Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Assembly and characterization of the mitochondrial genome of Camellia pitardii Coh.-St

**Juyan Chen**

jychen2008@163.com

Guizhou Academy of Forestry

**Rui Chen**

Guizhou Academy of Forestry

**He Li**

Guizhou Academy of Forestry

**Lang Huang**

Guizhou Academy of Forestry

**Research Article**

**Keywords:** Camellia pitardii, mitochondrial genome, repeat sequences, RNA editing, phylogenetic analysis

**Posted Date:** July 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9991553/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

**Version of Record:** A version of this preprint was published at BMC Genomics on November 3rd, 2026. See the published version at https://doi.org/10.1186/s12864-026-13508-x.

# Abstract

*Camellia pitardii* Coh.-St., an economically important tree species in the genus *Camellia* valued for both its seed oil and ornamental traits, currently lacks a systematically characterized mitochondrial genome. In this study, we assembled and annotated the complete mitogenome of *C. pitardii* using a hybrid approach that combined Illumina NovaSeq 6000 short reads and Oxford Nanopore long reads. Using bioinformatics tools, we systematically analyzed repetitive sequences, codon usage bias, and RNA editing sites. Furthermore, we constructed phylogenetic trees based on conserved protein-coding genes from both mitochondrial and chloroplast genomes to elucidate the phylogenetic position of these genes. The assembled *C. pitardii* mitogenome is 895,290 bp in length with a GC content of 45.68%. Annotation revealed a total of 38 protein-coding genes (PCGs), three rRNA genes, and 30 tRNA genes. We identified 260 simple sequence repeats (SSRs), 33 tandem repeats, and 620 pairs of dispersed repeats. Among these, three pairs of long repeats are predicted to mediate the formation of eight potential genomic substructures. The mitogenome exhibited a distinct codon usage bias, with 92.86% of the highly preferred codons (relative synonymous codon usage [RSCU] > 1) ending in A or U. Additionally, 667 RNA editing sites were predicted across the 38 PCGs, encompassing 14 types of amino acid substitutions; 46.48% of these editing events resulted in the conversion of hydrophilic residues to hydrophobic ones. Phylogenetic analyses revealed topological incongruence between the mitochondrial and chloroplast gene trees regarding the placement of several *Camellia* clades, suggesting a complex evolutionary history for the organelles of *C. pitardii* and its close relatives. In conclusion, this study characterizeed the fundamental features of the *C. pitardii* mitogenome, providing valuable baseline data for comparative organellar genomics, phylogenetic reconstruction, and the utilization of germplasm resources within the genus Camellia.


# 1. Introduction

*Camellia pitardii* a member of the family *Theaceae*, is a wild woody species primarily distributed across southwestern mountainous regions of China and adjacent territories. Its natural range encompasses the provinces of Guizhou, Sichuan, Yunnan, Guangxi, and Hunan. Notably, many wild forest stands remain intact in specific locations, such as Panzhou (Guizhou) and Panxi (Sichuan) [1]. *C. pitardii* is highly valued for both its oleaginous and ornamental traits. The seed kernels yield a substantial amount of oil rich in unsaturated fatty acids, highlighting its potential for commercial edible woody oil production. Furthermore, its vibrant floral pigmentation and extended blooming period confer significant ornamental value [2–5]. Given these distinct phenotypic traits and broad economic utility, *C. pitardii* represents a critical genetic resource for germplasm conservation, woody oil development, and ornamental plant breeding.

Currently, research on *C. pitardii* predominantly focuses on germplasm exploration, fruit morphology, lipid content, fatty acid profiling, floral phenotypic variation, and agronomic practices [6, 7]. These investigations reveal substantial phenotypic and genetic diversity across populations, establishing a foundational framework for germplasm evaluation and the selection of elite genotypes [8]. Despite these

advances in phenotypic and applied research, the molecular genetics of *C. pitardii* remain insufficiently explored. Although recent efforts have utilized ISSR markers to assess the genetic diversity of wild populations in Guizhou and have successfully generated a high-quality nuclear genome assembly, systematic investigations into its organelle genomes are notably absent [9, 10]. Specifically, the structural architecture, gene content, and evolutionary dynamics of the *C. pitardii* mitochondrial genome remain completely uncharacterized.

As essential organelles for cellular energy metabolism, mitochondria primarily synthesize ATP via oxidative phosphorylation to drive plant growth and development [11]. Plant mitochondrial genomes typically encode critical subunits of respiratory chain complexes, including NADH dehydrogenase, cytochrome c oxidase, and ATP synthase, which are indispensable for maintaining proper mitochondrial function [12, 13]. Consequently, elucidating the structural architecture and genetic composition of the *C. pitardii* mitogenome is a prerequisite for comprehensively characterizing the organellar genetics of this species. In contrast to their animal counterparts, plant mitogenomes exhibit highly complex structural features characterized by substantial size variations, abundant repetitive sequences, frequent recombination events, and diverse structural conformations [14–16]. Extensive research has demonstrated that intragenomic repeats frequently mediate homologous recombination, driving the formation of dynamic configurations in which multiple isoforms coexist. Thus, a single linear or circular sequence often fails to accurately capture its true *in vivo* architecture [17–19]. To address the current lack of genomic data for *C. pitardii*, this study presents a systematic assembly and structural analysis of its mitogenome. This work not only resolves the organellar genetic profile of *C. pitardii* but also fills a critical gap in the mitogenomic resources available for the genus *Camellia*. Ultimately, these findings provide essential foundational data for future comparative organellar genomics, phylogenetic analyses, investigations of genetic differentiation, and the utilization of elite *Camellia* germplasm.

## 2. Materials and Methods

# 2.1 Genome sequencing, assembly, and annotation

The *C. pitardii* samples used in this study were collected from Yunwu Mountain, Niaowang Village, Guiding County, Guizhou Province, and healthy fresh leaves were taken for total genomic DNA extraction. The geographical coordinates of the sampling site are 107°00'58.8228" E, 26°12'57.1870" N, with an elevation of 1352.5 m. The plant samples were identified as *C. pitardii* based on their morphological characteristics.All plant materials used in this study were formally authenticated by Professor Shixiong Yang, a *Camellia* taxonomist at the Kunming Institute of Botany, Chinese Academy of Sciences. A voucher specimen (accession No. 011209) has been deposited in the Herbarium of the Guizhou Academy of Forestry (GZAF). Total DNA was extracted using the CTAB method [20], DNA integrity was assessed using agarose gel electrophoresis [21], and DNA purity and concentration were measured using NanoDrop and Qubit [22, 23]. The qualified DNA was used for Illumina short-read and Oxford Nanopore long-read sequencing library construction [24, 25].

Paired-end genomic libraries were constructed and subsequently sequenced on an Illumina NovaSeq 6000 platform to generate short-read data [26]. Raw short reads were subjected to quality control using fastp [27] to remove adapter sequences, low-quality reads, and reads with a high proportion of undetermined bases (Ns). This process yielded clean short reads for downstream error correction and assembly validation. For long-read sequencing, high-molecular-weight (HMW) DNA was prepared according to the manufacturer's protocol, which included end repair, dA-tailing, and adapter ligation. The resulting library was subsequently sequenced on an Oxford Nanopore PromethION platform [28]. Raw long reads were filtered using Filtlong to obtain high-quality data for mitochondrial genome assembly and structural validation. Summary statistics for all sequencing datasets are provided in Supplementary Table S1.

To initiate mitochondrial genome assembly, Nanopore long reads were mapped against a plant mitochondrial reference database using minimap2 to identify candidate reads of mitochondrial origin. On the basis of the initial alignments, core mitochondrial gene distribution, and sequencing depth, high-confidence sequences were selected as seeds. The raw long reads were then iteratively mapped to these seeds via minimap2 to enrich the pool of mitochondria-associated reads. Following error correction with Canu [29], these enriched long reads were subjected to *de novo* hybrid assembly alongside Illumina clean reads using Unicycler [30]. Assembly graphs were visualized in Bandage, and contig connectivity and branching topologies were resolved by integrating long-read mapping profiles. Finally, ambiguous configurations near complex repeats were resolved using long reads that spanned the entire repetitive regions and their flanks.

The mitochondrial genome was annotated through a combination of homology-based sequence alignment and manual curation. Initially, the assembled sequences were aligned against protein-coding and rRNA genes from previously published mitochondrial genomes of the genus *Camellia* and other angiosperms using BLASTN [31]. These preliminary annotations were subsequently subjected to manual curation, utilizing the mitochondrial genome of the tea plant (*C. sinensis*, GenBank accession PP212895) and related species as reference frameworks. Putative tRNA genes were identified using tRNAscan-SE [32]. Open reading frames (ORFs) were predicted using ORFfinder [33], after which of redundant sequences and those overlapping with known genes were removed. The remaining ORFs exceeding 300 bp were queried against the NCBI non-redundant (nr) protein database to identify potential functional annotations. Finally, the mitochondrial genome map of the *C. pitardii* was visualized using OGDRAW [34].

# 2.2 Mitochondrial DNA repeat sequence analysis

To characterize the repeat sequence profile of the *C. pitardii* mitochondrial genome, simple sequence repeats (SSRs), tandem repeats, and dispersed repeats were systematically analyzed. SSRs were detected using MISA v1.0 [35, 36], with the minimum repeat thresholds set to 10 for mononucleotides, 5 for dinucleotides, 4 for trinucleotides, and 3 for tetra-, penta-, and hexanucleotides. Tandem repeats were identified using Tandem Repeats Finder v4.09 [37] under default parameters (Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50, MaxPeriod = 500). Sequences satisfying the length and alignment criteria were extracted to record their genomic position, repeat unit length, copy number, and

sequence identity. Dispersed repeats were identified via BLASTN v2.10.1, applying a word size of 7 and an E-value threshold of 1e-5. The resulting alignments were filtered to exclude self-hits, redundant records, and regions overlapping with tandem repeats [38]. On the basis of their sequence orientation and arrangement, the dispersed repeats were subsequently categorized as forward, reverse, or palindromic. Finally, Circos v0.69-5 was employed to visualize the genomic distribution of all identified repeat sequences across the *C. pitardii* mitochondrial genome[39].

## 2.3. Codon usage bias analysis

On the basis of the mitochondrial genome annotation of *C. pitardii*, the coding sequences (CDSs) of protein-coding genes (PCGs) were extracted using PhyloSuite v1.22 [40]. To ensure sequence integrity, pseudogenes and truncated sequences were excluded, and only intact, full-length PCGs were retained for downstream analysis. The relative synonymous codon usage (RSCU) values were subsequently calculated using MEGA v7.0 [41]. In this metric, an RSCU value > 1.0 denotes a positive codon usage bias (preferred codons), whereas an RSCU < 1.0 signifies a lower relative frequency among synonymous codons.

## 2.4 Prediction of RNA editing sites

To identify putative C-to-U RNA editing sites within the mitochondrial protein-coding genes (PCGs) of *C. pitardii*, the corresponding coding sequences (CDSs) were analyzed using the PREPACT 3.0 web server (http://www.prepact.de/) applying a stringent cutoff threshold of 0.001 [42]. The computational output detailes the specific gene loci of the editing sites, the pre- and post-editing codon conversions, and the resulting amino acid substitutions. Because empirical validation via RNA-seq was not performed in the current study, all the identified loci are strictly designated as putative RNA editing sites. Microsoft Excel 2021 was used to compile and visualize the frequency of these putative sites, the spectrum of amino acid substitutions, and the subsequent alterations in protein hydropathy.

## 2.5. Phylogenetic analysis

To determine the phylogenetic position of *C. pitardii* within the genus *Camellia*, independent phylogenetic trees were constructed using mitochondrial and chloroplast protein-coding genes (PCGs). The dataset included *C. pitardii* and representative *Camellia* species with available mitochondrial or chloroplast genomes in GenBank. Closely related taxa and other representative plant species were designated as outgroups [14–19, 43–52]. Comprehensive taxonomic information and GenBank accession numbers for all the sampled taxa are detailed in Supplementary Table S2.

For the mitochondrial phylogeny, PhyloSuite v2 [53] was employed to extract conserved PCGs from the assembled genomes. To mitigate potential phylogenetic biases introduced by missing data or fragmented genes, only intact PCGs shared across all sampled taxa were retained, yielding a core set of 37 mitochondrial coding sequences. Each gene was individually aligned using MAFFT v7.471 [54]. The resulting alignments were trimmed to remove poorly aligned regions using Gblocks v0.91b [55] and subsequently concatenated into a single supermatrix, resulting in a final alignment length of 37,236 bp.

Taxon sampling for the chloroplast phylogenetic analysis was consistent with that for the mitochondrial dataset. In total,64 shared protein-coding genes (PCGs) were extracted from the chloroplast genomes of the selected species using PhyloSuite. Each PCG was individually aligned using MAFFT and subsequently trimmed with Gblocks. The refined alignments were then concatenated to construct a chloroplast phylogenetic supermatrix with a total length of 41,730 bp.

Phylogenetic trees were reconstructed using both maximum likelihood (ML) and Bayesian inference (BI) approaches. The optimal nucleotide substitution models were determined using ModelFinder 2 [56] on the basis of the Akaike information criterion (AIC). ML analysis was performed in IQ-TREE v3.0.0, with nodal support evaluated via 1,000 bootstrap replicates [57]. BI analysis was conducted using MrBayes v3.2.7a. The Markov chain Monte Carlo (MCMC) algorithm was run for 1,000,000 generations, with sampling every 100 generations. The first 25% of the sampled trees were discarded as burn-in, and a 50% majority-rule consensus tree was generated after confirming convergence [58]. Ultimately, the ML bootstrap support (BS) and BI posterior probabilities (PP) were mapped onto the corresponding nodes of the resulting phylogenetic trees.

# 3 Results

## 3.1 *C. pitardii* mitochondrial genome assembly and annotation

Using a hybrid assembly strategy that integrates Illumina short reads and Nanopore long reads, we successfully assembled the mitochondrial genome (mitogenome) of *C. pitardii*. Topological analysis of the assembly graph revealed a complex, multibranched structure. On the basis of this topology and supported by long-read spanning evidence, the master conformation of the *C. pitardii* mitogenome was determined to be 895,290 bp in length, with a GC content of 45.68%. The resolved assembly path is as follows: contig3–contig7–LR11–contig2–LR10–contig4–LR12–contig8–LR11–contig6–contig1–LR10–contig9–LR12–contig5 (Fig. 1A, B). Within this sequence, LR10, LR11, and LR12 denote long repeat sequences—specifically comprising two pairs of forward repeats and one pair of inverted repeats—that mediate this structural complexity. Finally, mapping the raw sequencing reads back to the assembly demonstrated continuous coverage across the genome, validating the accuracy and reliability of the final assembly (Supplementary Figure S1).

Annotation of the *C. pitardii* mitochondrial genome (mitogenome) revealed a total of 38 protein-coding genes (PCGs), three ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes (Fig. 2, Table 1). The identified PCGs encode ATP synthase subunits (*atp1*, *atp4*, *atp6*, *atp8*,and *atp9*), cytochrome c biogenesis proteins (*ccmB*, *ccmC*, *ccmFc*,and *ccmFn*), cytochrome c oxidases (*cox1*, *cox2*,and *cox3*), NADH dehydrogenase subunits (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*,and *nad9*), and various ribosomal proteins. Additionally, single pseudogenized copies were detected for *cox2*, *rps19*, and *sdh3*. The three rRNA genes were identified as *rrn5*, *rrn18*, and *rrn26*. Several tRNA genes are present in

multiple copies; notably, *trn*M-CAT has the greatest redundancy, with five annotated copies, whereas *trn*C-GCA, *trn*I-GAT, *trn*P-TGG, t*rn*S-GCT, *trn*S-TGA, and *trn*T-TGT each have two copies.

Table 1
Gene content of the *C. pitardii* mitochondrial genome.

| | Gene Category | Gene Name |
|---|---|---|
| Core Genes | ATP synthase | *atp*1, *atp*4, *atp*6, *atp*8, *atp*9 |
| | Cytochrome c biogenesis | *ccm*B, *ccm*C, *ccm*Fc*, *ccm*Fn |
| | Ubiquinol cytochrome c reductase | *cob* |
| | Cytochrome c oxidase | #*cox*2, *cox*1, *cox*2, *cox*3 |
| | Maturases | *mat*R |
| | Transport membrane protein | *mtt*B |
| | NADH dehydrogenase | *nad*1****, *nad2*****, *nad*3, *nad*4**, *nad*4L, *nad*5****, *nad*6 *nad7*****, *nad*9 |
| Variable Genes | Ribosomal proteins (LSU) | *rpl*10, *rpl*16, *rpl*2,* *rpl*5 |
| | Ribosomal proteins (SSU) | #*rps*19, *rps*1, *rps*12, *rps*13, *rps*14, *rps*19, *rps*3*, *rps*4, *rps*7 |
| | Succinate dehydrogenase | #*sdh*3, *sdh*3, *sdh*4 |
| | Ribosomal RNAs | *rrn*18, *rrn*26, *rrn*5 |
| | Transfer RNAs | *trn*A-TGC*, *trn*C-GCA(2), *trn*D-GTC, *trn*E-TTC, *trn*F-GAA, *trn*G-GCC, *trn*H-GTG, *trn*I-GAT*(2), *trn*K-TTT, *trn*M-CAT(5), *trn*N-ATT, *trn*N-GTT, *trn*P-TGG(2), *trn*Q-TTG, *trn*S-GCT(2), *trn*S-TGA, *trn*S-TGA*, *trn*T-TGT*(2), *trn*V-GAC, *trn*W-CCA, *trn*Y-GTA |

Note: Numbers adjacent to gene names indicate copy numbers. The prefix "#" denotes pseudogenes. The number of asterisks "*" appended to a gene represents the number of introns it contains.

Structural analysis revealed the presence of introns within several mitochondrial genes. Specifically, single introns were identified in *ccmFc*, *rpl2*, *rps3*, and a subset of tRNA genes. Furthermore, *nad4* contains two introns, whereas *nad1*, *nad2*, *nad5*, and *nad7* each harbor four introns. Interestingly, structural variations were observed among multicopy tRNAs; for instance, one copy of *trn*S-TGA contains an intron, whereas the second copy is intronless.

## 3.2. Repeat sequence analysis of the *C. pitardii* mitochondrial genome

Analysis of the *C. pitardii* mitochondrial genome revealed three categories of repetitive sequences: simple sequence repeats (SSRs), tandem repeats, and dispersed repeats (Fig. 3). In total, 260 SSRs were identified, comprising 35 mono-, 68 di-, 38 tri-, 99 tetra-, 19 penta-, and single hexanucleotide repeats (Supplementary Table S3A). Tetranucleotide motifs were the most abundant, followed by dinucleotides, whereas hexanucleotide repeats were the rarest.

Furthermore, 33 tandem repeat loci were detected, with motif lengths ranging from 13 to 39 bp. The copy number per locus varied between 1.9 and 3.0, indicating a predominance of low-copy-number repeats. The locus situated within the 376,477–376,548 bp region presented the maximum copy number (3.0). Notably, two sets of overlapping tandem repeat units were identified within the 438,595–438,665 bp region; both motifs were 27 bp in length and shared 97% sequence identity (Supplementary Table S3B).

With respect to dispersed repeats, 620 sequence pairs were identified, consisting of 299 forward and 321 palindromic repeats. These elements ranged from 30 to 1,747 bp in length, with the vast majority (525 pairs) concentrated within the 30–100 bp size fraction (Supplementary Table S3C). Three exceptionally long and highly homologous repeat pairs were characterized: two forward and one palindromic. The two forward repeat pairs measured 1,747 bp and 505 bp, respectively, both of which exhibited > 99% sequence similarity. The single palindromic repeat spanned 1,514 bp with 99.67% sequence identity. These long, high-identity repetitive elements are likely to mediate homologous recombination and the subsequent generation of isomeric conformations within the *C. pitardii* mitogenome, making them primary targets for future structural dynamics analyses.

## 3.3 Potential mitogenome substructures mediated by long repeats

Within the mitochondrial genome of *C. pitardii*, LR10, LR11, and LR12 represent highly similar, long repeat sequences that are likely to drive homologous recombination and subsequent structural rearrangements. Specifically, LR10 and LR12 are direct repeats, whereas LR11 is an inverted repeat. On the basis of their genomic positions, orientations, and topological relationships within the assembly graph, we hypothesize that these three pairs of long repeats mediate the formation of multiple alternative substructures (Fig. 1C). In the primary conformation, the direct repeat LR10 connects contig2 to contig4 (contig2→LR10→contig4) and contig1 to contig9 (contig1→LR10→contig9). Homologous recombination across LR10 generates two alternative structural associations: contig2→LR10→contig9 and contig1→LR10→contig4. Similarly, the primary configurations surrounding the direct repeat LR12 (contig4→LR12→contig8 and contig9→LR12→contig5) could recombine to yield contig4→LR12→contig5 and contig9→LR12→contig8. Concurrent recombination events at both LR10 and LR12 consequently generate novel chimeric conformations. In contrast to direct repeats, the inverted repeat LR11 triggers different structural dynamics. Its native arrangement in the primary

genome is contig7→LR11→contig2→LR10→contig4→LR12→contig8→LR11→contig6. Recombination across this inverted repeat typically induces the inversion of the intervening segment, generating an alternative conformation: contig7→LR11→contig8→LR12→contig4→LR10→contig2→LR11→contig6. This event reverses the orientations of contig2, LR10, contig4, LR12, and contig8. Consequently, several previously direct repeats may be converted into inverted repeats, further amplifying the structural complexity of the mitogenome.

By integrating the combinatorial recombination events mediated by LR10, LR11, and LR12, we predict that the *C. pitardii* mitogenome can transition between eight distinct structural conformations. Although long-read mapping profiles provide preliminary evidence for these complex genomic associations, validating the *in vivo* presence and relative stoichiometry of these substructures will require more in-depth long-read sequencing or targeted experimental assays.

# 3.4 Codon usage bias analysis

We analyzed the relative synonymous codon usage (RSCU) of all intact protein-coding genes (PCGs) in the *C. pitardii* mitochondrial genome. For clarity, codons are denoted using the RNA alphabet (where U replaces T). All 64 standard codons, comprising 61 sense codons and three stop codons, were identified within the *C. pitardii* mitochondrial PCGs (Fig. 4; Supplementary Table S4).

Among the 61 sense codons, 28 had RSCU values > 1.0, reflecting a usage preference among their synonymous counterparts. Specifically, GCU (Ala) demonstrated the highest RSCU value (1.5906), followed by UAU (Tyr; 1.5389) and CAU (His; 1.5379). Notably, 26 of these 28 preferred codons terminated in either U (n = 16) or A (n = 10), whereas only two terminated in C (n = 1) or G (n = 1). This distribution reveals a pronounced A/U-ending bias among the preferred codons in the *C. pitardii* mitogenome.

With respect to the termination signals, UAA was the most frequently utilized stop codon (RSCU = 1.5789). These codon usage patterns are inherently linked to the overall base composition of the *C. pitardii* mitochondrial genome, which is characterized by a strong preference for A/U-terminating codons.

# 3.5 RNA editing prediction

PREPACT 3.0 was used to predict RNA editing sites within the mitochondrial protein-coding genes (PCGs) of *C. pitardii*. In total, 667 candidate RNA editing sites were identified across the analyzed PCGs. These modifications are predicted to yield 14 distinct types of amino acid substitutions: 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), Q(Gln)→X,R(Arg)→X, P(Pro)→S(Ser), P(Pro)→L(Leu), P(Pro)→F(Phe), L(Leu)→F(Phe) and A(Ala)→V(Val), where X denotes a premature stop codon.

S→L conversion was the most frequent, occurring at 143 sites and comprising 21.44% of the total predicted edits. Conversely, events generating stop codons (Q→X and R→X) were the least frequent,

with only three occurrences (0.45%) each (Fig. 5A; Supplementary Table S5A). Analysis of physicochemical property alterations revealed that 310 editing events (46.48%) converted hydrophilic residues to hydrophobic ones,whereas 50 events (7.50%) resulted in hydrophobic-to-hydrophilic shifts. The hydrophobicity of 301 sites (45.13%) remained unaltered post-editing. The six sites generating premature stop codons were excluded from the hydrophobicity analysis (Fig. 5B).

Analysis of the gene-specific distribution revealed that *ccmFn* contained the greatest number of candidate editing sites (45), closely followed by *nad2* (42). In contrast, sdh3 contained the fewest predicted sites, with only three (Fig. 5C; Supplementary Table S5B). Overall, the mitogenome of *C. pitardii* contains a substantial number of RNA editing sites that are heterogeneously distributed across its PCGs. However, as these findings are derived from in silico predictions, subsequent validation using transcriptomic data or empirical experimental approaches is necessary to confirm the actual editing events.

# 3.6 Phylogenetic analysis

To elucidate the phylogenetic placement of *C. pitardii* within the genus *Camellia*, phylogenetic trees were reconstructed using conserved mitochondrial and chloroplast protein-coding genes (PCGs). Both the mitochondrial and chloroplast PCG phylogenies supported the monophyly of the genus *Camellia*. However, topological incongruences were observed between the two organellar datasets regarding the specific placement of certain *Camellia* species.

The mitochondrial PCG phylogeny resolved the *Camellia* species into two primary clades (Fig. 6). Clade Ⅰ, which was well-supported (BS = 74, PP = 1.00), included *C. pitardii*, *C. drupifera*, *C. oleifera* var. *oleifera*, *C. sinensis* var. *sinensis*, and *C. sinensis* var. *pubilimba*. Clade Ⅱ contained *C. chekiangoleosa*, *C. granthamiana*, *C. nitidissima*, *C. huana*, and *C. sinensis* var. *assamica*. Notably, the Bayesian posterior probability for Clade II was distinctly low (PP = 0.32), indicating poor phylogenetic resolution within this group; consequently, the internal topology of this clade should be interpreted with caution.

Phylogenetic reconstruction based on chloroplast protein-coding genes (PCGs) strongly supports the monophyly of the genus *Camellia* (BS = 100, PP = 1.00); however, its intrageneric topology diverges from that of the mitochondrial PCG tree (Fig. 7). Within the chloroplast phylogeny, *C. huana* diverges early as an independent lineage, while all the other *Camellia* species form a highly supported major clade (Clade III, BS = 100, PP = 1.00). A distinct subclade within Clade III (Clade IV, BS = 100, PP = 1.00) comprises *C. pitardii*, *C. granthamiana*, *C. sinensis* var. *sinensis*, *C. chekiangoleosa*, *C. gigantocarpa*, and *C. drupifera*. Within this group, *C. pitardii*, *C. oleifera* var. *oleifer*, *C. chekiangoleosa*, and *C. gigantocarpa* form a further subdivision (Clade VI, BS = 90, PP = 1.00). *C. pitardii* occupies a basal position within Clade VI, whereas *C. chekiangoleosa* and *C. gigantocarpa* are recovered as sister taxa (BS = 78, PP = 1.00).

While both the mitochondrial and chloroplast PCG phylogenies consistently place *C. pitardii* within the genus *Camellia*, topological discordance between the two organellar trees is evident. This incongruence is particularly pronounced in terms of the phylogenetic placements of *C. pitardii*, *C. chekiangoleosa*, and

*C. huana*. This mito-plastid discordance suggests a complex evolutionary history for *Camellia* organellar genomes. Furthermore, given that organellar loci do not always reflect the true species tree and that certain nodes currently lack robust statistical support, clarifying the exact phylogenetic relationships of *Camellia* sect. *Camellia* and its closely related taxa will necessitate broader taxon sampling combined with nuclear genomic data.

## 4 Discussion

# 4.1. Structural features and comparative analysis of the *C. pitardii* mitochondrial genome

Plant mitochondrial genomes are typically characterized by structural complexity, extensive repetitive sequences, and frequent recombination events [14–19, 43–51]. Conventional short-read sequencing platforms are inherently limited by their read lengths, rendering them ineffective at spanning long repetitive sequences and complex genomic junctions. Consequently, assemblies relying solely on these technologies are often highly fragmented or erroneously collapse complex multipartite conformations into a single sequence representation. Conversely, third-generation long-read sequencing can bridge these extensive repeats, facilitating the accurate resolution of branched structures and alternative recombinational conformations inherent to plant mitogenomes [59, 60]. By integrating Illumina short-read and Oxford Nanopore long-read data, this study successfully assembled the master conformation of the *C. pitardii* mitogenome, yielding a total length of 895,290 bp with a GC content of 45.68%. These genomic data substantially expand the available mitogenomic resources for the genus *Camellia* and highlight the pronounced structural complexity of the *C. pitardii* mitochondrial genome.

Relative to previously published mitochondrial genomes within the genus *Camellia* [15, 16, 43–47, 51], the *C. pitardii* mitogenome falls at the larger end of the size spectrum. Size variations among plant mitogenomes are typically driven by the accumulation of repetitive sequences, the integration of foreign DNA, intergenic region expansion, and structural recombination. In the *C. pitardii* mitogenome, we identified an extensive array of repetitive elements, comprising 260 simple sequence repeats (SSRs), 33 tandem repeat loci, and 620 pairs of dispersed repeats. Notably, LR10, LR11, and LR12 are represent highly homologous long repeats that likely mediate the formation of alternative recombination conformations. Specifically, the forward repeats LR10 and LR12 may facilitate recombination between distinct genomic fragments, whereas the inverted repeat LR11 is likely involved in structural inversions [61]. On the basis of the physical positions and orientations of these long repeats, we propose that the *C. pitardii* mitogenome forms eight potential subgenomic conformations. However, because these substructures were inferred primarily from assembly graphs, repeat orientations, and long-read mapping profiles, their *in vivo* existence and relative stoichiometric abundances require further experimental validation using high-depth long-read sequencing or PCR assays.

With respect to gene content, the *C. pitardii* mitogenome encodes 38 protein-coding genes, 3 rRNA genes, and 30 tRNA genes, an inventory broadly consistent with most angiosperm mitogenomes [14–19,

43–51]. Interestingly, pseudogenized copies of *cox2*, *rps19*, and *sdh3* were identified, indicating structural alterations driven by gene duplication, degradation, or genomic recombination [62–64]. Among the tRNAs, *trn*M-CAT has the highest copy number (five copies), while several others—including *trn*C-GCA, *trn*I-GAT, *trn*P-TGG, *trn*S-GCT, *trn*S-TGA, and *trn*T-TGT—are also multicopy. The proliferation of these tRNA genes likely stems from repeat sequence expansion, localized fragment duplications, or inter-organellar DNA transfer. Elucidating their precise functional significance, however, will require future integration with transcriptomic data [62, 63, 65].

The mitochondrial genome of *C. pitardii* is characterized by substantial architectural complexity, driven by its variable size, repetitive sequence landscape, and the presence of recombination-mediated alternative conformations. Rather than relying solely on conventional metrics such as genome length and gene inventory, incorporating long-repeat analysis alongside assembly graph topologies provides a more rigorous understanding of this structural diversity. Nevertheless, as the current findings rely predominantly on *in silico* genome assembly and bioinformatics predictions, empirical validation is necessary. Future *in vivo* investigations are needed to determine the relative stoichiometry of these multipartite conformations and to elucidate their downstream effects on mitochondrial gene expression and cellular function.

## 4.2 Codon usage bias of mitochondrial protein-coding genes in *C. pitardii*

Codon usage bias (CUB) describes the non-random selection of synonymous codons, a phenomenon primarily driven by mutational pressure (e.g., base composition) and translational selection (e.g., tRNA availability) [65–67]. In this study, a relative synonymous codon usage (RSCU) analysis of the protein-coding sequences (CDSs) in the *C. pitardii* mitogenome revealed 28 preferred codons (RSCU > 1). Notably, 92.86% of these preferred codons terminate in A/U, whereas only 7.14% end in C/G, indicating a pronounced A/U-ending preference. This bias aligns with the overall A/T-rich nature of the *C. pitardii* mitogenome, which has a GC content of only 45.68%. Similar codon usage patterns have been documented across the mitogenomes of various angiosperms [14–19, 43–51]. These findings suggest that mutational pressure acting on base composition is a primary driver of this A/U preference, representing a conserved evolutionary feature among plant mitochondrial genes.

Beyond mutational pressure, translational selection via tRNA availability likely contributes to shaping this bias. However, mitochondrial CUB is not solely dictated by the endogenous tRNA pool. Because plant mitogenomes frequently lose tRNA genes during evolution, organellar translation heavily relies on the import of nucleus-encoded tRNAs, alongside expanded codon recognition facilitated by wobble base pairing [68, 69]. Consequently, the observed A/U bias in *C. pitardii* is likely a synergistic outcome of intrinsic base composition and complex tRNA utilization dynamics.

Furthermore, post-transcriptional RNA editing can alter codon composition, leading to non-synonymous amino acid substitutions [70]. Although our *in silico* analysis predicted an abundance of candidate RNA editing sites—indicating extensive transcript modification—these sites await experimental validation. The

lack of corresponding RNA-seq data, coupled with unmeasured tRNA expression profiles and translation efficiencies, currently precludes a definitive assessment of the causal relationships among RNA editing, CUB, and translational dynamics. Future multiomics studies integrating transcriptomic, tRNA profiling, and proteomic data are needed to systematically elucidate the mechanistic basis and functional significance of mitochondrial CUB in *C. pitardii*.

## 4.3 Prediction and potential biological significance of mitochondrial RNA editing sites in *C. pitardii*

RNA editing is a critical post-transcriptional modification in plant organelle gene expression, primarily manifested as cytidine-to-uridine (C-to-U) conversion within the mitochondrial genome [70, 71]. These conversions alter the codon identities of transcripts, potentially leading to amino acid substitutions or the creation of novel initiation and termination codons. In this study, PREPACT 3.0 was used to evaluate the mitochondrial protein-coding genes of *C. pitardii*, yielding a total of 667 candidate RNA editing sites. Because these loci were identified *in silico* and await experimental validation via RNA-seq or RT—PCR, they are strictly considered candidate sites.

These predicted sites encompass 14 distinct types of amino acid substitutions, with serine-to-leucine (S-to-L) transitions being the most prevalent. Hydropathy analysis indicated that several editing events likely mediate the conversion of hydrophilic residues to hydrophobic ones. This finding aligns with the general consensus that plant mitochondrial RNA editing frequently increases amino acid conservation and protein structural stability [72–74]. While these substitutions presumably preserve the functional integrity of mitochondrial respiratory chain complexes, their precise biological effects require further validation through protein structural modeling, transcriptomic profiling, and functional assays.

With respect to translation initiation, *in silico* predictions suggest that the start codons of *cox1* and *nad4L* undergo C-to-U editing, converting genomic ACG triplets to functional AUG codons in the mature transcripts. Such start codon reconstitution is a well-documented phenomenon in plant mitochondrial genomes and is essential for accurate translation initiation [75]. Although empirical transcriptomic evidence is currently lacking to definitively confirm these initiation corrections in *C. pitardii*, future validation can be readily achieved using RNA-seq alignments or RT—PCR coupled with Sanger sequencing. Furthermore, mitochondrial RNA editing has been increasingly implicated in plant stress responses. For instance, specific editing events in the *Arabidopsis nad3* transcript influence complex I functionality and subsequent stress adaptation [76]. We identified analogous candidate editing sites within the *C. pitardii nad3* homolog, highlighting these loci as promising targets for future investigations into the interplay between mitochondrial function and environmental stress tolerance in this species.

In summary, the substantial number of candidate RNA editing sites predicted within the *C. pitardii* mitochondrial genome suggests that its transcripts undergo extensive post-transcriptional modification. These findings establish a bioinformatics baseline for future empirical validation, mitochondrial gene expression profiling, and stress-response evaluations. Integrating multiomics approaches with protein

structural analyses and physiological stress assays will ultimately elucidate the definitive biological roles of these post-transcriptional modifications.

## 4.4 Comparative phylogenetic analysis of mitochondrial and chloroplast genomes

To determine the phylogenetic position of *C. pitardii* within the genus *Camellia*, we reconstructed phylogenetic trees using mitochondrial and chloroplast protein-coding genes (PCGs). While both organellar datasets confirmed the placement of *C. pitardii* within *Camellia*, topological incongruences emerged between the two phylogenies, indicating divergent evolutionary trajectories between the two organellar genomes.

On the basis of the mitochondrial PCG phylogeny, *C. pitardii* formed a highly supported clade (Bootstrap support [BS] = 74, Posterior probability [PP] = 1.00) with *C. drupifera*, *C. oleifera var. oleifera C. sinensis* var. *Sinensis* and *C. sinensis* var. *pubilimba*. Surprisingly, *C. chekiangoleosa*, a fellow member of sect. *Camellia*, failed to group as a sister to *C. pitardii* but instead fell into a distinct lineage. However, the low Bayesian posterior probability (PP = 0.324) for this divergence suggests that this topological placement lacks robust statistical support. Consequently, the monophyly of sect. *Camellia* and the precise evolutionary relationship between *C. chekiangoleosa* and *C. pitardii* cannot be definitively ascertained by relying solely on mitochondrial phylogenetics. Conversely, the chloroplast PCG tree placed *C. pitardii* in a subclade containing *C. chekiangoleosa*, *C. oleifera* var. *oleifera*, and *C. gigantocarpa*, revealing a stark topological contrast with the mitochondrial phylogeny.

This cytonuclear discordance highlights the intricate evolutionary history of *C. pitardii* and its close relatives. Such phylogenetic incongruence between mitochondrial and plastid signals in *Camellia* is frequently driven by factors such as introgressive hybridization, organellar capture, incomplete lineage sorting (ILS), or taxon sampling artifacts [77–79]. While plausible, determining the exact mechanisms underlying this is beyond the scope of the current organellar datasets.

Previous studies have demonstrated that traditional morphological sections within *Camellia* often fail to form monophyletic clades in molecular phylogenies [80]. Our findings corroborate this phenomenon, given the inconsistent placement of sect. *Camellia* species between the two organellar trees. Nevertheless, owing to limited taxon sampling and a strict reliance on organellar genomes, we cannot definitively reject the monophyly of sect. *Camellia* at this stage. Accurately delimiting the phylogenetic relationships within sect. *Camellia* and its allied taxa requires an integrative taxonomic approach incorporating broader sampling, nuclear genomic markers, and detailed morphological evidence. Ultimately, these organellar phylogenies provide a valuable baseline for deciphering the evolutionary framework of *C. pitardii*. The observed topological discordance highlights the reticulate evolutionary history of the genus and reinforces the paradigm that organellar gene trees do not strictly equate to species trees. Future studies integrating nuclear phylogenomics, expanded mitochondrial genome sampling, and population-level genetics are imperative to robustly elucidate the evolutionary history and phylogenetic relationships within the genus *Camellia*.

# 5. Conclusion

By integrating Illumina short-read and Nanopore long-read sequencing technologies, this study assembled, annotated, and comprehensively analyzed the mitochondrial genome of *C. pitardii*. The principal findings are as follows:

1. The *C. pitardii* mitogenome was assembled into a master circle conformation of 895,290 bp with a GC content of 45.68%. Genomic annotation revealed 38 protein-coding genes (PCGs), three rRNA genes, and 30 tRNA genes. Several tRNA genes are present in multiple copies, and pseudogenized paralogs were detected for *cox2*, *rps19*, and *sdh3*.

(2) Structural analysis revealed an extensive repeat landscape, including 260 simple sequence repeats (SSRs), 33 tandem repeats, and 620 pairs of dispersed repeats. Notably, three pairs of long repeats (two forward and one palindromic) were predicted to mediate homologous recombination, potentially generating eight alternative subgenomic conformations. While these findings highlight the structural complexity of the *C. pitardii* mitogenome, the *in vivo* presence and relative stoichiometry of these substructures require further experimental validation.

(3) Codon usage analysis revealed 28 highly preferred codons (RSCU > 1) in the mitochondrial PCGs, 92.86% of which ended in A or U, demonstrating a strong AT-ending bias. Furthermore, 667 putative RNA editing sites were predicted across the PCGs, encompassing 14 distinct types of amino acid substitutions, with the serine-to-leucine (S→L) transition being the most abundant. These data provide a valuable resource for future studies on mitochondrial post-transcriptional modifications and gene regulation.

(4) Phylogenetic reconstructions based on mitochondrial CDS and chloroplast PCG datasets confirmed the placement of *C. pitardii* within the genus *Camellia*. However, organellar discordance of topological placement was observed for certain taxa, notably *C. pitardii*, *C. chekiangoleosa*, and their close relatives.This incongruence underscores the complex evolutionary history of *Camellia* organelles and precludes definitive conclusions regarding the monophyly of section *Camellia* on the basis solely of organellar datasets. Comprehensive resolution of these relationships will require broader taxon sampling integrated with nuclear phylogenomics and morphological trait analysis.

Overall, this study provides foundational genomic resources for *C. pitardii* by elucidating its mitogenomic architecture, repeat landscape, codon preferences, RNA editing profiles, and phylogenetic context. These findings will facilitate future comparative organellar genomics and evolutionary studies within the genus *Camellia*, while supporting the conservation and utilization of *C. pitardii* germplasm resources.

# Declarations

**Ethics approval and consent to participate**

All the materials used in this study complied 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

**Author contributions**

Juyan Chen designed, wrote and revised this study. Rui Chen, He Li and Lang Huang participated in the collection and identification of plant material. All the authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this article. The authors alone are responsible for the content and writing of the paper.

**Funding**

This project was supported by the Basic Research of Guizhou Science and Technology - ZK [2024] General 620 and the research project of the director of the Forestry Science Institute of Guizhou Province(Guilin Kehe[2024]09) and the Key Laboratory of National Forestry and Grassland Administration on Biodiversity Conservation in Karst Mountainous Areas of Southwestern China.

**Data availability**

The raw transcriptome data generated and analyzed in this study were deposited in the SRA of the NCBI https://www.ncbi.nlm.nih.gov/under accession number PZ501402. The datasets analyzed during this study are included in this published article and its supplementary information files.

**Supplementary Information**

Supplementary data for this article can be found online at https://doi.

# References

1. XIONG D-w LIU, h, YIN G-l, y LIU. WU W-b, LI J-k: Discussion on the Protection and Utilization of Wild Germplasm Resourcesof *Camellia pitardii* Coh. St. of Panxi Area (in Chinese). Guizhou Forestry Sci Technol. 2019;47(01):11–5.
2. LI Y-f ZHANG, Y-x HUOD, WANG J. Analysis on Fruit Traits and Selection of Plus *Camellia pitardii* Indivuduals(in Chinese). J Zhejiang Sci Technol. 2018;38(01):1–5.

3. LI Y, HUO D, WANG J. Exploration of *Camellia pitardii* germplasm resources(in Chinese). Biotic Resour. 2018;40(04):377–81.
4. LIU X-l ZHANGM, ZHOU M. WANG D-p: Determination on total sterol in *Camellia pitardii* Coh.(in Chinese). FOOD Sci Technol. 2017;42(01):260–3.
5. LI Y-f LI, Z-j REN, H-d WANGJ. Study on the Fatty Acid Composition of the Seed Oil of *Camellia pitardii* Coh.Stuart(in Chinese). Guizhou Forestry Sci Technol. 2015;43(04):41–5.
6. LI F, HUO D, WANG J, XU J, ZHU Y, YANG B, XU J, LUO Y. Variation Analysis of Phenotypic Traits of Seed of Natural Populations of *Camellia pitardii* in Southwest China (in Chinese). Seed. 2021;40(04):109–14.
7. LI F, HUO D, WANG J. Variation of Phenotypic Traits of *Camellia pitardi* Flower(in Chinese). Guizhou Agricultural Sci. 2019;47(04):84–8.
8. YIN G-l WANG, X-q GUOC, ZHU M. A Preliminary Study of Cutting Seedling Technology of *Camellia pitardii* (in Chinese). J Sichuan Forestry Seience Technol. 2019;40(06):85–7.
9. Lv Z, Wang J, Zhou L, Zou SH, Ai LJ. Near telomere-to-telomere genome assembly of *Camellia pitardii*. Sci Data. 2025;12(1):9.
10. ZHU Y, HUO D, LUO Y, LI F, WANG J, LUO Z XUJ. Genetic Diversity Analysis of Wildly Southwest *Camellia pitardii* in Guizhou Plateau Based on ISSR Marker(in Chinese). Seed. 2020;39(11):80–2.
11. Ghifari AS, Saha S, Murcha MW. The biogenesis and regulation of the plant oxidative phosphorylation system. Plant Physiol. 2023;192(2):728–47.
12. Millar AH, Day DA, Whelan J. Mitochondrial biogenesis and function in Arabidopsis. Arabidopsis Book. 2008;6(1):e0111.
13. Eubel H, Jänsch L, Braun HP. New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. Plant Physiol. 2003;133(1):274–86.
14. Huang L, Jiang DZ, Yang YB, Chen R, Li Z, Zhou L, Xiao X. Report on the complete mitochondrial genome of the critically endangered and endemic Lauraceae plant *Syndiclis anlungensis* in China. Plant Gene. 2026;46:14.
15. Xiao X, Ran ZH, Yan C, Gu WH, Li Z. Mitochondrial genome assembly of the Chinese endemic species of *Camellia luteoflora* and revealing its repetitive sequence mediated recombination, codon preferences and MTPTs. BMC Plant Biol. 2025;25(1):18.
16. Jiang DZ, Zhou L, Ran ZH, Xiao X, Yang XH, Li Z. Assembly and analysis of the complete mitochondrial genome of an endemic *Camellia* species of China, *Camellia tachangensis*. BMC Genomics. 2025;26(1):18.
17. Sun MY, Zhang MY, Chen XN, Liu YY, Liu BB, Li JM, Wang RZ, Zhao KJ, Wu J. Rearrangement and domestication as drivers of Rosaceae mitogenome plasticity. BMC Biol. 2022;20(1):19.
18. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet. 2019;15(8):e1008373.

19. Arrieta-Montiel MP, Shedge V, Davila J, Christensen AC, Mackenzie SA. Diversity of the Arabidopsis Mitochondrial Genome Occurs via Nuclear-Controlled Recombination Activity. Genetics. 2009;183(4):1261–8.
20. Schenk JJ, Becklund LE, Carey SJ, Fabre PP. What is the modified CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. Appl Plant Sci. 2023;11(3):11.
21. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. Library construction for next-generation sequencing: Overviews and challenges. Biotechniques. 2014;56(2):61–.
22. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, Lawlor RT, Scarpa A. DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. PLoS ONE. 2013;8(6):8.
23. Desjardins P, Conklin D. NanoDrop Microvolume Quantitation of Nucleic Acids. JoVE 2010(45):2565.
24. Sauvage T, Cormier A, Delphine P. A comparison of Oxford nanopore library strategies for bacterial genomics. BMC Genomics. 2023;24(1):16.
25. Kozarewa I, Ning ZM, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G plus C)-biased genomes. Nat Methods. 2009;6(4):291–5.
26. Modi A, Stefania Vai, Caramelli D, Lari M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. Methods Mol Biol. 2021;2242:15–42.
27. Chen SF, Zhou YQ, Chen YR, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):884–90.
28. Wang YH, Zhao Y, Bollas A, Wang YR, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348–65.
29. 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(5):722–36.
30. 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):22.
31. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36:W5–9.
32. Lowe TM, Eddy SR. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 1997;25(5):955–64.
33. Rombel I, Sykes K, Rayner S, Johnston S. ORF-FINDER: a vector for high-throughput gene identification. Gene. 2002;282:33–41.
34. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.

35. Ran ZH, Li Z, Xiao X, An MT, 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(1):15.
36. Thiel T, Michalek W, Varshney RK, 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.
37. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
39. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
40. Zhang D, Gao FL, Jakovlic 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(1):348–55.
41. 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.
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;19:18.
43. Xiao Z, Gu YY, Zhou JQ, Lu MQ, Wang JF, Lu KZ, Zeng YL, Tan XF. 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):15.
44. Liang H, Qi HS, Chen JL, Wang YD, Liu MY, Sun XX, Wang CM, Xia TF, Feng XJ, Feng SL, et al. 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):19.
45. Li Z, Ran ZH, Xiao X, Yan C, Xu J, Tang M, An MT. Comparative analysis of the whole mitochondrial genomes of four species in sect. *Chrysantha* (*Camellia* L.), endemic taxa in China. BMC Plant Biol. 2024;24(1):18.
46. Li L, Li XR, Liu Y, Li JD, Zhen XY, Huang Y, Ye JH, Fan L. Comparative analysis of the complete mitogenomes of *Camellia sinensis* var. *sinensis* and C. *sinensis* var. *assamica* provide insights into evolution and phylogeny relationship. *Front Plant Sci* 2024, 15:16.
47. Li J, Tang H, Luo H, Tang J, Zhong N, Xiao LZ. Complete mitochondrial genome assembly and comparison of *Camellia sinensis* var. *Assamica* cv. *Duntsa*. *Front Plant Sci* 2023, 14:14.
48. Jiang M, Ni Y, Zhang JJ, Li JL, Liu C. Complete mitochondrial genome of *Mentha spicata* L. reveals multiple chromosomal configurations and RNA editing events. Int J Biol Macromol. 2023;251:12.
49. Feng LJ, Wang ZH, Wang CZ, Yang XM, An MM, Yin YL. Multichromosomal mitochondrial genome of *Punica granatum*: comparative evolutionary analysis and gene transformation from chloroplast

genomes. BMC Plant Biol. 2023;23(1):16.

50. Yang ZJ, Ni Y, Lin ZB, Yang LB, Chen GT, Nijiati N, Hu YZ, Chen XY. 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):12.
51. Lu C, Gao LZ, Zhang QJ. A High-Quality Genome Assembly of the Mitochondrial Genome of the Oil-Tea Tree *Camellia gigantocarpa* (Theaceae). Diversity-Basel. 2022;14(10):12.
52. 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.
53. Zhao D, Ye T, Gao FL, Jakovlic I, La Q, Tong YD, Liu X, Song R, Liu F, Lian ZM, et al. PhyloSuite v2: The development of an all-in-one, efficient and visualization-oriented suite for molecular dating analysis and other advanced features. iMeta. 2025;4(6):14.
54. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.
55. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564–77.
56. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–.
57. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32(1):268–74.
58. 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.
59. WANG Y, YU J, TANG H, ZHANG X. Research status and prospect of plant complex genomes and pan-genomes (in Chinese). Sci Sin Vitae. 2024;54(02):233–46.
60. Storchová H, Krüger M. Methods for assembling complex mitochondrial genomes in land plants. J Exp Bot. 2024;75(17):5169–74.
61. Tremblay-Belzile S, Lepage E, Zampini E, Brisson N. Short-range inversions: Rethinking organelle genome stability Template switching events during DNA replication destabilize organelle genomes. BioEssays. 2015;37(10):1086–94.
62. Christensen AC. Plant Mitochondrial Genome Evolution Can Be Explained by DNA Repair Mechanisms. Genome Biol Evol. 2013;5(6):1079–86.
63. Llorente B, Smith CE, Symington LS. Break-induced replication - What is it and what is it for? Cell Cycle. 2008;7(7):859–64.
64. 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.

65. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 2011;12(1):32–42.
66. Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. Mol Cell. 2015;59(2):149–61.
67. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. Mol Biol Rep. 2009;36(8):2039–46.
68. Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A, Seeger S, Schöttler MA, Ruf S, Bock R. The Contributions of Wobbling and Superwobbling to the Reading of the Genetic Code. PLoS Genet. 2012;8(11):16.
69. Schneider A, Maréchal-Drouard L. Mitochondrial tRNA import: are there distinct mechanisms? Trends Cell Biol. 2000;10(12):509–13.
70. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. Plant MolBiol. 2018;97(3):215–31.
71. 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.
72. Yura K, Go M. Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles. BMC Plant Biol. 2008;8:11.
73. Islas-Osuna MA, Silva-Moreno B, Caceres-Carrizosa N, García-Robles JM, Sotelo-Mundo RR, Yepiz-Plascencia GM. Editing of the grapevine mitochondrial cytochrome b mRNA and molecular modeling of the protein. Biochimie. 2006;88(5):431–5.
74. Giegé P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences* 1999, 96(26):15324–15329.
75. Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA Editing in Plants and Its Evolution. In: *Annual Review of Genetics, Vol 47.* Edited by Bassler BL, Lichten M, Schupbach G, vol. 47. Palo Alto: Annual Reviews; 2013: 335–352.
76. Yuan H, Liu D. Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in *Arabidopsis*. Plant J. 2012;70(3):432–44.
77. Vijayan K, Zhang WJ, Tsou CH. Molecular taxonomy of *Camellia* Theaceae inferred from nrITS sequences. Am J Bot. 2009;96(7):1348–60.
78. Wendel JF, Doyle JJ. Phylogenetic Incongruence: Window into Genome History and Molecular Evolution. In: *Molecular Systematics of Plants II: DNA Sequencing.* Edited by Soltis DE, Soltis PS, Doyle JJ. Boston, MA: Springer US; 1998: 265–296.
79. Rieseberg L. Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends Plants 1991, 5.
80. Zan T, He YT, Zhang M, Yonezawa T, Ma H, Zhao QM, Zhang WJ, Huang CH, Kuo WY. Phylogenomic analyses of *Camellia* support reticulate evolution among major clades. Mol Phylogenet Evol.

2023;182:13.

# Figures

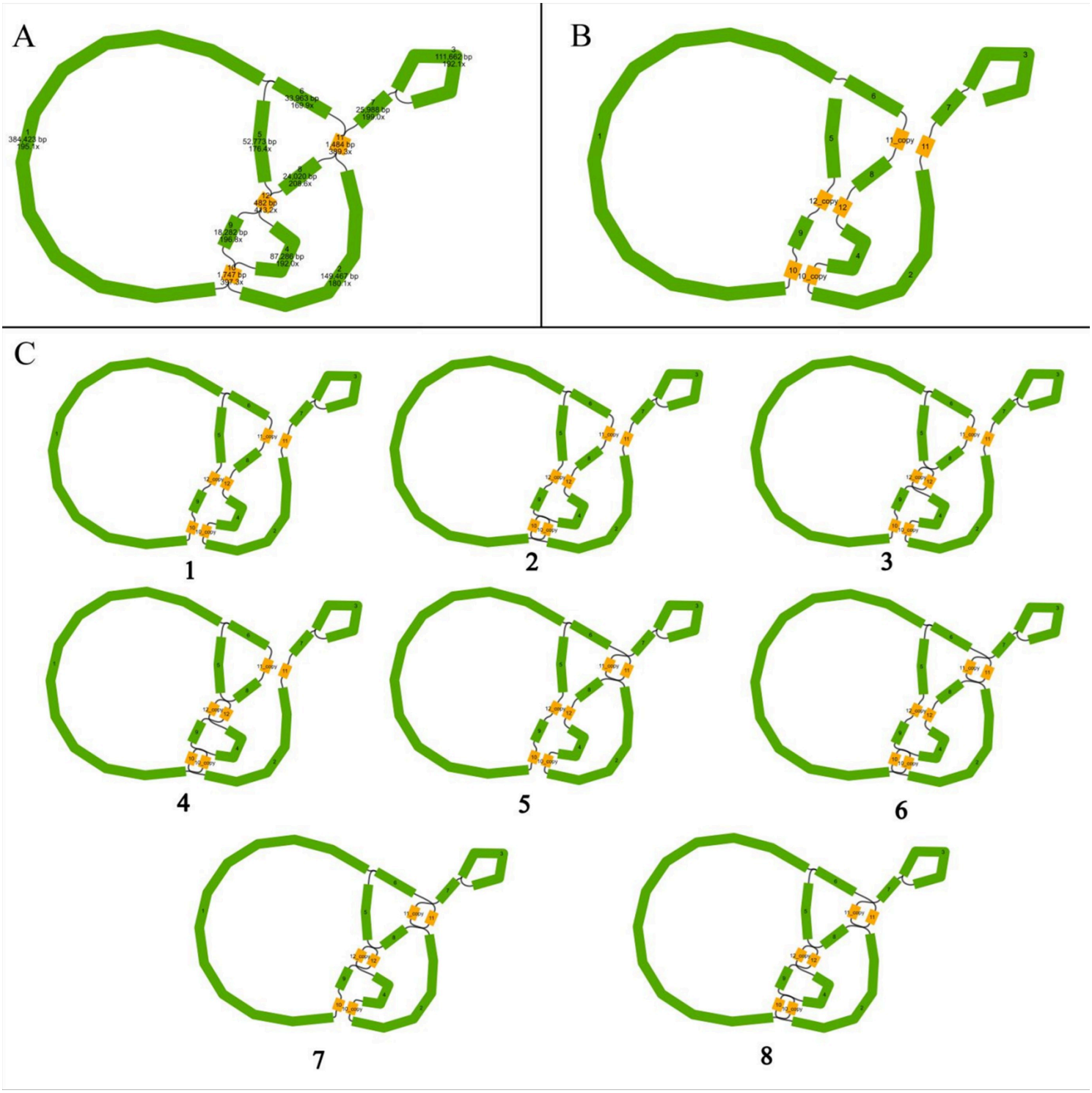



**Figure 1**

Assembly graph of the *C. pitardii* mitochondrial genome. Yellow rectangles indicate repeat sequences located at the branching nodes. (A) The initial branched configuration of the genome. (B) The complete

linearized sequence following the resolution of genomic branches. (C) Eight subgenomic structures mediated by two pairs of repeat sequences within the genome.

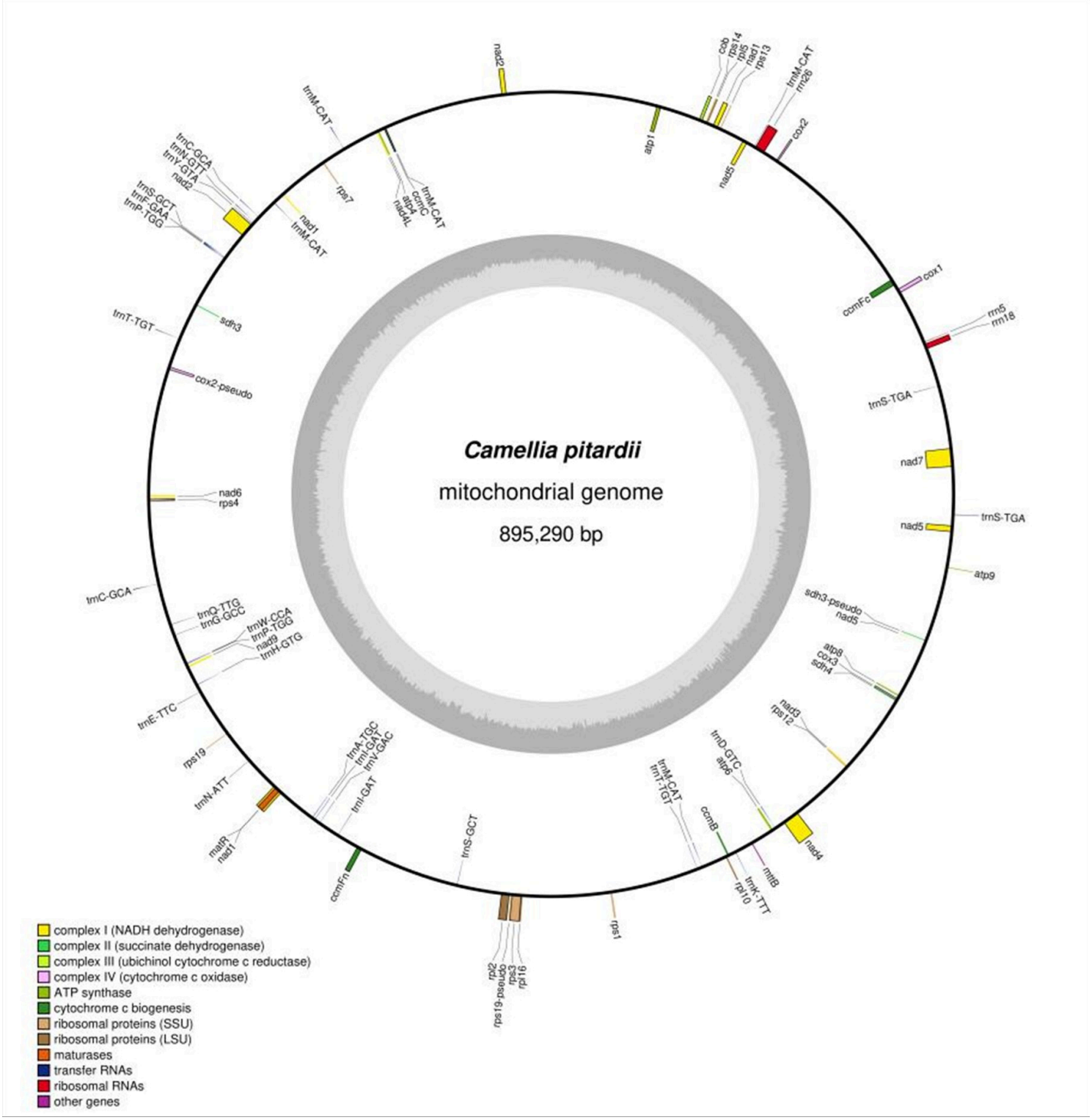
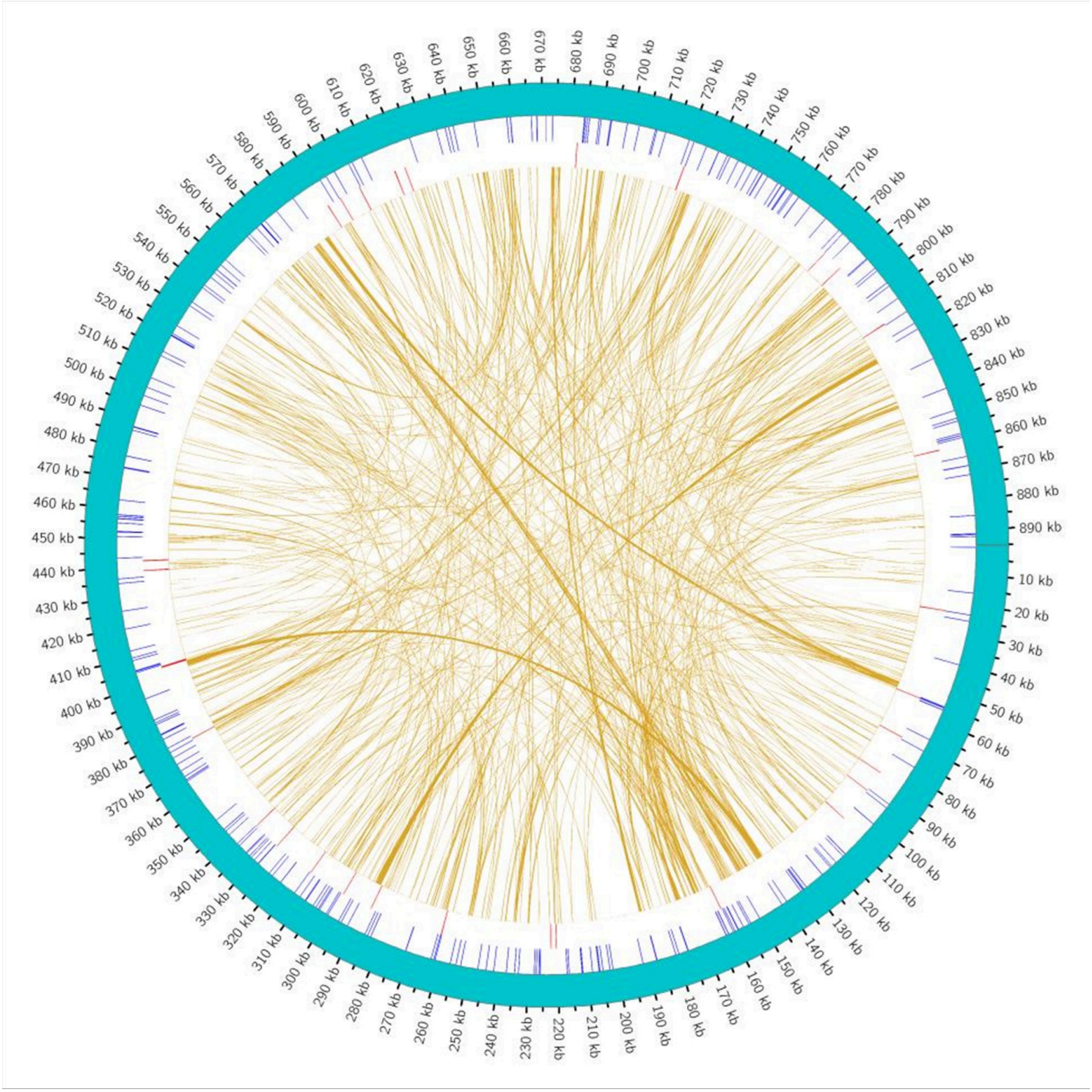



**Figure 2**

Circular map of the *C. pitardii* mitochondrial genome.




## Figure 3

Distribution of repeat sequences in the *C. pitardii* mitochondrial genome. The outer ring (turquoise) represents the mitochondrial DNA sequence. Tick marks on the inner rings indicate the locations of simple sequence repeats (SSRs, blue) and tandem repeats (red). The central ribbons (brown) denote dispersed repeats.

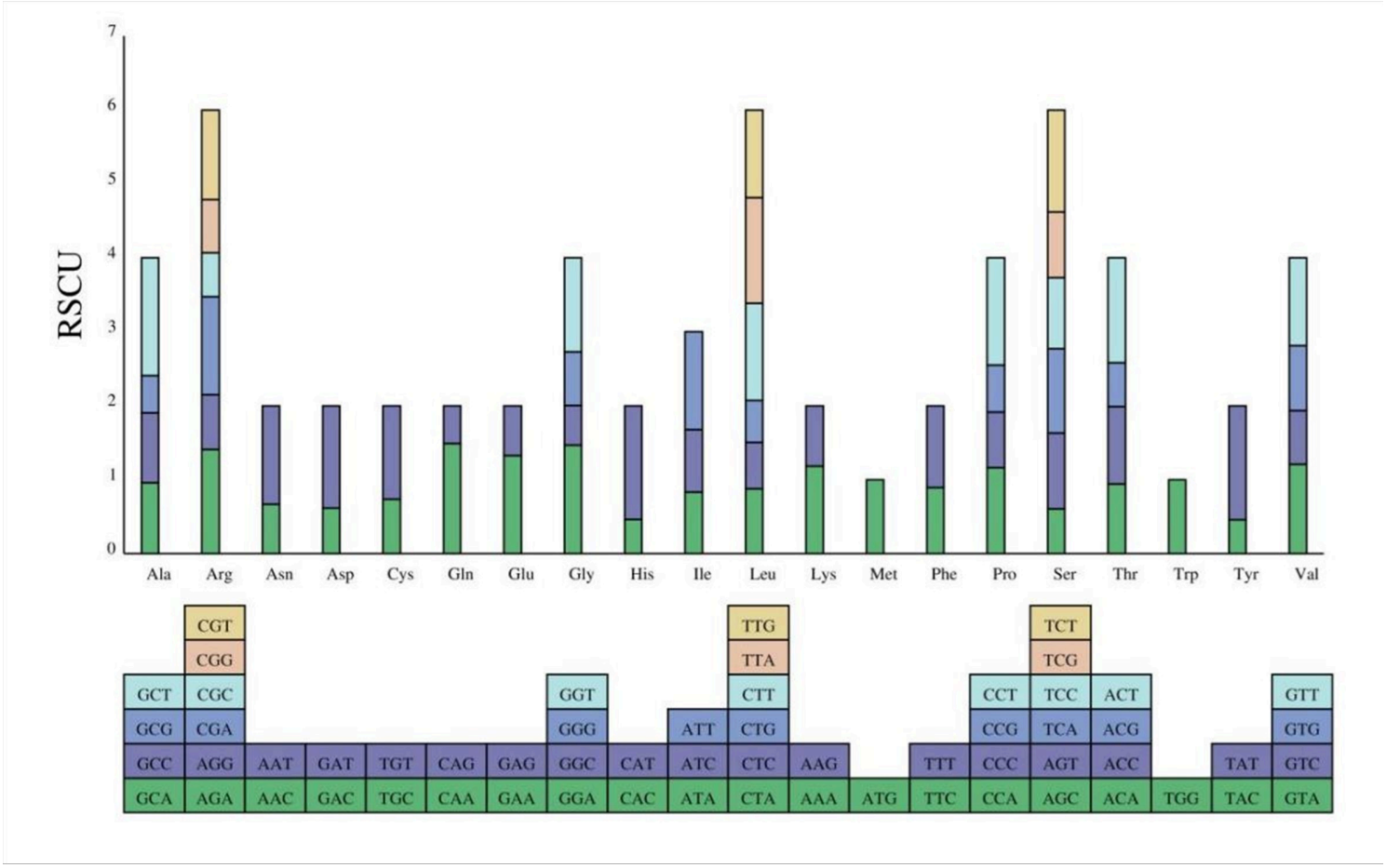



**Figure 4**

Relative synonymous codon usage (RSCU) in the mitochondrial genome of *C. pitardii*

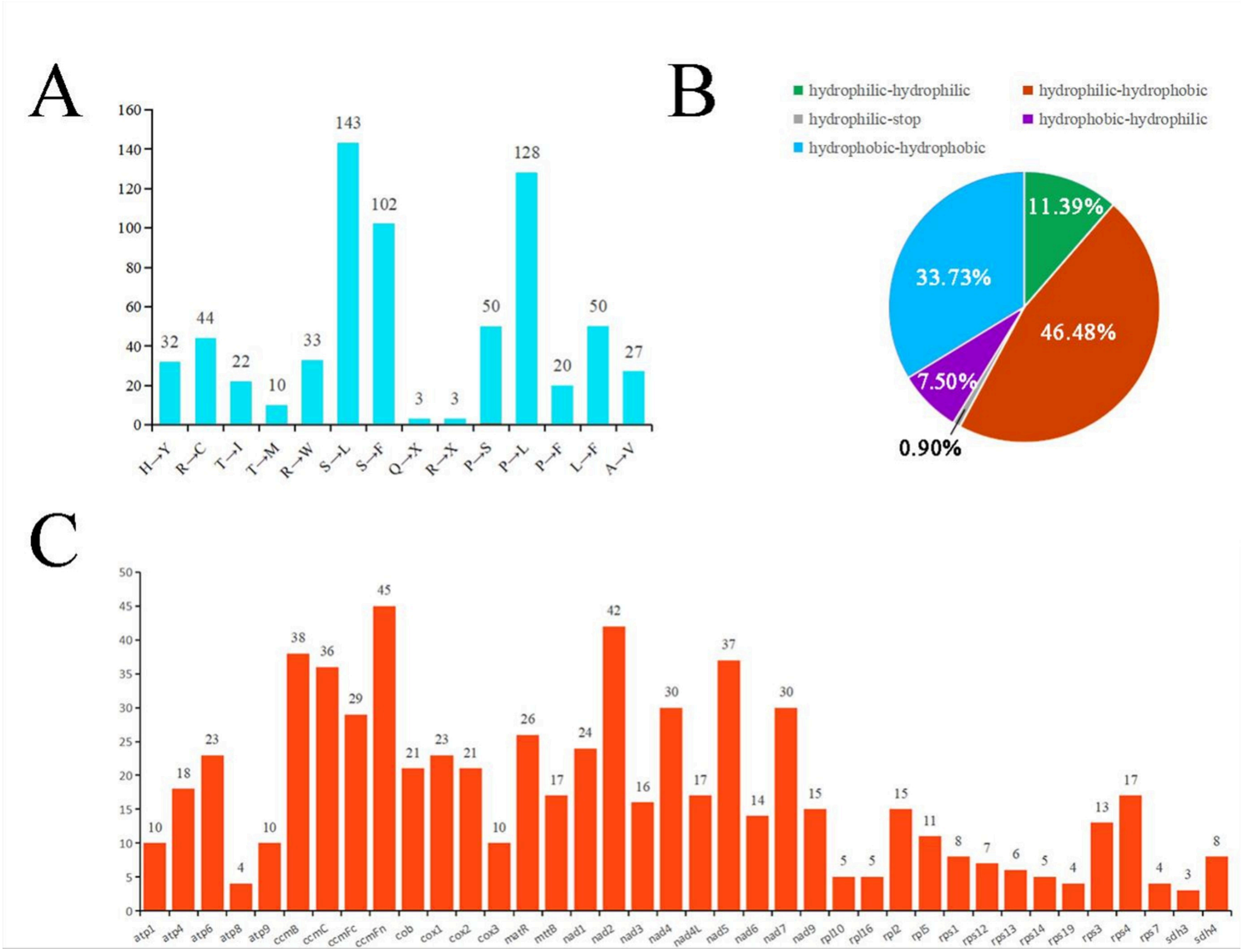



**Figure 5**

Predicted RNA editing sites in the *C. pitardii* mitogenome. (A) Characteristics of the RNA editing sites. (B) Relative proportions of different RNA editing types. (C) Number of RNA editing sites across the mitochondrial DNA.

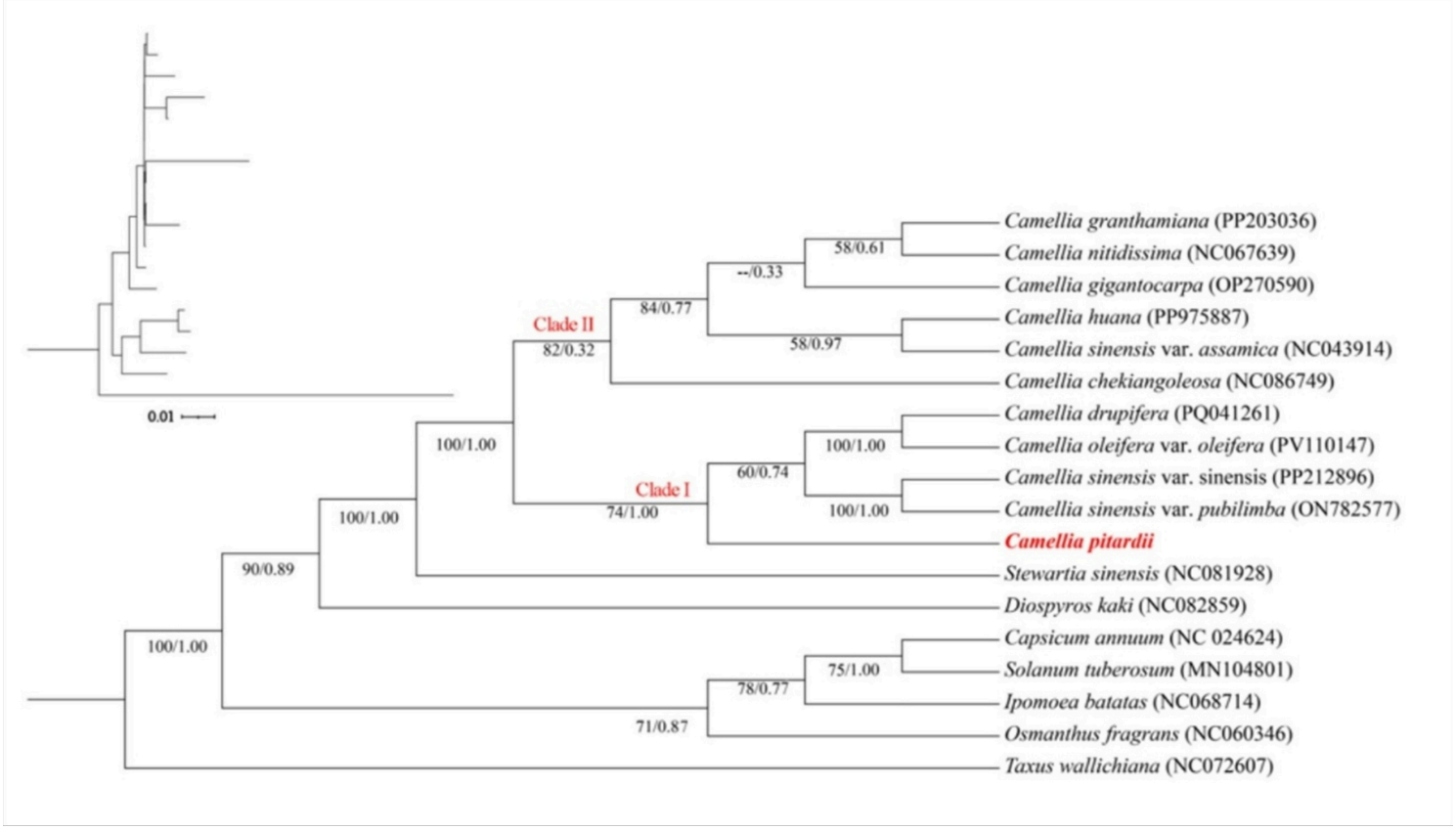



**Figure 6**

Phylogenetic tree inferred from mitochondrial protein-coding gene (PCG) sequences. The topology represents both Maximum Likelihood (ML) and Bayesian Inference (BI) analyses. Values above the branches indicate ML bootstrap support and BI posterior probability (BS/PP), respectively.

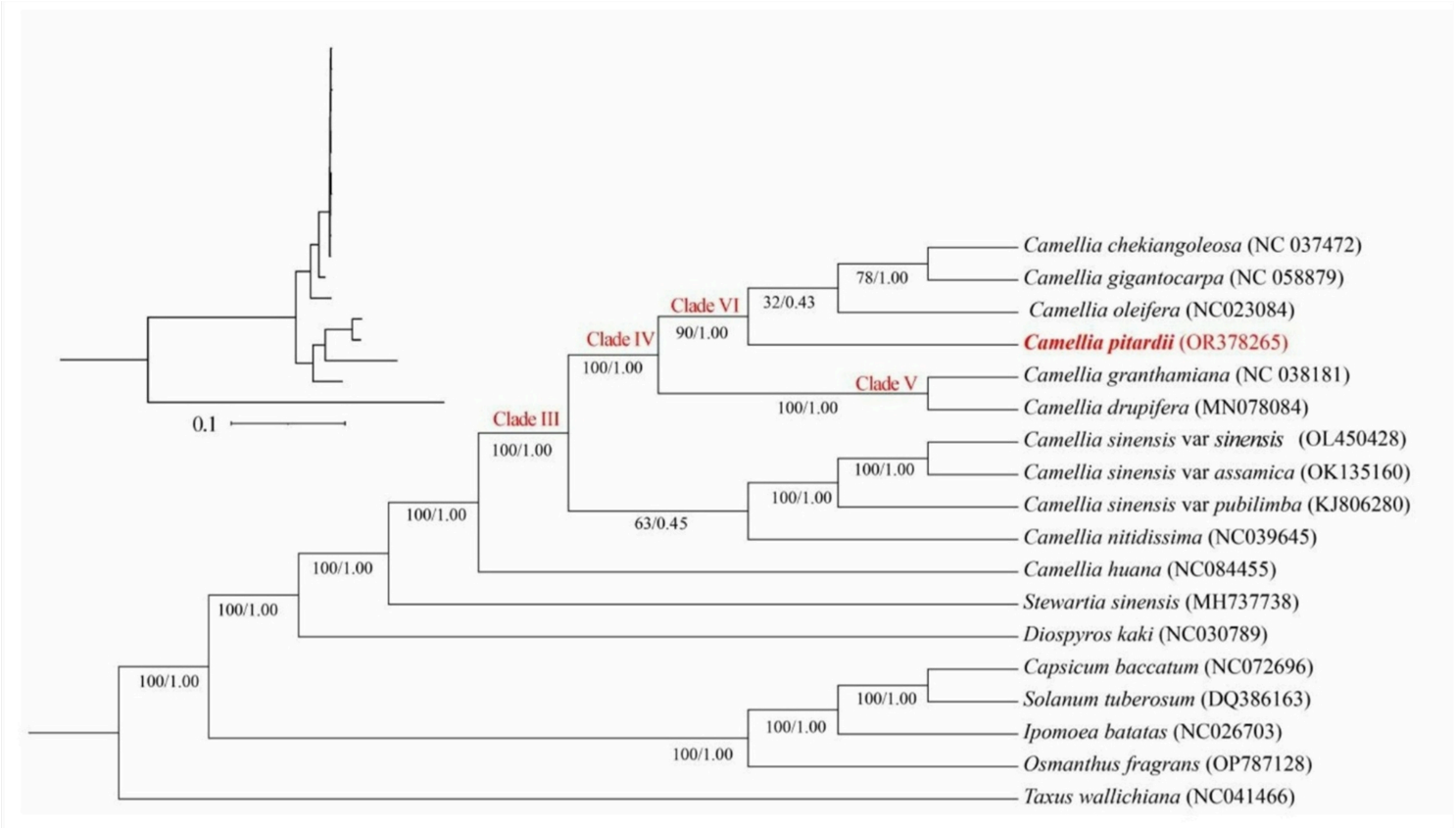



**Figure 7**

Phylogenetic tree inferred from chloroplast protein-coding genes (PCGs). The topology represents both Maximum Likelihood (ML) and Bayesian Inference (BI) analyses. Values above the branches indicate ML bootstrap support and BI posterior probability (BS/PP), respectively.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.Supplementarydata.docx