Assembly and analysis of the mitochondrial genome of *Prunella vulgaris*

Zhihao Sun[1], Ya Wu[1], Pengyu Fan[2], Dengli Guo[2], Sanyin Zhang[3] and Chi Song [1*]

[1]Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [2]Wuhan Benagen Technology Co., Ltd, Wuhan, Hubei, China, [3]Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China

Prunella vulgaris (Lamiaceae) is widely distributed in Eurasia. Former studies have demonstrated that *P. vulgaris* has a wide range of pharmacological effects. Nevertheless, no complete *P. vulgaris* mitochondrial genome has been reported, which limits further understanding of the biology of *P. vulgaris*. Here, we assembled the first complete mitochondrial genome of *P. vulgaris* using a hybrid assembly strategy based on sequencing data from both Nanopore and Illumina platforms. Then, the mitochondrial genome of *P. vulgaris* was analyzed comprehensively in terms of gene content, codon preference, intercellular gene transfer, phylogeny, and RNA editing. The mitochondrial genome of *P. vulgaris* has two circular structures. It has a total length of 297, 777 bp, a GC content of 43.92%, and 29 unique protein-coding genes (PCGs). There are 76 simple sequence repeats (SSRs) in the mitochondrial genome, of which tetrameric accounts for a large percentage (43.4%). A comparative analysis between the mitochondrial and chloroplast genomes revealed that 36 homologous fragments exist in them, with a total length of 28, 895 bp. The phylogenetic analysis showed that *P. vulgaris* belongs to the Lamiales family Lamiaceae and *P. vulgaris* is closely related to *Salvia miltiorrhiza*. In addition, the mitochondrial genome sequences of seven species of Lamiaceae are unconservative in their alignments and undergo frequent genome reorganization. This work reports for the first time the complete mitochondrial genome of *P. vulgaris*, which provides useful genetic information for further *Prunella* studies.

KEYWORDS

Prunella vulgaris, mitochondrial genome, codon usage, repeated sequence, evolution

1 Introduction

P. vulgaris is a low-growing herbaceous perennial widely distributed in Eurasia's temperate and tropical mountainous regions. The mature spikes of *P. vulgaris* are cylindrical and slightly flat. Its stems are relatively short. The panicle consists of several whorls of persistent calyxes and bracts, ranging from a few to ten. In southeastern China,

the fresh leaves of *P. vulgaris* are served as a vegetable (Bai et al., 2016). The spikes of dried fruits of *P. vulgaris* are considered to have anti-inflammatory action in traditional Chinese medicine (Li et al., 2015; Liu et al., 2020). Modern pharmacological studies have revealed the presence of natural compounds with anti-inflammatory, antibacterial, antioxidant, and immunomodulatory properties in *P. vulgaris*, e.g., triterpenes, phenolic acid, and oleanolic acid (Bai et al., 2016). These compounds are important for pharmaceutical research (Su et al., 2022). The study of *P. vulgaris* organelle genome is essential to better exploit its medicinal and economic value. Important plastid organelles, i.e., mitochondria and chloroplast, play key roles in plant development and reproduction, and their contributions to energy metabolism and material conversion depend on their own semi-autonomous genetic systems (Nielsen et al., 2010; Gualberto et al., 2014). Besides, compared with whole genome assembly, the assembly of plastid genome is more cost-effective and could provide useful information for the evolutionary analysis of the focal species. The complete chloroplast genome of *P. vulgaris* has already been assembled (Han and Zheng, 2018). The present study reveals the complete mitochondrial genome of *P. vulgaris*, providing the necessary genetic sequence for further phylogeny and resource utilization.

As an important component of most eukaryotic cells, mitochondria have energy conversion, biosynthetic, and signaling functions. Mitochondria can encode some proteins semi-autonomously, but these processes are regulated by nuclear-encoded genes (Mackenzie and McIntosh, 1999). Thus, abnormal mitochondrial gene expression in *Brassica napus* may lead to male sterility(Liu et al., 2017). The characteristics of plant mitochondrial genomes include the existence of highly conserved genes, a large number of genomic structural rearrangements, a wide range of non-coding sequences, and extensive RNA editing (Silvestris et al., 2020). Mitochondrial genomes usually exhibit matrilineal inheritance, which provides useful information about evolution and phylogeny of the focal species (Birky, 2001). For example, the mitochondrial genome sequence of *Brassica oleracea* facilitated the evolutionary analysis of this species(Shao et al., 2021). The structure of plant mitochondrial genomes may be linear or multi-branched (Wang et al., 2019; Jackman et al., 2020). The reasons for the structural diversity of plant mitochondrial genomes are still unclear. The transfer of DNA between the mitochondrial genome and the chloroplast genome is a common event in the plant genome. Some studies believe that this event usually leads to changes in the length of mitochondrial genome and changes the structure of mitochondrial genome (Allen, 2015; Turmel et al., 2016). With the rapid development of genome assembly and sequencing technologies, complete organelle genomes of plants are able to be assembled. This will facilitate our understanding of plants, for example, by detecting nucleotide fragments of gene insertions or deletions at the same position in different mitochondrial genomes to distinguish species (Chen et al., 2022).

In this study, we assembled the first complete *P. vulgaris* mitochondrial genome using a hybrid assembly strategy based on sequencing data from Illumina and Nanopore. The assembled mitochondrial genome was annotated from Illumina and Nanopore platforms. The characteristic information of the

mitochondrial genome of *P. vulgaris* was discussed in terms of codon usage preference, genome repeat sequence, and genes transfer between the mitochondrial genome and chloroplast genome. Phylogenetic tree and synteny analysis provide hints about the evolutionary history of *P. vulgaris*. The results of this study will provide useful information for the mitochondrial genome of *P. vulgaris*.

2 Materials and methods

2.1 *P. vulgaris* DNA extraction and mitochondrial genome assembly

The *P. vulgaris* plants were collected from wild in Hubei province, China, and cultured in Wuhan, China (31°68' N, 118° 45' E). High quality genomic DNA were isolated from fresh leaves using the standard CTAB method (Arseneau et al., 2017; Cheng et al., 2021). Illumina and Nanopore platforms were used for sequencing. Illumina sequencing and Oxford sequencing were performed by Wuhan Benagen Tech Solutions Company (http://en.benagen.com/). Illumina sequencing data was sequenced using the HiSeq Xten PE150 Illumina, San Diego, CA, USA sequencing platform and Nanopore sequencing was performed by Oxford Nanopore GridION × 5 Oxford Nanopore Technologies, Oxford, UK. GetOrganelle (v1.7.5) (Jin et al., 2020) was used to perform plant mitochondrial genome assembly (default parameters) and a graphical plant mitochondrial genome was obtained. Since the graphical genome generated by GetOrganelle comprised multiple nodes, with redundant fragments existing in the border of two neighbor nodes, Bandage (Wick et al., 2015) was used to visualize the graphical genome and Nanopore data was mapped to help manually check these redundant fragments. BWA (0.7.17) is used to map the third generation sequencing data to the graphical genome, followed by manually identification and removing of the redundant fragments (Li and Durbin, 2009).

2.2 Annotation of the mitochondrial genome of *P. vulgaris*

The *P. vulgaris* mitochondrial genomes were annotated using Geseq (Tillich et al., 2017) with Arabidopsis thaliana (Sloan et al., 2018) and Liriodendron tulipifera (Richardson et al., 2013) as reference genomes. The *P. vulgaris* mitochondrial genomes were annotated using Geseq (Tillich et al., 2017). The tRNA genes were annotated using the tRNAscan-SE (Lowe and Eddy, 1997). The rRNA genes were annotated using BLASTN (Chen et al., 2015). Each mitochondrial genome annotation error was manually corrected using Apollo (Lewis et al., 2002).

2.3 Relative synonymous codon usage

The protein coding sequences of the genome were extracted using Phylosuite (Zhang et al., 2020). Codon preferences of protein-

coding genes in the mitochondrial genome were analyzed by Mega (v7.0). The results of the analysis are expressed in relative synonymous codon usage (RSCU).

2.4 Analysis of repeated sequences

The online version of MISA (Beier et al., 2017) was used to analyze Simple Sequence Repeat (SSR) in assembled mitochondrial genome. To define the SSR locus, we searched that SSRs with a length of 1, 2, 3, 4, 5, and 6 bases have at least 10, 5, 4, 3, 3, and 3 repeats, respectively. The online version of TRF (Benson, 1999) was used to identify tandem repeat sequences in the mitochondrial genome, with alignment parameters: match = 2, mismatch = 5, and indels = 7. REPuter web server (Kurtz et al., 2001) was used to identify dispersed repeats with the following options: maximum computed repeats = 5000, hamming distance = 3, and minimal repeat size = 30.

2.5 Homologous fragments between chloroplast genome and mitochondrial genome

The chloroplast genome of *P. vulgaris* was reassembled by GetOrganelle (Jin et al., 2020) based on the Illumina and Nanopore sequencing data and then annotated by CPGAVAS2 (Shi et al., 2019). The assembled chloroplast genome size is 151, 346 bp (Figure S1). Identification of homologous fragments between chloroplast genome and mitochondria genome of *P. vulgaris* using the BLASTN online tool on the NCBI website (https://www.ncbi.nlm.nih.gov/), with the default parameters. Results with an identity value greater than 75 were retained (Table S1). The results were visualized using the RCircos package (Zhang et al., 2013).

2.6 Phylogenetic tree construction and synteny analysis

A phylogenetic tree was constructed for 22 species (Table S2) from 6 families of Lamiales based on the DNA sequences of 16 conserved mitochondrial PCGs (*atp1, atp4, ccmB, ccmC, ccmFC, ccmFN, cob, cox2, cox3, matR, nad1, nad2, nad3, nad5, nad6, rps13*). PhyloSuite (Zhang et al., 2020) and MAFFT(Katoh and Standley, 2013) were used to extract shared genes and align multiple sequences. IQ-TREE (Nguyen et al., 2015) was used to build the phylogenetic tree. ModelFinder (Kalyaanamoorthy et al., 2017; Zhang et al., 2020) was used to find the most suitable model from for out data based on Akaike information criterion (AIC). The 'TVM+F+I+G4' model was finally chosen for maximum likelihood tree construction. iTOL (https://itol.embl.de/) (Letunic and Bork, 2019) was used to visualize the results of phylogenetic analysis. The mitochondrial genomes of seven Lamiaceae species (*Scutellaria barbata, Pogostemon heyneanus, Salvia miltiorrhiza, P. vulgaris, Ajuga reptans, Rotheca serrata*, and *Vitex trifolia*) were compared

using the BLAST program. Then homologous sequences longer than 500 bp were retained as conserved collinearity blocks.

2.7 RNA editing event analysis methods

Prediction of RNA editing events was performed by the online version of PREPACT3 (http://www.prepact.de/) (Lenz et al., 2018). RNA editing sites of a total of 29 unique PCGs were predicted with a cutoff value of 0.001.

3 Results

3.1 Genomic features of the *P. vulgaris* mitochondrial genome

In this study, the *P. vulgaris* mitochondrial genome was assembled. It is composed of two circular structures (Figure 1A). We used Bandage (Wick et al., 2015) to visualize the mitochondrial genome assembled based on Illumina data. Duplicated regions were removed with the help of the third-generation sequencing reads. Manual method was used to remove the nodes formed by nuclear and chloroplast genes. The assembled raw mitochondrial genome contains 47 nodes that including predicted duplication regions and mitochondrial genomic regions migrating from chloroplast (Figure 1B). After manually removing of redundant fragments, two clear circular contigs were obtained (Figure 1C). The total length of the assembled genome was 297, 777 bp and the GC content was 43.92% (Table 1). The mitochondrial genome of *P. vulgaris* was annotated with 29 unique PCGs (Figure 1A), 13 tRNA genes and 3 rRNA genes (Table 2). The unique PCGs include 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 ubiquinol cytochrome c reductase genes (*ccmB, ccmC, ccmFC* and *ccmFN*), three cytochrome c oxidase genes (*cox1, cox2* and *cox3*), one transport membrane protein gene (*mttB*), one maturases gene (*matR*), one cytochrome c biogenesis gene (*cob*), one large subunit of ribosome gene (*rpl16*), three small subunit of ribosome genes (*rps3, rps12* and *rps13*), and one succinate dehydrogenase gene (*sdh4*).

3.2 Codon usage analysis of PCGs

Codon preference analysis was performed on 29 unique PCGs of *P. vulgaris* mitochondria. The codon usage by individual amino acids is shown in Table S3. Relative synonymous codon usage (RSCU) value greater than 1 indicated that the corresponding amino acid was preferentially used. As shown in Figure 2, the Methionine (Met) codon AUG and Tryptophan (Try) code UGG, which both have the RSCU value of 1. There is also a general preference for codon use in the PCGs of the mitochondria. For example, Alanine (Ala) has a high preference for GCU with the highest RSCU value of 1.61 among mitochondrial PCGs, followed by Leucine (Leu) with a usage preference for UUA. Notably, the

FIGURE 1
The mitochondrial genome structure and annotation of *P. vulgaris*. **(A)** Annotations of *P. vulgaris* mitochondrial genome. **(B)** Two circular contigs of *P. vulgaris* mitochondrial genome predicted by GetOrganelle. **(C)** The 2D structure of *P. vulgaris* mitochondrial genome after removing artificial chloroplast and nuclear gene fragments. In B and C, the red nodes represent the predicted duplication regions and the green nodes represent the predicted segments migrating to the mitochondrial genomes from chloroplast.

maximum RSCU values of Lysine (Lys) and Phenylalanine (Phe) were less than 1.2 and did not have a strong codon usage preference.

3.3 Repeated sequence analysis of the mitochondrial genome of *P. vulgaris*

In MISA online prediction for the chromosome 1, a hit was retained as a SSR when it met two criteria: the match score should be greater than 69% and length should be between 10 and 33 bp. A total of 47 SSRs were found on mitochondrial chromosome 1, with monomeric and dimeric forms accounting for 31.91% of the total SSRs. Adenine and thymine monomeric repeats accounted for 85.71% (6) of the 7 monomeric SSRs. There is a hexameric SSRs in chromosome 1 (Figure 3A). Tandem repeats are widely found in eukaryotic and prokaryotes genomes. Mitochondrial chromosome 1 contains 15 tandem repeats. Repetitive dispersed sequences in mitochondrial chromosome 1 were examined. A total of 57 pairs

TABLE 1   Information on the mitochondrial genome of *P. vulgaris*.

Contigs	Type	Length	GC content
Chromosome 1-2	Branched	297, 777 bp	43.92%
Chromosome 1	Circular	183, 505 bp	44.27%
Chromosome 2	Circular	114, 272 bp	43.36%

TABLE 2 The encoding genes of *P. vulgaris* mitochondrial genome.

Group of genes	Name of genes
ATP synthase	*atp1, atp4, atp6, atp8, atp9*
NADH dehydrogenase	*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9*
Cytochrome c biogenesis	*cob*
Ubiquinol cytochrome c reductase	*ccmB, ccmC, ccmFC, ccmFN*
Cytochrome c oxidase	*cox1, cox2, cox3*
Maturases	*matR*
Transport membrane protein	*mttB*
Large subunit of ribosome	*rpl16*
Small subunit of ribosome	*rps3, rps12, rps13*
Succinate dehydrogenase	*sdh4*
Ribosome RNA	*rrn5, rrn18, rrn26*
Transfer RNA	*trnC-GCA, trnD-GUC, trnE-UUC, trnfM-CAU, trnH-GUG, trnI-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnS-GGA, trnW-CCA, trnY-GUA*

of repetitive sequences with lengths greater than or equal to 30 bp were observed (Figure 3B). Among them, 28 pairs of palindromic repeats and 29 pairs of forward repeats were detected. The length of the longest palindromic repeat was 1, 392 bp and that of the longest forward repeat was 1, 429 bp.

In MISA online prediction for the chromosome 2, a hit was retained as a SSR when it met two criteria: the match score should be greater than 71% and length should be between 9 and 33 bp. A total of 29 SSRs were found in mitochondrial chromosome 2, with

monomeric and dimeric forms accounting for 37.93% of total SSRs. Adenine and thymine monomeric repeats accounted for 66.67% of the three monomeric SSRs. Mitochondrial chromosome 2 contains nine tandem repeats. A total of 30 pairs of repetitive sequences with lengths greater than or equal to 30 bp were observed (Figure 3B). Among them, 11 pairs of palindromic repeats and 19 pairs of forward repeats were detected. The length of the longest palindromic repeat was 54 bp and that of longest forward repeat was 51 bp.
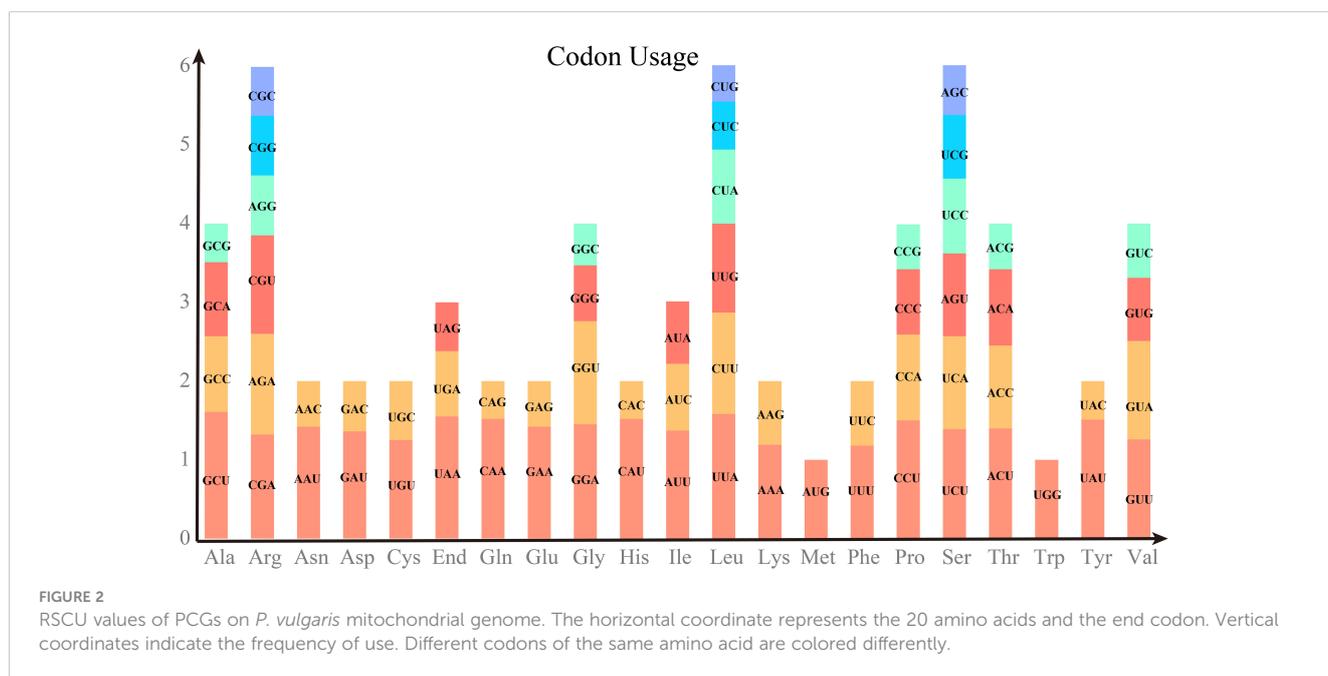
FIGURE 2
RSCU values of PCGs on *P. vulgaris* mitochondrial genome. The horizontal coordinate represents the 20 amino acids and the end codon. Vertical coordinates indicate the frequency of use. Different codons of the same amino acid are colored differently.

FIGURE 3
Horizontal coordinate indicates mitochondrial molecules and vertical coordinate indicates the number of repeat fragments. **(A)** Simple Sequence Repeat of *P. vulgaris* mitochondrial genome. **(B)** Repeated sequence of *P. vulgaris* mitochondrial genome.
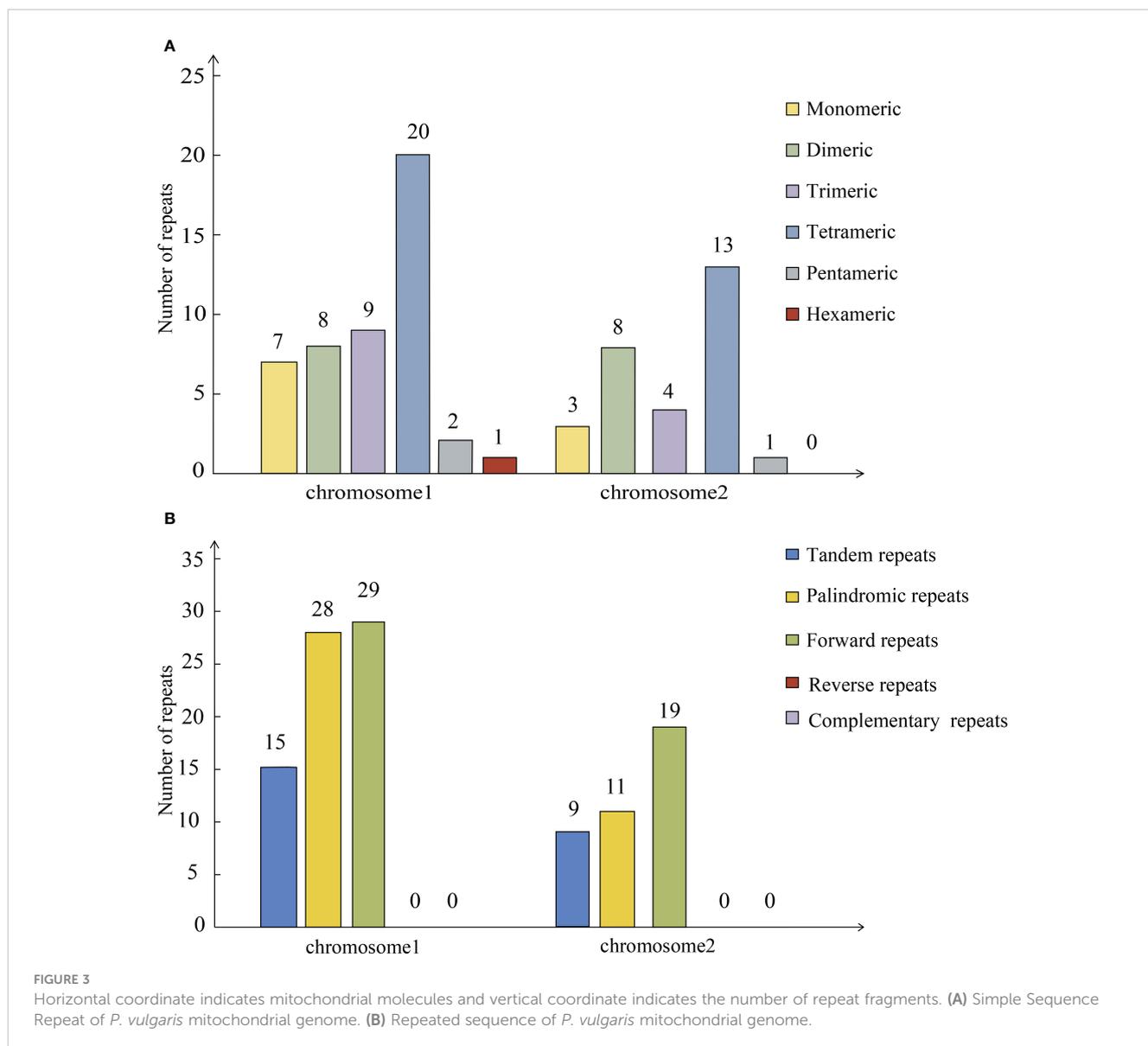
3.4 DNA migration from chloroplast to mitochondria

Based on the analysis of sequence similarity, a total of 36 fragments were homologous between the mitochondrial and chloroplast genomes (Figure 4). The total length of these homologous fragments was 28, 895 bp, accounting for 9.70% of the total length of the mitochondrial genome. The longest fragments were fragment 19 and fragment 20, both of which were 3, 276 bp (Table S1). By annotating these homologous sequences, 16 complete genes were identified on 36 homologous fragments (Table S1), including 10 PCGs (*ndhB, ndhI, psbJ, psbL, psbF, psbE, petL, petG, rps4,* and *ycf15*) and six tRNA genes (*trnD-GUC, trnH-GUG, trnM-CAU, trnP-UGG, trnS-GGA,* and *trnW-CCA*).

3.5 Phylogenetic analysis and synteny analysis based on mitochondrial genomes of higher plants

A phylogenetic analysis was performed with 22 species based on the DNA sequences of 16 conserved mitochondrial PCGs (*atp1, atp4, ccmB, ccmC, ccmFC, ccmFN, cob, cox2, cox3, matR, nad1, nad2, nad3, nad5, nad6,* and *rps13*).The two mitochondrial genomes of Oleaceae were set as outgroups (Figure 5A). The results showed that *P. vulgaris* belongs to the Lamiales family Lamiaceae and is closely related to *Salvia miltiorrhiza*. The topology of this mitochondrial DNA-based phylogeny is consistent with the Angiosperm Phylogeny Group IV (Bennett and Alarcón, 2015).
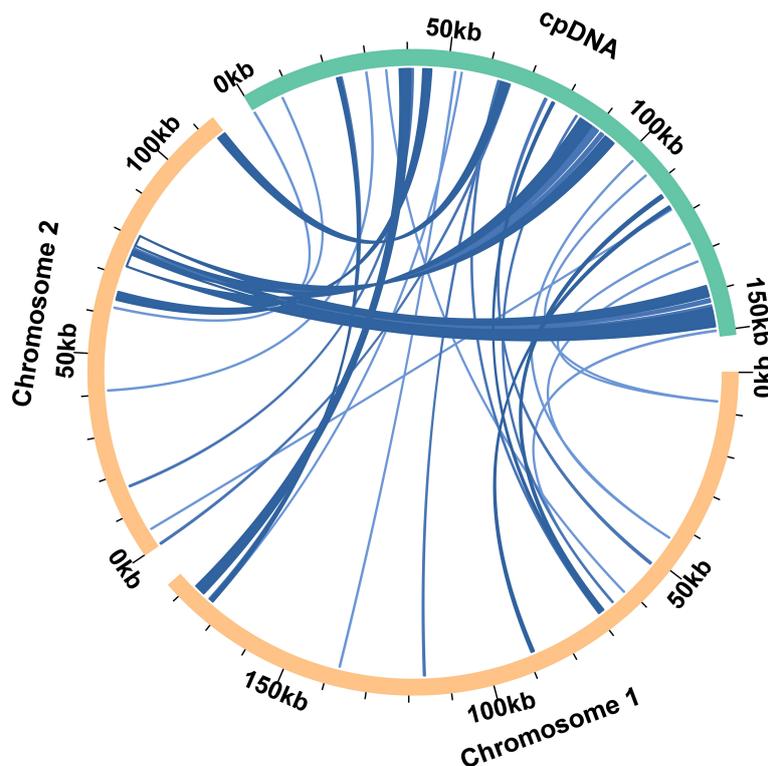
FIGURE 4
The brown arc in the figure represents the mitochondrial genome, the green arc represents the chloroplast genome. The genome fragments corresponding to the blue connecting lines between arcs are homologous fragments.

A large number of homologous collinearity blocks were detected in Lamiaceae species (Figure 5B). No collinearity blocks with lengths less than 0.5 kb were retained. In addition, some regions were unique to *P. vulgaris*, i.e., have no homologous region with any other species. The collinearity blocks are not in the same order. The mitochondrial genome sequences of these seven Lamiaceae species are not sequentially conserved and have undergone frequent genomic rearrangements.

3.6 The prediction of RNA editing events

A total of 379 potential RNA editing sites were identified on 29 mitochondrial PCGs (Figure 6), which are dominantly base C to U editing (Table S1). Both *ccmB* and *mttB* had the highest number of edits among all mitochondrial genes (35 RNA editing sites identified), followed by *ccmFN* with 31 RNA editing events. In addition, *rpl16* and *rps3* had the lowest number of edits among all mitochondrial genes (one RNA editing sites identified).

4 Discussion

In general, plants contain three genomes, the nuclear genome, the plastid genome and the mitochondrial genome. Recent advances in plant whole genome study such as genome assembly or single cell sequencing have greatly facilitated medicinal plant research; however,

plastid genome remains to be a powerful and cost-effective way(Guo et al., 2022; Chen et al., 2023; Sun et al., 2023). The *P. vulgaris* chloroplast genome has been sequenced (Han and Zheng, 2018), but no mitochondrial sequencing has been completed for this species. In this study, the *P. vulgaris* mitochondrial genome was assembled into two circular structures with a total length of 297, 777 bp. Due to the simplicity of codons, each amino acid corresponds to at least 1 codon, and there are up to six corresponding codons. The codons of one amino acid are often used at different frequencies. Codons may correlate with gene expression levels (Trotta, 2013; Hia and Takeuchi, 2021). The use of genetic codons varies greatly from species to species, which provides additional information on species-specific evolution. Gene expression levels and gene length, tRNA abundance and interactions, and codon position in the gene are some of the factors that influence codon preference. There is a clear codon usage preference in *P. vulgaris* mitochondria, with differences in the frequency of different codons of each amino acid being used, except for Met and Trp. For example, AAU and AAC are synonymous codons of Asn, in which AAU was used 71% and AAC was used 29%. Codon usage preference has been used to study phylogeny and molecular evolution of genes among organisms. By studying codon preference in *Brassica campestris* was found that selection pressure plays most of the role in mutational pressure(Paul et al., 2018; Parvathy et al., 2022). In addition, codon usage preference should be considered when designing high yield and resistance genes.

Tandem repeat sequences are one of the most prevalent features of genomic sequences. Tandem repeat sequences have important
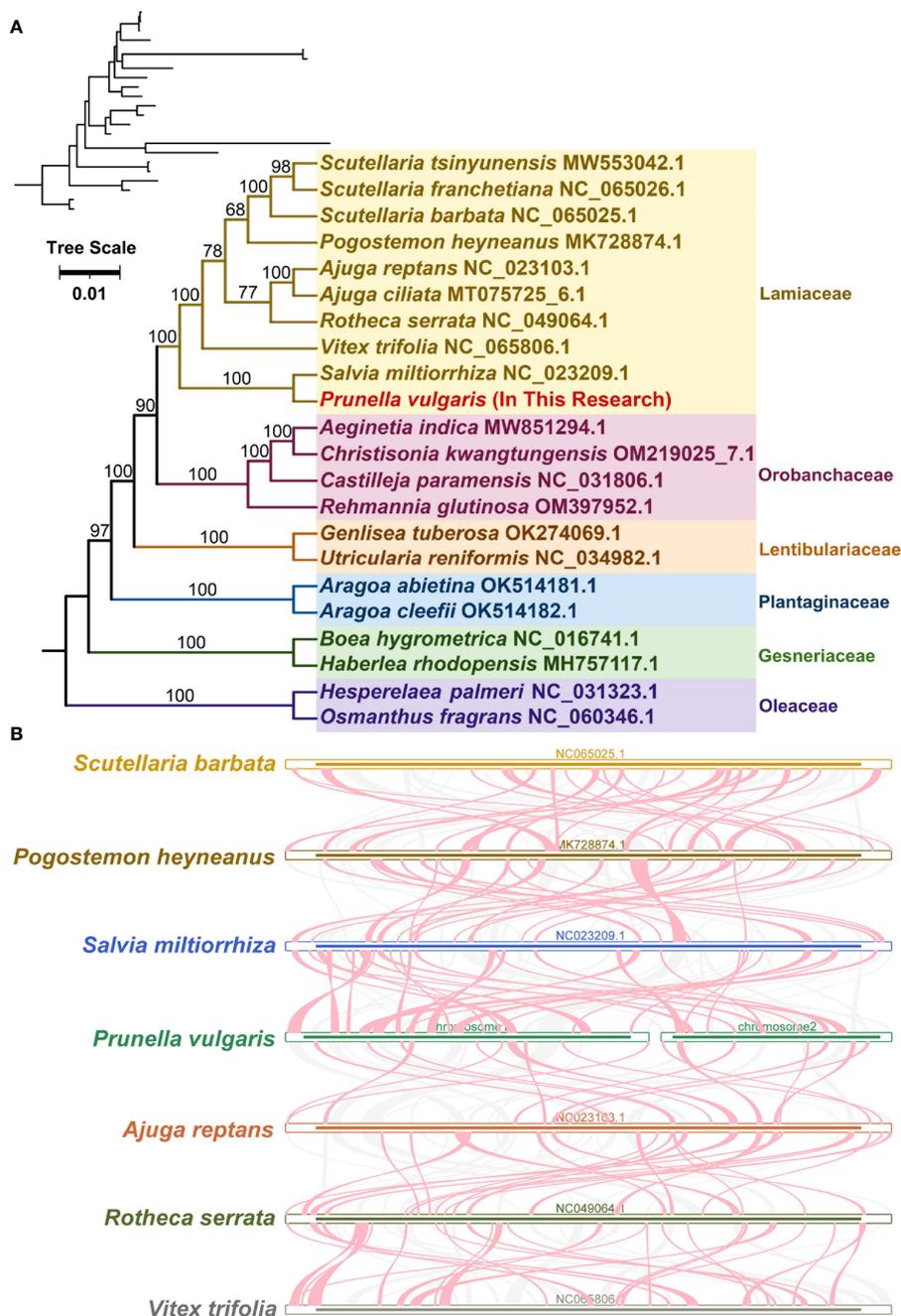
FIGURE 5
Evolution analysis of *P. vulgaris*. **(A)** The plants in the diagram belong to of Lamiales. Different families are represented by different colors, with *P. vulgaris* represented in red. **(B)** Red-curved regions indicate where inversions occur, gray regions indicate regions of good homology, and white regions indicate species-unique sequences.

roles in biological evolution, gene regulation, gene expression, and genome stability. A total of 58 SSRs were found in the mitochondria of *P. vulgaris*, which provided great convenience for genetic studies and species identification due to the maternal inheritance characteristics of mitochondria. SSRs have been used to classify different species, which is beneficial for species identification and breeding of superior varieties. Among tandem repeats, SSR is a special kind of tandem repeat sequence, which generally less than 6 bp. SSRs are often used for molecular marker of development due to

their characteristics such as dominant inheritance. They are independent of the external environment and growth conditions (Song et al., 2015). In addition, these markers have advantages such as large numbers, stable traits, simple operation, and rapid detection, and are widely used in the analysis and identification of herbal plants (Chen et al., 2022). The mitochondrial genomes of plants and animals have formed different evolutionary features.

In general, the mutation rate of plant mitochondrial genome is lower than that of animal mitochondrial genome (Darracq et al.,
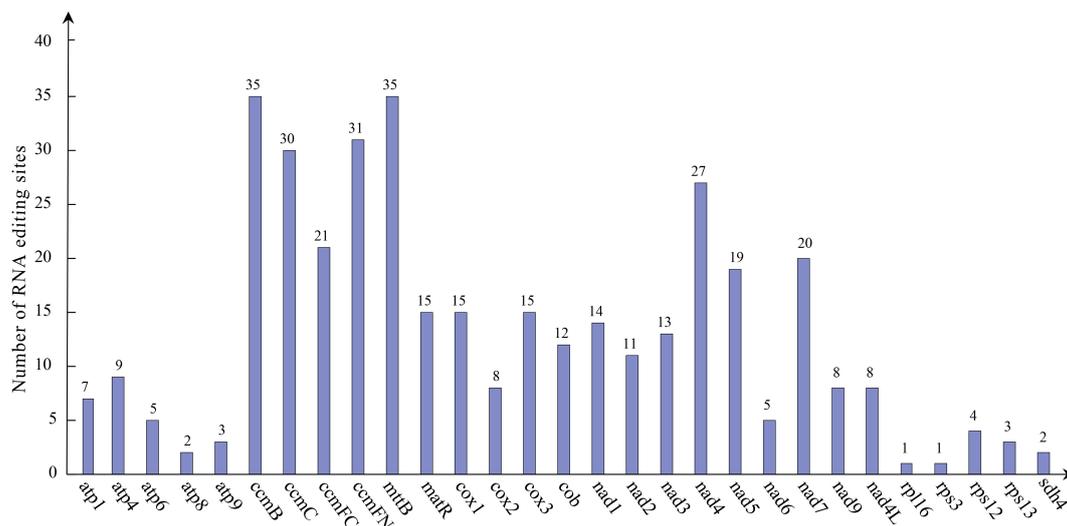
FIGURE 6
Predicted RNA editing events in *P. vulgaris* mitochondrial genes. Horizontal coordinate indicates represents different genes and vertical coordinate indicates the predicted number of RNA editing events.

2011). Plant mitochondrial genome can integrate exogenous DNA by migrating fragments with chloroplast genome (Law et al., 2022). The presence of fragments of chloroplast genes was found in the assembled mitochondrial genome of *P. vulgaris*. Mitochondrial and chloroplast gene migration is an important mechanism for biological evolution and diversity formation, which is important for evolution, adaptation, and diversity of organisms (Xiong et al., 2008). And this results in many structural variations in the plant mitochondrial genome.

Phylogenetic relationships among species are the basis for many biological studies. An accurate phylogenetic tree supports our understanding of key transitions in evolution (Kapli et al., 2020). Based on the phylogenetic tree constructed from 16 genes of mitochondria, *P. vulgaris* was more closely related to *Salvia miltiorrhiza* than other 20 species in this study. The tree matches the latest classification of the Angiosperm Phylogeny Group IV (Bennett and Alarcón, 2015). Collinearity research is a method to analyze the relationship between homologous genes or sequences. The collinearity of genes in plant genome usually decreases with the increase of evolutionary distance (Wicker et al., 2010). A large number of homologous collinearity blocks were detected in the *P. vulgaris* with the rest of Lamiales species, but these collinearity blocks were short in length. In addition, some blank regions were found. These sequences are unique to the species and have no homology with the rest of the species. The collinearity blocks were not in the same order among the mitochondrial genomes of Lamiaceae. The results indicate that the mitochondrial genome sequences of these seven Lamiaceae species are not conservative in their alignments and undergo frequent reorganization. RNA editing is the phenomenon of base insertion, deletion, or conversion that occurs in the coding region of post-transcriptional RNA, such as 441 C-to-U editing sites have been identified in *Arabidopsis thaliana* and 225 C-to-U editing sites in *Salvia miltiorrhiza*. Due to the lack of suitable transcriptome data, *P. vulgaris*

was predicted through the website that it has 379 RNA editing sites, all of which are C-to-U (Yang et al., 2022).

5 Conclusion

In this study, we have assembled the first complete mitochondrial genome of *P. vulgaris*. It has a total length of 297, 777 bp, a GC content of 43.92%, and 29 unique PCGs. We found 76 SSRs in the mitochondrial genome. The phylogenetic analysis showed that *P. vulgaris* is closely related to *Salvia miltiorrhiza*, consistent with the Angiosperm Phylogeny Group IV. The complete mitochondrial genome of *P. vulgaris* is useful to understanding Lamiales evolution and could benefit following works such as breeding of varieties of *P. vulgaris*.

Data availability statement

The original mitochondrial genome presented in the study are publicly available. This data can be found in NCBI (https://www.ncbi.nlm.nih.gov/) under the GenBank: OR113011.1 (https://www.ncbi.nlm.nih.gov/nuccore/OR113011.1/). The data are publicly available. The datasets presented in this study can be found in NCBI. The names of the repositories and accession numbers can be found in the Supplementary Material.

Author contributions

CS conceived the study. YW and SZ collected the data. PF and DG analyzed the data. ZS wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

Author PF and DG are employed by Wuhan Benagen Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1237822/full#supplementary-material

References

Allen, J. F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 112 (33), 10231–10238. doi: 10.1073/pnas.1500012112

Arseneau, J. R., Steeves, R., and Laflamme, M. (2017). Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Mol. Ecol. Resour* 17 (4), 686–693. doi: 10.1111/1755-0998.12616

Bai, Y., Xia, B., Xie, W., Zhou, Y., Xie, J., Li, H., et al. (2016). Phytochemistry and pharmacological activities of the genus Prunella. *Food Chem.* 204, 483–496. doi: 10.1016/j.foodchem.2016.02.047

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585. doi: 10.1093/bioinformatics/btx198

Bennett, B. C., and Alarcón, R. (2015). Hunting and hallucinogens: The use psychoactive and other plants to improve the hunting ability of dogs. *J. Ethnopharmacol* 171, 171–183. doi: 10.1016/j.jep.2015.05.035

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Birky, C. W.Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148. doi: 10.1146/annurev.genet.35.102401.090231

Chen, H., Guo, M., Dong, S., Wu, X., Zhang, G., He, L., et al. (2023). A chromosome-scale genome assembly of Artemisia argyi reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Commun.* 4 (3), 100516. doi: 10.1016/j.xplc.2023.100516

Chen, S., Li, Z., Zhang, S., Zhou, Y., Xiao, X., Cui, P., et al. (2022). Emerging biotechnology applications in natural product and synthetic pharmaceutical analyses. *Acta Pharm. Sin. B* 12 (11), 4075–4097. doi: 10.1016/j.apsb.2022.08.025

Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43 (16), 7762–7768. doi: 10.1093/nar/gkv784

Cheng, Y., He, X., Priyadarshani, S., Wang, Y., Ye, L., Shi, C., et al. (2021). Assembly and comparative analysis of the complete mitochondrial genome of Suaeda glauca. *BMC Genomics* 22 (1), 167. doi: 10.1186/s12864-021-07490-9

Darracq, A., Varré, J. S., Maréchal-Drouard, L., Courseaux, A., Castric, V., Saumitou-Laprade, P., et al. (2011). Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. *Genome Biol. Evol.* 3, 723–736. doi: 10.1093/gbe/evr042

Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100, 107–120. doi: 10.1016/j.biochi.2013.09.016

Guo, M., Pang, X., Xu, Y., Jiang, W., Liao, B., Yu, J., et al. (2022). Plastid genome data provide new insights into the phylogeny and evolution of the genus Epimedium. *J. Adv. Res.* 36, 175–185. doi: 10.1016/j.jare.2021.06.020

Han, Y. W., and Zheng, T. Y. (2018). The complete chloroplast genome of the common self-heal, Prunella vulgaris (Lamiaceae). *Mitochondrial DNA B Resour* 3 (1), 125–126. doi: 10.1080/23802359.2018.1424587

Hia, F., and Takeuchi, O. (2021). The effects of codon bias and optimality on mRNA and protein regulation. *Cell Mol. Life Sci.* 78 (5), 1909–1928. doi: 10.1007/s00018-020-03685-7

Jackman, S. D., Coombe, L., Warren, R. L., Kirk, H., Trinh, E., MacLeod, T., et al. (2020). Complete Mitochondrial Genome of a Gymnosperm, Sitka Spruce (Picea sitchensis), Indicates a Complex Physical Structure. *Genome Biol. Evol.* 12 (7), 1174–1179. doi: 10.1093/gbe/evaa108

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21 (1), 241. doi: 10.1186/s13059-020-02154-5

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589. doi: 10.1038/nmeth.4285

Kapli, P., Yang, Z., and Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* 21 (7), 428–444. doi: 10.1038/s41576-020-0233-0

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29 (22), 4633–4642. doi: 10.1093/nar/29.22.4633

Law, S. S. Y., Liou, G., Nagai, Y., Giménez-Dejoz, J., Tateishi, A., Tsuchiya, K., et al. (2022). Polymer-coated carbon nanotube hybrids with functional peptides for gene delivery into plant mitochondria. *Nat. Commun* 13 (1)2417. doi: 10.1038/s41467-022-30185-y

Lenz, H., Hein, A., and Knoop, V. (2018). Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinf.* 19 (1), 255. doi: 10.1186/s12859-018-2244-9

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–w259. doi: 10.1093/nar/gkz239

Lewis, S. E., Searle, S. M., Harris, N., Gibson, M., Iyer, V., Richter, J., et al. (2002). Apollo: a sequence annotation editor. *Genome Biol.* 3 (12), Research0082. doi: 10.1186/gb-2002-3-12-research0082

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, C., Huang, Q., Fu, X., Yue, X. J., Liu, R. H., and You, L. J. (2015). Characterization, antioxidant and immunomodulatory activities of polysaccharides from Prunella vulgaris Linn. *Int. J. Biol. Macromol* 75, 298–305. doi: 10.1016/j.ijbiomac.2015.01.010

Liu, Z., Dong, F., Wang, X., Wang, T., Su, R., Hong, D., et al. (2017). A pentatricopeptide repeat protein restores nap cytoplasmic male sterility in Brassica napus. *J. Exp. Bot.* 68 (15), 4115–4123. doi: 10.1093/jxb/erx239

Liu, Z., Hua, Y., Wang, S., Liu, X., Zou, L., Chen, C., et al. (2020). Analysis of the Prunellae Spica transcriptome under salt stress. *Plant Physiol. Biochem.* 156, 314–322. doi: 10.1016/j.plaphy.2020.09.023

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi: 10.1093/nar/25.5.955

Mackenzie, S., and McIntosh, L. (1999). Higher plant mitochondria. *Plant Cell* 11 (4), 571–586. doi: 10.1105/tpc.11.4.571

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi: 10.1093/molbev/msu300

Nielsen, B. L., Cupp, J. D., and Brammer, J. (2010). Mechanisms for maintenance, replication, and repair of the chloroplast genome in plants. *J. Exp. Bot.* 61 (10), 2535–2537. doi: 10.1093/jxb/erq163

Parvathy, S. T., Udayasuriyan, V., and Bhadana, V. (2022). Codon usage bias. *Mol. Biol. Rep.* 49 (1), 539–565. doi: 10.1007/s11033-021-06749-4

Paul, P., Malakar, A. K., and Chakraborty, S. (2018). Compositional bias coupled with selection and mutation pressure drives codon usage in Brassica campestris genes. *Food Sci. Biotechnol.* 27 (3), 725–733. doi: 10.1007/s10068-017-0285-x

Richardson, A. O., Rice, D. W., Young, G. J., Alverson, A. J., and Palmer, J. D. (2013). The "fossilized" mitochondrial genome of Liriodendron tulipifera: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 11, 29. doi: 10.1186/1741-7007-11-29

Shao, D., Ma, Y., Li, X., Ga, S., and Ren, Y. (2021). The sequence structure and phylogenetic analysis by complete mitochondrial genome of kohlrabi (Brassica oleracea var. gongylodes L.). *Mitochondrial DNA B Resour* 6 (9), 2714–2716. doi: 10.1080/23802359.2021.1966341

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47 (W1), W65–w73. doi: 10.1093/nar/gkz345

Silvestris, D. A., Scopa, C., Hanchi, S., Locatelli, F., and Gallo, A. (2020). *De Novo* A-to-I RNA Editing Discovery in lncRNA. *Cancers (Basel)* 12 (10). doi: 10.3390/cancers12102959

Sloan, D. B., Wu, Z., and Sharbrough, J. (2018). Correction of Persistent Errors in Arabidopsis Reference Mitochondrial Genomes. *Plant Cell* 30 (3), 525–527. doi: 10.1105/tpc.18.00024

Song, X., Ge, T., Li, Y., and Hou, X. (2015). Genome-wide identification of SSR and SNP markers from the non-heading Chinese cabbage for comparative genomic analyses. *BMC Genomics* 16 (1), 328. doi: 10.1186/s12864-015-1534-0

Su, X., Yang, L., Wang, D., Shu, Z., Yang, Y., Chen, S., et al. (2022). 1 K Medicinal Plant Genome Database: an integrated database combining genomes and metabolites of medicinal plants. *Hortic. Res.* 9, uhac075. doi: 10.1093/hr/uhac075

Sun, S., Shen, X., Li, Y., Li, Y., Wang, S., Li, R., et al. (2023). Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism. *Nat. Plants* 9 (1), 179–190. doi: 10.1038/s41477-022-01291-y

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (W1), W6–w11. doi: 10.1093/nar/gkx391

Trotta, E. (2013). Selection on codon bias in yeast: a transcriptional hypothesis. *Nucleic Acids Res.* 41 (20), 9382–9395. doi: 10.1093/nar/gkt740

Turmel, M., Otis, C., and Lemieux, C. (2016). Mitochondrion-to-Chloroplast DNA Transfers and Intragenomic Proliferation of Chloroplast Group II Introns in Gloeotilopsis Green Algae (Ulotrichales, Ulvophyceae). *Genome Biol. Evol.* 8 (9), 2789–2805. doi: 10.1093/gbe/evw190

Wang, S., Li, D., Yao, X., Song, Q., Wang, Z., Zhang, Q., et al. (2019). Evolution and Diversification of Kiwifruit Mitogenomes through Extensive Whole-Genome Rearrangement and Mosaic Loss of Intergenic Sequences in a Highly Variable Region. *Genome Biol. Evol.* 11 (4), 1192–1206. doi: 10.1093/gbe/evz063

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31 (20), 3350–3352. doi: 10.1093/bioinformatics/btv383

Wicker, T., Buchmann, J. P., and Keller, B. (2010). Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res.* 20 (9), 1229–1237. doi: 10.1101/gr.107284.110

Xiong, A. S., Peng, R. H., Zhuang, J., Gao, F., Zhu, B., Fu, X. Y., et al. (2008). Gene duplication and transfer events in plant mitochondria genome. *Biochem. Biophys. Res. Commun.* 376 (1), 1–4. doi: 10.1016/j.bbrc.2008.08.116

Yang, H., Chen, H., Ni, Y., Li, J., Cai, Y., Ma, B., et al. (2022). *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.* 23 (22). doi: 10.3390/ijms232214267

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour* 20 (1), 348–355. doi: 10.1111/1755-0998.13096

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. *BMC Bioinf.* 14, 244. doi: 10.1186/1471-2105-14-244