

Original Research

Assembly and Characterization of the Mitochondrial Genome of *Fallopia aubertii* (L. Henry) Holub

Xin Zhang^{1,2}, Yong-Chang Lu^{1,2,*}, Jiu-Li Wang^{3,*} 

¹Qinghai Key Laboratory of Phytochemistry on Qinghai-Tibet Plateau, Qinghai Minzu University, 810007 Xining, Qinghai, China

²College of Pharmacy, Qinghai Minzu University, 810007 Xining, Qinghai, China

³Country Qinghai Provincial Key Laboratory of High-value Utilization of Characteristic Economic Plants, Qinghai Minzu University, 810007 Xining, Qinghai, China

*Correspondence: qhlych@126.com (Yong-Chang Lu); wang_jiul@163.com (Jiu-Li Wang)

Academic Editor: Naeem Khan

Submitted: 12 June 2023 Revised: 29 August 2023 Accepted: 8 September 2023 Published: 28 September 2023

Abstract

Background: *Fallopia aubertii* (L. Henry) Holub is a perennial semi-shrub with both ornamental and medicinal value. The mitochondrial genomes of plants contain valuable genetic traits that can be utilized for the exploitation of genetic resources. The parsing of *F. aubertii* mitochondrial genome can provide insight into the role of mitochondria in plant growth and development, metabolism regulation, evolution, and response to environmental stress. **Methods:** In this study, we sequenced the mitochondrial genome of *F. aubertii* using the Illumina NovaSeq 6000 platform and Nanopore platform. We conducted a comprehensive analysis of the mitochondrial genome of *F. aubertii*, which involved examining various aspects such as gene composition, repetitive sequences, RNA editing sites, phylogeny, and organelle genome homology. To achieve this, we employed several bioinformatics methods including sequence alignment analysis, repetitive sequence analysis, phylogeny analysis, and more. **Results:** The mitochondrial genome of *F. aubertii* has 64 genes, including 34 protein-coding genes (PCGs), three rRNAs, and 27 tRNAs. There were 77 short tandem repeat sequences detected in the mitochondrial genome, five tandem repeat sequences identified by Tandem Repeats Finder (TRF), and 50 scattered repeat sequences observed, including 22 forward repeat sequences and 28 palindrome repeat sequences. A total of 367 RNA coding sites were predicted in PCGs, with the highest number (33) found within *ccmB*. *Ka/Ks* values estimated for mitochondrial genes of *F. aubertii* and three closely related species representing Caryophyllales were less than 1 for most of the genes. The maximum likelihood evolutionary tree showed that *F. aubertii* and *Nepenthes × ventrata* are most closely related. **Conclusions:** In this study, we obtained basic information on the mitochondrial genome of *F. aubertii* and this study investigated repeat sequences and homologous segments, predicted RNA editing sites, and utilized the *Ka/Ks* ratio to estimate the selection pressure on mitochondrial genes of *F. aubertii*. We also discussed the systematic evolutionary position of *F. aubertii* based on mitochondrial genome sequences. Our study revealed variations in the sequence and structure of mitochondrial genomes in Caryophyllales. These findings are of great significance for identifying and improving valuable plant traits and serve as a reference for future molecular studies of *F. aubertii*.

Keywords: Caryophyllales; organellar genomes; comparative genomics; evolution; phylogeny

1. Introduction

The mitochondrial genome, which is circular or linear DNA [1,2], usually contains dozens of genes that encode proteins and RNA molecules involved in the regulation of mitochondrial function and morphology. The primary function of mitochondria is the production of adenosine triphosphate (ATP), which provides energy to cells [3]. Mitochondrial DNA plays a crucial role in metabolic processes like respiration and photosynthesis, which have a direct or indirect impact on the growth and development of organisms [4]. However, mutations, deletions, or insertions in mitochondrial genes can alter this effect [5–7]. The parsing of higher plant mitochondrial genes can provide insight into the role of mitochondria in plant growth and development, metabolism regulation, and response to environmental stress [8,9]. Similarly, the mitochondrial genome contains valuable information that can be utilized for the devel-

opment of molecular markers, genetic engineering, and elucidation of the phylogenetic and evolutionary connections among plant species [10,11].

Fallopia aubertii (L. Henry) Holub is a perennial semi-shrub or deciduous vine plant that is widely distributed in Asia, Europe, and North America [12–14]. It is a heliophilous plant with strong adaptability but is intolerant of shade or water-logging [15], and is commonly found on slopes, riverbanks, forest edges, and beaches. Regarding population ecology, *F. aubertii* is a relatively dispersed species that is malleable and adaptable to various environments. In recent years, researchers have studied the growth and development, genetic variation, chemical constituents, and pharmacodynamic effects of *F. aubertii* [14,16,17]. Our research team has isolated multiple compounds from *F. aubertii*, studying its anti-gout efficacy, and sequenced and annotated its chloroplast genome [18–20].



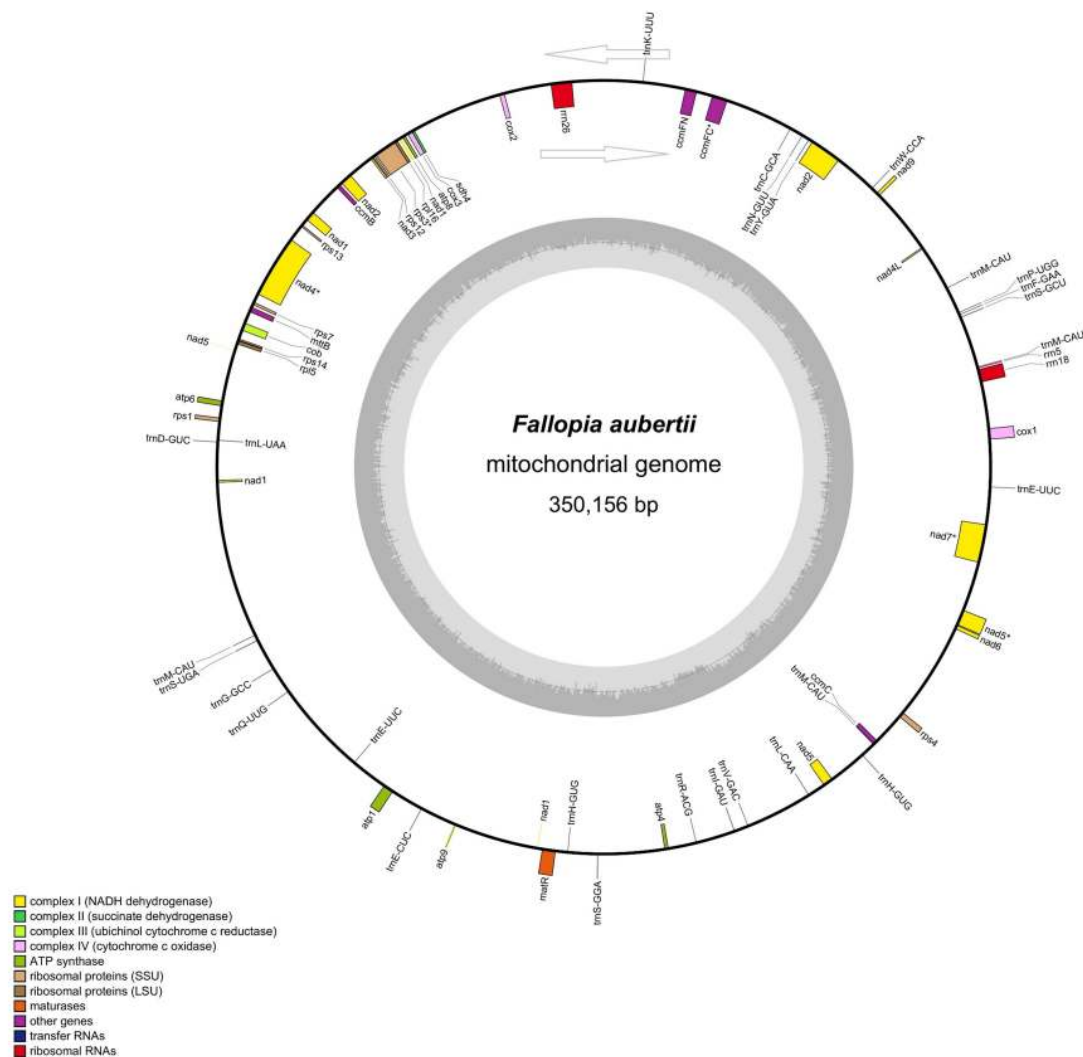


Fig. 1. A circular map of the *F. aubertii* mt genome. The counterclockwise genes are located on the outer side of the loop, whereas the clockwise genes are located on the inner side of the circle. The diagram shows the names of the contained genes. Colors are used to represent the different functional groups of genes. The grey circles inside represent GC content. The asterisks (*) represent intron-containing genes.

In this study, we utilize high-throughput sequencing technology and bioinformatics methods to analyze the mitochondrial genome features of *F. aubertii*. In addition, we reconstruct the phylogenetic relationships among *F. aubertii* and its closest relatives by using the mitochondrial genome sequence. The current study provides a scientific foundation and a point of reference for the exploitation and development of the species' resources, as well as the preservation of its biological diversity.

2. Materials and Methods

2.1 Sample Collection and DNA Sequencing

The sample of *F. aubertii* was collected from the campus of Qinghai Minzu University in Qinghai, China (36.59 °N, 101.82 °E) and identified by Professor Yong-chang Lu.

The fresh young leaves were dried in silica gel and stored at −20 °C. Genomic DNA was extracted from the whole sample of *F. aubertii* using the modified CTAB method and was subsequently evaluated for purity and integrity using agarose gel electrophoresis and a NanoDrop 2000c spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The classified DNA samples were sent to Shanghai Origine Biotechnology Pharmaceutical Technology Co., Ltd., where the genome DNA was sequenced using the Illumina NovaSeq 6000 platform (BIOZERON Co., Ltd., Shanghai, China) and Nanopore platform (Oxford Nanopore Technologies, Oxford, UK).

2.2 Assembly and Annotation of Mitochondrial Genome

We used GetOrganelle v1.7.1 (Max Planck Institute of Molecular Plant Physiology, Munich, Germany) [21] to

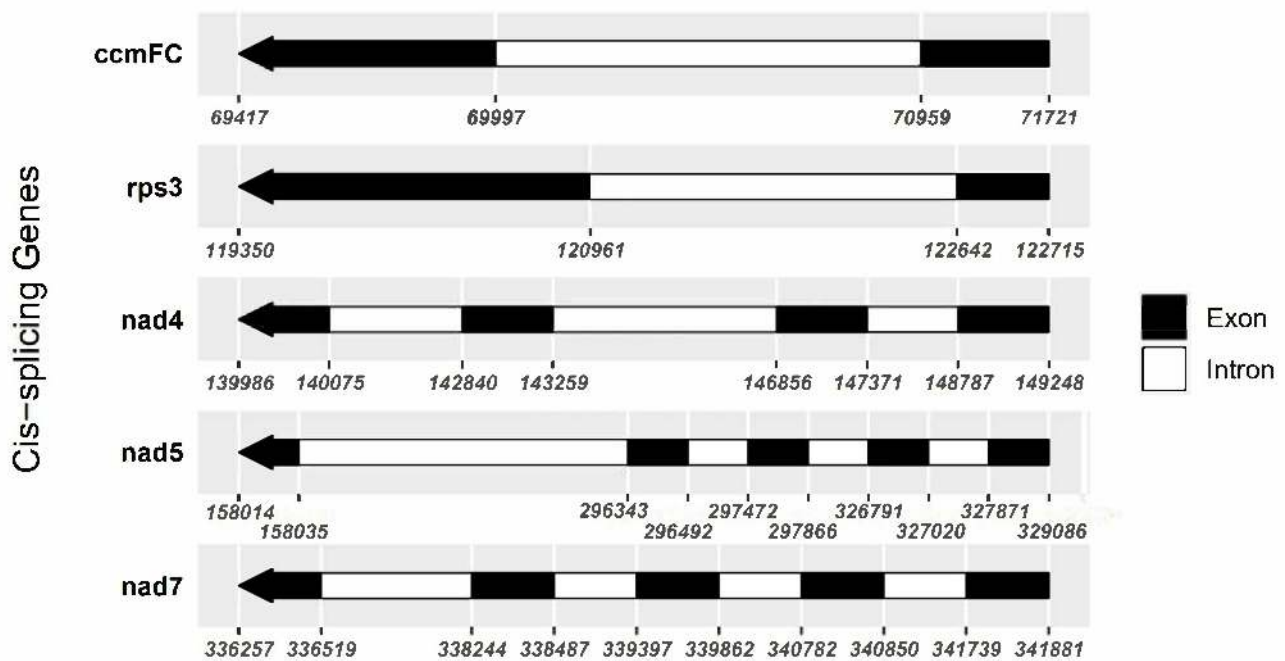


Fig. 2. Schematic map of the cis-splicing genes. Black blocks represent exons, and white blocks represent introns. The numbers beneath the protein-coding genes indicate the locations of introns and exons. One intron is present in both *ccmFC* and *rps3*. Three introns are in *nad4* and four in *nad5* and *nad7*.

Table 1. Characteristics of *F. aubertii* mt genome, CDS and Non-coding region.

Feature	A (%)	T (%)	G (%)	C (%)	GC (%)	Size (bp)	Proportion in Genome (%)
Genome	27.69	27.58	22.2	22.53	44.73	350,156	100
CDS	26.64	31.39	21.39	20.58	41.97	30,276	8.65
Cis-spliced intron	25.64	23.14	26.94	24.27	51.22	22,868	6.53
tRNA	23.05	26.34	28.28	22.32	50.61	2065	0.59
rRNA	26.18	22.41	28.82	22.59	51.41	5153	1.47
Non-coding region	27.85	27.81	22.08	22.26	44.34	289,955	82.81

Note: CDS: coding sequence.

perform *de novo* assembly of the mitochondrial genome. The mitochondrial genome of *Bougainvillea spectabilis* (GenBank Accession Number: MW167296), a closely related species, was used as the reference sequence. To select the potential mitochondrial reads, we used BLAST searches against the mitogenome of *B. spectabilis* and the GetOrganelle results from a pool of Illumina reads. The mitochondrial Illumina reads were assembled into ~50 mitogenome contigs using SPAdes v3.14.1 (Algorithmic Biology Lab, St. Petersburg Academic University of the Russian Academy of Sciences, St Petersburg, Russia) [22]. Nanopore reads were aligned against the GetOrganelle and SPAdes assembled scaffolds using BWA v0.7.1 (Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK) [23]. The aligned Nanopore reads were extracted to perform *de novo* assembly of the mitochondrial genome using Canu v2.0, and then used Pilon v1.23 (<https://github.com/broadinstitute/pilon/>; Broad Institute of MIT and Harvard, Cambridge, MA,

USA) [24] for error correction. The genes were predicted based on the method of homology alignment prediction. The coding genes, tRNAs, rRNAs, and possible pseudogenes were annotated by Blast+ 2.7.1 and tRNAscan-SE v2.0.7 (Todd Lowe Lab, Dept. of Biomolecular Engineering, School of Engineering University of California, Santa Cruz, CA, USA) [25]. tRNAscan-SE was also used to predict the secondary structure of tRNAs in *F. aubertii*. The mitochondrial genome map was drawn using OGDRAW (<https://chlorobox.mpimgolm.mpg.de/OGDraw.html>) [26].

2.3 Analysis of Mitochondrial Genome Repeat Sequences

Tandem repeat sequences in the genome were identified by Tandem Repeats Finder (TRF) [27] (<https://tandem.bu.edu/trf/trf.html>) with the following parameters: match +2, mismatch -7, indel -7. We reported only the repeats with scores greater than 50. Identification of short tandem repeat (STR) or simple sequence repeats

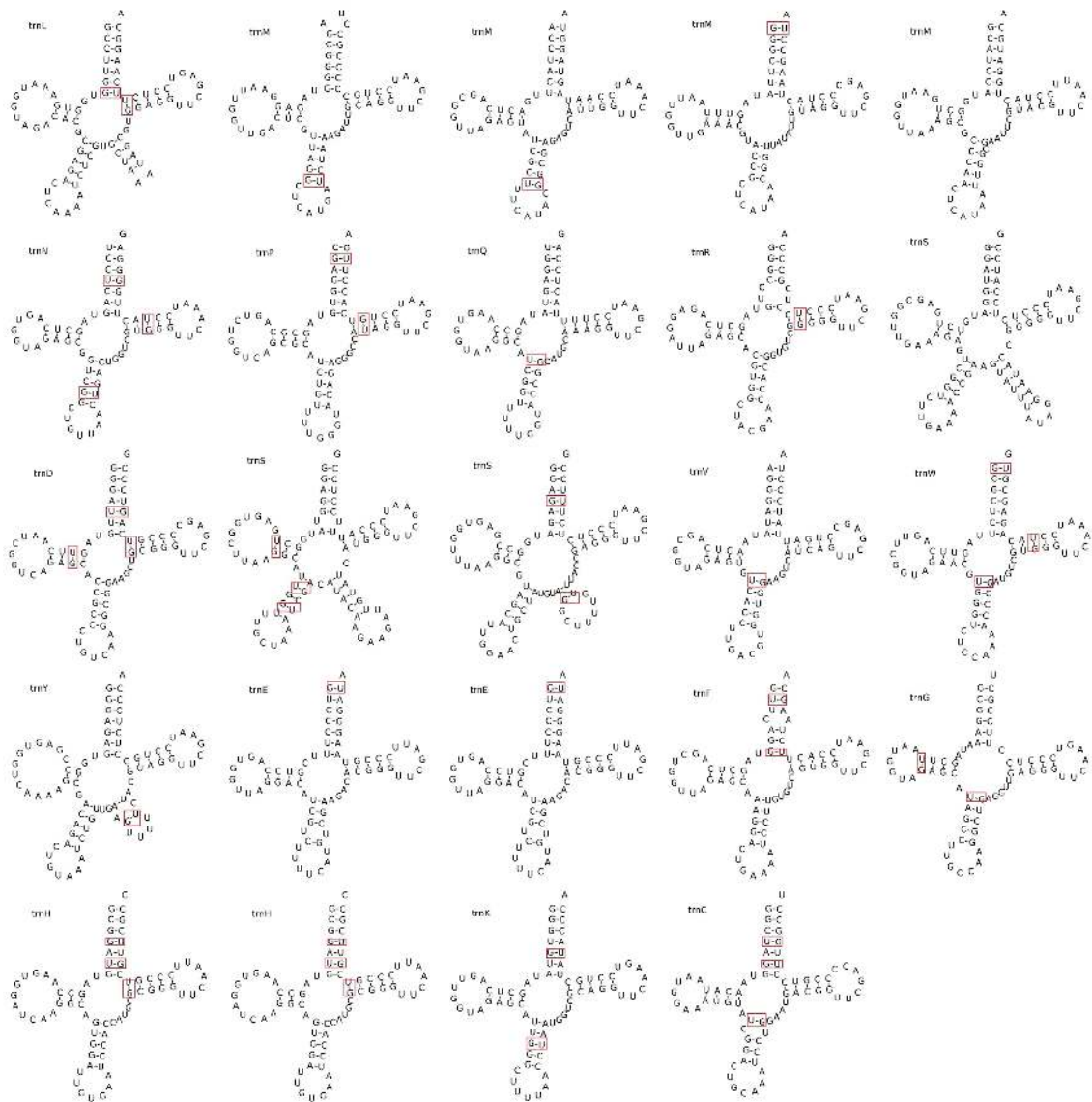


Fig. 3. tRNA secondary structure prediction. Twenty-four tRNA secondary structures were identified in the mitochondrial genome of *F. aubertii*. The top left of the structure is the amino acid abbreviation letter. Base mismatches exist, usually U-C base mismatches. The mismatched bases are marked in red boxes.

(SSR) was performed using MISA v2.1 MISA (Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany) [28] with the following parameters: 1–10, 2–5, 3–4, 4–3, 5–3, 6–3. The REPuter online version [29] (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) was used to search scattered repeats in the mitochondrial genome with the following parameters: hamming distance-3, maximum computed repeats-50, and minimal repeat size-8.

2.4 Analysis of RNA Editing

To predict the RNA-editing sites within the mitochondrial genes of *F. aubertii* the online PmtREP (<http://www.genepioneer.com/>) was used, with a threshold value of 0.2 for the mitochondrial sequences.

2.5 Analysis of GC Content and GC Skew

We applied the cloud platform (<http://www.genepioneer.com/>) to analyze the GC content of the *F. aubertii* mitochondrial coding sequence (CDS). To visualize the GC skew, we uploaded the CDS of *F. aubertii* to the online Proksee [30] tool and selected the “GC Content” and “GC Skew” tabs.

2.6 Selection Pressure Analysis

The ratio of nonsynonymous substitutions (K_a) to synonymous substitutions (K_s) of protein-coding genes (PCGs) between *F. aubertii* and three closely related species (*Nepenthes × ventrata* (MH798871.1), *Agrostemma githago* (MW553037.1), and *Tetragonia*

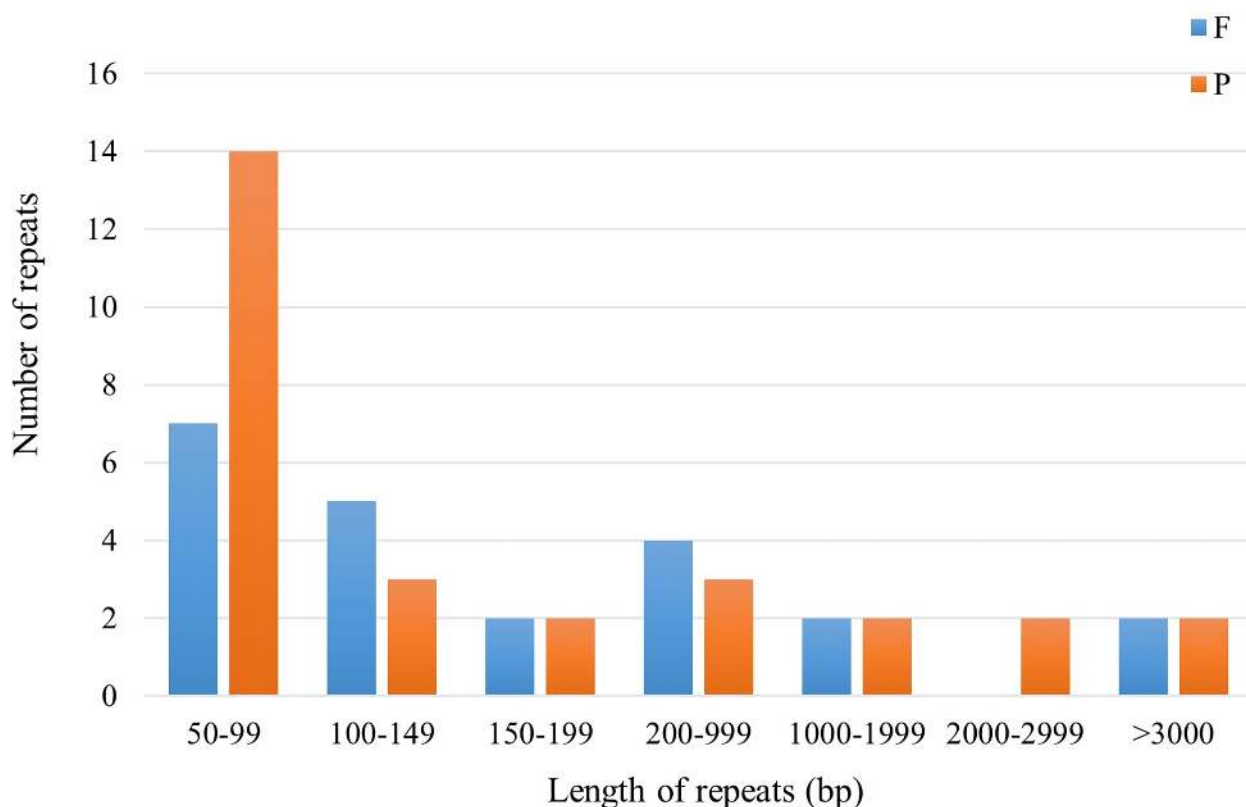


Fig. 4. Scattered repeats of *F. aubertii* mt genome. Different types of repeats are represented by different colors, blue represents forward repeats and orange represents palindromic repeats. The height of the colored bar represents the number of sequences. No scattered repeats in the length range of 2000–2999 bp were found.

Table 2. List of genes annotated in *F. aubertii* mt genome.

Group of genes	Gene name
ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFC*, ccmFN</i>
Ubiquinol cytochrome c reductase	<i>cob</i>
Cytochrome 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 proteins (LSU)	<i>rpl16, rpl5</i>
Ribosomal proteins (SSU)	<i>rps1, rps12, rps13, rps14, rps3*, rps4, rps7</i>
Succinate dehydrogenase	<i>sdh4</i>
Ribosomal RNAs	<i>rrn18, rrn26, rrn5</i>
Transfer RNAs	<i>trnC-GCA, trnD-GUC, trnE-CUC, trnE-UUC (2), trnF-GAA, trnG-GCC, trnH-GUG (2), trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnM-CAU (4), trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnS-GCU, trnS-GGA, trnS-UGA, trnV-GAC, trnW-CCA, trnY-GUA</i>

Notes: *: intron number. The number of asterisks (*) refers to the number of introns. The number in brackets next to the gene name indicates the number of copies of multi-copy genes.

tetragonoides (MW971440.1)) was calculated. The homologous protein-coding sequences were obtained by comparing protein-coding sequences among mitochondrial genomes of *F. aubertii* and its closely related species to find the best match using BLAST [31]. Mafft v7.427

(Research Institute for Microbial Diseases, Suita, Osaka, Japan) [32] (<https://mafft.cbrc.jp/alignment/server/>) was used for homologous protein sequence alignment, whereas a Perl script was utilized to map the aligned protein sequences to the coding sequence to obtain an aligned

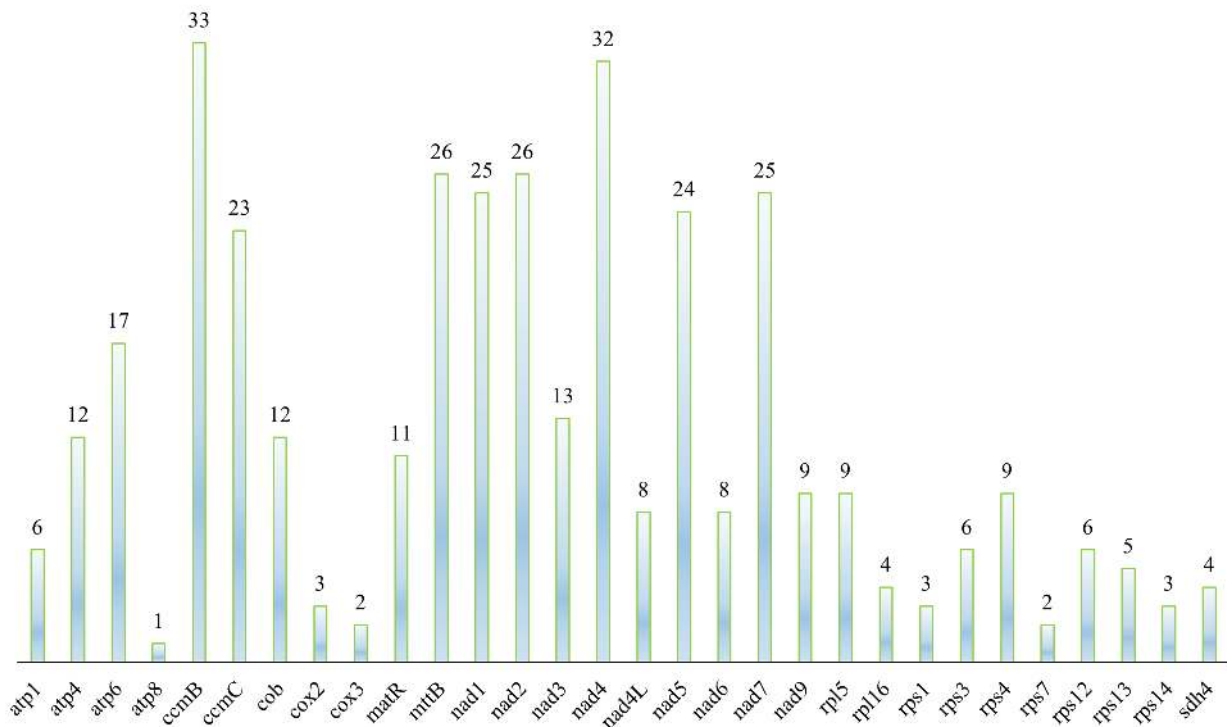


Fig. 5. Distribution of RNA editing sites in *F. aubertii* mt protein-coding genes (PCGs). RNA editing sites in 30 coding genes in *F. aubertii* were predicted. There are no RNA editing sites in *cox1*, *ccmFC*, *ccmFN*, or *atp9*. The horizontal axis is the protein-coding gene, and the number on the top of the color bar refers to the number of predicted RNA editing sites.

Table 3. Number of short tandem repeats in *F. aubertii* mt genome.

Motif	Number	Proportion (%)
Mononucleotide	28	36.4
Dinucleotide	16	20.8
Trinucleotide	7	9.1
Tetranucleotide	22	28.5
Pentanucleotide	4	5.2
Total	77	100

coding sequence. Finally, *Ka/Ks* Calculator 2.0 [33] 96–97 with MLWL model was applied to calculate *Ka* and *Ks* values.

2.7 Phylogenetic Analysis

A phylogenetic tree was constructed using the mitochondrial genome sequences of *F. aubertii*, *Malania oleifera* (Olacaceae, Santalales), and 15 other species from the order Caryophyllales, with *M. oleifera* used as an out-group. We extracted the shared coding sequences from the 17 genomes mentioned above and used BLAST to identify homologous sequences. The resulting sequences were aligned using Mafft, and then concatenated. A trimming percentage of 0.7 was set using the trim function to remove any poorly aligned bases. The maximum likelihood phylogenetic tree was constructed, and the substitution model

GTR+F+R3 was selected for the analysis using MEGA 7 [34]. The phylogenetic tree's step values were determined for each branch by performing 1000 spontaneous replicate analyses.

2.8 Analysis of Homologous Fragments of Chloroplast-Mitochondrial Genomes

A BLAST search was conducted between the mitochondrial and chloroplast genomes of *F. aubertii* to identify homologous regions. The resulting homologous segments were then compared and analyzed for sequence similarity, number of protein-coding genes, length, and composition of intergenic regions.

3. Results

3.1 Basic Features of the Mitochondrial Genome

The second-generation sequencing platform yielded a data set of 4645.7 Mb of raw reads and 4432 Mb of clean reads, whereas the third-generation sequencing platform generated 21,993 reads with a total size of 67 Mb.

The genome (NCBI accession number: MW664926.1) obtained in this study is a circular DNA molecule with a length of 350,156 bp (Fig. 1), and a GC content of 44.73%. The base composition of the mitochondrial genome is asymmetric, and the coding and non-coding regions' base composition, size, and proportion

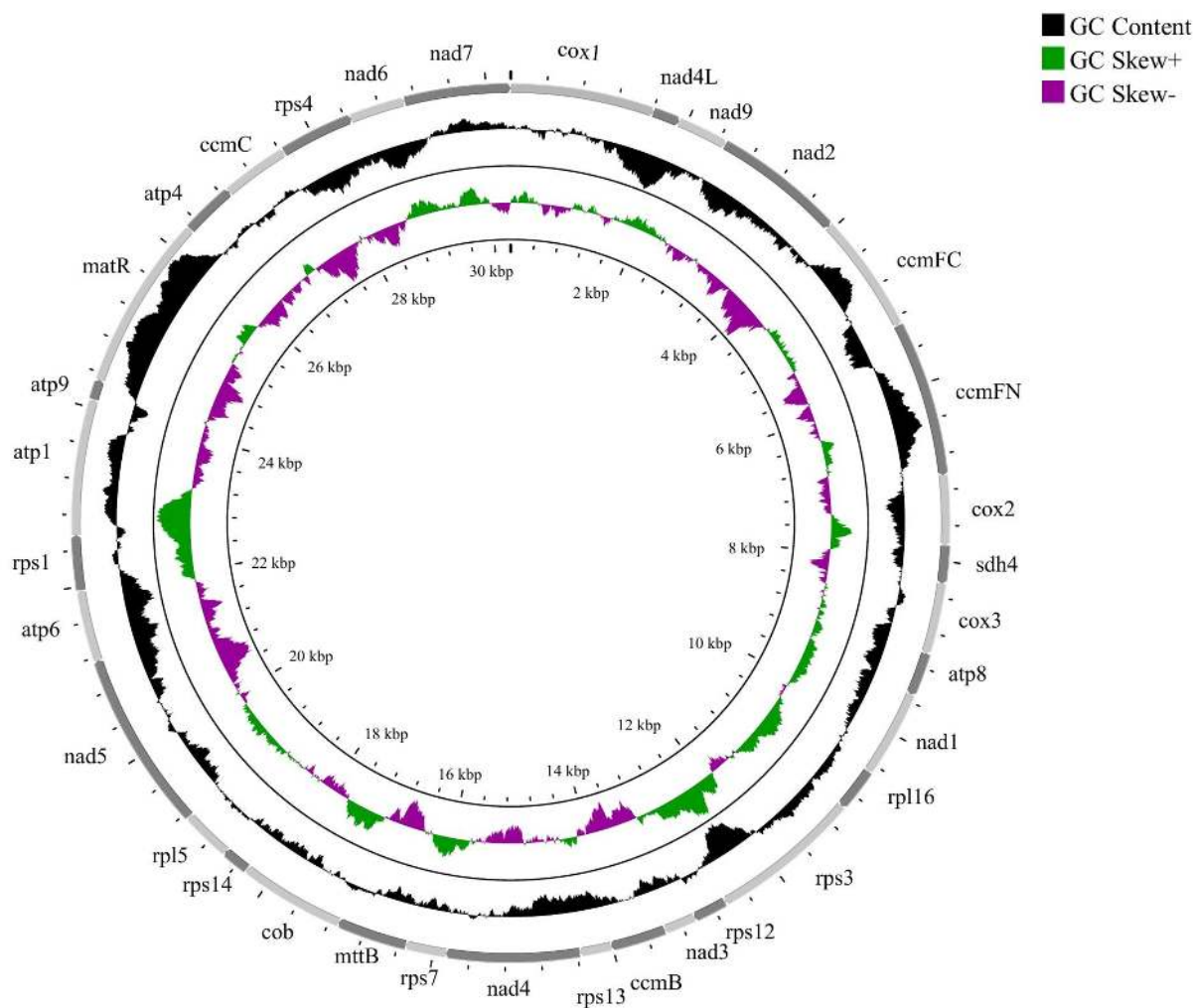


Fig. 6. GC skew of *F. aubertii* mt CDS. The outermost circle represents the coding area of a gene, the innermost circle represents the variation in GC content across the genome, and the innermost circle represents the GC skew, where green represents skew+ and purple represents skew-.

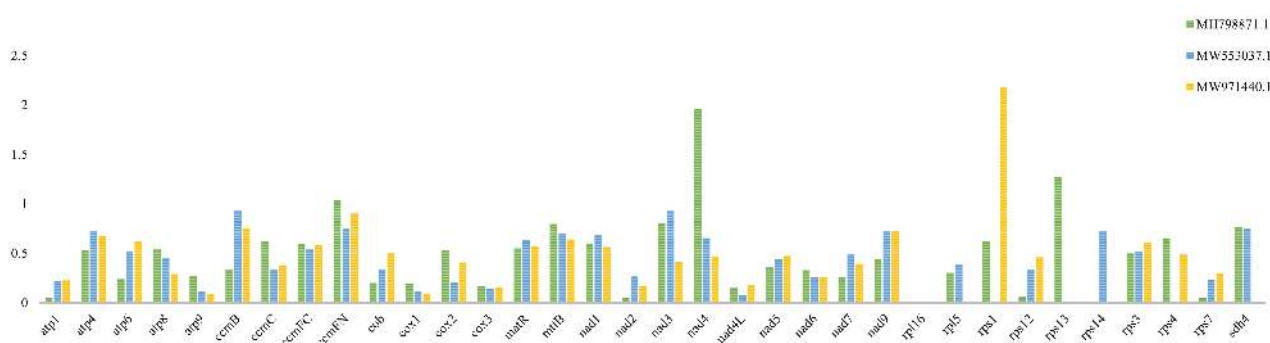


Fig. 7. K_a/K_s values of the PCGs. The mitochondrial protein-encoding genes of *F. aubertii* were compared with *N. ventrata* (MH798871.1), *A. githago* (MW553037.1), and *T. tetragonoides* (MW971440.1), respectively, for K_a/K_s analysis. The three closely related species contain protein-coding genes different from those in *F. aubertii*, so the K_a/K_s values of some genes in the figure are 0.

are shown in Table 1. The GC content in tRNA and rRNA is relatively high, accounting for 0.59% and 1.47% of the mitochondrial genome sequence, respectively.

The *F. aubertii* mitochondrial genome contains 64 genes, including 34 PCGs, three rRNAs, and 27 tRNAs (Table 2). Out of the 34 protein-coding genes, only five

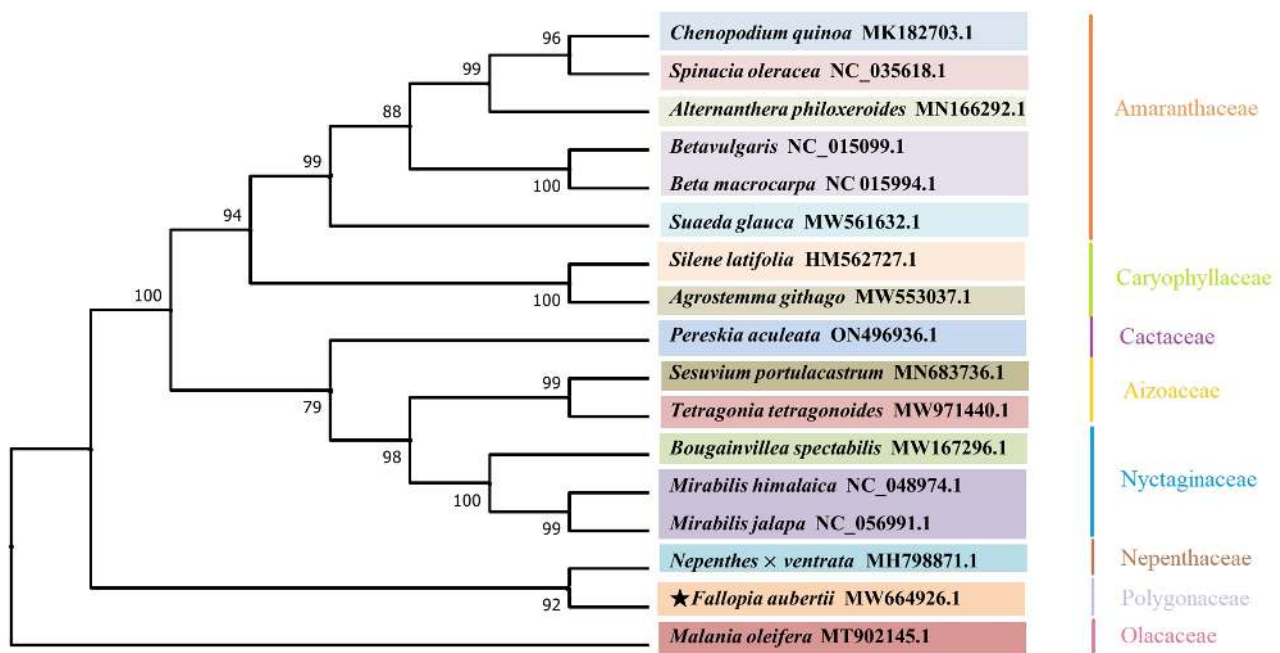


Fig. 8. A phylogenetic tree of 17 species. Reconstruction of phylogenetic relationships was based on the shared CDS using the maximum likelihood model with *M. oleifera* (Olacaceae) as an outgroup. *F. aubertii* is indicated using the star icon. The plants grouped together are marked according to the family.

Table 4. Long tandem repeats of *F. aubertii* mt genome.

Indices	Period size of the repeat	Copy number	Consensus size	Percent of matches	Percent of indels	A	C	G	T	Entropy
48,505–48,551	20	2.2	22	81	11	29	25	21	23	1.99
93,390–93,432	18	2.4	18	81	18	34	25	20	18	1.96
119,178–119,223	24	1.9	24	90	0	28	21	21	28	1.99
276,633–276,698	32	2.1	32	94	0	27	9	22	40	1.84
289,717–289,760	21	2.1	21	91	0	36	15	20	27	1.93

(*ccmFC*, *rps3*, *nad4*, *nad5*, and *nad7*) contain introns and were categorized as cis-splice genes (Fig. 2). We examined 27 tRNA genes and determined that they can transport 17 standard amino acids. The length of tRNA ranged from 71 bp to 88 bp, with a length of 1816 bp in total. Among these genes, multicopy genes were found for *trnE*-UUC, *trnH*-GUG, and *trnM*-CAU. Additionally, we predicted the secondary structures of tRNAs (Fig. 3). All genes except for *tRNA-Leu*, *tRNA-Ser*, and *tRNA-Tyr* were predicted to have typical cloverleaf structures, and the base mismatches were mostly G-U. The mitochondrial genome contains three different tRNAs, each with different anticodons that specifically transport serine (Ser). These results are of great significance for further study of the function and stability of tRNA.

3.2 Mitochondrial Repetitive Sequences Analysis

Repetitive sequences can serve as genome-specific genetic markers for phylogenetic relationships among species. STRs show a high degree of polymorphism and are com-

monly used as molecular markers for genetic diversity studies and germplasm characteristics, identification, and selection. In the analysis of STR in the mitochondrial genome of *F. aubertii*, 77 STR loci were detected. There are 28 single nucleotide motifs, 16 dinucleotide motifs, seven trinucleotide motifs, 22 tetranucleotide motifs, and four pentanucleotide motifs. No hexanucleotide motifs were detected (Table 3). Additionally, five long tandem repeat sequences were detected by TRF software in the mitochondrial genome of *F. aubertii* (Table 4).

Scattered repetitive sequences are crucial in the study of gene mutation, genome origin and evolution, and species formation [35]. In this study, a total of 50 scattered repeats were identified in the mitochondrial genome of *F. aubertii*, consisting of 22 forward repeats and 28 palindromic repeats. No reverse repeats, or complementary repeats were detected. The longest identified repeat sequence was 5272 bp, whereas most of the repeats were 50–200 bp in length (Fig. 4).

Table 5. Basic statistics on RNA editing sites predicted in *F. aubertii* mt genes.

Type	RNA-editing	Number	Percentage (%)
Hydrophobic	CCA (P) $\Rightarrow$ CTA (L)	31	28.07
	CCC (P) $\Rightarrow$ CTC (L)	3	
	CCG (P) $\Rightarrow$ CTG (L)	26	
	CCT (P) $\Rightarrow$ CTT (L)	18	
	CCC (P) $\Rightarrow$ TTC (F)	6	
	CCT (P) $\Rightarrow$ TTT (F)	2	
	GCT (A) $\Rightarrow$ GTT (V)	1	
	GCG (A) $\Rightarrow$ GTG (V)	2	
	GCA (A) $\Rightarrow$ GTA (V)	2	
	CTT (L) $\Rightarrow$ TTT (F)	9	
	CTC (L) $\Rightarrow$ TTC (F)	3	
Hydrophilic	CAT (H) $\Rightarrow$ TAT (Y)	15	14.17
	CAC (H) $\Rightarrow$ TAC (Y)	6	
	CGT (R) $\Rightarrow$ TGT (C)	23	
	CGC (R) $\Rightarrow$ TGC (C)	8	
Hydrophobic-hydrophilic	CCT (P) $\Rightarrow$ TCT (S)	14	7.90
	CCG (P) $\Rightarrow$ TCG (S)	6	
	CCC (P) $\Rightarrow$ TCC (S)	4	
	CCA (P) $\Rightarrow$ TCA (S)	5	
Hydrophilic-hydrophobic	CGG (R) $\Rightarrow$ TGG (W)	19	49.86
	TCT (S) $\Rightarrow$ TTT (F)	28	
	TCC (S) $\Rightarrow$ TTC (F)	25	
	TCG (S) $\Rightarrow$ TTG (L)	37	
	TCA (S) $\Rightarrow$ TTA (L)	56	
	ACT (T) $\Rightarrow$ ATT (I)	4	
	ACC (T) $\Rightarrow$ ATC (I)	6	
	ACG (T) $\Rightarrow$ ATG (M)	8	

3.3 Prediction of RNA Editing Sites

The prediction of RNA editing sites in the mitochondrial genome coding genes of *F. aubertii* resulted in a total of 367 sites in 30 genes (Fig. 5). Among them, *ccmB* and *nad4* contained the majority of RNA editing sites, with *ccmB* containing most of them (33 sites), accounting for 8.99% of the total RNA editing sites. The *atp8* had the fewest RNA editing sites, accounting for only 0.27% of the total number of sites. The results showed that 7.9% of hydrophobic amino acids were converted to hydrophilic amino acids, 49.86% of hydrophilic amino acids were converted into hydrophobic amino acids, and 42.24% of hydrophobic amino acids remained unchanged (Table 5). RNA editing sites were C to T conversion, and the first base of the codon accounted for 32.0% of the editing sites, whereas the second base accounted for 68.0%. Some RNA editing sites may result in the formation of stop codons, but we did not identify such phenomena in the mitochondrial genome of *F. aubertii*. Furthermore, the proportion of amino acid conversion to leucine after RNA editing was the largest, accounting for 46.59%.

3.4 Mitochondrial Genome Size and GC Content of *F. aubertii* Compared with Other Species

The CDS of *F. aubertii* were analyzed to determine the GC content of each gene and calculate the average GC content (Table 6). The *nad5* gene was the longest, possibly due to its introns. The GC content varied across PCGs, with *matR* gene having the highest GC content (51.5%) and the GC content of the *nad4L* being the lowest (34.65%). The AT skew average value was negative, and 13 genes had positive AT skew. Visual analysis of GC skew showed that 22 genes with positive GC skew had higher G-base content than C-base content, and 12 genes with negative GC skew and high C-base content (Fig. 6). Caryophyllales plants, including *F. aubertii*, had a mitochondrial genome size range of 247 Kb–509 Kb, with an average GC content of about 44% (Table 7).

3.5 *Ka*, *Ks* Analysis

Comparing the *Ka* and *Ks* in homologous genes during evolution is essential for the study of gene functions and evolutionary relationships across species, as well as for exploring issues such as adaptation and genetic diversity of plant communities. The *Ka/Ks* ratios of 34 PCGs in the mi-

Table 6. GC content of *F. aubertii* mt PCGs.

Genes	Length	GC1	GC2	GC3	GC all	AT skew	GC skew
<i>cox1</i>	1611	46	41.15	36.87	41.34	−0.17	0.02
<i>nad4L</i>	303	37.62	36.63	29.7	34.65	−0.24	0.1
<i>nad9</i>	573	49.74	40.31	31.41	40.49	0.01	0.07
<i>nad2</i>	1467	40.9	42.74	32.72	38.79	−0.21	−0.05
<i>ccmFC</i>	1344	48.21	43.53	42.63	44.79	−0.07	0.01
<i>ccmFN</i>	1740	49.66	46.55	40.52	45.57	−0.16	−0.01
<i>cox2</i>	783	51.72	36.4	30.65	39.59	−0.08	0.03
<i>sdh4</i>	420	42.14	34.29	42.14	39.52	−0.09	−0.02
<i>cox3</i>	798	50.75	41.73	33.08	41.85	−0.26	0.07
<i>atp8</i>	483	35.4	37.27	41.61	38.1	−0.02	0.03
<i>nad1</i>	978	53.99	41.1	28.83	41.31	−0.17	0.03
<i>rpl16</i>	480	49.38	53.12	31.87	44.79	0.09	0.18
<i>rps3</i>	1686	41.46	40.57	40.93	40.98	0.12	0.11
<i>rps12</i>	378	53.97	48.41	30.95	44.44	0.21	0.11
<i>nad3</i>	357	42.02	42.86	35.29	40.06	−0.3	−0.01
<i>ccmB</i>	621	44.44	44.44	33.82	40.9	−0.25	−0.17
<i>rps13</i>	351	49.57	39.32	26.5	38.46	0.15	0.14
<i>nad4</i>	1488	44.76	40.73	35.08	40.19	−0.19	−0.03
<i>rps7</i>	456	48.03	38.16	36.18	40.79	0.12	0.2
<i>mttB</i>	786	42.37	40.46	42.75	41.86	−0.2	−0.04
<i>cob</i>	1182	49.49	41.12	30.96	40.52	−0.22	0.03
<i>rps14</i>	303	42.57	46.53	33.66	40.92	−0.02	0.08
<i>rpl5</i>	561	50.27	37.97	40.64	42.96	0.1	0.02
<i>nad5</i>	2013	43.37	43.07	33.68	40.04	−0.23	−0.01
<i>atp6</i>	801	44.19	35.21	30.34	36.58	−0.18	−0.06
<i>rps1</i>	612	44.61	39.71	42.65	42.32	0.06	0.14
<i>atp1</i>	1524	57.28	41.54	33.66	44.16	0.06	0.07
<i>atp9</i>	225	48	42.67	37.33	42.67	−0.22	0.08
<i>matR</i>	1965	52.67	44.43	57.4	51.5	0.13	−0.03
<i>atp4</i>	597	47.24	41.71	36.18	41.71	0.02	0.03
<i>ccmC</i>	753	46.22	44.62	34.26	41.7	−0.21	−0.1
<i>rps4</i>	828	41.3	38.77	34.78	38.29	0.14	−0.05
<i>nad6</i>	624	41.83	37.02	35.58	38.14	−0.15	0.08
<i>nad7</i>	1185	55.7	43.8	32.66	44.05	0.02	0.05
All	30,276	47.24	41.78	36.9	41.97	−0.08	0.02

tochondrial genomes of *F. aubertii*, *N. ventrata*, *A. githago*, and *T. tetragonoides* were analyzed (Fig. 7). Among these 34 genes, the *rps16* gene had a zero *Ka/Ks* value. Overall, four of the 34 genes had *Ka/Ks* values greater than one, indicating that these four genes were under positive selective pressure during evolution, which included *nad4*, *rps13*, *rps1* and *ccmFN*. The remaining genes had values of *Ka/Ks* less than 1, which implies purifying selection and relative conservation.

3.6 Phylogenetic Analysis

To better understand the phylogenetic position of *F. aubertii*, 16 plant mitogenomes from the NCBI database were downloaded (Table 7). The phylogenetic tree showed that most branch nodes had high support values above 99%, and species from the same family were grouped together,

indicating high result reliability (Fig. 8). The phylogenetic tree also strongly demonstrated that *F. aubertii* and the *N. ventrata* clustered into one clade with a 92% bootstrap value (Fig. 7). Caryophyllales plants were further divided into two subgroups [36,37]. The *F. aubertii* and the *N. ventrata* were non-core taxa of Caryophyllales, and other species belonged to the core group of Caryophyllales.

3.7 Homologous Fragment Analysis of *F. aubertii* Chloroplast and Mitochondrial Genomes

By analyzing the homology of chloroplast and mitochondrial genomes, we can deeply explore their evolutionary history and their roles in evolution. Fifty-eight homologous fragments were observed in the organellar genome of *F. aubertii*, with a total length of 47,757 bp, accounting for 29.2% and 13.6% of the chloroplast and mitochondrial

Table 7. Mitochondrial genome and GC content in *F. aubertii* and 15 species of Caryophyllales.

Species	Accession number	Category	Genome size	GC
<i>Fallopia aubertii</i>	MW664926.1	Polygonaceae	350,156	44.73
<i>Nepenthes × ventrata</i>	MH798871.1	Nepenthaceae	520,764	44.17
<i>Alternanthera philoxeroides</i>	MN166292.1	Amaranthaceae	283,258	44.07
<i>Spinacia oleracea</i>	NC_035618.1	Amaranthaceae	329,613	43.41
<i>Beta macrocarpa</i>	NC_015994.1	Amaranthaceae	385,220	43.89
<i>Beta vulgaris</i>	NC_015099.1	Amaranthaceae	364,950	43.91
<i>Chenopodium quinoa</i>	MK182703.1	Amaranthaceae	315,003	43.83
<i>Suaeda glauca</i>	MW561632.1	Amaranthaceae	474,330	44.07
<i>Bougainvillea spectabilis</i>	MW167296.1	Nyctaginaceae	343,746	44.06
<i>Mirabilis jalapa</i>	NC_056991.1	Nyctaginaceae	267,334	44.75
<i>Mirabilis himalaica</i>	NC_048974.1	Nyctaginaceae	346,363	44.65
<i>Agrostemma githago</i>	MW553037.1	Caryophyllaceae	262,903	44.66
<i>Silene latifolia</i>	HM562727.1	Caryophyllaceae	253,413	42.56
<i>Pereskia aculeata</i>	ON496936.1	Cactaceae	515,187	44.05
<i>Sesuvium portulacastrum</i>	MN683736.1	Aizoaceae	392,221	44.16
<i>Tetragonia tetragonoides</i>	MW971440.1	Aizoaceae	347,227	43.84

genomes, respectively. There were 14 homologous genes in the mitochondrial genome, including *trnW-CCA*, *trnV-GAC*, *trnS-GGA*, *trnR-ACG*, *trnN-GUU*, *trnM-CAU*, *trnL-CAA*, *trnI-GAU*, *trnH-GUG*, *trnD-GUC*, *trnA-UGC*, *rrn26*, *rrn18*, and *ccmC* (Table 8).

4. Discussion

Analyzing mitochondrial genes can lead to the optimization of plant growth characteristics at the genetic level, thus increasing yield and enhancing plant adaptation to the environment, which is essential for agricultural production and food safety [38–41]. The study of genome size and GC content is crucial for understanding plant evolution and adaptability, as they greatly influence traits such as morphology, physiology, and biochemistry of plants [42]. The size of angiosperm mitochondrial genomes is usually 200–800 Kb [43], and the entire length of *F. aubertii* mitochondrial genome is 350,156 bp with a GC content of 44.73%, which is similar to other Caryophyllales plants [44,45]. Research on genome size and GC content can provide important insights into plants' genetic characteristics and biological functions [46] and form the scientific basis for applications such as plant breeding and biotechnology [47,48].

Repeat sequences are essential for gene expression and regulation [49]. They are the main driving force for gene diversification and evolution [50]—the present study identified STRs, long tandem repeats, and scattered repeats. A total of 77 STR loci, five long tandem repeats, and 50 scattered repeat sequences were detected by analyzing the *F. aubertii* mitochondrial genome repeats. Repetitive sequences play an important role in the processes of insertion, deletion, and other rearrangements that may occur within the mitochondrial genome [51]. Mitochondrial repetitive sequences have been widely used in plant genetic improvement and molecular marker-assisted breeding

[52,53]. Analysis of the mitochondrial repeat sequences of *F. aubertii* has helped us to gain insights into the organelle sequences, genetic diversity, and other relevant features of the plant, which are essential for species conservation and germplasm identification [54].

RNA editing is a type of genetic variation event in RNA post-transcriptional modification, which is crucial for maintaining the stability of gene expression [55,56]. Compared with DNA editing, RNA editing can respond more quickly to external stimuli or environmental changes, thereby adapting to the needs of survivors [57]. The diversity and complexity of RNA function, as well as its impact on gene expression and regulation, can be better understood by exploring the location and type of RNA editing, leading to further study into RNA function and regulatory mechanisms [58]. In this study, the RNA editing sites in the mitochondrial genome of *F. aubertii* were mainly C to T conversions, with the second base of the codon being the most frequently changed. The proportion of amino acid conversion to leucine was the largest, accounting for 46.59%. Leucine is the main amino acid produced after editing, similar to RNA editing results in higher angiosperms [59].

To understand the various selection pressures that took place during the evolution of the gene and to determine whether a gene has undergone rapid evolution, *Ka/Ks* analysis can be used [60]. In addition to predicting protein structure and function, *Ka/Ks* analysis can be used to research the gene families' functional diversity and evolutionary pathways [61]. When the value of *Ka/Ks* > 1, it indicates a positive selection pressure during the gene's evolution. Conversely, *Ka/Ks* < 1 denotes the negative selection. Most of the genes in *F. aubertii* are conserved. The four genes *rps1*, *nad4*, *rps13*, and *ccmFN* all have *Ka/Ks* > 1, which is significant for understanding the positive selection pressure these four genes experienced during evolution. We referred

Table 8. Homologous regions of organellar genomes of *F. aubertii*.

Sequence	Chloroplast genome		Mitochondrial genome		Identity
	Gene	Sequence position (bp)	Gene	Sequence position (bp)	
1	<i>trnR-ACG, rrn5, rrn4.5, rrn23, trnA-UGC, trnI-GAU, rrn16, trnV-GAC, rps7</i>	138,146–149,356	<i>trnR-ACG, trnA-UGC, trnI-GAU, trnV-GAC</i>	275,594–286,804	100
2	<i>rps7, trnV-GAC, rrn16, trnI-GAU, trnA-UGC, rrn23, rrn4.5, rrn5, trnR-ACG</i>	100,317–111,527	<i>trnR-ACG, trnA-UGC, trnI-GAU, trnV-GAC</i>	286,804–275,594	100
3	<i>ycf2, trnL-CAA, ndhB</i>	90,089–98,017	<i>trnL-CAA</i>	286,794–294,720	99
4	<i>trnL-CAA, ycf2, ndhB</i>	151,656–159,584	<i>trnL-CAA</i>	294,720–286,794	99
5	<i>rps19, rpl2, rpl23, trnM-CAU</i>	87,280–89,471	<i>trnM-CAU, ccmC</i>	303,457–305,648	100
6	<i>trnI-CAU, rpl23, rpl2</i>	160,202–162,393	<i>trnM-CAU, ccmC</i>	305,648–303,457	100
7	<i>psbA</i>	206–937	/	15,690–14,968	94
8	<i>trnH-GUG</i>	1–217	<i>trnH-GUG</i>	303,456–303,240	100
9	<i>rbcL</i>	58,741–58,955	/	241,189–240,975	97
10	<i>trnW-CCA</i>	68,795–68,976	<i>trnW-CCA</i>	44,932–44,752	90
11	<i>rrn23</i>	139,746–139,869	/	109,167–109,290	95
12	<i>rrn23</i>	109,804–109,927	/	109,290–109,167	95
13	<i>trnD-GUC</i>	31,665–31,802	<i>trnD-GUC</i>	171,543–171,407	92
14	<i>trnS-GGA</i>	47,519–47,611	<i>trnS-GGA</i>	261,692–261,784	96
15	<i>trnH-GUG</i>	2–78	<i>trnH-GUG</i>	257,385–257,461	100
16	<i>rrn16</i>	104,286–104,394	<i>rrn18</i>	13,207–13,315	92
17	<i>rrn16</i>	145,279–145,387	<i>rrn18</i>	13,315–13,207	92
18	<i>petG</i>	68,653–68,737	/	45,085–45,001	95
19	<i>rrn23</i>	139,686–139,933	<i>rrn26</i>	92,350–92,597	81
20	<i>rrn23</i>	109,740–109,987	<i>rrn27</i>	92,597–92,350	81
21	<i>trnN-GUU</i>	111,672–111,750	<i>trnN-GUU</i>	57,454–57,532	96
22	<i>trnN-GUU</i>	137,923–138,001	<i>trnN-GUU</i>	57,532–574,54	96
23	/	205–263	/	303,197–303,255	100
24	<i>trnM-CAU</i>	54,431–54,505	<i>trnM-CAU</i>	200,075–200,149	94
25	<i>rrn16</i>	103,420–103,478	/	2162–2220	100
26	<i>rrn16</i>	146,195–146,253	/	2220–2162	100
27	<i>rrn16</i>	103,730–103,815	<i>rrn18</i>	12,655–12,740	91
28	<i>rrn16</i>	145,858–145,943	<i>rrn18</i>	12,740–12,655	91
29	<i>rrn23</i>	140,317–140,410	<i>rrn26</i>	92,965–93,058	87
30	<i>rrn23</i>	109,263–109,356	<i>rrn26</i>	93,058–92,965	87
31	<i>psbD, psbC</i>	35,865–35,965	/	256,992–257,092	86
32	<i>trnP-UGG</i>	69,203–69,251	/	44,659–44,611	97
33	<i>ycf3</i>	46,462–46,537	/	260,822–260,897	88
34	<i>ndhA</i>	128,336–128,377	/	285,575–285,534	97
35	<i>rrn16</i>	104,701–104,818	<i>rrn18</i>	13,950–14,066	83
36	<i>rrn16</i>	144,855–144,972	<i>rrn18</i>	14,066–13,950	83
37	<i>rrn16</i>	104,580–104,649	<i>rrn18</i>	13,830–13,899	88
38	<i>rrn16</i>	145,024–145,093	<i>rrn18</i>	13,899–13,830	88

Table 8. Continued.

Sequence	Chloroplast genome		Mitochondrial genome		Identity
	Gene	Sequence position (bp)	Gene	Sequence position (bp)	
39	<i>rrn23</i>	140,568–140,612	/	93,214–93,258	95
40	<i>rrn23</i>	109,061–109,105	/	93,258–93,214	95
41	<i>rpl2</i>	87,768–87,803	/	101,974–102,009	100
42	<i>rpl2</i>	161,870–161,905	/	102,009–101,974	100
43	<i>rrn16</i>	103,910–103,981	<i>rrn18</i>	12,837–12,908	87
44	<i>rrn16</i>	145,692–145,763	<i>rrn18</i>	12,908–12,837	87
45	<i>ycf3</i>	45,725–45,763	/	285,573–285,535	97
46	/	46,265–46,315	/	260,632–260,682	92
47	<i>rrn16</i>	104,451–104,525	<i>rrn18</i>	13,371–13,445	86
48	<i>rrn16</i>	145,148–145,222	<i>rrn18</i>	13,445–13,371	86
49	<i>psbA</i>	248–285	/	303,209–303,272	97
50	/	68,036–68,103	/	19,391–19,458	86
51	<i>rrn23</i>	141,817–141,862	<i>rrn26</i>	94,788–94,833	91
52	<i>rrn23</i>	107,811–107,856	<i>rrn26</i>	94,833–94,788	91
53	<i>rrn16</i>	103,467–103,523	<i>rrn18</i>	12,351–12,407	87
54	<i>rrn16</i>	146,150–146,206	<i>rrn18</i>	12,407–12,351	87
55	<i>trnS-GCU</i>	8666–8693	<i>trnS-GGA</i>	261,781–261,754	100
56	<i>rpoC1</i>	22,600–22,627	/	110,671–110,698	100
57	<i>rrn16</i>	104,889–104,920	<i>rrn18</i>	14,184–14,215	96
58	<i>rrn16</i>	144,753–144,784	<i>rrn18</i>	14,215–14,184	96

Note: “/” indicates no gene at the locus.

to the literature on the mitochondrial genome of order Caryophyllales [62] and found that the phenomenon of four genes with $Ka/Ks > 1$ is unique to *F. aubertii*. Nevertheless, there are less studies concerning the selective pressure on the mitochondrial genome of Caryophyllales and our findings need to be confirmed by further studies. The positive selection of the above genes would contribute to a faster rate of evolution in organisms and would also increase the adaptive divergence of mitochondrial genes [63,64].

In this study, a maximum likelihood phylogenetic tree was constructed based on the shared mitochondrial CDSs. The *F. aubertii* and the *N. ventrata* formed a sister clade indicating a closer relationship between them, and agreed with previous research [62,65]. In the evolution of the chloroplast genome system in Caryophyllales, Polygonaceae plants are also closely related to Nepenthaceae plants [66–68]. Our study provides a scientific basis for further studies on the phylogeny of the Caryophyllales.

There is a horizontal transfer of genes between organelle genomes [69]. We made a comparison of the mitochondrial and chloroplast genomes of *F. aubertii*. Some genes were found to be similar, and these genes were highly homologous. Between the chloroplast and mitochondrial genomes of *F. aubertii*, there were 58 homologous segments and 14 homologous genes, including 11 tRNAs, two ribosomal protein genes, and a gene for cytochrome c biogenesis (*ccmC*). In angiosperms, tRNA genes are frequently transferred from the chloroplast genome to the mitochondrial genome [70,71]. The homology of chloroplast and mitochondrial genomes is an important research direction in biology and evolution. By studying the functional activity of migrating genes, we have gained a greater understanding of species evolution [72].

5. Conclusions

In this study, we provided a comprehensive analysis of the mitochondrial genome of *F. aubertii*. The genome was found to have a circular structure with a length of 350,156 bp and a GC content of 44.73%. The genome contains 64 genes, including 34 protein-coding genes, three rRNAs, and 27 tRNAs. Analyses of repeat sequences revealed the presence of short tandem repeats, long tandem repeats, and scattered repeats in the mitochondrial genome of *F. aubertii*. RNA editing sites were predicted, and the majority of editing events involved the conversion of C to T, with leucine being the most frequently converted amino acid. Analysis of the Ka/Ks ratios indicated that most mitochondrial genes in *F. aubertii* have undergone negative selection, whereas four genes (*rps1*, *nad4*, *rps13*, and *ccmFN*) showed evidence of positive selection. Phylogenetic analysis revealed that *F. aubertii* and *Nepenthes × ventrata* are closely related, supporting previous studies on the evolutionary relationships within the order Caryophyllales. Additionally, a comparison of the mitochondrial and chloroplast genomes

showed the presence of homologous regions and genes, suggesting potential gene transfer events between these organellar genomes.

This study provides valuable information on the mitochondrial genome of *F. aubertii* and serves as a reference for future molecular studies and the exploitation of genetic resources in this species. The findings contribute to the identification and improvement of valuable plant traits and highlight the significance of mitochondrial genomes in plant evolution and adaptation.

Availability of Data and Materials

The datasets used and/or analyzed during the current study can be obtained from the corresponding authors upon reasonable request. The raw data for the mitochondrial genomes sequencing is available on the website <https://www.ncbi.nlm.nih.gov/sra/> under the reference number SRX10859887. Additionally, the mitochondrial genome itself is available on the website <https://www.ncbi.nlm.nih.gov/genbank/> under the reference number MW167296.

Author Contributions

XZ performed the research, analyzed the data, and wrote the manuscript; JLW and YCL conceptualized and designed the study and reviewed the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was funded by Natural Science Foundation of Qinghai Province Science and Technology Department (Grant No. 2019-ZJ-915) and Foundation of Qinghai Minzu University (Grant No. 2022-JYQN-002, 2021XJG17).

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Yurina NP, Odintsova MS. Mitochondrial Genome Structure of Photosynthetic Eukaryotes. *Biochemistry. Biokhimiia*. 2016; 81: 101–113.
- [2] Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and

- p>visualizing expression data sets.
- Nucleic Acids Research*
- . 2013; 41: W575–W581.
- [3] He L, Maheshwari A. Mitochondria in Early Life. *Current Pediatric Reviews*. 2023; 19: 395–416.
 - [4] Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics*. 2004; 46: 123–139.
 - [5] Wallace DC. Mitochondrial diseases in man and mouse. *Science* (New York, N.Y.). 1999; 283: 1482–1488.
 - [6] Xia M, Zhang Y, Jin K, Lu Z, Zeng Z, Xiong W. Communication between mitochondria and other organelles: a brand-new perspective on mitochondria in cancer. *Cell & Bioscience*. 2019; 9: 27.
 - [7] Bohovych I, Khalimonchuk O. Sending Out an SOS: Mitochondria as a Signaling Hub. *Frontiers in Cell and Developmental Biology*. 2016; 4: 109.
 - [8] Ng S, De Clercq I, Van Aken O, Law SR, Ivanova A, Willems P, *et al.* Anterograde and retrograde regulation of nuclear genes encoding mitochondrial proteins during growth, development, and stress. *Molecular Plant*. 2014; 7: 1075–1093.
 - [9] Liberatore KL, Dukowicz-Schulze S, Miller ME, Chen C, Kianian SF. The role of mitochondria in plant development and stress tolerance. *Free Radical Biology & Medicine*. 2016; 100: 238–256.
 - [10] Rogalski M, do Nascimento Vieira L, Fraga HP, Guerra MP. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Frontiers in Plant Science*. 2015; 6: 586.
 - [11] Wu CC, Chang SH, Tung CW, Ho CK, Gogorcena Y, Chu FH. Identification of hybridization and introgression between *Cinnamomum kanehirae* Hayata and *C. camphora* (L.) Presl using genotyping-by-sequencing. *Scientific Reports*. 2020; 10: 15995.
 - [12] Jakubska-Busse A, Dziadas M, Gruss I, Kobyłka MJ. Floral Volatile Organic Compounds and a List of Pollinators of *Fallopia baldschuanica* (Polygonaceae). *Insects*. 2022; 13: 904.
 - [13] Schuster TM, Reveal JL, Bayly MJ, Kron KA. An updated molecular phylogeny of Polygonaceae: Relationships of Oxycoccus, Pteroxygonum, and Rumex, and a new circumscription of Koenigia. *Taxon*. 2015; 64: 1188–1208.
 - [14] Wang JF, Ma GX, Hu JZ. Phytochemical and chemotaxonomic study on *Fallopia aubertii* (L. Henry) Holub. *Biochemical Systematics and Ecology*. 2019; 85: 50–53.
 - [15] Schuster TM, Wilson KL, Kron KA. Phylogenetic relationships of Muehlenbeckia, Fallopia, and Reynoutria (Polygonaceae) investigated with chloroplast and nuclear sequence data. *International Journal of Plant Sciences*. 2011; 172: 1053–1066.
 - [16] Oлару OT, Venables L, VAN DE Venter M, Nitulescu GM, Margina D, Spandidos DA, *et al.* Anticancer potential of selected *Fallopia* Adans species. *Oncology Letters*. 2015; 10: 1323–1332.
 - [17] Zhao LQ, Zhang XP, Zhang CF. Separation and Identification of Chemical Constituents in *Fallopia auberti* Stems and Leaves. *Food Science*. 2012; 33: 1–5. (In Chinese)
 - [18] Ji D, Wang Q, Wang H, Ma Q, Wang M, Lu Y. Preparative separation of gallic acid from *Fallopia aubertii* using middle-pressure chromatogram isolated gel coupled with reversed-phase chromatography with hydrophilic groups. *RSC Advances*. 2021; 11: 27276–27282.
 - [19] Lu Y, Wang W, Yu Y, Ma Q, Wang H. The chloroplast genome of *Fallopia aubertii* (Polygonaceae) from Xining, China. *Mitochondrial DNA. Part B, Resources*. 2021; 6: 3458–3460.
 - [20] Ma Q, Zhao PC, Lu YC. Study on Inhibition of Xanthine Oxidase Activity in Different Extract Parts of *Fallopia auberti* L. Henry Holub. *Chinese Journal of Ethnomedicine and Ethnopharmacology*. 2020; 29: 16–20. (In Chinese)
 - [21] Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*. 2020; 21: 241.
 - [22] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology*. 2012; 19: 455–477.
 - [23] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* (Oxford, England). 2009; 25: 1754–1760.
 - [24] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014; 9: e112963.
 - [25] Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*. 2021; 49: 9077–9096.
 - [26] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*. 2019; 47: W59–W64.
 - [27] Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*. 1999; 27: 573–580.
 - [28] Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* (Oxford, England). 2017; 33: 2583–2585.
 - [29] Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*. 2001; 29: 4633–4642.
 - [30] Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, *et al.* Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*. 2023; 51: W484–W492.
 - [31] Hu G, Kurgan L. Sequence Similarity Searching. *Current Protocols in Protein Science*. 2019; 95: e71.
 - [32] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. 2019; 20: 1160–1166.
 - [33] Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics*. 2010; 8: 77–80.
 - [34] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*. 2016; 33: 1870–1874.
 - [35] Mehrotra S, Goyal V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics, Proteomics & Bioinformatics*. 2014; 12: 164–171.
 - [36] Zhang X, Shan Y, Li J, Qin Q, Yu J, Deng H. Assembly of the Complete Mitochondrial Genome of *Pereskia aculeata* Revealed That Two Pairs of Repetitive Elements Mediated the Recombination of the Genome. *International Journal of Molecular Sciences*. 2023; 24: 8366.
 - [37] Group TAP, Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, *et al.* An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*. 2016; 181: 1–20.
 - [38] Silva-Pinheiro P, Minczuk M. The potential of mitochondrial genome engineering. *Nature Reviews. Genetics*. 2022; 23: 199–214.
 - [39] Zhang X, Liu S, Takano T. Overexpression of a mitochondrial ATP synthase small subunit gene (AtMtATP6) confers tolerance to several abiotic stresses in *Saccharomyces cerevisiae* and *Ara-*

- bidopsis thaliana. Biotechnology Letters. 2008; 30: 1289–1294.
- [40] Moghadam AA, Ebrahimie E, Taghavi SM, Niazi A, Babgohari MZ, Deihimi T, *et al.* How the nucleus and mitochondria communicate in energy production during stress: nuclear MtATP6, an early-stress responsive gene, regulates the mitochondrial F₁F₀-ATP synthase complex. Molecular Biotechnology. 2013; 54: 756–769.
- [41] Zhang X, Takano T, Liu S. Identification of a mitochondrial ATP synthase small subunit gene (RMtATP6) expressed in response to salts and osmotic stresses in rice (*Oryza sativa* L.). Journal of Experimental Botany. 2006; 57: 193–200.
- [42] Trávníček P, Čertner M, Ponert J, Chumová Z, Jersáková J, Suda J. Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked to partial endoreplication, plant life-history traits and climatic conditions. The New Phytologist. 2019; 224: 1642–1656.
- [43] Li X, Zhe M, Huang Y, Fan W, Yang J, Zhu A. The Evolution of Mitochondrial Genomes between Two *Cymbidium* Sister Species: Dozens of Circular Chromosomes and the Maintenance and Deterioration of Genome Synteny. Genes. 2023; 14: 864.
- [44] Kim CK, Kim YK. The multipartite mitochondrial genome of *Fallopia multiflora* (Caryophyllales: Polygonaceae). Mitochondrial DNA. Part B, Resources. 2018; 3: 155–156.
- [45] Yuan F, Lan X. Sequencing the organelle genomes of *Bougainvillea spectabilis* and *Mirabilis jalapa* (Nyctaginaceae). BMC Genomic Data. 2022; 23: 28.
- [46] Bowers JE, Tang H, Burke JM, Paterson AH. GC content of plant genes is linked to past gene duplications. PloS One. 2022; 17: e0261748.
- [47] Sliwiska E, Loureiro J, Leitch IJ, Šmarda P, Bainard J, Bureš P, *et al.* Application-based guidelines for best practices in plant flow cytometry. Cytometry. Part a: the Journal of the International Society for Analytical Cytology. 2022; 101: 749–781.
- [48] Chaudhary J, Patil GB, Sonah H, Deshmukh RK, Vuong TD, Valliyodan B, *et al.* Expanding Omics Resources for Improvement of Soybean Seed Composition Traits. Frontiers in Plant Science. 2015; 6: 1021.
- [49] Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. Biological Reviews of the Cambridge Philosophical Society. 2005; 80: 227–250.
- [50] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science (New York, N.Y.). 2000; 290: 1151–1155.
- [51] Nardi F, Carapelli A, Frati F. Repeated regions in mitochondrial genomes: distribution, origin and evolutionary significance. Mitochondrion. 2012; 12: 483–491.
- [52] Baloch FS, Altaf MT, Liaqat W, Bedir M, Nadeem MA, Cömertpay G, *et al.* Recent advancements in the breeding of sorghum crop: current status and future strategies for marker-assisted breeding. Frontiers in Genetics. 2023; 14: 1150616.
- [53] Amiteye S. Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon. 2021; 7: e08093.
- [54] Jia HM, Jiao Y, Wang GY, Li YH, Jia HJ, Wu HX, *et al.* Genetic diversity of male and female Chinese bayberry (*Myrica rubra*) populations and identification of sex-associated markers. BMC Genomics. 2015; 16: 394.
- [55] Yan J, Zhang Q, Yin P. RNA editing machinery in plant organelles. Science China. Life Sciences. 2018; 61: 162–169.
- [56] Simpson L, Emeson RB. RNA editing. Annual Review of Neuroscience. 1996; 19: 27–52.
- [57] Sheridan C. Shoot the messenger: RNA editing is here. Nature Biotechnology. 2023; 41: 306–308.
- [58] Hao W, Liu G, Wang W, Shen W, Zhao Y, Sun J, *et al.* RNA Editing and Its Roles in Plant Organelles. Frontiers in Genetics. 2021; 12: 757109.
- [59] Suzuki H, Yu J, Ness SA, O'Connell MA, Zhang J. RNA editing events in mitochondrial genes by ultra-deep sequencing methods: a comparison of cytoplasmic male sterile, fertile and restored genotypes in cotton. Molecular Genetics and Genomics: MGG. 2013; 288: 445–457.
- [60] Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. Nature Reviews. Genetics. 2011; 12: 703–714.
- [61] Bellavista E, Santoro A, Galimberti D, Comi C, Luciani F, Mishto M. Current understanding on the role of standard and immunoproteasomes in inflammatory/immunological pathways of multiple sclerosis. Autoimmune Diseases. 2014; 2014: 739705.
- [62] Androsiuk P, Paukšto Ł, Jastrzębski JP, Milarska SE, Okorski A, Pszczółkowska A. Molecular Diversity and Phylogeny Reconstruction of Genus *Colobanthus* (Caryophyllaceae) Based on Mitochondrial Gene Sequences. Genes. 2022; 13: 1060.
- [63] Fourdrilis S, de Frias Martins AM, Backeljau T. Relation between mitochondrial DNA hyperdiversity, mutation rate and mitochondrial genome evolution in *Melarhaphe neritoides* (Gastropoda: Littorinidae) and other Caenogastropoda. Scientific Reports. 2018; 8: 17964.
- [64] Liu SL, Adams K. Molecular adaptation and expression evolution following duplication of genes for organellar ribosomal protein S13 in rosids. BMC Evolutionary Biology. 2008; 8: 25.
- [65] Zhu ZX, Wang JH, Chen CR, Zhao KK, Wang HF. Complete plastome sequence of *Nepenthes mirabilis* (Nepenthaceae): a “vulnerable” herb in China. Mitochondrial DNA. Part B, Resources. 2018; 3: 732–733.
- [66] Choi KS, Kwak M, Lee B, Park S. Complete chloroplast genome of *Tetragonia tetragonioides*: Molecular phylogenetic relationships and evolution in Caryophyllales. PloS One. 2018; 13: e0199626.
- [67] Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, *et al.* Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. Molecular Biology and Evolution. 2015; 32: 2001–2014.
- [68] Yao G, Jin JJ, Li HT, Yang JB, Mandala VS, Croley M, *et al.* Plastid phylogenomic insights into the evolution of Caryophyllales. Molecular Phylogenetics and Evolution. 2019; 134: 74–86.
- [69] Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biology. 2016; 17: 134.
- [70] Wang D, Wu YW, Shih ACC, Wu CS, Wang YN, Chaw SM. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Molecular Biology and Evolution. 2007; 24: 2040–2048.
- [71] Ala KG, Zhao Z, Ni L, Wang Z. Comparative analysis of mitochondrial genomes of two alpine medicinal plants of *Gentiana* (Gentianaceae). PloS One. 2023; 18: e0281134.
- [72] Richardson AO, Palmer JD. Horizontal gene transfer in plants. Journal of Experimental Botany. 2007; 58: 1–9.