


RESEARCH

Open Access



Repeat-mediated recombination results in Complex DNA structure of the mitochondrial genome of *Trachelospermum jasminoides*

Yisha Cai^{1,2}, Haimei Chen², Yang Ni², Jingling Li², Jinghong Zhang^{1*}  and Chang Liu^{2*} 

Abstract

Background *Trachelospermum jasminoides* has medicinal and ornamental value and is widely distributed in China. Although the chloroplast genome has been documented, the mitochondrial genome has not yet been studied.

Results The mitochondrial genome of *T. jasminoides* was assembled and functionally annotated using Illumina and nanopore reads. The mitochondrial genome comprises a master circular molecular structure of 605,764 bp and encodes 65 genes: 39 protein-coding genes, 23 transfer RNA (tRNA) genes and 3 ribosomal RNA genes. In addition to the single circular conformation, we found many alternative conformations of the *T. jasminoides* mitochondrial genome mediated by 42 repetitive sequences. Six repetitive sequences (DRS01–DRS06) were supported by nanopore long reads, polymerase chain reaction (PCR) amplifications, and Sanger sequencing of the PCR products. Eleven homologous fragments were identified by comparing the mitochondrial and chloroplast genome sequences, including three complete tRNA genes. Moreover, 531 edited RNA sites were identified in the protein-coding sequences based on RNA sequencing data, with *nad4* having the highest number of sites (54).

Conclusion To our knowledge, this is the first description of the mitochondrial genome of *T. jasminoides*. Our results demonstrate the existence of multiple conformations. These findings lay a foundation for understanding the genetics and evolutionary dynamics of Apocynaceae.

Keywords *Trachelospermum jasminoides*, Mitochondrial genome, MTPT, Recombination, RNA editing events

*Correspondence:

Jinghong Zhang

zjh@hqu.edu.cn

Chang Liu

cliu6688@hotmail.com; cliu@implad.ac.cn

¹School of Medicine, Huaqiao University, Fujian 362021, China

²Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, PR China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Background

Trachelospermum jasminoides, an evergreen liana of Apocynaceae, is widely distributed in the eastern, southern, and southeastern regions of China [1]. In traditional Chinese medicine, it has been used to treat inflammation-related diseases such as rheumatic arthralgia and traumatic injuries [1, 2]. Various secondary metabolites of *T. jasminoides*, including lignans [3], flavonoids [4], terpenoids [5], and alkaloids [6], have anti-inflammatory, antioxidant, analgesic, antitumor, and antibacterial activities, indicating the potential for the use of *T. jasminoides* in modern medicine.

High-throughput sequencing technology greatly facilitates the study of plant genome structure, which is crucial for exploring plant secondary metabolites and understanding plant genetic variation at the molecular level. Thus, numerous plant chloroplast genomes, such as *Brassica napus* [7], *Gelsemium elegans* [8], and *Glycine max* [9] have been reported. The chloroplast genome of *T. jasminoides* has also been reported [10]. However, owing to the complexity of plant mitochondrial genomes, only 11 complete mitochondrial genomes of Gentianales species have been reported and deposited in the NCBI nucleotide database ((July 15, 2023), namely, six Rubiaceae, three Apocynaceae, and two Gentianaceae species. To date, the mitochondrial genome sequence of *T. jasminoides* has not been reported, severely restricting follow-up research.

Mitochondria are semi-autonomous organelles of endosymbiotic origin derived from early alpha-archaeal bacteria [11]. Mitochondria are signalling organelles that regulate a wide variety of cellular functions and can dictate cell fate, such as generating ATP, releasing cytochrome c and activating transcription factors [12]. Mitochondria contain their own DNA (mitochondrial DNA) but are still regulated by the nuclear genome. Given that simple repetitive sequences in plant mitochondrial genomes have a high rate of evolution, they are widely used as molecular markers in modern molecular biology research [13]. RNA editing is common in plant mitochondria [14], and RNA editing events most often occur at the first or second codon positions in mitochondrial genomes. Therefore, they mainly affect the non-synonymous positions of protein-coding regions [15–17], changing the resulting amino acid sequences [18]. Furthermore, the plant mitochondrial genome is complex and contains various rearrangements [19, 20]. The plant mitochondrial genome is large, structurally complex, and genetically diverse. The structure of the mitochondrial genome is intricate in contrast to the master circular structure. It may involve a single or multiple chromosomes, each capable of exhibiting either a linear or circular architecture. However, some DNA repair mechanisms protect the stability of the mitochondrial

genome, especially base excision repair and homologous recombination (HR) [21]. Additionally, the process of genome evolution may lead to the horizontal migration of genes [22]. This horizontal migration of genes among chloroplast, mitochondrial, and nuclear DNA has been reported in some Apocynaceae species. The mitochondrial genes *sdh3* and *rps14* were transferred to the nucleus in *Rhazya stricta*, with *rps14* undergoing a single transfer to the nucleus followed by a duplication event [23]. The chloroplast gene *rpoC2* was transferred to the mitochondria in *Asclepias syriaca*, and the mitochondrial DNA insert *rpl2* migrated from the mitochondrial genome to the plastome of *A. syriaca* [24].

In the present study, we sequenced and assembled the complete mitochondrial and chloroplast genomes of *T. jasminoides*, deciphered the possible structure of *T. jasminoides* mitochondrial genome, and performed a series of comparative analyses of repetitive sequences, phylogenetic relationships, HR, gene migration in organelles, and RNA editing events. The results show that multiple mechanisms lead to the diversity of the *T. jasminoides* mitochondrial genome. This study provides an important theoretical foundation for understanding the evolution of *T. jasminoides* organelle genomes and genetic variations in medicinal plants.

Materials and methods

Plant materials and DNA and RNA extraction and sequencing

Fresh leaves of *T. jasminoides* were collected from the Medicinal Botanical Garden of the Institute of Medicinal Plant Development, Beijing, China (40°2'33.239" N, 116°16'35.867" E). The plant species was identified as *T. jasminoides* by Professor Jinghong Zhang of the School of Medicine, Huaqiao University. The total DNA and RNA were extracted using the Plant Genomic DNA Kit and the RNeasy Pure Plant Plus Kit (Qiagen Biotech, Co., Ltd., Beijing, China), respectively, and stored in a refrigerator at –80 °C. Total DNA was divided into two separate parts for different sequencing approaches. One part underwent library construction using the NEBNext DNA Library Prep Kit and sequenced with a 2500 platform (Illumina, San Diego, CA, USA). The other part, with an insert size of 10 kb, was prepared and sequenced with the PromethION sequencer (Oxford Nanopore Technologies, USA). The strand-specific RNA sequencing (RNA-seq) library was prepared according to ribosomal RNA (rRNA) depletion and stranded sequencing methods. First, rRNA was depleted from total RNA using an rRNA Removal Kit, following the manufacturer's instructions. Subsequently, a TruSeq Stranded mRNA Library Prep kit (Illumina) was used to construct a sequencing library that involved RNA fragmentation, first-strand complementary DNA (cDNA) synthesis, RNA

degradation, second-strand cDNA synthesis, blunt-end conversion, adenylation and adaptor ligation, polymerase chain reaction (PCR) enrichment, and library purification. The quantified library was subjected to Illumina PE150 sequencing.

Genome assembly and annotation

In the present study, the chloroplast genome was assembled using GetOrganelle software [25] with the parameters “-R 15 -k 21,45,65,85,105 -F embplant_pt”. The default reference genomes provided in GetOrganelle software were used to assemble the chloroplast genome. The process of assembling the mitochondrial genome involved a hybrid assembly strategy to mitigate potential false positives arising from nuclear and plastid DNA with mitochondrial origin [26]. First, mitochondria-derived Illumina short reads were obtained using GetOrganelle [25] with the parameters “-R 20 -k 21,45,65,85,105 -P 1000000 -F embplant_mt”. Second, the short reads were assembled using SPAdes software [27], which is embedded in Unicycler software [28]. The assembly resulted in a unitig graph. Third, Bandage [29], which can manually remove chloroplast- and nuclear-derived nodes, was used to visualize the unitig graph. Finally, the double bifurcating structures (DBSs) in the unitig graph were analyzed with nanopore long reads using Unicycler software [28].

The chloroplast genome was annotated by CPGAVAS2 [30], and the annotation was viewed by the CPGView web server tool [31]. Meanwhile, the mitochondrial genome was annotated using GeSeq [32] and IPMGA (<http://www.1kmpg.cn/mgavas/>). The annotation results were manually corrected using Apollo software [33]. The structures of the mitochondrial genome were plotted using PMGmap [34]. Finally, the organelle genome sequences and annotations were submitted to GenBank with accession numbers OR286427.1 for the chloroplast genome and OR333986.1 for the mitochondrial genome.

Identification of repetitive sequence

Three types of repetitive sequence, namely, simple sequence repeats (SSRs), tandem repeat sequences (TRSs), and dispersed repeat sequences (DRSs), were analyzed in the mitochondrial genome of *T. jasminoides*. SSRs were detected using MISA [35] with the following thresholds: 10 units for mononucleotides, 5 units for dinucleotides, 4 units for trinucleotides and 3 units for tetra-, penta- and hexa-nucleotides. TRSs were predicted using Tandem Repeats Finder [36] with the default settings of two for matches and seven for mismatches and indels. The minimum alignment score and maximum period size were set to 50 and 500, respectively. DRSs were analyzed using ROUSFinder2.0.py [37] with default parameters.

Identification and validation of repeat-mediated recombination

To identify recombination mediated by repeat sequences, we detected the DRSs of the mitochondrial genome using ROUSFinder2.0.py [37]. Then, we extracted each pair of repetitive sequences, as well as the 500 bp long flanking sequences on both sides, corresponding to one conformation. Subsequently, we recombined the two sequences in silico to generate sequences corresponding to the other recombined conformations. We then mapped the nanopore long reads for the four DNA sequences of the two conformations and counted the number of mapped reads using BWA [38] and Samtools [39]. Finally, the mapping results of the reads were visualized using IGV software [40].

Although reversing the position of the reverse primers could yield repeat sequence-mediated recombination products, we designed the optimal PCR primers to verify these hypothetical recombination products using Primer BLAST. These primer sequences are listed in Table S1. The PCR amplification was performed using approximately 1 µl DNA, 1 µl 10 µM each of the forward and reverse primers, 25 µl 2× Easy Taq SuperMix, and 22 µl ddH₂O for PCR under the following conditions: 94 °C for 2 min; 35 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 1 min; followed by a final extension at 72 °C for 2 min. PCR products were checked using 1% agarose gel electrophoresis. DNA markers ranging from 100 bp to 2000 bp (Sangon Biotech, Co., Ltd., China) were used to determine the size of the PCR products. Finally, the PCR products were further sequenced using the Sanger method, and the chromatograms were visualized using Seqman software [41].

Identification of mitochondrial plastid sequences

In the present study, mitochondrial plastid sequences (MTPTs) were identified by comparing the chloroplast and mitochondrial genomes using BLASTN software tool [42] with default parameters. BLASTN hits of >100 bp were retained. The MTPT gene cluster was annotated according to the annotation result of the *T. jasminoides* mitochondrial genome. To confirm the presence of MTPT sequences, we extracted the MTPT sequence along with its 1000 bp upstream and downstream sequences regions to create a reference sequence. We then aligned the nanopore long reads to these reference sequences using BWA software [39] with default parameters. Finally, the mapping results of nanopore long reads to MTPT regions were visualized using the IGV software [40]. In addition, we extracted MTPTs of the organelle genomes from 12 Gentianales species (*T. jasminoides*, *A. syriaca*, *R. stricta*, *Scyphiphora hydrophyllacea*, *Psychotria viridis*, *Psychotria serpens*, *Coffea arabica*, *Gentiana crassicaulis*, *Gentiana straminea*, *Neolamarckia*

cadamba, *Hoya lithophytica* and *Damnacanthus indicus*) and performed cluster analyses using CD-HIT software [43] with a parameter of 0.85. Finally, the MTPT genome map was plotted using Tbtools software [44].

Construction of phylogenetic tree

In our phylogenetic analysis, we employed a dataset consisting of 12 mitochondrial genomes from Gentianales species. *Salvia officinalis* (OQ001564.1-OQ001565.1), which belongs to Lamiales, was set as the outgroup. All mitochondrial genomes were obtained from the NCBI GenBank database with the following accession numbers: *T. jasminoides* (OR333986.1), *A. syriaca* (NC_022796.1), *R. stricta* (NC_024293.1), *S. hydrophyl-lacea* (NC_057654.1), *P. viridis* (NC_066984.1), *P. serpens* (NC_069806.1), *C. arabica* (OL789880.1-OL789881.1), *G. crassicaulis* (OM320814.1), *G. straminea* (OM328072.1), *N. cadamba* (MT320890.1-MT364442.1), *H. lithophytica* (MW719051.1), and *D. indicus* (MZ285076.1). For our phylogenetic analysis, we initially extracted the common coding sequence (CDS) of protein-coding genes (PCGs) from these species using PhyloSuite software [45]. Subsequently, the CDSs sequences were aligned using MAFFT software [46]. The resulting aligned sequences were concatenated and used to construct a phylogenetic tree using IQTREE2 software [47] with the maximum-likelihood method. Bootstrap analysis was assessed using UFBoot with 1,000 replicates [48]. Finally, the resulting phylogenetic tree was visualized using iTOL software [49].

Prediction and validation of RNA editing sites within PCGs of *T. jasminoides* mitochondrial genome

First, to identify RNA editing sites in the *T. jasminoides* mitochondrial genome, we mapped the strand-specific RNA-seq data to the sequences of 38 PCGs and their 5' and 3' flanking regions by 100 bp using BWA [38] and extracted alignment reads that match the reference sequence using Samtools [39]. Then, the RNA editing sites were predicted by REDIttools software [50] with the following parameters: coverage \geq 5 and frequency \geq 0.1 [51]. To enhance the accuracy of our predictions, we further visualized the RNA editing sites using the IGV software [40], focusing on sites with a frequency \geq 0.2. To eliminate the influence of single nucleotide polymorphism (SNP) sites on the results, we performed SNP prediction and excluded RNA editing sites that overlapped with SNP results. In parallel, for the identification of SNP in the *T. jasminoides* mitogenome, we mapped Illumina short reads of whole genome sequencing to the CDS sequences of the PCGs, along with their 5' and 3' flanking regions by 100 base pairs using BWA (Li and Durbin, 2009) and extracted alignment reads that match the reference sequence using Samtools (Li et al., 2009). Subsequently, SNP sites were determined using the REDIttools

software (Picardi and Pesole, 2013) with parameters requiring coverage \geq 5 and frequency \geq 0.1. Finally, predicted RNA editing sites that overlapped with the SNP sites were removed from subsequent analyses.

To validate RNA editing sites within the *T. jasminoides* mitochondrial genome, we randomly selected ten PCGs for PCR amplification. The PCR primers for all PCGs was designed using Primer BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome); the details are provided in Table S3. Subsequently, we extracted both total DNA and RNA from the same sample as described in section "Plant Materials and DNA and RNA Extraction and Sequencing". Total RNA was reverse transcribed into cDNA using the TransScript First-Strand cDNA Synthesis SuperMix kit. Total DNA and cDNA were then separately utilized for PCR amplification, following the experimental conditions specified in section "Identification and Validation of Repeat-mediated Recombination". Finally, the resulting PCR products were sequenced using the Sanger method, and the chromatograms were visualized using the Seqman software [41].

Results

Graph-based mitochondrial genome and general genomic features

We employed a hybrid assembly approach to construct the mitochondrial genome of *T. jasminoides*, using nanopore long and Illumina short reads obtained through whole genome sequencing. Initially, we acquired approximately 10 Gb of Illumina short reads and nanopore long reads (Table S2). The read length N50 for Illumina short reads and nanopore long reads were 150 and 23,739 bp, respectively. We then utilized the GetOrganelle software to extract and elongate the short reads derived from the mitochondrial genome. Subsequently, we employed SPAdes software, integrated within the Unicyclic software, to assemble them into a unitig graph (Fig. 1A). The unitig graph contained 18 contigs, including six DBSs. The longest contig, cg2, was 170,035 bp, while the shortest, cg6, was only 65 bp (Table 1). The nanopore long reads were then used to resolve the DBSs using Unicyclic software. Finally, a circular molecule was independently obtained and visualized using Bandage software (Fig. 1B). In the following description, we refer to this master circular molecule as "MC," which has 605,764 bp and a GC content of 44.65% (Fig. 2). It was closely related in size to the mitochondrial genome of *R. stricta* (Genome size: 548608 bp, NCBI Accession Number: NC_024293.1), *A. syriaca* (Genome size: 682498 bp, NCBI Accession Number: NC_022796.1) and *C. auriculatum* (Genome size: 426495 bp, NCBI Accession Number: NC_041494.1). To assess the accuracy of the assembly, both Illumina short reads and nanopore long reads were individually mapped

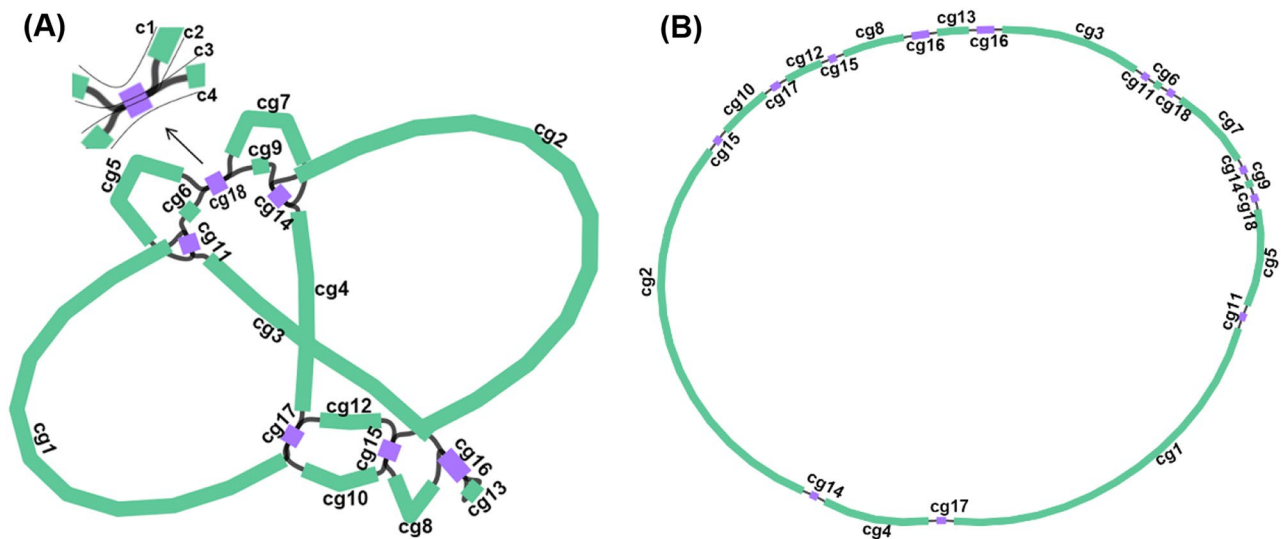


Fig. 1 The unitig graph of the *T. jasminoides* mitochondrial genome. **(A)** The green and purple contigs belongs to the master circular molecular (MC), and the purple contigs is double bifurcation structures (DBS). Each DBS has four conformations (c1, c2, c3, and c4), as shown in the top left corner with cg18 as an example. **(B)** The graph is presented after resolving the DBSs using long reads

Table 1 Attribution of assemble of contigs in the mitogenome of *T. jasminoides*

No.	Name	Length (bp)	Depth for Illumina short reads	Depth for Nanopore long reads	Bifurcation structures
1	cg1	152,646	946	470	No
2	cg2	170,035	1,250	758	No
3	cg3	57,069	1,155	644	No
4	cg4	45,415	1,217	711	No
5	cg5	40,517	1,058	693	No
6	cg6	65	948	406	No
7	cg7	34,448	1,064	542	No
8	cg8	25,683	919	423	No
9	cg9	2,678	1,149	527	No
10	cg10	22,286	1,063	607	No
11	cg11/DBS1	518	1,952	1,026	Yes
12	cg12	16,491	940	443	No
13	cg13	13,567	1,025	428	No
14	cg14/DBS2	266	2,017	926	Yes
15	cg15/DBS3	129	1,853	798	Yes
16	cg16/DBS4	7,460	2,395	1,044	Yes
17	cg17/DBS5	3,931	1,833	891	Yes
18	cg18/DBS6	128	1,993	938	Yes

DBS: double bifurcating structures

to the MC. The correctness of the assembly was then established by calculating the sequencing depth at each site. The sequencing depth across all genomic locations remained consistently with no observable gaps. In addition, the average depth of coverage was $>500\times$ (Figures S1A and Figure S1B).

The mitochondrial genome of *T. jasminoides* was annotated to contain a total of 65 genes (62 unique genes),

including 39 PCGs (38 unique genes), 23 tRNA genes (21 unique genes), and 3 rRNA genes. The *atp9* gene was duplicated, and the two copies have 100% identity. The PCGs were further categorized into 10 functional groups (Table 2). All the core genes were found within the *T. jasminoides* mitochondrial genome. We also identified three categories of variable genes, which encompassed nine small subunit ribosome genes (*rps1*, *rps3*, *rps4*, *rps7*, *rps10*, *rps12*, *rps13*, *rps19*), four large subunit ribosome genes (*rpl10*, *rpl16*, *rpl2*, *rpl5*), and two subunit succinate dehydrogenase genes (*sdh3* and *sdh4*) (Fig. 2; Table 2).

Repetitive sequence analysis

SSRs, also known as microsatellites or short tandem repeats, are repetitive DNA sequences that typically consist of one to six nucleotides [52]. These sequences are widely used as genetic markers in fields, such as forensics, population genetics, and disease diagnosis [53]. In this study, we employed the MISA online platform [35] to identify SSRs within the mitochondrial genome of *T. jasminoides*. A total of 57 SSRs were detected in the mitochondrial genome, including 39 monomers, 5 dimers and trimers, 6 tetramers, and 2 pentamers (Table S4).

TRSs are defined as a series of nucleotide sequences that repeat consecutively and are located adjacent to each other on a DNA molecule [54, 55]. These sequences can play important roles in gene regulation, chromosome stability, and genome evolution [56]. In the study, TRSs were predicted in the mitochondrial genome of *T. jasminoides* using Tandem repeats Finder program [36]. Based on these results, eight TRSs were predicted (Table S5). TRS3 had the longest repeat unit at 24 bp, whereas TRS1 had the shortest repeat unit of only two bp.

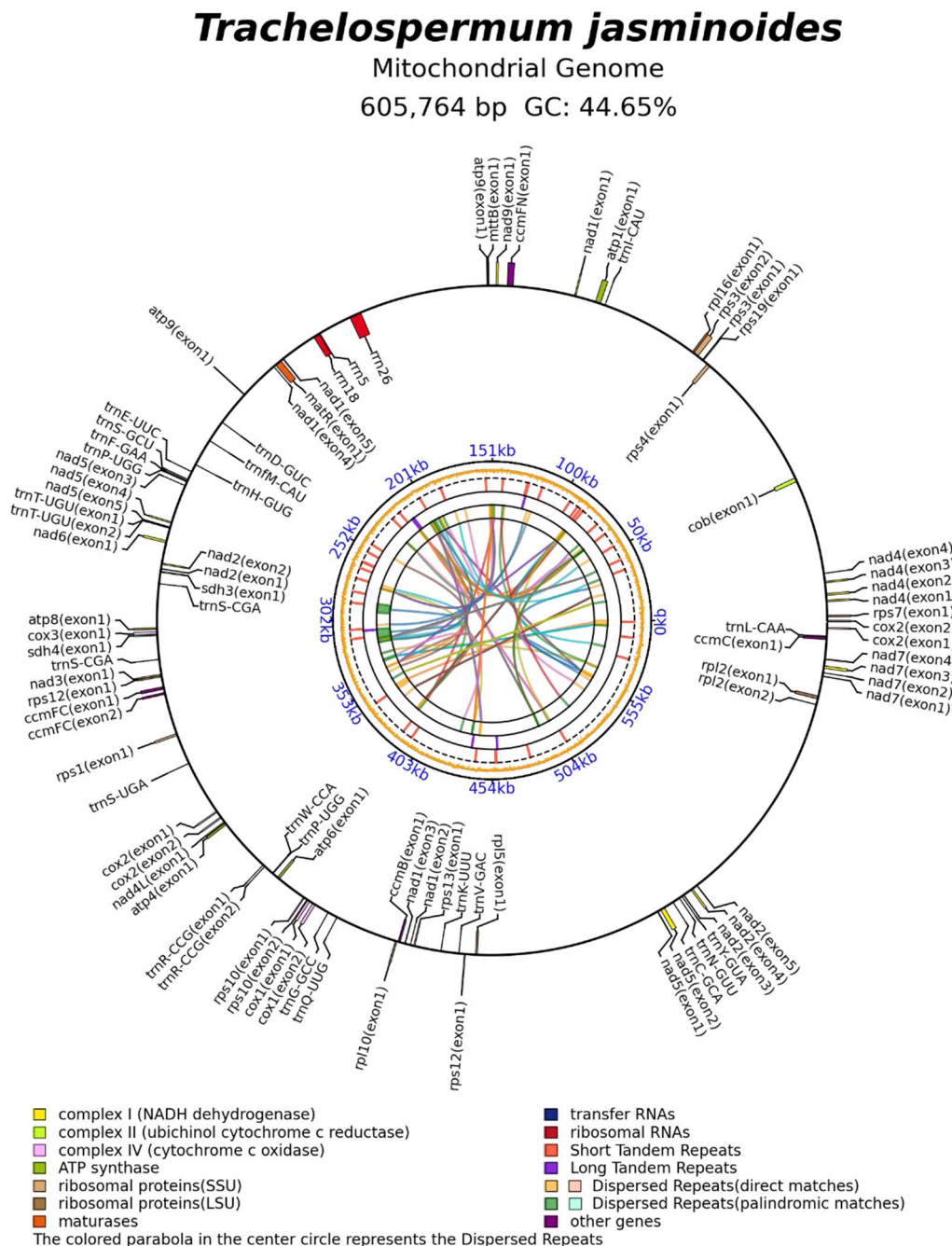


Fig. 2 Schematic representation of the *T. jasminoides* mitochondrial genome. The graph was drawn using PMGmap (<http://www.1kmpg.cn/pmgmap>). The graph is represented from inside out: (1) the relationship between dispersed repeat sequences; (2) the distribution of dispersed repeat sequences on the chromosome, where yellow represents direct dispersed repeat sequences and green represents inverted dispersed repeat sequences; (3) the distribution of tandem repeat sequences on the chromosome; (4) the distribution of simple sequence repeat sequences on the chromosome; (5) the distribution of GC content on the chromosome; (6) the scale coordinate axis; (7) genes located on the negative strand and positive strands

DRSs, also known as scattered repetitive sequences, are repeated at different locations throughout the genome and can be applied in various fields, such as biological evolution, disease diagnosis, and individual identification [57]. ROUSFinder2.0.py [37] was employed to predict DRSs in this study. Our analysis revealed a total of 42

DRSs within the mitochondrial genome of *T. jasminoides*. Among these, 14 were inverted repetitive sequences and 28 were direct repetitive sequences (Table S6). The longest DRS identified was DRS1, with 7,460 bp.

Table 2 Genes predicted in the mitogenome of *T. jasminoides*

Group of genes	Name of genes
Subunit of ATP synthase	<i>atp1, atp4, atp6, atp8, atp9(x2)</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFC, ccmFN</i>
Apocytochrome b	<i>cob</i>
Subunit of cytochrome c oxidase	<i>cox1, cox2, cox3</i>
Maturase related protein	<i>matR</i>
Transport membrane protein	<i>mttB</i>
Subunit of NADH dehydrogenase	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
Small subunit of ribosome	<i>rps10, rps12, rps13, rps19, rps3, rps4, rps1, rps7,</i>
Large subunit of ribosome	<i>rpl16, rpl2, rpl5, rpl10</i>
Subunit of succinate dehydrogenase	<i>sdh3, sdh4</i>
Transfer RNAs	<i>trnL-CAA, trnR-CCG, trnS-CGA(x2), trnT-UGU, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnI-CAU, trnI-M-CAU, trnN-GUU, trnP-UGG(x2), trnQ-UUG, trnS-GCU, trnS-UGA, trnV-GAC, trnW-CCA, trnY-GUA</i>
Ribosomal RNAs	<i>rrn18, rrn26, rrn5</i>

Repeat-mediated recombination

Subgenomic structures are primarily formed by repeat-mediated recombination within plant mitochondrial genomes [58]. Consequently, DRs have the potential to mediate HR [59]. To investigate the recombination of the *T. jasminoides* mitochondrial genome, we initially identified 42 DRs within the mitochondrial genome using the ROUSFinder software [37]. We then extracted each pair of repetitive sequences along with the 500 bp flanking sequences, creating reference sequences that represent the four potential conformations for recombination (Fig. 3A). Additionally, through comparative analysis of DBS and DRs, we observed that all six DBS sequences were consistent with certain repeat sequences (Table 1 and S6). These repetitive sequences corresponded to DBS04, DBS05, DBS01, DBS02, DBS03, and DBS06 (Table 3). We then mapped the nanopore reads to the 42 DR conformations, designating them as DRS01 to DRS42. Notably, we found corroborating evidence for six repetitive sequences (DRS01-DRS06) through the long nanopore reads (Table 3 and Figure S2). Moreover, nearly no nanopore long reads supported the conformations mediated by the remaining DRs (Table S7).

To validate the presence of recombination conformations mediated by DRS01, DRS02, DRS03, DRS04, DRS05, and DRS06, specific primers were designed for PCR amplification. These primers targeted DNA sequences corresponding to the c1-c4 conformations for each DR. Given the length of DRS01 and DRS02 exceeded the limits for PCR amplification, we were only able to verify their boundaries. Subsequently, the PCR products were subjected to Sanger sequencing, and the

results of gel electrophoresis and Sanger sequencing are presented in Fig. 3B and S3-S8. The sequencing results were aligned with the reference sequences of c1-c4 for each DR, confirming mitochondrial genome recombination events.

Moreover, we created a diagram illustrating the circular molecules that may have resulted from recombination events facilitated by the repetitive sequences DRS01, DRS02, DRS03, DRS04, DRS05 and DRS06 (Fig. 4). We then defined the genomic conformation as a set of chromosomes that can contain all genomic contents. We defined the genomic conformation prior to rearrangement as MC (Fig. 4). Other genomic conformation, namely, MC1, MC2, MC3, MC4, MC5, and MC6, can be formed after recombination, mediated by DRS01, DRS02, DRS03, DRS04, DRS05 and DRS06 alone, respectively. Moreover, if the genomic conformation has two circles, the large circle is named MCn-1, while the small circle is named MCn-2 ($n=2, 3, 4, 5$). MC1 and MC6 each consisted of a single chromosome mediated by inverted repeats, while MC2-5 comprised two chromosomes (MC2-1 and MC2-2, MC3-1 and MC3-2, MC4-1 and MC4-2, MC5-1 and MC5-2) mediated by directed repeats.

Chloroplast genome assembly, annotation, and MTPT prediction

MTPTs are plastid-derived DNA fragments in mitochondrial genomes [60]. In this study, we used the same dataset as that used for mitochondrial genome (OR333986.1) to assemble the chloroplast genome (OR286427.1) of *T. jasminoides*. The chloroplast genome of *T. jasminoides* spans 154,757 bp and has a GC content of 38.43%. The chloroplast genome encompasses a total of 130 genes (112 unique genes), including 85 protein-coding genes (78 unique genes), 37 tRNA genes (30 unique genes), and 8 rRNA genes (4 unique genes) (Table S8 and Figure S9).

Then, we identified a total of 11 MTPTs between the chloroplast and mitochondrial genome of *T. jasminoides* by conducting sequence similarity calculations. Among these, MTPT7 was found to be localized at two different positions on the mitochondrial genome and in the “small single copies” (SSC) region of chloroplast genome. MTPT1 (Complete: *ycf15, trnV-GAC*) and MTPT3 (partial: *rpl2*) were observed to have two copies localized within the “inverted repeat” (IR) region of the chloroplast genome, while only one copy migrated to the mitochondrial genome. The remaining MTPTs were located in the “large Single Copies” (LSC) region of chloroplast genome (Fig. 5). These MTPTs were 9,131 bp, representing approximately 1.51% of the total length of the *T. jasminoides* mitochondrial genome and 5.90% of the chloroplast genome. The longest MTPTs observed was MTPT1

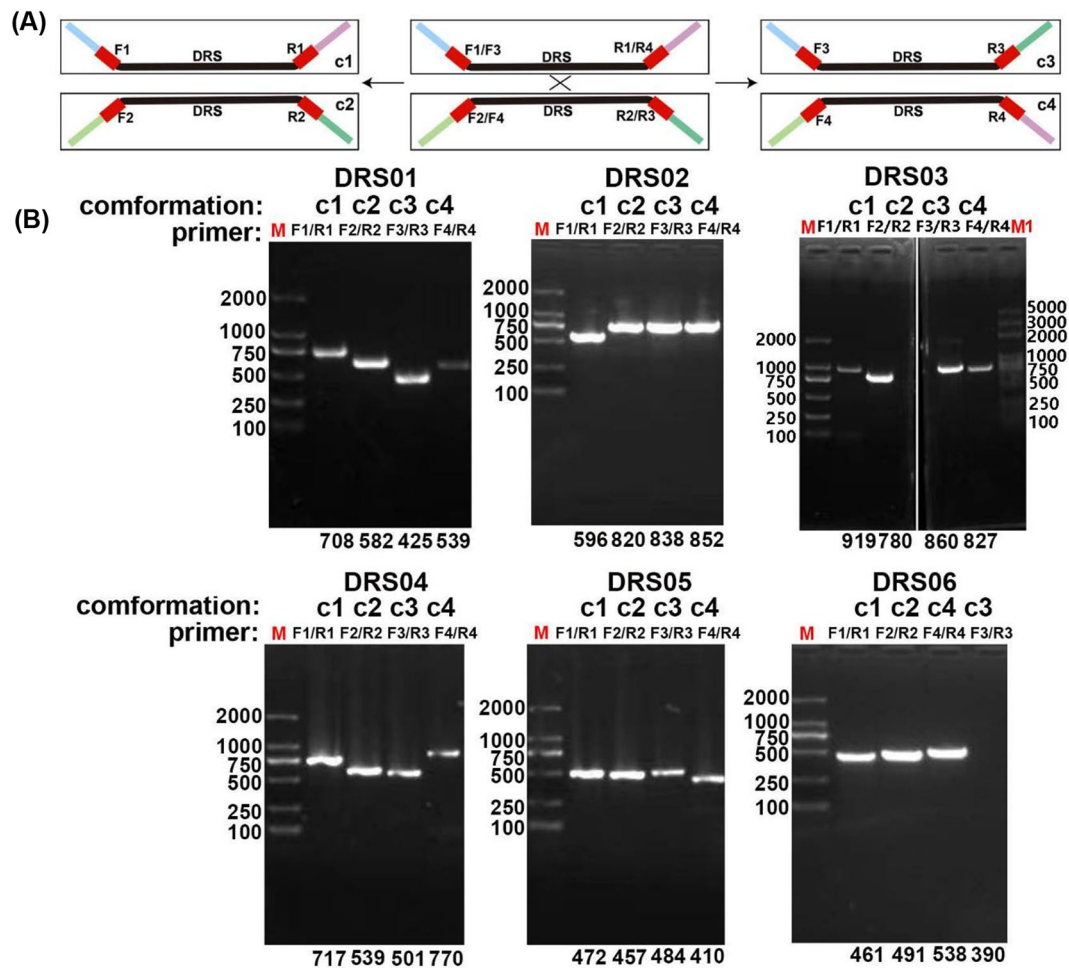


Fig. 3 PCR verification of recombination products by repetitive sequences. **(A)** Schematic representation of the four conformations (c1–c4) by recombination. The repetitive sequence and flanking sequence were composed of four configurations (c1, c2, c3 and c4). F1, F2, F3 and F4: Forward primers. R1, R2, R3 and R4: Reverse primers. c1, c2, c3 and c4: conformations. M: DL2000 marker, M1: DL5000 marker. **(B)** Electrophoretic gel of PCR products amplified with forward and reverse primers to DNA molecules corresponding to conformations c1–c4 for each DRS. The figure was cropped and the full-length electrophoretic gel was shown in the Figure S14A–C of Supplementary file. Four conformations (c1–4) were found in lanes 1–11 of Figure S14A, corresponding to recombination products of DRS01 and DRS03. Among these, the bands in DRS03 in the Fig. 3B were from different gels and were clearly delineated by white space, where the left was from lanes 6–8 of Figure S14A and the right was from lanes 9–11. Four conformations (c1–4) were found in Figure S14B, corresponding to DRS06 and DRS02, and four conformations (c1–4) were found in lanes 1–10 of Figure S14C, corresponding to DRS04 and DRS05.

with 3,857 bp, while the shortest MTPTs was 109 bp. The chloroplast and mitochondrial genomic positions of the MTPTs are shown in Table S9a. We further validated the accuracy of MTPTs by mapping nanopore long reads to MTPTs and flanking sequences (Figure S10).

To investigate the functions of the 11 MTPTs, we annotated their DNA fragments. These result revealed 13 unique genes within the MTPTs. Four complete genes were identified: one PCG and three tRNA genes (*yef15*, *trnV-GAC*, *trnW-CCA*, *trnD-GTC*, respectively). However, most PCGs were pseudogenized (*yef15*, *rpoC2*, *ndhJ*, *rps11*, *rpl36*, *ndhF*, *rpoB*, *psaB*, *psbE*), excluding *rpl2*. The phenomenon of non-functional genes in migrating fragments is a more common trait in plant mitochondrial genomes [60].

Moreover, we extracted MTPTs of the organelle genomes of 12 Gentianales species and performed cluster analyses using CD-HIT [43] with 85% identity. We observed that MTPT3 from *T. jasminoides*, MTPT4 from *G. straminea*, and MTPT25 from *H. lithophytica* clustered together. Additionally, MTPT9 from *T. jasminoides* clustered with MTPT2 from *C. arabica*, MTPT1 from *T. jasminoides* clustered with MTPT9 from *R. stricta*, MTPT8 from *T. jasminoides* clustered with MTPT20 from *C. arabica*, MTPT11 from *T. jasminoides* clustered with MTPT14 from *P. serpens*, and MTPT5 from *T. jasminoides* clustered with MTPT47 from *A. syriaca* (Table S9b–g). This finding indicates that the migration of chloroplast genes to the mitochondria, as observed in *T. jasminoides*, is present in other species of Gentianales.

Table 3 Mapping results for Nanopore long reads to the four possible conformations associated with 6 repetitive sequences (DRS01–DRS06)

ID of the DRS	ID of the DBS	Alignment Length (nt)	Positions of Repeat Copy 1		Positions of Repeat Copy 2		Type	Numbers of Long Reads Mapped to Each Conformation				Recombination Frequency (%)
			start	end	start	end		c1	c2	c3	c4	
DRS01	DBS4	7,460	288,982	296,441	310,009	317,468	Inverted	54	59	52	58	48.40%
DRS02	DBS5	3,931	359,772	363,702	601,834	605,764	Direct	65	73	72	74	51.41%
DRS03	DBS1	518	152,647	153,164	231,395	231,912	Direct	95	89	13	24	16.74%
DRS04	DBS2	266	196,488	196,753	556,153	556,418	Direct	90	91	3	6	4.74%
DRS05	DBS3	129	343,152	343,280	385,989	386,117	Direct	95	91	1	1	1.06%
DRS06	DBS6	128	193,682	193,809	231,202	231,329	Inverted	86	54	0	1	0.71%

Phylogenetic analysis

To analyze phylogenetic relationships of the 12 mitochondrial genomes of Gentianales species, we established a phylogenetic tree using the CDS of 23 shared PCGs from these mitochondrial genomes: 21 core genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *cox1*, *cox2*, *cox3*, *cob*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*) and two variable genes (*rps3*, *rps13*). The phylogenetic tree showed that *T. jasminoides*, *A. syriaca*, *H. lithophytica*, and *Rhazya stricta* clustered together with 100% bootstrap values. These results illustrated that species in Apocynaceae are closely related to each other. Meanwhile, all nodes had bootstrap support values >75%, demonstrating the higher reliability of the phylogenetic relationship among the 12 Gentianales species based on the mitochondrial genomes (Fig. 6).

The phylogenetic distribution of PCGs among the 11 Gentianales is shown in Fig. 6. This analysis revealed several shared and unique PCG and intron gains and losses within the Gentianales. The mitochondrial genomes of Gentianales shared 29 PCGs (Table S10): 24 core genes and 5 variable genes (*rpl5*, *rps3*, *rps12*, *rps13*, and *sdh4*). The presence of *rps14* was unique to *R. stricta* among the Gentianales species analyzed. Additionally, the absence of *sdh3* was exclusive to *R. stricta* among the Apocynaceae species analyzed. Similarly, the absence of *rps1* was distinctive to *Psychotria* species. Furthermore, *Gentiana* species lack several genes, including *rps14*, *rps7*, *rps19*, *rpl10*, *rpl2*, *sdh3*, and *rps10*. In all species within Gentianales, the introns *cox2i691* and *nad7i676* were absent (Table 4). Moreover, *cox1i729* was absent in the mitochondrial genome of *D. indicus* and *N. cadamba*, whereas *cox2i373* and *rps3i74* were absent in the mitochondrial genome of *C. arabica*. The phylogenetic distribution of the introns *cox1i729*, *cox2i373*, and *rps3i74* suggests independent losses of these introns. Among these, deletion of the start sites of the *atp4* and *ccmFC* genes occurred in the mitochondrial genome of *A. syriaca*, and *mttB* in *T. jasminoides* exhibited considerable sequence loss, accounting for more than half of its sequences compared to that in other Apocynaceae species (Figure S11). These findings provide a reference for further exploration of the mitochondrial genomes of other species in Apocynaceae.

RNA editing events analysis

RNA editing is an essential post-transcriptional biological process that specifically changes the nucleotide sequence of a primary transcript, enabling the alteration of genetic information at the transcriptional level. This process regulates translation and expands the information content of the original DNA template. In plant organelles, RNA editing occurs mainly in the form of cytidine-to-uridine conversion [61, 62].

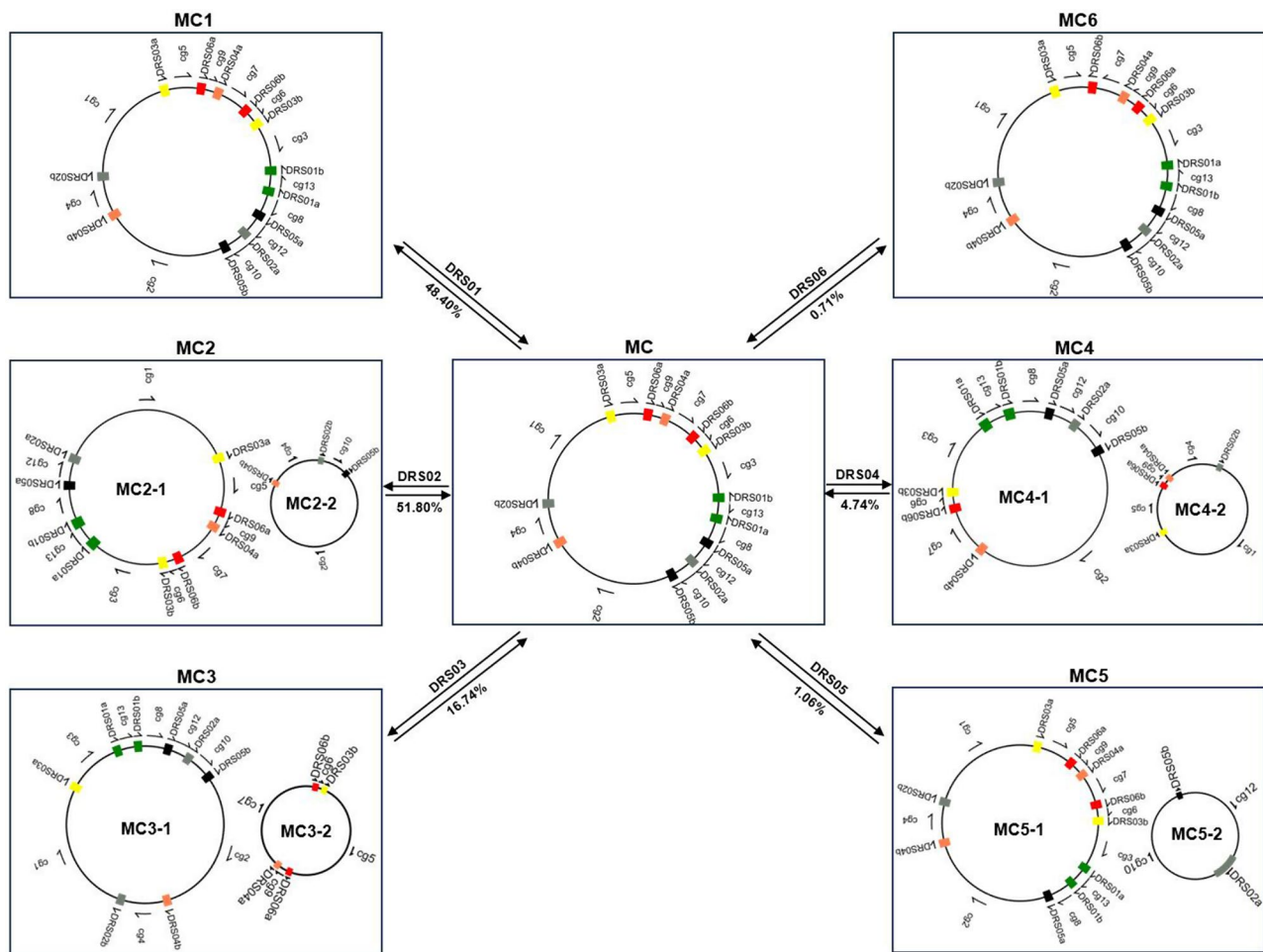


Fig. 4 The diagram illustrates the products of HR mediated by the repetitive sequences (DRS01-06). Each pair of repeat sequences is divided into DRSa and DRSb. The green fragments represent DRS01a and DRS01b. The moonlight silver fragments represent DRS02a and DRS02b. The lemon-yellow fragments represent DRS03a and DRS03b. The coral-colored fragments represent DRS04a and DRS04b. The black fragments represent DRS05a and DRS05b. The red fragments represent DRS06a and DRS06b. The black lines represent the contigs (cg) between repeats. The black box represents a set of genomic conformations. The circles show master circular molecule. MC is the genomic conformation prior to rearrangement. MC1-6 are the rearranged genomic conformation. MC1 and MC6 have only one chromosome, and MC2-5 have two chromosomes (MC2-1 and MC2-2, MC3-1 and MC3-2, MC4-1 and MC4-2, MC5-1 and MC5-2)

We identified 533 RNA-editing sites in the *T. jasminoides* mitochondrial genome based on RNA-seq data. To eliminate false positive results resulting from polymorphic sites in the mitochondrial genome, we identified single nucleotide polymorphisms (SNPs) through DNA-seq data obtained from the same sample as the RNA-seq data. We found two overlaps, namely, *nad3-102* and *nad5-87*, by comparing the detected RNA editing sites with the SNP sites (Table S11). Consequently, we identified 531 RNA editing sites in the *T. jasminoides* mitochondrial genome (Table S12 and Figure S12). Among these, 66 (12.43%) RNA-editing sites resulted in synonymous amino acid codons, while 465 (87.57%) resulted in non-synonymous amino acid codons. The non-synonymous codons were changed mainly to Ser, Pro, and Arg amino acids. Among them, 113 (21.28%) RNA editing

sites changed from Ser to Leu, 112 (21.09%) editing sites changed from Pro to Leu, 74 (13.94%) RNA editing sites changed from Ser to Phe, and 57 (10.73%) RNA editing sites changed from Arg to Cys (Table S12). RNA editing events in the *T. jasminoides* mitochondria occurred mainly at the first and second base positions of codons, accounting for 32.96% (175) and 62.52% (332) of all the RNA editing sites, respectively. Moreover, we found that all 38 unique protein-coding genes in the mitochondrial genome of *T. jasminoides* underwent RNA editing events. Among these, *nad4* exhibited the highest number of RNA editing sites at 54, followed by *ccmB* at 44 (Fig. 7A).

To confirm RNA editing at these sites, we randomly selected ten genes containing the 110 detected RNA editing sites. The corresponding genomic DNA (gDNA) and cDNA sequences were amplified using specific primers.

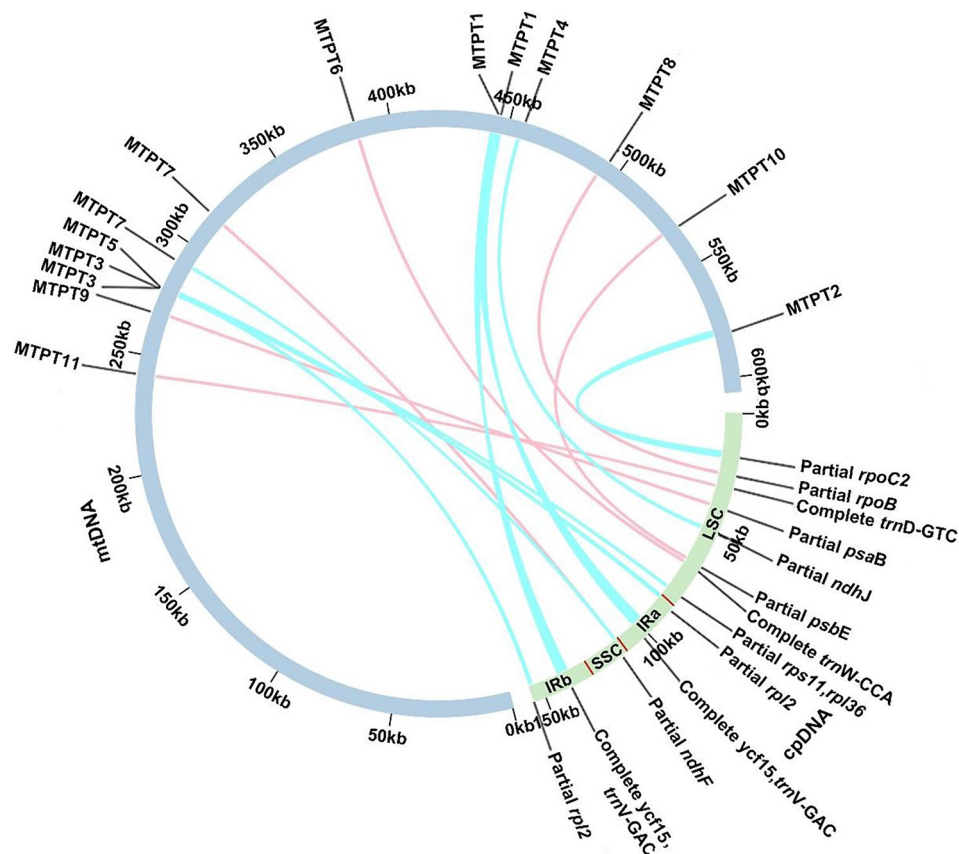


Fig. 5 Mitochondrial plastid DNAs (MTPTs) of the *T. jasminoides* mitochondrial genome. The blue arcs represent the mitochondrial genome, and the green arcs represent the chloroplast genome. The blue and pink arcs inside the circle represent the homologous regions between the mitochondrial and the chloroplast genome. The chloroplast genome of *T. jasminoides* is divided into LSC, IRa, SSC, and IRb by red line

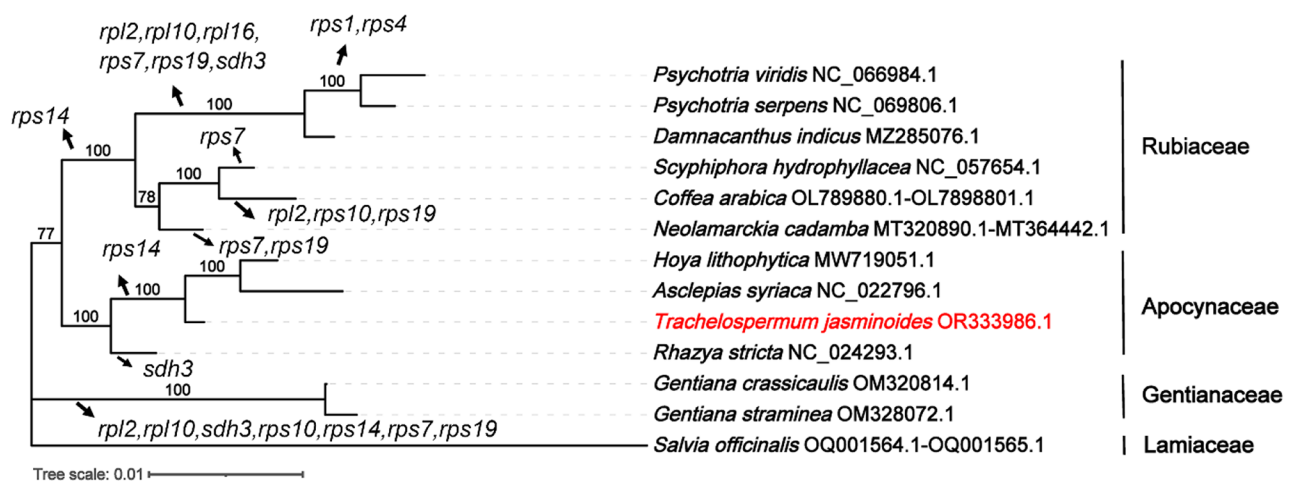


Fig. 6 Phylogenetic relationships of *T. jasminoides* and other 12 species. The phylogenetic tree based on the coding sequences of 23 PCGs from the mitogenome. The numbers indicate the bootstrap values for the maximum likelihood (ML) tree. The number on each branch node represents the bootstrap number. The mitochondrial genome NCBI accession number for each species follows the Latin name. The families of these species are shown on the right side. Lost genes were labelled on branches

Table 4 The comparison of the intron content of *T. jasminoides* with eleven other Gentianales species

Species	<i>P. viridis</i>	<i>P. serpens</i>	<i>D. indicus</i>	<i>S. hydrophyllacea</i>	<i>C. arabica</i>	<i>N. cadamba</i>	<i>H. lithophytica</i>	<i>A. syriaca</i>	<i>T. jasminoides</i>	<i>R. stricta</i>	<i>G. crassicaulis</i>	<i>G. straminea</i>
ccmfc829	●	●	●	●	●	●	●	●	●	●	●	●
cox1i729	●	●	○	●	●	○	●	●	●	●	●	●
cox2i373	●	●	●	●	●	●	●	●	●	●	●	●
cox2i691	○	○	○	○	○	○	○	○	○	○	○	○
nad1i394	□	□	□	□	□	□	□	□	□	□	□	□
nad1i477	●	●	●	●	●	●	●	●	●	●	●	●
nad1i669	□	□	□	□	□	□	□	□	□	□	□	□
nad1i728	□	□	□	□	□	□	□	□	□	□	□	□
nad2i156	●	●	●	●	●	●	●	●	●	●	●	●
nad2i542	□	□	□	□	□	□	□	□	□	□	□	□
nad2i709	●	●	●	●	●	●	●	●	●	●	●	●
nad2i1282	●	●	●	●	●	●	●	●	●	●	●	●
nad4i461	●	●	●	●	●	●	●	●	●	●	●	●
nad4i976	●	●	●	●	●	●	●	●	●	●	●	●
nad4i1399	●	●	●	●	●	●	●	●	●	●	●	●
nad5i230	●	●	●	●	●	●	●	●	●	●	●	●
nad5i1455	□	□	□	□	□	□	□	□	□	□	□	□
nad5i1477	□	□	□	□	□	□	□	□	□	□	□	□
nad5i1872	●	●	●	●	●	●	●	●	●	●	●	●
nad7i140	●	●	●	●	●	●	●	●	●	●	●	●
nad7i209	●	●	●	●	●	●	●	●	●	●	●	●
nad7i676	○	○	○	○	○	○	○	○	○	○	○	○
nad7i917	●	●	●	●	●	●	●	●	●	●	●	●
rpl2i846	×	×	×	×	×	×	×	×	×	×	×	×
rps10i235	●	●	●	●	●	●	●	●	●	●	●	●
rps3i74	●	●	○	●	●	●	●	●	●	●	●	●

●: cis-spliced intron, □: trans-spliced intron, ○: intron absent, X: gene loss. The mitochondrial genomes of *S. hydrophyllacea*, *P. serpens*, *H. lithophytica*, *G. crassicaulis*, *D. indicus*, and *C. arabica* mitochondrial genomes were re-annotated using mgavias (<http://www.ikmpg.cn/mgavias/>) in this study

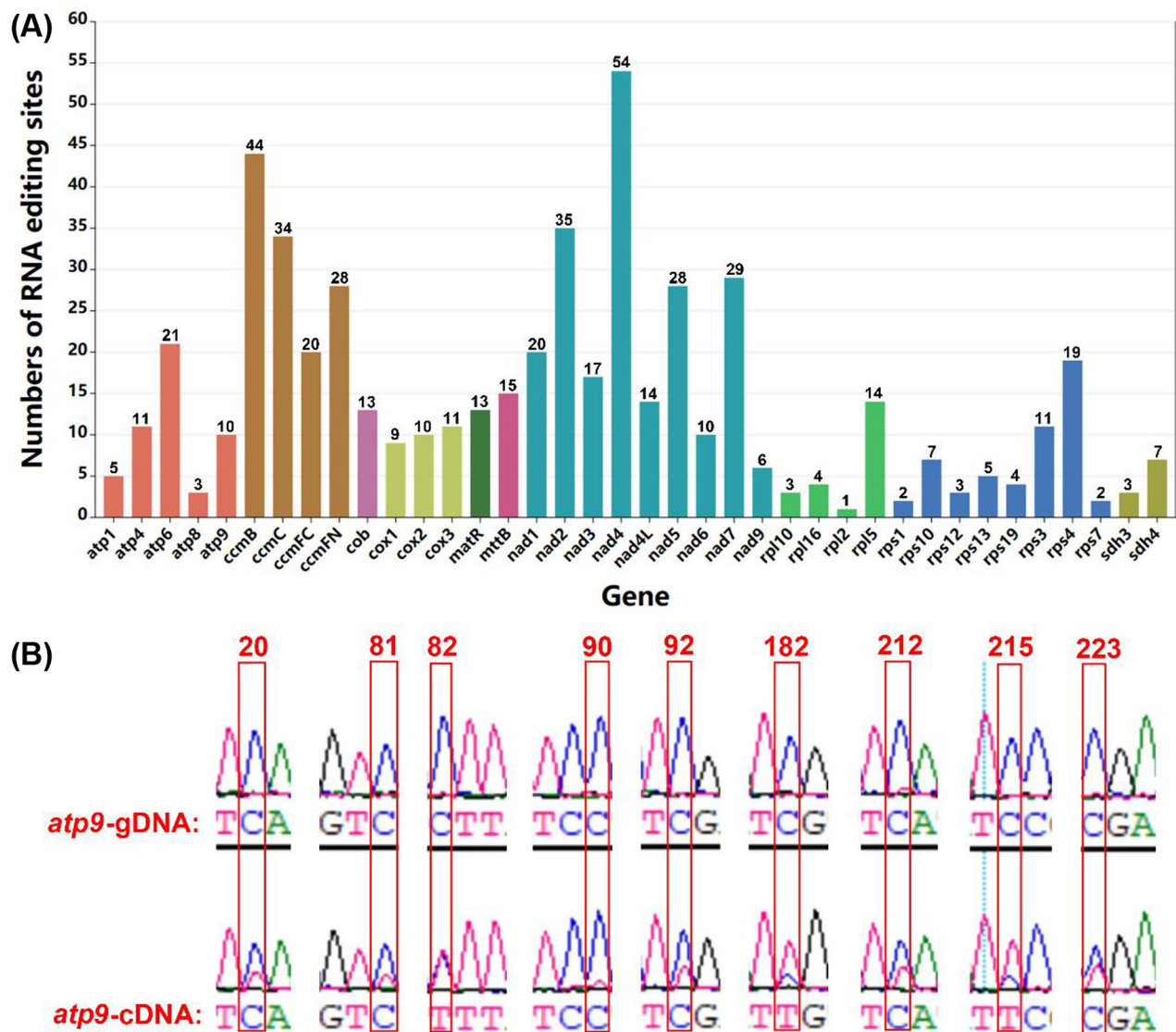


Fig. 7 Numbers of RNA editing sites and validation of nine predicted RNA editing sites in the *T. jasminoides* mitochondrial genome. **(A)** The graph was drawn using ChiPlot (<https://www.chiplot.online/>). The x-axis represents the name of the gene undergoing RNA editing. The y-axis represents the number of RNA editing for the gene. Different colours represent different family genes. **(B)** The figure above shows the sequence and chromatogram of the PCR amplified using gDNA as template (*atp9*). The upper panel shows the sequence and chromatogram of the PCR product amplified on the gDNA template. The lower panel shows the sequence and chromatogram of the PCR product amplified on the cDNA template. RNA editing sites are shown in red rectangle

The amplified products were sequenced by Sanger sequencing. The genomic DNA and cDNA-amplified sequences are shown in Fig. 7B and S13. Finally, we successfully identified 104 (94.5%) RNA editing sites. Notably, among these RNA editing site changes, a particularly remarkable outcome is the generation of start and stop codons. Specifically, codon changes in *atp6*-718 and *atp9*-223 transformed CAA into TAA and CGA into TGA, respectively, effectively creating stop codons (Fig. 7B and S13B).

Discussion

Overview of the *T. jasminoides* mitochondrial genome

Plant mitochondrial genomes are complex, rich in repetitive sequences, and have dynamically variable features, making their assembly remarkably challenging. In the present study, we identified the complete mitochondrial genome of *T. jasminoides*. To understand the diversity and evolution of the mitochondrial genome, we assembled and annotated the mitochondrial genome of *T. jasminoides* using a combination of Illumina short reads and Oxford Nanopore long reads. We then characterized the gene content and number, SSRs, TRSs, DRSs, MTPTs

and RNA editing events within the mitochondrial genome of *T. jasminoides*. Finally, we analyzed the phylogenetic relationships of 12 Gentianales species based on conserved PCGs of their mitochondrial genomes. This study provide a reference for the genomes of other species in Apocynaceae and an important theoretical basis for understanding the diversity and evolution of *T. jasminoides* mitochondrial genome.

Architecture of one master circular molecular structure and multiple variable forms for the *T. jasminoides* mitochondrial genome

The *T. jasminoides* mitochondrial genome contains one master circular molecular of 605,764 bp (Fig. 1B). Plant mitochondrial genomes generally have multiple alternative conformations owing to several repetitive sequences [58, 63]. In the present study, we identified six repetitive sequences (DRS01-06) in the mitochondrial genome of *T. jasminoides* that could mediate recombination, resulting in alternative conformations. We predicted six alternative genomic conformations MC1–6 based on independent repetitive sequences. Among these, the repetitive sequence-mediated recombinations DRS01-05 were double-ended recombinations, producing four conformations. However, DRS06 was different. Its mediated recombination was verified by nanopore long reads mapping, and PCR experiments showed it to be a single-ended recombination, producing only three conformations. Although we did not find single-ended recombination mediated by repetitive sequences in the mitochondrial genomes of plants such as *Salvia miltiorrhiza* [64] and *C. arabica* [65], this phenomenon is not erroneous and has been mentioned in previous studies [66]. Thus, single-ended recombination mediated by repetitive sequences in the mitochondrial genome of *T. jasminoides* provides new insights into its genetics and the evolution of mitochondrial genomes. Moreover, repeat-mediated HR is an intrinsic drivers of structural diversity in plant mitochondrial genomes. By comparing six pairs of repetitive sequences that mediate recombination in *R. stricta* and *T. jasminoides*, we discovered that they have no sequence similarity. Therefore, the mechanism of recombination mediated by repetitive sequences in other Apocynaceae species requires further experimental evidence. Although a series of mitochondrial genomic conformations of *T. jasminoides* was generated owing to recombination, the dominant conformation was a single circle.

Research trends of the *T. jasminoides* mitochondrial genome

Plant mitochondrial genomes are dynamically changing, and studies of the structure, function, and evolution of plant mitochondrial genomes must present

both their major and minor conformations [59, 67, 68]. Repetitive sequence-mediated recombination leads to the formation of secondary conformations. Plant mitochondrial genomes usually have repetitive sequences [58, 69]. Therefore, SSRs could potentially function as valuable molecular markers for future investigations. However, further experimental evidence is necessary to determine whether these repetitive sequences can facilitate HR. Additionally, future mitogenomic research should thoroughly investigate repeat sequence-mediated recombination.

MTPTs are relatively common phenomenon in plant mitochondrial genomes and show a high degree of diversity [60]. MTPTs contain numerous PCGs and tRNA. The same phenomenon was observed in *T. jasminoides*. The MTPTs of *T. jasminoides* mitochondrial genome contained three complete tRNA genes and some partial PCGs. Furthermore, MTPT7 from chloroplasts migrated to two different locations in the mitochondrial genome. However, MTPT3 (partial *rpl2*) is localized in the mitochondrial and chloroplast genomes, and we cannot accurately determine whether MTPT3 is derived from mitochondrial or chloroplast genome of the *T. jasminoides*. Nevertheless, *rpl2* is a horizontally transferred genes that migrates from the mitochondrial genome to the plastome of *A. syriaca* [24]. In addition, *A. syriaca* and *T. jasminoides* belong to Apocynaceae. Thus, MTPT3 (partial *rpl2*) is most likely a mitochondrial DNA insert derived from the mitochondria of *T. jasminoides*. However, the reasons for the migration of these fragments and their specific roles in the mitochondrial genome require further exploration.

RNA editing is a widespread phenomenon in higher plants organelles and has multiple biological functions, including influencing the biogenesis of plant organelles [70]. In addition, RNA editing events have been reported in the mitochondrial genomes of several higher plants; for example, 441 RNA editing sites exist in *Arabidopsis thaliana* [71]. Furthermore, RNA editing events occur most often at the first or second codon position in the mitochondrial genome. In the present study, we identified 531 single-base-altering RNA editing sites and discovered only 84 (15.8%) RNA editing sites occurring at the third codon position and 477 (84.2%) RNA editing sites occurring at the first or second codon position. All PCGs in the *T. jasminoides* had RNA editing events. Therefore, the regulatory role of RNA editing at the RNA level may considerably affect PCGs functions. However, the detailed reason for and timing of RNA editing occurrences require further study. We found that the coverage of the *atp9* is much greater than that of other genes (Table S12), and *atp9* had two copies in the mitochondrial genome of *T. jasminoides*. The results of the BLASTN analysis revealed 100% identity between the

two copies of *atp9*. Plant mitochondrial *atp9* is not only a key gene for energy synthesis in the inner mitochondrial membrane but also directly related to cytoplasmic male sterility in some plants such as tobacco and onion [72, 73]. We also compared *atp9* of *T. jasminoides* and tobacco by BLAST [42] and found that the shared 95.07% identity, indicating that *atp9* is likely able to regulate cytoplasmic male sterility in *T. jasminoides*. This finding can facilitate research on plant breeding and the evolution of reproductive systems.

Conclusion

In this study, we showed that the predominant conformation of the *T. jasminoides* mitochondrial genome is a master circular molecular structure. Although multiple conformations likely coexist, they are transformed from single conformations via repeat-mediated recombination. We identified 12 mitochondrial plastid DNAs and 531 RNA-editing sites in the *T. jasminoides* mitochondrial genome. Our results provide a reference for studying the evolutionary relationships between mitochondria in Apocynaceae species.

Abbreviations

CDS	Coding sequences
cDNA	Complementary DNA
DBSs	Double bifurcating structures
DRSs	Dispersed repeat sequences
HR	Homologous recombination
MTPTs	Mitochondrial plastid sequences
PCR	Polymerase chain reaction
PCGs	Protein-coding genes
rRNA	Ribosomal RNA
RNA-seq	RNA sequencing
SSRs	Simple sequence repeats
SNP	single nucleotide polymorphism
TRSs	Tandem repeat sequences
tRNA	Transfer RNA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05568-6>.

Supplementary Material 1

Acknowledgements

We would like to thank Jingwen Yue and Mei Jiang for helping collect the data, Novogene Technology Co., Ltd for technical support and genomic sequencing service, and Professor Jinghong Zhang for identifying the plant as *T. jasminoides*.

Author contributions

CL conceived the study; YN and JLL assembled and annotated the mitogenome; CYS performed the comparative analysis and designed the primers and conducted the PCR experiment. CL, CYS, and HMC wrote the manuscript; CL and JHZ critically reviewed the manuscript. All authors have read and agreed on the contents of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [81872966], CAMS Innovation Fund for Medical Sciences (CIFMS) (2021-I2M-1-071), the National Science & Technology Fundamental Resources Investigation

Program of China [2018FY100705], the National Science Foundation Funds [81173503], Guiding projects in Fujian Province [2018Y0062] and Xiamen Municipal Natural Science Foundation Project, 3502Z202373044. The funders were not involved in the study design, data collection, and analysis, decision to publish, or manuscript preparation.

Data availability

The organelle sequences supporting the conclusions of this article are available in GenBank (<https://www.ncbi.nlm.nih.gov/>) with accession numbers: OR333986.1 (mitogenome) and OR286427.1 (cpgenome), respectively. The raw data have been released through GenBank with the following accession numbers: (1) Nanopore DNA reads: SRA database SRR27865236. (2) Illumina DNA reads: SRA database SRR27865237.

Declarations

Ethics approval and consent to participate

All plant materials were not from endangered or protected species. All sample collection and analyses complied with local institutional guidelines and legislation. The experimental protocols in this study were approved by the Institute of Medicinal Plant Development, Beijing, China. All experiments were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 May 2024 / Accepted: 2 September 2024

Published online: 16 October 2024

References

1. Song H, Tan J, Ma R, Kennelly EJ, Tan Q. Anti-inflammatory constituents from *Caulis Trachelospermi*. *Planta Med*. 2022;88(9–10):721–8.
2. Sheu MJ, Chou PY, Cheng HC, Wu CH, Huang GJ, Wang BS, Chen JS, Chien YC, Huang MH. Analgesic and anti-inflammatory activities of a water extract of *Trachelospermum jasminoides* (Apocynaceae). *J Ethnopharmacol*. 2009;126(2):332–8.
3. Tan XQ, Chen HS, Liu RH, Tan CH, Xu CL, Xuan WD, Zhang WD. Lignans from *Trachelospermum jasminoides*. *Planta Med*. 2005;71(1):93–5.
4. Tan XQ, Guo LJ, Chen HS, Wu LS, Kong FF. Study on the flavonoids constituents of *Trachelospermum jasminoides*. *Zhong Yao Cai*. 2010;33(1):58–60.
5. Tan XQ, Guo LJ, Qiu YH, Chen HS, Tan CH. Chemical constituents of *Trachelospermum jasminoides*. *Nat Prod Res*. 2010;24(13):1248–52.
6. Fatima T, Crank G, Wasti S. Alkaloids from *Trachelospermum jasminoides*. *Planta Med*. 1988;54(4):364.
7. Sang SF, Mei DS, Liu J, Zaman QU, Zhang HY, Hao MY, Fu L, Wang H, Cheng HT, Hu Q. Organelle genome composition and candidate gene identification for Nsa cytoplasmic male sterility in *Brassica napus*. *BMC Genomics*. 2019;20(1):813.
8. Cheng Y, Zhang L, Qi J, Zhang L. Complete chloroplast genome sequence of *Hibiscus cannabinus* and comparative analysis of the Malvaceae Family. *Front Genet*. 2020;11:227.
9. Yue Y, Li J, Sun X, Li Z, Jiang B. Polymorphism analysis of the chloroplast and mitochondrial genomes in soybean. *BMC Plant Biol*. 2023;23(1):15.
10. Wang H-X, Cheng X-L, Chen W-S, Li L-M, Chen L. Complete plastome sequence of *Trachelospermum Jasminoides* (Lindley) Lemaire (Apocynaceae). *Mitochondrial DNA Part B*. 2019;4(1):2086–7.
11. Kan SL, Shen TT, Ran JH, Wang XQ. Both Conifer II and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. *BMC Biol*. 2021;19(1):146.
12. Martínez-Reyes I, Chandel NS. Mitochondrial TCA cycle metabolites control physiology and disease. *Nat Commun*. 2020;11(1):102.
13. Lanlan ZHANGCL, Yuzhu FANG, Yan SONG, Wanlin KANG, Zhiyu LI, Xiao ZHANG, Rui ZHANG. Progress on the application of mitochondrial SSR molecular markers in plants. *Curr Biotechnol*. 2023;13(06):821–6.
14. Takenaka M, Jörg A, Burger M, Haag S. RNA editing mutants as surrogates for mitochondrial SNP mutants. *Plant Physiol Biochem*. 2019;135:310–21.

15. Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.* 2010;38(14):4755–67.
16. Grimes BT, Sisay AK, Carroll HD, Cahoon AB. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics.* 2014;15:31.
17. Grewe F, Edger P, Keren I, Sultan L, Pires JC, Osterseker-Biran O, Mower JP. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion.* 2014;19(Pt B):135–43.
18. Gualberto JM, Lamattina L, Bonnard G, Weil J-H, Grienenberger J-M. RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature.* 1989;341(6243):660–2.
19. Wu Z, Liao XZ, Zhang X, Tembrock LR, Broz AK. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. *J Syst Evol* 2020, 60.
20. Zardoya R. Recent advances in understanding mitochondrial genome diversity. *F1000Res* 2020, 9.
21. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the Stability of the plant mitochondrial genome. *Int J Mol Sci* 2020, 21(1).
22. Kejnovsky E, Jedlicka P. Nucleic acids movement and its relation to genome dynamics of repetitive DNA: Is cellular and intercellular movement of DNA and RNA molecules related to the evolutionary dynamic genome components? Is cellular and intercellular movement of DNA and RNA molecules related to the evolutionary dynamic genome components? *Bioessays* 2022, 44(4):e2100242.
23. Park S, Ruhlman TA, Sabir JS, Mutwakil MH, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics.* 2014;15(1):405.
24. Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). *Genome Biol Evol.* 2013;5(10):1872–85.
25. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21(1):241.
26. Richly E, Leister D. NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol.* 2004;21(6):1081–4.
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77.
28. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13(6):e1005595.
29. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 2015;31(20):3350–2.
30. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 2019;47(W1):W65–73.
31. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. *Molecular Ecology Resources;* 2023.
32. Tillich M, Lehwarck P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 2017;45(W1):W6–11.
33. Lee E, Harris N, Gibson M, Chetty R, Lewis S. Apollo: a community resource for genome annotation editing. *Bioinformatics.* 2009;25(14):1836–7.
34. Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, Liu C. Plant mitochondrial genome map (PMGmap): a software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. *Mol Ecol Resour.* 2024;24(5):e13952.
35. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–5.
36. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
37. Wynn EL, Christensen AC. Repeats of unusual size in Plant mitochondrial genomes: identification, incidence and evolution. *G3 (Bethesda).* 2019;9(2):549–59.
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
40. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178–92.
41. Swindell SR, Plasterer TN. SEQMAN. Contig assembly. *Methods Mol Biol.* 1997;70:75–89.
42. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 2015;43(16):7762–8.
43. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28(23):3150–2.
44. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative Toolkit developed for interactive analyses of big Biological Data. *Mol Plant.* 2020;13(8):1194–202.
45. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2020;20(1):348–55.
46. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
47. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4.
48. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the Ultrafast bootstrap approximation. *Mol Biol Evol.* 2017;35(2):518–22.
49. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49(W1):W293–6.
50. Picardi E, Pesole G. REDIttools: high-throughput RNA editing detection made easy. *Bioinformatics.* 2013;29(14):1813–4.
51. Wu B, Chen H, Shao J, Zhang H, Wu K, Liu C. Identification of symmetrical RNA editing events in the Mitochondria of *Salvia miltiorrhiza* by strand-specific RNA sequencing. *Sci Rep.* 2017;7:1–11.
52. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet.* 2004;5(6):435–45.
53. Jian Y, Yan W, Xu J, Duan S, Li G, Jin L. Genome-wide simple sequence repeat markers in potato: abundance, distribution, composition, and polymorphism. *DNA Res* 2021, 28(6).
54. Raskina O, Barber JC, Nevo E, Belyayev A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenet Genome Res.* 2008;120(3–4):351–7.
55. Easterling KA, Pitra NJ, Morcol TB, Aquino JR, Lopes LG, Bussey KC, Matthews PD, Bass HW. Identification of tandem repeat families from long-read sequences of *Humulus lupulus*. *PLoS ONE.* 2020;15(6):e0233971.
56. Hannan AJ. Tandem repeat polymorphisms: mediators of genetic plasticity, modulators of biological diversity and dynamic sources of disease susceptibility. *Adv Exp Med Biol.* 2012;769:1–9.
57. Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science.* 2004;303(5664):1626–32.
58. Zhong Y, Yu R, Chen J, Liu Y, Zhou R. Highly active repeat-mediated recombination in the mitogenome of the holoparasitic plant *Aeginetia indica*. *Front Plant Sci.* 2022;13:988368.
59. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, Yu J. 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(2):36.
60. Wang XC, Chen H, Yang D, Liu C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. *Mitochondrial DNA DNA Mapp Seq Anal.* 2018;29(4):635–42.
61. Zheng P, Wang D, Huang Y, Chen H, Du H, Tu J. Detection and analysis of C-to-U RNA editing in Rice Mitochondria-encoded ORFs. *Plants (Basel)* 2020, 9(10).
62. Wu CS, Chaw SM. Evolution of mitochondrial RNA editing in extant gymnosperms. *Plant J.* 2022;111(6):1676–87.
63. Yang Z, Ni Y, Lin Z, Yang L, Chen G, Nijjati N, Hu Y, Chen X. De novo assembly of the complete mitochondrial genome of sweet potato (*Ipomoea batatas* [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination. *BMC Plant Biol.* 2022;22(1):285.
64. Yang H, Chen H, Ni Y, Li J, Cai Y, Ma B, Yu J, Wang J, Liu C. De Novo 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).

65. Ni Y, Zhang X, Li J, Lu Q, Chen H, Ma B, Liu C. Genetic diversity of *Coffea arabica* L. mitochondrial genomes caused by repeat-mediated recombination and RNA editing. *Front Plant Sci.* 2023;14:1261012.
66. Logan DC. Annual Plant Reviews. In: *Plant Mitochondria*. Edited by Logan DC, vol. 50: Wiley-Blackwell; 2017.
67. Fang B, Li J, Zhao Q, Liang Y, Yu J. Assembly of the Complete Mitochondrial Genome of Chinese Plum (*Prunus salicina*): characterization of genome recombination and RNA editing sites. *Genes.* 2021;12(12):1970.
68. Yang Z, Ni Y, Lin Z, Yang L, Chen G, Nijiati N, Hu Y, Chen X. De novo assembly of the complete mitochondrial genome of sweet potato (*Ipomoea batatas* [L.] Lam) revealed the existence of homologous conformations generated by the repeat-mediated recombination. *BMC Plant Biol.* 2022;22(1):285.
69. Zhou Q, Ni Y, Li J, Huang L, Li H, Chen H, Liu C. Multiple configurations of the plastid and mitochondrial genomes of *Caragana spinosa*. *Planta.* 2023;258(5):98.
70. Tang W, Luo C. Molecular and functional diversity of RNA editing in Plant Mitochondria. *Mol Biotechnol.* 2018;60(12):935–45.
71. Giegé P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A.* 1999;96(26):15324–9.
72. Chang Y, Liu B, Jiang Y, Cao D, Liu Y, Li Y. Induce male sterility by CRISPR/Cas9-mediated mitochondrial genome editing in tobacco. *Funct Integr Genomics.* 2023;23(3):205.
73. Yuan Q, Song C, Gao L, Zhang H, Yang C, Sheng J, Ren J, Chen D, Wang Y. Transcriptome de novo assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in onion. *Plant Physiol Biochem.* 2018;125:35–44.

Publisher's note

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