

RESEARCH

Open Access



# Mitochondrial genome assembly of the Chinese endemic species of *Camellia luteoflora* and revealing its repetitive sequence mediated recombination, codon preferences and MTPTs

Xu Xiao<sup>1</sup>, Zhaohui Ran<sup>1</sup>, Chao Yan<sup>1</sup>, Weihao Gu<sup>1</sup> and Zhi Li<sup>1\*</sup>

## Abstract

*Camellia luteoflora* Y.K. Li ex Hung T. Chang & F.A. Zeng belongs to the *Camellia* L. genus (Theaceae Mirb.). As an endemic, rare, and critically endangered species in China, it holds significant ornamental and economic value, garnering global attention due to its ecological rarity. Despite its conservation importance, genomic investigations on this species remain limited, particularly in organelle genomics, hindering progress in phylogenetic classification and population identification. In this study, we employed high-throughput sequencing to assemble the first complete mitochondrial genome of *C. luteoflora* and reannotated its chloroplast genome. Through integrated bioinformatics analyses, we systematically characterized the mitochondrial genome's structural organization, gene content, interorganellar DNA transfer, sequence variation, and evolutionary relationships. Key findings revealed a circular mitochondrial genome spanning 587,847 bp with a GC content of 44.63%. The genome harbors 70 unique functional genes, including 40 protein-coding genes (PCGs), 27 tRNA genes, and 3 rRNA genes. Notably, 9 PCGs contained 22 intronic regions. Codon usage analysis demonstrated a pronounced A/U bias in synonymous codon selection. Structural features included 506 dispersed repeats and 240 simple sequence repeats. Comparative genomics identified 19 chloroplast-derived transfer events, contributing 29,534 bp (3.77% of total mitochondrial DNA). RNA editing prediction revealed 539 C-to-T conversion events across PCGs. Phylogenetic reconstruction using mitochondrial PCGs positioned *C. luteoflora* in closest evolutionary proximity to *Camellia sinensis* var. *sinensis*. Selection pressure analysis (Ka/Ks ratios < 1 for 11 PCGs) and nucleotide diversity assessment (Pi values: 0–0.00711) indicated strong purifying selection and low sequence divergence. This study provides the first comprehensive mitochondrial genomic resource for *C. luteoflora*, offering critical insights for germplasm conservation, comparative organelle genomics, phylogenetic resolution, and evolutionary adaptation studies in *Camellia* species.

**Keywords** *Camellia luteoflora*, Mitochondrial genome, Homologous recombination, RNA editing events

\*Correspondence:

Zhi Li

zli7@gzu.edu.cn

<sup>1</sup>College of Forestry, Guizhou University, Guiyang 550025, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

*Camellia luteoflora* Y.K. Li ex Hung T. Chang & F.A. Zeng, a perennial evergreen shrub or small tree within the *Camellia* L. genus (Theaceae Mirb.), was first discovered in Chishui City, Guizhou Province, China, in November 1981. Prof. Chang Hongda formally described it as a novel species, designating the taxonomic classification “*Camellia* L.– sect. *Luteoflora* Chang” [1]. This species exhibits a highly restricted distribution, primarily limited to the Jinsha Gou and Sidong Gou valleys in Chishui, Guizhou [2, 3], with sporadic occurrences reported in adjacent regions such as Guihua Town, Gulin County, Sichuan [4]. Recognized for its ecological and botanical significance, *C. luteoflora* was designated as a nationally protected plant in China in 1983, with strict prohibitions on wild harvesting or transplantation [5]. Subsequent conservation efforts included the establishment of the Chishui *Alsophila spinulosa* Nature Reserve in 1984, which incorporated the species’ native habitat into its protected zones and prioritized its preservation [6]. By 1988, it was classified as a Grade I rare and endangered plant in Guizhou Province [6]. The 2013 Red List of China’s Biodiversity– Higher Plants Volume further assessed its conservation status as Vulnerable (VN), emphasizing its endemic status and escalating threats [7]. Currently, anthropogenic pressures, including habitat degradation and fragmentation, have precipitated a severe population decline in *C. luteoflora*. Both population size and individual numbers are diminishing rapidly, pushing this taxon toward critical endangerment. Urgent interdisciplinary conservation strategies integrating genomic insights, habitat restoration, and policy enforcement are imperative to mitigate its extinction risk.

Recent years have witnessed growing scientific interest in *C. luteoflora*, with studies spanning multiple aspects of its biology and ecology. Zhang Ting investigated the species’ asexual propagation potential, demonstrating that hormone treatments significantly influence rooting efficiency in cuttings derived from wild germplasm within its natural habitat [8]. Concurrently, Liu Haiyan identified optimal seed germination conditions through systematic propagation experiments, advancing cultivation protocols for this species [9]. Further contributions by Zou Tiancai et al. integrated comparative trait polarity analysis, biogeographical approaches, and phylogenetic evolutionary frameworks to elucidate the origin, distribution patterns, and adaptive features of *C. luteoflora*, complemented by investigations into its leaf anatomy and photosynthetic characteristics [10, 11]. While current research has prioritized unraveling the species’ evolutionary origins, refining breeding techniques, mapping spatial distributions, and diagnosing endangerment drivers [12, 13], critical gaps remain in genetic studies. Notably, comprehensive genomic analyses of its mitochondrial

and chloroplast genomes key resources for resolving phylogenetic relationships and genetic diversity are still lacking.

Mitochondria, the primary energy-producing organelles in eukaryotic cells, drive aerobic respiration and play multifaceted roles in regulating critical metabolic processes such as cell differentiation, apoptosis, proliferation, and stress response [14]. Beyond their metabolic functions, they are intrinsically linked to plant growth vigor and cytoplasmic male sterility (CMS), traits of significant agricultural and evolutionary importance [15]. These features position mitochondrial studies as vital tools for investigating eukaryotic evolution, species identification, genetic diversity, and molecular breeding strategies [16]. The mitochondrial genome exhibits distinctive characteristics: while relatively compact in size with a conserved gene repertoire and dense gene arrangement, it also contains hypervariable noncoding regions that contribute to genomic diversity [17]. This contrasts with chloroplast genomes in higher plants, which display minimal homologous recombination and maintain strict conservation in gene number, order, and composition [18]. Notably, mitochondrial genomes balance evolutionary conservation with unique divergence patterns, evolving at rates distinct from nuclear genes. Their relatively large size and structural complexity provide rich taxonomic information, enabling resolution of classification challenges among closely related species [19, 20].

Plant mitochondrial DNA (mtDNA) exhibits remarkable structural complexity, characterized by dynamic configurations including primary circular molecules, subgenomic circular forms, linear arrangements, and highly branched multigenomic architectures [21, 22]. For instance, mitochondrial genome assembly in *Salvia miltiorrhiza* Bunge revealed two distinct unit maps reflecting its branched multigenomic organization [23, 24]. This structural diversity is further amplified by abundant repetitive sequences within plant mtDNA, which drive frequent homologous recombination events [25, 26]. Consequently, mitochondrial genomes may exist as singular or multiple circular/linear DNA conformations, often coexisting within a single organism [27]. To explain this plasticity, researchers have proposed models such as the “master circle” hypothesis and the multichromosome framework. The former posits a dynamic multipartite system, where a primary “master circle” containing the full genomic content interconverts with smaller subgenomic circles via recombination at repetitive regions [28]. Advances in hybrid sequencing strategies (combining second- and third-generation technologies) have now enabled precise resolution of these intricate multipartite structures [29].

In this study, we present the first complete assembly and annotation of both mitochondrial and chloroplast

genomes for *C. luteoflora*. Through comparative analyses, we characterized fundamental genomic architecture, structural conformations, GC content, codon usage bias, repetitive elements, RNA editing sites, and evolutionary selection pressures (Ka/Ks ratios) in its mitochondrial genome. Furthermore, we systematically investigated interorganellar sequence transfers between mitochondrial and chloroplast genomes, providing critical insights into DNA exchange mechanisms and functional interdependencies between these organelles. These findings advance understanding of plant mitochondrial genome organization—particularly repeat-mediated recombination dynamics—while establishing a foundation for exploring organellar genome coevolution and horizontal gene transfer within the *Camellia* genus.

## Materials and methods

### Collection of plant material, DNA extraction and its sequencing

*C. luteoflora* specimens were collected from their natural habitat in Jinshagou, Yuanhou Town, Chishui City, Guizhou Province, China. Taxonomic identification was performed by Dr. Li Zhi (Professor of Plant Systematics, School of Forestry, Guizhou University). Voucher specimens (accession number: LZ-20240106) were deposited in the Guizhou University Forestry Herbarium (GZAC). Fresh, healthy leaves from disease-free individuals were selected for genomic analysis, snap-frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until processing. High-quality genomic DNA was extracted from leaf tissues using the Plant Genomic DNA Kit (DP305, Tiangen Biotech, Beijing, China). Sequencing was conducted on complementary platforms to ensure comprehensive coverage: short-read sequencing via the Illumina NovaSeq 6000 system (San Diego, CA, USA) and long-read sequencing using the Oxford Nanopore PromethION platform (Oxford, UK).

### Genome assembly and annotation

Long-read PromethION data underwent initial alignment to reference gene sequences using Minimap2 (v2.1) [30] for mitogenome reconstruction, followed by error correction with Canu (v2.2) [31]. To enhance assembly accuracy, second-generation sequencing (Illumina) reads were mapped to the corrected long-read sequences using Bowtie2 (v2.3.5.1) [32]. A hybrid assembly strategy was then implemented with Unicycler (v0.4.8) [33], integrating both short-read and corrected long-read datasets under default parameters. The resulting assembly was visualized and refined through manual inspection using Bandage (v0.8.1) [34] to resolve topological ambiguities. For mitochondrial genome annotation, GeSeq [35] was employed with the *Camellia sinensis* mitochondrial genome (GenBank: PP212896) as a reference. The final

annotation was manually curated to conform to a circular genome model. Chloroplast genome assembly utilized the same specimen's Illumina data processed through GetOrganelle (v1.7.7.0) [36], followed by annotation with the Plastid Genome Annotator (PGA) [37]. Genome maps were generated using OGDRAW (v1.3.1) [38] to illustrate structural features.

### Validation of mitochondrial genome repeat-mediated recombination in *Camellia luteoflora*

To elucidate recombination patterns in intra- and inter-loop structures, we implemented a multi-step analytical workflow. First, circular sequences assembled by Unicycler were subjected to pairwise alignment using BLASTN v2.12.0 (parameters:  $e\text{-value} \leq 1e-5$ ), with subsequent filtration retaining only those alignments spanning  $\leq 100$  bp to exclude low-complexity regions [39]. For long-read validation, we imposed stringent mapping criteria requiring  $\geq 500$  bp flanking sequence coverage on both sides of repetitive elements to ensure reliable resolution of repeat structures. Next-generation sequencing data from three distinct platforms were systematically aligned against the assembly using minimap2 v2.28 (preset:  $-x$  map-ont), followed by iterative realignment and interactive visualization of consensus regions through the same software pipeline. This multi-platform validation strategy enhanced the robustness of structural determination particularly in complex repetitive regions.

### *Camellia luteoflora* mitochondrial genome repeat sequence analysis

To characterize repetitive elements in the mitochondrial genome of *C. luteoflora*, we conducted systematic analyses through two computational approaches. Firstly, microsatellite identification was performed using MISA (<https://webblast.ipk-gatersleben.de/misa/>) [40] with optimized detection thresholds: mononucleotide repeats required  $\geq 10$  iterations, dinucleotides  $\geq 5$  iterations, trinucleotides  $\geq 4$  iterations, while tetra-, penta-, and hexanucleotide motifs required  $\geq 3$  iterations. These thresholds were established to ensure biological relevance while minimizing false positives from random sequence variations. Subsequently, complex repeat architecture was investigated using REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [41] with stringent parameters: an edit distance  $\leq 3$  and minimum repeat length  $\geq 300$  bp for detecting four types of sequence duplications - direct, inverted, palindromic, and complementary repeats. The dual analytical framework enabled comprehensive detection of both short tandem repeats and large-scale structural duplications, providing critical insights into the organization and evolutionary dynamics of this mitochondrial genome.

### ***Camellia luteoflora* mitochondrial genome codon preference analysis**

The mitochondrial genome of *C. luteoflora* underwent systematic nucleotide composition analysis, with annotated protein-coding genes (PCGs) being computationally extracted for downstream characterization. Codon usage bias was quantitatively assessed using CodonW v1.4.2 through three-dimensional metrics: (1) whole-genome base distribution patterns, (2) relative synonymous codon usage (RSCU) calculations, and (3) amino acid-specific codon preference profiling. The RSCU metric, defined as the ratio between observed codon frequency and its theoretical expectation under neutral evolution (RSCU = 1 indicates no bias), revealed distinct translational selection pressures. Codons with RSCU > 1 demonstrated preferential utilization, where magnitude positively correlated with selection intensity [41]. Notably, extreme RSCU values (> 1.6) were interpreted as molecular signatures of evolutionary optimization for translation efficiency or tRNA abundance adaptation.

### **Prediction of RNA editing sites**

In order to obtain the RNA editing information in the mitochondrial gene sequences of *C. luteoflora*, we utilized the PREPACT v3.0 software (<http://www.prepact.de/prepact-main.php>) to perform the calculations [42]. According to all the RNA editing sites of *C. luteoflora*, the RNA editing occurring at the first, second, and third codons was counted separately; combined with the amino acid changes, we analyzed the changes in hydrophilicity and hydrophobicity of amino acids induced by RNA editing and the changes in the start codon and the stop codon induced by RNA editing.

### **Ka/Ks and nucleotide diversity analysis of the mitochondrial genome of *Camellia luteoflora***

To elucidate the evolutionary trajectory of *Camellia sinensis* mitochondrial genome, we implemented a three-tiered comparative framework with six congeneric species (*C. tianeensis*, *C. fangchengensis*, *C. chekiangoleosa*, *C. sinensis* cv. Rougui, *C. sinensis* var. *pubilimba*, and *C. sinensis* reference genome). Firstly, core-gene phylogeny was reconstructed using MEGA11 (algorithm: Neighbor-Joining, bootstrap = 1000) through whole mitochondrial genome alignment, establishing the topological relationships among taxa. Secondly, evolutionary constraint metrics were calculated via DnaSP v6.12.03: (1) Non-synonymous/synonymous substitution ratios (Ka/Ks) with sliding window analysis (window = 150 bp, step = 30 bp) to detect selection signatures, and (2) Nucleotide diversity ( $\pi$ ) calculations (window = 1000 bp) identifying conserved versus hypervariable genomic regions. Thirdly, branch-specific selection patterns were decoded using site-model comparisons ( $dN/dS > 1$  indicating positive

selection). The integrated analysis revealed three evolutionary hotspots (*nad4L*, *cob*, *rps12*) showing strong purifying selection (Ka/Ks = 0.12–0.35), while *cox1* exhibited elevated  $\pi$  values ( $0.027 \pm 0.004$ ) suggesting adaptive diversification. Notably, *C. sinensis* cv. Rougui displayed unique selection signatures in *atp6* (Ka/Ks = 1.18,  $p < 0.05$ ), potentially linked to cultivar-specific mitochondrial-nuclear coevolution. These computational insights delineate how selection pressures sculpt mitochondrial genome architecture during *Camellia* speciation and domestication processes [43].

### **Fragments shared between mitotic and chloroplast genomes and phylogenetic tree analysis**

The assembled *C. luteoflora* chloroplast genome and mitochondrial genome sequences were used to identify homologous fragments between the mitochondrial and chloroplast genomes using BLAST v2.9.0 [44] available on NCBI, with the screening criteria of  $\geq 70\%$  match,  $E\text{-value} \leq 1e-5$ , and  $\geq 30$  bp in length, and the screened sequence fragments were visualized using Circos v0.65 [44]. The mitochondrial genomes downloaded from NCBI were from 27 species with *Ginkgo biloba* Linn as the outgroup. Gene sequences of 33 conserved genes from each species were identified and extracted using PhyloSuite v1.2.1 [45]. Conserved gene sequences were aligned using MAFFT v7.450 [46], and aligned sequences were concatenated for phylogenetic tree construction. The aligned sequences were concatenated and trimmed using trimAl (v1.4). Subsequently, model prediction was conducted with jmodeltest (v2.1.10), identifying the GTR model. The maximum likelihood phylogenetic tree was then constructed using RAxML (v8.2.10) with the GTRGAMMA model and a bootstrap value set to 1000 [47]. The chloroplast genome was selected for 28 species for whole genome sequence construction of the phylogenetic tree in the same way as for the mitochondrial genome. Finally, we used the iTOL v6 (<https://itol.embl.de/>) online website to visualize the phylogenetic tree [48].

## **Results**

### **Assembly and annotation of the *Camellia luteoflora* mitochondrial genome**

*C. luteoflora* mitochondrial genome was assembled using both long and short sequence assembly strategies. For both long and short reads, we obtained about 6.58 Gb of nanopore high-quality reads with a readsN50 of 2000 bp and a total mass of 5.17 Gb. The short reads were assembled ab initio into a unitary map, and the contiguous-repetitive-contiguous regions were disassembled by the long reads. Finally, we obtained a single cyclic molecule with an average long-read coverage depth of 158.4×. The initial assembly map presented a complex conformation due to the presence of a large number of repetitive





**Fig. 1** Mitochondrial genome assembly results of *Camellia luteoflora*. **(A):** Mitochondrial genome map; **(B):** Field photo of *Camellia luteoflora*; The genes located in the upper region of linear molecules and within the interior of circular molecules represent genes transcribed in a clockwise direction, while the genes in the lower region of linear molecules and the genes on the outside of circular molecules represent genes transcribed in a counterclockwise direction. Genes with different functions were depicted using different colors.)

**Table 1** Results of mitochondrial genome annotation in *Camellia luteoflora*

Group of genes	Gene name
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFc*, ccmFn</i>
Ubiquinol cytochrome c reductase	<i>cob</i>
Cytochrome oxidase	<i>cox1, cox2, cox3,</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1****, nad2****, nad3, nad4**, nad4L, nad5****, nad6, nad7****, nad9</i>
Ribosomal proteins (LSU)	<i>rpl10, rpl16, rpl2*, rpl5</i>
Ribosomal proteins (SSU)	<i>#rps19, rps1*, rps12, rps13, rps14, rps19, rps3*, rps4, rps7</i>
Succinate dehydrogenase	<i>#sdh3, sdh3, sdh4</i>
Ribosomal RNAs	<i>rrn18, rrn26, rrn5</i>
Transfer RNAs	<i>trnA-TGC*, trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-GCC, trnH-GTG, trnI-GAT*, trnK-TTT, trnM-CAT(6), trnN-ATT, trnN-GTT, trnP-TGG(2), trnQ-TTG, trnS-GCT, trnS-TGA, trnS-TGA*, trnT-TGT*, trnV-GAC, trnW-CCA, trnY-GTA</i>

Note: The \* symbol indicate the number of introns; The # symbol represent pseudo genes; Numbers after gene names are the Number of copies of multi-copy genes

sequences. However, the utilization of ONT long read segments enabled us to represent the mitochondrial genome as a single circular molecule of 78,3024 bp. The GC content of the mitochondrial genome was 44.63%, which was significantly higher than the GC content of the chloroplast genome of the same species (39.03%) (Fig. 1).

The mitochondrial genome includes a total of 40 protein-coding genes, 27 tRNA genes and three rRNA genes (Table 1). The coding genes included five ATP synthase genes (*atp1, atp4, atp6, atp8, and atp9*), nine NADH dehydrogenase genes (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, and nad9*), four cytochrome c biogenesis genes (*ccmB, ccmC, ccmFC and ccmFN*), three cytochrome c oxidase genes (*cox1, cox2, and cox3*), one maturases gene (*matR*), one membrane transporter protein gene (*mttB*), and one ubiquitin cytochrome c reductase gene (*cob*). The variant genes include four ribosomal protein large subunits (*rpl10, rpl16, rpl2, and rpl5*), nine ribosomal protein small subunits (*rps1, rps3, rps4, rps7, rps12, rps13, rps14, and two rps19* (one of which is a pseudogene)), and three succinate dehydrogenases (two *sdh3* (one of which is a pseudogene) and *sdh4*). 3 rRNAs (*rrn18, rrn26, and rrn5*) and 27 tRNAs. Of these, the *nad1, nad2, nad5, and nad7* genes contain 4 introns; the *nad4* gene has 2 introns, and *rpl2, rps3, rps1, and ccmFC* have 1 intron. *trnM-CAT* and *trnP-TGG* genes are multicopy genes. *trnA-TGC, trnI-GAT, trnS-TGA, and trnT-TGT* genes contain 1 intron.

In addition, we obtained the complete chloroplast genome, which is 157,172 bp in length (Figure S1, Table S1). It contains a large single-copy (LSC) region of 86,725 bp, a small-size copy (SSC) region of 18,293 bp, and two inverted repeat regions (IRs) of 26,077 bp each. 131 genes were identified, including 87 PCGs, 8 rRNAs, and 36 tRNA genes.

### Different configurations of mitochondrial genomes of *Camellia luteoflora*

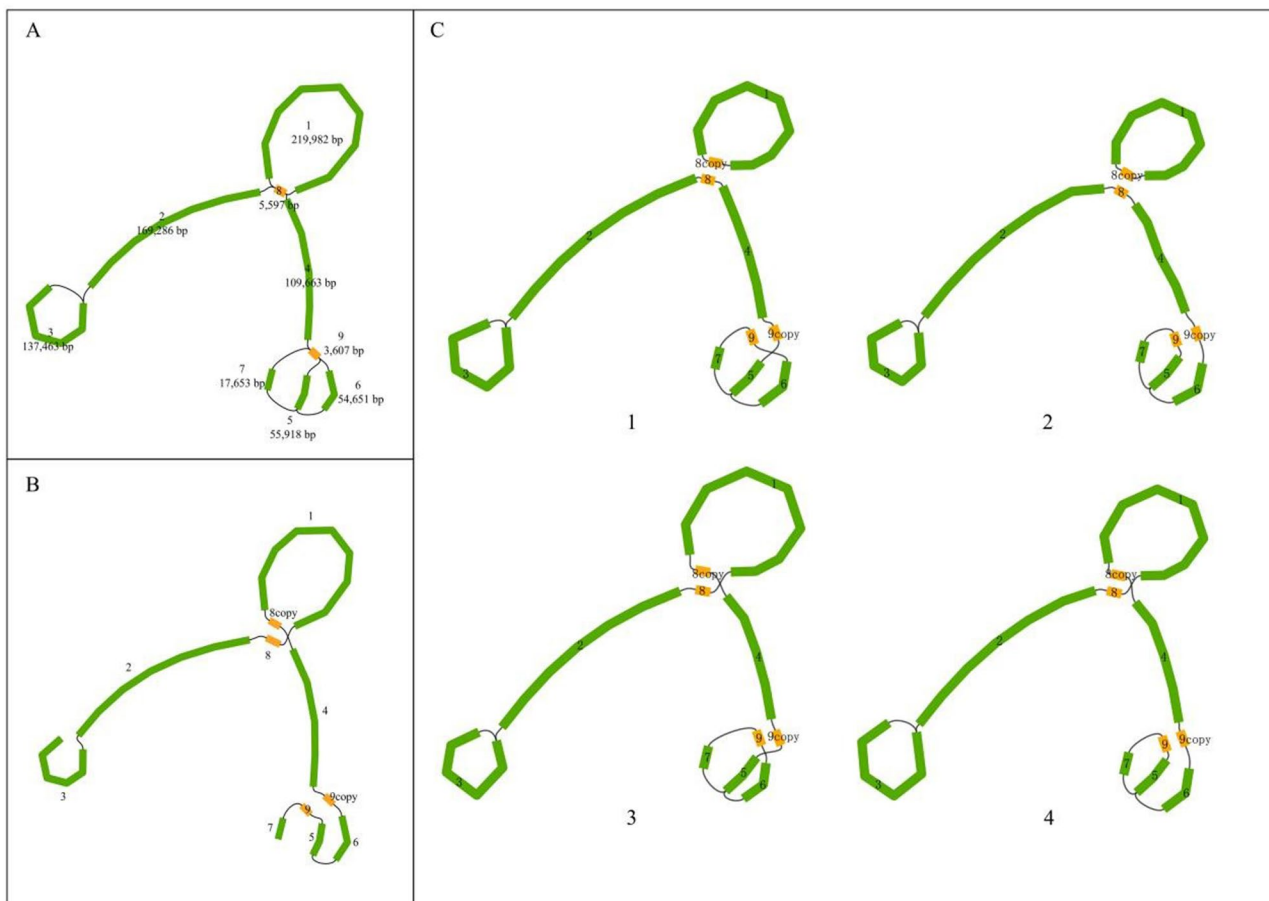
Repeat sequences are frequently involved in homologous recombination, and in Fig. 2. Two of the 8 and 9 repeat sequences were identified as mediating recombination of consecutive sequences on the boundary. For repeat sequence 8, the sequence configuration in the assembled mitochondrial genome sequence was  $1 \rightarrow 8 \rightarrow 1$ ,  $2 \rightarrow 8 \rightarrow 4$  (Fig. 2C-1, Figs. 2C and 3). After recombination, the sequence configuration changed to  $1 \rightarrow 8 \rightarrow 4$ ,  $2 \rightarrow 8 \rightarrow 1$ . Similarly, the sequence configuration of repeat sequence 8 was  $7 \rightarrow 9 \rightarrow 5$ ,  $4 \rightarrow 9 \rightarrow 6$  in the assembled mitochondrial genome sequence (Figs. 2C and C, 3 and 4). After recombination, the sequence configuration changed to  $7 \rightarrow 9 \rightarrow 6$ ,  $4 \rightarrow 9 \rightarrow 5$ . The results illustrate the complexity of the mitochondrial genome structure of *C. luteoflora* (Fig. 2C).

In order to verify the existence of different configurations of the repeated sequences 8 and 9, we performed the sequences to verify the sequences  $1 \rightarrow 8 \rightarrow 2 \rightarrow 4$  and  $4 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 9$ , respectively. The results show that different combinations are supported across the repeated

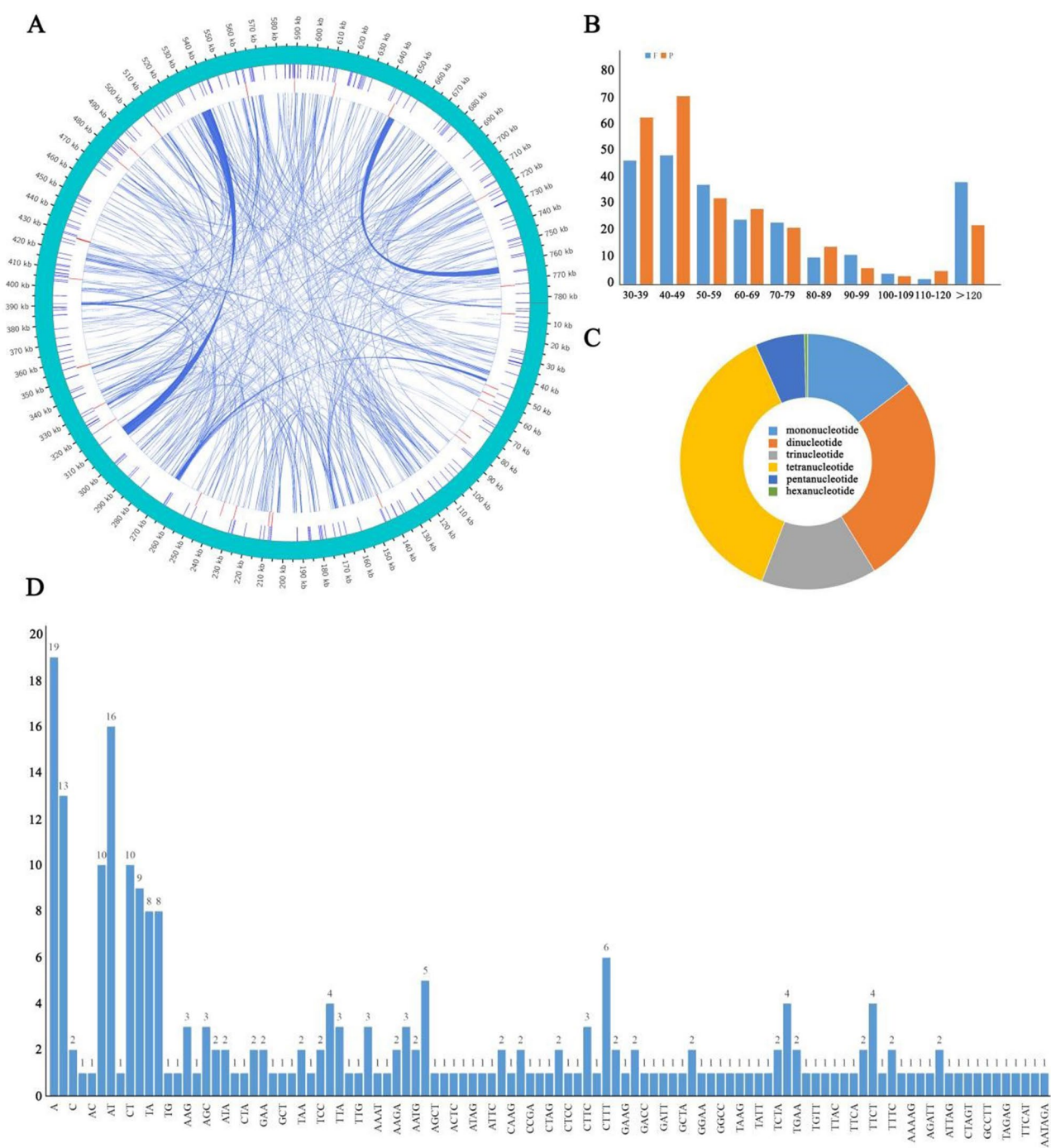
sequences, verifying the existence of different configurations of the repeated sequences. By mapping Nanopore longreads to sequences with different conformations associated with the repeat sequences, the study confirmed that intramolecular repeat sequences can mediate recombination (Figure S2-3). This phenomenon suggests that recombination mediated by specific homologous fragment pairs in the mitochondrial genome can influence the recombination patterns of different homologous fragment pairs, thereby enriching DNA substructures with different conformations.

### Repeat sequence analysis

In the *C. luteoflora* mitochondrial genome, two dispersed repeat sequence types, forward and palindromic, were identified, and a total of 506 dispersed repeats with lengths greater than or equal to 30 bp were obtained. Meanwhile, no forward and complementary repeat types were detected. These dispersed repeats were widely distributed throughout the intergenic region of the genome and were visualized by the Circos software package



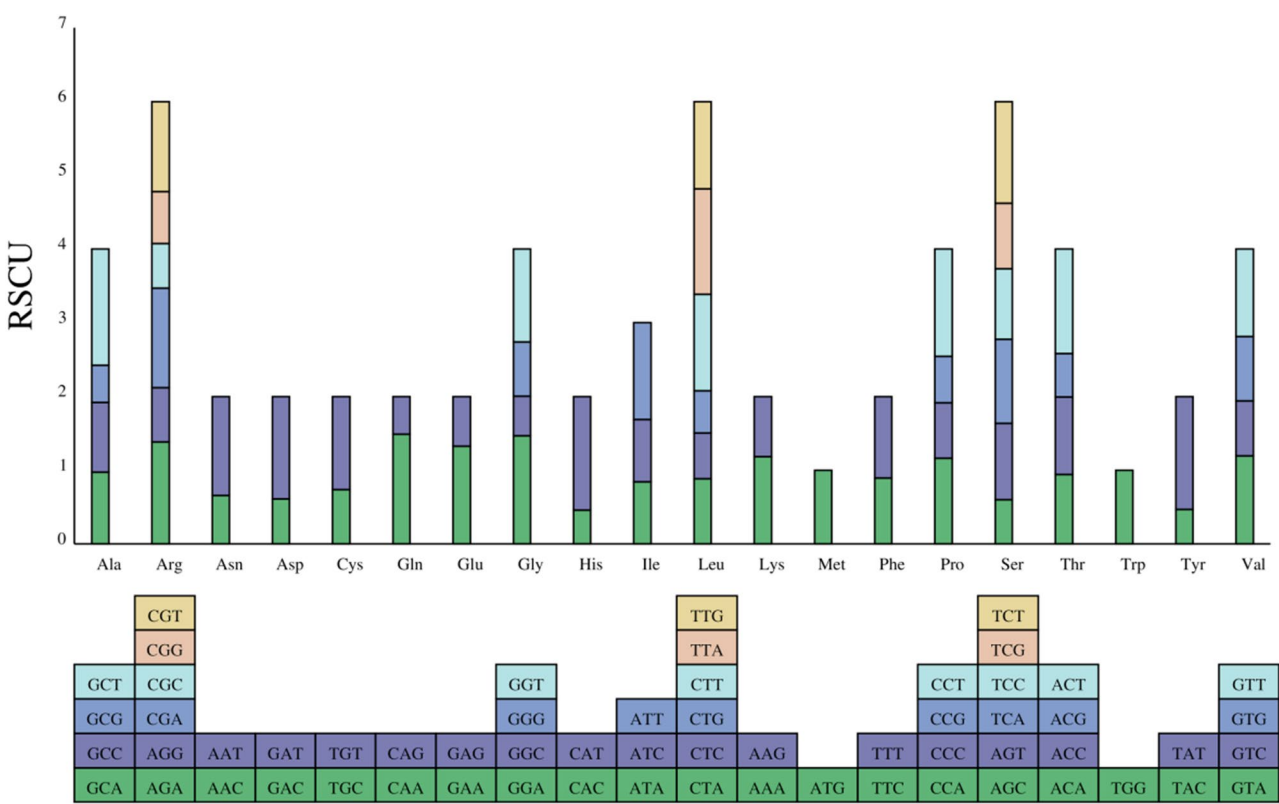
**Fig. 2** *Camellia luteoflora* recombination mediated by repetitive sequences. (The yellow rectangles indicate repetitive sequences at branch points. **A:** the initial structure of the genome is branching; **B:** the complete linear sequence of the genome branching unraveling; **C:** the four substructures arising from two pairs of repetitive sequences in the genome.)



**Fig. 3** Repetitive sequences of *Camellia luteoflora* mitochondrial genome. **(A):** Distribution of repetitive sequences; **(B):** length of forward repetitive sequence and palindromic sequence; **(C):** proportion of six SSR types; **(D):** number of SSR repetitive sequence types)

(Fig. 3A, Table S2). 506 repetitive sequences with a total length of 24,149 bp accounted for 3.08% of the total mitochondrial genome length. Among these repetitive sequences, there were 243 F repeats and 263 P repeats, and repeats of 40 to 49 bp were the most common (Figure. 3B, Table S3). A total of 240 SSRs were identified in the mitochondrial genome, and tetranucleotide repeats dominated with 37.50% (90) of the total number, followed

by dinucleotide repeats (64), mononucleotide repeats and trinucleotide repeats (both 35), pentanucleotide repeats (15), and hexanucleotide repeats (1). Among the single nucleotide SSRs, the highest percentage of A repeats (54.29%) and among the dinucleotide repeats, the highest percentage of AT base repeats (25.00%) were found (Figure. 3 C, 3D; Table S4,5).



**Fig. 4** Use of relative synonymous codons in the mitochondrial genomes of *Camellia luteoflora*

***Camellia luteoflora* mitochondrial genome codon preference analysis**

In the *C. luteoflora* mitochondrial genome, we found that the utilization rates of the termination codons UAA, UGA, and UAG were 50.00%, 31.58%, and 18.42%, respectively, while the termination codon TAG was not used. Relative synonymous codon usage (RSCU) can eliminate the effect of amino acid composition on codon usage and directly reflect the differences in codon usage patterns. The value of RSCU is equal to 1, which indicates an unbiased selection of codon usage. The value of RSCU is greater than 1, which means the frequency of usage is higher; the value of RSCU is less than 1, which means the frequency of usage is lower. The analysis of the RSCU method showed that there were two codons with RSCU equal to 1, which were AUG of Met and UGG of Trp; GCU codon encoding alanine (Ala) had the highest frequency, with an average RSCU value of 1.5766; and there were 29 codons with RSCU greater than 1, and most of them ended with A or U (Fig. 4, Table S6).

**Mitochondrial plastid DNA (MTPTs) in the mitochondrial genome of *Camellia luteoflora***

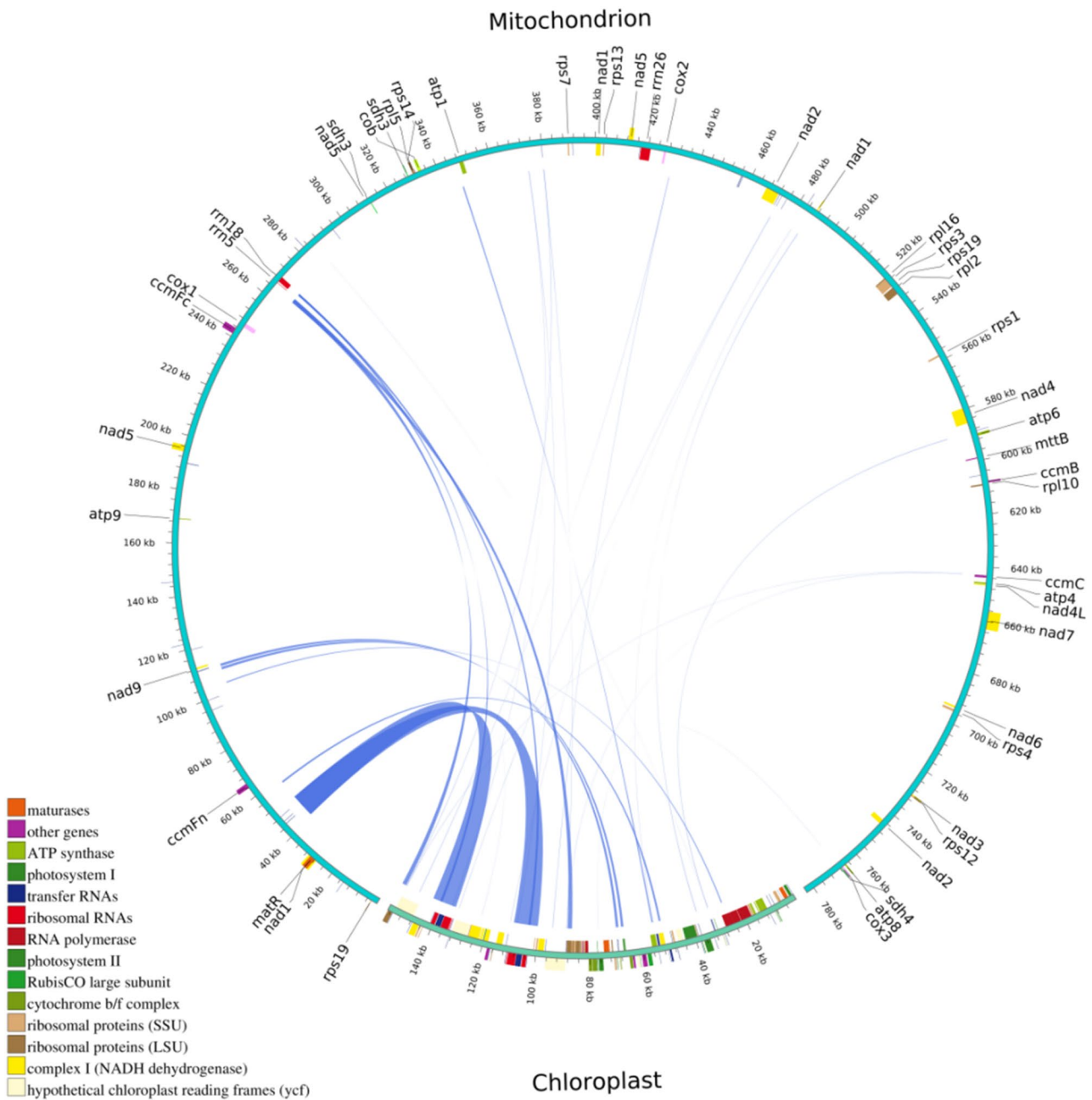
Intracellular transfer of genetic material has been a common phenomenon in mitochondrial genomes during the evolution of higher plants, and these sequence fragments originating from chloroplast organelles are

relatively poorly conserved. The mitochondrial genome of *C. luteoflora* is approximately 4.9 times larger than the chloroplast genome (157,172 bp). The distribution of mitochondrial genes was relatively sparse compared to chloroplasts (Fig. 5). Based on the sequence similarity between the chloroplast and mitochondrial genomes, 19 chloroplast gene fragments were identified to be transferred to the mitochondrial genome in this study. The total length of these transferred fragments was 29,534 bp, representing 3.77% of the entire mitochondrial genome. Referring to these fragments as MTPTs, they represent sequence migration from chloroplasts to mitochondrial organelles (Fig. 5). Among these 19 homologous fragments, the mitochondrial genome consists of nine CDS regions, two rRNA genes, and eight tRNAs. The longest sequence of MTPT1 was 9572 bp, and the shortest sequence of MTPT19 was 32 bp (Table 2). It is noteworthy that the two rRNA genes and eight tRNA genes present in the chloroplast genome of *C. luteoflora* may have been lost or undergone pseudogene changes in the chloroplast genome.

**RNA editing events in the mitochondrial genome of *Camellia luteoflora***

In this study, we revealed the characteristics of post-transcriptional modifications by predicting the RNA editing sites of all PCGs in the mitochondrial genome of *C.*





**Fig. 5** Fragments of *Camellia luteoflora* converted from chloroplasts to mitochondria

*luteoflora*. The results showed that a total of 539 RNA editing events were identified, which were mainly characterized by base C to T (corresponding to base C to U in RNA) transitions (Table S7). The *ccmFn* gene was the most significant, with 40 editing sites identified, followed by the *ccmB* gene (34 editing sites). In addition, *rps14* and *sdh3* each had 2 RNA editing sites (Fig. 6A). The study further observed that the editing events were concentrated at the first and second base positions of the start codon. These RNA editing events resulted in amino acid changes such as histidine (His) to tyrosine (Tyr), arginine

(Arg) to cysteine (Cys), threonine (His) to isoleucine (Iso), leucine (Leu) to phenylalanine (Phe), serine (Ser) to phenylalanine (Phe), arginine (Arg) to tryptophan (Try), etc. A total of five amino acid type shifts were also found, with the most being hydrophilic to hydrophobic conversions and the least being hydrophilic to stop conversions (Fig. 6B). The results showed that the ratio of hydrophilic and hydrophobic types before and after the amino acid changes was quite different, with the ratio changing from 1.63 to 0.27 before and after editing (Fig. 6C, D). It can be seen that many of the amino acid changes triggered

**Table 2** Transfer of *Camellia luteoflora* Chloroplast genome to mitochondrial genome gene fragments

No	Length	Identity	Mt-start	Mt-end	Cp-start	Cp-end	Type
1	9572	1.00	50,542	40,971	100,427	109,998	<i>trnV-GAC</i> tRNA
	9572	1.00	40,971	50,542	133,900	143,471	
2	1689	1.00	265,783	264,095	86,786	88,474	<i>rpl23</i> CDS
	1689	1.00	264,095	265,783	155,424	157,112	
3	1023	0.87	110,407	111,381	68,348	69,362	<i>trnW-CCA</i> tRNA
4	992	0.86	109,008	109,995	66,337	67,278	<i>psbJ</i> CDS
5	758	0.82	58,427	57,693	51,402	52,144	<i>ndhJ</i> CDS
6	505	0.87	346,892	347,383	54,578	55,064	<i>atpE</i> CDS
7	416	0.82	103,427	103,834	25,936	26,325	<i>rpoB</i> CDS
8	890	0.74	268,532	267,674	103,377	104,240	<i>rrn16</i> rRNA
	890	0.74	267,674	268,532	139,658	140,521	
9	147	0.96	485,957	486,103	36,687	36,833	<i>psbC</i> CDS
10	149	0.91	589,085	588,937	31,921	32,069	<i>trnD-GUC</i> tRNA
11	127	0.94	379,950	379,824	155,033	155,159	<i>trnI-CAU</i> tRNA
	127	0.94	379,824	379,950	88,739	88,865	
12	126	0.89	429,969	429,850	94,366	94,491	<i>ycf2</i> CDS
	126	0.89	429,850	429,969	149,407	149,532	
13	116	0.91	373,983	373,875	108,588	108,703	<i>rrn23</i> rRNA
	116	0.91	373,875	373,983	135,195	135,310	
14	84	0.96	473,229	473,146	132,470	132,552	<i>trnN-GUU</i> tRNA
	84	0.96	473,146	473,229	111,346	111,428	
15	77	0.94	482,871	482,947	54,264	54,340	<i>trnM-CAU</i> tRNA
16	76	0.93	643,141	643,066	155,033	155,108	<i>trnI-CAU</i> tRNA
	76	0.93	643,066	643,141	88,790	88,865	
17	41	1.00	49,911	49,871	123,002	123,042	<i>ndhA</i> CDS
18	34	0.97	768,270	768,237	76,283	76,316	<i>psbB</i> CDS
19	32	0.97	285,986	286,017	32,027	32,058	<i>trnD-GUC</i> tRNA

Note: Mt: mitochondrial genome; Cp: chloroplast genome

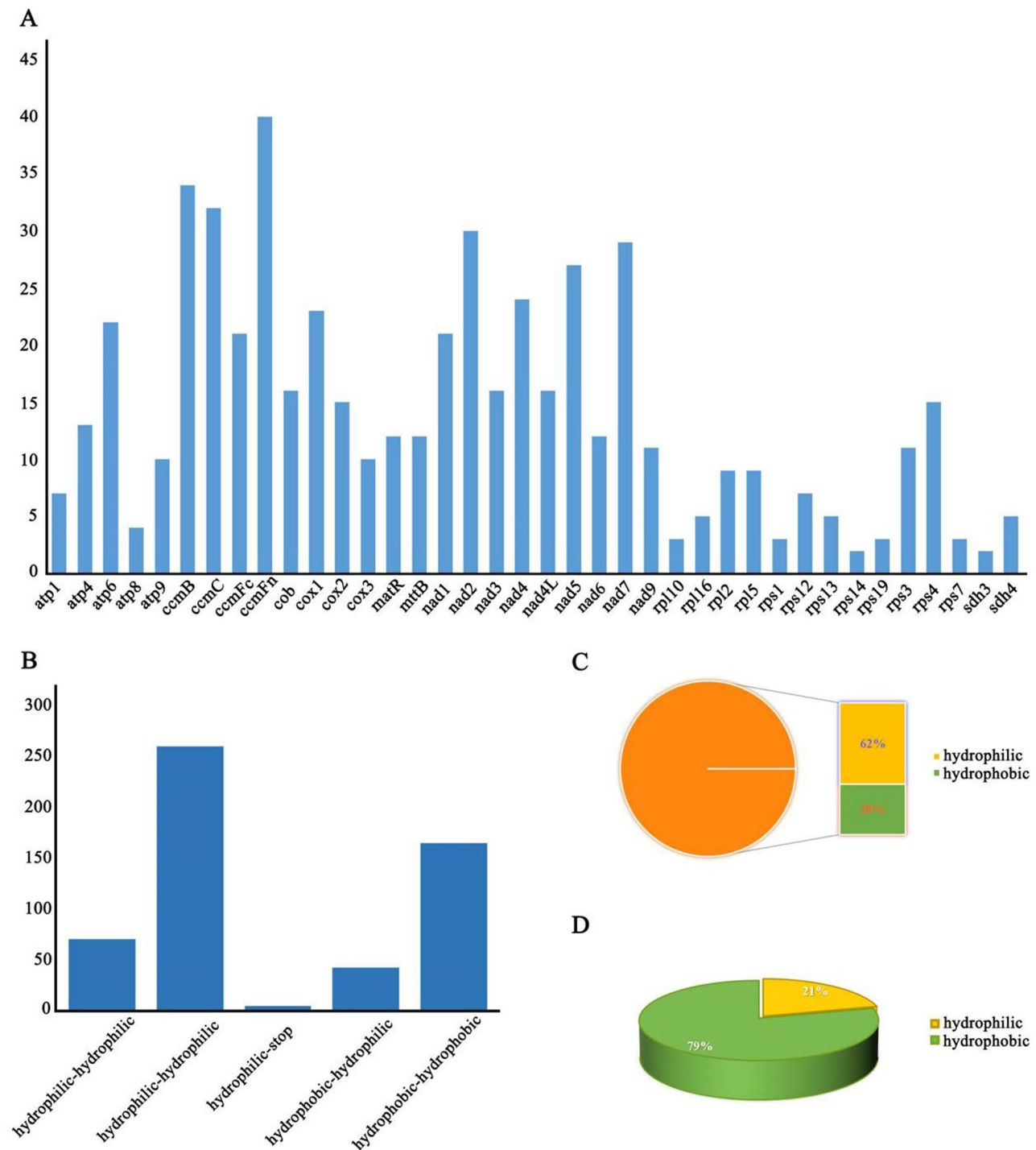
by RNA editing introduce more hydrophobic amino acids into the protein structure, thereby altering the hydrophilic nature of the protein, which may play a key role in maintaining the regulation of mitochondrial gene expression.

#### Selection pressure on the mitochondrial genome of *Camellia luteoflora* plants and its nucleotide diversity analysis

Ka/Ks (also known as dN/dS) denotes the ratio of the rate of nonsynonymous substitutions (Ka) to the rate of synonymous substitutions (Ks) and is used to measure the selective pressure on proteins during the evolution of different species. When Ka/Ks > 1, the gene is under positive selection. When Ka/Ks = 1, genes undergo neutral evolution. When Ka/Ks < 1, genes are subject to negative or purifying selection. In order to assess the selection pressure on seven *Camellia* L. was analyzed. The results showed (Fig. 7, Table S8) that the average Ka/Ks value of the 12 genes was 0.45. Only the *cox2* gene had a Ka/Ks value greater than 1, indicating that it was subject to strong positive selection. The remaining 11 genes had Ka/Ks values less than 1. This phenomenon reveals that

these genes have been subject to strong negative selection during evolution and have relatively stable protein functions. These 12 protein-coding genes of *C. luteoflora* mitochondria showed consistent conserved properties at the molecular evolutionary level with their counterparts in the six selected plants and were less susceptible to interspecies genetic variation. This analysis is of great scientific value for understanding the molecular evolutionary pathways of *C. luteoflora* and other species, the conservatism and innovativeness of gene functions, and the genetic basis of adaptive differences among species.

Nucleotide diversity (Pi) can be used to assess genetic differences in nucleotide sequences between species and populations and to select regions of high variability as potential molecular markers for populations. Pi analysis of organelle genes was performed on seven *Camellia* L. The results showed (Fig. 8, Table S9) that the mitochondrial gene with the highest variability was *cox2* (Pi = 0.02248), followed by *rrn18* (Pi = 0.00711) and *rrn26* (Pi = 0.00691). In mitochondrial PCG, the Pi values of all genes ranged from 0 to 0.00711, indicating that the nucleotide sequences of *C. luteoflora* mitochondrial genes are highly conserved.

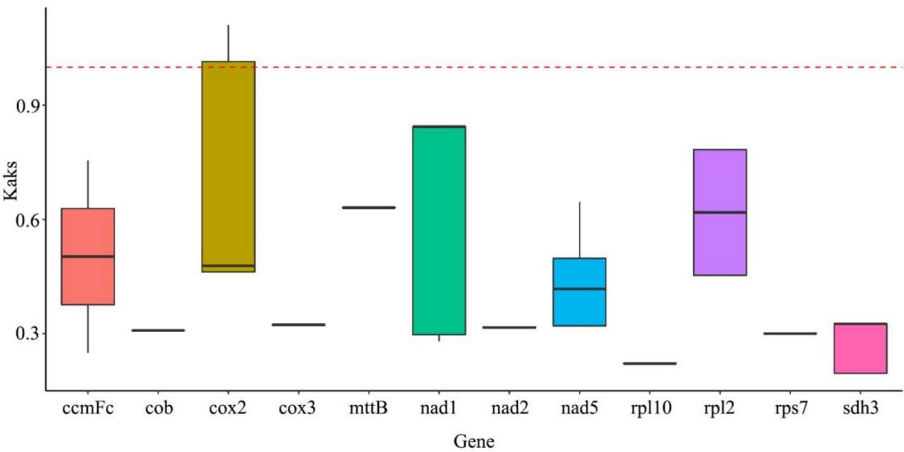


**Fig. 6** Prediction of RNA editing sites for PCGs in the mitochondrial genome of *Camellia luteoflora*. **(A)**: number of RNA editing sites for each gene; **(B)**: type of amino acid conversion **(C)**: percentage of hydrophilic and hydrophobic amino acids before RNA editing; **(D)**: percentage of hydrophilic and hydrophobic amino acids after RNA editing.)

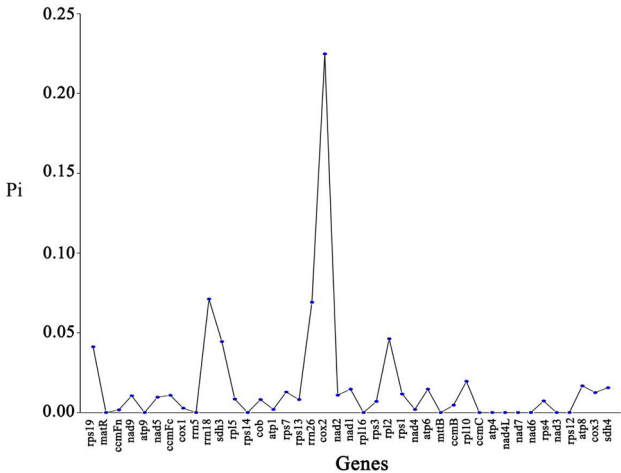
**Construction of phylogenetic tree of *Camellia luteoflora* mitochondrial genome**

To determine the phylogenetic position of *C. luteoflora*, we constructed a phylogenetic tree for green with protein-coding genes from the mitochondrial genome and

the whole chloroplast genome. The phylogenetic tree was constructed based on 33 shared coding genes of the mitochondrial genomes of 28 species using *Ginkgo biloba* as an outgroup. The results showed (Fig. 9) that the phylogenetic tree strongly supported the delineation of dicots



**Fig. 7** Box plots of Ka/Ks ratios of *Camellia luteoflora* with six plants



**Fig. 8** Nucleotide diversity values of *Camellia luteoflora* with six species

from monocots and the separation of angiosperms from gymnosperms, and that the phylogenetic relationships of these species were consistent with traditional classification. Meanwhile, the closest relative to *C. luteoflora* in the genus *Camellia* was *Camellia sinensis* var. *sinensis*, followed by *Camellia sinensis*, and the most distant species was *Stewartia sinensis* var. *sinensis*. using *Polyspora axillaris* and *Polyspora hainanensis* as outgroups, the phylogenetic tree constructed based on the whole chloroplast genes of 28 species (Fig. 10) showed that sect. *luteoflora* was in a separate position in *Camellia* L. and was more closely related to the sect. *Camellia*, while sect. *Thea* was more distantly related.

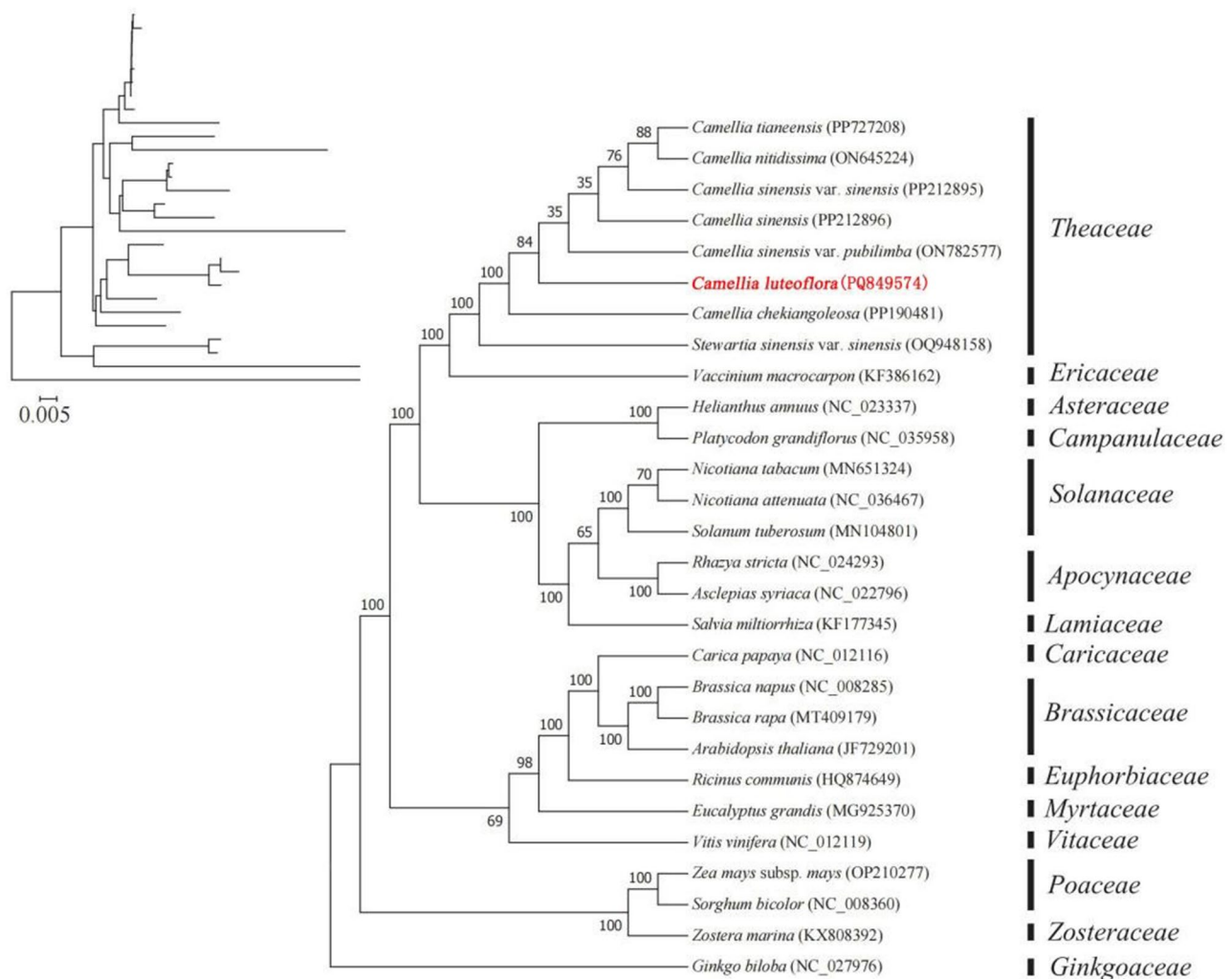
**Discussion**

With the continuous advancement of sequencing technologies, research on chloroplast genome sequences of higher plants has accumulated substantial achievements. In contrast, studies on plant mitochondrial genomes remain relatively limited, though this field

has now emerged as a burgeoning research focus [49]. The primary challenges in plant mitochondrial genome studies stem from their multifaceted complexity: intricate genomic composition, structural diversity, abundant recombinogenic repetitive sequences, and dynamic conformational changes. The interaction mechanisms between mitochondrial and nuclear genomes, coupled with their central roles in energy metabolism and biosynthesis, render functional and regulatory investigations particularly challenging. Therefore, deciphering mitochondrial structure and function is crucial for unraveling key physiological phenomena and evolutionary patterns in biology, yet requires overcoming multiple technical obstacles. In recent years, the increasing number of assembled plant mitochondrial genomes [50, 51] has enabled comprehensive comparative genomic studies. Significant interspecies variations have been observed in mitochondrial genome characteristics including total base length, gene composition, gene arrangement order, GC content, and overall architecture [52, 53]. Structurally, mitochondrial genomes typically exhibit complex circular or linear configurations. While most plants possess single circular genomes, some species such as *Salvia japonica* and *Paphiopedilum micranthum* demonstrate multipartite genome structures [54, 55].

In this study, we employed high-throughput sequencing to assemble the mitochondrial genome of *C. luteoflora*, which spans 587,847 bp. Through comprehensive annotation, we identified 42 protein-coding genes (PCGs), 3 rRNA-encoding genes, and 27 tRNA-encoding genes. The GC content of plant mitochondrial genomes has been recognized as an indicator of species adaptation [56], with documented values ranging from 23.9 to 50.5% in terrestrial plants [16]. Notably, the mitochondrial genome of *C. luteoflora* exhibited a GC content of 44.63%, significantly exceeding that of its chloroplast genome (39.03%). The mitochondrial genome architecture





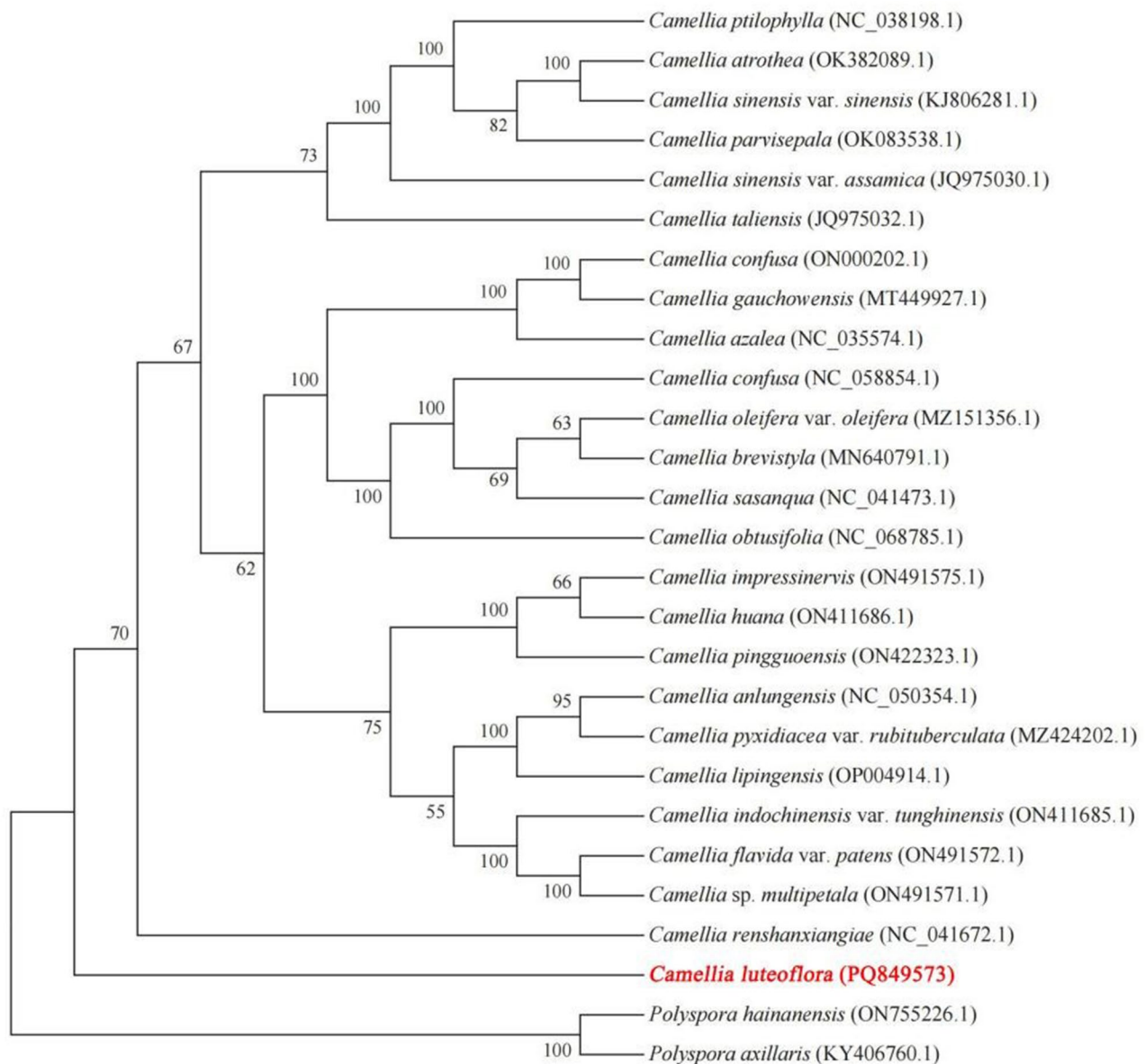
**Fig. 9** Phylogenetic relationships of *Camellia luteoflora* mitochondrial genomes. ( phylogenetic tree constructed on the basis of 28 protein coding genes)

of *C. luteoflora* features multiple interconnected gene segments organized into a continuous, high-resolution network. These structurally integrated components are functionally critical for mitochondrial operations, potentially encoding essential proteins for cellular energy production. Further analysis revealed intricate organizational details including gene quantity, spatial arrangement patterns, and putative regulatory elements. These structural insights establish a foundation for elucidating mitochondrial gene functionality and associated expression regulation mechanisms.

Repetitive sequences serve as crucial elements for investigating genome architecture, gene expression regulation, phenotypic characteristics, molecular marker development, functional annotation, and evolutionary processes [57–59]. Compared to chloroplast and nuclear genomes, plant mitochondrial genomes demonstrate significantly slower evolutionary rates [60], making their molecular markers particularly valuable for enhancing

species identification precision. In this study, we identified 506 repetitive sequences in the *C. luteoflora* mitochondrial genome, comprising 243 forward (F) and 263 palindromic (P) repeats. These repetitive elements suggest active molecular recombination events that likely drive dynamic structural rearrangements and conformational modifications during mitochondrial genome evolution. Additionally, 240 simple sequence repeats (SSRs) were detected across distinct genomic regions. The SSR profile exhibited tetranucleotide repeats as the predominant type (37.50%), followed by dinucleotide repeats (26.67%), a distribution pattern consistent with observations in most angiosperm mitochondrial genomes [41].

Codons serve as fundamental units in biological genetic expression, mediating protein synthesis, modulating gene expression, and facilitating genetic variation [61]. Organisms develop codon usage bias through selective pressures from environmental constraints and species-specific evolutionary trajectories [62], with these



**Fig. 10** Phylogenetic relationships of *Camellia luteoflora* chloroplast genomes. (phylogenetic tree constructed on the chloroplast genome)

preferential synonymous codon selections critically shaping genomic characteristics. Our analysis reveals that the *C. luteoflora* mitochondrial genome exhibits pronounced A/T nucleotide enrichment, demonstrating strong preferential usage of codons terminating with A/U bases. Comparative studies with other plant mitochondrial genomes [41, 63–65] indicate that while A/U-ending codon preference represents a conserved feature across green plant lineages, the degree of bias displays both interclade variations and intragroup fluctuations. The observed A/U predominance in *C. luteoflora* provides supporting evidence for ancestral codon usage patterns in early-diverging land plants. Phylogenetic conservation of mitochondrial codon preferences likely originates

from the monophyletic nature of plant mitochondria, whereas bias intensity variations may reflect terrestrial adaptation processes. Notably, the evolutionary transition from strong AT-bias toward gradual GC accumulation in derived lineages [65] potentially represents an adaptive strategy to mitigate UV-induced DNA damage in xeric environments.

Exogenous gene insertions in plant mitochondrial genomes predominantly localize to intergenic regions. The integration length of chloroplast-derived DNA sequences exhibits interspecific variation, typically constituting 1–12% of chloroplast genome content in angiosperms [40, 66, 67]. This transfer mechanism represents a primary driver of gene content divergence among plant

mitochondrial genomes, underscoring the necessity of tracking gene migration patterns for evolutionary studies [68]. Chloroplast-to-mitochondria tRNA gene transfers constitute a widespread phenomenon in plants [69]. Our investigation identified eight chloroplast-origin tRNA genes in the *C. luteoflora* mitochondrial genome, potentially fulfilling functional compensation roles. Additionally, we detected numerous chloroplast-derived gene fragments containing chloroplast-specific functional elements, though their mitochondrial operational significance remains unconfirmed.

Post-transcriptional RNA editing modifies genetic information, resulting in mitochondrial protein products often deriving from partially edited transcripts. Hydrophilic amino acids critically influence protein folding processes, with reduced proportions correlating with enhanced structural stability [70]. In *C. luteoflora*, mitochondrial RNA editing predominantly converts hydrophilic to hydrophobic residues, consequently increasing overall protein hydrophobicity. Ka/Ks calculations provide critical insights for phylogenetic reconstruction and protein evolution analysis [71]. Our results demonstrate prevalent negative selection pressure on mitochondrial coding genes, aligning with established patterns [41]. Notably, *cox2* exhibited  $Ka/Ks > 1$ , suggesting its pivotal evolutionary role in *Camellia* mitochondrial genome development.

Plant mitochondrial genomes exhibit frequent structural rearrangements and gene content flux, driven by mitochondrial-nuclear DNA transfers and suppressed mutation rates [72]. These evolutionary signatures offer unique phylogenetic markers [73], enabling species relationship reconstruction through mitochondrial gene homology analysis [74]. In this study, the phylogenetic relationships of *Camellia luteoflora* were constructed based on mitochondrial and chloroplast genomic information. The mitochondrial genome was able to strongly support the delineation of dicots from monocots and the separation of angiosperms from gymnosperms, and the species *Camellia sinensis* var. *sinensis*, which is closely related to *C. luteoflora*, was found. Meanwhile, phylogenetic analyses constructed from the chloroplast genome further confirmed the separate status of sect. *luteoflora* from *Camellia* L. sect. *Camellia* that are more closely related to it were found, but these still need to be further proved by biogeographical evidence.

## Conclusion

*C. luteoflora* is a plant with important economic and medicinal value, as well as a more specialized taxon of *Camellia* L. This study pioneers the assembly of *C. luteoflora* mitochondrial genome through integrated sequencing approaches and comprehensive bioinformatic analyses. These advancements enable systematic

comparisons of organellar genome architectures while expanding investigative perspectives on mitochondrial-plastid DNA transfer mechanisms. Phylogenetic reconstruction based on mitochondrial and chloroplast genomic datasets corroborates the species' evolutionary position within its taxonomic clade. The findings establish foundational references for understanding *C. luteoflora*'s genetic characteristics, molecular variation patterns, evolutionary origins, and systematic classification, while concurrently informing cultivation practices and resource utilization strategies. To refine phylogenetic resolution within *Camellia* L., future investigations should prioritize mitochondrial genome sequencing across broader taxonomic representatives.

## Abbreviations

PCGs	Protein-Coding Genes
mtDNA	Mitochondrial Genome
cpDNA	Chloroplast Genome
Ka/Ks	Non-Synonymous/Synonymous Mutation Ratio
RSCU	Relative Synonymous Codon Usage
MTPT	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/s12870-025-06461-6>.

Supplementary Material 1: Fig. S1: Chloroplast genome map of *Camellia luteoflora*;

Supplementary Material 2: Figure S2. Verification of repeat 8 mediated recombination coverage in *Camellia luteoflora*;

Supplementary Material 3: Figure S3. Verification of repeat 9 mediated recombination coverage in *Camellia luteoflora*;

Supplementary Material 4: Table S1. Characteristics of the chloroplast genome of *Camellia luteoflora*;

Supplementary Material 5: Table S2. Distribution of repetitive sequences;

Supplementary Material 6: Table S3. Length analysis of dispersed repetitive sequences;

Supplementary Material 7: Table S4. Proportion of six SSR types;

Supplementary Material 8: Table S5. Number of SSR repetitive sequence types;

Supplementary Material 9: Table S6. Use of relative synonymous codons in the mitochondrial genomes of *Camellia luteoflora*;

Supplementary Material 10: Table S7. Prediction of RNA editing sites for PCGs in the mitochondrial genome of *Camellia luteoflora*;

Supplementary Material 11: Table S8. Ka/Ks ratios of *Camellia luteoflora* with six plants;

Supplementary Material 12: Table S9. Nucleotide diversity values of *Camellia luteoflora* with six species;

Supplementary Material 13: Table S10. Phylogenetic tree constructed on the basis of 28 Mitochondrial genome;

Supplementary Material 14: Table S11. Chloroplast genome sequence numbers of the 28 species used to construct the phylogenetic tree

## Acknowledgements

Not applicable.

## Author contributions

Z.L. conceived the study. X.X. were responsible for analyzing and writing this manuscript; X.X. and Z.R. were responsible for collecting and identifying the material; W.G. and C.Y. helped with the analysis of the data; Z.L. revised the manuscript. All authors reviewed and approved the final manuscript.

## 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

No datasets were generated or analysed during the current study.

## Declarations

### Ethics approval and consent to participate

The plant materials collected in this study are in accordance with international and national legal standards. The collected plant material does not pose a threat to other species and the collection of the species was recognized by the relevant authorities. The collected material was not subjected to medical experiments and only chloroplast and mitochondrial genes were extracted from the plant material.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 18 January 2025 / Accepted: 25 March 2025

Published online: 05 April 2025

## References

- Zhang HD, Zeng FA. *Camellia* L. new group sect. *Luteoflora* Chang. Acta Scientiarum Naturalium Universitatis Sunyatseni. 1982; (3):74–5.
- Zou TC. Inquire into species origin of *Camellia luteoflora* Y. K. Li, an endemic species in Guizhou. J Guizhou Normal Univ (Social Science). 2002;20(1):6–10.
- He QQ. The distribution pattern of *Camellia luteoflora* Y.K. Li population. Environ Prot Technol. 2012;18(3):28–30.
- Chen F, Wang X. *Camellia luteoflora* Li ex Chang, A newly recorded species of theaceae from Sichuan Province. J Fujian Forestry Sci Technol. 2016;43(4):167–8.
- Liu QB, Liu BY, Liang S. Discussion on the endangered causes and countermeasures of *Camellia luteoflora*. Environ Prot Technol. 2005;11(3):18–20.
- Guo NB, Deng YH, Liu QB. Observation and preliminary study of *Camellia luteoflora*. Environ Prot Technol. 2006;12(1):18–20.
- Ministry of Ecology and Environment of the People's Republic of China and Chinese Academy of Sciences. China Biodiversity Red List - Higher Plants Volume. 2013. [https://www.mee.gov.cn/gkml/hbb/bgg/201309/t20130912\\_260061.htm](https://www.mee.gov.cn/gkml/hbb/bgg/201309/t20130912_260061.htm)
- Zhang T, Liu HY, Zou TC. Main chemical components in leaves of 8 wild *Camellia* species in Guizhou. Guizhou Agricultural Sci. 2010;38(11):78–80.
- Zhang T, Zhou XL, Liu HY. Study on Cottage Propagation Technology of *Camellia luteoflora* Y. K. Li. Seed. 2010; 29(4):86–89. <https://doi.org/10.16590/j.cnki.1001-4705.2010.04.074>
- Zou TC, Zhang ZL, Zhou HY. Photosynthetic properties of five wild plants in Guizhou *Camellia*. Acta Horticulturae Sinica. 1994; (4):366–70.
- Zou TC. Studies on narrow limited distribution and cultivation expansion of 10 endemic species in Guizhou. J Guizhou University(Natural Sciences). 2001;18(4):277–84.
- Rong S, Luo P, Yi H et al. Predicting Habitat Suitability and Adaptation Strategies of an Endangered Endemic Species, *Camellia luteoflora* Li ex Chang (Ericales: Theaceae) under Future Climate Change, Forests. 2023, 14, 2177. <https://doi.org/10.3390/f14112177>
- Gu ZJ, Sun XF. A Karyomorphological study of seventeen species of Chinese *Camellia*. Plant Divers. 1997;19(2):159–70.
- Wang Y, Chen S, Chen JJ, et al. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Photinia serratifolia* Li. Sci Rep. 2023;13(1):770. <https://doi.org/10.1038/s41598-022-24327-x>.
- Tan GF, Li MY, Luo Q, et al. Creation of A male sterility line and identification of its candidate mitochondrial male sterile gene in celery. J Plant Genetic Resour. 2022;23(6):1807–15. <https://doi.org/10.13430/j.cnki.jpgr.20220517003>.
- Lu G, Wang W, Mao J, et al. Complete mitogenome assembly of *Selenicereus monacanthus* revealed its molecular features, genome evolution, and phylogenetic implications. BMC Plant Biol. 2023;23(1):541. <https://doi.org/10.1186/s12870-023-04529-9>.
- Ma Q, Wang Y, Li S, et al. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer truncatum* Bunge: a Woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022;22(1):29. <https://doi.org/10.1186/s12870-021-03416-5>.
- Wu Y, Zhou H. Research progress of *Sugarcane* Chloroplast genome. Agricultural Sci Technol. 2013;14(12):1693–7. <https://doi.org/10.16175/j.cnki.1009-4229.2013.12.015>.
- Janouskovec J, Liu SL, Martone PT, et al. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. PLoS ONE. 2013;8:e59001. <https://doi.org/10.1371/journal.pone.0059001>.
- Kan S, Shen T, Ran H, Wang X. Both conifer LI and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. BMC Biol. 2021;19(1):146. <https://doi.org/10.1186/s12915-021-01096-z>.
- Mower JP, Case AL, Floro ER, et al. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a Monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol. 2012;4(5):670–86. <https://doi.org/10.1093/gbe/evs042>.
- Manchekar M, Scissum-Gunn K, Song D, et al. DNA recombination activity in soybean mitochondria. J Mol Biol. 2006;356(2):288–99. <https://doi.org/10.1016/j.jmb.2005.11.070>.
- Yang H, Chen H, Ni Y, et al. Denovo hybrid assembly of the *Salvia miltiorrhiza* mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial DNA structure of salvia species. Int J Mol Sci. 2022;23(22):14267. <https://doi.org/10.3390/ijms232214267>.
- Yang H, Chen H, Ni Y, et al. Mitochondrial genome sequence of *Salvia officinalis* (Lamiales: Lamiaceae) suggests diverse genome structures in cogenetic species and finds the stop gain of genes through RNA editing events. Int J Mol Sci. 2023;24(6):5372. <https://doi.org/10.3390/ijms24065372>.
- Alverson AJ, Zhuo S, Rice DW, et al. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. PLoS ONE. 2011;6:e16404. <https://doi.org/10.1371/journal.pone.0016404>.
- Arrieta-Montiel MP, Shedje V, Davila J, et al. Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics. 2009;183:1261–8. <https://doi.org/10.1534/genetics.109.108514>.
- Kozik A, Rowan BA, Lavelle D, et al. Christensen the alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15:e1008373. <https://doi.org/10.1371/journal.pgen.1008373>.
- Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the master circle model of plant mitochondrial DNA structure. New Phytol. 2013;200:978–85. <https://doi.org/10.1111/nph.12395>.
- Li J, Xu Y, Shan Y, et al. Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta. 2021;254:36. <https://doi.org/10.1007/s00425-021-03684-3>.
- Sahlin K, Baudeau T, Cazaux B, et al. A survey of mapping algorithms in the long-reads era. Genome Biol. 2023;24(1):133. <https://doi.org/10.1186/s13059-023-02972-3>.
- Koren S, Walenz BP, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res.



- 2017;27(5):722–36. <https://doi.org/10.1101/gr.215087.116>. <http://www.genome.org/cgi/doi/>.
32. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
  33. Wick RR, Judd LM, Gorrie CL, et al. 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>.
  34. Wick RR, Schultz MB, Zobel J, et al. *Bioinformatics*. 2015;31(20):3350–2. <https://doi.org/10.1093/bioinformatics/btv383>. Bandage: interactive visualization of de novo genome assemblies.
  35. Tillich M, Lehwark P, Pellizzer T, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–11. <https://doi.org/10.1093/nar/gkx391>.
  36. Yu JJ, Yang WB, Song JB, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 2020;21(1):241. <https://doi.org/10.1186/s13059-020-02154-5>.
  37. Qu XJ, Moore MJ, Li DZ, et al. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 2019;15:50. <https://doi.org/10.1186/s13007-019-0435-7>.
  38. 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>.
  39. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*. 1990;215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2). PMID: 2231712.
  40. Ran ZH, Li Z, Xiao X, et al. Complete Chloroplast genomes of 13 species of sect. *Tuberculata* Chang (*Camellia* L.): genomic features, comparative analysis, and phylogenetic relationships. *BMC Genomics*. 2024;25:108. <https://doi.org/10.1186/s12864-024-09982-w>.
  41. Li Z, Ran Z, Xiao X, et al. Comparative analysis of the whole mitochondrial genomes of four species in sect. *Chrysanth* (*Camellia* L.), endemic taxa in China. *BMC Plant Biol*. 2024;24:955. <https://doi.org/10.1186/s12870-024-0567-3>.
  42. 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;19:255. <https://doi.org/10.1186/s12859-018-2244-9>.
  43. Julio R, Albert FM, Juan CSD, et al. *Mol Biol Evol*. 2017;34(12):3299–302. <https://doi.org/10.1093/molbev/msx248>. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.
  44. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45. <http://www.genome.org/cgi/doi/10.1101/gr.092759.109>.
  45. Zhang D, Gao FL, Jakovlić C, et al. PhyloSuite: an integrated and scalable desktop Platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020;20(1):348–55. <https://doi.org/10.1111/1755-0998.13096>.
  46. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinformatics*. 2019;20(4):1160–6. <https://doi.org/10.1093/bib/bbx108>.
  47. Hohler D, Pfeiffer W, Ioannidis V, et al. RAXML-Groove: an empirical phylogenetic tree database. *Bioinformatics*. 2022;38(6):1741–2. <https://doi.org/10.1093/bioinformatics/btab863>.
  48. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(1):78–82. <https://doi.org/10.1093/nar/gkx268>.
  49. Wang J, Kan SL, Liao XZ, et al. Plant organellar genomes: much done, much more to do. *Trends Plant Sci*. 2024;29:754–69. <https://doi.org/10.1016/j.tplants.2023.12.014>.
  50. Wei SY, Wang XL, Bi CW, et al. Assembly and analysis of the complete *Salix purpurea* L. (Salicaceae) mitochondrial genome sequence. *Springer Plus*. 2016;5(1):1894. <https://doi.org/10.1186/s40064-016-3521-6>.
  51. Niu YF, Gao CW, Liu J. Mitochondrial genome variation and intergenomic sequence transfers in *Hevea* species. *Front Plant Sci*. 2024;15:1234643. <https://doi.org/10.3389/fpls.2024.1234643>.
  52. Richardson AO, Rice DW, Young GJ, et al. The fossilized mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol*. 2013;11:29. <https://doi.org/10.1186/1741-7007-11-29>.
  53. He X, Zhang X, Deng Y, et al. Structural reorganization in two *Alfalfa* mitochondrial genome assemblies and mitochondrial evolution in medicago species. *Int J Mol Sci*. 2023;24(24):17334. <https://doi.org/10.3390/ijms242417334>.
  54. Gao CW, Wu CH, Zhang Q, et al. Characterization of Chloroplast genomes from two *Salvia medicinal* plants and gene transfer among their mitochondrial and Chloroplast genomes. *Front Genet*. 2020;11:574962. <https://doi.org/10.3389/fgenet.2020.574962>.
  55. Yang JX, Dierckxens N, Bai MZ, et al. Multichromosomal mitochondrial genome of *Paphiopedilum micranthum*: compact and fragmented genome, and rampant intracellular gene transfer. *Int J Mol Sci*. 2023;24(4):3976. <https://doi.org/10.3390/ijms24043976>.
  56. Cheng Y, He X, Priyadarshani SVGN, Aslam M, Qin Y, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. *BMC Genomics*. 2021;22(1):167. <https://doi.org/10.1186/s12864-021-07490-9>.
  57. Zhang LY, Hu J, Han XL, et al. A high-quality Apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun*. 2019;10(1):1494. <https://doi.org/10.1038/s41467-019-09518-x>.
  58. Li C, Tang J, Hu Z, et al. A novel maize Dwarf mutant generated by Ty1-copia LTR-retrotransposon insertion in Brachytic2 after spaceflight. *Plant Cell Rep*. 2020;39(3):393–408. <https://doi.org/10.1007/s00299-019-02498-8>.
  59. Li J, Chen Y, Liu Y, et al. Complete mitochondrial genome of *Agrostis stolonifera*: insights into structure, codon usage, repeats, and RNA editing. *BMC Genomics*. 2023;24(1):466. <https://doi.org/10.1186/s12864-023-09573-1>.
  60. Fan WS, Liu F, Jia QY, et al. *Fragaria* mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by microinversions. *New Phytol*. 2022;236(2):745–59. <https://doi.org/10.1111/nph.18334>.
  61. Wang L, Liu X, Xu YJ, et al. Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine *Angelica biserrata* (Shan et Yuan) Yuan et Shan. *Int J Biol Macromol*. 2024;257(1):128571. <https://doi.org/10.1016/j.ijbiomac.2023.128571>.
  62. Krasovec M, Filatov DA. Codon usage bias in phytoplankton. *J Mar Sci Eng*. 2022;10:168. <https://doi.org/10.3390/jmse10020168>.
  63. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol Biol Rep*. 2009;36:2039–46. <https://doi.org/10.1007/s11033-008-9414-1>.
  64. Xu W, Xing T, Zhao M, et al. Synonymous codon usage bias in plant mitochondrial genes is associated with intron number and mirrors species evolution. *PLoS ONE*. 2015;10(6):e0131508. <https://doi.org/10.1371/journal.pone.0131508>.
  65. Wang B, Yuan J, Liu J, et al. Codon usage bias and determining forces in green plant mitochondrial genomes. *J Integr Plant Biol*. 2011;53(4):324–34. <https://doi.org/10.1111/j.1744-7909.2011.01033.x>.
  66. Wee CC, Muhammad NAN, Subbiah VK, et al. Mitochondrial genome of *Garcinia Mangostana* L. variety mesta. *Sci Rep*. 2022;12(1):9480. <https://doi.org/10.1038/s41598-022-13706-z>.
  67. Alverson AJ, Wei XX, Rice DW, et al. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita Pepo* (Cucurbitaceae). *Mol Biol Evol*. 2010;27(6):1436–48. <https://doi.org/10.1093/molbev/msq029>.
  68. Mower JP, Sloan DB, Alverson AJ. *Plant Genome Diversity Volume 1: Plant Mitochondrial Genome Diversity: The Genomics Revolution*, Vienna: Springer. 2012: 123–144. <https://doi.org/10.1007/978-3-7091-1130-7>.
  69. Bergthorsson U, Adams K, Thomason B, et al. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*. 2003;424(6945):197–201. <https://doi.org/10.1038/nature01743>.
  70. Zhang D, Yu H, Gao L, et al. Genetic diversity in oilseed and *Vegetable mustard* (*Brassica juncea* L.) accessions revealed by nuclear and mitochondrial molecular markers. *Agronomy*. 2023;13(3):919. <https://doi.org/10.3390/agronomy13030919>.
  71. Zhang Z, Li J, Zhao XQ, Genomics, et al. *Proteom Bioinf*. 2006;4(4):259–63. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2).
  72. Jérôme D, Besnard G. Utility of the mitochondrial genome in plant taxonomic studies. *Methods Mol Biol*. 2021;2222:107–18. [https://doi.org/10.1007/978-1-0716-0997-2\\_6](https://doi.org/10.1007/978-1-0716-0997-2_6).
  73. Christensen AC. Plant mitochondria are a riddle wrapped in a mystery inside an enigma. *J Mol Evol*. 2021;89(3):151–6. <https://doi.org/10.1007/s00239-020-09980-y>.

74. Kozik A, Rowan BA, Lavelle D, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. *PLoS Genet.* 2019;15(8):e1008373. <https://doi.org/10.1371/journal.pgen.1008373>.

### **Publisher's note**

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