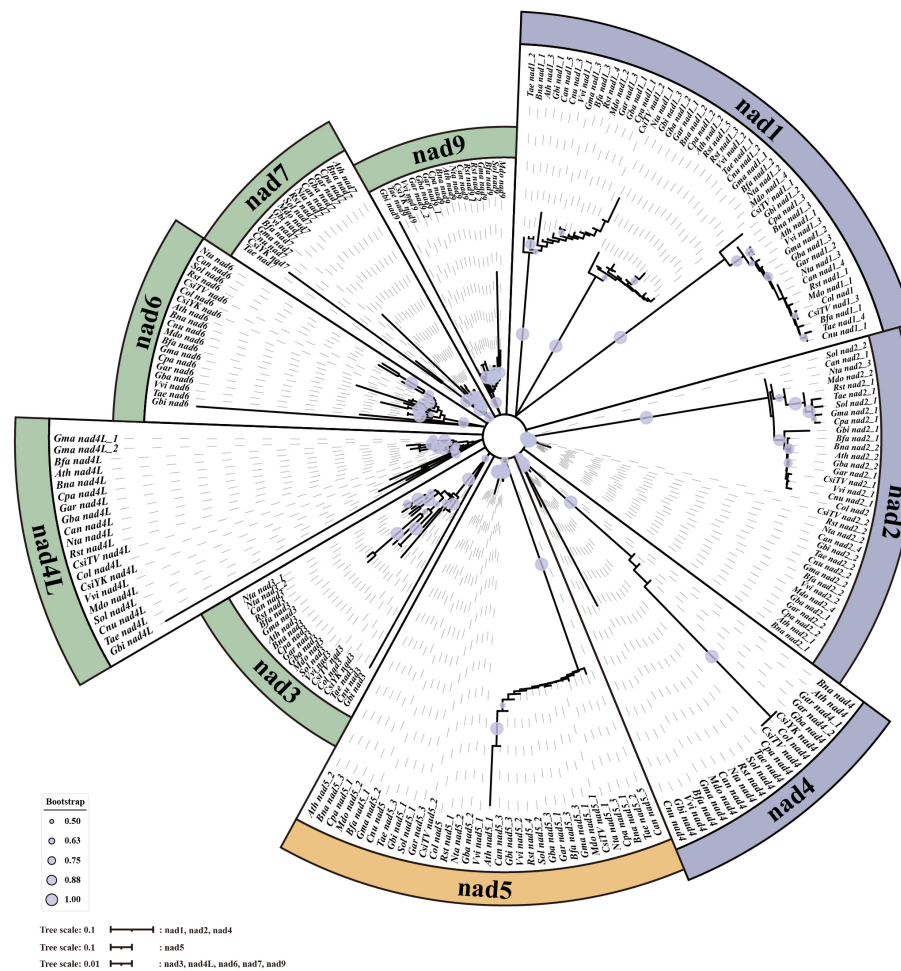


FIGURE 8
The phylogenetic tree of *nad* genes in mt genomes.

arranging was common in all the mt genomes. For example, *nad5–nad4* in *C. oleifera* and *C. sinensis*, *nad5–nad9*, *nad1–nad2*, and *nad5–nad7* in *A. thaliana* were located in adjacent locations in the mt genome (Unsel'd et al., 1997). With the phylogenetic tree, there were common *nad* tandem pairs in the species with close parental relationship. For example, *nad1/2/3/4/5/9/4L* in *G. barbadense* and *G. arboretum* has genomic tandem (Tang et al., 2015). Also, *nad1/2/3/4/5* in the mt genomes of *N. tabacum* and *C. annuum* existed in tandem (Sugiyama et al., 2005; Magee et al., 2010). In the mt genomes of gymnosperms, such as *G. biloba*, all *nad–nad* tandem gene pairs contain *nad5* (*nad5–nad1* and *nad5–nad4L*). It was found that, except for *Oryza minuta*, the mt genomes of most species contain *nad5* (Sajjad et al., 2016; Niu et al., 2022), and the *nad5* genes were arranged in tandem. Similarly, in the mt genome of *Dalbergia odorifera*, the *nad5* gene was found to be linked in tandem with *nad1*, as was the case with the *nad5* gene in *Leucaena trichandra* (Kovar et al., 2018; Zhou et al., 2021). Therefore, other *nad* genes may be required to induce or enhance the function of *nad5* in mitochondria life activity. *nad5* may have an ability to enhance its tandem genes' function.

4.3 Software parameters on gene annotation in mt genomes

When we annotated the *Camellia oleifera* mitochondrial genome, we found that the gene number of *nads* and other gene families was correlated with the software parameters that we set. At first, when the annotated genes were screened, the “coverage” and “match” were set to more than 40%. Twelve *nad* genes and all 48 protein-coding genes were obtained (Supplementary Figure 1). But when the two parameters were increased to 60% or even 90%, only seven *nad* genes were annotated (Supplementary Figure 1). In our manuscript, parameters with 60% were used. It was also found when we reannotated the mitochondrial genome of TV-1. In addition, the evolution-related regularities we found in our manuscript were mainly related to these seven *nad* genes. Combined with the subsequent analysis of the paper, we found that adjusting the values of coverage and matching degree during screening could help in obtaining more accurate results in comparative genomic analysis.

5 Conclusion

In summary, by sequencing and analyzing the mt genome of *Camellia oleifera*, we found the general tandem array of the *nad* gene family in the mt genome. Through the analysis of *nad* gene families in different species, we found that the *nad* gene with tandem relationship was species specific. In addition, it was noteworthy that the *nad5* has tandem gene pairs of *nad5–nads* in all species. The discovery of conserved genes and conserved gene pairs will provide a new direction for the study of the mitochondria.

Data availability statement

The data that support the findings of this study have been deposited into [CNGB Sequence Archive \(CNSA\)](#) of [China National GeneBank DataBase \(CNGBdb\)](#) with accession number CNP0005996.

Author contributions

YG: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. LY: Data curation, Writing – original draft, Methodology. JZ: Writing – review & editing, Supervision, Conceptualization. ZX: Writing – original draft, Validation, Software. ML: Writing – original draft, Visualization, Investigation. YZ: Writing – review & editing, Project administration. XT: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Hunan Province Innovative Province construction project (Project Number: 2021JC0007); and the

Hunan Provincial Innovation Foundation for Postgraduate (Project Number: CX20220702).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1396635/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Gene annotation results with different parameters.

SUPPLEMENTARY TABLE 1

SSR motifs annotation of *C. oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 2

Long repeat sequences annotation of *C. oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 3

Statistics of RNA-editing sites in the *Camellia oleifera* cv. Huashuo mt genomes.

SUPPLEMENTARY TABLE 4

RNA-editing sites annotation of *C. oleifera* cv. Huashuo mt genomes.

References

- Al-Nakeeb, K., Petersen, T. N., and Sicheritz-Pontén, T. (2017). Norgal: extraction and *de novo* assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinf.* 18, 1–7. doi: 10.1186/s12859-017-1927-y
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Burns, J. A., Paasch, A., Narechania, A., and Kim, E. (2015). Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. *Genome Biol. Evol.* 7, 3047–3061. doi: 10.1093/gbe/evv144
- Cai, X. F., Jiao, C., Sun, H. H., Wang, X. L., Xu, C. X., Fei, Z. J., et al. (2017). The complete mitochondrial genome sequence of spinach, *Spinacia oleracea* L. *Mitochondrial DNA B. Resour.* 2, 339–340. doi: 10.1080/23802359.2017.1334518
- Chang, S. X., Wang, Y. K., Lu, J. J., Gai, J. Y., Li, J. J., Chu, P., et al. (2013). The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *PLoS One* 8, E56502. doi: 10.1371/journal.pone.0056502
- Chaumont, F., Bernier, B., Buxant, R., Williams, M. E., Levings, C. S., and Boutry, M. (1995). Targeting the maize T-urf13 product into tobacco mitochondria confers methomyl sensitivity to mitochondrial respiration. *P. Natl. A. Sci.* 92, 1167–1171. doi: 10.1073/pnas.92.4.1167
- Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, Z., Nie, H., Grover, C. E., Wang, Y., Li, P., Wang, M., et al. (2017). Entire nucleotide sequences of *Gossypium raimondii* and *G. arboreum* mitochondrial genomes revealed A-genome species as cytoplasmic donor of the allotetraploid species. *Plant Biol.* 19, 484–493. doi: 10.1111/plb.12536
- Clifton, S. W., Minx, P., Fauron, C. M.-R., Gibson, M., Allen, J. O., Sun, H., et al. (2004). Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* 136, 3486–3503. doi: 10.1104/pp.104.044602
- Cui, P., Liu, H. T., Lin, Q., Ding, F., Zhuo, G. Y., Hu, S. N., et al. (2009). A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. *J. Genet.* 88, 299–307. doi: 10.1007/s12041-009-0043-9
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45, e18–e18. doi: 10.1093/nar/gkw955

- Duroc, Y., Gaillard, C., Hiard, S., DeFrance, M., Pelletier, G., and Budar, F. (2005). Biochemical and functional characterization of *ORF138*, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassicaceae. *Biochimie* 87, 1089–1100. doi: 10.1016/j.biochi.2005.05.009
- Edgar, R. C. (2021). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13, 6968. doi: 10.1038/s41467-022-34630-w
- Gong, W. F., Xiao, S. X., Wang, L. K., Liao, Z. Y., Chang, Y. H., Mo, W. J., et al. (2022). Chromosome-level genome of *Camellia lanceoleosa* provides a valuable resource for understanding genome evolution and self-incompatibility. *Plant J.* 110, 881–898. doi: 10.1111/tpj.15739
- Goremykin, V. V., Lockhart, P. J., Viola, R., and Velasco, R. (2012). The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant J.* 71, 615–626. doi: 10.1111/j.1365-3113.2012.05014.x
- Goremykin, V. V., Salamini, F., Velasco, R., and Viola, R. (2009). Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* 26, 99–110. doi: 10.1093/molbev/msn226
- Green, R. E., Malaspina, A. S., Krause, J., Briggs, A. W., Johnson, P. L., Uhler, C., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134, 416–426. doi: 10.1016/j.cell.2008.06.021
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Gualberto, J. M., Mieshina, D., Wallet, C., Niaz, A. K., Weber-Lotfi, F., and Dietrich, A. (2013). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016
- Hahn, C., Bachmann, L., and Chevreaux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41, e129–e129. doi: 10.1093/nar/gkt371
- Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* 31, 5907–5916. doi: 10.1093/nar/gkg795
- He, F., and He, B. (2002). Cultural distribution and site classification for *Camellia oleifera*. *Sci. Silv. Sin.* 38, 64–72. doi: 10.1007/s11769-002-0026-8
- Hong, Z., Liao, X. Z., Ye, Y. J., Zhang, N. N., Yang, Z. J., Zhu, W. D., et al. (2021). A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics* 22, 1–13. doi: 10.1186/s12864-021-07967-7
- Hunter, S. S., Lyon, R. T., Sarver, B. A., Hardwick, K., Forney, L. J., Settles, M. L., et al. (2015). Assembly by reduced complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *Biorxiv* 014662. doi: 10.1101/014662
- King, J. L., LaRue, B. L., Novroski, N. M., Stoljarova, M., BumSeo, S., Zeng, X. P., et al. (2014). High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci. Int.: Gen.* 12, 128–135. doi: 10.1016/j.fsigen.2014.06.001
- Kovar, L., Nageswara-Rao, M., Ortega-Rodriguez, S., Dugas, D. V., Straub, S., Cronn, R., et al. (2018). PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intragenetic assessment of mitochondrial RNA editing. *Genome Biol. Evol.* 10, 2501–2517. doi: 10.1093/gbe/evy179
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., et al. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, J., Bi, C. W., Tu, J., and Lu, Z. H. (2018). The complete mitochondrial genome sequence of *Boechera stricta*. *Mitochondrial DNA B. Resour.* 3, 896–897. doi: 10.1080/23802359.2018.1501323
- Li, L., Hu, Y. F., He, M., Zhang, B., Wu, W., Cai, P. M., et al. (2021). Comparative chloroplast genomes: insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. *BMC Genomics* 22, 1–22. doi: 10.1186/s12864-021-07427-2
- Li, S. S., Xue, L. F., Su, A. G., Lei, B. B., Wang, Y. M., and Hua, J. P. (2011). Progress on sequencing and alignment analysis of higher plant mitochondrial genomes. *J. China Agric. Univ.* 16, 22–27. doi: 10.3724/SP.J.1011.2011.00353
- Liu, Y., Medina, R., and Goffinet, B. (2014). 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. *Mol. Biol. Evol.* 31, 2586–2591. doi: 10.1093/molbev/msu199
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955
- Lu, C., Gao, L. Z., and Zhang, Q. J. (2022). A high-quality genome assembly of the mitochondrial genome of the oil-tea tree *camellia gigantocarpa* (Theaceae). *Diversity* 14, 850. doi: 10.3390/d14100850
- Lu, M. Q., Zhou, J. Q., Liu, Y. Y., Yang, J., and Tan, X. F. (2021). *CoNPR1* and *CoNPR3.1* are involved in SA- and MeSA-mediated growth of the pollen tube in *Camellia oleifera*. *Physiol. Plant.* 172, 2181–2190. doi: 10.1111/ppl.13410
- Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Sémon, M., Perry, A. S., et al. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20, 1700–1710. doi: 10.1101/gr.111955.110
- Mower, J. P. (2005). PREP-Mt: predictive RNA editor for plant mitochondrial genes. *BMC Bioinf.* 6, 1–15. doi: 10.1186/1471-2105-6-96
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Niu, L., Zhang, Y. Y., Yang, C. M., Yang, J., Ren, W., Zhong, X. F., et al. (2022). Complete mitochondrial genome sequence and comparative analysis of the cultivated yellow nutsedge. *Plant Genome* 15, e20239. doi: 10.1002/tpg2.20239
- Park, S., Ruhlman, T. A., Sabir, J. S., Mutwakil, M. H. Z., Baeshen, M. N., Sabir, M. J., et al. (2014). Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics* 15, 405. doi: 10.1186/1471-2164-15-405
- Rawal, H. C., Kumar, P. M., Bera, B., Singh, N. K., and Mondal, T. K. (2020). Decoding and analysis of organelle genomes of Indian tea (*Camellia assamica*) for phylogenetic confirmation. *Genomics* 112, 659–668. doi: 10.1016/j.ygeno.2019.04.018
- Sajjad, A., Latif, K. A., Rahim, K. A., Muhammad, W., Kang, S. M., Aqil, K. M., et al. (2016). Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. *PLoS One* 11, e0152937. doi: 10.1371/journal.pone.0152937
- Sugiyama, Y., Watake, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A., et al. (2005). The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genomics* 272, 603–615. doi: 10.1007/s00438-004-1075-8
- Tan, X. F. (2013). Advances in the molecular breeding of *Camellia oleifera*. *J. Cent. South Univ. Forest. Tech.* 43, 1–24. doi: 10.14067/j.cnki.1673-923x.2023.01.001
- Tang, M. Y., Chen, Z. W., Grover, C. E., Wang, Y. M., Li, S. S., Liu, G. Z., et al. (2015). Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. *BMC Genomics* 16, 770. doi: 10.1186/s12864-015-1988-0
- Tian, X. J., Zheng, J., Hu, S. N., and Yu, J. (2006). The rice mitochondrial genomes and their variations. *Plant Physiol.* 140, 401–410. doi: 10.1104/pp.105.070060
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Unsel, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* 15, 57–61. doi: 10.1038/ng0197-57
- Wang, R., Fang, Y. N., Wu, X. M., Qing, M., Li, C. C., Xie, K. D., et al. (2020). The miR399-CsUBC24 module regulates reproductive development and male fertility in citrus. *Plant Physiol.* 183, 1681–1695. doi: 10.1104/pp.20.00129
- Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Wu, F. Y., Li, Z., Liu, F. D., Wang, H., Shi, K., Sun, Y., et al. (2018). Comparison of photosynthetic characteristics of *Camellia oleifera* cultivars. *Nonwood. For. Res.* 36, 7–13. doi: 10.14067/j.cnki.1003-8981.2018.02.002
- Wu, L. L., Li, J. A., Gu, Y. Y., Zhang, F. H., Li, Z., and Tan, X. F. (2020). Complete chloroplast genome sequences and phylogenetic analysis of three *Camellia oleifera* cultivars. *Int. J. Agric. Biol.* 24, 743–750. doi: 10.17957/IJAB/15.1495
- Wu, Z. Q., Liao, X. Z., Zhang, X. N., Tembrock, L. R., and Broz, A. (2022). Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. *J. Syst. Evol.* 60, 160–168. doi: 10.1111/jse.12655
- Zhang, F., Li, W., Gao, C. W., Zhang, D., and Gao, L. Z. (2019). Deciphering tea tree chloroplast and mitochondrial genomes of *Camellia sinensis* var. *assamica*. *Sci. Data* 6, 1–10. doi: 10.1038/s41597-019-0201-8
- Zhang, F. F., Li, Z., Zhou, J. Q., Gu, Y. Y., and Tan, X. F. (2021). Comparative study on fruit development and oil synthesis in two cultivars of *Camellia oleifera*. *BMC Plant Biol.* 21, 1–16. doi: 10.1186/s12870-021-03114-2
- Zhang, J., Fu, X. X., Li, R. Q., Zhao, X., Liu, Y., Li, M. H., et al. (2020). The hornwort genome and early land plant evolution. *Nat. Plants* 6, 107–118. doi: 10.1038/s41477-019-0588-4
- Zhao, J. M., Chen, J., Xiong, Y., He, W., Xiong, Y. L., and Xu, Y. D. (2023). Organelle genomes of *Indigofera amblyantha* and *Indigofera pseudotinctoria*: comparative genome analysis, and intracellular gene transfer. *Ind. Crop Prod.* 198, 0926–6690. doi: 10.1016/j.indcrop.2023.116674
- Zhou, H., Liao, X. Z., Ye, Y. J., Zhang, N. N., Yang, Z. J., Zhu, W. D., et al. (2021). A complete mitochondrial genome for fragrant Chinese rosewood (*Dalbergia odorifera*, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. *BMC Genomics* 22, 672. doi: 10.1186/s12864-021-07967-7
- Zhou, J. Q., Lu, M. Q., Yu, S. S., Liu, Y. Y., Yang, J., and Tan, X. F. (2020). In-depth understanding of *Camellia oleifera* self-incompatibility by comparative transcriptome, proteome and metabolome. *Int. J. Mol. Sci.* 21, 1600. doi: 10.3390/ijms21051600
- Zhou, Y. Z., Zheng, R. Y., Peng, Y. K., Chen, J. M., Zhu, X. Y., Xie, K., et al. (2023). The first mitochondrial genome of *Melastoma dodecandrum* resolved structure evolution in Melastomataceae and micro inversions from inner horizontal gene transfer. *Ind. Crop Prod.* 205, 0926–6690. doi: 10.1016/j.indcrop.2023.117390