

RESEARCH

Open Access



# Assembly and analysis of the complete mitochondrial genome of an endemic *Camellia* species of China, *Camellia tachangensis*

Dongzhen Jiang<sup>1</sup>, Lei Zhou<sup>1</sup>, Zhaohui Ran<sup>1</sup>, Xu Xiao<sup>1</sup>, Xuehang Yang<sup>1</sup> and Zhi Li<sup>1\*</sup>

## Abstract

**Background** *Camellia tachangensis* F. C. Zhang is an endemic *Camellia* species of the junction of Yunnan, Guizhou and Guangxi Provinces in China. It is characterized by a primitive five-chambered ovary morphology and serves as the botanical source of the renowned “Pu’an Red Tea”. Unfortunately, the populations of the species have declined due to the destruction of their habitats by human activities. The lack of mitochondrial genomic resources has hindered research into molecular breeding and phylogenetic evolution of *C. tachangensis*.

**Result** In this study, we had sequenced, assembled, and annotated the mitochondrial genome of *C. tachangensis* to reveal its genetic characteristics and phylogenetic relation with other *Camellia* species. The assembly result indicated that the mitochondrial genome sequence of *C. tachangensis* was 746,931 bp (GC content = 45.86%). It consisted of one multibranch sequence (Chr1) and one circular sequence (Chr2), with Chr1 capable of producing 7 substructures. The comparative analysis of the mitochondrial and chloroplast DNA of *C. tachangensis* revealed 23 pairs of chloroplast homologous fragments, with 10 fully preserved tRNA genes within them. Interspecies comparison of Ka/Ks ratios revealed that mutations in mitochondrial protein-coding genes (PCGs) of *C. tachangensis* were predominantly shaped by purifying selection throughout its evolution ( $Ka/Ks < 1$ ). The mitochondrial CDS-based phylogenetic tree indicated that within the *Camellia* lineage, *C. tachangensis* was phylogenetically independent of the species of sections *Oleifera*, *Camellia*, *Heterogenea*, and *Chrysantha*. However, it also did not support the clustering of *C. tachangensis* with certain variants of *C. sinensis*, due to the extremely low support (BS = 22, PP = 0.41). Meanwhile, the chloroplast PCG-based phylogenetic analysis revealed that *C. tachangensis* formed a strongly supported basal clade (BS = 100, PP = 1.00), alongside *C. makuanica* (NC\_087766), *C. taliensis* (NC\_022264), and *C. gymnogyna* (NC\_039626).

**Conclusions** Our study deciphered the mitochondrial genome and its multibranch structure of *C. tachangensis*. These findings not only enhanced our comprehension of the complexity and diversity of mitochondrial genome structures in *Camellia* species, but also established a foundational genetic data framework for future research on molecular breeding programs and phylogenetic relationship involving *C. tachangensis* and its related species.

**Keywords** *Camellia tachangensis*, Horizontal transfer, Mitochondrial genome, Evolution, Phylogenetic analysis

## Introduction

*Camellia tachangensis* F. C. Zhang, belonging to sect. *Thea* (L.) of the genus *Camellia*, is an endemic species native to the border regions of Yunnan, Guizhou, and Guangxi Provinces in China [1]. This species has a distinctive morphological features of a capsule with five-locular ovary and been recognized as the primitive plant

\*Correspondence:

Zhi Li  
zli7@gzu.edu.cn

<sup>1</sup> College of Forestry, Guizhou University, Guiyang, China



within sect. *Thea* Dyer in genus *Camellia* [2, 3]. Some studies on the systematic taxonomy of *Camellia* plants showed *C. tachangensis* has high scientific value for exploring the origin and evolution of *Camellia* in south-west China [4–7]. At the same time, as a member of sect. *Thea*, *C. tachangensis* was served as a processing raw material for “Pu’an red tea” with unique aroma and flavor that it is a product of China’s national geographical indication [8]. Due to environmental degradation and human interference, coupled with its relatively low reproductive capacity, the survival of *C. tachangensis* was facing a significant crisis [9]. To improve the survival status of *C. tachangensis*, researchers had conducted comprehensive and in-depth studies on its physiological, breeding, propagation, biochemical characteristics and so on [10–13]. However, there were few studies in genome biology, and most of them only focused on population genetics and chloroplast genomes [14–16]. Studies concerning the complete mitochondrial genome were still lacking.

Mitochondria in plant cells are essential organelles that not only participate in cellular respiration but also play crucial roles in regulating intracellular metabolic networks [17, 18]. In botanical research, the mitochondrial genes of plants hold significant value for study. First, mitochondrial genes are involved in the synthesis of essential enzymatic components such as ATP synthase, cytochrome c oxidase, and NADH dehydrogenase, which are crucial for the respiratory metabolism of plants [19–21]. Investigating the composition and evolution of these genes will provide novel insights into the genetic improvement of *C. tachangensis* and its related species. Furthermore, most plant mitochondrial DNA exhibits maternal inheritance characteristics, with a lower level of heterozygosity than nuclear genes [22, 23]. Consequently, the mitochondrial genome can be effectively utilized for the classification and identification of species within the genus *Camellia*, which display significant genetic heterozygosity and phenotypic diversity [24–26]. Additionally, structural variations and the insertion of chloroplast-derived fragments occur relatively frequently in mitochondrial DNA. These processes typically take place during different stages of plant lineage differentiation. By comparing the similarities and differences in mitochondrial genome structures and chloroplast-derived fragments across different species, we can gain deeper insights into the evolutionary history of the *Camellia* genus [27–29].

This study employed a combination of high-throughput sequencing and long-read sequencing to achieve the first complete sequencing, assembly, and annotation of mitochondrial genome of *C. tachangensis* while also exploring its substructure. Comprehensive analyses were performed to investigate multiple genomic features

including codon usage patterns, chloroplast-derived homologous sequences, repetitive element distribution, RNA editing sites, evolutionary selection pressures through Ka/Ks ratio calculations, and phylogenetic relationships. The findings of this work established a foundational genetic data framework for future research on molecular breeding programs and evolutionary dynamics involving *C. tachangensis* and its related species.

## Materials and methods

### Material collection, DNA extraction and sequencing

In this study, we selected the leaves of *C. tachangensis* from the forestry center in Longli County, Guizhou Province (N 26°24′49″–26°44′30″, E 106°48′12″–107°8′50″), as the research materials. Fresh leaves were collected and preserved in liquid nitrogen at –80 °C. Total DNA was extracted by CTAB method [30].

The mitochondrial genome was sequenced and assembled via a combination of high-throughput sequencing (Illumina Novaseq 6000) and long-read sequencing (Oxford Nanopore R10.4). The high-throughput sequencing strictly conformed to the standard operating procedures provided by Illumina. First, DNA quality and concentration were assessed via 1% agarose gel electrophoresis and a NanoDrop 2000. The high-quality DNA samples were then fragmented through ultrasonic mechanical disruption. The fragmented DNA subsequently underwent purification, end repair, and ligation of sequencing adapters. Finally, the selected DNA was amplified via PCR to construct the sequencing library. The constructed library underwent quality control, and those that conformed with the quality assessment were sequenced via the Illumina Novaseq 6000 platform (Illumina, San Diego, California, United States) [31]. The raw data were subsequently filtered via fastp v0.23.4 software (<https://github.com/OpenGene/fastp>) [32]. The specific filtering criteria were as follows: (1) sequencing adapters and primer sequences were removed from the reads; (2) reads with average quality scores below Q5 were excluded; and (3) reads containing more than 5 N bases were eliminated. In the process of the long-read sequencing, genomic DNA was first randomly fragmented. Subsequently, magnetic beads were utilized for enrichment and purification to obtain large DNA fragments. Following this step, gel extraction was performed on the large fragments, and damage repair was conducted on the fragmented DNA. The purified fragments subsequently underwent end repair and A-tailing. The SQK-LSK109 kit was used to ligate adapters, thereby constructing a DNA library for quantitative assessment. Next, an appropriate amount of the DNA library was loaded onto the flow cell and subjected to real-time single-molecule sequencing on the Oxford Nanopore PromethION sequencer

[33]. Finally, the long-read sequencing data were filtered via Filtlong v0.24 software, and a Perl script was used for data analysis.

#### Assembly and annotation of mitochondrial genome

First, the raw long-read sequencing data were aligned with the plant mitochondrial gene database via minimap2 v2.1. Subsequently, sequences with alignment lengths greater than 50 bp were selected as candidate sequences. Among these candidates, those exhibiting a greater number of aligned genes (a single sequencing read containing multiple core genes) and superior alignment quality (exhibiting a relatively complete coverage of core genes) were chosen as seed sequences. Minimap2 was subsequently employed to align the original sequencing data against the seed sequences, filtering for overlaps greater than 1 kb and similarity exceeding 70% to incorporate additional sequences into the seed sequences [34]. The third-generation assembly software Canu v2.2 was employed to correct the obtained third-generation data [35]. Subsequently, Bowtie2 v2.3.5.1 was utilized to align the second-generation data with the corrected sequences. The paired high-throughput data and the corrected long-read sequencing data were then assembled via the default parameters of Unicycler v0.4.8 [36]. Subsequently, Bandage software v0.8.1 was employed to visualize the assembly results and make manual adjustments as necessary. Owing to the presence of multiple subcircular structures or even non-circular complex physical configurations in mitochondrial genomes, the corrected third-generation sequencing data were aligned to the contigs generated by Unicycler via minimap2. The branch directions were subsequently manually determined to obtain the final assembly results.

The annotation of the mitochondria genes was carried out through the following steps: utilizing the Basic Local Alignment Search Tool-Nucleotide (BLASTN) (<https://blast.ncbi.nlm.nih.gov/>) [37], the protein-coding genes and rRNA genes were compared with publicly available reference mitochondrial genome sequences from plants. Manual adjustments were subsequently made using the closely related species *C. sinensis* var. *sinensis* cv. Dahongpao (Genbank ID: PP212895) as a reference genome [2–7, 29]. In addition, tRNA genes were annotated via tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) [38]. The Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) was used for ORF annotation [39]. The minimum length was set to 102 bp, with redundant sequences and overlapping known genes excluded. Sequences longer than 300 bp were aligned against the NR database for annotation. The mitochondrial genome map was constructed via OGDRAW (<https://chlorobox.mpimp-goelm.mpg.de/OGDraw.html>) [40].

#### Analysis of repeat sequences and RNA editing prediction

To clarify the 3 types of repetitive sequences in the mitochondrial genome of *C. tachangensis*, simple sequence repeat (SSR), tandem repeat, and dispersed repeat, MISA v1.0 was employed (<https://webblast.ipk-gatersleben.de/misa/>) for the identification of SSRs [41, 42]. Tandem repeats were identified via Tandem Repeats Finder v4.09 (<http://tandem.bu.edu/trf/trf.submit.options.html>) [43]. The identification of dispersed repeats was conducted via BLASTN software (v2.10.1, parameters: -word size 7, evaluate 1e-5). During this process, redundant and tandem repeat sequences were removed [44]. Ultimately, all the identification results were visualized via Circos v0.69–5 [45]. To further investigate the RNA editing sites, we utilized the online tool PREPACT3 (<http://www.prepact.de/>) to predict RNA editing events, setting a critical threshold of 0.001 [46]. In Excel 2021, the distribution of RNA editing sites for different genes and the number of amino acid variation types were visualized using bar charts; while the proportion of hydrophilic and hydrophobic group change types was presented through pie charts to show their ratio relationship.

#### Analysis of relative synonymous codon usage (RSCU) and mitochondrial plastid DNAs (MTPTs) in the mitochondrial genome

The protein-coding sequences were obtained via the default settings of Phylosuit v1.22 software [47]. The relative synonymous codon usage rates (RSCU) based on mitochondrial genome protein-coding genes were calculated via MEGA v7.0 software [48].

Mitochondrial plastid DNAs (MTPTs) refer to DNA fragments of plasmid origin present in the mitochondrial genome. In this study, chloroplast genome sequences from the same samples were extracted, and BLASTN software was used to identify homologous sequences between the chloroplast and mitochondrial genomes, with a similarity threshold set at 70% and an E value of 1e-5. To visualize the homologous segments between the chloroplast and mitochondrial genomes more intuitively, Circos v0.69–5 was used [49].

#### Analysis of nucleotide diversity (Pi) and selection pressure

To comprehensively analyze the diversity and the impact of selection between *C. tachangensis* and other species within sect. *Thea*, as well as between *C. tachangensis* and species from other sections of the genus *Camellia*, this study selected mitochondrial genomes from 9 representative species (including *C. tachangensis*) within 5 related taxonomic groups of genus *Camellia*: sect. *Thea*, sect. *Chrysantha*, sect. *Camellia*, and sect. *Heterogonea*, and sect. *Oleifera* for comparison. Subsequently,

the Pi analysis and the Ka/Ks analysis were conducted as follows: mitochondrial sequences of 8 *Camellia* species were downloaded from the NCBI database (<http://www.ncbi.nlm.nih.gov/genome/organelle/>): *C. sinensis* var. *sinensis* cv. Dahongpao (PP212895), *C. sinensis* var. *sinensis* cv. Rougui (PP212896), *C. sinensis* var. *pubilimba* (ON782577), *C. sinensis* var. *assamica* cv. TV-1 (NC\_043914), *C. tianeensis* (PP727208), *C. chekiangoleosa* (NC\_086749), *C. gigantocarpa* (OP270590), and *C. oleifera* (PP579569). The MAFFT v7.427 software was used for global alignment of these plant mitochondrial genomes along with that of *C. tachangensis* [50, 51]. The resulting alignment file was used to calculate Pi values for each shared gene with DnaSP v6.12.03 and Ka/Ks ratios for shared PCGs with KaKs\_Calculator v3.0 [52, 53]. The Ka/Ks ratio data were visualized in the form of a box plot using Excel 2021.

### Phylogenetic analysis

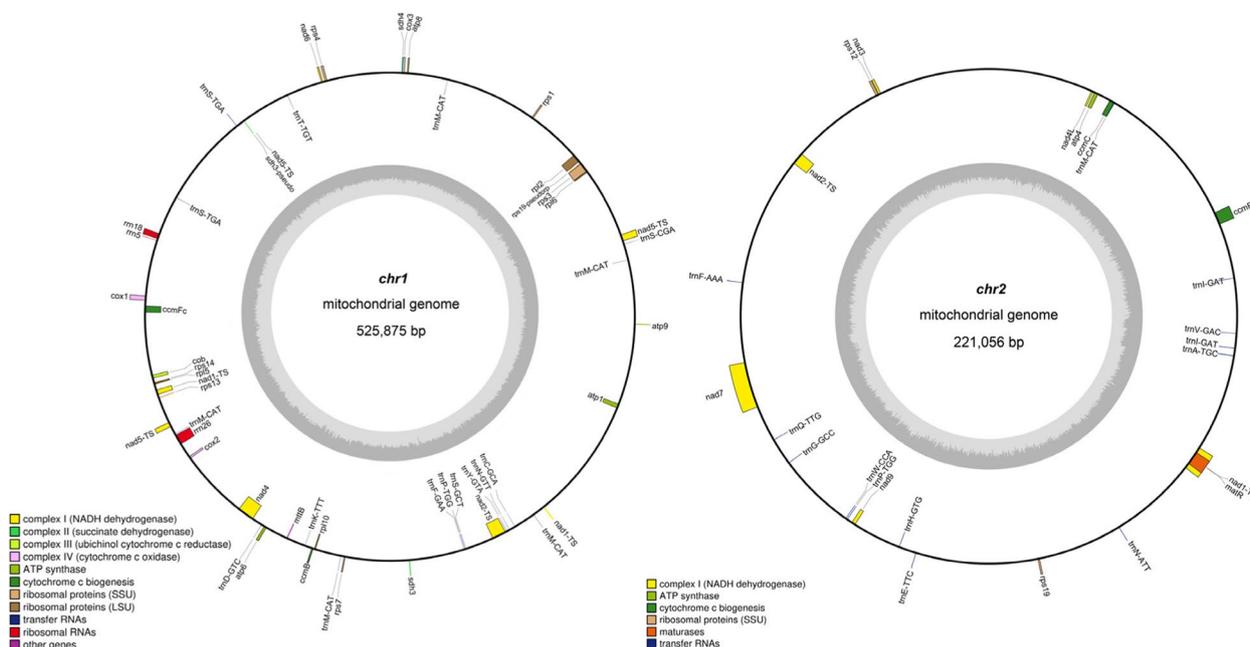
In the phylogenetic analysis of mitochondrial and chloroplast genomes, while we both primarily focused on species within the genus *Camellia*, the study adopted differentiated phylogenetic tree construction strategies due to the data imbalance between the 2 organelle genomes in the NCBI database (<http://www.ncbi.nlm.nih.gov/genome/organelle/>): For mitochondrial genomes with relatively scarce data, the mitochondrial CDS based phylogenetic tree incorporated 15 represented *Camellia* species (including *C. sinensis* variants, *C. oleifera*, *C. chekiangoleosa*, etc.) to investigate the phylogenetic position of *C. tachangensis*. It also included 14 species from different angiosperm families (e.g., Ericaceae, Solanaceae and Apiaceae) and the gymnosperm *Taxus wallichiana* as outgroups. Tbttools software (<https://github.com/CJ-Chen/TBtools/releases>) was utilized to extract 24 conserved mitochondrial protein-coding genes (PCGs) among these species [54], including *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*. The coding sequences (CDSs) of these genes within the mitochondrial genomes of 28 species were aligned via MAFFT v7.427 for interspecies sequence comparison [50, 51]. The aligned sequences were concatenated and trimmed via trimAl v1.4. Model prediction was subsequently conducted with jmodeltest v2.1.10 to identify the GTR model. The maximum likelihood phylogenetic tree was then constructed via RAxML v8.2.10 with the GTRGAMMA model and a bootstrap value of 1000 [55]. The Bayesian phylogenetic tree was constructed via MrBayes v3.2.7 with the Markov chain Monte Carlo method for 1,000,000 generations, and sampling trees every 100 generations [56]. In the phylogenetic analysis based on chloroplast PCGs, 23 representative

*Camellia* species (covering 10 significant sections including sect. *Thea*, sect. *Chrysantha*, and sect. *Oleifera*), due to they are closely related in *Camellia* genus. Meanwhile, the sister group (*Polyspora axillaris* and *Schima superba*) of *Camellia* genus was selected as the outgroup. The maximum likelihood phylogenetic tree and the Bayesian phylogenetic tree were constructed based on 53 conserved chloroplast PCGs among these species: *accD*, *atpA*, *atpE*, *atpF*, *atpH*, *atpI*, *matK*, *petA*, *petB*, *petD*, *petG*, *petL*, *petN*, *psaA*, *psaB*, *psaC*, *psbA*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*. The analysis methods employed were identical to those used for mitochondrial genomes. Finally, visualization was performed via Interactive Tree Of Life (ITOL) software v4.0 (<https://itol.embl.de/>) [57].

## Results

### Genomic features of *C. tachangensis* mitochondrial genome

The total mitochondrial DNA of *C. tachangensis* was sequenced, and the raw data were prepared for assembly, resulting in 16.34 Gb Illumina sequencing data (Q20 = 96.36%, Q30 = 90.75%) and 20.5 Gb Nanopore PromethION sequencing data with a N50 read length of 21,799 bp. The assembly results indicated that the mitochondrial genome sequence of *C. tachangensis* was 746,931 bp (GC content = 45.86%), consisting of one multibranch sequence and one circular sequence, which were designated chromosome 1 (Chr1) and chromosome 2 (Chr2), with lengths of 525,875 bp and 221,056 bp, respectively (Fig. 1). The results of the read mapping indicated that there were no reads present between Chr1 and Chr2 (Fig. S1), suggesting that Chr1 and Chr2 were relatively independent. Additionally, Chr1 was capable of producing 7 kinds of substructures, whereas Chr2 didn't exhibit any substructures. A total of 24 core protein-coding genes, 16 variable protein-coding genes, 3 ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes were identified. The core protein-coding genes could be categorized into 7 functional groups: ATP synthase (*atp1*, *atp4*, *atp6*, *atp8*, and *atp9*), Cytochrome c maturation proteins (*ccmB*, *ccmC*, *ccmFc*, and *ccmFn*), Ubiquinol cytochrome c reductase (*cob*), Cytochrome c oxidases (*cox1*, *cox2*, and *cox3*), Maturases (*matR*), Transport membrane proteins (*mttB*), and NADH dehydrogenases (*nad1*, *nad2*, and *nad3*) (Table 1). Notably, the exons of *nad1* and *nad2* were distributed on both Chr1 and Chr2; these segments require post-transcriptional RNA splicing to assemble into complete gene sequences. The analysis of 14 variable protein-coding genes revealed 8 types of small subunit ribosomal proteins (*rps1*, *rps13*, *rps14*, *rps3*, *rps4*, *rps7*, *rps12*, and *rps20*), 4 types of large subunit ribosomal proteins (*rpl10*, *rpl16*, *rpl2*, and *rpl5*), and 2 types of succinate dehydrogenases (*sdh3* and *sdh4*). Notably, both *sdh3*



**Fig. 1** Circular map of the mitochondrial genome of *C. tachangensis*

**Table 1** List of genes in the mitochondrial genome of *C. tachangensis*

	Group of genes	Gene name(Chr1)	Gene name(Chr2)
Core genes	ATP synthase	<i>atp1, atp6, atp8, atp9</i>	<i>atp4</i>
	Cytochrome c biogenesis	<i>ccmB, ccmFc</i>	<i>ccmC, ccmFn</i>
	Ubichinol cytochrome c reductase	<i>cob</i>	
	Cytochrome c oxidase	<i>cox1, cox2, cox3</i>	
	Maturases		<i>matR</i>
	Transport membrane protein	<i>mttB</i>	
	NADH dehydrogenase	<i>nad1-TS, nad2-TS, nad4, nad5-TS, nad6</i>	<i>nad1-TS, nad2-TS, nad3, nad4L, nad7, nad9</i>
Variable genes	Ribosomal proteins (LSU)	<i>rpl10, rpl16, rpl2, rpl5</i>	
	Ribosomal proteins (SSU)	<i>rps1, rps13, rps14, rps3, rps4, rps7, #rps19</i>	<i>rps12, rps19</i>
	Succinate dehydrogenase	<i>#sdh3, sdh3, sdh4</i>	
	Ribosomal RNAs	<i>rrn18, rrn26, rrn5</i>	
	Transfer RNAs	<i>trnC-GCA, trnD-GTC, trnF-GAA, trnK-TTT, trnM-CAT(5), trnN-GTT, trnP-TGG, trnS-CGA, trnS-GCT, trnS-TGA(2), trnT-TGT, trnY-GTA</i>	<i>trnA-TGC, trnE-TTC, trnF-AAA, trnG-GCC, trnH-GTG, trnI-GAT(2), trnM-CAT, trnN-ATT, trnP-TGG, trnQ-TTG, trnV-GAC, trnW-CCA</i>

Note: Numbers after gene names are the number of copies. Genes preceded by the # symbol represent pseudogenes. Genes followed by -TS need to be spliced into complete genes by RNA splicing after transcription

and *rps19* appeared twice in the genome: once as functional genes and another as pseudogenes. A total of 30 tRNA genes were annotated. Among these annotations, *trnM-CAU* was recorded 5 times on Chr1 and once on Chr2. Additionally, *trnS-UGA* was annotated twice on

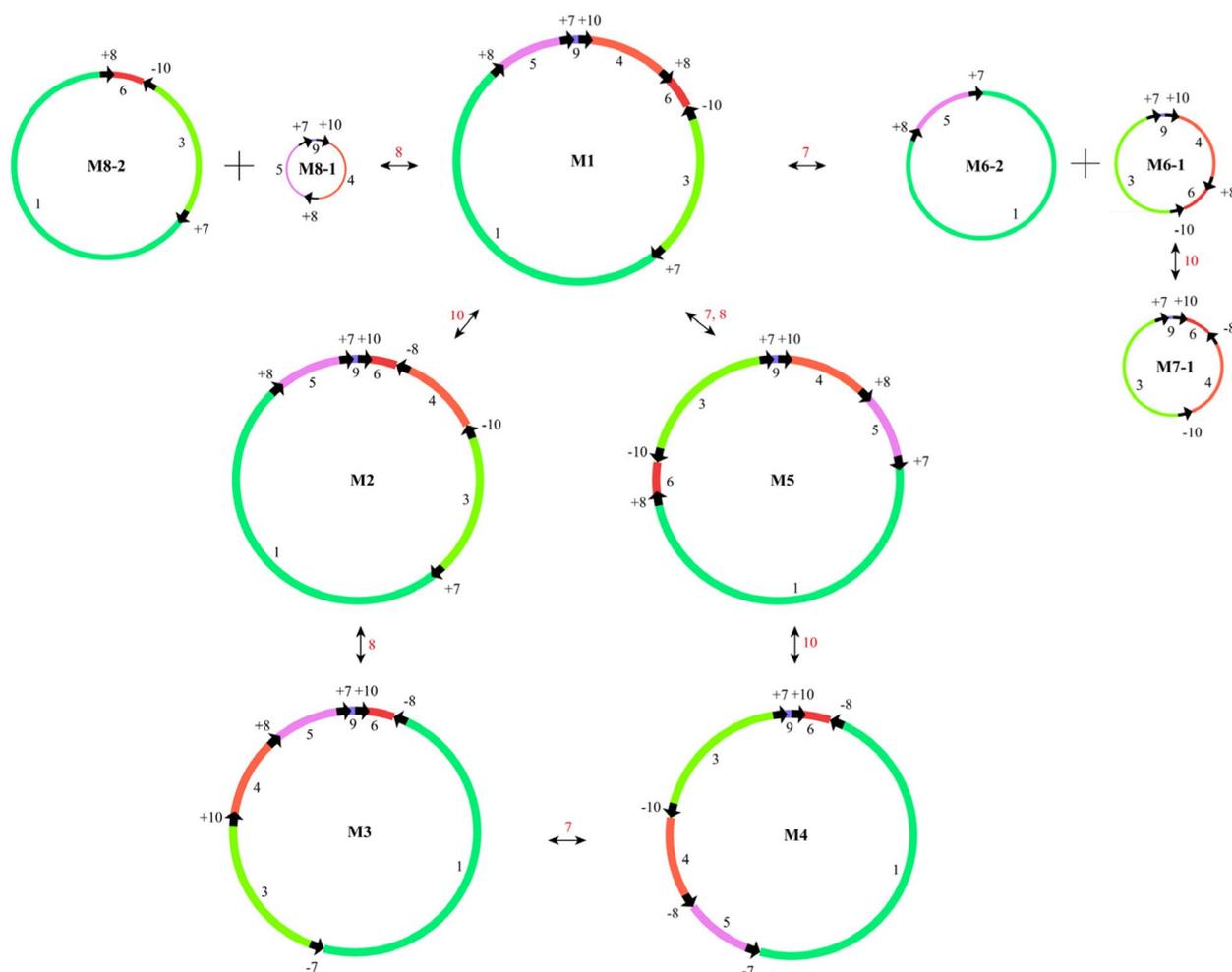
Chr1, *trnI-GAU* was annotated twice on Chr2, and *trnP-UGG* was noted once on both Chr1 and Chr2. Furthermore, 14 genes contained introns. Among these genes, ten possess one intron each (*ccmFc, rpl2, rps1, rps3, trnA-UGC, trnF-AAA, trnI-GAU (2), trnS-UGA, and*

*trnT*-UGU); one gene contained 2 introns (*nad4*); and 4 genes had 4 introns (*nad1*, *nad2*, *nad5*, and *nad7*).

**Different configurations of the *C. tachangensis* mitochondrial genome**

The phenomenon of recombination mediated by homologous fragments was commonly observed in the mitochondrial genome of cells. On Chr1, 3 pairs of dispersed repeats (homologous fragments), designated R7, R8, and R10, with lengths ranging from 1,190 to 8,440 bp, were identified. The similarity between the paired repeat units reached as high as 99.965% to 100%. Among these sequences, R7 and R8 were classified as direct repeats, whereas R10 was categorized as a palindromic repeat. Collectively, these 3 pairs of repeats facilitated the formation of 7 substructures within Chr1 (Fig. 2).

The homologous fragments of the mitochondrial genome in *C. tachangensis* mediated recombination through 2 distinct mechanisms, which were primarily determined by the orientation differences between these segments: (1) In M1, the arrangement directions of the homologous segments R7 and R8 within their respective groups were identical. When either R7 or R8 broke, the 2 homologous segments recombined in a head-to-tail manner, resulting in M1 splitting from a large ring into 2 smaller rings, producing either M8 or M6. If both R7 and R8 break simultaneously, M1 could recombine to form a new large ring designated M5. At this point, in a clockwise direction, the sequence order of the non-homologous fragments changes from C1 → C5 → C9 → C4 → C6 → C3 to C1 → C6 → C3 → C9 → C4 → C5. The distinction between M1 and M5 lies in the reciprocal positioning of C6 and C3 relative to C5. However, there



**Fig. 2** Hypothetical products generated by recombination mediated by R7, R8, and R10. The black arrows indicate the repetitive sequences R7, R8, and R10 (simply written as 7, 8, and 10) involved in recombination, with the arrow direction showing their orientation. The colored segments represent the DNA fragments C1, C3, C4, C5, C6, C9, and C10 (simply written as 1, 3, 4, 5, 6, 9, and 10) located between these repetitive sequences

is no alteration in the arrangement direction of each sequence. (2) In contrast, within M1, owing to opposing orientations between the 2 homologous segments associated with R10, an inversion phenomenon occurs among adjacent non-homologous fragments. Specifically, during the recombination from M1 to M2 mediated by R10, segment C6 → C4 experiences a 180° inversion in its clockwise orientation, thus altering the sequence order of the non-homologous fragments from C1 → C5 → C9 → C6 → C4 → C3 to C1 → C5 → C9 → C4 → C6 → C3. Furthermore, one homologous segment of R8 is located between fragments C4 and C6. Consequently, this inversion involving fragment pairings from C4 to C6 resulted in an inverse arrangement for that particular homologous segment: it is transformed from R+ 8 to R-8. Both homologs of R8 now exhibit opposite orientations, leading to an inverted recombination mechanism for R8 that allows for the transformation of M2 into M3. A similar outcome is observed during the transition from M5 to M4 through recombination processes. The above hypothesis can be observed in the coverage validation map aligned to the assembly results of long reads, confirming the presence of 7 potential substructures on Chr1 (Fig. S1).

This phenomenon revealed that recombination mediated by specific pairs of homologous segments within mitochondrial genomes could influence alternative pairs' modes of recombination, thereby enriching DNA with diverse substructures.

### Analysis of repeat sequences

A total of 223 SSRs were identified in the mitochondrial genome of *C. tachangensis* (Fig. 3, Table S1 A), with 160 located on Chr1 and 63 on Chr2. On Chr1 and Chr2, there are 19 and 12 mononucleotide (mono-), 48 and 12 dinucleotide (di-), 22 and 7 trinucleotide (tri-), and 61 and 25 tetranucleotide (tetra-) SSRs, respectively. There is 1 hexanucleotide (hexa-) SSR present on both Chr1 and Chr2. The highest proportion of SSRs on both chromosomes were found to be tetranucleotides, accounting for approximately 38.125% on Chr1 and 39.682% on Chr2. Furthermore, we observed that A/T is the most prevalent type among the mononucleotide SSRs. In total, there were also 28 tandem repeat sequences within the mitochondrial genome; the longest sequence was located on Chr1, with a copy number of 2, measuring 78 bp in length, whereas the shortest sequence resided on Chr2, with a copy number of 2, measuring only 24 bp in length (Table S1B). The mitochondrial genome contains a substantial number of dispersed repetitive sequences, totaling 479. This dataset included 266 palindromic repeat sequences and 213 forward repeat sequences, with lengths ranging from 29 to 8,452 bp (Table S1 C). Notably, 88.28% of these sequences were shorter than 100 bp,

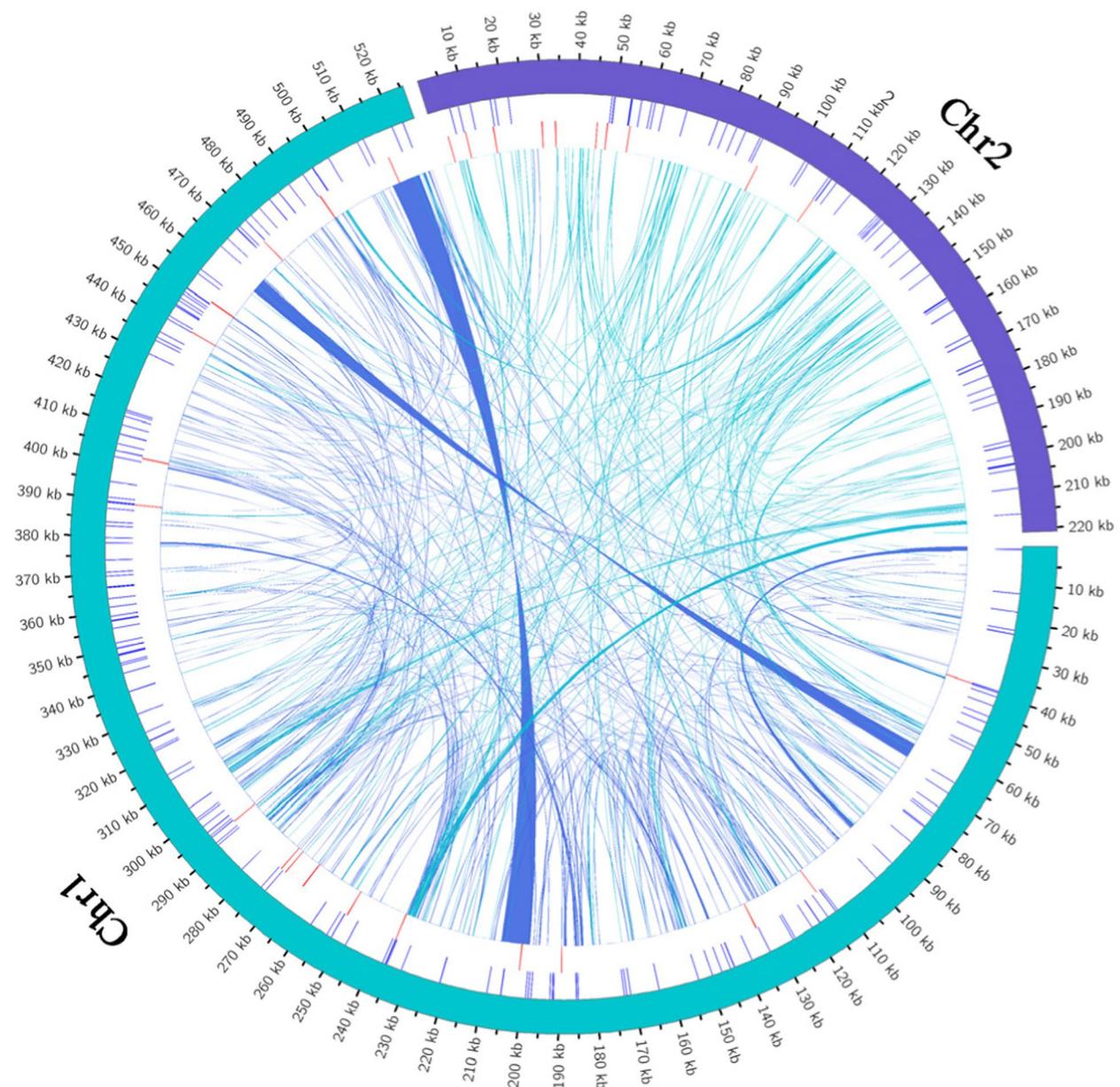
with the most common lengths falling between 29 and 49 bp. Furthermore, the analysis of dispersed repetitive sequences indicated that transposon exchange between Chr1 and Chr2 occurred quite frequently; 161 sequences were copied from Chr1 to Chr2, and only 28 sequences were transferred in the opposite direction (from Chr2 to Chr1). Among all the homologous fragments identified, only 3 segments exceeded a length of 1,000 bp, each exhibiting greater than or equal to 99.96% similarity. Of these longer segments, 2 were classified as direct repeat sequences, whereas one was categorized as a palindromic repeat sequence; these specific sequences played a role in mediating recombination within the mitochondrial genome.

### Prediction of RNA editing sites

In this study, we predicted RNA editing sites in the mitochondrial genome of *C. tachangensis*, focusing on 38 protein-coding genes. A total of 537 non-synonymous editing sites were identified (Fig. 4A, Table S2 A), involving changes in 14 amino acids, including H(His) → Y(Tyr), R(Arg) → C(Cys), T(Thr) → I(Ile), T(Thr) → M(Met), R(Arg) → W(Trp), S(Ser) → L(Leu), S(Ser) → F(Phe), P(Pro) → S(Ser), P(Pro) → L(Leu), P(Pro) → F(Phe), L(Leu) → F(Phe), A(Ala) → V(Val), Q(Gln) → \*, and Arg(R) → \* (\* represents a stop codon). Among these changes, the most common alteration was Ser to Leu, with a total of 128 RNA editing sites accounting for 23.84% of the total. Among all the amino acid changes observed, 259 (48.23%) of the hydrophilic amino acids were converted to hydrophobic ones; conversely, 39 (7.26%) of the hydrophobic amino acids were transformed into hydrophilic ones, whereas 235 (43.76%) amino acids exhibited no change in hydrophobicity. Additionally, 4 (0.74%) codons encoding hydrophilic amino acids were converted into stop codons (Fig. 4B). Specifically, 3 instances of CGA(R)—UGA(\*) conversion occurred at the last codon positions of the *ccmFc*, *atp9*, and *sdh4* genes; one instance of CAG(Q)—TAG(\*) conversion was found at the 13 codon position of the *rpl16* gene, which might lead to premature termination of mRNA translation. In terms of genetic analysis, the *ccmFn* gene presented the highest frequency of RNA editing sites, with a total of 39 occurrences. This was followed by the *ccmB* and *ccmC* genes, which had 34 and 32 instances, respectively. In contrast, the *sdh3* gene had the lowest frequency of RNA editing sites, with only 2 identified editing locations (Fig. 4C, Table S2B).

### Codon usage analysis of protein-coding genes (PCGs)

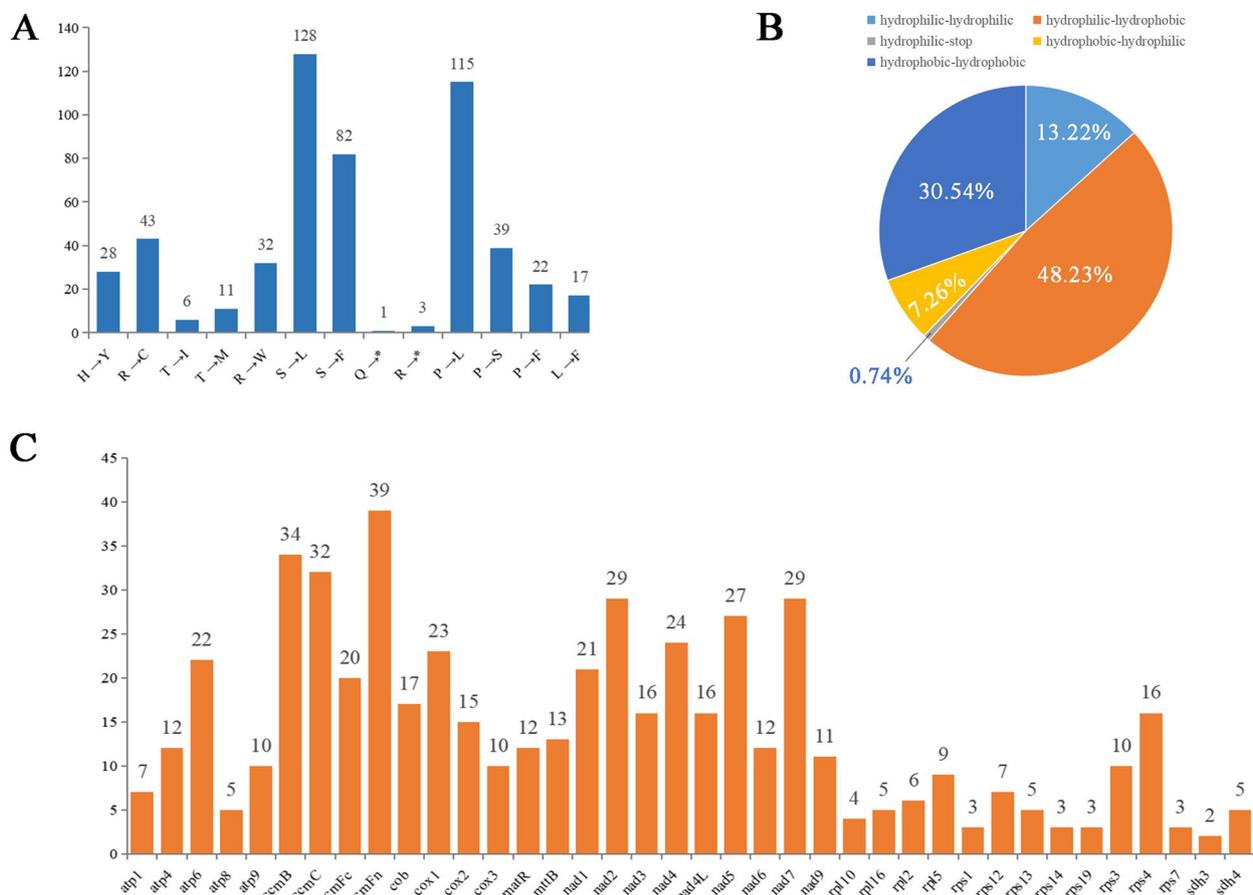
Relative synonymous codon usage frequency (RSCU) analysis was conducted on 64 codons of the mitochondrial genome of *C. tachangensis*. The results



**Fig. 3** Repeat analysis of the mitochondrial genome in *C. tachangensis*. The arc represents Chr1 (green) and Chr2 (purple). The ticks inner circles are SSR (Blue) and tandem repeat (red). The ribbons represents dispersed repeat

indicated that all 64 codons are expressed in PCGs (Fig. 5, Table S3). Among these, the GCU (encoding alanine) exhibited the highest RSCU value of 1.5743, whereas the CAC (encoding histidine) had the lowest RSCU value of 0.4586. In the PCGs, the start codon was consistently ATG, with no codon usage bias (RSCU = 1). The stop codons included UAA, UAG, and UGA, among which only UAA had an RSCU value greater than 1. Among the 61 coding amino acid codons

analyzed, 29 had RSCU values exceeding 1, indicating a strong preference for their use. There were 10 A-ending codons and 17 U-ending codons; conversely, there was only one C-ending or G-ending codon each. The proportion of high-frequency codons (RSCU > 1) ending with A or U reached 93.103%, whereas those ending with C or G accounted for only 6.897%. Therefore, it could be concluded that in the mitochondrial genome of *C. tachangensis*, there was a notable preference for the use of A- or U-ending codons.



**Fig. 4** Prediction of RNA editing sites in the *C. tachangensis* mitochondrial genome. (A: Characterization of RNA-editing sites; B: Proportion of different RNA-editing types; C: Numbers of RNA-editing sites in the mtDNA)

### Mitochondrial plastid DNAs (MTPTs) in the mitochondrial genome

To investigate the sequence transfer between the mitochondrial and chloroplast genomes of *C. tachangensis*, we conducted a comparative analysis of both organellar genomes. The results indicated that there were a total of 23 groups of chloroplast homologous fragments within the *C. tachangensis* mitochondrial genome, with MTPT 1–7 located on Chr1 and MTPT 8–23 located on Chr2 (Fig. 6, Table S4). These fragments collectively spanned a length of 16,396 bp, accounting for approximately 2.1951% of the total length of the mitochondrial genome. Among these fragments, MTPT1 was the longest at 9,556 bp and was located within the range of 221,056–211,509 bp on Chr2. In contrast, MTPT23 was the shortest fragment, with a length of only 32 bp, and was found within the range of 216,692–216,661 bp on Chr1. The annotation results indicate that these fragments originate from protein-coding genes, rRNA genes, tRNA genes, and intergenic regions of the chloroplast genome. However, all the chloroplast protein-coding genes and rRNA

coding genes were not retained after the insertion of the mitochondrial sequences, whereas the tRNA coding genes were preserved relatively intact within the mitochondria. A total of 7 completed tRNA genes were distributed across 7 homologous sequences: *trnA*-UGC, *trnI*-GAU, *trnV*-GAC, *trnW*-CCA, *trnP*-UGG, *trnM*-CAU, and *trnN*-GUU.

### Analysis of Pi and Ka/Ks

To analyze the sequence differences between *C. tachangensis* and its related species, we calculated the nucleotide diversity (Pi) values for 41 common genes across 9 species of the genus *Camellia* (Table S5). The data indicate significant differences in nucleotide diversity (Pi values) among the mitochondrial genomes of 9 *Camellia* species across different genes, ranging from 0 to 0.06435. The highest Pi value was observed in *rrn18* at (0.06345), followed by *nad5* (0.01662) and *cox2* (0.00253). While, 8 genes (e.g., *nad6*, *cox1*, and *nad4L*) exhibited Pi values of 0, indicating their high conservation. Some shorter regions showed a higher number of mutations despite

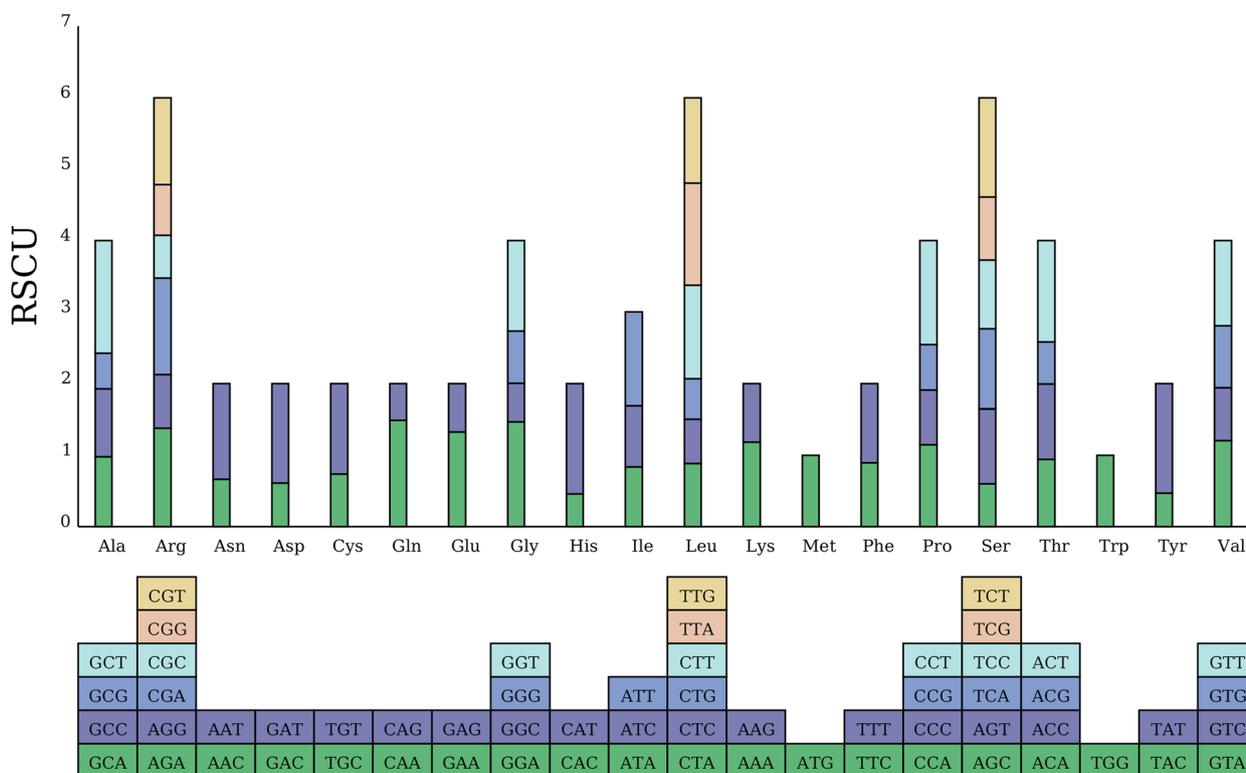


Fig. 5 Relative synonymous codon usage in the *C. tachangensis* mitochondrial genome

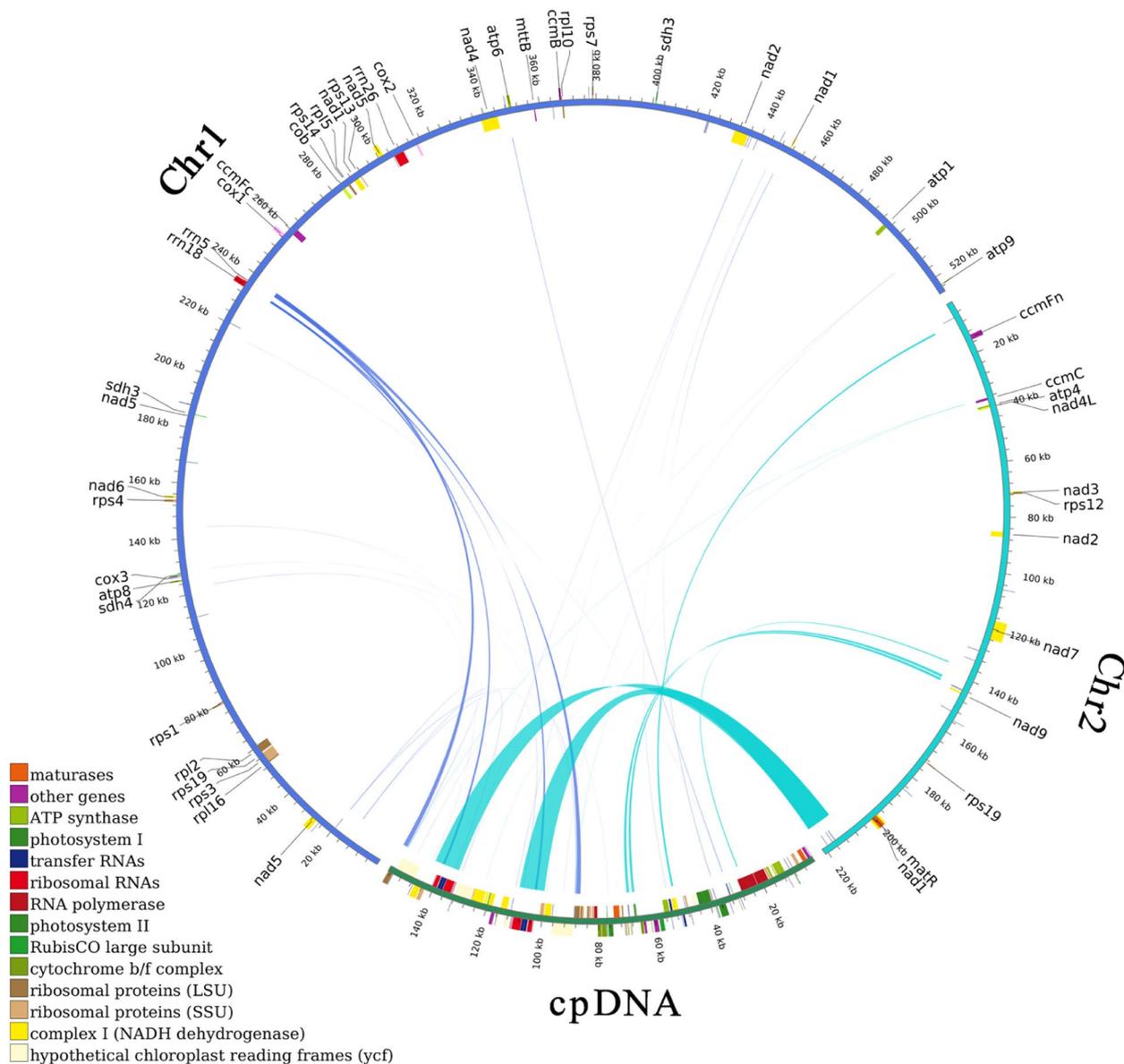
their limited length (e.g., *sdh3* [length: 321 bp; mutations: 6;  $P_i = 0.00848$ ]), while longer regions like *rrn26* (length: 3614 bp; mutations: 48;  $P_i = 0.00564$ ) displayed lower mutation density.

To further investigate the impact of environmental stress on mitochondrial PCG mutations in the aforementioned species, we conducted Ka/Ks analysis and screened 22 genes with Ka/Ks ratios. The results (Fig. 7, Table S6) showed that nearly all Ka/Ks ratios of mitochondrial PCGs in *C. tachangensis* were less than 1 when compared with *Camellia* species (only *cox2* showed  $Ka/Ks = 1.01438$  in *C. tachangensis* vs. *C. pubilimba* (ON782577), *nad1* showed  $Ka/Ks = 1.41208$  in *C. tachangensis* vs *C. assamica* (NC\_043914), and *rpl2* showed  $Ka/Ks = 1.03902$  in *C. tachangensis* vs. *C. gigantocarpa*). This indicated that *C. tachangensis*, as an endemic species of China, had undergone predominantly purifying selection during evolution. Notably, some genes exhibited a wide range of Ka/Ks ratios due to their higher mutation rates. For example, the protein-coding gene *nad5*, which had the highest  $P_i$  value, yielded 16 distinct Ka/Ks ratios, primarily ranging between 0.3 and 0.65. In contrast, certain genes such as *rpl2* and *sdh3*, demonstrated both conservation and heterogeneity across species comparisons, resulting in relatively limited Ka/Ks variations.

Taking *rpl2* as an example, its sequence exhibited either complete identity or divergence across different species comparisons, yielding only 3 distinct Ka/Ks ratios: 1.03902 (*C. tachangensis*/*C. chekiangoleosa*/*C. oleifera* vs. *C. gigantocarpa*), 0.783174 (*C. tachangensis*/*C. chekiangoleosa*/*C. oleifera* vs. *C. sinensis* variants PP212895/PP212896/ON782577), and 0.304929 (*C. sinensis* variants PP212895/PP212896/ON782577 vs. *C. gigantocarpa*).

### Phylogenetic analysis

The phylogenetic tree based on mitochondrial coding sequences (CDS) was constructed with *Taxus wallichiana* as the outgroup. This analysis revealed that *Diospyros kaki* (NC\_082859) and *Rhododendron simsii* (NC\_053763), both belonging to Ericales, formed a well-supported clade (BS = 100, PP = 1.00) alongside various *Camellia* species. Within the *Camellia* lineage, *C. tachangensis* was phylogenetically independent of species of sections *Oleifera*, *Camellia*, *Heterogenea* and *Chrysantha*. However, this phylogenetic tree also did not support the clustering of *C. tachangensis* with certain variants of *C. sinensis*, as their placement in Clade I exhibited extremely low support (BS = 22, PP = 0.41) (Fig. 8). The chloroplast PCG-based phylogenetic tree



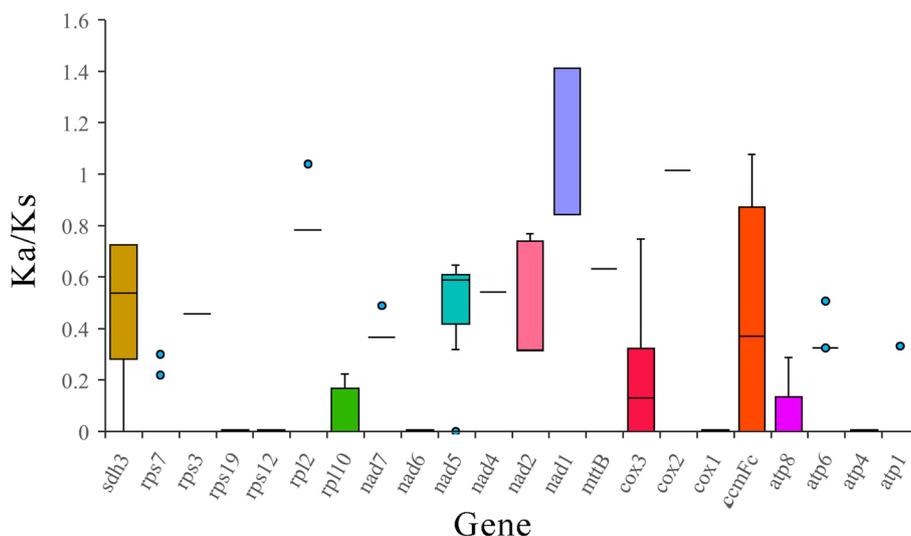
**Fig. 6** Schematic for the chloroplast-to-mitochondrial sequence transfer of *C. tachangensis*. The blue arc represents Chr1. The light green represents Chr2. The dark green arc represents chloroplast DNA. The homologous fragments are indicated by the connecting ribbons between the blue (light green) and dark green arcs

utilizing *Polyspora axillar*is (NC\_035709) and *Schima superba* (NC\_035545) as outgroups demonstrated high phylogenetic resolution for *C. tachangensis*: Among the *Camellia* species analyzed, this species formed a strongly supported basal group in Clade II (BS = 100, PP = 1.00), alongside *C. makuanica* (NC\_087766), *C. taliensis* (NC\_022264), and *C. gymnogyna* (NC\_039626) (Fig. 9). Although different species selections were employed for the two organellar PCGs' phylogenetic tree, the comparison between them revealed that the mitochondrial

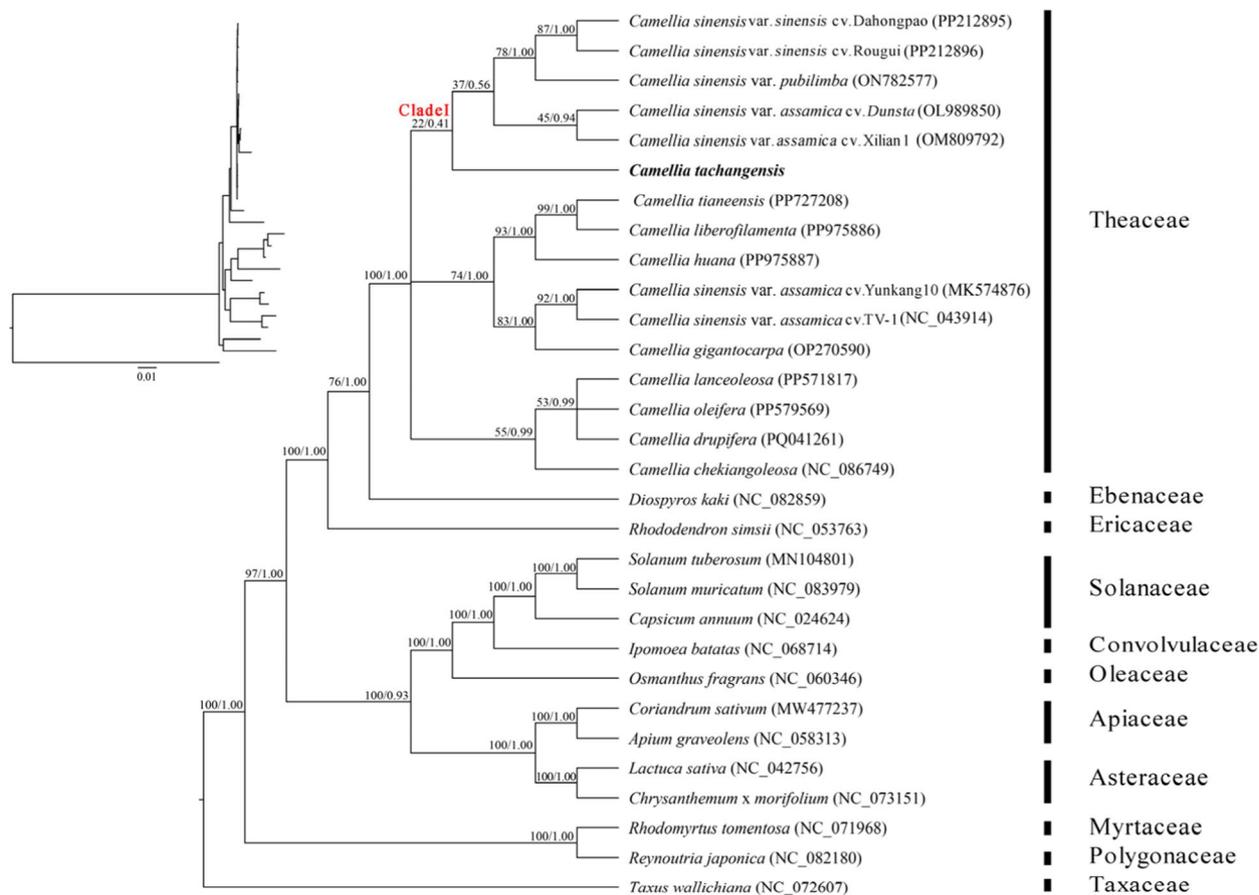
CDS-based phylogenetic tree did not support a basal position for *C. tachangensis* within the genus *Camellia*. Meanwhile, chloroplast PCGs demonstrated superior phylogenetic resolution for clarifying species relationships within the genus.

**Discussion**

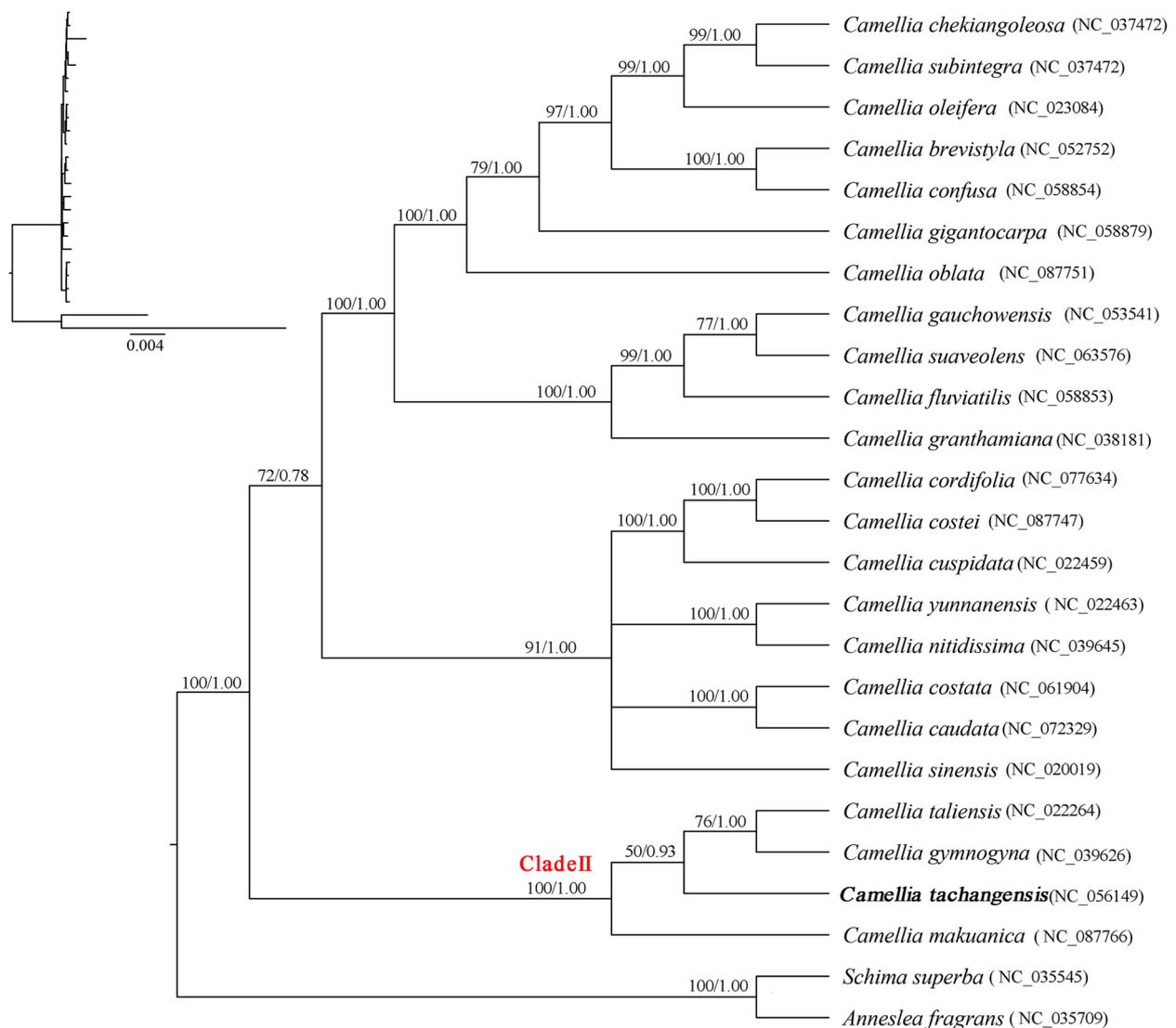
Although branched structures had been widely reported in plant mitochondrial genomes [58–60], the mitochondrial DNA of most *Camellia* species (e.g., *C. sinensis*,



**Fig. 7** Ka/Ks ratios of conserved PCGs in the mitochondrial genomes of 9 species of the genus *Camellia*



**Fig. 8** Construction of a phylogenetic tree based on mitochondrial CDS. (Maximum likelihood (ML) and Bayesian (BI) trees; BS and PP are indicated above the branches as BS/PP)



**Fig. 9** Construction of a phylogenetic tree based on chloroplast PGC. (Maximum likelihood (ML) and Bayesian (BI) trees; BS and PP are indicated above the branches as BS/PP)

*C. assamica*, and *C. nitidissima*) were traditionally presented as circular structures due to differing assembly strategies [29, 61–64]. Only 3 species of sect. *Oleifera* (*C. drupifera*, *C. oleifera*, and *C. lanceoleosa*) exhibit mitochondrial genomes with multibranch configurations [27, 28]. In this study, we achieved the first complete mitochondrial genome sequencing and assembly for *C. tachangensis*, which revealed its unique dual-component architecture comprising a multibranch sequence and a circular sequence. Subsequent structural analysis further identified the multibranch sequence could form 7 substructures via 3 pairs of dispersed repeats over 1000 bp. These findings not only enhanced our comprehension

of the complexity and diversity of mitochondrial genome structures in *Camellia* species, but also established a foundational genetic data framework for future research on molecular breeding programs targeting *C. tachangensis*.

Beyond structural variations, the mitochondrial genome length of *C. tachangensis* significantly differed from other *Camellia* species. The total mtDNA length of *C. tachangensis* was 746,931 bp, while other *Camellia* species range from 1,082,025 bp in *C. sinensis* var. *sinensis* cv. Dahongpao (CSSDHP, PP212895) (longest) to 707,441 bp in *C. sinensis* var. *assamica* cv. TV-1 (NC\_043914) (shortest), representing a 374,584 bp difference [27–29,

61–64]. Among these, *C. huana* (733,752 bp) showed the closest genome length to *C. tachangensis*, differing by only 13,179 bp. However, the length of repetitive sequences identified in the studies cannot directly explain the differences in mitochondrial DNA length between these *Camellia* species. This is evidenced by the contrasting repetitive sequence lengths between CSSDHP and *C. tachangensis*: SSR (4,548 vs. 2,688 bp, including tandem repeats) and dispersed repeats (33,871 vs. 46,509 bp). Even when combining the total lengths of simple sequence repeats and dispersed repeats, CSSDHP still exhibits a reverse correlation in total repetitive sequence content compared to *C. tachangensis* (38,419 vs. 49,197 bp). This paradoxical phenomenon may be attributed to 3 possible factors: first, differential loss and transfer of mitochondrial DNA fragments between the 2 species; second, frequent recombination and mutation events in intergenic regions of plant mitochondrial DNA that obscure detection of original repetitive sequences; third, CSSDHP's mitochondrial DNA had acquired longer chloroplast-derived homologous sequences compared to *C. tachangensis* (20,733 vs. 16,448 bp) [29, 65, 66]. However, the first 2 factors still require further exploration and validation specifically for *Camellia* species.

Although *C. tachangensis* possessed complete mitochondrial PCG composition, it only retained pseudogene copies for *rps19* and *sdh3*. In contrast, other *Camellia* species exhibited varying PCG duplications. For instance, CSSDHP contained 8 duplicated PCGs (e.g., *atp8*, *atp9*, *nad6*, *cox3*), *C. tianeensis* shows 2 duplicated PCGs (*ccmFn* and *rps16*), and *C. oleifera* retained 4 duplicated PCGs (*cox1*, *rpl16*, *rps3*, and *rpl2*) [28, 29, 64]. These differences likely resulted from combined effects of transposon activity and environmental adaptation [67], offering new perspectives for exploring plant mitochondrial genome evolution through further investigation. While other tRNA gene copy numbers vary among *Camellia* species, all of them (including *C. tachangensis*) exhibited high copies of the *trnM*-CAU gene. This might relate to its role in transporting the initiation codon AUG. *trnM*-CAU likely enhances its expression to competitively occupy ribosomal P-sites, preventing non-initiator tRNA misbinding and ensuring protein synthesis fidelity [68]. The GC content (45.86%) and codon usage bias of *C. tachangensis* showed remarkable conservation, being highly similar to both *Camellia* and other species, reflecting the evolutionary stability of these genetic features in Angiosperms [27–29, 58–64].

RNA editing is a widely occurring post-transcriptional mechanism that modified RNA by altering the types of nucleotides present within it [69]. To determine the final protein sequences of the mitochondrial genes in *C. tachangensis*, it is essential to predict RNA editing events

for each gene. In this study, 537 RNA editing sites across 38 genes in *C. tachangensis* were identified. Previous indicated that these editing sites played a crucial role in gene expression. RNA editing in plants could restore codons altered by mutations, thereby ensuring that mRNAs encode proteins with normal functionality [70]. In addition, RNA editing is a prerequisite for the proper translation of certain mRNAs. In the mitochondrial genome of *C. tachangensis*, the initial codons for *cox1* and *nad4L* are ACG. Through RNA editing, these start codons could be converted from ACG to ATG, thereby ensuring the proper function of mRNA translation. More importantly, RNA editing had been demonstrated to play a crucial role in regulating responses to environmental stress in certain plant species. Research indicated that specific RNA editing modifications—such as enhanced editing of mitochondrial genes *nad3*, *nad7*, and *ccmFn* in *Oryza sativa* L., alongside deficient editing of *nad4* and *cox3* in *Arabidopsis thaliana* (L.) Heynh.—are correlated with improved tolerance to salt and drought stress, respectively [71, 72]. However, although previous studies had described possible interactions between PLS-CsPPR proteins and target sequences of RNA editing sites in mitochondrial and chloroplast genes in *Camellia* species (*C. sinensis*) [73], a direct link between mitochondrial RNA editing and environmental stress adaptation in *Camellia* plants had not yet been established. To address this research gap, future studies could employ multi-omics correlation analysis methods, integrating the predicted RNA editing site data from this study, to investigate the potential mechanisms and roles of mitochondrial gene RNA editing in stress physiological responses of *C. tachangensis* and other *Camellia* species. This approach would provide theoretical support for developing precise conservation strategies.

DNA could be transferred between the mitochondrial and chloroplast genomes within cells [74]. This process was accompanied by the insertion of exogenous tRNA genes to support the translation of mitochondrial PCGs [75–77]. In this study, we identified a total of 23 MTPTs in the mitochondrial genome of *C. tachangensis*, among which 7 MTPTs contained 1–3 tRNA genes. By comparing these fragments with those from other species in the sect. *Thea* and sect. *Oleifera*, we found partial MTPTs shared similarities across these species: these fragments were highly similar in length, and their tRNA gene compositions were entirely consistent (e.g., *trnM*-CAT, *trnA*-UGC–*trnI*-GAU–*trnV*-GAC, and *trnD*-GUC). Therefore, we hypothesized that the transfer events of these fragments could be traced back to before the divergence of these 2 taxonomic groups. In contrast, MTPTs identified in 4 species of the sect. *Chrysanthia* were relatively scarce, with only 5–14 MTPTs per species. Moreover, only 1–2

MTPTs in each species contained tRNA genes, which might reflect the unique evolutionary trajectory of sect. *Chrysantha*.

The Ka/Ks analysis indicated that the mitochondrial PCGs of *C. tachangensis* had primarily undergone purifying selection during the course of evolution (Ka/Ks < 1), which aligned with previous Ka/Ks analysis results between *C. drupifera* (sect. *Oleifera*) and species from sect. *Thea* and sect. *Chrysantha* [27]. These phenomena suggest that purifying selection may play a dominant role in the evolution of mitochondrial PCGs in *Camellia* plants. It might stem from mitochondrial genes predominantly functioning in core metabolic pathways like oxidative phosphorylation. Non-synonymous mutations in these genes were often deleterious, resulting in their persistent purging through natural selection to maintain functional evolutionary conservation [78]. Furthermore, sequences of *rpl2* and *sdh3* genes detected in *C. tachangensis* completely match those of *C. chekiangoleosa* (sect. *Camellia*) but differed from cultivated variants (CSDHP, CSSRG) within the sect. *Thea*. This pattern might arise because *C. tachangensis*, as an early-diverged species of sect. *Thea*, retained primitive sequence characteristics of *rpl2* and *sdh3* genes shared with *C. chekiangoleosa* from the initial differentiation stage between sect. *Thea* and Sect. *Camellia* groups. In contrast, later-diverged *Camellia* species like CSDHP and CSSRG had accumulated mutations in these genes during evolution, ultimately resulting in sequence divergence from their corresponding genes in *C. tachangensis*.

Although the species selected for the 2 phylogenetic trees were not entirely consistent, we can still observe that the chloroplast PCG phylogenetic tree exhibits higher resolution compared to the mitochondrial CDS-based phylogenetic tree. Additionally, the phylogenetic position of *C. tachangensis* in the mitochondrial CDS-based tree did not appear at the base of *Camellia* species as observed in the chloroplast PCG-based phylogenetic tree. This discrepancy may be related to several factors: First, the lack of available mitochondrial genome data for closely related species such as *C. taliensis* and *C. gymnogyna* likely reduced the phylogenetic support for the branch containing *C. tachangensis*. Future studies should prioritize generating mitochondrial genome data for these species to resolve the phylogenetic placement of *C. tachangensis*. Additionally, studies had shown that mitochondrial genomes evolved at a slower rate compared to chloroplast genomes, resulting in smaller genetic distances among related species in mitochondrial genomes [79, 80]. Furthermore, although mitochondrial and chloroplast genomes were predominantly maternally inherited, both might undergo paternal leakage during inheritance, resulting in discrepancies in the genetic

lineages of these 2 organellar genomes within the same species [81]

## Conclusion

This study reported the first sequencing and annotation of the mitochondrial genome of *C. tachangensis*, which exhibited a multichromosomal structure, comprising a 525,875 bp branched molecule (resolvable into 7 sub-structures) and a 221,056 bp circular molecule. A total of 63 functional elements were annotated, including 30 protein-coding genes (PCGs), 30 tRNAs, and 3 rRNAs. Comparative analysis identified 23 homologous chloroplast-derived fragments in the mitochondrial genome, introducing 10 intact tRNA genes. Ka/Ks analysis indicated that PCGs evolved predominantly under purifying selection (Ka/Ks < 1). Phylogenetic analysis based on chloroplast genome analysis strongly supported *C. tachangensis* close relationship with *C. makuanica*, *C. taliensis*, and *C. gymnogyna* (BS = 100, PP = 1.00). However, the phylogenetic tree based on mitochondrial CDS failed to identify species closely related to *C. tachangensis* due to the current lack of comprehensive mitochondrial genome data for the genus *Camellia*. Despite this limitation, our study filled a critical gap in organelle genomics of *Camellia*, offering valuable genomic resources for elucidating evolutionary mechanisms, advancing genetic improvement programs, and informing conservation strategies for this ecologically and economically important genus.

## Abbreviations

PCGs	Protein-coding genes
mtDNA	Mitochondrial genome
cpDNA	Chloroplast genome
Ka/Ks	Non-synonymous/synonymous mutation ratio
RSCU	Relative synonymous codon usage
MTPPT	Mitochondrial plastid DNA sequence
tRNA	Transfer RNA
rRNA	Ribosomal RNA
SSR	Simple sequence repeat
Pi	Nucleotide diversity
BS	Bootstrap support value
PP	Posterior probabilities
PPR	Pentatricopeptide repeat

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11673-z>.

Supplementary Material 1  
 Supplementary Material 2  
 Supplementary Material 3  
 Supplementary Material 4  
 Supplementary Material 5  
 Supplementary Material 6  
 Supplementary Material 7

## Acknowledgements

We thank the Editors and the anonymous reviewers for their insightful comments and suggestions on the manuscript.

## Authors' contributions

Z.L.: Conceptualization, D.Z.J.: Writing—original draft, Data curation, Formal analysis, Software. Z.L.: Funding acquisition, Resources, Review & editing, Investigation. L.Z.: Investigation, Methodology. Z.H.R.: Resources, Supervision. X.X.: Visualization, Investigation. X.H.Y.: Methodology, Validation.

## Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 32400179), the Guizhou Provincial Basic Research Program (Natural Science) 2022 (072), the Guizhou University Student Innovation Project 2024 (302), and the 2024 Guizhou Science and Technology Innovation Talent Team Construction Project: Wildlife Innovation Team of the Forestry college of Guizhou University (Qian ke he ren cai CXTD[2025]053).

## Data availability

The mitogenome sequences supporting the conclusions of this article are available in GenBank (<https://www.ncbi.nlm.nih.gov/>) with accession numbers: PQ658231 and PQ658232.

## Declarations

### Ethics approval and consent to participate

All materials used in this study comply with international and national legal standards. The collected species material does not pose a threat to other species, and the collection of the species is recognized by the relevant authorities.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 13 December 2024 Accepted: 5 May 2025

Published online: 15 May 2025

## References

- Chang H. *Thea*—A Section of Beveragial Tea-Trees of the Genus *Camellia*. Acta Sci Nat. Univ. Sunyatseni. 1981;20(1):89–101.
- Chen L, Yu F, Tong Q. Discussions on Phylogenetic Classification and Evolution of Sect. *Thea*. J Tea Sci. 2000;02:89–94. <https://doi.org/10.13305/j.cnki.jts.2000.02.003>.
- Min T, Zhang W. The Evolution and Distribution of genus *Camellia*. Plant Diversity. 1996;01:1–13.
- Lu H, Shen J, Lin X, Fu J. Relevance of Fourier transform infrared spectroscopy and leaf anatomy for species classification in *Camellia* (Theaceae). Taxon. 2008;57(4):1274–8. <https://doi.org/10.1002/tax.574018>.
- Jin J, Dai W, Zhang C, Lin Z, Chen L. Genetic, morphological, and chemical discrepancies between *Camellia sinensis* (L.) O. Kuntze and its close relatives. J Food Compos Anal. 2022;108:104417. <https://doi.org/10.1016/j.jfca.2022.104417>.
- Chen K, Zhurbenko P, Danilov L, Matveeva T, Otten L. Conservation of an Agrobacterium cT-DNA insert in *Camellia* section *Thea* reveals the ancient origin of tea plants from a genetically modified ancestor. Front Plant Sci. 2022;13: 997762. <https://doi.org/10.3389/fpls.2022.997762>.
- Jiang Y, Yang J, Folk RA, Zhao J, Liu J, He Z, Peng H, Yang S, Xiang C, Yu X. Species delimitation of tea plants (*Camellia* sect. *Thea*) based on super-barcodes. BMC Plant Biol. 2024;24(1):181. <https://doi.org/10.1186/s12870-024-04882-3>.
- Liu W, Deng C, Chen X, Lu Y, Liao D. Determination of free amino acid and volatile aromatic compound in *Camellia tachangensis*. J Zhejiang Forestry Sci Technol. 2021;41(03):1–14. <https://doi.org/10.3969/j.issn.1001-3776.2021.03.001>.
- Liu W, Chen X, Deng C, Zhao L, Liao D, Lu Y. Research progress and protection status of *Camellia tachangensis*. Agric Technol. 2021;41(04):10–15. <https://doi.org/10.19754/j.nyyjs.20210228004>.
- Li C, Song Q, Fan Q, He Y, Zhao Z, Li F, Niu S, Chen Z. Comparative analysis of agronomical and quality traits of ancient trees and their clonal progenies in *Camellia tachangensis*. J South Agric. 2022;53(2):343–55. <https://doi.org/10.3969/j.issn.2095-1191.2022.02.007>.
- Yang C, Yang D, Su S, Liang S, Li Y, Guo Y, Qiao D, Mi X, Chen Z. Comparison of purine alkaloids and catechin components of wild *Camellia tachangensis* in Pu'an County and Panzhou County, Guizhou Province. Acta Agric Zhejiangensis. 2024;36(6):1232–44. <https://doi.org/10.3969/j.issn.1004-1524.20230972>.
- Ren N, Cheng L, Zhao Y, Zhao D. Tea plant  $\beta$ -1, 4-glycosyltransferase enhances the propagation of *Camellia tachangensis* F. C. Zhang by promoting graft wound healing. Sci Hortic. 2024;331:113156. <https://doi.org/10.1016/j.scienta.2024.113156>.
- Lu W, Huang C, Jin F. Ecological and chemical stoichiometric characteristics of tea tree leaves at different ages in *Camellia tachangensis*. Agric Technol. 2024;44(19):4–8. <https://doi.org/10.19754/j.nyyjs.20241015002>.
- Hao W, Ma J, Ma C, Jin J, Chen L. The complete chloroplast genome sequence of *Camellia tachangensis* F. C. Zhang (Theaceae). Mitochondrial DNA Part B. 2019;4(02): 3344–3345. <https://doi.org/10.1080/23802359.2019.1673247>.
- Huang D, Niu S, Bai D, Zhao Z, Li C, Deng X, Wang Y. Analysis of population structure and genetic diversity of *Camellia tachangensis* in Guizhou based on SNP markers. Mol Biol Rep. 2024;51(01):715. <https://doi.org/10.1007/s11033-024-09632-0>.
- Niu S, Song Q, Koiwa H, Qiao D, Zhao D, Chen Z, Liu X, Wen X. Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. BMC Plant Biol. 2019;19(1):328. <https://doi.org/10.1186/s12870-019-1917-5>.
- Logan DC. Plant mitochondrial dynamics. Biochim Biophys Acta, Mol Cell Res. 2006;1763(5–6):430–41. <https://doi.org/10.1016/j.bbamcr.2006.01.003>.
- Millar AH, Whelan J, Soole KL, Day DA. Organization and regulation of mitochondrial respiration in plants. Annu Rev Plant Biol. 2011;62:79–104. <https://doi.org/10.1146/annurev-arplant-042110-103857>.
- Boyer PD. The ATP synthase—a splendid molecular machine. Annu Rev Biochem. 1997;66:717–49. <https://doi.org/10.1146/annurev.biochem.66.1.717>.
- Millar AH, Eubel H, Jansch L, Kruff V, Heazlewood JL, Braun HP. Mitochondrial cytochrome c oxidase and succinate dehydrogenase complexes contain plant-specific subunits. Plant Mol Biol. 2004;56:77–90. <https://doi.org/10.1007/s11103-004-2316-2>.
- Møller IM, Rasmussen AG, Van Aken O. Plant mitochondria - past, present and future. Plant J. 2021;108(4):912–59. <https://doi.org/10.1111/tpj.15495>.
- Sato M, Sato K. Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochim Biophys Acta, Mol Cell Res. 2013;1833(08):1979–84. <https://doi.org/10.1016/j.bbamcr.2013.03.010>.
- Su X, Yang X, Hou T, Zeng X, Jian G. Comparative genomic and phylogenetic analyses of mitochondrial genome in oil palm. Genomics Appl Biol. 2024;43(08):1340–52. <https://doi.org/10.13417/j.gab.043.001340>.
- Lu H, Jiang W, Ghiassi M, Lee S, Nitin M. Classification of *Camellia* (Theaceae) species using leaf architecture variations and pattern recognition techniques. PLoS One. 2012;7(1):e29704. <https://doi.org/10.1371/journal.pone.0029704>.
- Xiao X, Li Z, Ran Z, Yan C, Tang M, Huang L. Taxonomic studies on five species of sect. *Tuberculata* (*Camellia* L.) based on morphology, pollen morphology, and molecular evidence. Forests. 2024; 15(10): 1718. <https://doi.org/10.3390/f15101718>.
- Ran Z, Li Z, Xiao X, Tang M. *Camellia nerifolia* and *Camellia ilicifolia* (Theaceae) as separate species: Evidence from morphology, anatomy, palynology, molecular systematics. Bot Stud. 2024;65(23). <https://doi.org/10.1186/s40529-024-00430-2>.
- Liang H, Qi H, Chen J, Wang Y, Liu M, Sun X, Wang C, Xia T, Feng X, Feng S, Chen C, Zheng D. Assembly and analysis of the first complete mitochondrial genome sequencing of main Tea-oil *Camellia* cultivars *Camellia drupifera* (Theaceae): revealed a multi-branch mitochondrial

- conformation for *Camellia*. BMC Plant Biol. 2025;25(1):13. <https://doi.org/10.1186/s12870-024-05996-4>.
28. Xiao Z, Gu Y, Zhou J, Lu M, Wang J, Lu K, Zeng Y, Tan X. De novo assembly of the complete mitochondrial genomes of two *Camellia* - oil tree species reveals their multibranch conformation and evolutionary relationships. Sci Rep. 2025;15(1):2899. <https://doi.org/10.1038/s41598-025-86411-2>.
  29. Li L, Li X, Liu Y, Li J, Zhen X, Huang Y, Ye J, Fan L. Comparative analysis of the complete mitogenomes of *Camellia sinensis* var. *sinensis* and *C. sinensis* var. *assamica* provides insights into evolution and phylogeny relationship. Front Plant Sci. 2024;15:1396389. <https://doi.org/10.3389/fpls.2024.1396389>.
  30. Pahlich E, Gerliz C. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry. 1980;19(1):11–3. [https://doi.org/10.1016/0031-9422\(80\)85004-7](https://doi.org/10.1016/0031-9422(80)85004-7).
  31. Modi A, Vai S, Caramelli D, Lari M. The Illumina sequencing protocol and the NovaSeq 6000 system. Methods Mol Biol. 2021;2242:15–42. [https://doi.org/10.1007/978-1-0716-1099-2\\_2](https://doi.org/10.1007/978-1-0716-1099-2_2).
  32. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
  33. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, Bioinformatics And Applications. Nat Biotechnol. 2021;39(11):1348–65. <https://doi.org/10.1038/s41587-021-01108-x>.
  34. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
  35. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36. <https://doi.org/10.1101/gr.215087.116>.
  36. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
  37. Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36(suppl\_2):W5–9. <https://doi.org/10.1093/nar/gkn201>.
  38. Lowe UM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64. <https://doi.org/10.1093/nar/25.5.955>.
  39. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-FINDER: a vector for high-throughput gene identification. Gene. 2002;282(1–2):33–41. [https://doi.org/10.1016/s0378-1119\(01\)00819-8](https://doi.org/10.1016/s0378-1119(01)00819-8).
  40. Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64. <https://doi.org/10.1093/nar/gkz238>.
  41. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003;106(3):411–22. <https://doi.org/10.1007/s00122-002-1031-0>.
  42. Ran Z, Li Z, Xiao X, An M, Yan C. Complete chloroplast genomes of 13 species of sect. *Tuberculata* Chang (*Camellia* L.): genomic features, comparative analysis, and phylogenetic relationships. BMC Genomics. 2024;25(108):108. <https://doi.org/10.1186/s12864-024-09982-w>.
  43. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80. <https://doi.org/10.1093/nar/27.2.573>.
  44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
  45. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45. <https://doi.org/10.1101/gr.092759.109>.
  46. Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinformatics. 2018;19(1):255. <https://doi.org/10.1186/s12859-018-2244-9>.
  47. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(2):348–55. <https://doi.org/10.1111/1755-0998.13096>.
  48. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870–4. <https://doi.org/10.1093/molbev/msw054>.
  49. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:244. <https://doi.org/10.1186/1471-2105-14-244>.
  50. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66. <https://doi.org/10.1093/nar/gkf436>.
  51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–800. <https://doi.org/10.1093/molbev/mst010>.
  52. Rozas J, Ferrer-Mata A, Sánchez-DeBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol. 2017;34(12):3299–302. <https://doi.org/10.1093/molbev/msx248>.
  53. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8(1):77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3).
  54. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>. (Epub 2020 Jun 23 PMID: 32585190).
  55. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
  56. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42. <https://doi.org/10.1093/sysbio/sys029>.
  57. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–9. <https://doi.org/10.1093/nar/gkz239>.
  58. Feng L, Wang Z, Wang C, Yang X, An M, Yin Y. Multichromosomal mitochondrial genome of *Punica granatum*: comparative evolutionary analysis and gene transformation from chloroplast genomes. BMC Plant Biol. 2023;23(1):512. <https://doi.org/10.1186/s12870-023-04538-8>.
  59. Jiang M, Ni Y, Zhang J, Li J, Liu C. Complete mitochondrial genome of *Mentha spicata* L. reveals multiple chromosomal configurations and RNA editing events. Int J Biol Macromol. 2023;251:126257. <https://doi.org/10.1016/j.ijbiomac.2023.126257>.
  60. Yang Z, Ni Y, Lin Z, Yang L, Chen G, Nijati N, Hu Y, Chen X. De novo assembly of the complete mitochondrial genome of sweet potato (*Ipomoea batatas* [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination. BMC Plant Biol. 2022;22(1):285. <https://doi.org/10.1186/s12870-022-03665-y>.
  61. Lu C, Gao L-Z, Zhang Q-J. A High-Quality Genome Assembly of the Mitochondrial Genome of the Oil-Tea Tree *Camellia gigantocarpa* (Theaceae). Diversity. 2022;14(10):850. <https://doi.org/10.3390/d14100850>.
  62. Rawal HC, Kumar PM, Bera B, Singh NK, Mondal TK. Decoding and analysis of organelle genomes of Indian tea (*Camellia assamica*) for phylogenetic confirmation. Genomics. 2020;112(1):659–68. <https://doi.org/10.1016/j.ygeno.2019.04.018>.
  63. Li J, Tang H, Luo H, Tang J, Zhong N, Xiao L. Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. *assamica* cv. *Dunsa*. Frontiers in Plant Science. 2023; 14: 1117002. <https://doi.org/10.3389/fpls.2023.1117002>.
  64. Li Z, Ran Z, Xiao X, Yan C, Xu J, Tang M, An M. Comparative analysis of the whole mitochondrial genomes of four species in sect. *Chrysantha* (*Camellia* L.), endemic taxa in China. BMC Plant Biology. 2024; 24(1): 955. <https://doi.org/10.1186/s12870-024-05673-6>.
  65. Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol. 2013;5(6):1079–86. <https://doi.org/10.1093/gbe/evt069>.
  66. Llorente B, Smith CE, Symington LS. Break - induced replication: what is it and what is it for? Cell Cycle. 2008;7(7):859–64. <https://doi.org/10.4161/cc.7.7.5613>.

67. Hassan AH, Mokhtar MM, El Allali A. Transposable elements: multifunctional players in the plant genome. *Front Plant Sci.* 2024;14:1330127. <https://doi.org/10.3389/fpls.2023.1330127>.
68. Kapoor S, Das G, Varshney U. Crucial contribution of the multiple copies of the initiator tRNA genes in the fidelity of tRNA(fMet) selection on the ribosomal P-site in *Escherichia coli*. *Nucleic Acids Res.* 2011;39(1):202–12. <https://doi.org/10.1093/nar/gkq760>.
69. Lo Giudice C, Hernández I, Ceci LR, Pesole G, Picardi E. RNA editing in plants: A comprehensive survey of bioinformatics tools and databases. *Plant Physiol Biochem.* 2019;137:53–61. <https://doi.org/10.1016/j.plaphy.2019.02.001>.
70. Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA editing in plants and its evolution. *Annu Rev Genet.* 2013;47:335–52. <https://doi.org/10.1146/annurev-genet-111212-133519>.
71. Chen G, Zou Y, Hu J, Ding Y. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. *BMC Genomics.* 2018;19(1):720. <https://doi.org/10.1186/s12864-018-5088-9>.
72. Sechet J, Roux C, Plessis A, Effroy D, Frey A, Perreau F, Biniek C, Krieger - Liszkay A, Macherel D, North HM, Mireau H, Marion - Poll A. The ABA - deficiency suppressor locus HAS2 encodes the PPR protein LOI1/MEF11 involved in mitochondrial RNA editing. *Molecular Plant.* 2015;8(4):644 - 656. <https://doi.org/10.1016/j.molp.2014.12.005>.
73. Zhang M, Li Z, Wang Z, Xiao Y, Bao L, Wang M, An C, Gao Y. Exploring the RNA Editing Events and Their Potential Regulatory Roles in Tea Plant (*Camellia sinensis* L.). *International Journal of Molecular Sciences.* 2022;23(21):13640. <https://doi.org/10.3390/ijms232113640>.
74. Dietrich A, Small I, Cosset A, Weil JH, Maréchal-Drouard L. Editing and import: Strategies for providing plant mitochondria with a complete set of functional transfer RNAs. *Biochimie.* 1996;78(6):518–29. [https://doi.org/10.1016/0300-9084\(96\)84758-4](https://doi.org/10.1016/0300-9084(96)84758-4).
75. Ceci LR, Veronico P, Callerani R. Identification and mapping of tRNA genes on the *Helianthus annuus* mitochondrial genome. *DNA Seq.* 1996;6(3):159–66. <https://doi.org/10.3109/10425179609010203>.
76. Sangaré A, Weil JH, Grienemberger JM, Fauron C, Lonsdale D. Localization and organization of tRNA genes on the mitochondrial genomes of fertile and male sterile lines of maize. *Mol Gen Genet.* 1990;223(2):224–32. <https://doi.org/10.1007/BF00265058>.
77. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet.* 1997;15(1):57–61. <https://doi.org/10.1038/ng0197-57>.
78. Mower JP, Sloan DB, Alverson AJ. Plant Mitochondrial Genome Diversity: The Genomics Revolution. In: Wendel J, Greilhuber J, Dolezel J, Leitch I, eds. *Plant Genome Diversity Volume 1*. Springer, Vienna; 2012:123 - 144. [https://doi.org/10.1007/978-3-7091-1130-7\\_9](https://doi.org/10.1007/978-3-7091-1130-7_9).
79. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA.* 1987;84(24):9054–8. <https://doi.org/10.1073/pnas.84.24.9054>.
80. Palmer JD, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol.* 1988;28(1–2):87–97. <https://doi.org/10.1007/BF02143500>.
81. Bentley KE, Mandel JR, McCauley DE. Paternal leakage and heteroplasmy of mitochondrial genomes in *Silene vulgaris*: evidence from experimental crosses. *Genetics.* 2010;185(3):961–8. <https://doi.org/10.1534/genetics.110.115360>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.