Zhang *et al. BMC Genomics (2024) 25:1162*
https://doi.org/10.1186/s12864-024-11086-4

BMC Genomics

RESEARCH

Open Access

Comprehensive analysis of the first complete mitogenome and plastome of a traditional Chinese medicine *Viola diffusa*

Chenshuo Zhang[1], Aamir Rasool[2], Huilong Qi[1], Xu Zou[1], Yimeng Wang[1], Yahui Wang[1], Yang Wang[1*], Yan Liu[1*] and Yuan Yu[1*]

Abstract

Background *Viola diffusa* is used in the formulation of various Traditional Chinese Medicines (TCMs), including antiviral, antimicrobial, antitussive, and anti-inflammatory drugs, due to its richness in flavonoids and triterpenoids. The biosynthesis of these compounds is largely mediated by cytochrome P450 enzymes, which are primarily located in the membranes of mitochondria and the endoplasmic reticulum.

Results This study presents the complete assembly of the mitogenome and plastome of *Viola diffusa*. The circular mitogenome spans 474,721 bp with a GC content of 44.17% and encodes 36 unique protein-coding genes, 21 tRNA, and 3 rRNA. Except for the RSCU values of 1 observed for the start codon (AUG) and tryptophan (UGG), the mitochondrial protein-coding genes exhibited a codon usage bias, with most estimates deviating from 1, similar to patterns observed in closely related species. Analysis of repetitive sequences in the mitogenome demonstrated potential homologous recombination mediated by these repeats. Sequence transfer analysis revealed 24 homologous sequences shared between the mitogenome and plastome, including nine full-length genes. Collinearity was observed among *Viola diffusa* species within the other members of Malpighiales order, indicated by the presence of homologous fragments. The length and arrangement of collinear blocks varied, and the mitogenome exhibited a high frequency of gene rearrangement.

Conclusions We present the first complete assembly of the mitogenome and plastome of *Viola diffusa,* highlighting its implications for pharmacological, evolutionary, and taxonomic studies. Our research underscores the multifaceted importance of comprehensive mitogenome analysis.

Keywords *Viola diffusa*, Mitogenome, Chloroplast genome, RNA editing, Gene transfer, Recombination

*Correspondence:
Yang Wang
wangy@ncst.edu.cn
Yan Liu
liuysm@ncst.edu.cn
Yuan Yu
yuyuan@ncst.edu.cn
[1] College of Life Sciences, North China University of Science and Technology, 21 Bo Hai Road, Tangshan, People's Republic of China
[2] Institute of Biochemistry, University of Balochistan, Quetta 87300, Pakistan

Introduction

Viola diffusa Ging, a member of the *Violaceae* family commonly found in East Asia, is used in Traditional Chinese Medicines (TCMs) due to its antiviral, antimicrobial, antitussive, and anti-inflammatory properties [1]. Recently, researchers have shown that the flavonoid- and triterpenoid-rich extract of *V. diffuse* exhibits potent anti-hepatitis B activity [2].

The mitochondria and chloroplasts within eukaryotic cells are believed to have originated from free-living alphaproteobacteria and cyanobacteria, respectively, through endosymbiosis [3, 4]. These organelles have their own genetic material (DNA), which encodes essential genes; allowing them to perform certain functions independently, such as energy production (mitochondria) and photosynthesis (chloroplasts) [5].

The mitochondrion is a vital organelle for the survival of eukaryotic cells, as it is involved in various functions such as energy production, cell proliferation, differentiation, apoptosis, and stress response [6]. The structure of the mitogenome in plants differs from that in animals, and a wide range of mitochondrial complexities has been reported in many plant species [7, 8]. The largest plant mitogenome has been sequenced from angiosperms, and the general plant mitogenome size ranges from 66 kb to over 11.3 Mb [9, 10]. The length of the mitogenome of seed plants is larger than that of other plants because they can contain non-coding sequences, including those originated from intracellular gene transfer (IGT) or horizontal gene transfer (HGT) sequences. The presence of repetitive sequences in the mitogenome enhances the chances of homologous recombination, which subsequently contributes to the formation of multipart structures of the mitogenome of flowering plants [11].

The mitogenome is large, complex, and non-conservative compared to the plastome. The heterogeneity of mitogenome reflects the acquisition of foreign genetic material through intracellular gene transfer (IGT) and horizontal gene transfer (HGT) [12, 13]. Gene migration has been reported between mitochondria and plastids and between mitochondria and chloroplasts in different plants [14–16]. However, there is no report on gene transfer between the mitogenome and plastome of *V. diffusa*.

The GC content of the mitogenome may indicate how well species have adapted to global environmental fluctuations through mutations and alterations in gene expression profiles [17]. Furthermore, the plant mitogenome is a crucial tool for studying species' origin, classification, and phylogeny. It is characterized by extreme variation in size, sparse gene distribution, a large number of non-coding sequences, abundant repetitive sequences, the ability to incorporate foreign DNA, highly conserved gene sequences, and numerous RNA editing sites [18–20].

Since the structure of the plant mitogenome is more complex than that of chloroplast or plastid. By December 25, 2022, fewer mitogenomes (560) have been published in the NCBI database compared with chloroplast genomes (9,479) and plastid genomes (1,253).

Similar to mitogenome, plastome sequence analysis is also a very useful tool for studying the evolution and phylogenetic relationships among various lineages because it is haploid, maternally inherited, and possesses highly conserved genes [21, 22]. The plastome's single-parent inheritance (maternal) and recombination-free nature make it a valuable tool for phylogenetic studies, understanding gene flow, cytoplasmic diversity, and population differentiation [23, 24].

The combined analysis of the mitogenome and plastome will significantly enhance our understanding of the evolution of *V. diffusa* and other similar species. The sequencing data from the mitogenome and plastome will help elucidate the evolutionary history of *V. diffusa* and promote research on its phylogeny, population, and genetic engineering to optimize the production of flavonoids and terpenoids in this medicinal plant.

In this study, Illumina and Nanopore high-throughput sequencing technologies were employed to sequence the whole mitogenome and plastome of *V. diffusa*. Additionally, functional annotation, codon usage analysis, repeat sequence identification, comparative mitogenome and plastome analysis, gene transfer, and RNA editing analysis were performed. This study provides a theoretical basis for understanding the evolution, classification, and better identification of *V. diffusa*, which can also facilitate the exploitation of its germplasm resources.

Material and methods

Plant materials, DNA extraction, and sequencing

The entire plant of *Viola diffusa* was harvested from Baishajiang area in Shuangpai County, Yongzhou, Hunan Province, China (111°38′47.31″ E, 25°54′59.238″ N). Dr. Yuan Yu, a professor at the North China University of Science and Technology, taxonomically identified the collected plants and snap-froze them using liquid nitrogen. A voucher specimen (NCST20240327YY) is deposited at the Herbarium of North China University of Science and Technology. Genomic DNA was subsequently extracted using the CTAB method [25]. The integrity of the extracted genomes was determined using gel electrophoresis and quantified using a Nanodrop spectrophotometer. A Qubit assay (Thermo Fisher Scientific, USA) was then performed to quantify DNA for library construction and sequencing.

The sequencing of the mitogenome and plastome was performed on the NovaSeq 6000 platform (Illumina, San Diego, CA, USA) and the PromethION 48 (Oxford Nanopore Technologies, Oxford, UK) high-throughput sequencing platforms, with the services provided by Qiantang Biotechnology Co., LTD (China). Illumina sequencing was performed using the NovaSeq 6000 to generate short paired-end reads, with the genomic libraries prepared using the NEBNext® library preparation kit (New England Biolabs) [26]. For single long-read sequencing, genomic DNA was sheared into fragments averaging 10 kb using the g-TUBE device (Covaris, USA) and centrifugation at 8000 rpm for 60 s [27]. The long-read library was prepared and loaded onto a Flow Cell for sequencing on the MinION platform, following the manufacturer's instructions. The sequencing was performed by BioMarker [28].

Assembly and annotation of mitogenome and plastome

GetOrganelle software (v1.7.7.0) [29] was employed to assemble high-throughput sequencing data of plastome (cpDNA). We assembled *V. diffusa* chloroplast sequencing data using the Higher Plant Chloroplast Genome database (embplant_pt) as a seed in the software [30]. We assembled the complete mitogenome of *V. diffusa* using long-read sequencing data, using Flye software [31] with its default settings to assemble the long reads into a complete genome. The results were generated in GFA format, which provides a graphical representation of the assembly [32].

We used makeblastdb to build a BLAST database from all assembled contigs. Then, we employed the BLASTn program [33] to identify contigs corresponding to the mitochondrial genome, using conserved plant mitochondrial genes from *Arabidopsis* as query sequences. The parameters -perc_identity 80 -evalue 1e-5 -outfmt 6 -max_hsps 10 -word_size 7 -task blastn-short were used to identify high-confidence mitochondrial contigs. Bandage software (v0.8.1) [34] was used to visualize the GFA file and filter the mitochondrial contigs, which facilitated the assembly of the draft mitogenome of *V. diffusa* from the library of contigs. Then, the long and short reads were aligned to the mitochondrial contigs using BWA (v0.7.17) with the MEM algorithm [35]. The reads aligned with mitochondrial contigs were filtered, exported, and saved using the BWA (v0.7.17) and Samtools (v1.6) [35, 36] for subsequent hybrid assembly into the mitogenome of *V. diffusa*. The hybrid assembly of the mitogenome was performed using Unicycler [37] with default parameters. The long reads were aligned to the repetitive sequences and determined if they spanned the repetitive regions. The alignment results and coverage of repetitive regions by long reads allowed us to deduce the most probable structure of the mitogenome from these fragmented sequences. Bandage software was again used to visualize, explore, and verify the structure and completeness of the assembled mitogenome [34].

The cpDNA of *V. diffusa* was annotated using the 2544-plastomes library in CPGAVAS2 [38]. The CPG-view-RSG was employed to evaluate the annotation's accuracy and remove potential irregularities such as missing transcription origins, uncertain intron boundaries, the presence of pseudogenes, and incorrect annotation of genes [39]. AGORA was used to identify and annotate the genes and other features of mitogenome [40]. The accuracy of protein-coding genes (PCGs) and rRNAs of mitogenome annotations was evaluated using Apollo [41]. Finally, the graphical map of the mitogenome was created and visualized by OGDRAW (1.3.1) [42].

Analysis of repetitive and homologous sequences in the mtDNA

The repeated sequences present in DNA can be categorized into two types: tandem repeat sequences and scattered (or dispersed) repeat sequences [43]. The key difference between tandem repeats and scattered repeats lies in the arrangement of the repeated segments [44]. Tandem repeats refer to adjacent repeats, while scattered repeats are non-adjacent. Microsatellite repeats, which are no longer than 6 bp, represent a specific type of tandem repeats. Microsatellite repeats are commonly used in the development of molecular markers due to their codominant inheritance, high polymorphism, abundance in genomes, and ease of detection [45].

The BLASTn program with an e-value of 1e-6 was used to compare the mitochondrial genome to itself in order to identify homologous and repetitive sequences [46]. MISA (Microsatellite Identification Tool) (v2.1) (https://webblast.ipk-gatersleben.de/misa/), Tandem Repeats Finder (https://tandem.bu.edu/trf/trf.unix.help.html), and the REPuter web server (v4.09) (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) were used to identify microsatellite, tandem repeats, and duplicated repeat sequences [47–49]. The results were visualized using the Circos package (v0.69–9) [50].

To improve the completeness of the assembled mitogenome, the long reads were first aligned to the selected repeats, and the 1000 bp flanking regions on either side were used as reference sequences for the major conformation analysis. One side of the flanking region was subsequently exchanged to simulate the reconstituted reads, which served as reference sequences for the secondary conformation. We extracted sequences with homology > 80% from the BLASTn comparison results to verify homology and identify repetitive sequences. Subsequently, we used Primer-BLAST [51] to design primers to amplify selected regions.

Then, the upstream and downstream primers of repeats (F1R1 and F2R2 in Fig. 3) were exchanged (F1R2 and F2R1) to verify mediating recombination, respectively. The PCR reaction mixture (25 μl) was prepared using the following reagents: mitochondrial DNA (2.5 μl), PrimeStar HS DNA polymerase (0.25 μl), forward primer (0.5 μl), reverse primer (0.5 μl), dNTPs (2.5 μl), 5 × buffer (5 μl), and ddH_2O (13.75 μl). The PCR amplification was carried out for 30 cycles, with each cycle consisting of three steps: denaturation at 98 °C for 10 s, annealing at 56 °C for 15 s, and extension at 72 °C at a rate of 1 kb per minute.

Analysis of intracellular horizontal gene transfer (IHGT)

To date, there is no report on the sequencing of the nuclear genome of *V. diffusa*. Therefore, we could only detect intracellular horizontal gene transfer (IHGT) between the mitogenome and plastome. BLASTn was performed between mtDNA and cpDNA to investigate the incidence of intracellular horizontal gene transfer (IHGT) between the mitogenome and plastome. The parameters to run the BLASTn program were set as follows: percentage identity greater than 80%, e-value 1e-5, word size 9, gap open 5, gap extend 2, reward 2, penalty −3, and dust set to 'no' [52–56]. The results obtained from running the BLASTn program were visualized using the Circos package (v0.69–9) [50]. The DNA sequences showing evidence of IHGT were identified and extracted from the overall dataset based on their gene locations within the mitochondrial or plastome genomes. These sequences were then annotated using GeSeq [57].

Analysis of codon *bias* and RNA editing events

The preference for codon usage reflects the complex interplay of evolutionary forces, such as natural selection and genetic drift. These forces fine-tune codon preferences to balance translation efficiency, accuracy, and gene expression, ultimately contributing to the survival of organisms [58, 59].

Three independent laboratories from Canada, France, and Germany first reported the C-to-U RNA editing phenomenon in plant mitochondrial RNA [60]. Subsequent studies revealed that RNA editing is prevalent in higher plant mitochondria and plays a crucial role in gene expression within the plant mitogenome.

RNA editing is a post-transcriptional modification process, categorized under RNA processing. This process involves the deamination of cytosine (C) into uracil (U) at specific sites, typically at the second position of codons. Most of these sites are fully edited, enhancing the homology of mitochondrial protein sequences across different species. It has been reported that approximately 92% of RNA editing sites result in amino acid changes, often converting a hydrophilic amino acid to a hydrophobic one, thereby enhancing protein folding and functionality [61]. Additionally, RNA editing can create start and stop codons that are not encoded in the original genome sequence. The generation of these new start and stop codons during mRNA processing increases the conservation of the encoded proteins and enhances their homology with corresponding proteins in other species. This process also contributes to the evolutionary stability, functional integrity, and expression of mitochondrial genes.

For codon bias determination, Phylosuite software (v1.1.16) [62] was used to extract the protein-coding sequences, and the relative synonymous codon usage (RSCU) values of the amino acid composition of protein-coding genes from the mitogenome were calculated using MEGA (v7.0) [63]. Deepred-mt [64], a tool based on a convolutional neural network (CNN), was employed to predict RNA editing sites in all protein-coding genes (PCGs) of the *V. diffusa* mitogenome. We focused solely on the prediction of C-to-U RNA editing sites in PCGs, and results with a probability greater than 0.9 were retained for further analysis.

Phylogenetic and Collinearity analyses

We downloaded the mitogenomes of 23 previously reported species belonging to the order Malpighiales from NCBI [65] and used them as references. These species include members of families such as *Salicaceae* Mirb., *Euphorbiaceae* Juss., and *Rhizophoraceae* Pers. Additionally, two mitogenomes from the *Zygophyllaceae* family were used as an outgroup. MAFFT software (v7.505) [66] was used for multiple sequence alignment analysis. Subsequently, IQ-TREE software (v1.6.12) [67] was employed to reconstruct the phylogenetic tree using concatenated sequences of 24 PCGs (*atp1, atp4, atp6, atp8, ccmB, ccmC, ccmFC, ccmFN, cob, cox1, cox2, cox3, matR, nad1, nad2, nad4, nad5, nad6, nad7, nad9, rps3, rps4, rps12, sdh4*) shared by the 26 species. The parameters for phylogenetic tree construction were set to 'GTR + F + I + G4' with a bootstrap of 1000. The results of the phylogenetic analysis were visualized using ITOL software (v6) [68]. The mitogenomes of closely related species, including *Salix wilsonii* Seemen, *Populus davidiana* Dode, *Bruguiera sexangula* (Lour.) Poir., and *Ricinus communis* L., were used for homology analysis. To visualize synteny and identify conserved genomic regions, multiple synteny plots were generated for *V. diffusa* and its closely related species using the MCscanX toolkit [69].

Results

Characterization of the mitogenome of *V. diffusa*

We obtained 41,790,910 sequences of the mitogenome, each 150 bp in length, through sequencing on the Illumina platform. Additionally, we obtained 17,943,395 sequences from the Oxford Nanopore platform, with an average length of 14,063 bp, including 1,217,030 sequences longer than 2,000 bp. The combination of short and long reads was processed to distinguish between two mitogenome conformations: a single circular genome structure and a more complex structure that could consist of multiple intersecting circles. The results of mitogenome obtained from both short and long reads exhibited a high degree of sequence similarity. Illumina short read sequencing assembled a complete mitogenome of *V. diffusa*, which spans 474,721 bp and comprises 44.17% GC content (GenBank accession number: PP952082). Additionally, a draft plastome of 157,904 bp was also obtained through GetOrganelle assembly (GenBank accession number: PP952083).

The analysis of the long reads revealed a single circular mitogenome structure (Figure S1A). In contrast, the examination of short reads revealed the presence of numerous repetitive sequences in the mitogenome of *V. diffusa*, that could result in the formation of the complex structure of multiple intersecting circles (Figure S1B). During the assembly process, long reads can effectively cover the entire span of repetitive regions, resulting in a uniquely resolved path. However, this does not imply that repetitive sequences are absent in the assembly results. In contrast, short reads lack this advantage, as repetitive sequences longer than the read length can complicate the selection of the best assembly path, leading to multiple possible outcomes.

To overcome the issue associated with the presence of repetitive regions in the mitogenome, Unicycler pipeline used in this study aligns long reads to these repetitive sequences and helps resolve the complex regions that challenge short read assembly alone. The alignment results provide the most probable structure of the mitogenome (Figure S1C).

While the sequence of a circular contig was resolved based on long-read data, particular attention was given to the branching nodes generated due to repetitive sequences, which are depicted as red nodes, with each red node representing a potential repetitive sequence. This focus on branching nodes highlights the challenges posed by repetitive regions in the assembly process (Figure S1D).

The annotation of the mitogenome of *V. diffusa* reveals its intricate genetic architecture, highlighting 36 unique protein-coding genes. Among these, 24 genes are classified as mitochondrial core genes, while 12 are classified as non-core genes. The core genes encompass five ATP synthase genes (*atp1, atp4, atp6, atp8, and atp9*), nine NADH dehydrogenase genes (*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7,* and *nad9*), four cytochrome C biogenesis genes (*ccmB, ccmC, ccmFC,* and *ccmFN*), three cytochrome C oxidase genes (*cox1, cox2,* and *cox3*), one membrane transport protein gene (*mttB*), one maturase gene (*matR*), and one pantothenate-cytochrome C reductase gene (*cob*). Conversely, the non-core genes comprise four ribosomal large subunits (*rpl2, rpl5, rpl10,* and *rpl16*), six ribosomal small subunits (*rps3, rps4, rps10, rps12, rps14,* and *rps19*), and two succinate dehydrogenase genes (*sdh3* and *sdh4*) (Fig. 1A, Table 1). The annotation results were similar to those of related species in the Malpighiales order, but only quantitative differences were observed in most genes. These quantitative differences refer to variations in the number of specific genes annotated across related species. This includes differences in the copy number of certain genes or variations

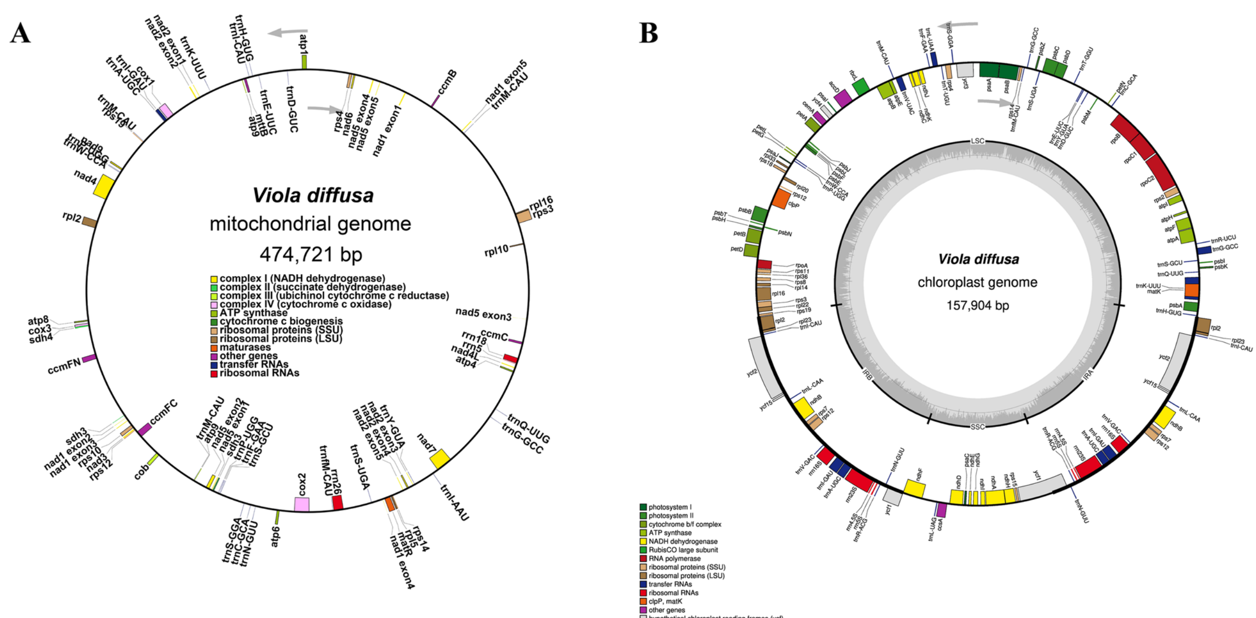

Fig. 1 Organelle genome map. A. Map of mitogenome. The map denotes annotated genes grouped based on different functions, which are color-coded on the outer circle as transcribed clock-wise (inside) and counter clock-wise (outside). B. Map of plastome. The map denotes annotated genes grouped based on different functional, which are color-coded on the outer circle as transcribed clock-wise (inside) and counter clock-wise (outside)

Table 1 Mitochondrial coding genes in *V. diffusa*

Group of protein/RNAs	Name of genes
ATP synthases	*atp1, atp4, atp6, atp8, atp9* (×2)
NADH dehydrogenases	*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*
Cytochrome b	*cob*
Cytochrome *c* biogenesis	*ccmB, ccmC, ccmFC, ccmFN*
Cytochrome *c* oxidases	*cox1, cox2, cox3*
Maturase	*matR*
Protein transport subunit	*mtt*B
Ribosomal protein large subunits	*rpl*2, *rpl*5, *rpl*10, *rpl*16
Ribosomal protein small subunits	*rps*3, *rps*4, *rps*10, *rps*12, *rps*14, *rps*19
Succinate dehydrogenases	*sdh*3, *sdh*4
Ribosome RNAs	*rrn*5, *rrn*18, *rrn*26
Transfer RNAs	*trnA-UGC, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnH-GUG, trnI-AAU, trnI-CAU, trnI-GAU, trnK-UUU, trnM-CAU* (×2), *trnN-GUU, trnP-UGG* (×2), *trnQ-UUG, trnS-GCU, trnS-GGA, trnS-UGA, trnW-CCA, trnY-GUA*

Numbers in parentheses indicate the copy number of each gene, such as (×2) to denote two copies

in the presence of specific non-coding regions, while the overall gene content remains relatively consistent [70–73].

Similarly, chloroplast assembly using GetOrganelle resulted in a single circular plastome containing 131 genes, including *ycf1, ycf2, ycf3, ycf4, psaA, psaB, psaI, psbA, psbB, psbC,* and *psbD* (Fig. 1B).

Analysis of repetitive and homologous sequences in the mitogenome

We identified approximately 244 simple sequence repeats (SSRs) in the mitogenome of *V. diffusa*. The monomeric (single repeat units) and dimeric (two repeat units) forms accounted for about 48.77% of the total SSRs (Fig. 2A). The adenine-rich monomeric repeats constituted 54.76% (46) of the 84 monomeric SSRs. The tandem repeat sequences (satellite DNA) consist of core repeat units ranging from ~7 to 200 bp, found in tandem multiple times. These sequences are prevalent in both eukaryotic and prokaryotic genomes. The results demonstrated that the mitogenome of *V. diffusa* possesses 9 tandem repeats with a sequence identity greater than 81%, with lengths ranging from 14 to 36 bp.

Further examination of scattered repeat sequences revealed 321 repeat pairs with a length of 30 bp or greater. The longest palindromic repeat was 360 bp, while the longest forward repeat measured 455 bp. In summary, we identified 9 tandem repeats, 170 palindromic repeats, and 151 forward repeats in the mitogenome of *V. diffusa* (Fig. 2B and Tables S1-3). A chord plot of repetitive sequence homology was generated based on the position of each repeat type on the mitogenome of *V. diffusa* (Fig. 2C).

We selected four pairs of repetitive sequences, R2, R6, R7, and R11-through a comprehensive analysis of the repetitive sequences in the mitogenome of *V. diffusa*. These sequences span over 1000 bp flanking regions of mitogenome (Table S4). DNA sequencing of PCR-amplified products indicated that the fragment between the R6 repeat was reversed relative to the mitogenome sequence. In contrast, sequencing of the R2 repeat sequence showed the amplification of two relatively complete ring structures through cross-primer PCR, suggesting that R2, most likely, mediates the conversion from single-ring to two subgenomic structures. These findings indicate that R6 has a distinct function, facilitating the reversal of fragment orientation between repetitive sequences, while R2 enables the conversion of single-loop mitochondria to two subgenomic structures (Fig. 3).

Analysis of intracellular horizontal gene transfer (IHGT)

During the annotation of the mitogenome, we observed the presence of chloroplast-like gene sequences. This observation indicated the putative occurrence of the IHGT phenomenon between these two organelles and was confirmed by BLASTn. The results of the BLASTn displayed the presence of 24 homologous sequences spanning over 12,894 bp, accounting for 2.72% of the mitogenome (Table S5).

The annotation of these homologous sequences revealed the presence of nine complete genes, including one protein-encoding gene (*ycf15*) and eight tRNAs (*trnD-GUC, trnH-GUG, trnI-GAU, trnN-GUU, trnM-CAU, trnP-UGG, trnS-GGA, trnW-CCA*) (Fig. 4). Among these sequences, MTPT22 was the longest, measuring 2,891 bp. We found that not all of the mitochondrial and

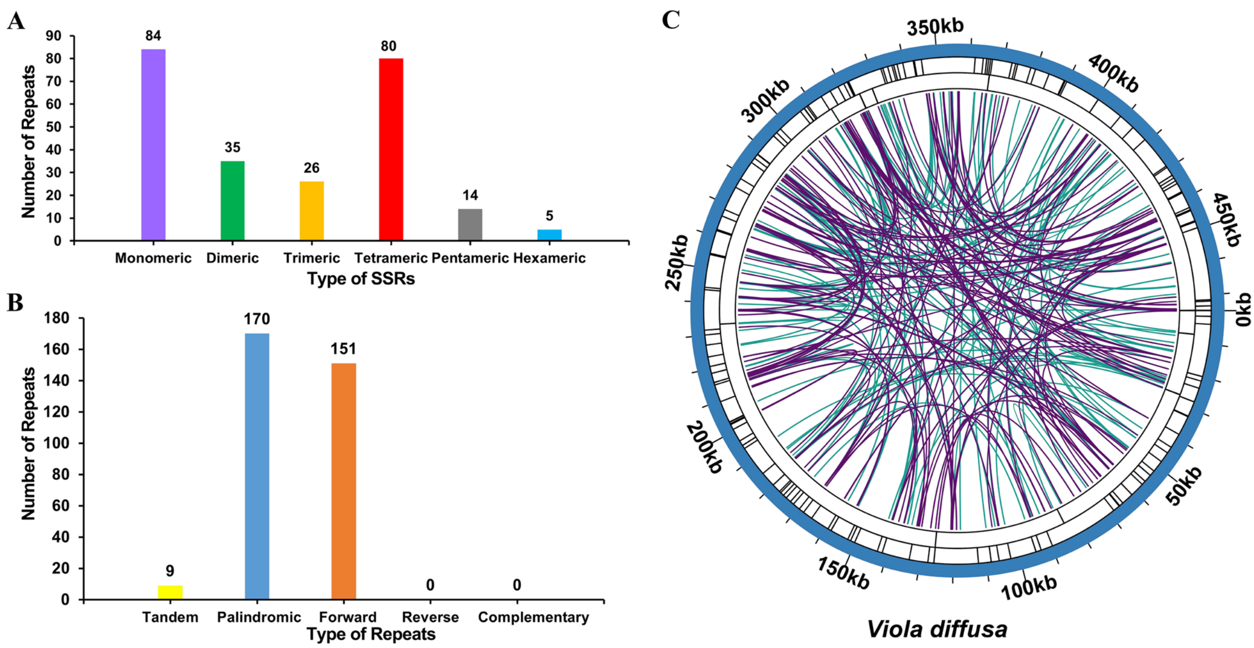

Fig. 2 Analysis of mitochondrial genome repeat sequence. A. Bar chart of SSRs. The abscissa represents the type of SSRs, ordinate represents the number of repeats sequences, purple legend represents the monomer SSRs, green legend represents the dimer SSRs, yellow legend represents the trimer SSRs, red legend represents the tetramer SSRs, gray legend represents the pentamer SSRs, and blue legend represents the hexamer SSRs. B. Bar chart of repeat sequence. The abscissa indicates the type of repeat sequence, ordinate indicates the number of repeat sequences, yellow legend indicates tandem repeats, blue legend indicates palindromatic repeats, and orange legend indicates forward repeats. Inverted and complementary repeats sequences were not detected in the mitogenome. C. Chordogram of repeats of mitogenome. The colored lines on the innermost circles connect the two repeats of the dispersed repeats, with purple lines representing Palindromic repeats and green lines representing Forward repeats. The black line segment on the second circle represents tandem repeats, and the black line segment on the outermost circle represents microsatellite repeats

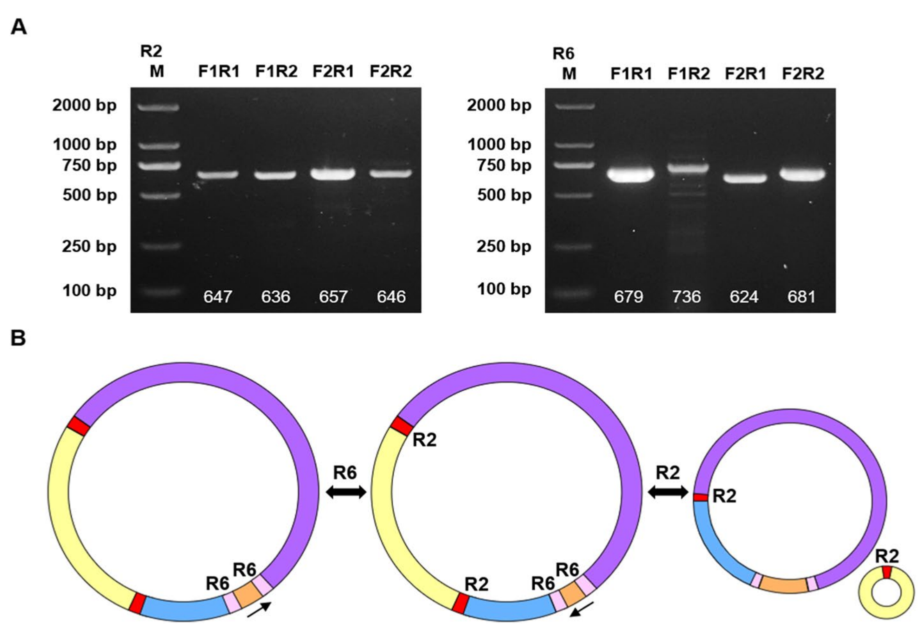

Fig. 3 Validation of homologous mediated recombination by repeats. A: Validation of repeat sequence in the mitogenome of *V. diffusa*; B: Schematic representation of repeat sequence-mediated conformational recombination of the mitogenome of *V. diffusa*

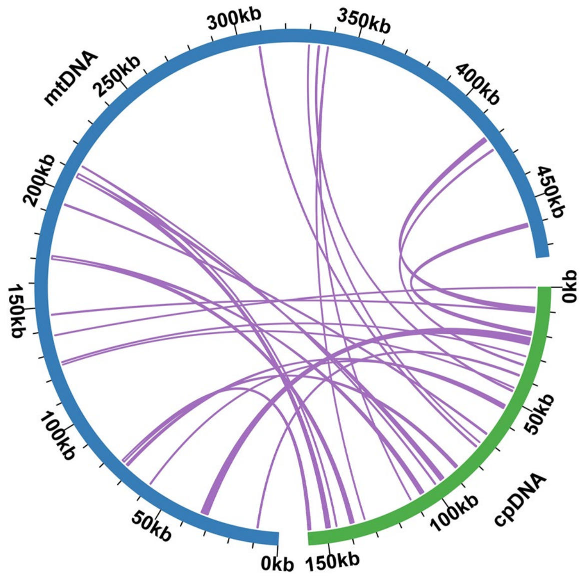

Fig. 4 Sequence migration analysis. The blue arc represents the mitogenome, green arc represents plastome, and corresponding genomic sequences of the purple lines between the arcs are homologous sequences

chloroplast homologous sequences identified were 100% similar, but most were greater than 90% similar. This observation indicates that the majority of homologous sequences have undergone significant evolutionary changes, leading to a loss of integrity. Consequently, only partial sequences of these genes are retained in the mitochondrial and chloroplast in contemporary species. The sequences *atpA* and *atpF*, *trnW-CCA* and *trnP-UGG*, and *trnI-GAU* and *trnA-UGC* were found in close genomic proximity in the mitogenome, suggesting their critical role in IHGT.

Analysis of codon *bias* and RNA editing events

The codon usage frequency exhibit the synonymous codon usage ratio in protein coding genes (RSCU), and values greater than 1 indicating amino acid bias (Fig. 5). The RSCU value for the start (AUG) and tryptophan (UGG) codons were equal to 1, but a notable preference was observed for other codons. For example, the stop codon UAA had the highest RSCU value of 1.63 among mitochondrial PCGs, indicating a preference for this specific codon. Similarly, alanine (Ala) showed a preference for the GCU codon, with an RSCU value of 1.58, indicating a bias towards using GCU to encode alanine. Additionally, the calculation results from MEGA showed that the RSCU values for phenylalanine (Phe) and cysteine (Cys) codons were less than 1.2, ranging from 0.8 to 1.1, indicating no obvious codon usage bias and a relatively flexible codon usage pattern.

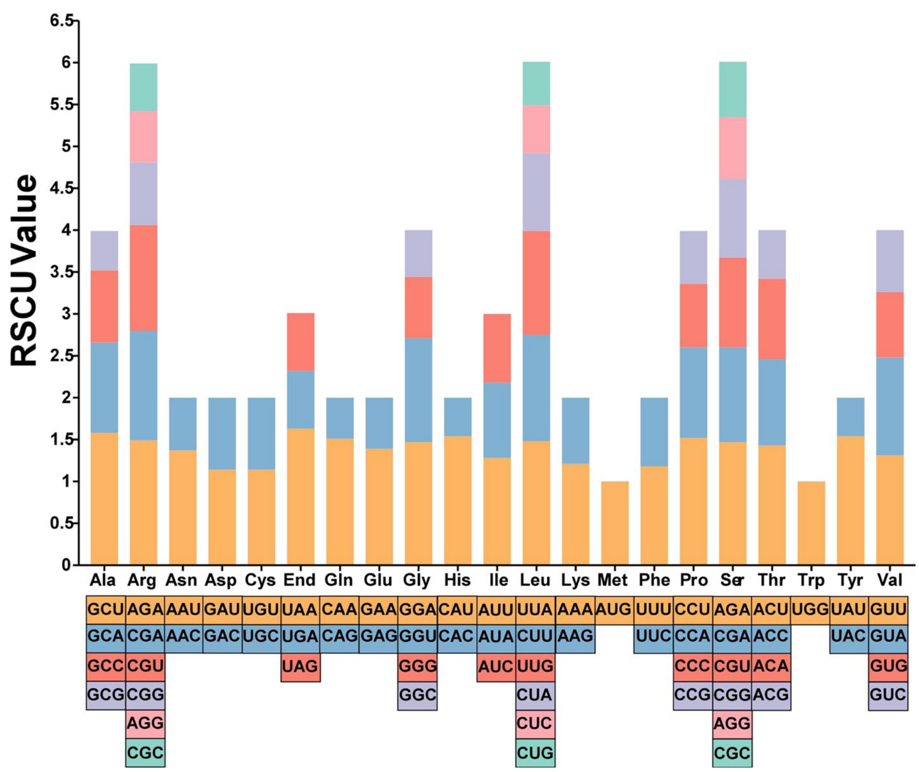

Fig. 5 Codon bias analysis of mitogenome of *V. diffusa*

In this study, we identified 430 potential RNA editing sites in the 36 protein-coding genes (PCGs) of *V. diffusa*, with the most common mutations being from cytidine (C) to uridine (U) (Table S6). The highest number of RNA editing sites (37) was observed in the *ccmB* and *nad4* genes. The second highest number of RNA editing sites (28) was observed in the *ccmC* gene (Fig. 6). The majority of RNA editing sites (80 in total) displayed a 100% probability of mutation, while all the remaining RNA editing sites demonstrated a probability higher than 90%. Mutations at the second codon position accounted for the highest number of 259 sites or 60.2%, followed by the first codon position (151 or 35.1%). Most of the RNA editing events resulted in transitions to leucine (189 times), with serine to leucine accounting for the highest number (98), followed by proline to leucine (85), and five synonymous mutations. The presence of a higher number of RNA editing sites causing mutations into leucine in the mitogenome of *V. diffusa* also corroborates the findings of Sheng and Wang et al. [16, 74].

Phylogenetic and collinearity analyses

The mitogenome sequences of specific plant species used in this study are listed in Table S7, which contains the mitogenomic data necessary for the phylogenetic analysis. This analysis focused on 24 conserved mitochondrial PCGs, including *atp1, atp4, atp6, atp8, ccmB, ccmC, ccmFC, ccmFN, cob, cox1, cox2, cox3, matR, nad1, nad2, nad4, nad5, nad6, nad7, nad9, rps3, rps4, rps12,* and *sdh4*. The resulting topology of the phylogenetic tree, based on mitochondrial DNA, aligns with the latest classification proposed by the Angiosperm Phylogeny Group (APG). In this phylogenetic tree, *V. diffusa* belongs to the family *Violaceae* within the order Malpighiales, and clusters closely with species from the family *Salicaceae* (Fig. 7A).

The multiple synteny plot of *V. diffusa* with closely related species highlights regions of inversion (red arced areas) and good homology (grey areas) (Fig. 7B). Short collinear blocks (< 0.5 kb) were excluded from the analysis to reduce noise, focus on significant conserved regions, and improve the clarity and interpretability of the results. While several homologous collinear blocks were identified between *V. diffusa* and closely related species within the Malpighiales order; the lengths of these collinear blocks were relatively short. This suggests that while synteny exists, the regions of conserved gene order are relatively small, indicating potential mitogenomic rearrangements and evolutionary changes that have occurred over time. Furthermore, distinct regions were observed in the *V. diffusa* mitogenome that exhibited no homology with the mitogenomes of other species. These findings demonstrate that the arrangement of collinear blocks among the five species was inconsistent, and the mitogenome of *V. diffusa* has undergone significant rearrangements compared to its close relatives. The mitogenome sequences of these five species exhibit substantial variation in organization and have probably undergone frequent recombination (Table S8).

Discussion

In previous studies by Dai J-J (2015) and Dai Nin (2023), *Viola diffusa* was shown to possess anti-inflammatory, antiviral, detoxifying, and other medicinal activities, primarily attributed to its triterpenoids, flavonoids, and polysaccharide compounds [2, 75]. Mitochondria express a various key enzymes involved in numerous metabolic pathways, including the biosynthesis of triterpenoids, steroids, and flavonoids. For example, mitochondria provide ATP and reducing equivalents to the mevalonate pathway, which is critical for the synthesis of triterpenoids. Additionally, mitochondria supply energy and reducing equivalents necessary for flavonoid biosynthesis. Therefore, analyzing the mitochondrial genome of *V. diffusa* can provide valuable insights into the biosynthetic pathways of these medicinal compounds, ultimately

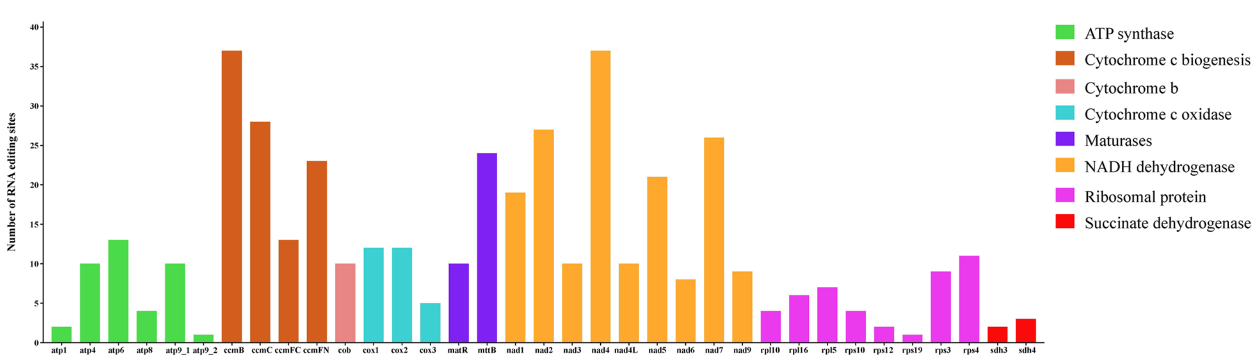

Fig. 6 Number of RNA editing sites in individual mitochondrial PCGs

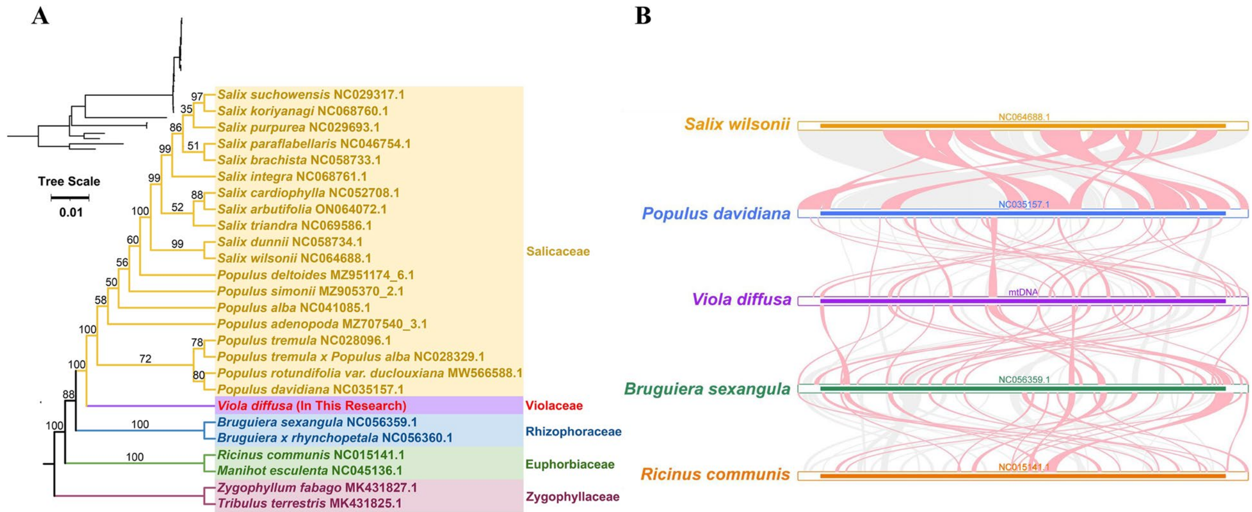

Fig. 7 Evolutionary analysis of *V. diffusa*. **A**. Phylogenetic tree based on 24 conserved protein-coding genes from 26 typical plant mitogenomes. The number on each node is the bootstrap support value. *Tribulus terrestris* and *Zygophyllum fabago* were used as outgroups. **B**. Pairwise synteny analysis of mitogenome of five species of the genus *Cinerea*. Bars represent the mitogenome, and bands represent homologous sequences between adjacent species

enhancing our capacity to develop effective plant-derived medicines for disease management.

The duplicate sequences in higher plant mitogenomes contribute to their complex ring or linear structures and facilitate gene recombination [76]. Recent studies have revealed structural variations across plant species' mitogenomes. For instance, *Populus simoni's* mitogenome has three ring chromosomes, while *Aeginetia indica's* includes a ring chromosome and a linear chromosome. *Thonningia sanguinea's* mitogenome is notable for its 18-ring chromosomes [77–79]. The main structure of the mitogenome of *V. diffusa* was found to be a single-circle, similar to previously sequenced plant mitogenomes in the Malpighiales order. The annotation of sequencing data indicated the presence of 321 pairs of dispersed repetitive sequences in the mitogenome of *V. diffusa;* however, only 13 of these repeats were longer than 200 bp. A set of R2 and R6 sequences that may support homologous recombination was also identified within these repetitive sequences through long-read alignment, flanking sequence extension, and cross-primer PCR. Using cross-primer PCR, two relatively complete circular structures were amplified: one corresponding to the main mitochondrial circular sequence and the other to the sequence between the R2 repeats. The amplification of two distinct circular structures suggests that the mitogenome of *V. diffusa* can exist in more than one structural form, implying subgenomic conformations. Based on the structural variability of mtDNA, it can be concluded that repetitive and homologous sequences may be involved in mitochondrial gene recombination events, leading to the formation of non-dominant mitogenome structures [43, 48, 49]. Zhou Hong et al. proposed the existence of sub-loop or polyloop structures in the mitogenome, suggesting that a single loop can appear at different stages of plant growth to support development [80]. This aspect may not have been considered in most mitogenome sequencing studies; however, we plan to sequence and assemble the mitogenome of *V. diffusa* at different growth stages to investigate the impact of these stages on the configuration of the mitogenome in our forthcoming study.

Mitochondria are maternally inherited organelles that contain genetic information distinct from that of the nucleus. The study of mitochondrial phylogeny can reveal different evolutionary patterns compared to nuclear genome, which is significant for understanding hybridization and incomplete lineage sorting during species evolution [81, 82]. The comparison of the mitogenome of *V. diffusa* with those of 23 species from the same order, along with 2 species from the *Zygophyllaceae* family as an outgroup, revealed that *V. diffusa* and species from the *Salicaceae* family were grouped together. This grouping pattern aligns with the latest APG IV classification, indicating a close evolutionary relationship between these species. Furthermore, we observed significant inconsistencies in short homology blocks and their arrangement between the mitogenome of *V. diffusa* and those of related species. This suggests that gene rearrangements occurred over evolutionary time within this group of plants. The comparative analysis of homologous blocks between *V. diffusa* and *P. davidiana* helped to identify 30 conserved genes (*nad1, nad2, nad3, nad4, nad5, nad6, nad7, rps3, rps4, rps19, cox1, cox2, ccmB, rpl2, rpl5, rpl10, rpl16, atp4, atp6, rrn5, rrn18, rrn26, sah3, cob, trnM-CAU, trnN-GUU, trnS-UCA, trnI-CAU, trnP-UGG, trnA-UGC*). However, many of these genes within homologous blocks were gene fragments, and their sequence identity ranged from 80 to 100%. These findings support the hypothesis of gene rearrangements in the mitogenome of *V. diffusa*, while highlighting the important role of these large-scale rearrangements in plant evolution [83]. The sequence similarity analysis helped to identify 24 homologous sequences transferred from the plastome to the mitogenome due to IHGT. The annotation of these genes revealed the presence of nine complete genes, including one protein-coding gene (*ycf15*) and eight tRNA (*trnD-GUC, trnH-GUG, trnI-GAU, trnN-GUU, trnM-CAU, trnP-UGG, trnS-GGA, trnW-CCA*) in the mitogenome. The discovery of *ycf15*, a highly conserved protein-coding gene in the plastome, suggests that gene migration from the plastome to the mitogenome has occurred, similar to the observation in other higher plants [84–86]. Additionally, eight translocated tRNAs in the plastome may have become pseudogenes [87, 88]. Previous studies have reported the loss or alteration of similar pseudogenes in the mitogenome of *Angelica biserrata* and the *Cistanche* genus [76, 89].

In addition to sequence migration and gene recombination, RNA editing is also a significant contributor to variations in the mitogenome of angiosperms [90, 91]. RNA editing primarily involves changes in nucleotide sequences through insertions, deletions, and substitutions, resulting in modifications of genetic information [92].

Studies have demonstrated that RNA editing is primarily facilitated by enzymes called deaminases, which catalyze base substitution reactions, such as cytosine to uracil (C to U), uracil to cytosine (U to C), and adenine to hypoxanthine (A to I) transitions in organelles [93]. RNA editing alters the sequences, even introns, making them more compatible with the splicing machinery. Additionally, RNA editing plays a vital role in the evolutionary adaptation and development of plants [94]. Furthermore, alterations in the proportion of hydrophilic amino acids are vital for the proper folding of proteins [95–97]. The sequencing results of the mitogenome of *V. diffusa* demonstrated the presence of 430 RNA editing sites, with ~ 95.59% located in the first and second codons. The primary RNA editing events identified involved changes from cytosine to thymine (C to T). The remaining 4.41% of editing sites were situated in the third codon, all of which also involved transitions from cytosine to thymine (C to T). It was observed that RNA editing sites were extensively found in most protein-coding genes of the mitogenome, suggesting that these editing events may influence gene expression. Among the identified RNA editing sites, there were 2 stop codon mutations and 24 synonymous amino acid mutations, including 17 sites with alterations in the third codon position. This finding suggests that RNA editing events may be associated with codon bias and the optimization of codon usage for more efficient gene expression.

The codon is crucial for gene expression as it plays a key role in translating genetic information. The study of codons becomes even more important in the context of mutations in genes, which can lead to variations in protein structure and function [98, 99]. The utilization of specific synonymous codons is influenced by species-specific preferences, which play a vital role in shaping the genetic traits of organisms [59, 98].

Understanding these preferences can provide insights into the molecular mechanisms of protein synthesis and the genetic characteristics of *V. diffusa* at the mitochondrial level.

Conclusions

In this study, the mitochondrial genome (mitogenome) of *Viola diffusa* was sequenced, assembled, and annotated with high precision. A mitogenome of 474,721 bp in length, with a GC content of 44.17%, was obtained, including 36 unique protein-coding genes, 21 tRNAs and 3 rRNAs. Additionally, 430 RNA editing sites, with a bias toward C to U, were predicted in the mitogenome. Homology analysis suggested a potential complex conformation of the mitogenome, and alignment analysis revealed collinearity between organelles, including evidence of IHGT from the chloroplast. Phylogenetic and collinearity analyses uncovered numerous gene rearrangement events in the mitogenome of *Viola diffusa*, indicating an evolutionary trend. This study provides valuable insights into the evolutionary history and phylogenetic analysis of *Viola diffusa*, laying a foundation for future research.

Abbreviations

V. diffusa	*Viola diffusa*
PCGs	Protein-coding genes
bp	Base pairs
SSRs	Microsatellite repeats
RSCU	Relative synonymous codon usage
APG	Angiosperm Phylogeny Group

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11086-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.

Acknowledgements
Not applicable.

Authors' contributions
YY, YL, YW, CSZ, AR, and HQ developed the research plan. YY, XZ, and YMW performed the experiments. YY, YHW, and CSZ collected and analyzed data. YY, YL, YW, and CSZ wrote the manuscript. All authors commented on and revised the manuscript.

Funding
The authors would like to acknowledge the funding support from the National Natural Science Foundation of China (No. 32171430) and Natural Science Foundation of Hebei Province (No. B2021209008).

Data availability
The sequence and annotation of Viola diffusa mitogenome and plastome have been submitted to the NCBI. The accession numbers in Gene Banks are PP952082 and PP952083.

Declarations

Ethics approval and consent to participate
The collection and cultivation of *V. diffusa* complied with relevant institutional, national, and international guidelines and legislation. The *V. diffusa* plants used in this experiment were grown in the Baishajiang area in Shuangpai County, Yongzhou, Hunan Province, China. Ethical approval or consent was not required for this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 July 2024 Accepted: 25 November 2024
Published online: 02 December 2024

References

1. Xu G. Herbological study of some medicinal plants from Viola. Zhong Yao Cai. 1997;20(7):371–3.
2. Dai J-J, Tao H-M, Min Q-X. Anti-hepatitis B virus activities of friedelolactones from Viola diffusa Ging. Phytomedicine. 2015;22(7–8):724–9.
3. Lang BF, Seif E, Gray MW, O'Kelly CJ, Burger G. A comparative genomics approach to the evolution of eukaryotes and their mitochondria. J Eukaryot Microbiol. 1999;46(4):320–6.
4. Douglas SE. Development: Plastid evolution: origins, diversity, trends. Curr Opin Genet Dev. 1998;8(6):655–61.
5. Christensen AC. Mitochondrial DNA repair and genome evolution. Annual Plant Reviews. 2017;50:11–31.
6. Chandel NS. Mitochondria as signaling organelles. BMC Biol. 2014;12:1–7.
7. Choi I-S, Schwarz EN, Ruhlman TA, Khiyami MA, Sabir JS, Hajarah NH, Sabir MJ, Rabah SO, Jansen RK. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol. 2019;19:1–15.
8. Li S, Xue L, Su A, Lei B, Wang Y. Hua JJ. Progress on sequencing and alignment analysis of higher plant mitochondrial genomes. JCAU. 2011;16(2):22–7.
9. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Taylor DR. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012;10(1):e1001241.
10. 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 U S A. 2015;112(27):E3515–24.
11. Gandini CL, Garcia LE, Sanchez-Puerta MV. Evolution: The complete organelle genomes of Physochlaina orientalis: Insights into short sequence repeats across seed plant mitochondrial genomes. Mol Phylogenet Evol . 2019;137:274–84.
12. Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. Sci Rep. 2017;7(1):43402.
13. Zhao N, Wang Y, Hua JJ. The roles of mitochondrion in intergenomic gene transfer in plants: a source and a pool. Int J Mol Sci. 2018;19(2):547.
14. Li J, Li J, Ma Y, Kou L, Wei J, Wang WJ. The complete mitochondrial genome of okra (Abelmoschus esculentus): using nanopore long reads to investigate gene transfer from chloroplast genomes and rearrangements of mitochondrial DNA molecules. BMC Genomics. 2022;23(1):481.
15. Cao P, Huang Y, Zong M, Xu ZJ. De Novo assembly and comparative analysis of the complete mitochondrial genome of Chaenomeles speciosa (Sweet) Nakai revealed the existence of two structural isomers. Genes (Basel). 2023;14(2):526.
16. Wang Y, Chen S, Chen J, Chen C, Lin X, Peng H, Zhao Q, Wang X. Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of Photinia serratifolia. Sci Rep. 2023;13(1):770.
17. Li Z, Ran Z, Xiao X, Yan C, Xu J, Tang M, An M. Comparative analysis of the whole mitochondrial genomes of four species in sect. Chrysantha (Camellia L.), endemic taxa in China. BMC Plant Biol. 2024;24(1):1–18.
18. Lu G, Zhang K, Que Y, Li Y. Assembly and analysis of the first complete mitochondrial genome of Punica granatum and the gene transfer from chloroplast genome. Front Plant Sci. 2023;14:1132551.
19. Fan W, Liu F, Jia Q, Du H, Chen W, Ruan J, Lei J, Li DZ, Mower JP, Zhu A. Fragaria mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by microinversions. New Phytol. 2022;236(2):745–59.
20. Cheng Y, He X, Priyadarshani S, Wang Y, Ye L, Shi C, Ye K, Zhou Q, Luo Z, Deng F. Assembly and comparative analysis of the complete mitochondrial genome of Suaeda glauca. BMC Genomics. 2021;22:1–15.
21. Khan AL, Asaf S, Lee I-J, Al-Harrasi A, Al-Rawahi A. First chloroplast genomics study of Phoenix dactylifera (var Naghal and Khanezi): A comparative analysis. PloS one. 2018;13(7):e0200104.
22. Singh NV, Patil PG, Sowjanya RP, Parashuram S, Natarajan P, Babu KD, Pal RK, Sharma J, Reddy UK. Chloroplast genome sequencing, comparative analysis, and discovery of unique cytoplasmic variants in Pomegranate (Punica granatum L.). Front Genet. 2021;12:704075.
23. Liu H, He J, Ding C, Lyu R, Pei L, Cheng J, Xie L. Comparative analysis of complete chloroplast genomes of Anemoclema, Anemone, Pulsatilla, and Hepatica revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). Front Plant Sci. 2018;9:1097.
24. Zhang Y, Iaffaldano BJ, Zhuang X, Cardina J, Cornish K. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. BMC Plant Biol. 2017;17:1–14.
25. Mavrodiev EV, Dervinis C, Whitten WM, Gitzendanner MA, Kirst M, Kim S, Kinser TJ, Soltis DE. A new, simple, highly scalable, and efficient protocol for genomic DNA extraction from diverse plant taxa. Appl Plant Sci. 2021;9(3):e11413.
26. Modi A, Vai S, Caramelli D, Lari M. The Illumina sequencing protocol and the NovaSeq 6000 system. In: Bacterial Pangenomics: Methods and Protocols. New York: Springer US; 2021. p. 15–42.
27. Lu H, Giordano F, Ning Z. Proteomics, bioinformatics: Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinform. 2016;14(5):265–79.

28. Lin B, Hui J, Mao H. Nanopore technology and its applications in gene sequencing. Biosensors (Basel). 2021;11(7):214.
29. Li X, Lin C-Y, Yang J-B, Yu W-B. De novo assembling a complete mitochondrial genome of Pedicularis rex (Orobanchaceae) using GetOrganelle toolkit. Mitochondrial DNA B Resour. 2020;5(1):1056–7.
30. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.
31. Mochizuki T, Sakamoto M, Tanizawa Y, Nakayama T, Tanifuji G, Kamikawa R, Nakamura Y. A practical assembly guideline for genomes with various levels of heterozygosity. Brief Bioinform. 2023;24(6):bbad337.
32. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6.
33. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Madden TL. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. BMC Bioinformatics. 2019;20:1–19.
34. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.
35. Houtgast EJ, Sima V-M, Bertels K, Al-Ars Z. Chemistry: Hardware acceleration of BWA-MEM genomic short read mapping for longer read lengths. Comput Biol Chem. 2018;75:54–64.
36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
37. 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.
38. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47(W1):W65–73.
39. Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 2023;23(3):694–704.
40. Jung J, Kim JI, Jeong Y-S, Yi G. AGORA: organellar genome annotation from the amino acid and nucleotide references. Bioinformatics. 2018;34(15):2661–3.
41. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. Apollo: democratizing genome annotation. PLoS Comput Biol. 2019;15(2):e1006790.
42. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
43. Nishikawa T, Vaughan DA, Kadowaki K-i. Genetics A: Phylogenetic analysis of Oryza species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. Theor Appl Genet. 2005;110:696–705.
44. Sperisen C, Büchler U, Gugerli F, Mátyás G, Geburek T, Vendramin GGG. Tandem repeats in plant mitochondrial genomes: application to the analysis of population differentiation in the conifer Norway spruce. Mol Ecol. 2001;10(1):257–63.
45. Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, Latif MA. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. Int J Mol Sci. 2013;14(11):22499–528.
46. Dong S, Chen L, Liu Y, Wang Y, Zhang S, Yang L, Lang X, Zhang S. The draft mitochondrial genome of Magnolia biondii and mitochondrial phylogenomics of angiosperms. PLoS One. 2020;15(4):e0231020.
47. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
48. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich SR. REPuter: the manifold applications of repeat analysis on a genomic. Nucleic Acids Res. 2001;29(22):4633–42.
49. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
50. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:1–5.
51. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:1–11.
52. Niu Y, Gao C, Liu J. Complete mitochondrial genomes of three Mangifera species, their genomic structure and gene transfer from chloroplast genomes. BMC Genomics. 2022;23(1):1–8.
53. Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes. Proc Natl Acad Sci U S A. 1999;96(24):13863–8.
54. Choi K-S, Park S. Complete plastid and mitochondrial genomes of Aeginetia indica reveal intracellular gene transfer (IGT), horizontal gene transfer (HGT), and cytoplasmic male sterility (CMS). Int J Mol Sci. 2021;22:6143.
55. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, Zhang L, Liu Y. The complete mitochondrial genome of the early flowering plant Nymphaea colorata is highly repetitive with low recombination. BMC Genomics. 2018;19(1):1–12.
56. Yang H, Chen H, Ni Y, Li J, Cai Y, Ma B, Yu J, Wang J, Liu C. De novo hybrid assembly of the salvia miltiorrhiza mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial DNA structure of salvia species. Int J Mol Sci. 2022;23(22):14267.
57. 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.
58. Fages-Lartaud M, Hundvin K, Hohmann-Marriott MF. Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of Chlamydomonas reinhardtii. Plant J. 2022;112(4):919–45.
59. Parvathy ST, Udayasuriyan V, Bhadana V. Codon usage bias. Mol Biol Rep. 2022;49(1):539–65.
60. Gray MW. RNA editing in plant mitochondria: 20 years later. IUBMB Life. 2009;61(12):1101–4.
61. He P, Huang S, Xiao G, Zhang Y, Yu J. Abundant RNA editing sites of chloroplast protein-coding genes in Ginkgo biloba and an evolutionary pattern analysis. BMC Plant Biol. 2016;16:1–12.
62. 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.
63. Kumar S, Stecher G, Tamura K. Evolution: MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
64. Edera AA, Small I, Milone DH, Sanchez-Puerta MV. Medicine: Deepred-Mt: Deep representation learning for predicting C-to-U RNA editing in plant mitochondria. Comput Biol Med. 2021;136:104682.
65. Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants APG. Botanical Journal of the Linnean Society 2009 161(2);105-121.
66. Katoh K, Standley DM. Evolution: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
67. von Haeseler A, Schmidt HA, Bui MQ, Nguyen LT. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32(1):268-74.
68. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–9.
69. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Jin H, Marler B, Guo H. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49–e49.
70. Zhang X, Gu C, Zhang T, Tong B, Zhang H, Wu Y, Yang C. Chloroplast (Cp) Transcriptome of P davidiana Dode× P bolleana Lauch provides insight into the Cp drought response and Populus Cp phylogeny. BMC Evol Biol. 2020;20:1–14.
71. Moon H, Kim S. A complete chloroplast genome sequence of Viola albida Palibin 1899 (Violaceae), a member of VIOLA ALBIDA complex. Mitochondrial DNA B Resou. 2023;8(6):673–7.
72. Xu N, Du X, Zhang X-X, Yang H-L. The complete chloroplast genome of Salix lindleyana (salicaceae), a plateau plant species. Mitochondrial DNA B Resour. 2023;8(8):877–81.
73. Dong Y, Du C, Yue J, Li W. The complete chloroplast genome and phylogenetic analysis of Acalypha hispida Burm. f.(Euphorbiaceae), an ornamental and medicinal plant. Mitochondrial DNA B Resour. 2023;8(11):1285–9.

74. Sheng W, Deng J, Wang C, Kuang Q. The garden asparagus (Asparagus officinalis L.) mitochondrial genome revealed rich sequence variation throughout whole sequencing data. Front Plant Sci. 2023;14:1140043.
75. Dai N, Li G, Ni J, Li F, Tong H, Liu Y. A novel galactoxylan derived from Viola diffusa alleviates LPS-induced acute lung injury via antagonizing P-selectin-mediated adhesion function. YJIJoBM: Int J Biol Macromol. 2023;242:124821.
76. 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.
77. Zhou S, Wei N, Jost M, Wanke S, Rees M, Liu Y, Zhou R. Evolution: The mitochondrial genome of the Holoparasitic plant Thonningia sanguinea Provides Insights into the Evolution of the Multichromosomal Structure. Genome Biol Evol. 2023;15(9):155.
78. Zhong Y, Yu R, Chen J, Liu Y, Zhou R. Highly active repeat-mediated recombination in the mitogenome of the holoparasitic plant Aeginetia indica. Front Plant Sci. 2022;13:988368.
79. 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.
80. Hong Z, Liao X, Ye Y, Zhang N, Yang Z, Zhu W, Gao W, Sharbrough J, Tembrock LR, Xu D. A complete mitochondrial genome for fragrant Chinese rosewood (Dalbergia odorifera, Fabaceae) with comparative analyses of genome structure and intergenomic sequence transfers. BMC Genomics. 2021;22:1–13.
81. Tang D, Huang S, Quan C, Huang Y, Miao J, Wei F. Mitochondrial genome characteristics and phylogenetic analysis of the medicinal and edible plant Mesona chinensis Benth. Front Genet. 2023;13:1056389.
82. Yang S, Huang J, Qu Y, Zhang D, Tan Y, Wen S, Song Y. Phylogenetic incongruence in an Asiatic species complex of the genus Caryodaphnopsis (Lauraceae). BMC Plant Bio. 2024;24(1):616.
83. Pfeifer M, Martis M, Asp T, Mayer KF, Lübberstedt T, Byrne S, Frei U, Studer B. The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. Plant Physiol. 2013;161(2):571–82.
84. Rodda M, Niissalo MA. Plastome evolution and organisation in the Hoya group (Apocynaceae). Sci Rep. 2021;11(1):14520.
85. Sun Z, Wu Y, Fan P, Guo D, Zhang S, Song C. Assembly and analysis of the mitochondrial genome of Prunella vulgaris. Front Plant Sci. 2023;14:1237822.
86. Lu G, Wang W, Mao J, Li Q. Complete mitogenome assembly of Selenicereus monacanthus revealed its molecular features, genome evolution, and phylogenetic implications. BMC Plant Biol. 2023;23(1):541.
87. Morais da Silva G, de Santana Lopes A, Gomes Pacheco T, Lima de Godoy Machado K, Silva MC, de Oliveira JD, de Baura VA, Balsanelli E, Maltempi de Souza E, de Oliveira Pedrosa F. Genetic and evolutionary analyses of plastomes of the subfamily Cactoideae (Cactaceae) indicate relaxed protein biosynthesis and tRNA import from cytosol. Braz J Bot. 2021;44:97–116.
88. Xue J-Y, Liu Y, Li L, Wang B, Qiu Y-L. The complete mitochondrial genome sequence of the hornwort Phaeoceros laevis: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. Curr Genet . 2010;56:53–61.
89. Miao Y, Chen H, Xu W, Liu C, Huang LJG. Cistanche species mitogenomes suggest diversity and complexity in lamiales-order mitogenomes. Genes (Basel). 2022;13(10):1791.
90. Hao Z, Zhang Z, Zhang J, Cui X, Li J, Luo L, Li Y. The complete mitochondrial genome of Aglaia odorata, insights into its genomic structure and RNA editing sites. Front Plant Sci. 2024;15:1362045.
91. Kovar L, Nageswara-Rao M, Ortega-Rodriguez S, Dugas DV, Straub S, Cronn R, Strickler SR, Hughes CE, Hanley KA, Rodriguez DN, et al. PacBio-based mitochondrial genome assembly of Leucaena trichandra (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. Genome Biol Evol. 2018;10(9):2501–17.
92. Maas S, Rich A. Changing genetic information through RNA editing. Bioessays. 2000;22(9):790–802.
93. Zhu L. Xian F-J, Zhang Q-N. Hu J. Research progress of RNA editing. aBIOTECH. 2022;38(1):1.
94. Jiang M, Ni Y, Li J, Liu C. Characterisation of the complete mitochondrial genome of Taraxacum mongolicum revealed five repeat-mediated recombinations. Plant Cell Rep. 2023;42(4):775–89.
95. Yang H, Liu L, Li J, Chen J, Du G. Rational design to improve protein thermostability: recent advances and prospects. ChemBioEng Rev. 2015;2(2):87–94.
96. Jaenicke R. Protein stability and molecular adaptation to extreme conditons. Eur J Biochem. 1991;202(3):715–28.
97. Dill KA. Dominant forces in protein folding. Biochemistry. 1990;29(31):7133–55.
98. Komar AA. The Yin and Yang of codon usage. Hum Mol Genet. 2016;25(R2):R77–85.
99. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics. 1991;129(3):897–907.

Publisher's Note