

RESEARCH

Open Access



Characterization of the complete mitogenome of *Tiarella polyphylla*, commonly known as Asian foamflower: insights into the multi-chromosomes structure and DNA transfers

Bo Liu^{1,2†}, Qian Long^{2†}, Weiwei LV², Yu Shi^{2,3,4}, Pan Li^{5*} and Luxian Liu^{2,3,4*}

Abstract

Background *Tiarella polyphylla* D. Don has been traditionally used to cure asthma and skin eruptions. However, the sequence and the structure of the mitogenome of *T. polyphylla* remained elusive, limiting the genomic and evolution analysis based on the mitogenome.

Results Using a combination of Illumina and Nanopore sequencing reads, we *de novo* assembled the complete mitogenome of *T. polyphylla*. In addition to unveiling the major configuration of the *T. polyphylla* mitogenome was three circular chromosomes with lengths of 430,435 bp, 126,943 bp, and 55,269 bp, we revealed five (R01-R05) and one (R06) repetitive sequence could mediate the intra- and inter-chromosomal recombination, respectively. Furthermore, we identified 208 short and 25 long tandem segments, seven cp-derived mtDNAs, 106 segments of mtDNAs transferred to the nuclear genome, and 653 predicted RNA editing sites. Based on the sequence of the mitogenomes, we obtained the resolved phylogeny of the seven Saxifragales species.

Conclusions These results presented the mitogenome features and expanded its potential applications in phylogenetics, species identification, and cytoplasmic male sterility (CMS) in the future.

Keywords *Tiarella polyphylla*, Mitochondrial genome, Recombination, DNA transfers, Saxifragales

[†]Bo Liu and Qian Long contributed equally to this work.

*Correspondence:

Pan Li

panli_zju@126.com

Luxian Liu

liushuangcx2007@126.com

¹School of Agriculture, Henan Institute of Science and Technology, Xinxiang 453003, China

²State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng 475000, China

³Henan Dabieshan National Field Observation and Research Station of Forest Ecosystem, Zhengzhou 450000, China

⁴Xinyang Academy of Ecological Research, Xinyang 464000, China

⁵Systematic & Evolutionary Botany and Biodiversity Group, MOE Key Laboratory of Biosystems Homeostasis and Protection, College of Life Sciences, Zhejiang University, Hangzhou 310058, China

Background

Tiarella polyphylla D. Don (Saxifragaceae) is geographically distributed in Asia from the eastern Himalayas through China to Japan [1]. Traditionally, *T. polyphylla* has been used to treat asthma, skin eruptions, bruises, and hearing difficulties [2, 3]. Due to its medicinal value, some research focused on the main active components and pharmacological study of *T. polyphylla*. The compounds isolated from the whole plant of *T. polyphylla* include some flavonols, triterpenoids, and steroids, such as Tiarellic acid (TA), β -sitosterol, and ergosterol endoperoxide [3]. TA has anti-asthmatic properties [2], and β -sitosterol and ergosterol endoperoxide



show anti-complementary activity [3]. Moreover, the 3, 23-dihydroxy-20(29)-lupen-27-oic acid isolated from *T. polyphylla* regulated the Type 1-procollagen and MMP-1 (matrix metalloproteinases) on the UV-induced damage of cultured old age human dermal fibroblasts [4]. Together, these studies indicated the high medicinal potential of *T. polyphylla*. Consequently, this underscores the necessity of initiating molecular breeding programs for this species to enhance its therapeutic qualities.

With the increasing number of studies using different sequencing technologies in plants, genomic research on *T. polyphylla* has also been developed. Plant cells contain three genomic compartments, including the nuclear, mitochondrial (mt), and plastid (pt) genomes, and they interact with each other. Functions of the organellar proteins rely on the import of nuclear-encoded proteins and mutation accumulation in one compartment, causing the nuclear-cytoplasmic co-evolution [5]. The chromosome-level nuclear genome assembly and the complete plastome of *T. polyphylla* have been constructed [6, 7]. The chromosome-level nuclear genome of *T. polyphylla* featured seven chromosomes and was 403.10 Mb. Sizes of the seven chromosomes ranged from 44.08 Mb to 79.84 Mb. The size of the complete plastome of *T. polyphylla* was 154,850 bp. Although previous efforts to assemble the nuclear genome and plastome of *T. polyphylla* have provided valuable data, the mitogenome has not been reported yet.

Apart from energy production, mitogenomes also hold a crucial role in phylogenetics, species authentication, and cytoplasmic male sterility (CMS). Given the importance of CMS characteristics, the mitochondrial genome becomes crucial for molecular breeding in plants, facilitating the development of varieties with desirable traits. Contrarily to the plastome, whose structure is conserved among most angiosperms, the plant mitogenome is variable in size, structure, and gene order due to recombination through repeated sequences. Recombination mediated by the repeated sequences produced sub-stoichiometric molecules within the same plant cell referred to as “substoichiometric shifting” [8]. Many other features have been reported previously in plant mitogenomes, such as higher RNA editing frequency than the plastome. Moreover, plastid DNA insertions in the mitogenome and mitochondrial DNA insertions in the nuclear genome were commonly found in plants.

Here, we assembled the complete mitogenome of *T. polyphylla*, combining Illumina and Nanopore sequencing data. We characterized the main structural feature of the mitogenome of *T. polyphylla* and documented the repeat-mediated recombination supported by the Nanopore reads. Also, we identified the transfer of plastomic DNA to the mitogenome and mitochondrial DNA segments to the nuclear genome. Finally, based on the

mitogenome sequences, we performed a phylogenetic analysis of the seven Saxifragales species. The complete mitogenome sequence of *T. polyphylla* was the first sequenced mitogenome of the family Saxifragaceae, provided new sequences that can be used to infer evolutionary relationships, and offered a valuable genetic resource for comparative genomics within the order Saxifragales.

Methods

Plant materials and total DNA extraction

The fresh leaves were collected from the whole plants of *T. polyphylla* in Chongdugou scenic spot, Henan, China (111°39′41.64″E, 33°56′23.87″N). Dr. Luxian Liu (School of Life Sciences, Henan University, China) formally identified the plant material. *T. polyphylla* is neither classified as an endangered species nor subject to protection regulations, so no special permissions were required for its collection. A voucher specimen (LLX20160814) is deposited at the Herbarium of Henan University (HHU). The fresh leaves were used to extract total DNA using a plant genomic DNA kit (Tiangen Biotech, Beijing, Co., Ltd.). The purity and quantity of DNA were assessed using the Nanodrop spectrophotometer 2000 (Thermo Fisher Scientific, America).

DNA sequencing and mitogenome assembly

PromethION libraries were constructed from 1 µg of intact genomic DNA using the LSK108 ligation kit (Oxford Nanopore Technologies) following the manufacturer’s protocol. Sequencing was performed on a PromethION Flow Cell (Oxford Nanopore Technologies, UK).

Next-generation sequencing (NGS) sequencing libraries were prepared using Illumina’s Nextera DNA Flex library. Libraries were sequenced on Illumina’s NovaSeq 6000 System, obtaining 2×150 paired-end reads using the NovaSeq 6000 S2 Reagent Kit. All sequencing data utilized in this study were derived from our previous whole-genome sequencing research [6].

The mitogenome of *T. polyphylla* was assembled using a hybrid assembly method. We used Unicycler (v0.4.9) [9] to assemble the NGS and Nanopore reads. In brief, the hybrid assembly method for the *T. polyphylla* mitogenome combines the accuracy of short reads with the structural resolving power of long reads to create a complete mitogenome assembly. Initially, a primary assembly graph was constructed using NGS reads with SPAdes software, and Unicycler simplified this graph using NGS and Nanopore reads, resolving ambiguities and spanning repetitive regions. The NGS and Nanopore reads were mapped back to the assembled sequences using BWA (v0.7.12-r1039) [10] to validate the assembly results of the Unicycler. The number of reads mapped to the mtDNA was counted using the samtools (v1.3.1) [11].

Mitogenome annotation

We annotated the *T. polyphylla* mitogenome using GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) with manual checking using Apollo [12]. We used IPMGA (<http://www.1kmpg.cn/ipmga/>), which provides origin information from 423 tRNA genes across 12 species, to determine the origin of each tRNA. A map of the *T. polyphylla* mitogenome was constructed using OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>). The sequences and the annotation of the *T. polyphylla* mitogenome were deposited in GenBank under the accession number OR866907 for chromosome 1, OR866908 for chromosome 2, and OR866909 for chromosome 3 respectively.

Identification of the repeat-mediated recombination

To identify the recombination events mediated by repetitive sequences in the *T. polyphylla* mitogenome, we detected the intra-chromosomal repetitive sequences using ROUSfinder v2 [13] and inter-chromosomal repetitive sequences using BLASTN with the e -value $< 1e-5$. Then, the 1000 bp-long non-repeat regions flanking the repetitive sequences and the repetitive sequences were extracted and generated the four configurations corresponding to both 5'- and 3'-ends of each non-repeat contig encompassing one single recombination event. Based on the inferred alternative genome configurations, we calculated the stoichiometry of each configuration using Nanopore long reads. The configurations with the more supporting reads were the major configurations (Mac1, 2), and those with the less supporting reads were the minor configurations (Mic1, 2). To calculate the recombination efficiency, we mapped the Nanopore long reads to the four configurations of each recombination event. We extracted all mapped reads and calculated the read numbers for each configuration. The mapped reads of each configuration for the long reads supporting the recombination event were visualized using IGV (v 2.15.1) [14].

Analysis of the SSRs and long tandem repeats

Simple sequence repeats (SSRs) in the *T. polyphylla* mitogenome were searched by MISA [15]. The thresholds applied were 10, 5, 4, 3, 3, and 3 base pairs for all mono-, di-, trinucleotide, tetra-, penta-, and hexanucleotide repeat units, respectively. The tandem repeats in the *T. polyphylla* mitogenome were identified by the Tandem Repeats Finder (v 4.09) [16]. The parameters applied were two base pairs for matches, seven base pairs for mismatches and indels, and 50 and 500 base pairs for the minimum alignment score and maximum period size, respectively.

Identification of mitochondrial plastid and nuclear DNA sequences

We assembled the plastome of the *T. polyphylla* using GetOrganelle with the parameters: “-R 15 -k 21,45,65,85,105 -F embplant_pt”. To find the mitochondrial DNA sequences that are derived from inserted plastid DNA (MTPTs), we searched the sequence of *T. polyphylla* plastome against a custom database constructed by the sequence of the mitogenome, using BLASTN with the e -value $< 1e-5$ and removing hits < 100 bp with $< 80\%$ identity [17]. We extracted the 1000 bp-long regions flanking the MTPTs, and the sequences of MTPTs were extracted. We mapped the Nanopore long reads to these sequences and visualized the mapping results using IGV. We annotated these sequences using CPGAVAS2 (<http://www.1kmpg.cn/cpgavas2>) to check the genes located in the MTPTs.

To identify the mitochondrial DNA insertions in the nucleus (NUMTs), we searched the sequence of the *T. polyphylla* mitogenome against a custom database constructed by the sequence of the nuclear genome of *T. polyphylla*. The nuclear genome was downloaded from the CNCB-NGDC database (<https://ngdc.cncb.ac.cn/gwh/Assembly/26102/show>) under the accession number GWHBKAH00000000. The NUMTs were identified using the same procedure as MTPTs. Circos implemented in TBtools [18] were used to plot the MTPTs and NUMTs located in the assembled mitochondrial chromosomes of *T. polyphylla*.

Prediction C-to-U RNA editing sites

To predict the cytidines (C) for uridines (U) RNA editing sites in the protein-coding regions of the *T. polyphylla* mitogenome, we applied Deepred-mt [19] to analyze the coding sequences. The RNA editing sites predicted in the protein-coding regions with the > 0.9 probability were included in the current study.

Phylogenetic analysis of the seven Saxifragales species based on mitogenomes

For phylogenetic reconstructions based on the mitogenomes, the sequence of the complete mitogenomes of *T. polyphylla* was obtained in this study, and fourteen other species of Saxifragales were downloaded from the GenBank database (Table 1). The coding sequences of twelve shared genes among these fifteen species were extracted using PhyloSuite (v1.2.1) [20] and were aligned with MAFFT (v7.450) [21]. We selected the best-fit evolutionary model and partitioning scheme using PartitionFinder2 [22]. The twelve shared genes were *atp1*, *atp4*, *atp6*, *atp8*, *ccmC*, *ccmFc*, *cox3*, *matR*, *nad4*, *nad4L*, *nad9*, and *rpl5*. The phylogenetic tree was reconstructed using the Maximum Likelihood (ML) method in RAxML (v8.2.4) [23] with the model, and the partitioning scheme

Table 1 Summary of the sequences used in phylogenetic analysis

Family	Species Name	Accession Number of mitochondrial genomes	Length of mitochondrial genomes	GC content of mitochondrial genomes	Accession Number of plastid genomes	Length of plastid genomes	GC content of plastid genomes
Grossulariaceae	<i>Ribes nigrum</i>	NC_082306.1	450,227	45.80%	NC_080517.1	157,450	38.10%
Saxifragaceae	<i>Tiarella polyphylla</i>	OR866907-8.1	430,435	45.60%	NC_042928.1	154,902	37.70%
Saxifragaceae	<i>Heuchera parviflora</i>	KR559021.1	542,954	45.80%	KR478645.1	154,696	37.80%
Crassulaceae	<i>Sedum plumbizincicola</i>	NC_069572.1	212,159	44.50%	MN185459.1	149,397	37.70%
Crassulaceae	<i>Umbilicus rupestris</i>	OY787630.1	241,617	45.20%	MN794335.1	150,995	37.60%
Crassulaceae	<i>Rhodiola wallichiana</i>	OP312068-9.1	118,787/82,073	45.2%/44.9%	NC_060683.1	151,358	37.70%
Crassulaceae	<i>Rhodiola tangutica</i>	NC_072122.1	257,378	44.20%	NC_060605.1	151,707	37.80%
Crassulaceae	<i>Rhodiola crenulata</i>	NC_070303.1	194,106	45.20%	NC_050881.1	151,912	37.70%
Crassulaceae	<i>Rhodiola juparensis</i>	NC_082108.1	202,019	44.60%	NC_082109.1	151,073	37.70%
Crassulaceae	<i>Rhodiola sacra</i>	OP312070-1.1	128,593/80,447	45.3%/44.9%	NC_052735.1	150,941	37.80%
Hamamelidaceae	<i>Loropetalum chinense</i>	ON553696.1	758,747	45.90%	NC_060831.1	159,444	38%
Paeoniaceae	<i>Paeonia lactiflora</i>	NC_070189.1	181,688	46.20%	NC_040983.1	152,747	38.40%
Paeoniaceae	<i>Paeonia suffruticosa</i>	NC_084129.1	203,077	45.90%	NC_037879.1	153,119	38.40%
Cynomoriaceae	<i>Cynomorium coccineum</i>	KX270753-801.1	None	None	None	None	None
Vitaceae(outgroup)	<i>Vitis riparia x Vitis cinerea</i>	LR738918.1	755,068	44.30%	LR738917.1	161,008	37.40%
Vitaceae(outgroup)	<i>Vitis vinifera</i>	NC_012119.1	773,279	44.10%	NC_007957.1	160,928	37.40%

was obtained from PartitionFinder2 and 1000 bootstrap replicates. We also performed the Bayesian (BI) phylogenetic analysis using MrBayes (v3.2.7) [24] with the model and the partitioning scheme obtained from PartitionFinder2. Phylogenetic trees were displayed using the Interactive Tree of Life (iTOL v4) web server (<https://itol.embl.de>).

For phylogenetic reconstructions based on the plastid genomes, the sequences of the complete mitogenomes of thirteen species of Saxifragales were downloaded from the GenBank database (Table 1). These thirteen species' whole plastid genome sequences were aligned with MAFFT (v7.450). The phylogenetic tree was reconstructed using the Maximum Likelihood (ML) method in RAxML (v8.2.4) with the GTRGAMMA substitution model and 1000 bootstrap replicates. We also performed the BI phylogenetic analysis using MrBayes with the model parameters obtained from jModelTest (v2.1.0) [25]. Phylogenetic trees were displayed using the iTOL.

Results

General features of the *T. polyphylla* mitogenome

We obtained the assembled mitogenomes of *T. polyphylla*, three circular chromosomes with lengths of 430,435 bp, 126,943 bp, and 55,269 bp (Fig. 1). They were the dominant configurations of *T. polyphylla* mitogenome. The average Illumina sequencing depth of three mitochondrial chromosomes was estimated at 326×, 404×, and 313×, respectively (Figure S1A-C). The average Nanopore sequencing depth of three mitochondrial chromosomes was estimated at 32×, 41×, and 22×, respectively (Fig S1D-F).

T. polyphylla mitogenome encoded the conserved set of genes consisting of 36 protein-coding genes (three subunits of cytochrome oxidase, four cytochrome c biogenesis genes, nine subunits of NADH dehydrogenase, five subunits of ATP synthase, one cytochrome b gene, one transport membrane protein gene, and one maturases gene), three rRNA genes, and 16 tRNA genes (Table 2, Table S1). 27 protein-coding genes were intronless genes, and nine protein-coding genes were intron-containing gene. In the *T. polyphylla* mitogenome, the coding sequence of *rps3* overlaps with that of *rpl16*, with an overlap of 29 bp. This overlap is also observed in other plant mitogenomes [26, 27]. Chromosome 1 of the *T. polyphylla* mitogenome encoded 26 protein-coding genes and 15 tRNA genes. Chromosome 2 of the *T. polyphylla* mitogenome encoded nine protein-coding genes, one tRNA gene, and three rRNA genes. Chromosome 3 of the *T. polyphylla* mitogenome encoded four protein-coding genes and one tRNA gene (Fig. 1).

The plant mitochondrial tRNAs have been reshaped through a series of gene transfer events over time [28]. We identified the origins of the 17 tRNAs in the *T. polyphylla* mitogenome. Among the 17 tRNA genes identified, 12 were native mt-tRNA genes (Table S1). Four tRNAs (*trnN*-GUU, *trnD*-GUC, *trnW*-CCA, and *trnM*-CAU) originated from plastids, signifying the presence of intracellular gene transfers. Among them, only *trnD*-GUC was identified as the MTPT (mitochondrial plastid DNA) transferred from the plastid genome of *T. polyphylla*. This finding supports previous conclusions showing that plastid-derived tRNA genes have been retained for hundreds of millions of years [29, 30]. Additionally,

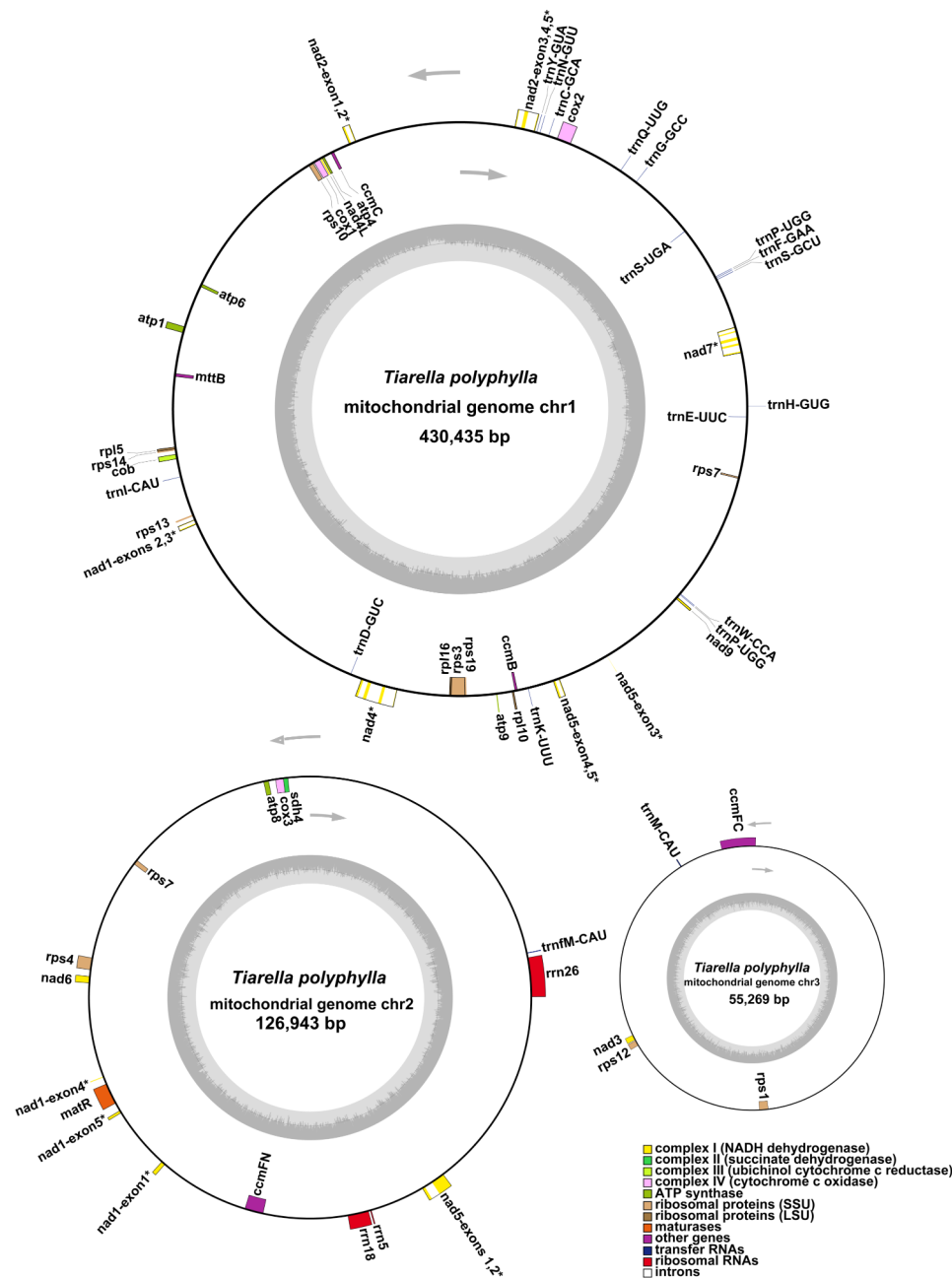


Fig. 1 Genome maps of the three circular chromosomes of *T. polyphylla* mitogenome. Genes shown on the inside were on the negative strand, whereas those on the outside were on the positive strand. Genes with introns were highlighted using “*”. The gray circle represents the GC contents. The circle inside the GC content graph marks the 50% threshold. The colors indicate different functional categories shown in the legend

trnP-UGG was identified as an unknown source gene, suggesting it may have originated from bacteria.

Repeat-mediated recombination in the *T. polyphylla* mitogenome

Plant mitogenome showed a high repeat-mediated recombinational activity, producing structurally dynamic genome configurations [31]. Homologous sequence analysis and Nanopore sequencing read showed that the

T. polyphylla mitogenome contains six recombinationally active repetitive sequences (Table S2, Figure S2). The lengths of the longest and shortest repetitive sequences of these six repetitive sequences were 9,703 bp and 350 bp. Four repetitive sequences mediated recombination within chromosome 1 (R01-R04), and the other two mediated recombination between chromosomes 1 and 2 (R05-R06). Five repetitive sequences were direct

Table 2 Summary of genes annotated in *T. polyphylla* mitogenome

Group of genes	Name of genes
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFC*, ccmFN</i>
Ubichinol cytochrome c reductase	<i>cob</i>
Cytochrome c oxidase	<i>cox1, cox2*, cox3</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1*, nad2*, nad3, nad4*, nad4L, nad5*, nad6, nad7*, nad9</i>
Ribosomal protein large subunit	<i>rpl5, rpl10, rpl16</i>
Ribosomal protein small subunit	<i>rps1, rps3*, rps4, rps7, rps10*, rps12, rps13, rps14, rps19</i>
Ribosomal RNAs	<i>rrn5, rrn18, rrn26</i>
Transfer RNA	<i>trnY-GUA, trnN-GUU, trnC-GCA, trnQ-UUG, trnG-GCC, trnP-UGG, trnF-GAA, trnS-GCU, trnH-GUG, trnE-UUC, trnW-CCA, trnK-UUU, trnI-CAU, trnM-CAU, trnM-CAU, trnP-UGG</i>

*** labeled the intron-containing genes

repeats (R01, R02, R04, R05, and R06), and one repetitive sequence was an inverted repeat (R03).

The relative abundance of alternative genome configurations generated by the six repetitive sequences ranged from 4.35 to 50%, quantified using the Nanopore sequencing reads (Table S2). Two longer repetitive sequences (R01 and R06) with lengths of 9,446 bp and 9,703 bp showed recombination frequencies of 21.43% and 45.45%, respectively. The shortest repetitive sequence (R05) had a lower recombination frequency of 4.35%. The repetitive sequence R02 exhibited the highest recombination frequency at 50.00%. A partial sequence of the *cox2* and *nad6* genes, located within the direct repetitive sequences R01 and R05, mediated genome recombination with frequencies of 21.43% and 4.35%, respectively (Table S2). This resulted in the duplication of *cox2* and *nad6* gene fragments within two copies of R01 and R05. Notably, the entire sequence of the *rps7* gene, found within the direct repetitive sequence R06, mediated genome recombination with a frequency of 45.45%, leading to the duplication of the *rps7* gene within two copies of R06 (Table S2).

Starting with a dominant configuration of three circular chromosomes and assuming each repeat pair recombines only once, we can predict the six alternative configurations (Fig. 2). Each alternative configuration maintained the same genomic content but varied in the number of chromosomes, ranging from an increase to four circular chromosomes to a decrease to two, diverging from the original three-chromosome structure.

SSRs and long tandem repeats in the *T. polyphylla* mitogenome

Simple sequence repeats (SSRs) with a tandem repeat motif of 1–6 nucleotides and long tandem repeats with a tandem repeat motif of >7 nucleotides have been used for genetic fingerprinting, phylogenetic analysis, and population genetic studies [32, 33]. Mitochondrial SSRs (mt-SSRs) have specific advantages over nuclear SSRs due to their uniparental inheritance, providing clearer insights into phylogenetic relationships between individuals [34]. In addition, mt-SSRs are particularly valuable when used as mitotype-specific markers to differentiate mitotypes or cytoplasm, such as distinguishing normal CMS lines from the lines with floral deformities in plant breeding [35]. A total of 208 SSR loci were mined in three chromosome sequences of the *T. polyphylla* mitogenome (Tables S3), with an average frequency of 0.34 SSR per kb. The total SSRs were 138, 36, and 34, identified in chromosomes 1, 2, and 3, respectively (Fig. 3). The mononucleotide repeat was the most abundant repeat type in the *T. polyphylla* mitogenome, comprising 30.29%, followed by tetra-, di-, tri-, penta, and hexa-nucleotide repeat that comprised 28.85%, 21.15%, 12.98%, 4.81%, and 1.92% respectively. The A/T motif was the most abundant mononucleotide repeat in the *T. polyphylla* mitogenome. Several genes with associated SSRs are featured (Table S4). Notably, the *rps3* gene exhibits the highest count of mt-SSRs, including a monomeric repeat of 11 thymine bases (T)11 and a tetranucleotide repeat (AATG)3. Additionally, the *nad1* gene possesses a monomeric repeat of 10 thymine bases (T)10. The *matR* gene is associated with a pentanucleotide microsatellite repeat (CTAGT)3, while the *rps1* gene contains a trinucleotide repeat (GAA)4.

A total of 25 long tandem repeats were identified in three chromosome sequences of the *T. polyphylla* mitogenome (Tables S5). 12, 7, and 3 long tandem repeats were identified in chromosomes 1, 2, and 3, respectively. Moreover, the length of all the tandem repeats in *T. polyphylla* mitogenome was <100 bp.

Mitochondrial plastid and nuclear DNA sequences in the *T. polyphylla* mitogenome

DNA transfers of MTPTs and mtDNAs to the nuclear genome (NUMTs) were documented and suggested to be signatures of the evolutionary processes [36, 37]. In the *T. polyphylla* mitogenome, we identified seven MTPTs listed in Table S6 and shown in Fig. 4A. The length of these transferred organelle genome copies ranged from 133 bp (MTPT03) to 2,023 bp (MTPT01). These transferred copies were 80.46% (MTPT04) to 100% (MTPT01), identical to their ptDNA counterparts. Most of these MTPTs were part of sequences of some protein-coding genes, indicating that they are non-functional. However, MTPT02 and MTPT03 contained the complete

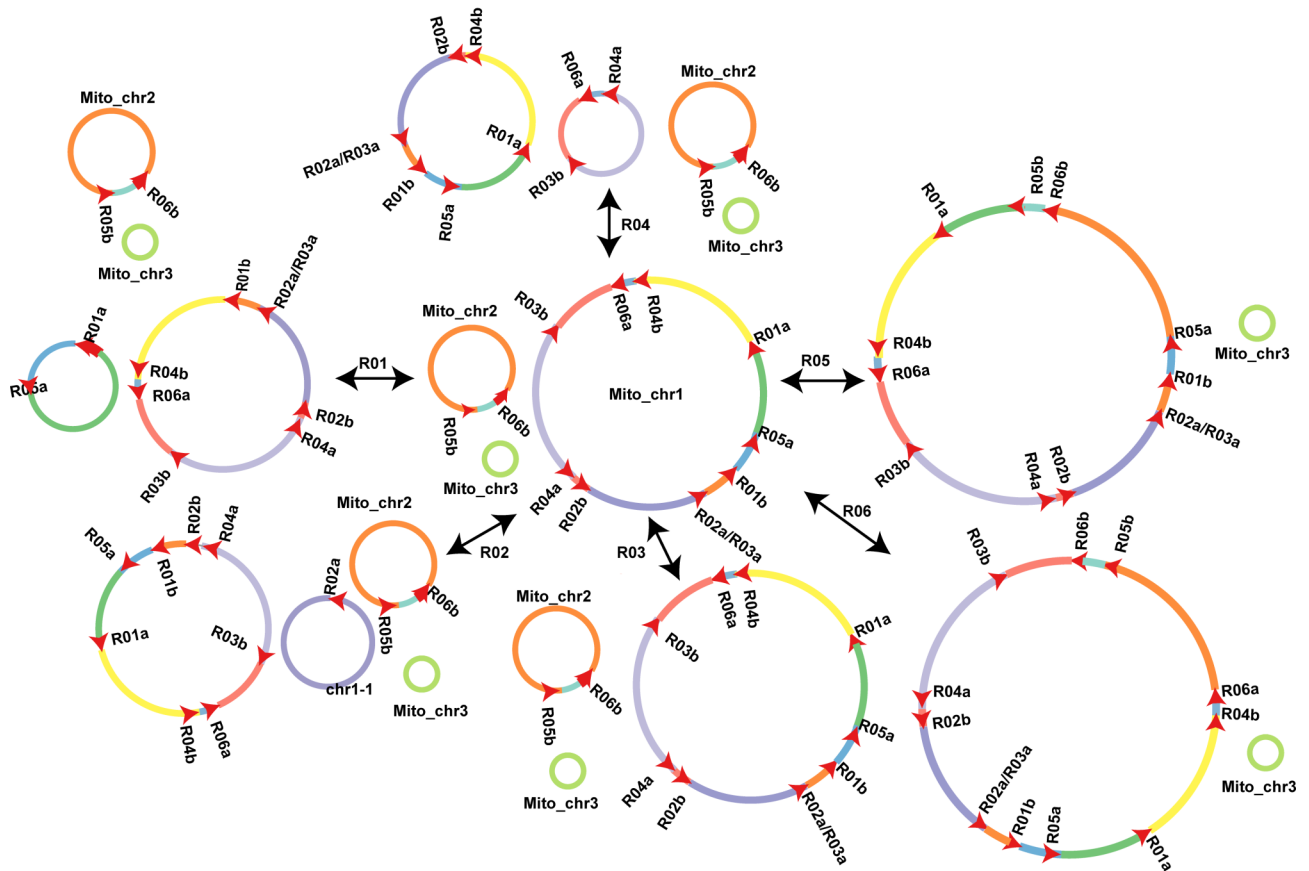


Fig. 2 Predicted alternative configurations (ACs) generated through recombination of six pairs of repetitive sequences in the *T. polyphylla* mitogenome. The dominant configuration of three circular chromosomes was shown in the middle (Mito_chr1/2/3: mitogenome chromosome 1/2/3). This figure assumes that each pair of repeats recombines only once, as the black arrows indicate. Red arrows on genome maps show six pairs of repetitive sequences, with the direction of each arrow indicating the orientation of these sequences. Single-copy genomic regions are represented in gradients of color, with consistent coloring across all configurations to represent each specific single-copy region

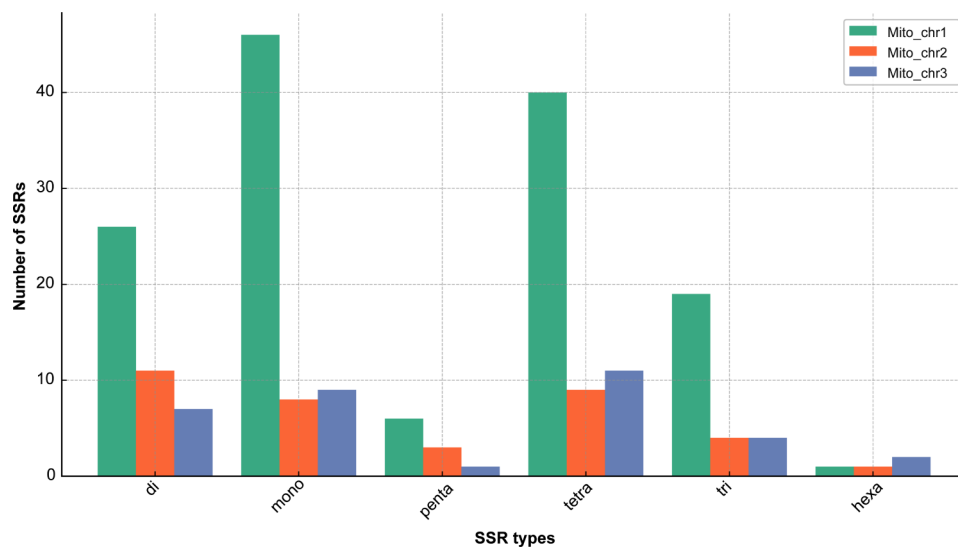


Fig. 3 Number of short tandem repeats of 1–6 base for SSRs in the three circular chromosomes of *T. polyphylla* mitogenome

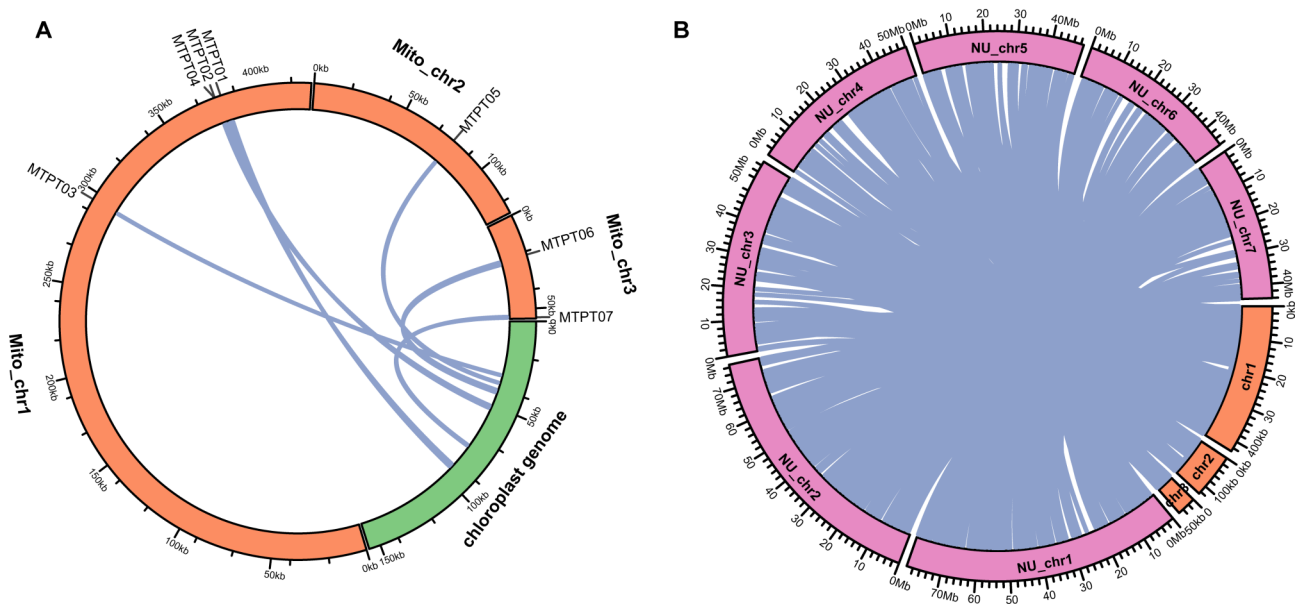


Fig. 4 Circos plot shows the seven MTPTs in the three circular chromosomes of *T. polyphylla* mitogenome (A) and the NUMTs in the seven chromosomes of *T. polyphylla* nuclear genome (B)

coding sequences of the *ndhK* and one cp-derived tRNA (*trnD*-GUC). Moreover, Nanopore long reads confirmed the presence of these seven MTPTs in the *T. polyphylla* mitogenome (Figure S3).

As depicted in Table S7 and Fig. 4B, we identified 1,012 NUMTs in the seven chromosomes of the *T. polyphylla* nuclear genome involving all three chromosomes of the *T. polyphylla* mitogenome. 646, 262, and 104 NUMTs transferred from chromosomes 1, 2, and 3 of the *T. polyphylla* mitogenome, respectively. The lengths of the longest and shortest sequences of these NUMTs were 24,772 bp and 34 bp. After removing the duplicate statistics, these NUMTs have a total length of 234,840 bp, accounting for 38.33% of the mitogenome. The divergence of the largest NUMTs in the seven nuclear chromosomes of *T. polyphylla* was relatively low (2.45–8.53%), indicating their recent origin. However, none of these largest NUMTs exhibited the lowest divergence among the NUMTs in each chromosome (Table S7, Figure S4).

RNA editing sites predicted in the *T. polyphylla* mitogenome

In plant organelles, post-transcriptionally RNA editing (numerous C-to-U and some U-to-C sites) is highly prevalent [38], resulting in changes in the coding sequences of transcripts and creating AUG start sites or eliminating premature stop codons or changes in the RNA structure [39]. The complete mitogenome of *T. polyphylla* contained 37 protein-coding genes, and 653 edited sites were predicted for all these genes (Table S8, Fig. 5A). The frequency of editing across these genes ranged from 0.15% (1 site in 303 bp) for *rps14* to 7.35% (48 sites in 1,488 bp)

for *nad4* (Table S8). RNA editing produced new start codons for *cox1*, *nad1*, and *nad4L* and new stop codons for *atp6*, *atp9*, *ccmFc*, and *rps10*.

The most frequent amino acid changes are Pro to Leu, occurring at 151 sites (23.12% of the changes), followed by Ser to Leu at 137 sites (20.98%), and Ser to Phe at 92 sites (14.09%) (Fig. 5B). In the *T. polyphylla* mitogenome, the proportion of RNA editing at silent and nonsilent editing sites was biased. 33 of 653 editing sites were silent (5.05%), and 620 were nonsilent sites (94.95%).

The phylogeny analysis based on the mitogenomes and plastid genomes of Saxifragales species

The mitogenome has been one of the informative loci of molecular evolutionary biology [40]. Phylogenetic analysis based on coding sequences of twelve protein-coding genes of the fourteen Saxifragales species was conducted in this study using the ML and BI methods. The results suggested robust phylogenetic inferences, with many nodes having ML bootstrap support values >80 and BI posterior probabilities = 1 (Fig. 6). We also performed the phylogenetic analysis based on the complete plastid genomes of the same species except for one species that didn't have the available plastid genome in Genbank. We found the incongruence between mitochondrion and plastid datasets. Paeniaceae was sister to Hamamelidaceae in the plastid tree (Fig. 6B), the same as the APG IV system. However, it was sister to Cynomoriaceae in the mitochondrion tree (Fig. 6A). And we also found the congruence between mitochondrion and plastid datasets. Saxifragaceae was sister to Grossulariaceae in both mitochondrion and plastid tree.

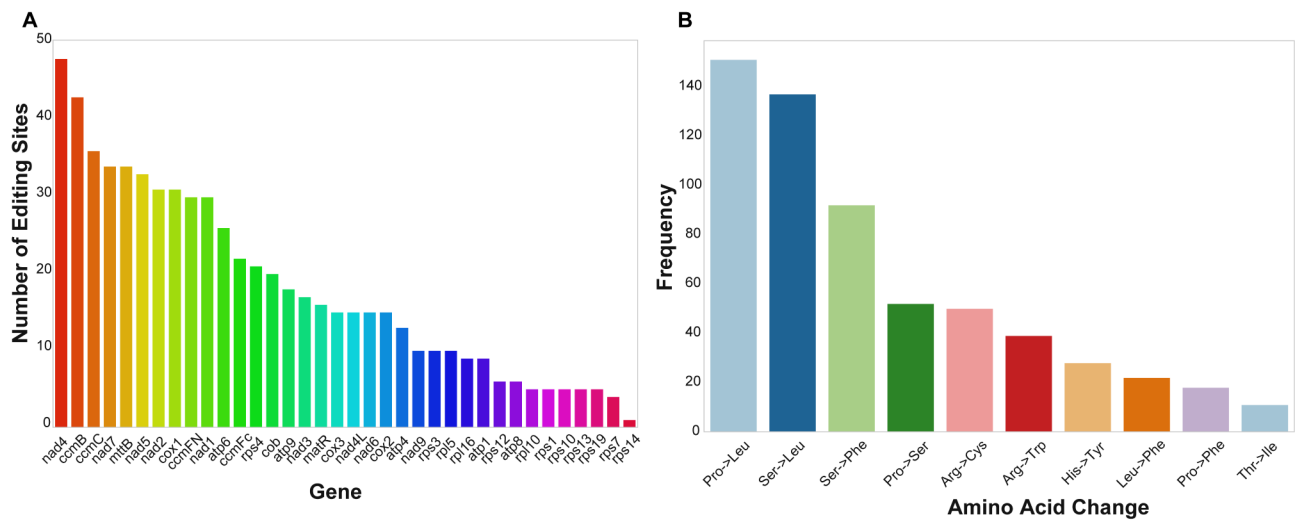


Fig. 5 Bar plots show the numbers of RNA editing sites of protein-coding genes (A) and the frequency of the top ten amino changes generated by the RNA editing events (B)

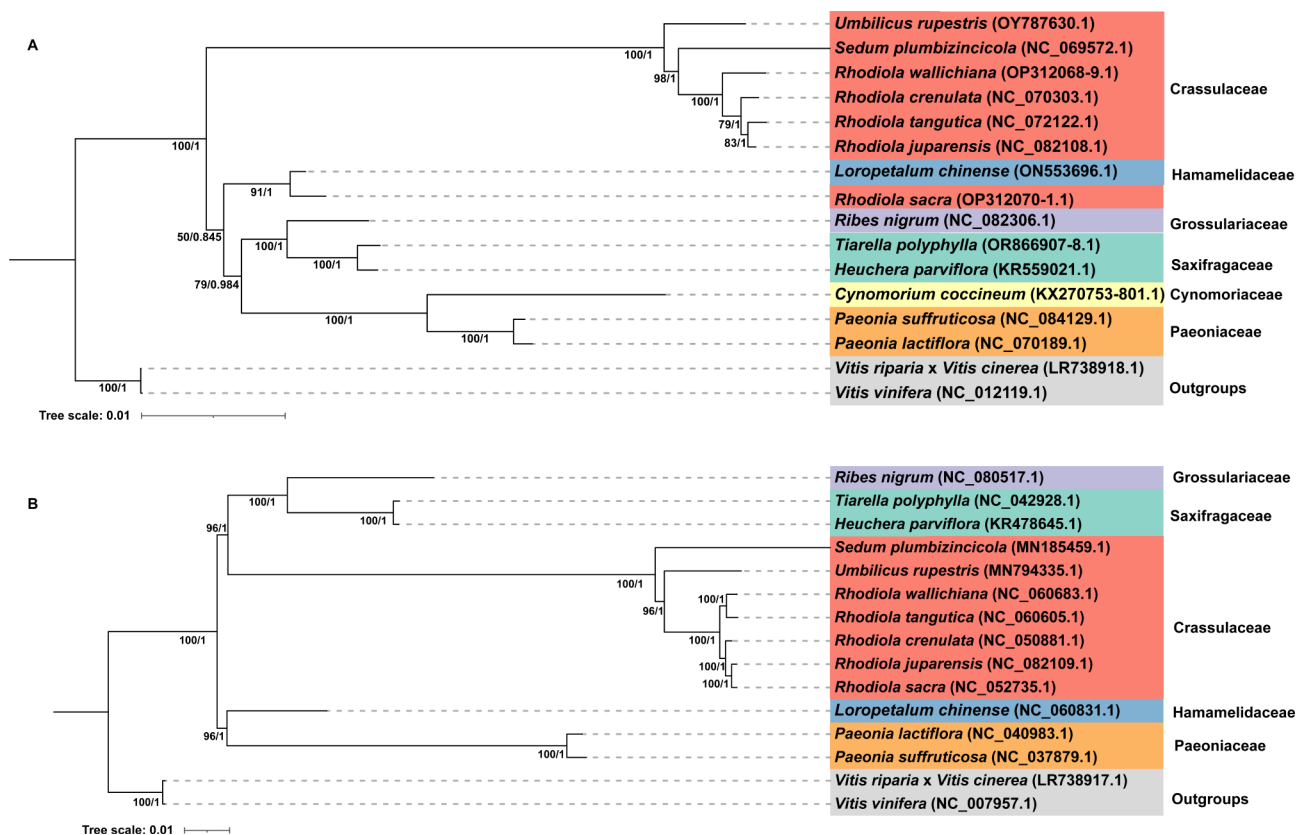


Fig. 6 Phylogenetic trees reconstructed using the coding sequences of 26 conserved genes of the seven species of Saxifragales. The ML bootstrap support values and BI posterior probabilities were labeled. Two species of superasterids, including *Chrysanthemum boreale* (NC_039757.1) and *Synotis nagensium* (NC_082299.1), were used as outgroups

Discussions

Multi-chromosome organizations of *T. polyphylla* mitogenome

Although often assembled as one master circle, the plant mitogenome structure physically existed as a variety of

circles, linear molecules, and complex branching structures [41]. Moreover, a few cases in the genus *Silene* L., *Cucumis sativus* L., and sugarcane identified the mitogenome consisting of complex multichromosomal structures instead of finding a single master circle [31, 42].

Our study showed that three circular-mapping chromosomes were also found to be the dominant configuration of the *T. polyphylla* mitogenome, and our sequencing data did not support the presence of a master circle.

In the mitogenome of *Silene* species and *Cucumis sativus*, no repetitive sequences that could mediate recombination were found between at least two circular chromosomes. Similar to our study, the Nanopore long reads only supported two repetitive sequences (R05 and R06) that mediated the inter-chromosomal recombination in chromosomes 1 and 2.

Imbalanced stoichiometry in recombination products of repetitive sequences in *T. polyphylla* mitogenome

Frequent homologous recombination involving repetitive sequences could generate recombination products as the alternative genome configurations in plant mitogenome [43]. These recombination products showed balanced and imbalanced stoichiometry, indicating reciprocal and non-reciprocal recombination events [44]. In the *T. polyphylla* mitogenome, one of the six repetitive sequences (R02) that mediate homologous recombination displayed precisely balanced stoichiometry, with an equal number of reads supporting both the principal and alternative configurations, resulting in a 50% recombination frequency. The occurrence of imbalanced stoichiometry in recombination products with more reads supporting the major configuration was observed in most of the repetitive sequences (R01, R03, 04, 05, and 06), with recombination frequencies varying between 4.35% and 47.06%. These non-reciprocal recombination events of repeat pairs were also found in the mitogenomes of *Oenothera* species and *Mimulus guttatus* [17, 45].

The mechanism underlying these reciprocal and non-reciprocal homologous recombination events was complex. Specifically, non-reciprocal recombination events may be induced by the break-induced replication pathway within the homologous recombination process [46]. Therefore, we postulated that non-reciprocal homologous recombination events in *T. polyphylla* mitogenome were associated with the break-induced replication of homologous recombination. Moreover, nuclear genes involving the DNA repair pathways by homologous recombination and mismatch repair modulated these processes [43]. For example, it has been reported in *Arabidopsis* that OSB1 (for Organellar Single-stranded DNA Binding protein 1) played a role in controlling the stoichiometry of alternative mitogenome configurations generated by homologous recombination [47]. Further empirical research is essential to elucidate these processes fully in the *T. polyphylla* mitogenome in the future.

DNA transfers of cp-derived mtDNAs (MTPTs) and mtDNAs to nuclear genome (NUMTs)

MTPTs have been shown to transfer randomly from any position in the ptDNAs, and some of the tRNA transferred from the ptDNA was still functional in mtDNA [29]. In agreement with these conclusions, the MTPTs in the *T. polyphylla* mitogenome were randomly distributed across the plastome, and the pt-derived tRNA sequence (trnD-GUC) was intact, suggesting its function. However, in this study, the coding sequence of the *ndhK* protein-coding gene was intact, which appears to differ from the results, indicating that all the protein-coding MTPTs degenerate [29].

It has been recognized that some large mtDNA segments are transferred into nuclear genomes through evolution [48]. For example, a 620-kb long NUMT was found in *Arabidopsis thaliana* [49]. The present study found a large mtDNA segment, 24.7 kb in length, in the *T. polyphylla* nuclear genome.

RNA editing events were conserved

RNA editing in plant organelle plays a critical role in correct mtDNA expression [50]. In our study, new start codons for *cox1*, *nad1*, and *nad4L* and new stop codons for *atp6*, *atp9*, *ccmFc*, and *rps10* were generated by RNA editing. These start and stop gain RNA editing events were described previously [51–53]. It has been shown that the most frequent amino acid changes are Pro to Leu, Ser to Leu, and Ser to Phe [54, 55], similar to this study's findings.

Conclusions

The multichromosomal structure of the *T. polyphylla* mitogenome described here offers unique insights into both variable structures and the recombination events mediated by the repetitive sequences, achieving a better understanding of the mitogenome of Saxifragaceae species. Using sequences of our *T. polyphylla* mitogenome and six other Saxifragales species, we constructed the resolved phylogenetic tree consistent with the APG IV system. The investigation of the *T. polyphylla* mitogenome expanded our genomic resources to investigate the structure and evolution of the plant mitogenome.

Abbreviations

CMS	Cytoplasmic Male Sterility
NGS	Next-Generation Sequencing
MTPTs	Mitochondrial Plastid DNAs
NUMTs	Nuclear Mitochondrial DNAs
tRNA	Transfer RNA
rRNA	Ribosomal RNA
C	TCytidines
U	Uridines

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10790-5>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8

Acknowledgements

We sincerely thank Yonghua Zhang and Xinjie Jin from Wenzhou University for the help with mitochondrial genome assembly and annotation.

Author contributions

PL and LXL designed the study. BL and QL conducted the field sampling. BL, QL and WWL produced and analyzed the data. BL and QL wrote the manuscript. YS, PL and LXL revised the manuscript. All authors approved the final manuscript.

Funding

This research was supported by Xinyang Academy of Ecological Research Open Foundation (Grant No. 2023DBS09), the National Natural Science Foundation of China (Grant No. 31900188).

Data availability

The raw sequencing reads for *T. polyphylla* genome assembly have been deposited in the NCBI database with BioProject ID PRJNA870970. The Nanopore and Illumina raw data are available under the accession nos. SRR21133002 and SRR21133003. The *T. polyphylla* mitogenome assembly has been deposited in the NCBI database with the accession nos. OR866907, OR866908, and OR866909.

Declarations

Ethics approval and consent to participate

All plant materials were collected following national and international standards and local laws and regulations. No specific permission is required to collect all samples described in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 December 2023 / Accepted: 10 September 2024

Published online: 19 September 2024

References

1. Ying TS. The floristic relationships of the temperate forest regions of China and the United States. *Ann Mo Bot Gard.* 1983;70:597–604.
2. Lee MY, Ahn KS, Lim HS, Yuk JE, Kwon OK, Lee KY, Lee HK, Oh SR. Tiarelic acid attenuates airway hyperresponsiveness and inflammation in a murine model of allergic asthma. *Int Immunopharmacol.* 2012;12(1):117–24.
3. Shen G, Oh SR, Min BS, Lee J, Ahn KS, Kim YH, Lee HK. Phytochemical investigation of *Tiarella Polyphylla*. *Arch Pharm Res.* 2008;31(1):10–6.
4. Moon HI, Lee J, Zee OP, Chung JH. Triterpenoid from *Tiarella polyphylla*, regulation of type 1 procollagen and MMP-1 in ultraviolet irradiation of cultured old age human dermal fibroblasts. *Arch Pharm Res.* 2004;27(10):1060–4.
5. Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havidr JC. Cytonuclear integration and co-evolution. *Nat Rev Genet.* 2018;19(10):635–48.
6. Liu L, Chen M, Folk RA, Wang M, Zhao T, Shang F, Soltis DE, Li P. Phylogenomic and syntenic data demonstrate complex evolutionary processes in early radiation of the rosids. *Mol Ecol Resour.* 2023;23(7):1673–88.
7. Liu L, Wang Y, He P, Li P, Lee J, Soltis DE, Fu C. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics.* 2018;19(1):235.
8. Small ID, Isaac PG, Leaver CJ. Stoichiometric differences in DNA molecules containing the atpA gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *Embo J.* 1987;6(4):865–9.
9. 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.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
11. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S: The sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
12. Lee E, Harris N, Gibson M, Chetty R, Lewis S. Apollo: a community resource for genome annotation editing. *Bioinformatics.* 2009;25(14):1836–7.
13. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: Identification, Incidence and Evolution. *G3 (Bethesda).* 2019;9(2):549–559.
14. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D. Tablet—next generation sequence assembly visualization. *Bioinformatics.* 2010;26(3):401–2.
15. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33(16):2583–5.
16. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27(2):573–80.
17. 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.
18. 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.
19. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Deepred-Mt: deep representation learning for predicting C-to-U RNA editing in plant mitochondria. *Comput Biol Med.* 2021;136:104682.
20. 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.
21. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20(4):1160–6.
22. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 2017;34(3):772–3.
23. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
24. 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.
25. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9(8):772–772.
26. Forgione I, Bonavita S, Regina TMR. Mitochondria of *Cedrus Atlantica* and allied species: a new chapter in the horizontal gene transfer history. *Plant Sci.* 2019;281:93–101.
27. Bonavita S, Regina TMR. The evolutionary conservation of rps3 introns and rps19-rps3-rpl16 gene cluster in *Adiantum capillus-veneris* mitochondria. *Curr Genet.* 2015;62(1):173–84.
28. Warren JM, Sloan DB. Interchangeable parts: the evolutionarily dynamic tRNA population in plant mitochondria. *Mitochondrion.* 2020;52:144–56.
29. Wang D, Wu YW, Shih AC, Wu CS, Wang YN, Chaw SM. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Mol Biol Evol.* 2007;24(9):2040–8.

30. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. The fossilized mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 2013;11:29.
31. Sloan DB, Alverson AJ, Chukalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 2012;10(1):e1001241.
32. Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, et al. A simple sequence repeat-based linkage map of barley. *Genetics.* 2000;156(4):1997–2005.
33. Vincens MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science.* 2009;324(5931):1213–6.
34. Kuntal H, Sharma V. In silico analysis of SSRs in mitochondrial genomes of plants. *OMICS.* 2011;15(11):783–9.
35. Singh S, Bhatia R, Kumar R, Behera TK, Kumari K, Pramanik A, Ghemeray H, Sharma K, Bhattacharya RC, Dey SS. Elucidating mitochondrial DNA markers of Ogura-based CMS lines in Indian Cauliflowers (*Brassica oleracea* var. *botrytis* L.) and their floral abnormalities due to Diversity in Cytonuclear interactions. *Front Plant Sci.* 2021;12:631489.
36. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet.* 2004;46(3):123–39.
37. Huang CY, Grünheit N, Ahmadinejad N, Timmis JN, Martin W. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to Angiosperm nuclear chromosomes. *Plant Physiol.* 2005;138(3):1723–33.
38. Edera AA, Gandini CL, Sanchez-Puerta MV. Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol Biol.* 2018;97(3):215–31.
39. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.* 2020;101(5):1040–56.
40. Mueller RL. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in bayesian phylogenetic analysis. *Syst Biol.* 2006;55(2):289–300.
41. Backert S, Börner T. Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L). *Curr Genet.* 2000;37(5):304–14.
42. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell.* 2011;23(7):2499–513.
43. Gualberto JM, Newton KJ. Plant mitochondrial genomes: Dynamics and mechanisms of Mutation. *Annu Rev Plant Biol.* 2017;68:225–52.
44. Lonsdale D, Brears T, Hodge T, Melville SE, Rottmann W. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. *Philos Trans R Soc Lond B Biol Sci.* 1988;319(1193):149–63.
45. Mower JP, Case AL, Floro ER, Willis JH. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biol Evol.* 2012;4(5):670–86.
46. Llorente B, Smith CE, Symington LS. Break-induced replication: what is it and what is it for? *Cell cycle (Georgetown, Tex.)*. 2008;7(7):859–64.
47. Zaegel V, Guermann B, Ret ML, Andrés C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P. The plant-specific ssDNA binding protein OSB1 is involved in the Stoichiometric transmission of mitochondrial DNA in Arabidopsis. *Plant Cell.* 2006;18(12):3548–63.
48. Noutsos C, Richly E, Leister D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res.* 2005;15(5):616–28.
49. Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 2004;5(2):123–35.
50. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.* 2020;101(5):1040–56.
51. Bi C, Lu N, Xu Y, He C, Lu Z. Characterization and Analysis of the Mitochondrial Genome of Common Bean (*Phaseolus vulgaris*) by Comparative Genomic Approaches. In: *Int J Mol Sci.* Edited by Bi C, Lu N, Xu Y, He C, Lu Z, vol. 21; 2020.
52. Chen TC, Su YY, Wu CH, Liu YC, Huang CH, Chang CC. Analysis of mitochondrial genomics and transcriptomics reveal abundant RNA edits and differential editing status in moth orchid, *Phalaenopsis Aphrodite* subsp. *formosana*. *Sci Hortic.* 2020;267:109304.
53. Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B, Knoop V. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 2011;39(7):2890–902.
54. Bonnard G, Gualberto JM, Lamattina L, Grienberger JM, Brennlcke A. RNA editing in plant mitochondria. *Crit Rev Plant Sci.* 1992;10(6):503–24.
55. Freyer R, Kiefer-Meyer MC, Kössel H. Occurrence of plastid RNA editing in all major lineages of land plants. *Proc Natl Acad Sci U S A.* 1997;94(12):6285–90.

Publisher's note

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