De-novo assembly and comparative analysis of the complete mitogenome of traditional Chinese medicine *Strobilanthes sarcorrhiza*

Yujie Shi[1], Ping Li[2], Jian Sun[3], Meixin Li[1], Jingyong Jiang[4], Yue Xin[1], Zhen Chen[1] and Wei Zeng[1*]

Abstract

Background *Strobilanthes sarcorrhiza* is a traditional medicinal plant known for its heat-clearing and kidney-nourishing properties. While its plastid genome has been reported, there is a scarcity of genetic information regarding its mitogenome, leading to unclear phylogenetic relationships. We sequenced and assembled the complete its mitogenome and conducted a series of genetic analyses in conjunction with the plastid genome to gain a better understanding of the species' genetic background.

Results The mitogenome comprised a linear structure spanning 617,134 bp. It included 35 protein-coding genes (PCGs), 19 transfer RNAs (tRNAs), and 3 ribosomal RNAs (rRNAs) that have been annotated. Additionally, 122 simple sequence repeats (SSRs) and 25 tandem repeats were identified. A total of 1,482 pairs of dispersed repeats were detected, which account for 17.58% of the entire mitogenome. Furthermore, 37 migration fragments between the mitochondrial and plastid genomes were discovered, consisting of 5 complete PCGs, 7 tRNAs, and 1 rRNA. Based on the analysis of 38 mitogenomes and 46 plastid genomes, the evolutionary relationship and phylogenetic position of *S. sarcorrhiza* were elucidated.

Conclusions This study has, for the first time, provided insights into the mitochondrial genomic characteristics of *S. sarcorrhiza* and clarified its phylogenetic position. These findings offered significant insights for the future identification and classification of this genus, as well as for the genetic breeding of medicinal plants.

Keywords De-novo assembly, *Strobilanthes sarcorrhiza*, Mitochondrial genome, Molecular marker, Chinese medicine identification

*Correspondence:
Wei Zeng
zengw@tzc.edu.cn
[1]Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, College of Life Sciences, Taizhou University, Taizhou 318000, China
[2]Huangyan Forestry Technology Promotion Station, Taizhou 318000, China
[3]Zhejiang Research Institute of Traditional Chinese Medicine, Hangzhou 310023, China
[4]Institute of Horticulture, Taizhou Academy of Agricultural Sciences, Taizhou 318000, China

Background

The Acanthaceae is one of the largest families within the Lamiales order [1]. Statistics indicate that it comprises approximately 250 genera and over 4,000 species worldwide, primarily found in tropical regions [2, 3]. The family exhibits a high degree of diversity in survival forms, encompassing xerophytes, hygrophytes, herbs, shrubs, vines, and even small trees, and plays a significant role in the flora of many areas [4]. Acanthaceae plants are prized for their ornamental value and are frequently utilized as horticultural flowers, examples being *Thunbergia grandiflora*, *Barleria cristata*, and *Cyrtanthera carnea*. Additionally, medicinal properties are a prominent feature of Acanthaceae plants; for instance, *Strobilanthes cusia* and *Andrographis paniculata* are known for their heat-clearing, detoxifying, blood-activating, and anti-swelling effects [5, 6]. Consequently, members of the Acanthaceae family are plants of significant ecological, medicinal, and ornamental importance, with vast potential for development and utilization. However, due to the extensive number of genera and species, widespread distribution, and unique lifestyles, the Acanthaceae has long been regarded as a challenging group to classify [7, 8]. As a result, there is no comprehensive monograph on the family available worldwide. Moreover, the varying nutritional and reproductive organs of these plants make it difficult to achieve a consensus on the classification system proposed by different scholars based on various morphological characteristics [4, 9], particularly for the two largest genera, *Strobilanthes* and *Justicia*, which have been subject to differing opinions both domestically and internationally [1, 7, 10].

The *Strobilanthes* genus is one of the largest within the Acanthaceae family, comprising approximately 400 species [7, 11]. China stands as a center of diversity for this genus, hosting around 130 species [12, 13]. Many of these species possess significant medicinal properties. The *S. sarcorrhiza* is one of the representative species of the genus, with its substantial fleshy roots utilized in traditional Chinese medicine to nourish yin, clear heat, and address kidney deficiencies [14, 15]. It is a widely circulated remedy in southwestern Zhejiang and a distinctive plant of the region [14–16]. In recent times, the demand and price for it have escalated. Nonetheless, the taxonomy and evolutionary relationships within this genus remain ambiguous, resulting in the marketing of roots from other plants, such as *Pseudostellaria heterophylla*, *Lophatherum gracile*, and *Silene tatarinowii*, as *S. sarcorrhiza*. This practice has compromised the safety of clinical treatments and has impeded the development and utilization of this plant genus.

Mitochondria generate energy via aerobic respiration within eukaryotic cells [17]. They are involved not only in regulating key metabolic processes such as cell differentiation, apoptosis, growth, and division but are also closely associated with significant traits like stress tolerance, plant growth vitality, and cytoplasmic male sterility [18, 19]. Mitochondria are an essential tool for studying eukaryotic evolution, species identification, genetic diversity, and breeding [18, 20, 21]. The mitogenome is smaller than the nuclear genome, with coding genes that are highly conserved and non-coding regions that are highly variable [22, 23]. Currently, research on complete plant mitogenomes lags significantly behind that of complete plastid genomes. Although the NCBI database contains nearly 15,218 complete plastid genomes, there are only 1,102 complete plant mitogenomes [24]. Different plant mitogenomes exhibit significant structural and content variations, nucleotide substitution rates, and repetitive sequences, resulting in complex structural types such as circular, branched, and reticular structures within the mitogenomes [22, 25]. The genome length varies from 66 Kb for *Viscum scurruloideum* to 12 Mb for *Larix sibirica*, making the assembly of plant mitogenomes challenging [26, 27]. Consequently, most plant systematics research concentrates on nuclear and plastid genomes, and the complete assembly of plant mitogenomes remains a bottleneck in evolutionary biology. With the advancement of high-throughput sequencing technology and the emergence of the next generation of systematic genomics, numerous software programs suitable for mitochondrial genome sequencing and assembly have been developed, such as GetOrganelle [28], GSAT [29] and PMAT [30]. These advancements have made mitogenome sequencing and assembly more accurate and efficient, providing robust technical support for a deeper understanding of the genetic characteristics and phylogeny of species.

Only two mitogenomes (*Avicennia marina* and *Echinacanthus longipes*) have been reported in the Acanthaceae family [31]. However, to date, the mitogenome of the genus *Strobilanthes* has not been documented, and this gap in knowledge has significantly impeded phylogenetic research and the utilization of molecular resources for this genus. Consequently, there is an urgent need to analyze the mitogenome of *S. sarcorrhiza* to further acquire information that will aid in future genetic evolution, phylogeny, and conservation strategies.

We assembled and annotated the mitogenome of *S. sarcorrhiza*, and analyzed its genetic characteristics, repetitive sequences, codon usage bias, RNA editing sites, non-synonymous substitutions (Ka)/synonymous substitutions (Ks) ratios, and sequence transfers. Multicollinearity analysis was used to compare it with six related species to determine sequence rearrangements during evolution. Furthermore, based on 38 mitogenomes and 46 plastid genomes respectively, the evolutionary relationship and genetic background of the family

Acanthaceae were further clarified. The purpose of this study is to analyze the mitogenome and provide insights for future research on genetic variation, phylogeny, and breeding of this species. Ultimately, it will also lay the groundwork for the protection and utilization of medicinal economic plants.

Methods

Plant sample collection and sequencing

The leaves of *S. sarcorrhiza* were collected from the planting greenhouse of Taizhou University (120°23.37' N, 28°39.46' E) in Taizhou City, Zhejiang Province, China, and stored in a -80℃ refrigerator for subsequent use. We utilized a modified CTAB method to extract total genomic DNA from the plants [32]. The DNA concentration was measured using a Nanodrop and a Qubit Fluorometer, and its purity and integrity were tested with 1% agarose gel electrophoresis. Subsequently, high-quality genomic DNA was sequenced using the Illumina NovaSeq X plus platform for second-generation sequencing and the Pacific Biosciences Revio platform for third-generation sequencing. In the end, we obtained WGS short reads and HiFi long reads, respectively.

Assembly and annotation of mitogenome

Utilizing HiFi sequencing data, we assembled the mitogenome of *S. sarcorrhiza* using PMAT v1.5.3 [30]. The parameters and pattern employed by PMAT were "-st hifi -g 820m -m" and the "autoMito" model, respectively. The frequency of 19-kmers was generated from clean WGS reads using the jellyfish v2.2.7 tool [33]. Subsequently, the GenomeScope software [34] was utilized to evaluate the genome's characteristics. Furthermore, we inferred the ploidy of the its genome based on k-mer data using the Smudgeplot v0.4.0 software [35].

We utilized the Bandage v0.8.1 tool [36] to visualize the original assembly graph and removed contigs corresponding to the nuclear and chloroplast genomes based on depth. Subsequently, we untangled the original assembly graph to obtain the *S. sarcorrhiza* mitogenome. The mitogenome was annotated using the online programs PMGA and GeSeq [37, 38]. We uploaded the assembled sequences to the PMGA online website (http://47.96.249.172:16084/annotate.html) in FASTA format and selected 319 plant mitogenomes as reference sequences for annotations. Simultaneously, the assembly data was uploaded to the Geseq online website (https://chlorobox.mpimp-golm.mpg.de/geseq.html), where the mitogenomes of Lamiales species were used as reference sequences to annotate the *S. sarcorrhiza* mitogenome, and other parameters were set to their default values. Additionally, tRNA and rRNA were annotated using tRNAscan-SE v2.0 and BLASTN [39, 40], respectively. All annotations were manually reviewed and corrected using Geneious

v11.0.18 software [41]. Finally, the online tool PMGmap [42] was used to visualize the complete mitogenome with default parameters, including *cis-* and *trans-*splicing gene maps.

Identification and analysis of sequence repeats

The SSRs in the mitogenome were identified using the online program MISA [43]. The minimum repeat numbers for mononucleotide (Mono), dinucleotides (Di), trinucleotides (Tri), tetranucleotides (Tetra), pentanucleotides (Penta) and hexanucleotides (Hexa) were set to 10, 5, 4, 3, 3, and 3, respectively. Dispersed repeats were detected through the online tool REPuter [44], which included four types: forward (F) repeat, reverse (R) repeat, palindrome (P) repeat and complementary (C) repeat. The Hamming distance was set to 3, and the maximum number of repeats and minimum repetition size were set to 5,000 and 30, respectively. Additionally, tandem repeats (TR) in the mitogenome were detected using the online program Tandem Repeats Finder with default parameters [45]. Finally, all repetitive components were manually inspected and corrected using the "Advanced Circos" module in Tbtools v2.142 software [46].

RNA editing event detection

All PCGs in the mitogenome of *S. sarcorrhiza* were extracted. The editing events of C-to-U RNA from the PCGs were predicted through Deepred-Mt software [47], which employs a convolutional neural network (CNN) model. Finally, results with a threshold greater than 0.9 were considered reliable editing sites.

Analysis of codon usage bias

The PCGs of eight mitogenomes (*S. sarcorrhiza*, *A. marina*, *Salvia miltiorrhiza*, *Osmanthus fragrans*, *Rehmannia glutinosa*, *Utricularia reniformis*, *Plantago ovata*, and *Boea hygrometrica*) from Lamiales were extracted using PhyloSuite v1.2.3 [48]. The "relative synonymous codon usage (RSCU) analysis" module in the online tool library (http://cloud.genepioneer.com:9919) was used to compare the codon usage preferences of PCGs in the different mitogenome, and their RSCU values were calculated respectively.

Identification of migration sequences between mitochondria and plastids

There are frequent migration events between the mitochondrial genome and the plastid genome in higher plants. To detect mitochondrial plastid DNA transfers (MTPTs) in *S. sarcorrhiza*, we utilized Getorganelle v1.7.7 software [28] to assemble the complete plastid genome (accession number PQ631129). Subsequently, BLASTN software was employed to identify the homologous sequences between the mitogenome and the plastid

genome, with the e-value $\geq 1 \times 10^{-5}$. The MTPTs with matching rates $\geq 70\%$ and the lengths ≥ 30 bp were chosen for further manual annotation analysis. Ultimately, the results of the MTPTs were visualized based on the "Advanced Circos" module in Tbtools v2.142 software [46].

Collinearity analysis of mitogenome

To compare collinearity between the mitogenomes of *S. sarcorrhiza* and those of other related plants, we retrieved the mitogenomes of six Lamiales species (*A. marina* PP908999; *S. miltiorrhiza* NC_023209; *U. reniformis* NC_034982; *R. glutinosa* NC_086689; *E. longipes* PQ164709, PQ164710, PQ164711, PQ164712, PQ164713; *O. fragrans* NC_060346) from the NCBI database. Each mitogenome was compared in pairs with that of *S. sarcorrhiza* via MUMmer4 software [49], and only collinear fragments with an alignment length of at least 500 bp were kept for further analysis. The collinearity results were then visualized with NGenomeSyn v1.39 software [50].

Phylogenetic analysis

To ascertain the phylogenetic position of *S. sarcorrhiza*, we constructed phylogenetic trees using both plastid and mitochondrial sequences, respectively. We retrieved the plastid genomes of an additional 46 species from the NCBI database. However, since the *Strobilanthes* genus lacks mitogenomes, we could only obtain sequences for 38 species (Table S1). We extracted 10 conserved mitochondrial PCGs and 20 conserved plastid PCGs using PhyloSuite v1.2.3 [48]. Subsequently, these shared PCGs were multiple aligned through MAFFT v7.149b software [51] and sequence trimming was performed using trimAl v1.2 software [52]. We used IQ-TREE v2.2.0.3 software [53] to build the maximum likelihood (ML) trees, with a bootstrapping value of 5,000 repeats. Based on the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) scores, the GTR + F + I + G4 model was selected as the optimal model for ML phylogenetic tree construction. Finally, the online platform ChiPlot [54] was used to visualize the phylogenetic tree of both plastid and mitochondrial genomes.

Results

General characteristics of the mitogenome

A total of 62.63 Gb (76.6×) Illumina clean reads, and 10.5 Gb (12.8×) of HiFi reads with an N50 of 18,337 bp, were generated (Table S2-S3). Initially, based on short clean reads, the genome size of *S. sarcorrhiza* was predicted to be 820 Mb and it might be a diploid species (Fig. 1A and B). Subsequently, a draft mitogenome was assembled using high-precision long-read sequence data. The repetitive sequences were sequentially unwrapped based

on the coverage of contigs, resulting in a linear structure with a size of 617,134 bp (Fig. 1C). The GC content of the mitogenome was 42.56%, and the intergenic region accounted for 93.7% of the mitogenome, indicating a very sparse gene density (Fig. 1D). A total of 35 PCGs were annotated, including 24 core mitochondrial genes and 11 variable genes (Table S4). Additionally, 3 rRNAs and 19 tRNAs genes were annotated. The core genes comprised 5 ATP synthase genes, 4 cytochrome c biogenesis genes, 1 ubichinol cytochrome c reductase (*cob*), 3 cytochrome c oxidase genes, 1 maturase (*matR*), 1 transport membrane protein (*mttB*) and 9 NADH dehydrogenase genes. The variable genes included 3 large ribosomal proteins, 6 small ribosomal proteins, and 2 succinate dehydrogenase genes (*sdh3*, *sdh4*). Notably, the genes *ccmFC*, *rrn5*, *trnA-UGC*, *trnH-GUG*, *trnI-CAU* and *trnM-CAU* each had 2 copies. Among the 35 PCGs, 10 genes contained introns, with 5 genes (*ccmFC*, *cox1*, *cox2*, *rps10*, *rps3*) containing only 1 intron, and 5 genes (*nad1*, *nad2*, *nad4*, *nad5*, *nad7*) each containing multiple introns (Fig. 1D). Furthermore, additional analysis of the mitogenome revealed 7 cis-splicing genes within PCGs and 3 trans-splicing genes (*nad1*, *nad2*, *nad5*) within NADH dehydrogenase genes, indicating that these three genes may be related to adaptive evolution.

Repeated sequence analysis of the mitogenomes between *S. sarcorrhiza* and related species

Repetitive sequences are prevalent in plant mitogenomes and are primarily categorized into three types: SSRs, dispersed repeats and TRs (Fig. 2A). The mitogenome of *S. sarcorrhiza* contained 122 SSRs, encompassing 5 types, with Tetra repeats being the predominant repeating unit (constituting 41.8% of all repeats). This is comparable to the mitogenomes of the other four Lamiales species (*A. marina*, *R. glutinosa*, *S. miltiorrhiza*, *U. reniformis*) (Fig. 2B). In contrast, the other two related species (*O. fragrans*, *P.ovata*) predominantly feature Mono repeats, and *P. ovata* had a significantly higher total number of SSRs (258 loci) than the other seven related species (ranging from 87 to 164 loci).

In the mitogenome of *S. sarcorrhiza*, we identified 25 TRs, ranging from 5 in *B. hygrometrica* to 65 in *O. fragrans* (Fig. 2C). These TRs were short, with lengths varying between 25 bp and 74 bp, and were sparsely distributed. In contrast, dispersed repeats exhibited characteristics opposite to those of tandem repeats. They were evenly dispersed throughout the mitogenome and had longer lengths (Fig. 2A). The *S. sarcorrhiza* mitogenome contained 1,482 dispersed repeats, exceeding the number found in seven other similar species, ranging from 244 repeats in *B. hygrometrica* to 1,395 in *P. ovata*. The total length of these dispersed repeats was 108,514 bp, constituting 17.58% of the entire *S. sarcorrhiza* mitogenome.
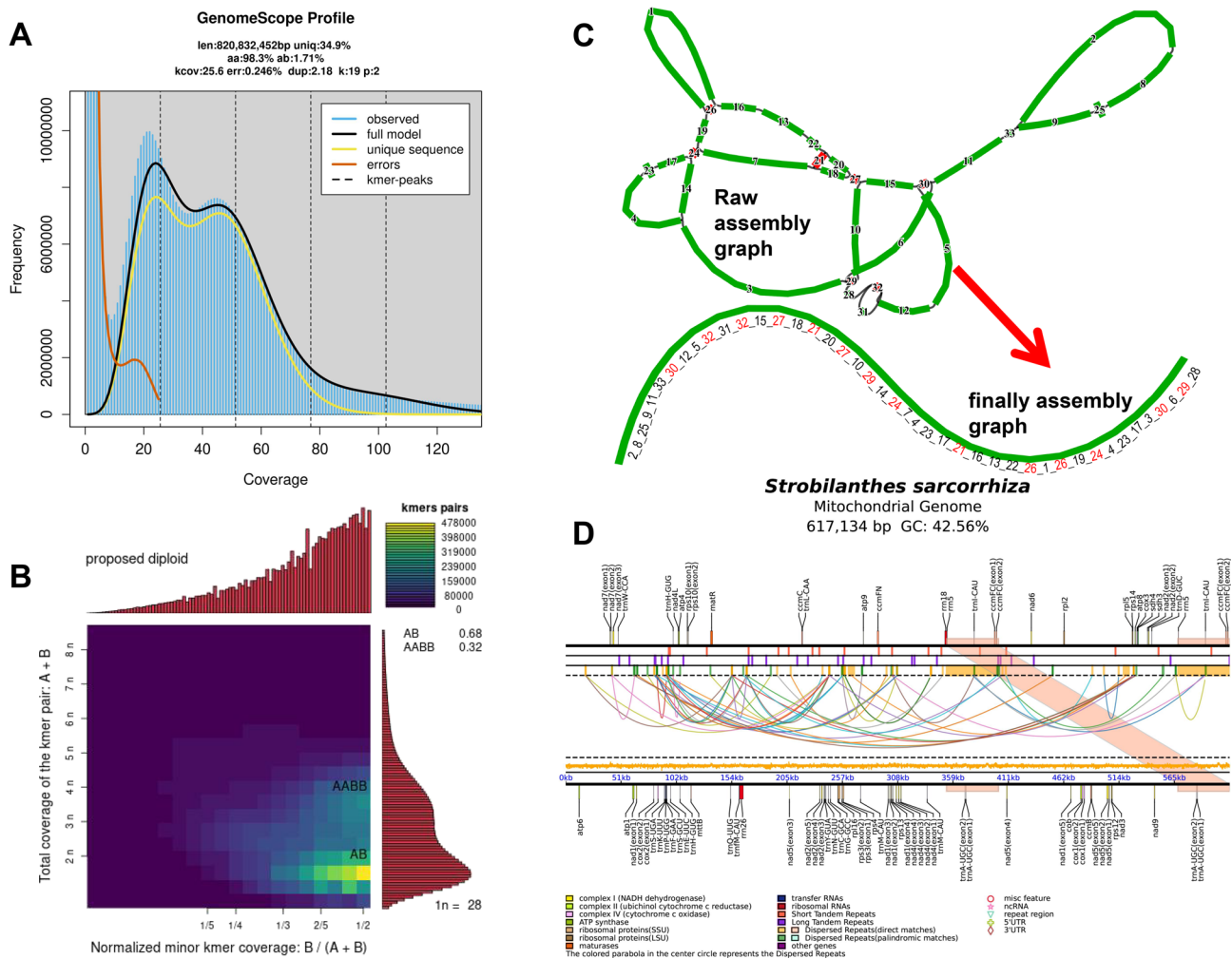
Fig. 1 Assembly and annotation of the mitogenome of *S. sarcorrhiza*. (**A**) The genomic size was estimated based on k-mer analysis. (**B**) The genomic ploidy was inferred based on k-mer data. (**C**) The original and finally assembly graphs of the mitogenome. (**D**) The annotation of mitogenome

Most sequences were shorter than 50 bp (1,348 repeats, 90.96%), and only 7 repeats (0.47%) were larger than 200 bp, a pattern similar to that observed in the mitogenomes of seven other related species (Fig. 2D). Moreover, the types of dispersed duplication exhibited similar patterns among the eight related species. Except for *P. ovata* and *S. militorhiza*, which contained 4 types (F, R, P, C) and 3 types (F, R, P) respectively, all other species contained only F and P types. However, among the eight species, the F (accounting for 32.90-63.82% of all) and P (accounting for 34.55-55.74% of all) types still had the highest occurrence frequencies, a consistency observed across these species (Fig. 2C).
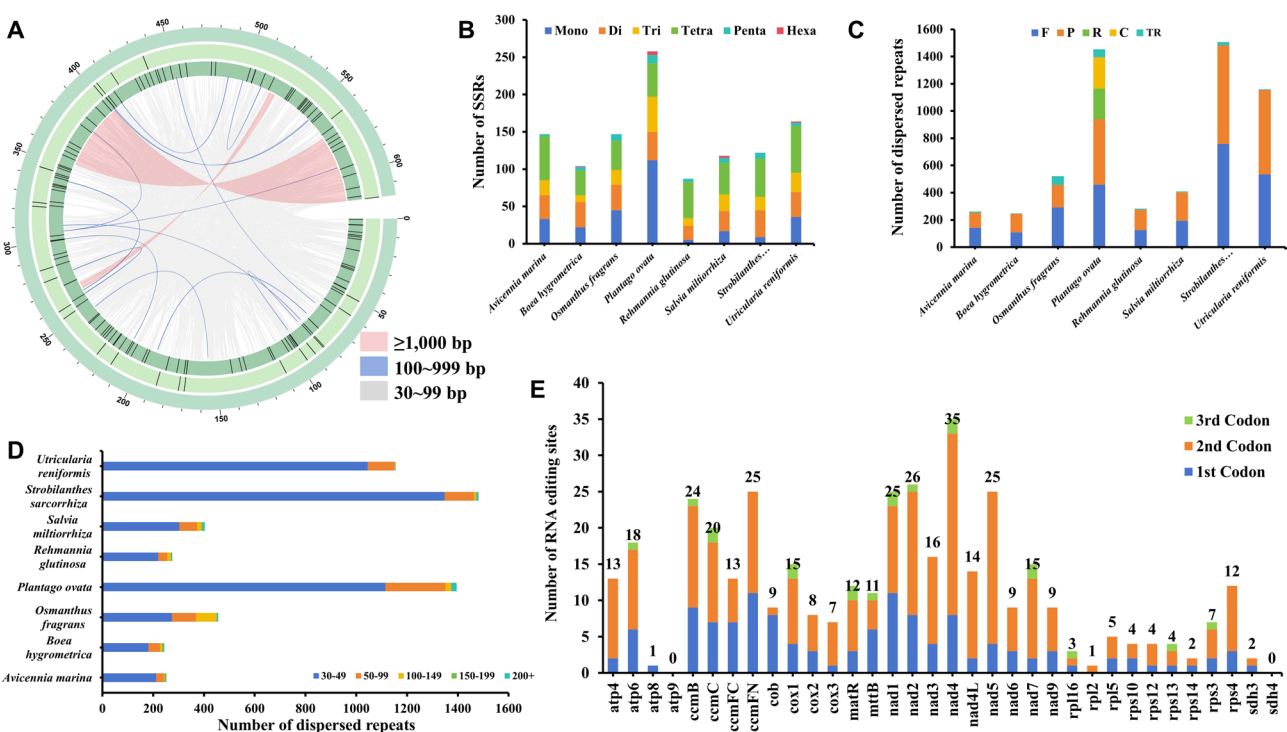
Prediction of RNA editing sites in the mitogenome

We predicted potential C-to-U RNA editing sites for 35 PCGs in the *S. sarcorrhiza* mitogenome to gain a deeper understanding of gene expression (Fig. 2E). A total of 394 RNA editing sites were identified, with the *nad4* gene

containing the most RNA editing sites (35 sites), followed by *nad2* gene (26 sites). Most ribosomal proteins had fewer editing sites, and the *rpl2* gene had just one editing site. Interestingly, the second codon was edited most frequently, with 248 sites (62.94% of the total), whereas the third codon had the fewest edits (19 sites, 62.94% of the total), significantly fewer than the first and second codons. Typically, these RNA editing events resulted in changes to the encoded amino acids, with most involving the conversion of hydrophobic amino acids, which can enhance protein stability.

Analysis of selection pressure of shared PCGs from Lamiales species

To investigate the impact of selection pressure on the evolution of the mitogenome of *S. sarcorrhiza*, we selected seven closely related species and analyzed the ratio of Ka to Ks in the 34 PCGs shared with *S. sarcorrhiza* mitogenome (Fig. 3). The analysis of selection

Fig. 2 Identification of repetitive sequences and RNA editing sites in the mitogenomes. (**A**) Distribution of repetitive sequences in the mitogenome of *S. sarcorrhiza*. From the outside to the inside were the mitogenome, TRs, SSRs, and dispersed repeats. (**B**) Comparative analysis of SSRs loci among the mitogenomes of *S. sarcorrhiza* and 7 related species. (**C**-**D**) Comparative analysis of dispersed repeats loci among the mitogenomes of *S. sarcorrhiza* and 7 related species. (**E**) Identification of RNA editing events in the mitogenome
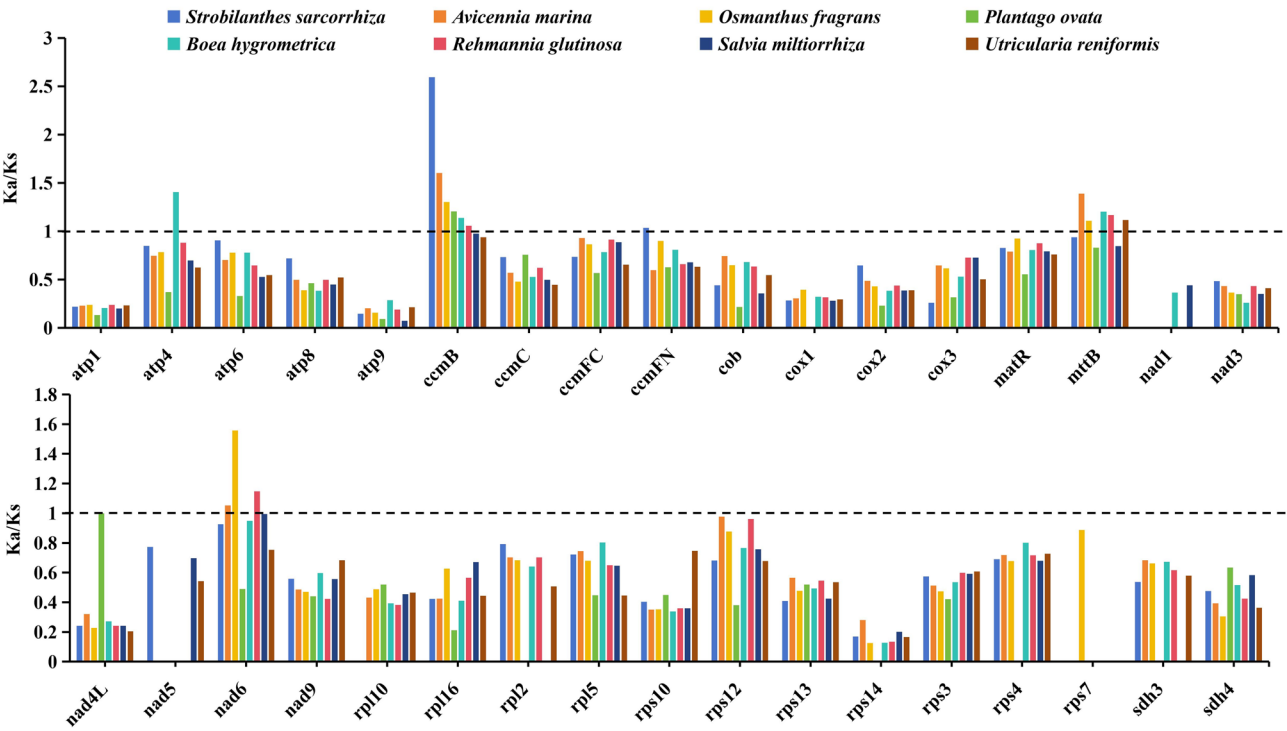
Fig. 3 Selection pressure analysis of 34 PCGs from the mitogenomes of *S. sarcorrhiza* and 7 related species

pressure revealed that the Ka/Ks ratio trend in the 34 shared PCGs was similar among the eight species, with most genes experiencing purifying selection, and the Ka/Ks ratio values were less than 1. Interestingly, only the Ka/Ks ratio values of three genes (*ccmB*, *mttB*, *nad6*) were higher than 1 in most species, and the *ccmB* gene in *S. sarcorrhiza* mitogenome exhibited the highest value (2.59), indicating that it had undergone potential positive selection in response to environmental adaptation.
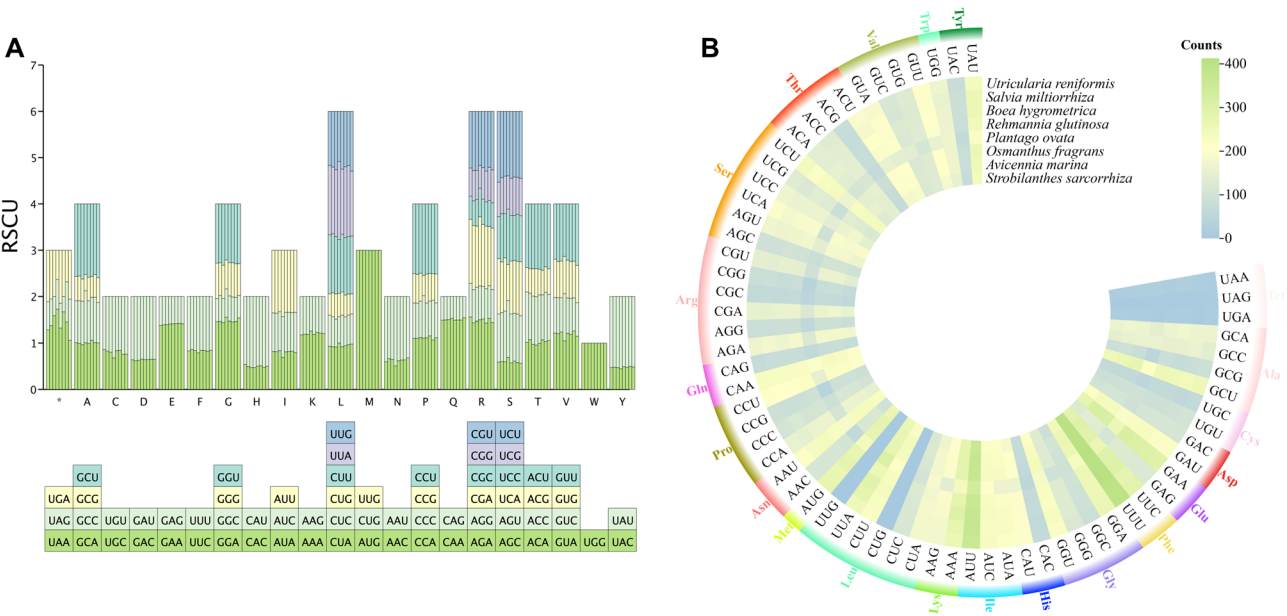
Analysis of codon usage bias

In total, 82,646 codons were identified from the mitochondrial PCGs of 8 Lamiales species, distributed in 8,817 in *P. ovata* and 10,829 in *A. marina*. There were 64 different types of these codons, encoding a total of 20 amino acids and 1 stop codon. Of these, UUU was the most commonly used codon, with 400–413 occurrences. Among the 20 amino acids, the trends of codon usages bias were consistent across the 8 species (Fig. 4B). Leucine exhibited the highest codon usage bias, with 1,071 to 1,152 codons, which accounted for 10.78% of the all. Serine followed, with 693-1,003 codons, representing 9.11% of the total, whereas cysteine had the fewest codons, with 144–168, constituting 1.48% of the total.

In addition, we calculated the RSCU values for the 64 codons across various species. The results indicated that 32 codons were utilized more frequently than anticipated, as evidenced by RSCU values exceeding 1 (Fig. 4A). Conversely, the remaining 32 codons were used less frequently than expected (RSCU < 1). Interestingly,

both the lowest and highest preferences were observed in the same amino acid, methionine. Specifically, the codon AUG exhibited the most pronounced usage bias, whereas codons CUG and UUG demonstrated the weakest preference. Notably, tryptophan exhibited no codon preference (RSCU = 1). Overall, with the exception of tryptophan, all amino acids displayed codon usage preferences. Most amino acids possessed at least two distinct codons, with arginine, serine, and leucine each having six unique codons.

Genes transfer within the organelle genome

The mitogenomes of higher plants contained a substantial number of sequences that have migrated from nuclear and plastid genomes. In the organelle genome of *S. sarcorrhiza*, we identified 37 MTPTs (total length of 34,978 bp), all with a similarity greater than 73%, accounting for 24.17% of the plastid genome and 5.67% of the mitogenome, respectively. Among these migration sequences, 21 fragments were larger than 500 bp, with the longest being 5,120 bp (Fig. 5). By annotating these fragments, a total of 34 plastid gene fragments were identified to be located in MTPTs (Table S4), including 22 PCGs fragments (*ycf1*, *ycf2*, *ycf4*, *psbA*, *psaA*, *psaB*, *psaI*, *petA*, *atpA*, *atpB*, *atpE*, *rpl16*, *rpl20*, *rpl23*, *rps18*, *rpoB*, *rpoC1*, *ndhA*, *ndhF*, *ndhH*, *accD*, *cemA*), 9 tRNA gene fragments (*trnA-UGC*, *trnD-GUC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnM-CAU*, *trnN-GUU*, *trnW-CCA*, *trnV-GAC*), and 3 rRNA gene fragments (*rrn4.5 S*, *rrn16S*, *rrn23S*). However, only 13 genes (*psbA*, *psaI*,

Fig. 4 Codon usage bias analysis of mitogenomes from *S. sarcorrhiza* and 7 related species. (**A**) The RSCU values of *S. sarcorrhiza* and 7 related species. Each bar represents a species, from left to right, namely *S. sarcorrhiza*, *A. marina*, *O. fragrans*, *P. ovata*, *B. hygrometrica*, *R. glutinosa*, *S. miltiorrhiza*, and *U. reniformis*. (**B**) Statistics on the number of amino acids and codons corresponding to PCGs in different species
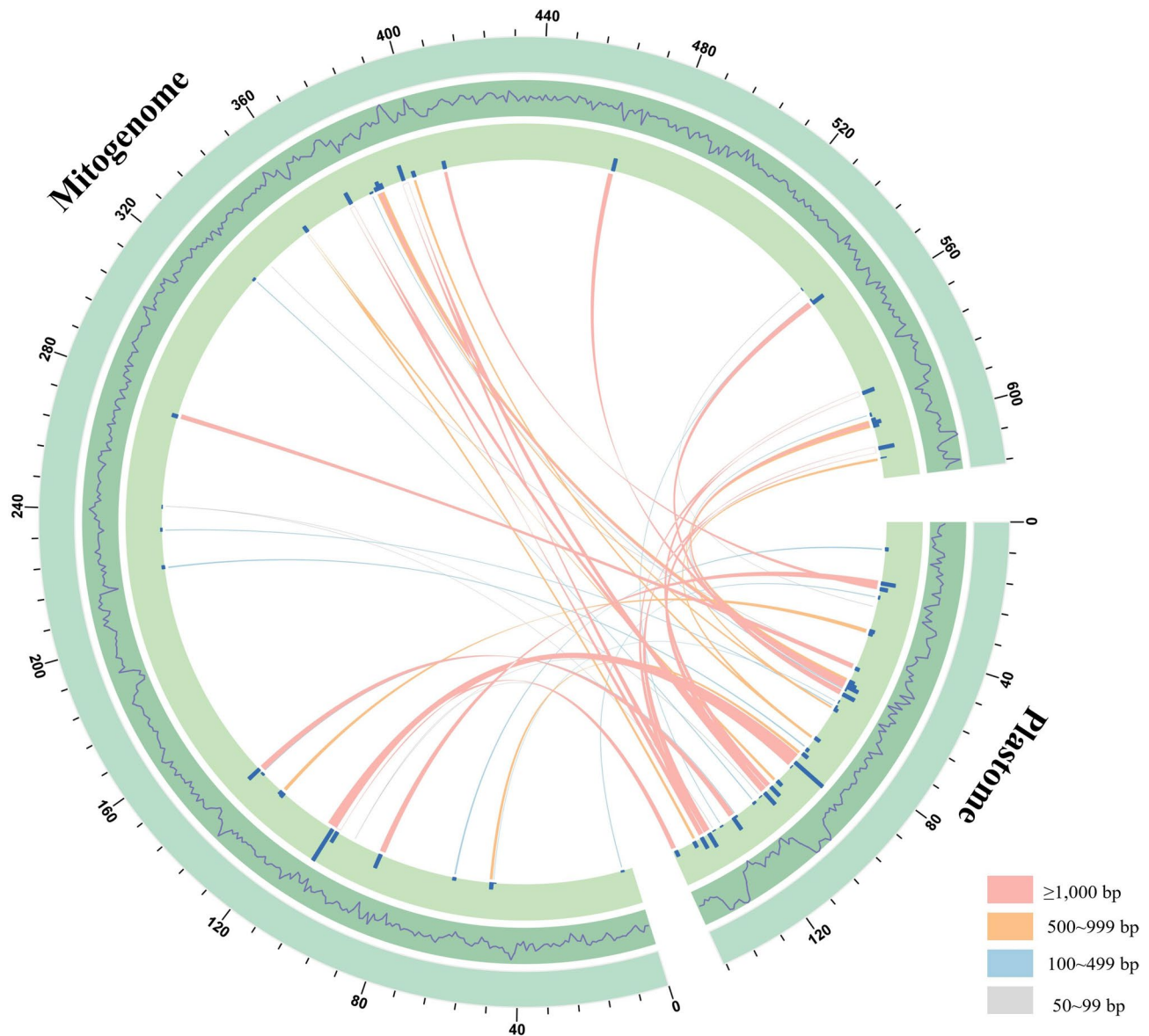
Fig. 5 Detection of migration fragments between mitogenome and plastid genome in *S. sarcorrhiza*. Two arcs of different lengths represented the mitochondrial and plastid genomes. The curve on the outer circle represented the GC content of the two genomes, and the histogram on the inner circle represented the length of MTPTs. The line between the two arcs represents MTPTs. The detailed information was recorded in Table S4

ycf4, atpE, cemA, trnA-UGC, trnD-GUC, trnH-GUG, trnM-CAU, trnN-GUU, trnI-CAU, trnW-CCA, rrn4.5 S) remained intact after transfer.

Collinearity analysis of mitochondrial genomes

In this study, we used the mitogenome of *S. sarcorrhiza* as the primary focus and conducted a collinearity analysis with the mitogenomes of six other related species (Fig. 6). Upon comparison with the mitogenome of *S. sarcorrhiza*, we identified 31, 43, 24, 38, 57, and 35 homologous fragments in *S. miltiorrhiza, A. marina, U. reniformis, R. glutinosa, E. longipes,* and *O. fragrans*, respectively, all of which were larger than 500 bp. Notably, the comparison

between *E. longipes* and *S. sarcorrhiza* revealed the highest number of homologous fragments, with a total length of 90,241 bp, constituting 11.14% of the *E. longipes* mitogenome. Conversely, the collinearity between the *U. reniformis* and *S. sarcorrhiza* mitogenomes was the poorest, with only 24 homologous fragments (total length of 27,608 bp), representing 3.22% of the *U. reniformis* mitogenome. Furthermore, numerous sequence rearrangements were also identified within the mitogenomes of *S. sarcorrhiza* and the other species.
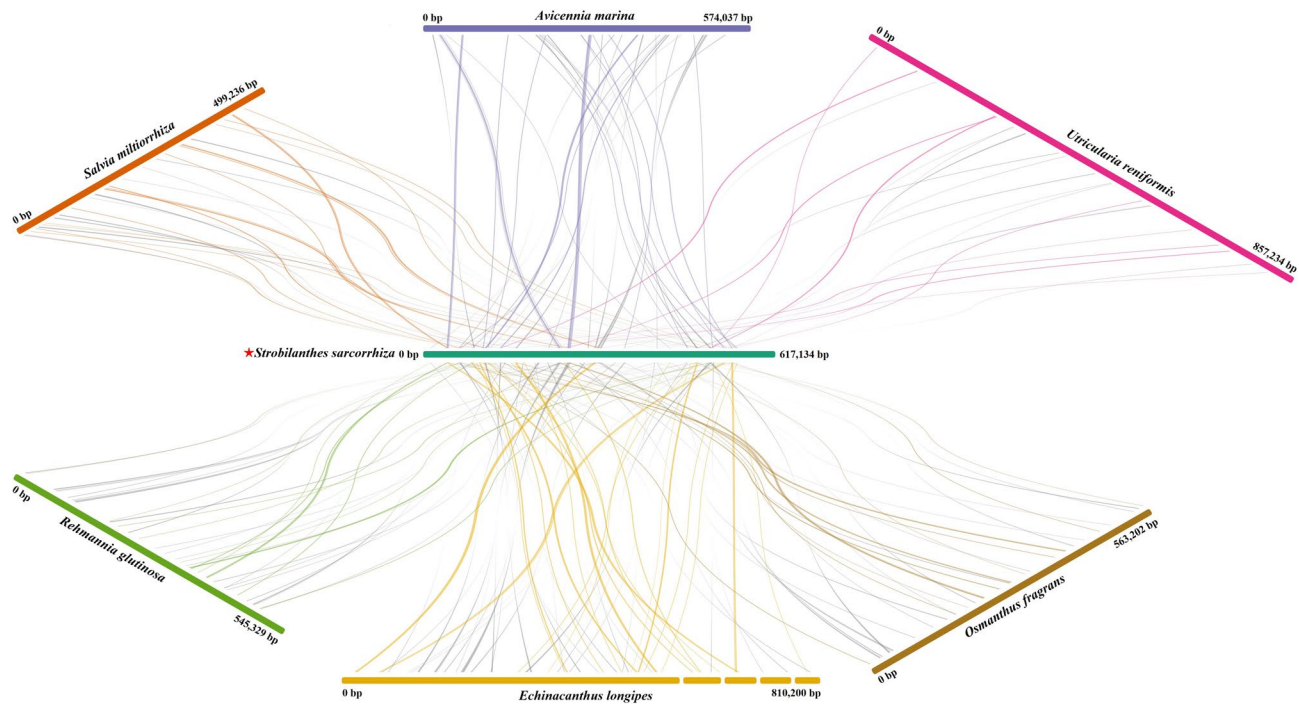
Fig. 6 Collinearity analysis among the mitogenomes of *S. sarcorrhiza* and 6 related species. Different colored bars represented the mitogenomes of different plants. Among them, the colored line between the mitogenome of *S. sarcorrhiza* and other 6 species represented good similarity, and the gray line represented reversal. All colinear fragments were greater than 500 bp
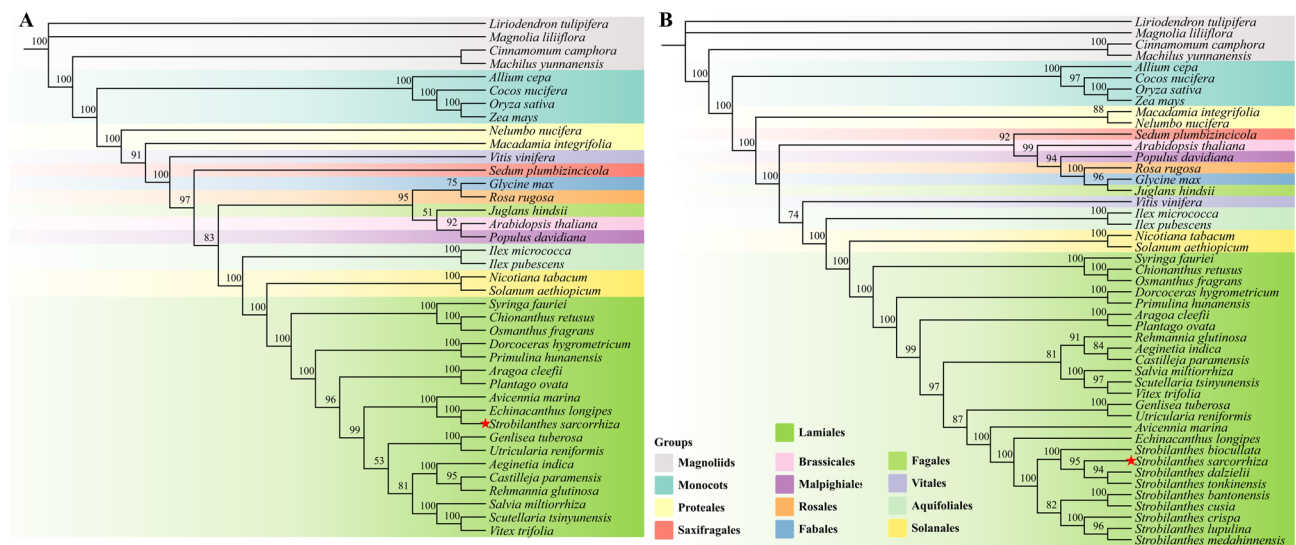
Fig. 7 Phylogenetic analysis. (**A**) The ML phylogenetic tree constructed based on PCGs shared in the mitogenomes of *S. sarcorrhiza* and 38 species. (**B**) The ML phylogenetic tree constructed based on PCGs shared in the plastid genomes of *S. sarcorrhiza* and 46 species. Support rates for different nodes were displayed on branches

Phylogenetic analysis

To further investigate the phylogenetic status of *S. sarcorrhiza*, an ML tree was constructed based on 10 conserved mitochondrial PCGs from 39 plant species (Fig. 7A). Phylogenetic analysis revealed that *S. sarcorrhiza*, *A. marina*, and *E. longipes* formed a clade (belonging to Acanthaceae), and together with other family species, constituted

the Lamiales. The reconstructed ML tree proved that the phylogeny of *S. sarcorrhiza* was fully supported (bootstrap = 100), and the topology of the entire tree exhibited a high degree of consistency with the APG IV classification system. Due to the absence of mitogenomes from plants of the same genus as *S. sarcorrhiza*, we further examined the phylogenetic relationships within the genus

using plastid sequences to reconstruct the phylogenetic tree. An ML phylogenetic tree was constructed based on 20 conserved plastid PCGs from 47 plant species (Fig. 7B). The results indicated that *S. sarcorrhiza* and 8 species from the same genus initially clustered into a branch, which then formed a clade with *A. marina* and *E. longipes*, and exhibited a closer genetic relationship with *S. tonkinensis*, *S. dalzielii*, and *S. biocullata*. The ML tree constructed from the mitogenomes and the plastid genomes showed good consistency in the phylogenetic relationship about *S. sarcorrhiza*, thus verifying the accuracy of the phylogenetic analysis.

Discussion

Plant mitogenomes exhibit complex structural characteristics, with variations across different species, including linear, circular, branched and network structures [55]. In some species, multiple structures can coexist simultaneously due to high-frequency recombination mediated by repeated sequences [56, 57]. Numerous studies have elucidated the diversity of mitogenome conformations in various plants. For instance, the mitogenome of *Cinnamomum camphora* is a linear structure spanning 900,894 bp [58], whereas the *Populus simonii* mitogenome consists of three circular structures of different sizes [59]. The mitogenome of *Lilium tsingtauense* is particularly complex, spanning 1,125,108 bp and containing 27 independent circular chromosomes [60]. In this study, we successfully assembled the mitogenome of *S. sarcorrhiza* for the first time. Its linear structure diverges from those of other plants within the same family. For example, *A. marina* possesses a typical circular structure [31], and *E. longipes* features a multi-circular compound structure, suggesting that the mitogenomes of Acanthaceae plants may have undergone divergent evolution, contributing to the high plant diversity observed. The sizes of mitogenomes are influenced by the accumulation of various repetitive sequences and horizontal gene migration [61]. The mitogenome of *S. sarcorrhiza* spans 617,134 bp, which is comparable to that of *A. marina* (574,037 bp) and falls within the middle range of Lamiales mitogenome sizes (274,779 bp to 857,234 bp). Repeated and migrating sequences dictate the diversity in the sizes of Lamiales mitogenomes. In the *S. sarcorrhiza* mitogenome, 57 functional genes have been annotated, including 35 PCGs, 19 tRNAs and 3 rRNA genes. Notably, GC content is a crucial indicator for evaluating species evolution and impacts amino acids composition [62]. The GC content of the *S. sarcorrhiza* mitogenome is found to be approximately 42.56%, which aligns with the GC content range of Lamiales species (42.3-45.6%), suggesting that despite the structural and size diversity of Lamiales mitogenome throughout evolution, the GC content has remained relatively stable. Comparable

outcomes are observed in Apiales species and Rosoideae species [55, 63]. In genetic research, the ratio of Ka to Ks for a gene can reflect its relative molecular evolution rate. Ka/Ks > 1, Ka/Ks = 1, and Ka/Ks < 1 indicate that the gene has been subjected to positive selection, neutral selection and purifying (negative) selection, respectively [64]. In our study, among the 34 PCGs from *S. sarcorrhiza* and 7 related species, most genes appeared to be under purifying selection, while a few showed signs of positive selection, with slight variations among different species. Notably, the Ka/Ks values for *ccmB*, *mttB* and *nad6* genes in most species were greater than 1, and the *ccmB* gene in *S. sarcorrhiza* had the highest value, suggesting it was under the least selection pressure and might be a vital gene for life processes.

Codons are the fundamental units responsible for accurately identifying and transmitting genetic information [65]. Codon usage bias refers to the uneven distribution of synonymous codons that encode the same amino acid. This phenomenon is widespread among natural organisms and is influenced by various factors, including natural selection and mutation pressure [66]. By examining codon usage preferences, we can uncover the evolutionary patterns of genes and predict regulatory mechanisms during gene expression [67]. ATG serves as a common start codon in all plants, although a small number of PCGs in the mitogenome use ACG or CTG as their start codon [68]. For instance, the *nad4L* gene in the *S. sarcorrhiza* mitogenome initiates with ACG. Similar findings have been reported in other plants, such as *Rubus ideus* [55], *Ilex rotunda* [69] and *Lycopodium japonicum* [70]. This phenomenon may be attributed to RNA editing events. The RSCU value for the encoded amino acid was greater than 1, indicating that A/T nucleotide were significantly enriched at the third codon position of the *S. sarcorrhiza* gene sequence. The usage frequency of NNA and NNT codons was similar to that observed in many other species, including *Rubus suavissimus* [71], *Punica granatum* [72], and *Angelica sinensis* [63]. The pronounced bias towards A/T nucleosides at the third codon position may be a shared characteristic among the mitogenomes of various plants. Codon usage patterns reveal the non-random preference for synonymous codons by various organisms or genes, influencing mRNA stability, translation efficiency, and protein expression levels [73, 74]. In plant mitochondria, unique evolutionary pressures and genomic characteristics, such as high AT content and RNA editing, make codon usage patterns particularly significant for gene regulation. The first impact to consider is that of codon usage on translation efficiency. High-frequency codons typically correspond to abundant tRNAs in cells, enabling ribosomes to decode swiftly and minimize translation pauses, for instance, CCA (Pro), UUU (Phe), and others in plant

mitochondria. The abundance of tRNAs can cause ribosomes to hesitate, potentially affecting the expression levels of protein folding box genes. Furthermore, codon usage aligns with that of highly expressed genes, which are generally more efficient in translation [75], such as the respiratory chain complex genes *cox2*, *cob*, and others. Conversely, some regulatory genes may be constrained by translation efficiency and thus regulate expression levels.

A large number of repetitive sequences are often found in the mitochondria of higher plants, including simple sequence repeats, tandem repeats and dispersed repeats [69–71]. These sequences play a crucial role in evolution. Among them, irreversible recombination mediated by short repeats with low recombination activity plays an important role, while reversible recombination mediated by long repeats with high recombination activity leads to the continuous expansion and complexity of the genome [76]. Compared to seven related species, the proportion of repeated sequences in *S. sarcorrhiza* was notably higher. Specifically, in the *S. sarcorrhiza* mitogenome, repetitive sequences made up 17.97%, which was comparable to the proportion in *Plantago ovata* (19.30%), while in other plants, this figure ranges from 2.39% to 8.49%. Additionally, extensive genetic material exchanges between organelle genomes can be observed in angiosperms, including sequence fragments transferred from plastids to mitochondria (MTPTs) and gene fragments transferred from nuclei to mitochondria [77]. MTPTs typically constitute 1-12% of the mitogenome. In *S. sarcorrhiza*, 34,978 bp MTPTs were found, which account for 5.67% of the mitogenome size, classifying them as medium-sized. Therefore, it was further proved that the size of the *S. sarcorrhiza* mitogenome was mainly affected by repetitive sequences and migrating sequences. Compared to previous studies, we observed variations in migrating fragments among different species. Annotations of MTPTs in *S. sarcorrhiza* revealed that a large number of fragments originated from plastids, providing the mitogenome with rich unnatural sequences. The complete gene may remain functional within the mitogenome. However, migrated fragments often experience varying degrees of sequence loss, and the transfer of sequences to mitochondria can also lead to the formation of pseudogenes [55]. In this study, we detected 7 complete tRNA genes within MTPTs, which represented 36.8% of mitochondrial tRNA genes. Compared to intact PCGs (14.2%), tRNA genes were more conserved in the mitogenome, indicating that they might play an indispensable role in the mitogenome. In general, sequence transfer plays a significant role in plant evolution, impacting not only the size of the mitogenome, but also increasing its genetic diversity.

RNA editing events, which involve post-transcriptional modifications of RNA sequences through substitutions, deletions or insertions, significantly contribute to transcriptome diversity and are essential steps in plant mitochondrial gene expression [78, 79]. Current research indicates that substitution is the most frequent type of editing, with the primary form being C-to-U [70, 80]. Previous studies have found that there are approximately 491 RNA editing sites in 34 PCGs of *Oryza sativa* [81], approximately 421 RNA editing sites in 35 PCGs of *Acer truncatum* [82], approximately 283 RNA editing sites in *Spodiopogon sagittifolius* [83], around 543 RNA editing sites in *Ilex metabaptista* [84], and roughly 401 RNA editing sites in *E. longipes* [85]. In the mitogenome of *S. sarcorrhiza*, 394 RNA editing sites were identified. We have observed that the number of RNA editing sites varies significantly between different species, while the number tends to be relatively conservative among closely related species. The number of RNA editing sites for different genes can vary greatly, even among closely related species. For instance, the *ccmB* gene of *E. longipes* has the most editing sites (36), whereas *nad4L* in *S. sarcorrhiza* has the highest number of RNA editing sites (35). A similar phenomenon has been noted in *R. suavissimus* and *C. camphora* [58, 71]. Such editing may affect the stability of the mitochondrial respiratory chain complex enzyme. Collinearity analysis is crucial to elucidating the evolutionary relationships among species [86]. Additionally, RNA editing can cause alterations in protein function [87]. For instance, editing within the coding region may result in non-synonymous mutations that modify protein activity; it can also disrupt enzyme active sites or protein interaction interfaces; furthermore, editing frequently introduces premature stop codons, leading to a loss of function. In the mitogenome of *S. sarcorrhiza*, RNA editing primarily occurs at the first and second codons, and the start codon of *nad4L* is altered to ACG as a result of RNA editing, a phenomenon similar to that observed in most plants [82, 85]. We conducted the collinearity tests between *S. sarcorrhiza* and six related species, respectively. The results indicated that *S. sarcorrhiza* and plants of the same family shared more homogeneous fragments, but the orders of the mitogenome sequences were extremely non-conservative, suggesting that they had experienced frequent genome recombination over a long evolutionary period. Previous studies have found that gene rearrangement is a driving factor for the remodeling of the Rosaceae mitogenome [88]. We had also observed similar results in the Lamiales mitogenome, which may be a potential driving force for the evolution of mitogenome.

The mitogenome is proof of the unique evolutionary trajectory of angiosperms [89]. The content, structure and genetic arrangement of plastomes are important for exploring the evolutionary relationships among plants [90]. In most terrestrial plants, the plastid genome size

remains relatively constant, and gene losses are infrequent during organism evolution [91]. Mitochondrial genomes, on the other hand, are relatively complex and highly variable in size, yet their PCGs exhibit lower substitution rates and higher homogeneity [92]. These characteristics make the mitogenome more adept at uncovering older and deeper phylogenetic relationships within the tree of life [93]. In certain plants, discrepancies in topological structures between phylogenetic trees arise due to the distinct evolutionary paths of plastid and mitochondrial genomes [69, 94]. This study explored the phylogenetic placement of *S. sarcorrhiza* from both mitochondrial and plastids. The results showed that the phylogenetic trees constructed from plastids perspectives. The findings indicated that phylogenetic trees constructed from both plastids and mitochondria supported the classification of *S. sarcorrhiza* within the family Acanthaceae, with the plastid genome further indicating its genus as *Strobilanthes*, forming a sister group with *S. tonkinensis*, *S. dalzieli*, and *S. biocullata*, rather than *Championella*. Intriguingly, despite the high consistency in the phylogenetic position of *S. sarcorrhiza* and a very high support rate, the trees revealed minor inconsistencies in the overall phylogenetic relationships of angiosperms. For instance, the phylogenetic placement of *Vitis vinifera* was closely related to Aquifoliales in the plastid genome but closely related to Saxifragales in the mitogenome. Overall, the phylogenetic relationships inferred from mitochondrial data align more closely with the APG IV classification system. These inconsistencies might be attributed to the relatively independent genetic system of mitochondria, or they could be a result of species hybridization and ILS [95]. More accurate guesses require a more comprehensive sampling strategy and more detailed evolutionary analysis.

Conclusions

This study assembled and annotated the mitogenome of *S. sarcorrhiza* using HiFi reads, and examined the detailed characteristics of the genome, such as GC content, codon usage preference, repetitive sequences, RNA editing events, phylogenetic relationships, sequence migration, and rearrangement. The mitogenome of *S. sarcorrhiza* featured a linear structure spanning 617,134 bp, comprising 35 PCGs, 19 tRNAs, and 3 rRNA genes, with a GC content of 42.56%. A comprehensive phylogenetic analysis indicated that it belongs to the genus *Strobilanthes*, not *Championella*. This study offers a crucial foundation for future molecular breeding, germplasm resource innovation, and phylogeny research on *S. sarcorrhiza*.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06731-3.

Supplementary Material 1

Author contributions
Y.J.S., P.L., J.S., M.X.L., J.Y.J., Y.X., Z.C., and W.Z., conceived and performed the original research project. Y.J.S., P.L., M.X.L., J.Y.J., Y.X., collected samples and performed the experiments. Y.J.S., Z.C., and J.S. designed the experiments and analyzed the data. Y.J.S. refined the project and wrote the manuscript with contributions from all authors. Z.C. and W.Z. supervised the experiments and revised the writing. Z.C. and W.Z. obtained the funding for the research project. All authors have read and agreed to the published version of the manuscript.

Data availability
The accession numbers of mitochondrial and plastid genomes of Strobilanthes sarcorrhiza in Gene Bank are PQ601285 and PQ631129, respectively.

Declarations

Ethics approval and consent to participate
We collected fresh leaf samples of *Strobilanthes sarcorrhiza* for this study. The plant samples and experimental research comply with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Li L, Deng Y. Chromosome numbers of six species of Championella group in Strobilanthes (Acanthaceae) from China. CYTOLOGIA. 2024;89(3):257–60.
2. Grant WF. A cytogenetic study in the Acanthaceae. Brittonia. 1955;8(2):121–49.
3. Good R. The geography of the flowering plants. 1947.
4. Manzitto-Tripp EA, Darbyshire I, Daniel TF, Kiel CA, McDade LA. Revised classification of Acanthaceae and worldwide dichotomous keys. Taxon. 2022;71(1):103–53.
5. Yu H, Li T-n, Ran Q, Huang Q-w, Wang J. Strobilanthes Cusia (Nees) Kuntze, a multifunctional traditional Chinese medicinal plant, and its herbal medicines: a comprehensive review. J Ethnopharmacol. 2021;265:113325.
6. Akbar S. Andrographis paniculata: a review of Pharmacological activities and clinical effects. Altern Med Rev. 2011;16(1):66–77.
7. Carine MA, Scotland RW. Classification of strobilanthinae (Acanthaceae): trying to classify the unclassifiable? *TAXON* 2002, 51(2):259–79.
8. McDade LA, Daniel TF, Kiel CA, Vollesen K. Phylogenetic relationships among acantheae (Acanthaceae): major lineages present contrasting patterns of molecular evolution and morphological differentiation. Syst Bot. 2005;30(4):834–62.
9. Scotland R, Vollesen K. Classification of acanthaceae. Kew Bull 2000:513–89.
10. McDade LA, Daniel TF, Masta SE, Riley KM. Phylogenetic relationships within the tribe Justicieae (Acanthaceae): evidence from molecular sequences, morphology, and cytology. Ann Mo Bot Gard. 2000;87(4):435–58.
11. Moylan EC, Bennett JR, Carine MA, Olmstead RG, Scotland RW. Phylogenetic relationships among Strobilanthes s.l. (Acanthaceae): evidence from ITS NrDNA, trnL-F CpDNA, and morphology. Am J Bot. 2004;91(5):724–35.

12. Chen F-l, Deng Y-f. Xiong Z-b, Ran J-c: Strobilanthes hongii, a new species of Acanthaceae from Guizhou, China. 2019.

13. Chen J-T, Huang X-H, Lv Z-Y, Kuang T-H, Luo J, Deng Y-F, Deng T. Strobilanthes sunhangii (Acanthaceae), a new species from Tibet, China. *PhytoKeys* 2020, 166.

14. Chen R, Fan J, Wu Y, Huang X, Zhang W, Xu Y, Zhang Y, Li L, Wang C, Yu M. Strobilanthes sarcorrhiza root phenolic extract prevent diabetic nephropathy in mice by regulating NF-κB/IL-1β signaling and glycerophospholipid metabolism. J Pharm Biomed Anal. 2025;253:116534.

15. Tian Z, Zhang Y, Wang X, Xb C, Li W, Tan Xh C, Jj L, Hz. Wen Hm: chemical profile and difference analysis of four parts of Strobilanthes sarcorrhiza by ultrafast flow liquid chromatography coupled with quadrupole-time-of-flight mass spectrometry and multivariate statistical analysis. Biomed Chromatogr. 2024;38(5):e5838.

16. Yunfei D, Nianhe X. Validation of the name Strobilanthes sarcorrhiza (Acanthaceae). Novon: J Bot Nomenclature. 2007;17(2):154–5.

17. Nunnari J, Suomalainen A. Mitochondria: in sickness and in health. Cell. 2012;148(6):1145–59.

18. Lu G, Wang W, Mao J, Li Q, Que Y. Complete mitogenome assembly of Selenicereus monacanthus revealed its molecular features, genome evolution, and phylogenetic implications. BMC Plant Biol. 2023;23(1):541.

19. Møller IM, Rasmusson AG, Van Aken O. Plant mitochondria– past, present and future. Plant J. 2021;108(4):912–59.

20. Bi C, Sun N, Han F, Xu K, Yang Y, Ferguson DK. The first mitogenome of Lauraceae (*Cinnamomum chekiangense*). Plant Divers. 2024;46(1):144.

21. Xiong Y, Yu Q, Xiong Y, Zhao J, Lei X, Liu L, Liu W, Peng Y, Zhang J, Li D. The complete mitogenome of *Elymus sibiricus* and insights into its evolutionary pattern based on simple repeat sequences of seed plant mitogenomes. Front Plant Sci. 2022;12:802321.

22. Wang R, Luo Y, Lan Z, Qiu D. Insights into structure, codon usage, repeats, and RNA editing of the complete mitochondrial genome of Perilla frutescens (Lamiaceae). Sci Rep. 2024;14(1):13940.

23. Li R, Ren X, Bi Y, Ding Q, Ho VWS, Zhao Z. Comparative mitochondrial genomics reveals a possible role of a recent duplication of NADH dehydrogenase subunit 5 in gene regulation. DNA Res. 2018;25(6):577–86.

24. Li S, Wang Z, Jing Y, Duan W, Yang X. Graph-based mitochondrial genomes of three foundation species in the Saccharum genus. Plant Cell Rep. 2024;43(8):191.

25. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014;100:107–20.

26. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all Nad genes. Proc Natl Acad Sci. 2015;112(27):E3515–24.

27. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, Konstantinov YM, Shmakov VN, Belkov VI, Sadovsky MG. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics. 2020;21:1–12.

28. Jin J-J, Yu W-B, Yang J-B, Song Y, CW d, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. Genome Biol. 2020;21(21):241.

29. He W, Xiang K, Chen C, Wang J, Wu Z. Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. Brief Bioinform. 2023;24(1):bbac522.

30. Bi C, Shen F, Han F, Qu Y, Hou J, Xu K, Xu L-a, He W, Wu Z, Yin T. PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. Hortic Res. 2024;11(3):uhae023.

31. Natarajan P, Murugesan AK, Govindan G, Gopalakrishnan A, Kumar R, Duraisamy P, Balaji R, Tanuja, Shyamli PS, Parida AK. A reference-grade genome identifies salt-tolerance genes from the salt-secreting Mangrove species Avicennia marina. Commun Biology. 2021;4(1):851.

32. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bull. 1987;19:11–5.

33. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.

34. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33(14):2202–4.

35. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432.

36. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de Novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.

37. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.

38. Li J, Ni Y, Lu Q, Chen H, Liu C. PMGA: A plant mitochondrial genome annotator. *Plant Communications*.

39. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.

40. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated megablast search tool. Nucleic Acids Res. 2015;43(16):7762–8.

41. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.

42. Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, Liu C. Plant mitochondrial genome map (PMGmap): A software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. Mol Ecol Resour. 2024;24:e13952.

43. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.

44. 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 Res. 2001;29(22):4633–42.

45. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.

46. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y. TBtools-II: A one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42.

47. 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.

48. 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.

49. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14(1):e1005944.

50. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics. 2023;39(3):btad121.

51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

52. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

53. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–4.

54. Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51(W1):W587–92.

55. Zhang H, Yan M, Li L, Jiang Z, Xiong Y, Wang Y, Akogwu CO, Tolulope OM, Zhou H, Sun Y. Assembly and comparative analysis of the complete mitochondrial genome of red raspberry (Rubus Idaeus L.) revealing repeat-mediated recombination and gene transfer. BMC Plant Biol. 2025;25(1):85.

56. Jiang M, Ni Y, Zhang J, Li J, Liu C. Complete mitochondrial genome of *Mentha spicata* L. reveals multiple chromosomal configurations and RNA editing events. Int J Biol Macromol. 2023;251:126257.

57. Wang H, Wu Z, Li T, Zhao J. Highly active repeat-mediated recombination in the mitogenome of the aquatic grass Hygroryza aristata. BMC Plant Biol. 2024;24(1):644.

58. Han F, Bi C, Zhao Y, Gao M, Wang Y, Chen Y. Unraveling the complex evolutionary features of the *Cinnamomum camphora* mitochondrial genome. Plant Cell Rep. 2024;43(7):183.

59. Bi C, Qu Y, Hou J, Wu K, Ye N, Yin T. Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. Front Plant Sci. 2022;13:914635.

60. Qu K, Chen Y, Liu D, Guo H, Xu T, Jing Q, Ge L, Shu X, Xin X, Xie X. Comprehensive analysis of the complete mitochondrial genome of *Lilium Tsingtauense* reveals a novel multichromosome structure. Plant Cell Rep. 2024;43(6):150.

61. Donnelly K, Cottrell J, Ennos RA, Vendramin GG, A'Hara S, King S, Perry A, Wachowiak W, Cavers S. Reconstructing the plant mitochondrial genome for marker discovery: a case study using Pinus. Mol Ecol Resour. 2017;17(5):943–54.

62. Du M-Z, Liu S, Zeng Z, Alemayehu LA, Wei W, Guo F-B. Amino acid compositions contribute to the proteins' evolution under the influence of their abundances and genomic GC content. Sci Rep. 2018;8(1):7382.

63. Wang L, Liu X, Xu Y, Zhang Z, Wei Y, Hu Y, Zheng C, Qu X. Assembly and comparative analysis of the first complete mitochondrial genome of a traditional Chinese medicine *Angelica biserrata* (Shan et Yuan) Yuan et Shan. Int J Biol Macromol. 2024;257:128571.

64. Wang D, Liu F, Wang L, Huang S, Yu J. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. Biol Direct. 2011;6:1–17.

65. Li Y, Hu X, Xiao M, Huang J, Lou Y, Hu F, Fu X, Li Y, He H, Cheng J. An analysis of codon utilization patterns in the Chloroplast genomes of three species of Coffea. BMC Genomic Data. 2023;24(1):42.

66. Huang Y, Ma Q, Sun J, Zhou LN, Lai CJ, Li P, Jin XJ, Zhang YH. Comparative analysis of Diospyros (Ebenaceae) plastomes: insights into genomic features, mutational hotspots, and adaptive evolution. Ecol Evol. 2023;13(7):e10301.

67. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65.

68. Bock H, Brennicke A, Schuster W. Rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GTG as a potential translation initiation codon in plant mitochondria? Plant Mol Biol. 1994;24:811–8.

69. Wang Y, Cui G, He K, Xu K, Liu W, Wang Y, Wang Z, Liu S, Bi C. Assembly and comparative analysis of the complete mitochondrial genome of *Ilex rotunda* thunb. Forests. 2024;15(7):1117.

70. Sun N, Han F, Wang S, Shen F, Liu W, Fan W, Bi C. Comprehensive analysis of the Lycopodium Japonicum mitogenome reveals abundant tRNA genes and cis-spliced introns in Lycopodiaceae species. Front Plant Sci. 2024;15:1446015.

71. Shi Y, Chen Z, Jiang J, Wu W, Xin Y, Zeng W. Assembly and comparative analysis of the complete mitogenome of Rubus Chingii Var. Suavissimus, an exceptional berry plant possessing sweet leaves. Front Plant Sci. 2024;15:1504687.

72. Feng L, Wang Z, Wang C, Yang X, An M, Yin Y. Multichromosomal mitochondrial genome of Punica granatum: comparative evolutionary analysis and gene transformation from Chloroplast genomes. BMC Plant Biol. 2023;23(1):512.

73. Quax TE, Claassens NJ, Soll D, van der Oost J. Codon Bias as a means to Fine-Tune gene expression. Mol Cell. 2015;59(2):149–61.

74. Bazzini AA, Del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, Giraldez AJ. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. EMBO J. 2016;35(19):2087–103.

75. Baeza M, Alcaino J, Barahona S, Sepulveda D, Cifuentes V. Codon usage and codon context bias in Xanthophyllomyces dendrorhous. BMC Genomics. 2015;16(1):293.

76. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68(1):225–52.

77. Wu ZQ, Liao XZ, Zhang XN, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. J Syst Evol. 2022;60(1):160–8.

78. Lukeš J, Kaur B, Speijer D. RNA editing in mitochondria and plastids: Weird and widespread. Trends Genet. 2021;37(2):99–102.

79. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8(1):5–14.

80. 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.

81. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 2002;268(4):434–45.

82. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, Du Y, Ren J, Li S, Chen Z, et al. Assembly and comparative analysis of the first complete mitochondrial genome of Acer truncatum Bunge: a Woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022;22(1):29.

83. Xu C, Bi W, Ma RY, Li PR, Liu F, Liu ZW. Assembly and comparative analysis of the complete mitochondrial of Spodiopogon sagittifolius, an endemic and protective species from Yunnan, China. BMC Plant Biol. 2025;25(1):373.

84. Zhou P, Zhang Q, Li F, Huang J, Zhang M. Assembly and comparative analysis of the complete mitochondrial genome of Ilex metabaptista (Aquifoliaceae), a Chinese endemic species with a narrow distribution. BMC Plant Biol. 2023;23(1):393.

85. Gao C, Wang S, Huang Y, Deng Y. Assembly and comparative analysis of the complete mitochondrial genome of Echinacanthus longipes (Acanthaceae), endemic to the Sino-Vietnamese karst flora. BMC Genomics. 2025;26(1):251.

86. Wang X, Wang J, Jin D, Guo H, Lee T-H, Liu T, Paterson AH. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. Mol Plant. 2015;8(6):885–98.

87. Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA editing in plants and its evolution. Annu Rev Genet. 2013;47(1):335–52.

88. Sun M, Zhang M, Chen X, Liu Y, Liu B, Li J, Wang R, Zhao K, Wu J. Rearrangement and domestication as drivers of Rosaceae mitogenome plasticity. BMC Biol. 2022;20(1):181.

89. Li Y-X, Li Z-H, Schuiteman A, Chase MW, Li J-W, Huang W-C, Hidayat A, Wu S-S, Jin X-H. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. Mol Phylogenet Evol. 2019;139:106540.

90. Yang Z, Ferguson DK, Yang Y. New insights into the plastome evolution of Lauraceae using herbariomics. BMC Plant Biol. 2023;23(1):387.

91. Liu C, Chen H-H, Tang L-Z, Khine PK, Han L-H, Song Y, Tan Y-H. Plastid genome evolution of a monophyletic group in the subtribe lauriineae (Laureae, Lauraceae). Plant Divers. 2022;44(4):377–88.

92. Dong W, Xu C, Cheng T, Lin K, Zhou S. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol. 2013;5(5):989–97.

93. QIU YL, Li L, Wang B, XUE JY, Hendry TA, LI RQ, Brown JW, Liu Y, Hudson GT. CHEN ZD: Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 2010, 48(6):391–425.

94. Pearl SA, Welch ME, McCauley DE. Mitochondrial heteroplasmy and paternal leakage in natural populations of Silene vulgaris, a gynodioecious plant. Mol Biol Evol. 2009;26(3):537–45.

95. McLay TG, Fowler RM, Fahey PS, Murphy DJ, Udovicic F, Cantrill DJ, Bayly MJ. Phylogenomics reveals extreme gene tree discordance in a lineage of dominant trees: hybridization, introgression, and incomplete lineage sorting blur deep evolutionary relationships despite clear species groupings in Eucalyptus subgenus Eudesmia. Mol Phylogenet Evol. 2023;187:107869.

Publisher's note