BMC Genomics

RESEARCH Open Access

# De novo assembly of the mitochondrial genome of *Glycyrrhiza glabra* and identification of two types of homologous recombination configurations caused by repeat sequences

Guowang Zhou[1,2], Meiling Qin[3], Xiuli Liu[2,4], Yonghui Qi[2,5], Xiaobin Ou[2,4,6*] and Min Tang[3*]

## Abstract

**Background** *Glycyrrhiza glabra*, which is widely used in medicine and therapy, is known as the 'king of traditional Chinese medicine'. In this study, we successfully assembled and annotated the mitochondrial and chloroplast genomes of *G. glabra* via high-throughput sequencing technology, combining the advantages of short-read (Illumina) and long-read (Oxford Nanopore) sequencing.

**Results** We revealed the ring structure of the mitochondrial genome, which spans 421,293 bp with 45.1% GC content and 56 annotated genes. Notably, we identified 514 repetitive sequences, including 123 Simple sequence repeats (SRs), 3 Tndem sequence repeats (TSRs), and 388 Dispersed sequence repeats (DSRs). We identified 79 out of the 388 DSRs as potentially involved in homologous recombination. We identified five forward repeats and four palindromic repeats that facilitate homologous recombination and induce alterations in the mitochondrial genome structure. We corroborated this finding via polymerase chain reaction (PCR). Furthermore, we identified chloroplast-derived sequence fragments within the mitochondrial genome, offering novel insights into the evolutionary history of plant mitochondrial genomes. We predicted 460 potential RNA editing sites, primarily involving cytosine-to-uracil transitions. This study reveals the complexity of repetitive sequence-mediated homologous recombination in the mitochondrial genome of *G. glabra* and provides new insights into its structure, function, and evolution.

**Conclusions** These findings have important implications for conservation biology, population genetics, and evolutionary studies, underscoring the role of repetitive sequences in genome dynamics and highlighting the need for further research on mitochondrial genome evolution and function in plants.

**Keywords** *Glycyrrhiza glabra*, Mitochondrial genome, Repeat sequences, Homologous recombination, Plant evolution

*Correspondence:
Xiaobin Ou
xbou@zju.edu.cn
Min Tang
mt3138@ujs.edu.cn

Full list of author information is available at the end of the article

## Background

*Glycyrrhiza glabra*, a perennial plant belonging to the genus *Glycyrrhiza* of the family Leguminosae, has significant medicinal importance; it has been cultivated in China for more than two millennia and has been revered as the "king of traditional Chinese medicines" [1]. The main production area of *G. glabra* is northwestern China [2]. The roots and rhizomes of *G. glabra* contain a plethora of biologically active compounds, including glycyrrhizic acid [3], glycyrrhizin [4], and licochalcone A [5], which have diverse pharmacological effects, including anti-inflammatory, antitumour, antiviral, and immunomodulatory effects [6].

Mitochondria are essential organelles in eukaryotic cells, play pivotal roles in cellular energy metabolism, whereas chloroplasts are primarily found in plant cells and are essential for photosynthesis. Both organelles have their own DNA, exhibiting the capacity for autonomous replication and a gene expression system, which suggests a certain degree of semi-autonomy [7]. This feature enables mitochondria to transmit and express specific genetic information independently, without relying solely on the nucleus. It has been demonstrated that mitochondria evolved from the endosymbiosis of *Alphaproteobacteria*, whereas chloroplasts originated from cyanobacteria, which were engulfed by ancestral eukaryotic cells during endosymbiosis [8].

By November 2024, the NCBI database had recorded 13,346 chloroplast genomes, while only 688 plant mitochondrial genomes had been cataloged, highlighting the significant disparity in the number of available sequences. This disparity underscores the challenges faced in mitochondrial genome research. The primary divergence between chloroplast genomes and the mitochondrial genome is evident in their structural composition: the chloroplast genome generally maintains a relatively uniform and circular genome structure [9, 10], whereas the mitochondrial genome is characterized by a high incidence of repetitive sequences and exhibits considerable structural heterogeneity [11], including polycyclic [12–14], linear [15], and branched forms [16]. The intricate nature of plant mitochondrial genomes, with their abundance of repetitive elements and diverse configurations [17], introduces substantial difficulties in assembling plant mitochondrial genomes. In plant mitochondria, homologous recombination mediated by repetitive sequence fragments represents a critical mechanism for genomic rearrangement, resulting in structural diversity and alterations in the relative positions of genes, thereby affecting gene expression regulation [18]. These recombination events may result in the formation of new gene combinations or influence gene expression levels through position effects. Homologous recombination plays a significant role in plant evolution [19], providing flexibility in genomic structure and enabling plants to adapt to environmental changes and evolutionary pressures [20]. The mitochondrial genomes of plants are considerably larger than those of other eukaryotes and have undergone rapid evolution in terms of their structural composition.

The mitochondrial genomes of the plant is considerably larger than that of other eukaryotes and has undergone a rapid evolution in terms of its structural composition [21]. This phenomenon can be attributed to the recombination activities that give rise to alternative mitochondrial genome configurations, which serve as an important reservoir of genetic diversity and facilitate the rapid evolution of mitochondrial genomes. Conversely, the high prevalence of ectopic recombination results in mitochondrial genome instability and the expression of gene chimeras, which may have adverse consequences [22]. In contrast to the structural plasticity of the genome, the mitochondrial genome coding sequences of most plant species evolve at a markedly slower pace, despite the considerable variability in genome [23].

The gradual maturation of sequencing technology in recent years has led to the reporting of a number of plant organelle genomes. However, studies of the organelle genome of *G. glabra* have primarily focused on the chloroplast, with minimal attention directed towards the mitochondrial genome. Some scholars have developed DNA barcodes for molecular identification [24] and chloroplast microsatellite markers [25] on the based on *G. glabra* chloroplast DNA sequences.

In this study, we completed a high-quality mitochondrial genome assembly of *G. glabra* with in-depth analyses. These results advance our understanding of *G. glabra* mitochondrial genome structure and function and provide critical insights for conservation biology, population genetics and evolution research on this species.

## Materials and methods

### Materials and sequencing

The plant *G. glabra* was sourced from the Bio-Agroforestry Park of Longdong University, QingYang, Gansu, China (35.7289°N, 107.6842°E). Fresh leaves were collected and sterilised in 75% alcohol, then rapidly frozen in liquid nitrogen and stored at -80 ℃.

*G. glabra* leaf material was sent to Wuhan Bena Technology Co (Wuhan, China) for sequencing. Total genomic DNA was extracted using a plant genomic DNA kit (Tiangen Biotech, Beijing, China). The short-read sequencing platform used was an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) and the paired end sequencing (PE) read length was 150 bp. Fastp (v0.21.0) was used to filter the original data and obtain high-quality reads [26]. The long-read sequencing platform used was Nanopore GridION (Nanopore, Oxford, UK) and the

sequencing data was subsequently filtered via NanoFilt (v2.8.0) [27].

#### Mitochondrial and chloroplast genomes assemble and annotation

Initially, Minimap2 (v2.28) was used to align all long reads to the reference genome, ensuring accurate read placement [28]. These aligned long reads were then assembled into a preliminary genome via Canu (v2.2) [29]. To refine the assembly, Bowtie2 (v2.5.4) was employed to align short reads to the preliminary assembly generated by Canu [30], which improved both the accuracy and completeness of the assembly. Finally, Unicycler was used to perform a hybrid assembly [31], effectively integrating the information from both long and short reads, resulting in a high-quality assembly of the *G. glabra* mitochondrial and chloroplast genomes. Annotation was performed via Geseq (https://chlorobox.mpimp-golm.mpg.de/geseq.html). The mitochondrial genomes of *Glycyrrhiza uralensis* (NC_053919.1), *Medicago sativa* (NC_029641.1), *Lotus japonicus* (NC_016743.2), *Trillium grandiflorum* (NC_048501.1), *Trifolium pratense* (NC_048499.1), and *Lotus corniculatus* (NC_048499.1) were used as reference genomes to annotate mitochondrial genomes.

#### Codon usage analysis

Evolutionary analysis was conducted via PhyloSuite (v1.2.3) [32]. This involved both phylogenetic analysis and the calculation of relative synonymous codon usage (RSCU) to assess the evolutionary relationships and codon usage patterns within the mitochondrial genome.

#### Repeat-mediated homologous recombination prediction and pcr amplification validation

Microsatellite repeats in the genome were identified using MISA (v2.1) with the following parameters: motif lengths and minimum repeat counts (unit_size, min_repeats) of 1–10, 2–5, 3–4, 4–3, 5–3, and 6–3, and the maximum allowed base number between two SSRs (max_difference_for_2_SSRs) set to 100 [33]. Transposable repeat sequences (TRSs) were identified using REpter (v4.10) with the parameters '2 7 7 80 10 50 500 -f -d -m' [34]. Diverse repeat sequences (DRSs) were detected using REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) with a Hamming distance of 3, a maximum of 500 computed repeats, and a minimum repeat size of 30 [35]. Subsequently, Circos (v0.69-9) was employed to create circular visualizations [36], illustrating the distribution and types of repetitive sequences.

In order to ascertain the potential of repetitive sequences in the mitotic genome to act as mediators of homologous recombination, we proceeded to extract pairs of repetitive sequences and their 100 bp flanking sequences, subsequently constructing two conformations. Thereafter, nanopore reads were aligned to these two conformations, thus enabling us to determine the potential impact of the repeat sequences on recombination.

In consideration of the frequency of repetitive sequences and their length, selected for further analysis were F-repeats 2, 3, 4, 5, and P-repeats 4, 6, 7 (200–400 bp), along with F-repeat 51 (shorter but more frequent) and P-repeat 2 (longer but more frequent). Primers were designed on the flanking 100 bp of these sequences using the Primer6 software (Table S1), and the primers were validated through polymerase chain reaction (PCR) cross-validation.

#### Mitochondrial plastid DNA (MTPT) identification

Using BLAST (v2.16.0) [37], a homology comparison was conducted between the mitochondrial and chloroplast genomes, retaining homologous sequences longer than 500 bp as conserved syntenic blocks to generate a synteny map. The MCScanX program was subsequently used to analyze homologous gene relationships between the mitochondrial and chloroplast genomes [38].

#### RNA editing site prediction

The RNA editing sites were predicted via the PREPACT3 web tool [39]. The analysis employed reference sequences from *Lotus japonicus* and *Millettia pinnata*. The predicted RNA editing sites were subsequently visualized using R (4.4.0) to facilitate interpretation and analysis of the results.

#### Phylogenetic and collinearity analysis

A total of 24 complete mitochondrial genomes, representing three lower subfamilies of the Leguminosae family (Papilionoideae, Caesalpinioideae, and Detarioideae) were retrieved from the NCBI database (Table S2). Using PhyloSuite version (v1.2.3) [32], common genes among these genomes were identified and subsequently subjected to multiple sequence alignment via MAFFT (v7.505) [40]. The aligned sequences were then analyzed for phylogenetic relationships with IQ-TREE (v1.6.12) [41]. To assess the statistical support for the tree topology, 1000 bootstrap replicates were performed, and the resulting bootstrap values were used to estimate the reliability of each branch. Finally, the phylogenetic outcomes were graphically represented via iTOL (v6) [42], providing a visual synopsis of the evolutionary connections among the species studied.

On the basis of the BLAST (v2.16.0) procedure [37], the individual mitochondrial genomes were compared and any homologous sequences exceeding 500 bp were retained as conserved collinear blocks, thus enabling the generation of multiple synteny plots. Multiple synteny

plots were created via the MCScanX source program [38].

## Results

### Mitochondrial genome assembly and annotation

Characterization and annotation of the *G. glabra* mitochondrial genome revealed that the mitochondrial genome of *G. glabra* has a classic cyclic structure (Fig. 1). It is a circular structure with a total length of 421,293 bp and a GC content of 45.1%. A total of 56 genes were annotated in the genome, including 34 protein-coding genes (PCGs), 17 tRNA genes, and 5 rRNA genes. The predicted genes of the G. glabra mitochondrial genome are shown in Table S3. The 34 protein-coding genes are divided into ten categories: ATP synthases (5), cytochrome c biogenesis (4), ubiquinol cytochrome c reductase (1), cytochrome c oxidase (3), cytochrome b (1), maturases (1), transport membrane protein (1), NADH dehydrogenases (9), ribosomal proteins (LSU) (2), ribosomal proteins (SSU) (6), and Twin-arginine translocation (1).

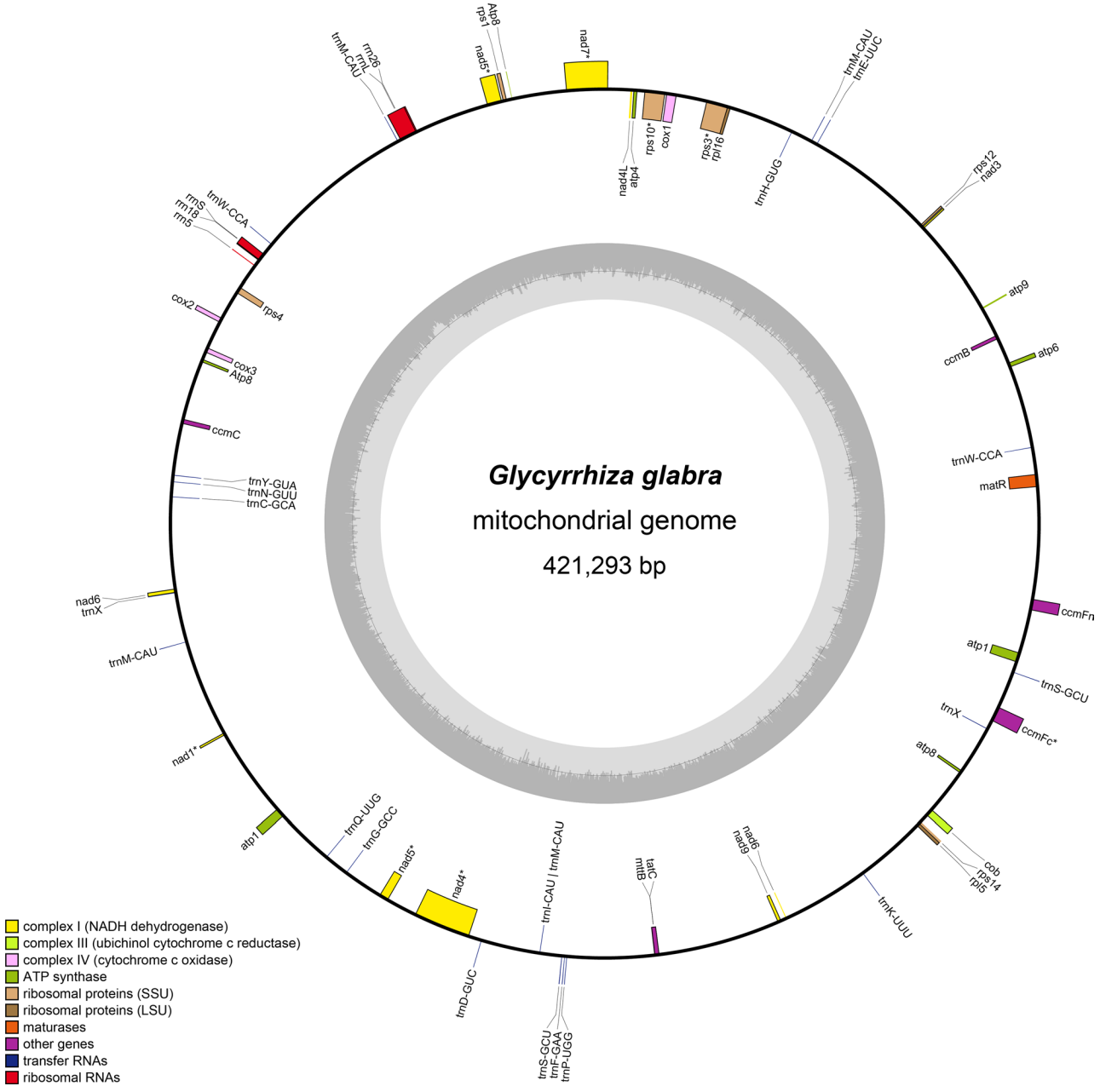


**Fig. 1** Mitochondrial annotation map of *Glycyrrhiza glabra*. The dark gray region represents GC content, while the light gray region represents AT content

### Gene copy number

The genes that are responsible for coordinating electron transfer, oxidative phosphorylation and energy metabolism have remained largely unchanged in the majority of legumes. Furthermore, the mitochondrial gene copy numbers of Fabaceae were found to be consistent with those of the vast majority of other legumes. However, the most notable divergence was observed in the *rps3* gene of *Ammopiptanthus mongolicus*, a plant belonging to the subfamily Papilionoideae (Fig. 2). The high diversity of mitochondrial DNA genes observed in higher plants is attributed primarily to the amplification of ribosomal protein genes, which is consistent with these findings.

### Codon usage of PCGs

Relative synonymous codon usage (RSCU) denotes the ratio of specific synonymous codon utilization to the total usage. The relative RSCU in the *G. glabra* mt genome is shown in Fig. 3. The RSCU value of a total of 31 codons exceeds one, thereby indicating that they are used with greater frequency than other synonymous codons. Of the aforementioned codons, 28 were terminated by A or T, representing a proportion of 90.32%. This suggests that frequently used codons tend to end with the A/T base. It is noteworthy that among the amino acids, methionine (Met) and tryptophan (Trp) are distinctive in that they lack synonymous codons. Furthermore, serine (Ser) is of interest due to the observation that one of its codons, UCC, has an RSCU of 1.00. This RSCU value indicates that the usage of UCC aligns with the expected frequency of synonymous codons, without displaying any significant adaptive or selective traits. This study employs an in-depth analysis of codon usage patterns and preferences in the mitochondrial genome of *G. glabra* with the objective of elucidating the gene coding and expression mechanisms of this species (Fig. 3).

### Repeat sequence analysis

We identified 514 repeat sequences in the *G. glabra* mitochondrial genome, including 123 simple sequence repeats (SSRs), 3 tandem sequence repeats (TSRs), and 388 dispersed sequence repeats (DSRs) (Fig. 4). Among the SSRs, monomeric, dimeric, trimeric, and tetrameric SSRs were the most prevalent, accounting for 22.8% (28),

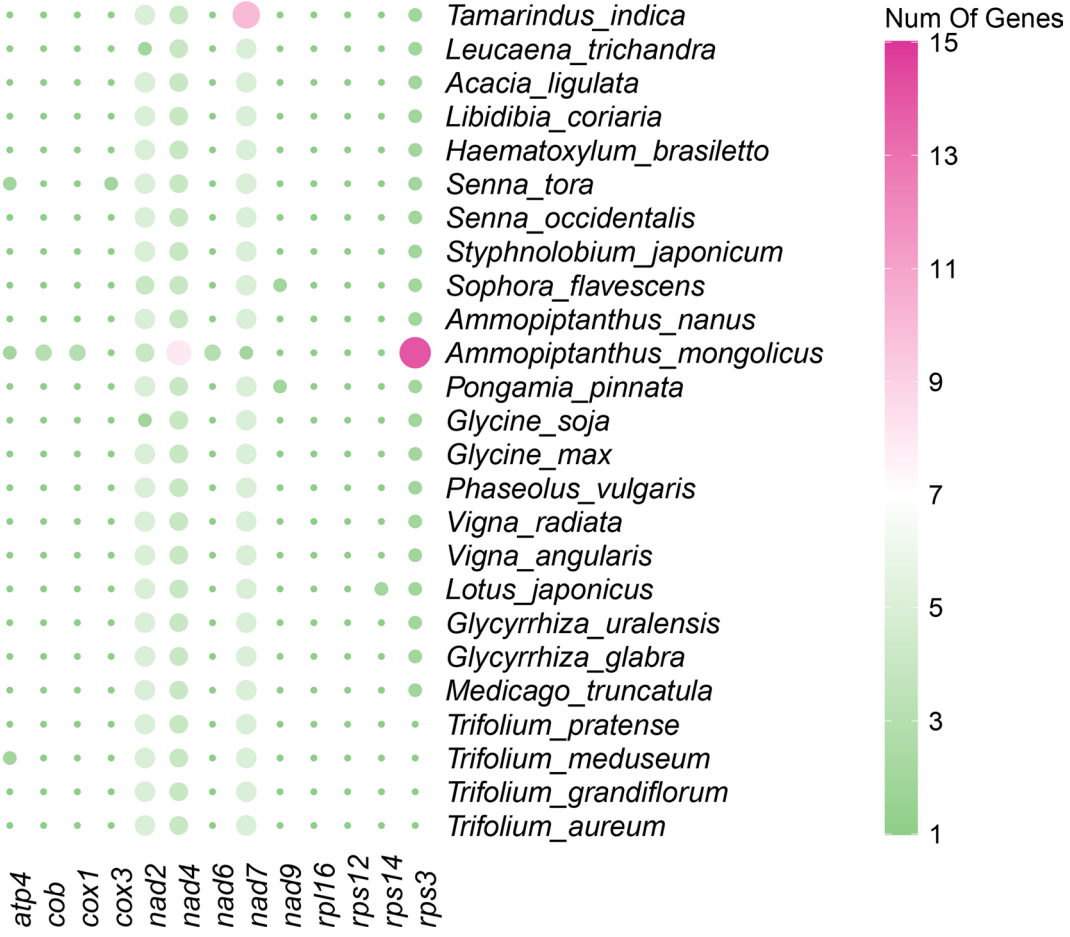


**Fig. 2** Distribution of protein-coding genes in plant mitochondrial genomes. The dimensions and chromaticity of the circle serve to indicate the quantity of copy number

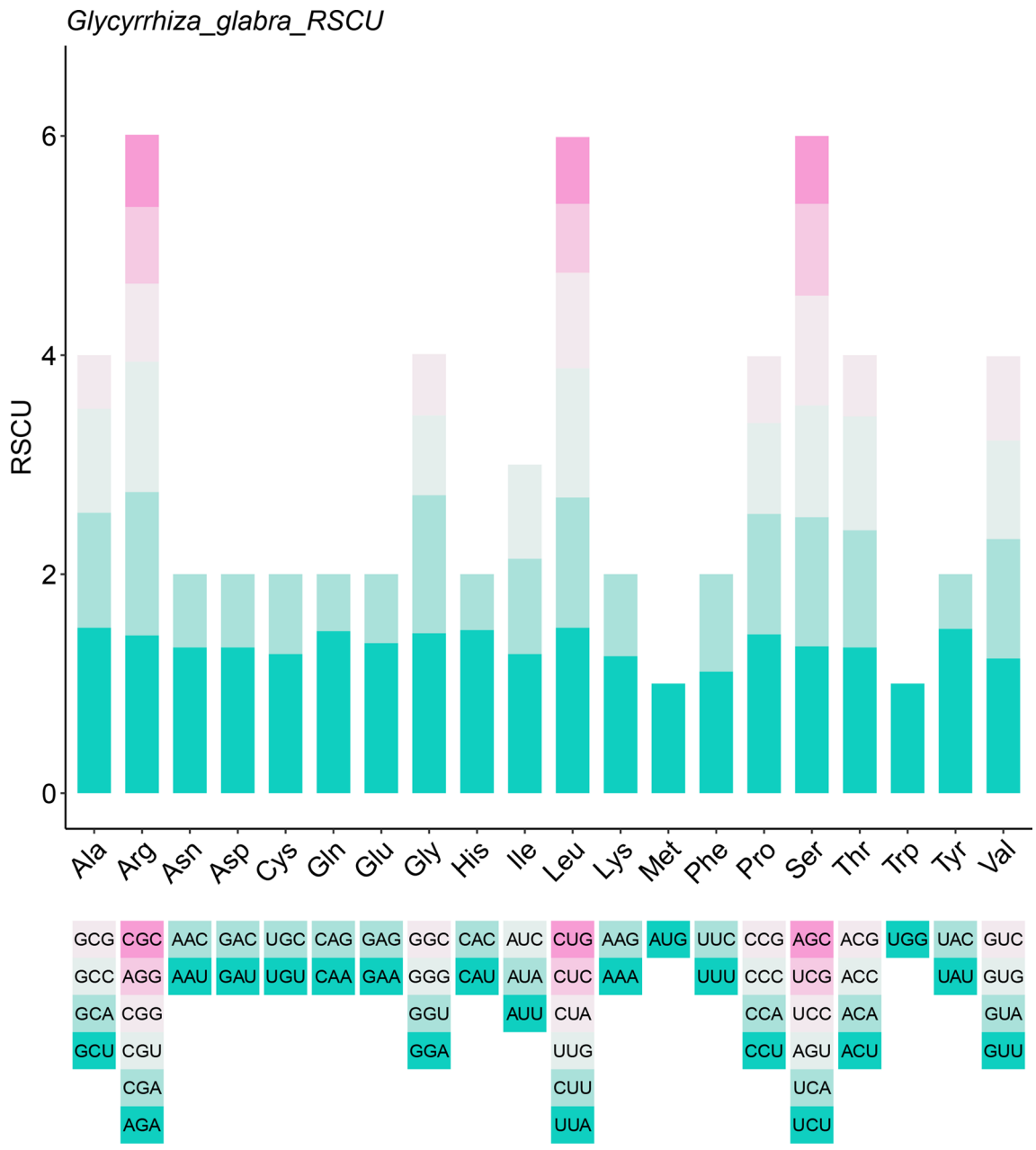


**Fig. 3** Relative synonymous codon usage in *Glycyrrhiza glabra* mitogenome. Note: The following blocks represent all codons encoding each amino acid, and the height of the upper column represents the sum of RSCU values of all codons

26.0% (32), 10.6% (13), and 35.0% (43) of the total SSRs, respectively. In contrast, pentamers and hexamers were less common, representing 3.3% (4) and 2.4% (3) of the total SSRs, respectively. DSRs include forward repeats (186), palindromic repeats (171) and reverse repeats (31). In the present study, we identified 388 repeat sequences that ranged from 30 to 4907 bp and accounted for 7.84% of the total *G. glabra* mt genome.

### Repeat-mediated homologous recombination prediction and PCR amplification validation

An analysis of the repeat sequences revealed that forward and palindromic repeat sequences constituted the largest proportion of DSRs, at 92.01%. Prior research has indicated that repeat sequences can facilitate homologous recombination in mitochondrial genomes [18]. A total of 40 forward and 39 palindromic repeat sequences were identified in the *G. glabra* mitochondrial genome that may play a role in homologous recombination, as supported by long-read sequencing data (Table S4). Both

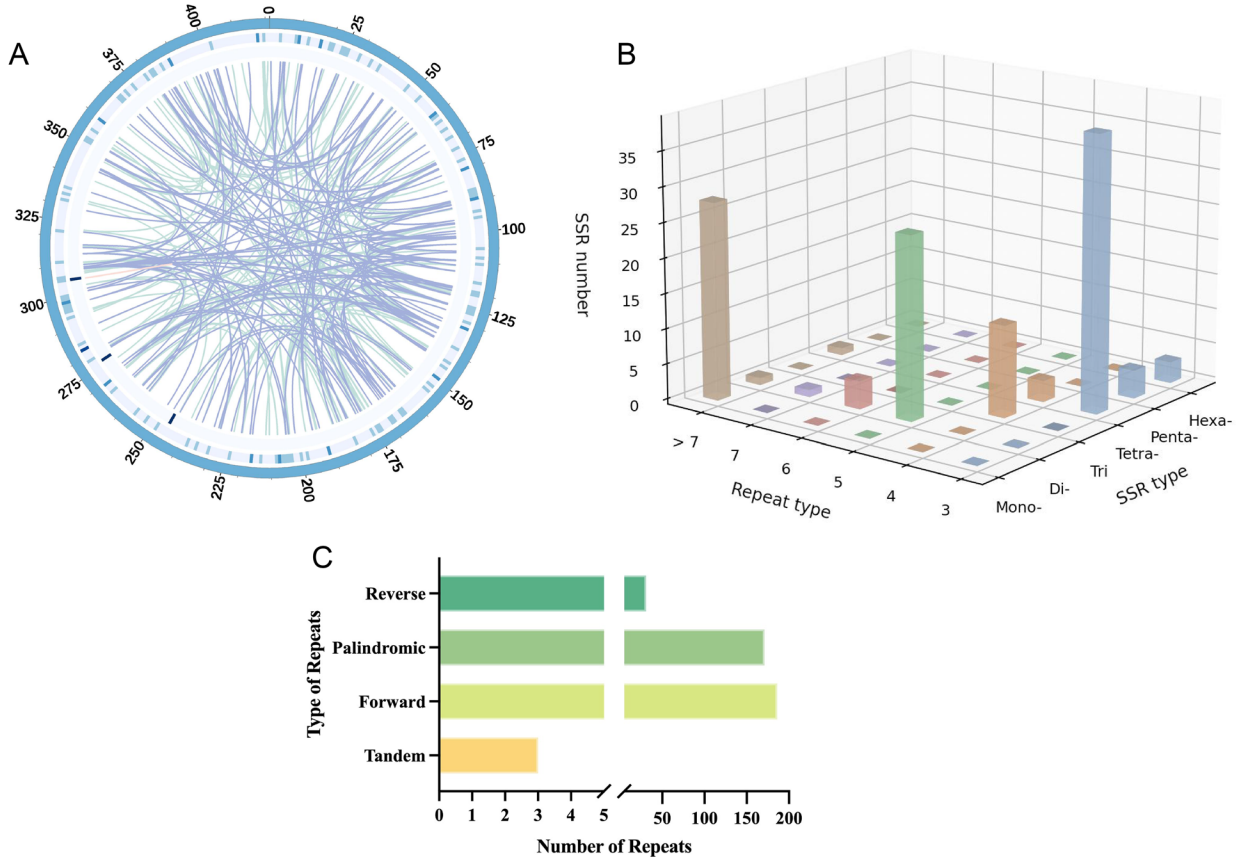


**Fig. 4** The Distribution map, type, frequency, and distribution of simple sequence repeats and long repeated sequences in the mitochondrial genome of *Glycyrrhiza glabra*.**A.** Distribution map of repeat sequences in the *Glycyrrhiza glabra* mitochondrial genome. **B.** Distribution of simple sequence repeats types and repeat numbers. **C.** Dispersed and tandem repeat sequence frequency. Note: The outermost circle represents simple repeat sequences(4 A), followed by tandem repeat sequences, and the innermost line represents scattered repeat sequences

forward [43] and palindromic [44] repeat sequences can mediate homologous recombination of the mitochondrial genome, but the processes and resulting conformations are quite different (Fig. 5B).

To investigate their potential role in mediating homologous recombination, we designed primers based on 100 bp bases at either end of the repeated sequence. Then, the primers were cross-used in PCR to assess their ability to facilitate recombination events. The experimental findings revealed that a subset of these repeat sequences indeed played a mediating role in homologous recombination within the mitochondrial genome of *G. glabra*. The majority of repeats involved in homologous recombination were 300–500 bp in length (F-repeat2, 3, 4, 5 and P-repeat4, 6, 7), with some of the shorter (F-repeat51) and longer (P-repeat2) repeats also facilitating homologous recombination (Fig. 5B and C and Figure S1).

### Diversity of mitochondrial plastid DNAs (MTPTs)

The chloroplast genome was assembled and annotated using a methodology analogous to that employed for the mitochondrial genome. The chloroplast genome has a total length of 127,887 bp (Fig. 6A). During the course of evolution, the sequences of the chloroplast genome became integrated into the mitochondrial genome. A total of 12 homologous DNA fragments were identified within the mitochondrial genome of *G. glabra* (Fig. 6B). The summary insert length was 3,860 bp, representing 0.92% of the mitochondrial genome. This integration encompasses four unique tRNA genes: *trnW-CCA*, *trnD-GUC*, *trnN-GUU*, and *trnM-CAU*, with *trnM-CAU* being present twice.

### RNA-editing site prediction

An investigation into RNA editing was conducted on 34 PCGs of the *G. glabra* mitochondrial genome via the PREPACT3 method. This analysis identified 460 putative RNA editing sites in 31 PCGs (Fig. 7A). Notably, the *nad4* gene presented the highest frequency of 33 potential RNA editing sites, exceeding the number observed in all other mitochondrial genes. This was closely followed by the *ccmB*, *nad7*, *mttB*, *nad1*, *nad2*, and *ccmC* genes, all of which presented more than 20 RNA editing sites. Each

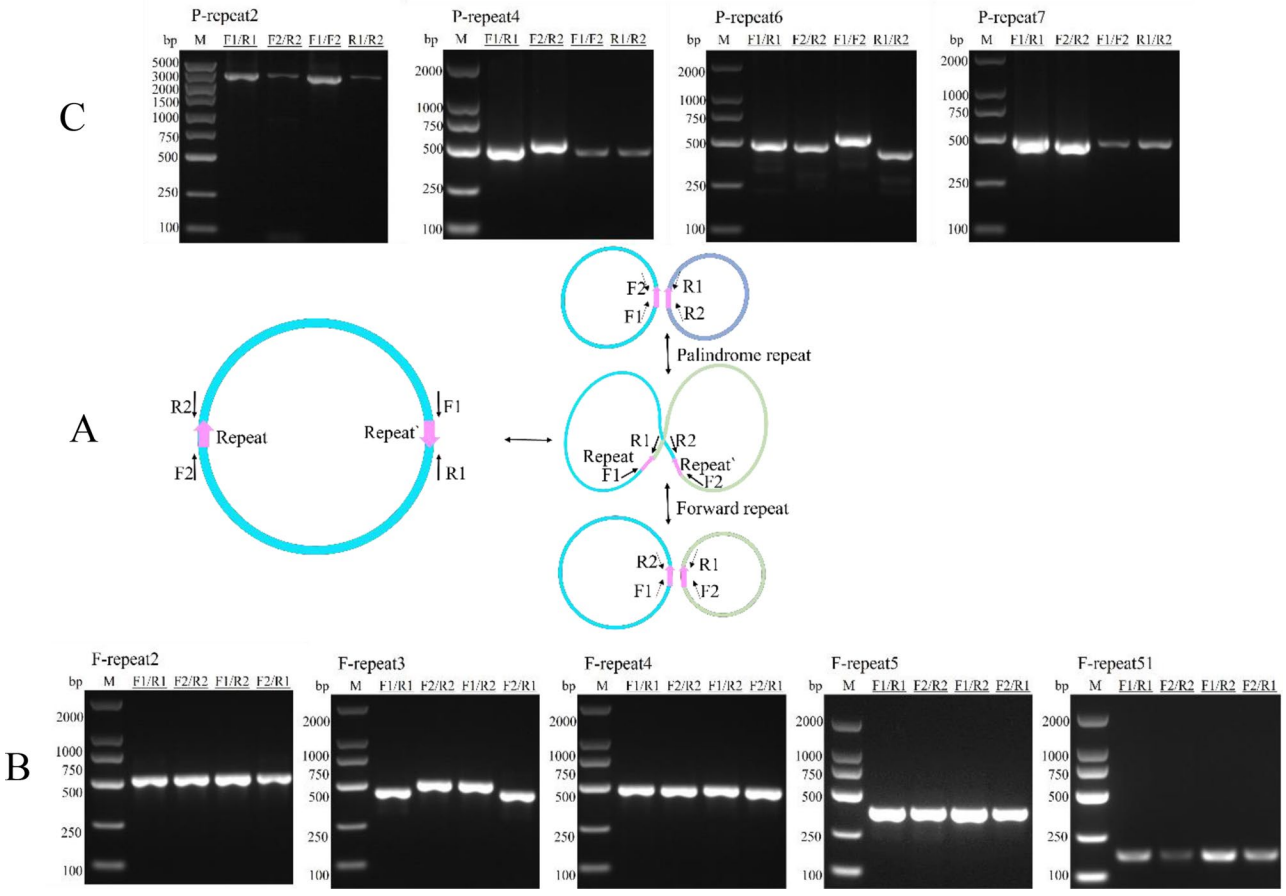


**Fig. 5** Homologous recombination and PCR validation in the mitochondrial genome of *Glycyrrhiza glabra*. **(A)** Schematic representation of homologous recombination in the *Glycyrrhiza glabra* mitochondrial genome, based on forward and palindromic repeat analysis. **(B)** PCR validation of homologous recombination mediated by forward repeats. **(C)** PCR validation of homologous recombination mediated by palindromic repeats

of these sites involves cytidine-to-uridine (C-to-U) base modifications. We predicted that one editing event would change ACG (encoding threonine) to AUG (encoding methionine) in the genes *nad1*, *nad4L*, *nad5*, *nad7*, and *rps10*. Additionally, *cox2* has two such events.

We also identified one editing event in which CAA (encoding glutamine) was changed to UAA (a stop codon) in the *atp6* gene, and another in which CGA (encoding arginine) was changed to UGA (another stop codon) in the *ccmFC* and *rps10* genes. RNA editing has been shown to induce changes in amino acid composition, with the potential to affect 460 unique sites. The most prevalent amino acid modifications observed are Ser → Leu and Por → Leu, occurring 108 and 106 times, respectively (Fig. 7B). Notably, Leu represents the most abundant amino acid in the RNA editing results. Concurrently, the RNA editing site resulted in alterations in the hydrophilicity and hydrophobicity of amino acids, with 48.26% of amino acids undergoing a transition from hydrophilic to hydrophobic and 7.61% from hydrophobic to hydrophilic (Table S5).

### Phylogenetic analysis

A phylogenetic analysis based on the mitochondrial genome was conducted to elucidate the phylogenetic position of *G. glabra* within the legume family. A phylogenetic tree was constructed using 13 conserved protein-coding genes (PCGs) from 25 different species (Fig. 8). In this analysis, mitochondrial genomes from plants in the Tamarix subfamily were used as outgroups. The resulting phylogenetic tree topology aligns with the current classification system proposed by the Angiosperm Phylogeny Group (APG). Notably, *G. glabra* and *G. uralensis* are grouped together on a single branch, with bootstrap support of 100%.

### Collinearity analysis

Covariance analysis revealed that the mitochondrial genomes of *G. glabra* and *G. uralensis* were widely conserved, with a total of 30 sequences with a total length of 433,157 bp, with identity of more than 95%, and 13 sequences with a length of more than 1,000 bp, of which the longest conserved region covered 54,344 nucleotides, with a sequence identity of 99.77%. In addition, shorter

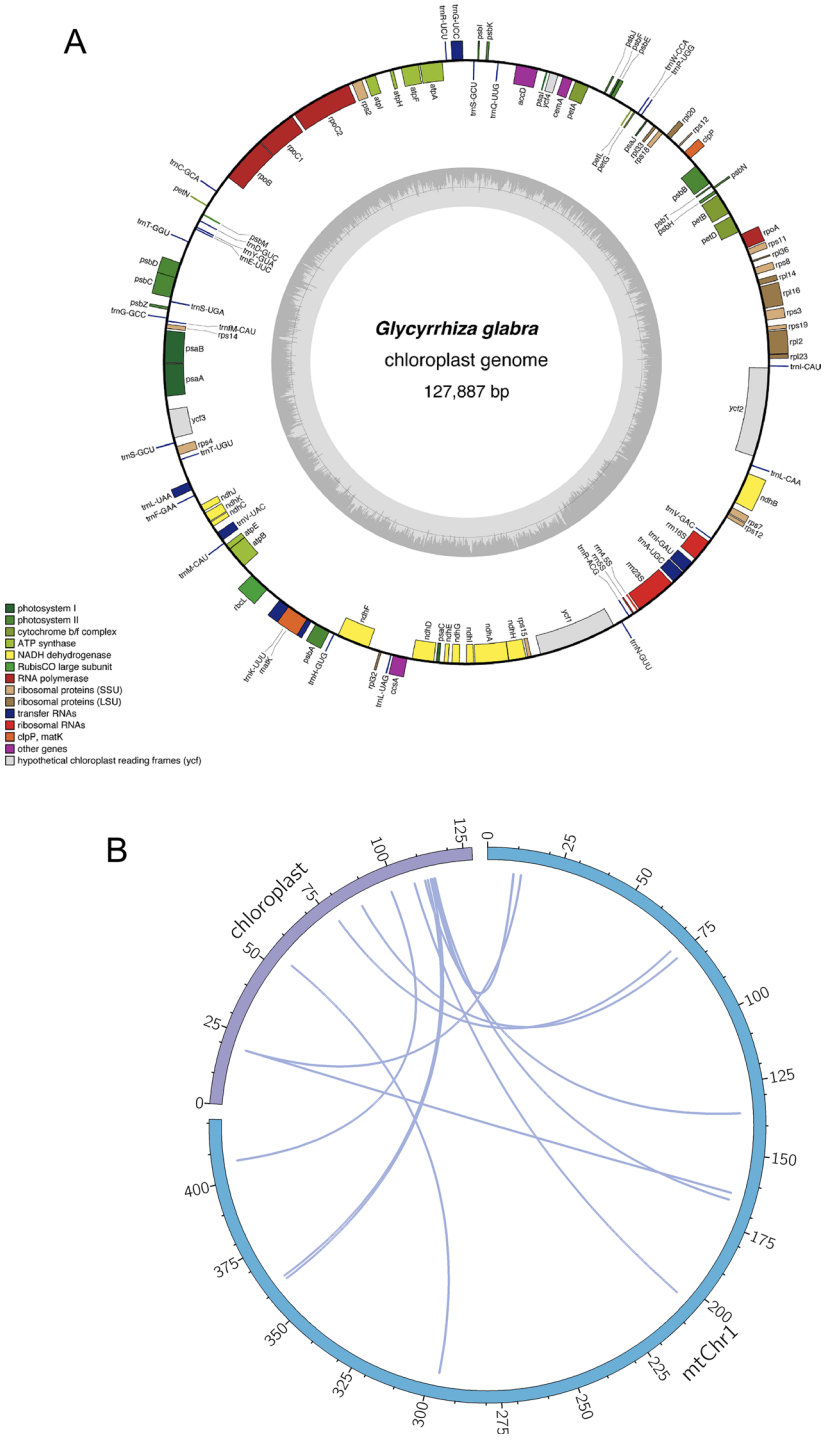


**Fig. 6** Chloroplast Genome annotation and its covariance analysis with Mitochondrial Genome. **(A)** Chloroplast annotation map of *Glycyrrhiza glabra*. **(B)** DNA transfer between the chloroplast and mitochondrial genome. Note: The blue circular segment represents the mitochondrial genome, and the purple circular segment represents the chloroplast genome

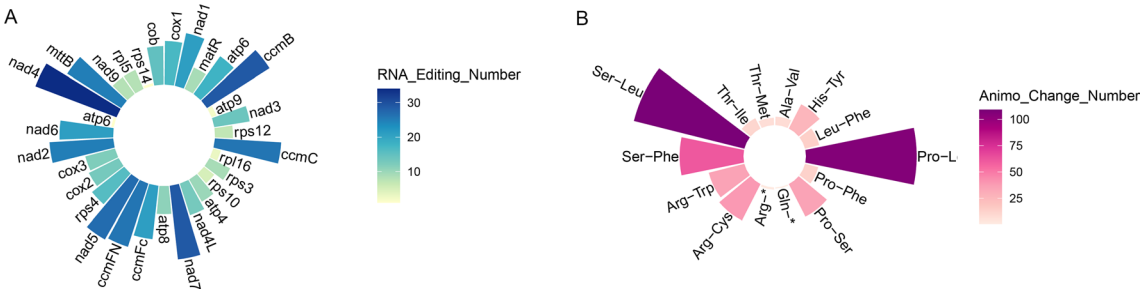


**Fig. 7** Characteristics of the RNA editing sites identified in mitochondrial PCGs of *G. glabra*. **(A)** The ordinate shows the number of RNA editing sites identified in PCGs, the abscissa shows the name of PCGs identified in the mitogenome of *G. glabra* Panel. **(B)** shows the potential effect of all RNA editing events on amino acids. Most of these RNA editing events lead to changes in amino acids

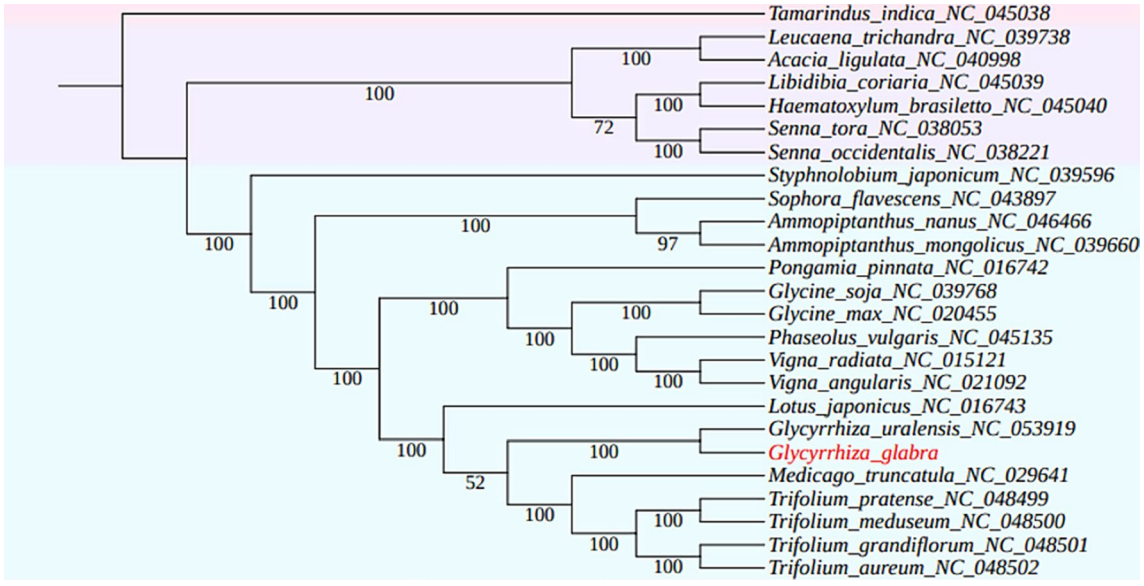


**Fig. 8** Maximum-likelihood (ML) tree based on 25 species of mitogenomes in the Leguminosae with Tamarindus indica of subfamily Detarioideae as outgroup. Note: The *G. glabra* is marked in red. Background colours light red, purple and blue for plants of the subfamilies Detarioideae, Caesalpinioideae and Papilionoideae, respectively. Bootstrap support values (%) are shown next to the nodes, with a value of 100 indicating full support

colinear regions were found throughout the genome, highlighting the overall conservation of the genome structure (Fig. 9).

## Discussion

Mitochondria, which generate the energy needed for essential life processes, are considered the powerhouses of eukaryotic cells [45]. The genomes of plant mitochondria are characterised by a notable degree of complexity, with abundant sequence-related variation. They encompass a multitude of repetitive sequences and exhibit a relatively high level of conservation in coding sequences [46]. The rapid development of genome sequencing technology has greatly accelerated the pace of mitochondrial genome research in medicinal plants [47–49]. This study presents the inaugural description of the fundamental characteristics of the *G. glabra* mitochondrial genome. Our findings furnish a crucial foundation for the comprehension of the mt genome's functional, hereditary, and evolutionary trajectory. The *G. glabra* mitochondrial genome is a circular sequence with a length of 421,291 bp and a GC content of 45.1%. We performed Geseq annotation of sequences, and identified 34 protein-coding genes, 17 tRNA genes and 5 rRNA genes in the mt genome. The codon preferences of the 34 protein-coding genes showed a clear A/T bias, as did the vast majority of plant mitochondria.

It has been demonstrated that plant mitochondria contain many repeat sequences [50]. Repeats play an important role in mitochondrial intergenic sequences [18]. Large plant mitochondrial repeat fragments, exceeding 1,000 bp, frequently recombine with each other, not only subdividing the genome and increasing recombination dynamics, but also leading to the coexistence of mini and macrocyclic structures [51]. In the mitochondrial genome of *G. glabra*, a significant presence of repeat sequences has been observed, with the most prevalent being the DSRs, totaling 388. To empirically assess the

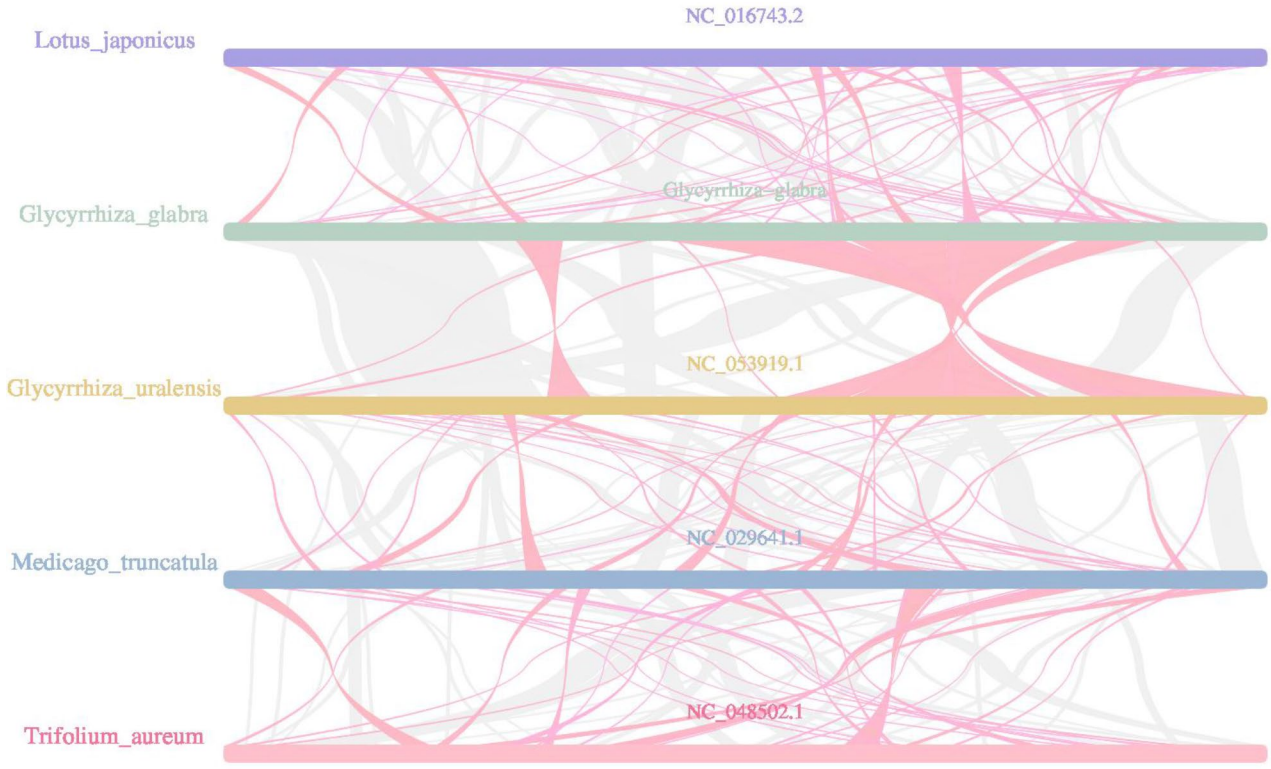


**Fig. 9** Synteny analysis of the *G. glabra* mitochondrial genome. Note: Purple, light green, yellow, cyan and red lines represent *Lotus japonicus*, *Glycyrrhiza glabra*, *Glycyrrhiza uralensis*, *Medicago truncatula* and *Trifolium aureum*, respectively. Red curved areas indicate regions where inversions occur, and the gray areas indicate regions with high homology

role of these repetitive sequences in mediating homologous recombination, we conducted PCR validation. Our findings revealed that certain repetitive sequences indeed play a role in this process, thereby contributing to the dynamic changes observed in the genome structure. Homologous recombination patterns have been shown to be important for reproductive diversity in higher plants, particularly for DNA repair by homologous recombination [23].

In addition, during mitochondrial evolution, some chloroplast fragments migrate into the mitochondrial genome [52], and the length and sequence similarity of the migrated fragments change over time. Thus, higher plant mitochondrial genomes contain sequences homologous to chloroplast DNA, facilitating the movement of genetic material throughout the organism [53]. In our study, we identified 12 mitochondrial genome fragments with a total length of 3,860 bp that are homologous to the chloroplast genome, including four complete genes (*trnW-CCA*, *trnD-GUC*, *trnN-GUU* and *trnM-CAU*), a finding that demonstrates the existence of gene transfer between chloroplasts and mitochondria.

RNA editing is a deamination reaction that is essential for gene expression in the mitochondrial genomes of higher plants and occurs through posttranscriptional processes [54]. Earlier studies reported this phenomenon, for example 455 RNA editing sites in 37 genes in *P. hunanensis* [55], 508 RNA editing sites in 31 genes in *S. plumbizincicola* [56], and 451 RNA editing sites in 31 genes in *S. officinalis* [57]. In our study, we predicted 460 RNA editing sites across 34 PCGs in *G. glabra* mitochondria, all of which involved C-U RNA editing. This C-U RNA editing event occurred mainly at the second codon position and most of them were completely edited, a process that increases the homology of mitochondrial protein sequences between different species and promotes plant growth and development [58].

Currently, phylogenetic analyses based on organelle genomes are mainly performed on the basis of chloroplast genomes [59, 60], and it has also been shown that mitochondrial genomes can also be used for evolutionary analyses [61]. Phylogenetic analyses of species based on plant mitochondrial genomes have been hampered by the structural complexity of these genomes and the difficulty of their assembly. To elucidate the phylogeny of G. glabra on the basis of mitochondrial genomic data and clarify taxonomic relationships, we constructed a phylogenetic tree using PCGs. The analysis demonstrated that *G. glabra* is closely related to *G. uralensis*. This result provides further evidence of the consistency between the

phylogenetic tree constructed from mitochondrial PCGs and the classification system set out in APG IV.

Analysis of the mitochondrial genome covariance of Glycyrrhiza glabra and its relatives showed that the mitochondrial genome of Glycyrrhiza glabra underwent a major genomic rearrangement, resulting in significant structural variability compared to its evolutionary relatives. This has allowed the mitochondrial genome to evolve and diversify. Certain regions of the mitochondrial genome of Glycyrrhiza glabra do not share homology with the mitochondrial genomes of other species, suggesting their unique presence in the mitogenome.

## Conclusions

In this study, we utilized short-read sequencing and long-read sequencing technologies to assemble the mitochondrial and chloroplast genomes of *G. glabra*. The following procedures were conducted: genome annotation, analysis of repetitive sequences, evaluation of organelle genome exchange, codon usage bias analysis, identification of RNA editing sites, comparative genomic analyses with closely related species, and phylogenetic relationship assessment. These findings will significantly advance our understanding of the structure and function of the mitochondrial genome in *G. glabra*, providing critical insights for the fields of conservation biology, population genetics, and evolutionary studies pertaining to this species.

### Abbreviations

| | |
|---|---|
| PCR | polymerase chain reaction |
| PCGs | protein coding genes |
| GC | Guanine-cytosine |
| tRNAs | Transfer RNAs |
| rRNA | ribosomal RNA |
| SSRs | Simple sequence repeats |
| TSRs | Tndem sequence repeats |
| DSRs | Dispersed sequence repeats |
| MTPTs | Mitochondrial plastid DNAs |
| RSCU | Relative synonymous codon usage |
| Met | methionine |
| Trp | tryptophan |
| Arg | arginine |
| Ser | serine |
| Por | proline |
| Leu | leucine |
| APG | Angiosperm Phylogeny Group classification |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11190-5.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

### Acknowledgements

We thank all the individuals who helped us in this study. We acknowledge the valuable work of the many investigators whose published articles we were unable to cite owing to space limitations.

### Author contributions

This study was designed by M.T and X.O. Data analysis was performed and the manuscript was written by G.Z. and M.Q. X.L. and Y.Q. revised the manuscript. All the authors made direct and intellectual contributions to this topic and approved the article for publication.

### Funding

This work was financially supported by grants from the Science and Technology Program of Qingyang City (Grant No. QY-STK-2023 A-030), and the Industrial Support Project of Colleges and Universities in Gansu Province (Grant No. 2020 C-24).

### Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics2021) of the National Genomics Data Center (NucleicAcids Res 2022), China National Center for Bioinformation /Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA019594) , which are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

## Declarations

### Ethics approval and consent to participate

No applicable.

### Consent for publication

No applicable.

### Competing interests

The authors declare no competing interests.

### Author details

[1]College of Agronomy & Resources and Environment, Tianjin Agricultural University, Tianjin 300384, China
[2]Gansu Key Laboratory of Protection and Utlization for Biological Resources and Ecological Restoration in Longdong, Longdong University, Qingyang, Gansu 745000, China
[3]School of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu 212013, China
[4]School of Agriculture and Bioengineering, Longdong University, Qingyang, Gansu 745000, China
[5]College of Agronomy, Gansu Agricultural University, Lanzhou, Gansu 730070, China
[6]Gansu Key Laboratory of Protection and Utilization for Biological Resources and Ecological Restoration in Longdong, School of Agriculture and Bioengineering, Longdong University, Qingyang 745000, Gansu, China

Received: 21 October 2024 / Accepted: 27 December 2024
Published online: 06 January 2025

### References

1. Lu D, Yang Y, Du Y, Zhang L, Yang Y, Tibenda JJ, Nan Y, Yuan L. The potential of glycyrrhiza from medicine food homology in the fight against digestive system tumors. Molecules. 2023;28(23).
2. Huang S, Zhang W, Hong Z, Yuan Y, Tan Z, Wang Y, Chen Z, Zheng J, Zhang Z, Zhang L, Chen M. Geographic distribution and impacts of climate change on the suitable habitats of Glycyrrhiza species in China. Environ Sci Pollut Res Int. 2023;30(19):55625–34.
3. Zhang Y, Sheng Z, Xiao J, Li Y, Huang J, Jia J, Zeng X, Li L. Advances in the roles of glycyrrhizic acid in cancer therapy. Front Pharmacol. 2023;14:1265172.
4. Li A, Zhao Z, Zhang S, Zhang Z, Shi Y. Fungicidal activity and mechanism of action of Glabridin from Glycyrrhiza glabra L. Int J Mol Sci. 2021;22(20).

5. Liu M, Du Y, Gao D. Licochalcone A: a review of its pharmacology activities and molecular mechanisms. Front Pharmacol. 2024;15:1453426.
6. Pastorino G, Cornara L, Soares S, Rodrigues F, Oliveira M. Liquorice (Glycyrrhiza glabra): a phytochemical and pharmacological review. Phytother Res. 2018;32(12):2323–39.
7. Kwasniak-Owczarek M, Janska H. Experimental approaches to studying translation in plant semi-autonomous organelles. J Exp Bot. 2024;75(17):5175–87.
8. Schiff JA, origin and evolution of the plastid and its function. Ann N Y Acad Sci. 1981;166–92.
9. Tian S, Lu P, Zhang Z, Wu JQ, Zhang H, Shen H. Chloroplast genome sequence of Chongming lima bean (Phaseolus lunatus L.) and comparative analyses with other legume chloroplast genomes. BMC Genomics. 2021;22(1):194.
10. Yanfei N, Tai S, Chunhua W, Jia D, Fazhong Y. Complete chloroplast genome sequences of the medicinal plant Aconitum Transsectum (Ranunculaceae): comparative analysis and phylogenetic relationships. BMC Genomics. 2023; 24(1).
11. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.
12. Hao Z, Zhang Z, Jiang J, Pan L, Zhang J, Cui X, Li Y, Li J, Luo L. Complete mitochondrial genome of Melia azedarach L., reveals two conformations generated by the repeat sequence mediated recombination. BMC Plant Biol. 2024;24(1):645.
13. 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.
14. Liu J, Yu S, Lu P, Gong X, Sun M, Tang M. De novo assembly and characterization of the complete mitochondrial genome of Phellodendron amurense reveals three repeat-mediated recombination. Gene. 2025;935:149031.
15. Handa H. Linear plasmids in plant mitochondria: peaceful coexistences or malicious invasions? Mitochondrion. 2008;8(1):15–25.
16. Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, MacLeod T, Pleasance S, Pandoh P, Zhao Y, Coope RJ, et al. Complete mitochondrial genome of a Gymnosperm, Sitka Spruce (Picea sitchensis), indicates a Complex Physical structure. Genome Biol Evol. 2020;12(7):1174–9.
17. Gray MW, Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. Rapid Evolution of enormous, multichromosomal genomes in Flowering Plant Mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10(1).
18. Cole LW, Guo W, Mower JP, Palmer JD, Purugganan M. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. Molecular Biology and Evolution; 2018.
19. Wu Z, Waneka G, Sloan DB. The Tempo and Mode of Angiosperm Mitochondrial Genome Divergence Inferred from Intraspecific Variation inArabidopsis Thaliana. G3 Genes|Genomes|Genetics. 2020;10(3):1077–86.
20. Zeng Z, Zhang Z, Tso N, Zhang S, Chen Y, Shu Q, Li J, Liang Z, Wang R, Wang J, Qiong L. Complete mitochondrial genome of Hippophae Tibetana: insights into adaptation to high-altitude environments. Front Plant Sci. 2024; 15.
21. Wu ZQ, Liao XZ, Zhang XN, Tembrock LR, Broz A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. J Syst Evol. 2020;60(1):160–8.
22. Štorchová H, Stone JD, Sloan DB, Abeyawardana OAJ, Müller K, Walterová J, Pažoutová M. Homologous recombination changes the context of cytochrome b transcription in the mitochondrial genome of Silene vulgaris KRA. BMC Genomics. 2018;19(1).
23. Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the Stability of the plant mitochondrial genome. Int J Mol Sci. 2020;21(1).
24. Liu Q, Guo S, Zheng X, Shen X, Zhang T, Liao B, He W, Hu H, Cheng R, Xu J. Licorice Germplasm resources Identification using DNA barcodes innervariants. Plants (Basel). 2021; 10(10).
25. Lee KJ, Raveendar S, Choi JS, Gil J, Lee JH, So Y-S, Chung J-W. Development of chloroplast microsatellite markers for identification of Glycyrrhiza species. Plant Genetic Resources: Charact Utilization. 2018;17(1):95–8.
26. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. Imeta. 2023;2(2):e107.
27. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666–9.
28. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37(23):4572–4.
29. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
30. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics. 2019;35(3):421–32.
31. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
32. Xiang CY, Gao F, Jakovlic I, Lei HP, Hu Y, Zhang H, Zou H, Wang GT, Zhang D. Using PhyloSuite for molecular phylogeny and tree-based analyses. Imeta. 2023;2(1):e87.
33. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
34. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27.
35. Kurtz S, Choudhuri JV, Ohlebusch E, Chris Schleiermacher1, Stoye2 J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research. 2001; 29:4633–4642.
36. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ. Marra MA: Circos: an Information Aesthetic for Comparative Genomics. 2024.
37. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
38. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
39. Lenz H, Hein A, Knoop V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. BMC Bioinformatics. 2018;19(1):255.
40. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics. 2018;34(14):2490–2.
41. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
42. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
43. Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, Yu J. Assembly of the complete mitochondrial genome of an endemic plant, Scutellaria tsinyunensis, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta. 2021; 254(2).
44. Tourasse NJ, Shtaida N, Khozin-Goldberg I, Boussiba S, Vallon O. The complete mitochondrial genome sequence of the green microalga Lobosphaera (Parietochloris) incisa reveals a new type of palindromic repetitive repeat. BMC Genomics. 2015;16(1):580.
45. Igamberdiev AU, Bykova NV. Mitochondria in photosynthetic cells: coordinating redox control and energy balance. Plant Physiol. 2023;191(4):2104–19.
46. Galtier N. The intriguing evolutionary dynamics of plant mitochondrial DNA. BMC Biol. 2011:61.
47. Xie P, Wu J, Lu M, Tian T, Wang D, Luo Z, Yang D, Li L, Yang X, Liu D, et al. Assembly and comparative analysis of the complete mitochondrial genome of Fritillaria Ussuriensis Maxim. (Liliales: Liliaceae), an endangered medicinal plant. BMC Genomics. 2024;25(1):773.
48. Min XJ, 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(1).
49. Yu X, Duan Z, Wang Y, Zhang Q, Li W. Sequence analysis of the complete mitochondrial genome of a Medicinal Plant, Vitex rotundifolia Linnaeus f. (Lamiales: Lamiaceae). Genes. 2022;13(5).
50. Wynn EL, Christensen AC. Repeats of unusual size in Plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
51. Cechova J, Lysek J, Bartas M, Brazda V. Complex analyses of inverted repeats in mitochondrial genomes revealed their importance and variability. Bioinformatics. 2018;34(7):1081–5.
52. Wang D, Wu YW, Shih AC, Wu CS, Wang YN, Chaw SM. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. Mol Biol Evol. 2007;24(9):2040–8.
53. Zhao N, Grover CE, Chen Z, Wendel JF, Hua J. Intergenomic gene transfer in diploid and allopolyploid Gossypium. BMC Plant Biol. 2019;19(1):492.
54. Hao W, Liu G, Wang W, Shen W, Zhao Y, Sun J, Yang Q, Zhang Y, Fan W, Pei S, et al. RNA editing and its roles in Plant Organelles. Front Genet. 2021;12:757109.

55. Chen L, Dong X, Huang H, Xu H, Rono PC, Cai X, Hu G. Assembly and comparative analysis of the initial complete mitochondrial genome of Primulina hunanensis (Gesneriaceae): a cave-dwelling endangered plant. BMC Genomics. 2024;25(1):322.
56. Ding H, Bi D, Zhang S, Han S, Ye Y, Yi R, Yang J, Liu B, Wu L, Zhuo R, Kan X. The Mitogenome of Sedum Plumbizincicola (Crassulaceae): insights into RNA editing, lateral gene transfer, and phylogenetic implications. Biology (Basel). 2022;11(11).
57. Yang H, Chen H, Ni Y, Li J, Cai Y, Wang J, Liu C. Mitochondrial genome sequence of Salvia officinalis (Lamiales: Lamiaceae) suggests diverse genome structures in Cogeneric species and finds the stop gain of genes through RNA editing events. Int J Mol Sci. 2023;24(6).
58. Small ID, Schallenberg-Rudinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101(5):1040–56.
59. wang y, Wang S, Liu Y, Yuan Q, Sun J, Guo L. 2020.
60. Wang L, Li F, Wang N, Gao Y, Liu K, Zhang G, Sun J. Characterization of the Dicranostigma leptopodum chloroplast genome and comparative analysis within subfamily Papaveroideae. BMC Genomics. 2022;23(1).
61. Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N. The mitochondrial genome of the Moss Physcomitrella patens sheds New Light on mitochondrial evolution in land plants. Mol Biol Evol. 2007;24(3):699–709.

## Publisher's note