



OPEN Insights into structure, codon usage, repeats, and RNA editing of the complete mitochondrial genome of *Perilla frutescens* (Lamiaceae)

Ru Wang^{2,4}, Yongjian Luo^{2,4}, Zheng Lan³ & Daoshou Qiu^{1✉}

Perilla frutescens (L.) Britton, a member of the Lamiaceae family, stands out as a versatile plant highly valued for its unique aroma and medicinal properties. Additionally, *P. frutescens* seeds are rich in γ -linolenic acid, holding substantial economic importance. While the nuclear and chloroplast genomes of *P. frutescens* have already been documented, the complete mitochondrial genome sequence remains unreported. To this end, the sequencing, annotation, and assembly of the entire Mitochondrial genome of *P. frutescens* were hereby conducted using a combination of Illumina and PacBio data. The assembled *P. frutescens* mitochondrial genome spanned 299,551 bp and exhibited a typical circular structure, involving a GC content of 45.23%. Within the genome, a total of 59 unique genes were identified, encompassing 37 protein-coding genes, 20 tRNA genes, and 2 rRNA genes. Additionally, 18 introns were observed in 8 protein-coding genes. Notably, the codons of the *P. frutescens* mitochondrial genome displayed a notable A/T bias. The analysis also revealed 293 dispersed repeat sequences, 77 simple sequence repeats (SSRs), and 6 tandem repeat sequences. Moreover, RNA editing sites preferentially produced leucine at amino acid editing sites. Furthermore, 70 sequence fragments (12,680 bp) having been transferred from the chloroplast to the mitochondrial genome were identified, accounting for 4.23% of the entire mitochondrial genome. Phylogenetic analysis indicated that among Lamiaceae plants, *P. frutescens* is most closely related to *Salvia miltiorrhiza* and *Platostoma chinense*. Meanwhile, inter-species Ka/Ks results suggested that Ka/Ks < 1 for 28 PCGs, indicating that these genes were evolving under purifying selection. Overall, this study enriches the mitochondrial genome data for *P. frutescens* and forges a theoretical foundation for future molecular breeding research.

Keywords Lamiaceae, Sequence characteristics, Phylogenetic analysis, Mitochondrion, RSCU, RNA editing

The mitochondrial genome is relatively small, highly conserved, and densely packed with genes, while also containing highly variable non-coding regions? In plant research, in addition to its role in energy metabolism, the mitochondrial genome generally affects plant fertility, also playing an important role in plant development and stress resistance¹. Comparative analyses of mitochondrial genome sequences, structures, and functions reveal the complexity of plant mitochondrial genomes, providing a basis for deciphering their classification and evolution, and facilitating the exploration of cytoplasmic male sterility mechanisms^{2,3}. However, research on complete plant mitochondrial genomes lags far behind that of complete plastid genomes. The NCBI database hosts nearly 673 complete plant mitochondrial genomes⁴, which is attributed to 13,000 complete plastid genomes, contrasting with only 673 complete plant mitochondrial genomes⁴, which is attributed to the complex structure of plant Mitochondria. They are challenging to purify and often face interference from chloroplasts and other organelles, complicating genome assembly. In addition, significant differences are observed in the structure and content of plant mitochondrial genomes, as well as in nucleotide substitution rates and repeat sequences^{3,5}.

¹Key Laboratory of Crops Genetics and Improvement of Guangdong Province, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China. ²Hubei Minzu University, School of Forestry and Horticulture, Enshi 445000, China. ³Heilongjiang Bayi Agricultural University, Daqing 163319, China. ⁴These authors contributed equally: Ru Wang and Yongjian Luo. ✉email: qiudaoshou@gdaas.cn

Therefore, a complete description of the plant mitochondrial genome remains a bottleneck in evolutionary biology, yet most plant systematic studies focus on nuclear and plastid genomes⁶. For example, while the nuclear and plastid genomes of *Perilla frutescens* have been sequenced and assembled, its mitochondrial genome remains unknown^{7–9}. With the development of high-throughput sequencing technology and the rise of next-generation systems genomics, many software programs applicable to mitochondrial genome sequencing and assembly, such as GetOrganelle¹⁰, Mitofiner¹¹, GSAT¹², and PMAT¹³ (<https://github.com/bichangwei/PMAT>), have been developed, making mitochondrial genome sequencing and assembly more accurate and efficient¹⁴.

Besides, in land plants, mitochondrial genomes typically contain exogenous genes or fragments due to horizontal gene transfer or intracellular gene transfer^{15,16}. As the mitochondrial genome undergoes continuous recombination, this sequence transfer between genomes also occurs¹⁷. This can result in cytoplasmic male sterility (CMS), where plants cannot produce viable pollen grains due to a mismatch between the nuclear and cytoplasmic genomes. However, female fertility is maintained, as observed in Brassica, rice, and other plants^{17–19}. Additionally, CMS is often inherited maternally and is commonly associated with abnormal open reading frames (ORFs) and RNA editing in the mitochondrial genome²⁰. These ORFs are often found near functional mitochondrial genes and are co-transcribed with them, affecting the mitochondrial function²¹. CMS-based hybridization techniques have been utilized to breed offspring with significantly improved yield, stress resistance, and adaptability, making this approach a promising way to maintain crop productivity²². Therefore, mitochondrial genomes are a valuable source of genetic information for plant systematics and necessary cellular process research. They have significant implications for species evolution research, species identification, and genetic transformation²³.

Perilla frutescens (L.) Britt. belongs to the Lamiaceae family, and is an annual upright herb. It is renowned as a traditional dual-purpose medicinal and edible plant, prized for its distinct aroma²⁴. Notably, its leaves, stems, and fruits all possess medicinal properties. *P. frutescens* is widely cultivated in China and Southeast Asia, ranging from Indonesia in the south to Japan in the east²⁵. Current research on *P. frutescens* mainly focuses on its pharmacological effects and chemical composition of its compound perillaldehyde²⁶. Currently, 400 compounds have been isolated from *P. frutescens* leaves, including terpenoids, flavonoids, alkaloids, steroids, quinines, and phenolic compounds, all of which hold significant potential for various applications^{27–29}. People have discovered the pharmacological functions of *P. frutescens* attributed to secondary metabolites present in different parts of the plant, such as anti-allergic, anti-depressive, lipid-lowering, hepatoprotective, neuroprotective, anti-inflammatory, anti-cancer, antioxidant and antibacterial activities^{30–33}. In addition, common oil crops such as soybean, rapeseed, peanut, and olive generally have an α -linolenic acid content of less than 5%^{34,35}. Conversely, *P. frutescens* seeds boast a high oil content of 45–55%, with a rich content of unsaturated fatty acids, accounting for more than 90% of the total oil content. Among which, α -linolenic acid has the highest content, reaching 55.65%³⁶. Consequently, *P. frutescens* is considered a natural resource possessing significant economic and medicinal value. Despite being a plant of significant medicinal and economic value, the mitochondrial genome of *P. frutescens* has not yet been assembled and analyzed. This crucial step is essential for comprehending its genetic composition and realizing its full potential for diverse applications.

In this study, sequencing and annotation of the *P. frutescens* mitochondrial genome were conducted, and mitochondrial genome features, RNA editing, and codon bias were comprehensively analyzed. Besides, systematic evolutionary analysis was also carried out to provide essential background information for further understanding the genetics of this plant. Overall, this research lays the groundwork for future studies on molecular breeding strategies for *P. frutescens*.

Results

Mitochondrial gene organization and features of *P. frutescens*

In this study, the mitochondrial genome was sequenced using Illumina and Pacbio sequencing platforms. A total of 248,479,765 clean reads (Q30 = 89.78%) were generated from the second-generation sequencing platform, while the third-generation sequencing platform produced 7,853,744 clean reads. A total of 63,122,358,854 bases were generated. The subreads possessed an N50 value of 9597 bp and an N90 value of 4224 bp. The longest sub-read possessed a length of 65,388 bp. The complete circular mitochondrial DNA molecule of *P. frutescens* with a length of 299,551 bp was obtained through de novo assembly (Fig. 1). The nucleotide composition of *P. frutescens* mitochondrial DNA was 27.3% A, 27.5% T, 22.5% G, and 22.7% C, with a GC content of 45.23%. Ultimately, the *P. frutescens* mitochondrial genome sequence was archived in CNGB with the accession number AA059311.1.

The mitochondrial genome of *P. frutescens* encompasses a total of 59 genes, consisting of 37 protein-coding genes (PCGs), 20 transfer RNAs (*trnAs*), and 2 ribosomal RNAs (rRNAs) (refer to Table 1). The essential genes include 6 ATP synthase genes (*atp1*, *atp4*, *atp6*, *atp8*, and *atp9*), 9 NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*), 4 cytochrome c biogenesis genes (*ccmB*, *ccmC*, *ccmFC*, and *ccmFN*), 4 cytochrome c oxidase genes (*cox1*, *cox2*, and *cox3*), 1 Maturases gene (*matR*), 1 membrane transport protein gene (*mttB*), and 1 ubiquinol cytochrome c reductase gene (*cob*). The variable genes include 3 large subunits of ribosomal protein (*rpl16*, *rpl10*, *rpl5*), 6 small subunits of ribosomal protein (*rps10*, *rps12*, *rps13*, *rps14*, *rps3*, *rps4*), and 2 succinate dehydrogenases (*sdh4*). Herein, notably, the genes *atp8*, *cox3*, and *sdh4* were observed in duplicate copies. Introns were found in the *nad1*, *nad2*, *nad5* (with four introns each), *nad4*, and *nad7* genes (with three introns each), as well as in *ccmFC*, *cox1*, *cox2*, *rps10*, and *rps3* (each containing one intron). Furthermore, the *sdh4* sequences were identified as pseudogenes. Among the transfer RNA genes, *trnM*-CAU was duplicated 5 times, while the remaining *trnA* genes existed as identical copies within the mitochondrial genome.

Repeat sequence analysis

In the plant mitochondrial genome, there are abundant repetitive sequences, including simple sequence repeats (SSRs), tandem repeats, and dispersed repeats³⁷. In the *P. frutescens* mitochondrial genome, two different types

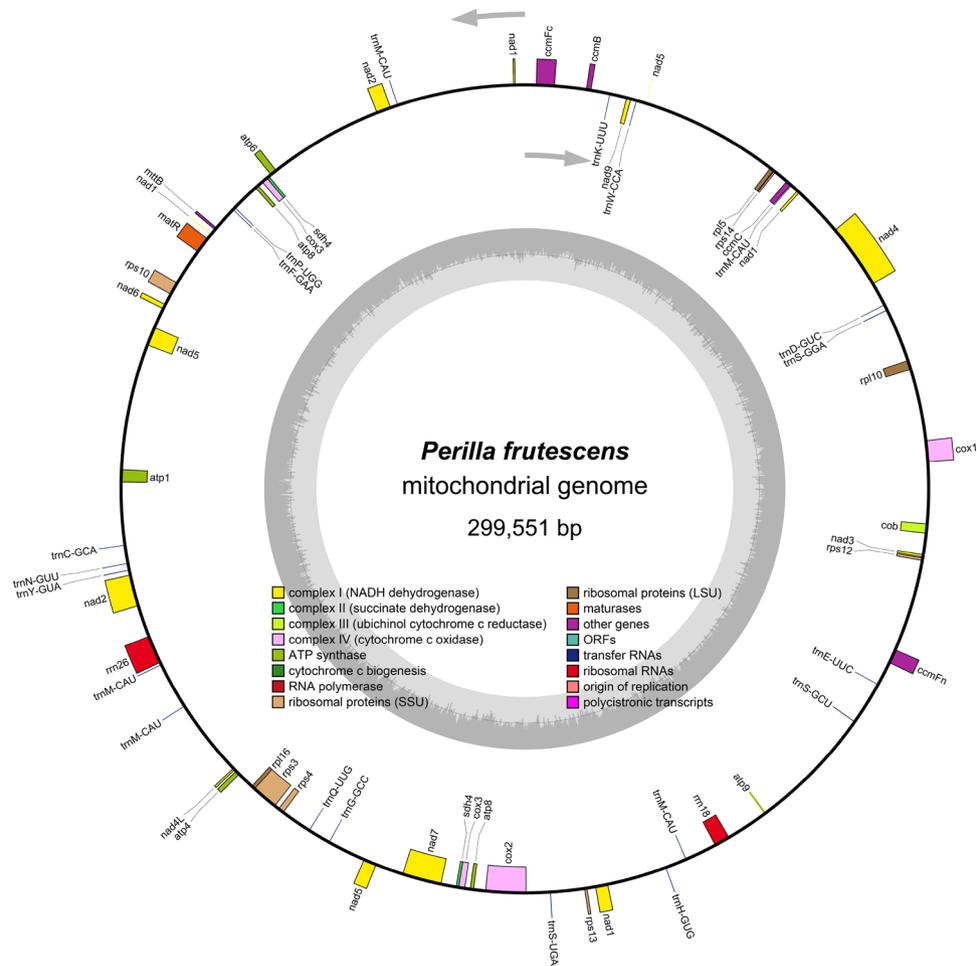


Figure 1. The mitogenome map of *P. frutescens*.

Group of genes	Gene name
ATP synthase	<i>atp1, atp4, atp6, atp8(2), atp9</i>
Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFc*, ccmFn</i>
Ubichinol cytochrome c reductase	<i>cob</i>
Cytochrome c oxidase	<i>cox1*, cox2*, cox3(2)</i>
Maturases	<i>matR</i>
Transport membrane protein	<i>mttB</i>
NADH dehydrogenase	<i>nad1****, nad2****, nad3, nad4***, nad4L, nad5****, nad6, nad7***, nad9</i>
Ribosomal proteins (LSU)	<i>rpl16, rpl10, rpl5</i>
Ribosomal proteins (SSU)	<i>rps10*, rps12, rps13, rps14, rps3*, rps4</i>
Succinate dehydrogenase	<i>sdh4(2)</i>
Ribosomal RNAs	<i>rrn18, rrn26</i>
Transfer RNAs	<i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnM-CAU(5), trnN-GUU, trnP-UGG, trnQ-UUG, trnS-GCU, trnS-GGA, trnS-UGA, trnW-CCA, trnY-GUA</i>

Table 1. Classification of genes in the *P. frutescens* mitochondrial genome. Notes: (*): intron number; (#): Pseudo gene; Gene (2): Number of copies of multi-copy genes.

of dispersed repeats, i.e., Forward and Palindromic, were identified, totaling 293 dispersed repeats with a length greater than or equal to 20 bp. No reverse or complementary repeat sequences were detected. The dispersed repeat sequences were visualized using the Circos software package. The results uncovered the wide distribution of these repeats in the intergenic regions of these repeats in the intergenic regions of the entire genome (Fig. 2A and Supplementary Table S1). Besides, a total of 293 dispersed repeat sequences were identified, with a combined

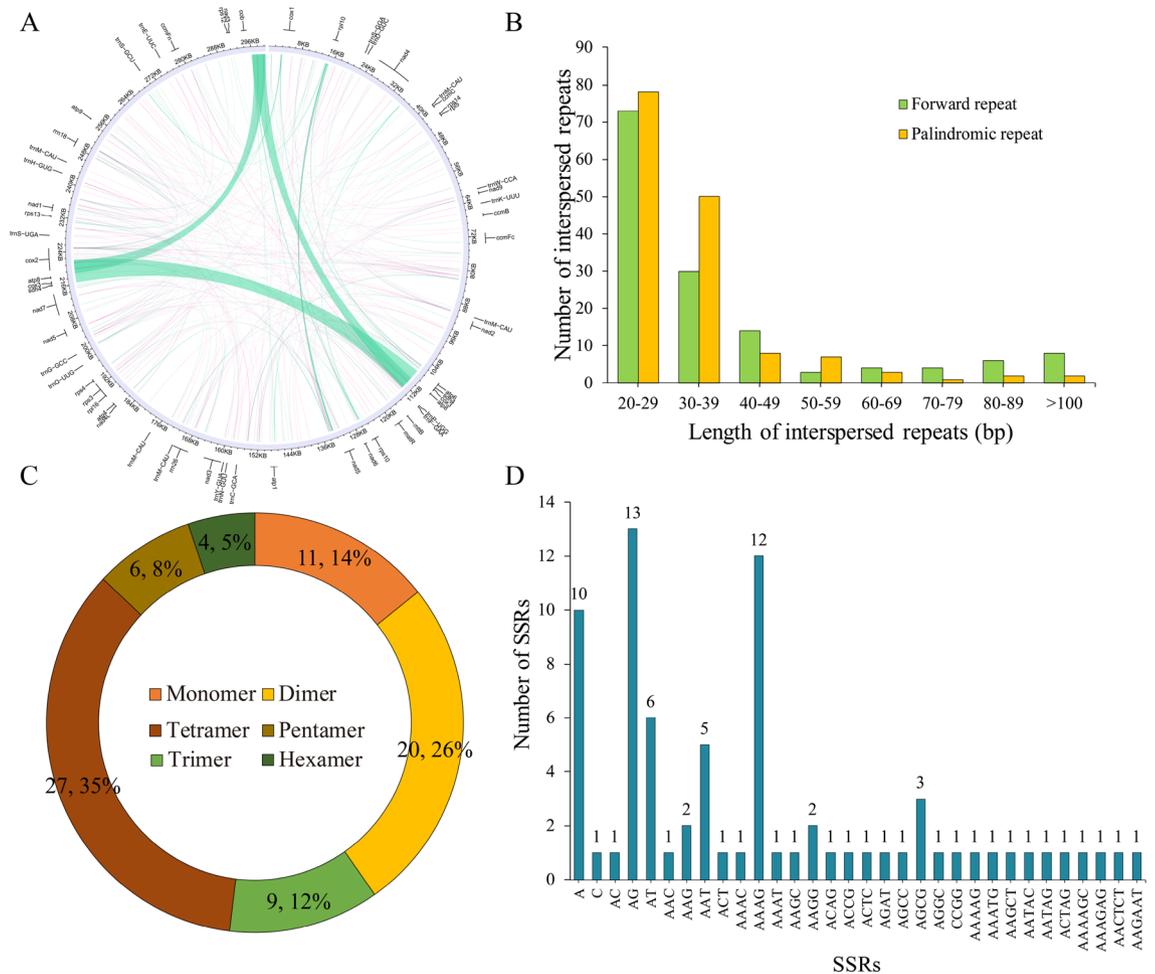


Figure 2. Mitochondrial genomic repeat sequences of *P. frutescens*. (A) Distribution of repeat sequences on the genome; (B) Allocation of the lengths of dispersed repeats in the *P. frutescens* mitogenome; (C) Statistical histogram of the *P. frutescens* mitochondrial genome with repeat sequences of different lengths; (D) SSR motif analysis.

length of 22,2651 bp, accounting for 3.76% of the total mitochondrial genome length. Among these repeats, there were 151 forward repeats and 142 palindromic repeats (Fig. 2B). The longest forward repeat sequence was 5545 bp, while the longest palindromic repeat sequence was 222 bp. Further analysis demonstrated 20–29 bp repeats as the most common type, accounting for 51.53% of the total repeat occurrences. The presence of these dispersed repeat sequences might be related to the structure and function of the mitochondrial genome.

Microsatellites (simple repeat sequences, SSRs) are typically tandem sequences of up to 6 base pairs in eukaryotic genomes³⁸. In this investigation, a total of 77 SSRs were identified in the mitogenome of *P. frutescens* (Fig. 2C), dominated by tetrameric repeats, accounting for 27.35% (27) of the total count, followed by dimeric repeats (20), monomeric repeats (11), trimeric repeats (9), pentameric repeats (6), and hexameric repeats (5). Adenine (A) repeats took up the highest proportion (90.91%) within the monomeric SSRs, while AG repeats occupied the highest proportion among dimeric repeats (65.00%) (Fig. 2D). None of the repeat types exhibited sequences with a repeat length exceeding 20 base pairs. These widely distributed SSRs provided rich potential molecular markers for the identification and genetic research of *P. frutescens* plants.

Additionally, tandem repeats, also known as satellite DNA, are characterized by repeat lengths of 1–200 bases with varying numbers of repeats³⁹. Herein, 6 tandem repeat sequences with lengths between 34 and 60 base pairs were identified in the *P. frutescens* mitochondrial genome, all of which were present in the intergenic regions (Supplementary Table S2).

Homologous sequence analysis of organellar genomes

The mitochondrial genome of *P. frutescens* (299,551 bp) is approximately 1.96 times larger than the chloroplast genome (152,593 bp). However, compared to the chloroplast genes, the distribution of mitochondrial genes in *P. frutescens* is relatively sparse (Fig. 3 and Supplementary Table S3). In this study, based on sequence similarity between chloroplast and mitochondrial genomes, 70 chloroplast-like segments potentially involved in gene transfer were identified in the mitochondrial genome (Fig. 3). These inserted segments were distributed throughout the mitochondrial genome, with a total length of 12,680 bp, accounting for 4.23% of the entire mitochondrial genome. The longest sequence (875 bp) was transferred from the cp genome's *psbA* gene to the intergenic regions

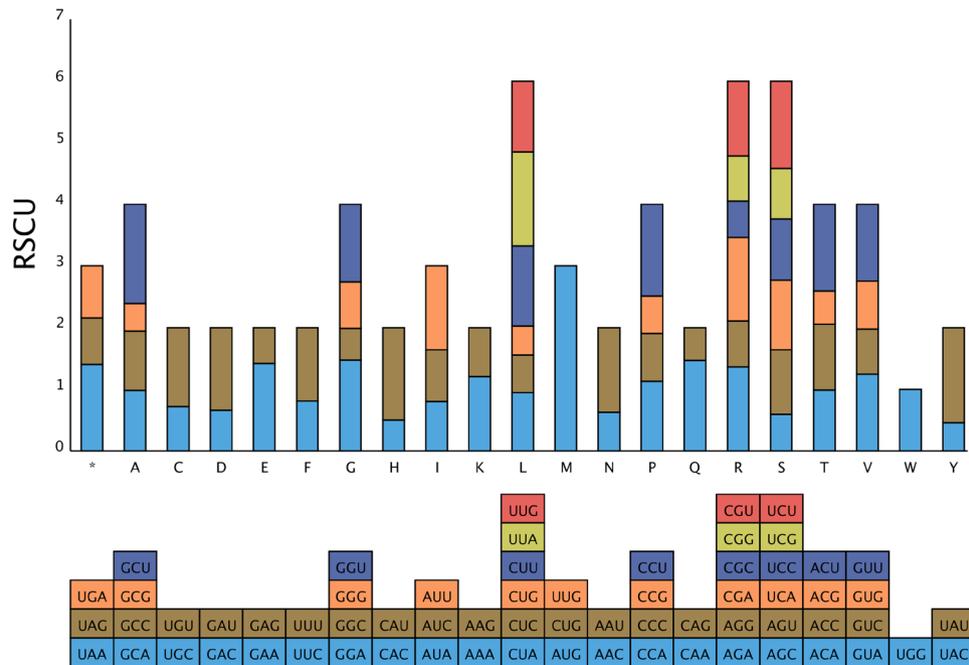


Figure 4. Analysis of relative synonymous codon usage (RSCU) in the *P. frutescens* mitochondrial genome.

ending with C or G had RSCU values mostly less than 1. Additionally, codon usage was typically strongly biased towards A or T(U) at the third codon position, which was also observed in other plant mitochondrial genomes.

Selective pressure analyses

The Ka/Ks values of common protein-coding genes in the *P. frutescens* mitochondrial genome were calculated and compared with 9 Lamiaceae species (*Ajuga reptans*, *Lavandula angustifolia*, *Platostoma chinense*, *Pogostemon heyneanus*, *Prunella vulgaris*, *Rotheca serrata*, *Scutellaria barbata*, *Scutellaria franchetiana*, and *Vitex trifolia*) (Fig. 5). The average Ka/Ks value for 32 identical protein-coding genes was 0.33. Deeper investigation into the selective pressures on specific genes uncovered that except for *ccmFn*, *mttB*, *rps10*, and *matR* showed Ka/Ks ratios > 1 compared to the other 9 plants, indicating that they might be undergoing positive selection. Most genes exhibited a negative selection effect (Ka/Ks < 1) compared to the other 9 plants, suggesting that most protein-coding genes in the *P. frutescens* mitochondrial genome were highly conserved during molecular evolution.

RNA editing sites prediction

RNA editing refers to the process of adding, removing, or substituting bases in the coding region of transcribed RNA, which occurs in all eukaryotes¹⁵. In the present study, a total of 559 RNA editing sites were predicted within the mitochondrial genome's 34 protein-coding genes (PCGs) of *P. frutescens*. The distribution of these RNA editing sites varied among different genes, with the *nad4* gene having the highest number at 48, followed by *ccmB* with 42 sites, making them the top two genes in terms of RNA editing sites (Fig. 6, Supplementary Table S4). Additionally, *mttB*, *ccmC*, *nad7*, *nad2*, *ccmFn*, and *nad5* each exhibited over 30 editing sites. Conversely, genes like *rps14*, *rpl2*, and *rps7* presented the fewest editing sites, with only two C-to-U edits observed. Meanwhile, a total of 56 types of codon transitions were identified, yielding 25 different amino acid changes. Among these codon transition types, the UCA→UUA transition occurred the most frequently, occurring at 64 sites. The translated amino acids were significantly affected by RNA editing events, with 558 RNA editing sites leading to non-synonymous substitutions and alterations in encoded amino acids, accounting for 99.83% of the total changes. Furthermore, the most common substitutions involved proline (Pro) and serine (Ser) being replaced by leucine (Leu) 100 and 99 times, respectively. Additionally, serine (Ser) was frequently substituted by phenylalanine (Phe) in 76 cases. Collectively, these three substitutions accounted for 49.23% of all observed changes. In comparison to the one synonymous substitution RNA editing event, these widespread RNA editing events held considerable significance in protein translation.

Phylogenetic analysis

Reconstructing the phylogenetic tree based on conservative mitochondrial protein-coding genes is instrumental in elucidating the molecular evolutionary relationships among green plants. To determine the classification status of the *P. frutescens* mitochondrial genome, *P. frutescens* was hereby analyzed phylogenetically with 59 species, including 44 dicotyledonous plants (including 20 species of Lamiaceae), 13 monocotyledonous plants, 1 gymnosperm, and 1 bryophyte (outgroup). A total of 24 protein-coding genes corresponding to each species were also analyzed.

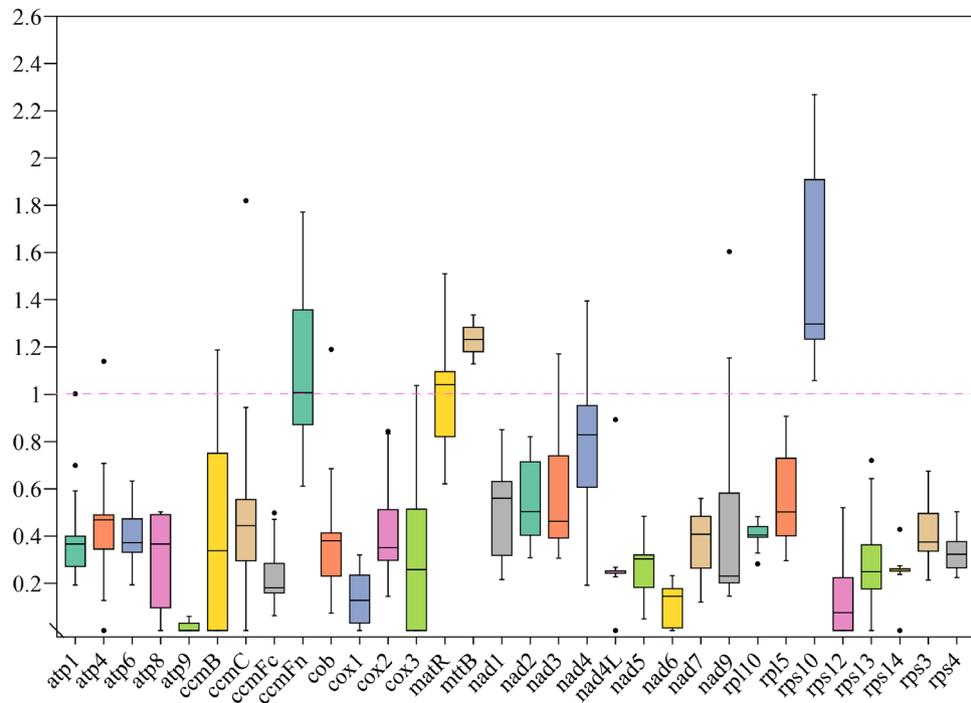


Figure 5. Boxplots of the Ka/Ks ratio of *P. frutescens* and 9 other plant species.

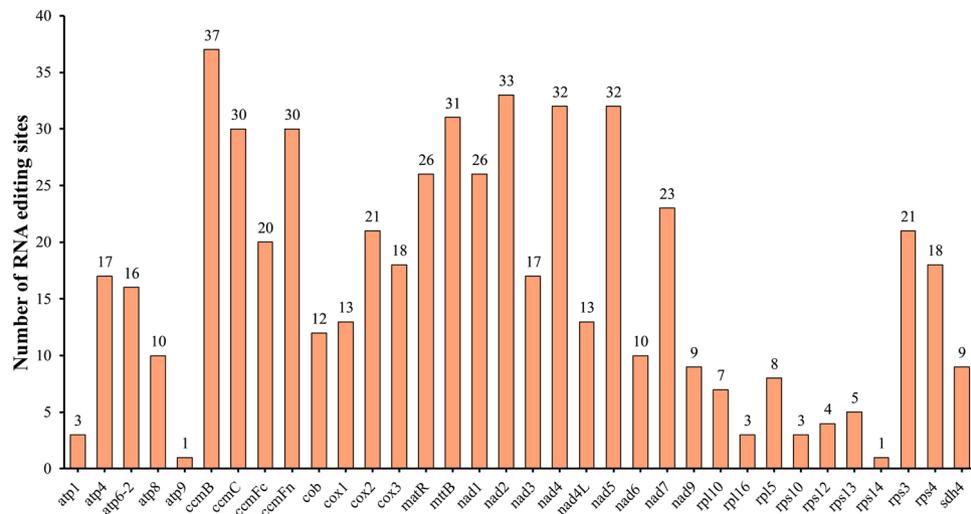


Figure 6. Distribution of RNA editing sites in the mt PCGs of *P. frutescens*.

Based on the concatenated dataset, both ML and BI phylogenies generated identical tree topologies, with nearly all nodes exhibiting high support values (ultrafast bootstrap (UFboot) = 100; SH-aLRT values (SH-aLRT) = 100; posterior probability (PP) = 1.00). The results indicated that the phylogenetic tree strongly supported the separation of dicots and monocots, as well as that of angiosperms and gymnosperms (Fig. 7). Furthermore, the clade distributions of Asparagales, Lamiales, Fabales, Brassicales, Alismatales, Poales, Cucurbitales, Gentianales, Malvales, Caryophyllales, Vitales, Rosales, Solanales, Apiales, Butomaceae, Zosteraceae, Marchantiophyta, Malpighiales, Ginkgoales and Arecales were well resolved in the phylogenetic tree. Relationships in the phylogenetic tree were found to be consistent with the traditional classification relationships of these species, indicating the congruence of traditional and molecular classifications at the family level. *P. frutescens* belonging to the Nepetoideae subfamily of the Lamiaceae family shared a highly similar topology to that of the phylogenetic reconstruction inferred from plastid genome and nuclear genome sequences. The species tree topology implemented in ASTRAL was basically congruent with the result of the concatenation analyses, with high local posterior probability (LPP) values at most nodes. Overall, the present analysis of the *P. frutescens* mitochondrial genome provides a valuable foundation for future studies on the phylogenetic relationships of species.

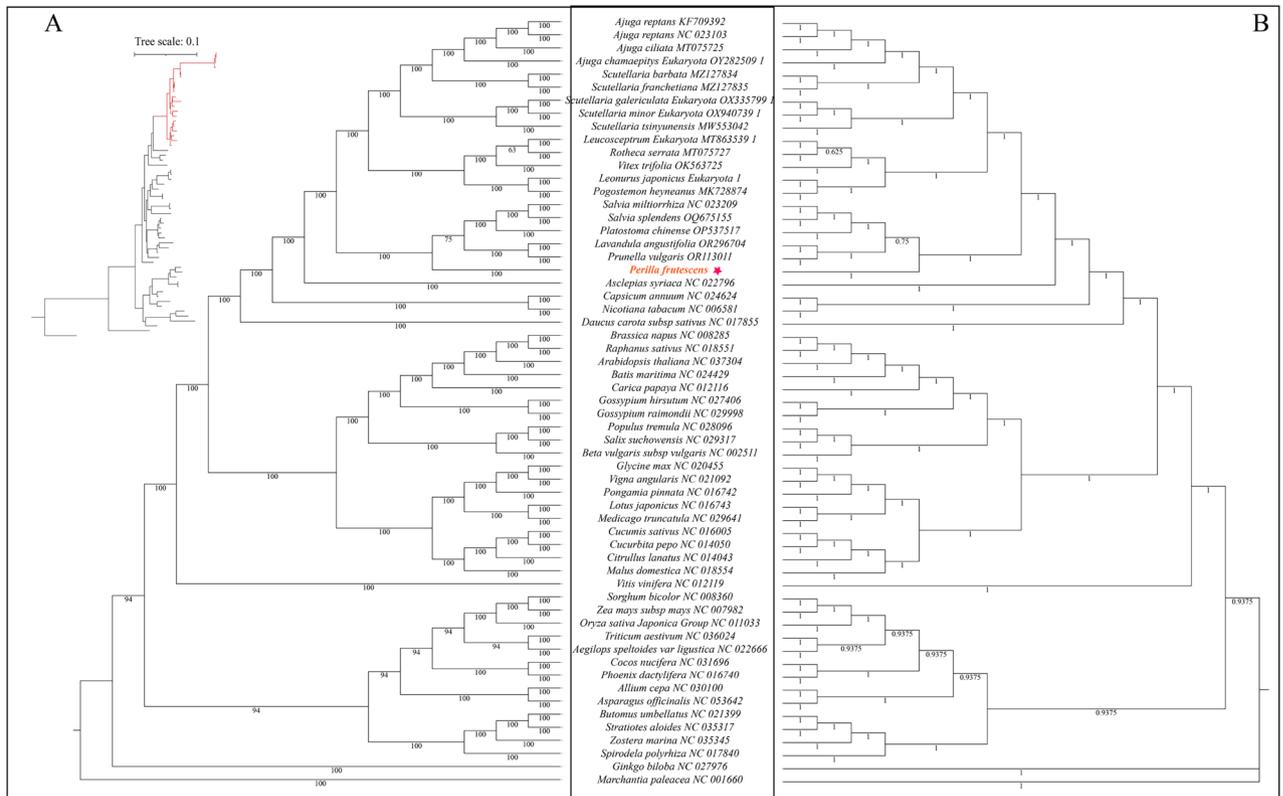


Figure 7. Phylogenies of 59 species inferred from whole mitogenomes. **(A)** Phylogeny of *P. frutescens* reconstructed by analyses of 24 concatenated plastid PCGs using maximum likelihood (ML) and Bayesian inference (BI) methods. Values at nodes represent the ultrafast bootstrap support values (UFBoot) of the maximum likelihood analysis and the posterior probabilities (PP) of the BI analysis values. **(B)** Species tree of *P. frutescens* reconstructed by ASTRAL based on 24 plastid protein-coding genes. Values at nodes represent local posterior probabilities (LPP).

Discussion

Mitochondria are essential organelles in eukaryotic cells. They are crucial for cellular respiration and energy metabolism, playing key roles in regulating important cellular activities such as differentiation, apoptosis, growth, and division^{40,41}. The complex organization, diverse structures, dynamic non-coding sequences, and high levels of repetitive sequences in plant mitochondrial genomes all pose challenges in the study of plant mitochondria^{4,15,42,43}. Herein, considering the much higher than copy number of plant organelle genomes compared to the corresponding nuclear genome, Illumina and PacBio sequencing technologies were combined to first describe the basic characteristics of the *P. frutescens* genome. The mitochondrial genome of *P. frutescens* was demonstrated as a circular structure with a total length of 299,551 bp. The genome sizes of Lamiaceae plants ranged from 271,618 to 729,504 bp, and the mitochondrial genome of *P. frutescens* was relatively small compared to other species in the Lamiaceae family. Its genome size was similar to the published mitochondrial genome of *Prunella vulgaris* (274,779 bp) in the Lamiaceae family⁴⁴. The mitochondrial genome of *P. frutescens*, similar to most Lamiaceae plants, predominantly exhibited a circular structure, which might be attributed to the relative rarity of homologous recombination events, leading to structural conservation. However, some previously assembled mitochondrial genomes of Lamiaceae plants exhibited multi-chromosomal or multi-branch structures. For instance, *Ajuga decumbens* (374,491 bp) and *Teucrium ornatum* (608,646 bp) each had up to 8 and 9 isolated scaffolds, respectively. The occurrence of such diverse chromosomal structures in mitochondrial genomes within the same family highlighted the evolutionary adaptability and complexity of plant mitochondrial genomes⁴⁵. Nevertheless, The mitochondrial genome of angiosperms averages 475 kb, and the mitochondrial genome of Lamiaceae has a smaller and more conserved structure compared to other angiosperms? Besides, the depth and length of sequencing data during the assembly process also significantly impacted the assembly results⁴, warranting further investigation upon the mitochondrial genome structures of *Ajuga decumbens* and *Teucrium ornatum*. The GC content of the mitotic genome of *P. frutescens* was 45.23%, which was highest in the rRNA genes. The result was consistent with other Lamiaceae species such as *Scutellaria tsinyunensis* (45.09%)⁴⁶, *Rotheca serrata* (45.53%)⁴⁷, and *P. vulgaris* (43.92%)⁴⁴. This suggested that the GC content of the mitotic genome was relatively conserved during the evolutionary process in plants. In angiosperms, the mitotic genome encoded 24 core protein-coding genes, including *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*, most being respiratory protein genes⁴⁸. In the mitochondrial genome of *P. frutescens*, a total of 59 genes were identified, including 37 protein-coding genes (26 were core genes), 20 tRNA genes, and 2 rRNA genes. The types and quantities of these genes

were consistent with the core mitochondrial genes of most Lamiaceae plants (Supplementary Table S5)^{44,46,49}. The coding region of the *P. frutescens* mitochondrial genome accounted for only 19.23% of its total length, with over 80% being non-coding regions. This might result from the gradual increase of repeat sequences in the mitochondrial genome during evolution⁴⁷.

The calculation of non-synonymous substitutions (Ka) and synonymous substitutions (Ks) plays a critical role in phylogenetic reconstruction and understanding the evolutionary dynamics of protein-coding sequences among closely related species⁴⁴. In genetics, the Ka/Ks ratio serves as a key indicator to determine whether specific protein-coding genes are subject to selection pressure during evolution. A Ks value equaling Ka, or the Ka/Ks ratio equaling 1 suggests neutral selection. A Ka/Ks ratio greater than 1 indicates positive selection, where the non-synonymous substitution rate (Ka) exceeds the synonymous substitution rate (Ks). Conversely, negative selection appears if Ks surpasses Ka or the Ka/Ks ratio is less than 1⁵⁰. Herein, few protein-coding genes were affected by positive selection, aligning with existing research reports^{3,43,46,51,52}. The Ka/Ks analysis of the mitochondrial genome indicated that widely present PCGs (protein-coding genes) in mitochondrial were retained. However, genes with Ka/Ks > 1, such as *ccmFn*, *mttB*, *rps10*, and *matR* were also detected, indicating that positive selection played a role in the evolutionary history of these coding genes. Besides, high Ka/Ks genes were considered important in the study of gene selection and evolution in Lamiaceae plants.

The transfer of genes between chloroplast and mitochondrial genomes is an important feature of plant mitochondrial genome evolution⁵³. When exogenous genes are inserted into the mitotic genome, they are preferentially inserted into intergenic regions^{54,55}. This phenomenon is a major reason for the observed differences in the number of coding genes in mitochondrial genomes among different plant species. Therefore, tracking gene transfer matters considerably in exploring the evolution of plant mitochondrial genomes. The length of chloroplast genome-integrated DNA into the mitotic genome varies among plant species, but generally falls within the range of 1–12% of the angiosperm plastome sequence⁵⁶, such as in *Garcinia mangostana* L. (1.7%)⁵⁷ and *cucurbita pepo* (11.6%)⁵⁸. In this study, 70 migrating segments with a total length of 12,680 bp were identified, accounting for 4.23% of the entire mitochondrial genome. In the evolutionary history of mitochondrial genomes, widespread loss events of tRNA were observed, and the lost tRNAs in mitochondrial genomes could be replaced by tRNAs from other organelles⁵⁹. In this study, it was discovered that the chloroplast *rps11* and 15 tRNA genes completely transferred to the Mitochondrial genome, a common phenomenon in plants featuring frequent transfer of tRNA from chloroplasts to mitochondria⁶⁰, and that chloroplast-derived tRNAs might have potential functional complementation. Furthermore, a significant amount of chloroplast gene fragments migrating to the mitochondrial genome were also identified. These chloroplast-derived fragments contained genes playing important roles in chloroplast function. However, their functionality in the mitochondrial genome remains unclear in this study. Two possible explanations were proposed for the fate of exogenous genes in the mitochondrial genome: (1) Transferred genes typically lost their functionality, while naturally functional copies coexisted in the mitochondrial genome⁶¹; (2) Natural genes were lost from the mitochondrial genome, and exogenous copies might have a function in maintaining normal cellular operations⁶². Hence, understanding the patterns of sequence transfer plays an important role in tracing ancient recombination events and structural variations in plant mitochondrial genomes, requiring more attention.

In plants, RNA editing enzymes catalyze conversion of specific nucleotide positions in mitochondrial RNA sequences from C to U or U to C. RNA editing holds much significance in regulating mitochondrial gene expression and function by inducing alterations in RNA sequences, thereby influencing the translated protein products^{63–65}. RNA editing is an important source of variation in angiosperm mitochondrial genomes⁶⁶. Additionally, the mechanism of crop male sterility has been associated with RNA editing processes. Herein, in the mitochondrial genome of *P. frutescens*, 559 RNA editing sites were identified, primarily occurring at the first and second positions of codons, with the editing method being cytosine \rightleftharpoons uracil (C \rightleftharpoons U). The preference for G/C codons might be attributed to their high binding energy, which played a role in maintaining translation accuracy⁶⁷. By identifying RNA editing sites, researchers could gain valuable insights into the functional implications of predicting newly encoded codons. Notably, RNA editing events produced stop codons in the *rps10* gene, and start codons in three genes (*cox2*, *nad4L*, and *rps10*), which could often be related to the production of highly conserved and homologous proteins found in other species, thereby promoting efficient gene expression in mitochondria. Furthermore, RNA editing events were found to be least common in ribosomal protein genes like *rps14*, while genes like *nad5*, *mttB*, and *ccmB* exhibited higher frequencies of RNA editing events. These findings suggested the crucial role of RNA editing in plant adaptation to environmental changes and signal transduction^{68,69}. Codon usage bias is an important factor reflecting the evolution of mitochondrial genomes⁷⁰. Generally, factors such as mutation, natural selection, and phylogenetic relationships may lead to differences in codon usage preferences⁷¹. In the mitochondrial genome of *P. frutescens*, 30 codons with bias (RSCU > 1) were hereby identified, and these codons used more A/T bases. Therefore, upcoming research should account for codon usage biases influenced by mitochondrial RNA editing sites in the development of male sterile materials, thereby facilitating advancements in genetic breeding programs.

Repeat sequences are widely present in mitochondrial genomes and are usually crucial for molecular recombination, structural variation, and extreme differences in mitotic genome size⁷². These sequences also serve as important sources of information for population development and evolutionary analysis markers⁷³. In this study, a total of 293 repeats, including 141 pairs of reverse complementary repeats and 152 pairs of forward repeats, were observed, possibly revealing frequent molecular recombination in mitochondrial genomes. This frequent recombination could play a crucial role in dynamically changing the structure and conformation of the mitochondrial genome during the evolutionary process. Simple sequence repeats (SSRs) represent a distinctive class of repetitive DNA sequences characterized by high repetitiveness, extensive polymorphism, and frequent co-dominance, and are abundantly dispersed throughout plant genomes³⁸. SSR markers provide useful information for analyzing genetic diversity, genetic relationships, and population structures in crop species germplasm. In the

present study, 77 SSRs were distributed in different genomic regions of *P. frutescens* mitotic genomes. The most abundant SSR type was tetrameric repeats (27.35%), followed by dimeric repeats (20%), consistent with previous studies on mitochondrial genomes. Moreover, a nearly universal A/T bias was also observed in the mitochondrial genome of *P. frutescens*. These SSRs consisted of motifs rich in A and T, also consistent with previous observations. The findings further confirmed the correlation between the AT content of the complete mitotic genome and SSRs. Extensive collection of genetic resources in *P. frutescens* could be found⁷⁴. While numerous plant genetic resources have been collected in gene banks worldwide, the large sample size and lack of sufficient information on population structure and genetic diversity still terribly hinder the successful utilization of the genetic potential of plant genetic resources^{25,74,75}. Currently, efforts have shifted from the collection of plant genetic resources to the identification of genetic diversity and core germplasm in *P. frutescens*^{74,76–78}. The identified SSR sequences will provide useful information for the effective protection and utilization of genetic resources and the selection of useful genetic resources for *P. frutescens* breeding programs.

Methods

Plant sampling, DNA extracting and sequencing

In December 2023, fresh leaves of *P. frutescens* were collected at the Baiyun Experimental Base of Guangdong Academy of Agricultural Sciences (N23°07′03″, E113°08′36″). The plant had been grown for 6 months. The collected leaves were immediately cooled with liquid nitrogen and stored in an ultra-low temperature freezer at −80 °C. Total DNA was extracted using a plant genomic DNA kit (TSINGKE Biotech, Beijing, China), and the quality and quantity of the extracted DNA were assessed using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) and the Qubit dsDNA HS assay kit on a Qubit 3.0 fluorometer (TSINGKE Biotech, Beijing, China), as well as agarose gel electrophoresis. Subsequently, the extracted DNA samples were shipped in dry ice to Wuhan BMK Tech Solutions Company Limited (Qingdao, China) for SMRT and Illumina sequencing.

Genome assembly and annotation

The raw read data were filtered and corrected using the Pacbio RS II sequencing technology, while sequencing adapters and low-quality sequences were filtered out using the SMRT Analysis (v2.3.0) with default settings to obtain clean reads. Then, the mt genome sequence of *P. frutescens* was extracted from the filtered reads containing chloroplast and mitochondrial genomes. Besides, the NCBI chloroplast genome data were used for BLAST⁷⁹ filtering of reads containing chloroplast genomes, and reads with a match percentage greater than 90% were removed. Following that, the third-generation assembly software Canu⁸⁰ was employed to correct the obtained third-generation data, and Bowtie2 (v2.3.5.1) was adopted to align the second-generation data with the corrected sequences. Using default parameters, Unicycler (v0.4.8)⁸¹ concatenated the aforementioned second-generation data and the corrected third-generation data, yielding a circular mitochondrial genome of *P. frutescens*. The average depth of the assembled mitochondrial genome was 152.25× (Supplementary Table S6). Furthermore, for mitochondrial genome annotation, the following steps were performed: BLAST was used to compare encoded proteins and rRNA with published plant mitochondrial sequences, followed by manual adjustments based on closely related species. tRNA annotation was conducted using tRNAscan-SE with default settings (<http://lowelab.ucsc.edu/tRNAscan-SE/>)⁸². ORFs were annotated using Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Besides, the circular mitochondrial genome map was visualized using the online software Draw Cellelle Genome Map⁸³. Following these steps, the assembly and annotation of the *P. frutescens* mitochondrial genome could be accomplished, providing essential foundational data for further research.

Repeat sequence analysis

Three types of repeat sequences (simple, tandem, and dispersed) were detected in the mitochondrial genome. Simple repeat sequence analysis was performed using the online software MISA (<https://webblast.ipk-gatersleben.de/misa/>)⁴⁴. In this analysis, 10, 5, 4, 3, 3, and 3 repeat sequences were identified with 1, 2, 3, 4, 5, and 6 base pairs, respectively. Tandem repeat sequences with a length exceeding 6 bp and a repeat unit match of over 95 % were detected using the online tool Tandem Repeats Finder v4.09 (<http://tandem.bu.edu/trf/trf.submit.optio ns.html>)⁸⁴. The following parameters were involved: 2 7 7 80 10 50 2000 -f -d -m. Dispersed repeat sequences were identified using BLASTN (v2.10.1) with parameters of word size 7, e-value 1e-5. Redundancy, and tandem repeats were removed. Circos v0.69-5 (<http://circos.ca/software/download/>)⁸⁵ was used for visualizing these repeat sequences. Thorough repeat sequence analyses could provide valuable insights into the repeat elements in the *P. frutescens* mitochondrial genome, illuminating its genomic organization and evolution.

Exploration of chloroplast to mitochondrion DNA transformation, RNA editing, codon usage patterns, and selection pressure

In this investigation, an online cloud platform was utilized to analyze the codon composition of the *P. frutescens* mitochondrial genome, screen for unique CDS, and determine the relative synonymous codon usage (RSCU) of each gene. The chloroplast genome sequence of *P. frutescens* was obtained from the NCBI Organelle Genome Resources database. By employing the BLAST software on NCBI, homologous fragments between the mitochondrial and chloroplast genomes were identified, and the screening criteria were set at match rate ≥ 70%, E-value ≤ 1e-5, and length ≥ 30 bp. The results were visualized using Circos (v0.69-5). The Deepred-mt was used for the prediction of RNA editing of the mitochondrial genomes. Furthermore, genbank files of 9 Lamiaceae species were imported into the Ka/Ks cloud tool to calculate Ka/Ks values for 32 shared proteins. Subsequently, one-way analysis of variance on the Ka/Ks ratios of the 32 protein-coding genes was conducted using the R programming language.

Phylogenetic inference

A phylogenetic tree was plotted using the mitochondrial genomes of 50 species downloaded from NCBI, with *Marchantia paleacea* taken as the outgroup. To ensure comparability, these mitochondrial genomes were re-annotated using the previously described tools. PhyloSuite (v.1.2.1)⁸⁶ was utilized to identify and extract 30 orthologous mitochondrial genes from the analyzed species. Besides, the corresponding nucleotide sequences were aligned using MAFFT (v.7.450)⁸⁷. Subsequently, these aligned sequences were concatenated and utilized to construct the phylogenetic tree. ModelFinder was used to build the best model with default parameters, and the maximum likelihood (ML) analysis with 1000 bootstrap replicates was performed in RAxML (v.8.2.4)⁸⁸. Furthermore, Bayesian inference (BI) analysis was conducted using MrBayes (v.3.2.6)⁸⁹. The Markov chain Monte Carlo method was adopted for 200,000 generations, with tree sampling conducted every 1000 generations for precision. The first 20% of trees were discarded as burn-in, while the remaining were used to generate the consensus tree. For the ASTRAL⁹⁰ analysis, 24 mitochondrial gene trees were inferred separately using RAxML, utilizing the GTRGAMMA model alongside 100 bootstrap replicates. To enhance the accuracy of species-tree inference, branches (<10% bootstrap support) within these gene trees were pruned subsequently, the collapsed gene trees were imported into ASTRAL-III with the default settings. This process yielded an estimated topology of the species tree with branch lengths and local posterior probabilities (LPP) as branch support values.

Statement of plant collection

We hereby declare that *Perilla frutescens* is not a plant species covered by the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. The botanical collection work involved in this research has obtained the necessary permits and approvals from relevant local institutions, and strict compliance with applicable laws and guidelines has been ensured. Moreover, we have minimized the impact on the environment and ecosystems during the collection process, and made every effort to maintain the survival and reproductive capacity of the *P. frutescens* plants.

Conclusions

In this research, the sequencing and assembly of the mitochondrial genome of *P. frutescens* were successfully completed, shedding light on a comprehensive comparison of the organelle genome. This breakthrough is anticipated to provide a broader perspective for the investigation of gene transfer between mitochondrial and plastids. Additionally, through phylogenetic analysis of the mitochondrial genomes of this species and 50 other taxa, the evolutionary status of *P. frutescens* was definitively determined. Overall, the findings offer valuable insights to lay the groundwork for future research on genetic variation, systematic evolution, and breeding of *P. frutescens*, thereby facilitating the cultivation, development, and utilization of this valuable plant.

Data availability

The complete mitochondrial genome of *P. frutescens* generated in this study was submitted to the NGDC database (https://ngdc.cncb.ac.cn/genbase/search/gb/C_AA059311.1) with GeneBase accession number C_AA059311.1. The associated BioProject are PRJNA1108956.

Received: 1 March 2024; Accepted: 10 June 2024

Published online: 17 June 2024

References

- Aliyari Rad, S., Dehghanian, Z., Asgari Lajayer, B., Nobaharan, K. & Astatkie, T. Mitochondrial respiration and energy production under some abiotic stresses. *J. Plant Growth Regul.* **41**, 3285–3299. <https://doi.org/10.1007/s00344-021-10512-1> (2022).
- Maréchal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *N. Phytol.* **186**, 299–317. <https://doi.org/10.1111/j.1469-8137.2010.03195.x> (2010) (PMID: 20180912).
- Gualberto, J. M. *et al.* The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **100**, 107–120. <https://doi.org/10.1016/j.biochi.2013.09.016> (2014) (PMID: 24075874).
- Wang, J. *et al.* Plant organellar genomes: Much done, much more to do. *Trends Plant Sci.* <https://doi.org/10.1016/j.tplants.2023.12.014> (2024).
- Day, D. A. Mitochondrial structure and function in plants. In *Plant Mitochondria: From Genome to Function. Collection-title: Advances in Photosynthesis and Respiration*, vol. 17 (eds. Day, D. A., *et al.*) 1–11 (Springer Netherlands, 2004). https://doi.org/10.1007/978-1-4020-2400-9_1.
- Møller, I. M., Rasmussen, A. G. & Van Aken, O. Plant mitochondria—Past, present and future. *Plant J.* **108**, 912–959. <https://doi.org/10.1111/tj.15495> (2021).
- Lang, Y., Zhang, Z., Jiang, J., Cao, T. & Tian, J. Characterization of the chloroplast genome of *Perilla frutescens* (Lamiaceae), an herb plant from China. *Mitochondrial DNA Part B* **5**, 1550–1551. <https://doi.org/10.1080/23802359.2020.1742597> (2020).
- Zhang, Y. *et al.* Incipient diploidization of the medicinal plant perilla within 10,000 years. *Nat. Commun.* **12**, 5508. <https://doi.org/10.1038/s41467-021-25681-6> (2021).
- Tamura, K. *et al.* A highly contiguous genome assembly of red perilla (*Perilla frutescens*) domesticated in Japan. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes*. **30**, dsac044. <https://doi.org/10.1093/dnares/dsac044> (2023).
- Jin, J.-J. *et al.* GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241. <https://doi.org/10.1186/s13059-020-02154-5> (2020).
- Allio, R. *et al.* Mitofinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol. Ecol. Resour.* **20**, 892–905. <https://doi.org/10.1111/1755-0998.13160> (2020).
- He, W., Xiang, K., Chen, C., Wang, J. & Wu, Z. Master graph: An essential integrated assembly model for the plant mitogenome based on a graph-based framework. *Brief. Bioinform.* **24**, bbac522. <https://doi.org/10.1093/bib/bbac522> (2023) (PMID: 36644898).
- Bi, C. *et al.* Pmat: An efficient plant mitogenome assembly toolkit using low coverage hifi sequencing data. *Hortic. Res.* <https://doi.org/10.1093/hr/uhae023> (2024).
- Kuang, W. M. & Yu, L. Mitogenome assembly strategies and software applications in the genome era. *Yi Chuan = Hereditas* **41**, 979–993. <https://doi.org/10.16288/j.ycz.19-227> (2019).

15. Ma, Q. *et al.* Assembly and comparative analysis of the first complete mitochondrial genome of acer truncatum bunge: A woody oil-tree species producing nervonic acid. *BMC Plant Biol.* **22**, 29. <https://doi.org/10.1186/s12870-021-03416-5> (2022).
16. Roulet, M. E. *et al.* Multichromosomal structure and foreign tracts in the ombrophytum subterraneum (balanophoraceae) mitochondrial genome. *Plant Mol. Biol.* **103**, 623–638. <https://doi.org/10.1007/s1103-020-01014-x> (2020) (PMID: 32440763).
17. Wang, M. *et al.* Mitochondrial genome comparison and phylogenetic analysis of dendrobium (orchidaceae) based on whole mitogenomes. *BMC Plant Biol.* **23**, 586. <https://doi.org/10.1186/s12870-023-04618-9> (2023).
18. Singh, S. *et al.* Elucidating mitochondrial DNA markers of Ogura-based CMS lines in Indian cauliflowers (brassica oleracea var botrytis l) and their floral abnormalities due to diversity in cytonuclear interactions. *Front. Plant Sci.* **12**, 631489. <https://doi.org/10.3389/fpls.2021.631489> (2021).
19. Wu, Z. *et al.* Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in brassica juncea. *BMC Genom.* **20**, 348. <https://doi.org/10.1186/s12864-019-5721-2> (2019).
20. Okazaki, M., Kazama, T., Murata, H., Motomura, K. & Toriyama, K. Whole mitochondrial genome sequencing and transcriptional analysis to uncover an rt102-type cytoplasmic male sterility-associated candidate gene derived from oryza rufipogon. *Plant Cell Physiol.* **54**, 1560–1568. <https://doi.org/10.1093/pcp/pct102> (2013) (PMID: 23852329).
21. Cheng, Q. *et al.* Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in celery (apium graveolens l). *Int. J. Mol. Sci.* **22**, 8584. <https://doi.org/10.3390/ijms22168584> (2021).
22. Melonek, J. *et al.* The genetic basis of cytoplasmic male sterility and fertility restoration in wheat. *Nat. Commun.* **12**, 1036. <https://doi.org/10.1038/s41467-021-21225-0> (2021).
23. Wang, Y. *et al.* Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of photinia serratifolia. *Sci. Rep.* **13**, 770. <https://doi.org/10.1038/s41598-022-24327-x> (2023).
24. Yu, H. *et al.* Phytochemical and phytopharmacological review of *Perilla frutescens* L. (labiatae), a traditional edible-medicinal herb in China. *Food Chem. Toxicol.* **108**, 375–391. <https://doi.org/10.1016/j.fct.2016.11.023> (2017).
25. Park, D. H., Sa, K. J., Lim, S. E., Ma, S. J. & Lee, J. K. Genetic diversity and population structure of *Perilla frutescens* collected from Korea and China based on simple sequence repeats (SSRS). *Genes Genom.* **41**, 1329–1340. <https://doi.org/10.1007/s13258-019-00860-4> (2019) (PMID: 31468347).
26. Liu, S. *et al.* A comprehensive review of the botany, ethnopharmacology, phytochemistry, pharmacology, toxicity and quality control of *Perillae fructus*. *J. Ethnopharmacol.* **304**, 116022. <https://doi.org/10.1016/j.jep.2022.116022> (2023).
27. Wu, X. *et al.* *Perilla frutescens*: A traditional medicine and food homologous plant. *Chin. Herb. Med.* **15**, 369–375. <https://doi.org/10.1016/j.chmed.2023.03.002> (2023).
28. Hou, T. *et al.* *Perilla frutescens*: A rich source of pharmacological active compounds. *Molecules* **27**, 3578. <https://doi.org/10.3390/molecules27113578> (2022).
29. Aochen, C. *et al.* *Perilla frutescens* L.: A dynamic food crop worthy of future challenges. *Front. Nutr.* <https://doi.org/10.3389/fnut.2023.1130927> (2023).
30. Ahmed, H. M. Ethnomedicinal, phytochemical and pharmacological investigations of *Perilla frutescens* (L.) britt. *Molecules* **24**, 102. <https://doi.org/10.3390/molecules24010102> (2018).
31. Zhou, X.-J. *et al.* Structural characterisation and antioxidant activity evaluation of phenolic compounds from cold-pressed *Perilla frutescens* var. arguta seed flour. *Food Chem.* **164**, 150–157. <https://doi.org/10.1016/j.foodchem.2014.05.062> (2014).
32. Ha, T. J. *et al.* Isolation and identification of phenolic compounds from the seeds of *Perilla frutescens* (l.) and their inhibitory activities against α -glucosidase and aldose reductase. *Food Chem.* **135**, 1397–1403. <https://doi.org/10.1016/j.foodchem.2012.05.104> (2012).
33. Meng, L., Lozano, Y. F., Gaydou, E. M. & Li, B. Antioxidant activities of polyphenols extracted from *Perilla frutescens* varieties. *Molecules (Basel, Switzerland)* **14**, 133–140. <https://doi.org/10.3390/molecules14010133> (2008).
34. Lee, D.-S., Noh, B.-S., Bae, S.-Y. & Kim, K. Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Anal. Chim. Acta* **358**, 163–175. [https://doi.org/10.1016/S0003-2670\(97\)00574-6](https://doi.org/10.1016/S0003-2670(97)00574-6) (1998).
35. Li, S.-S. *et al.* Fatty acid composition of developing tree peony (paeonia section moutan dc) seeds and transcriptome analysis during seed development. *BMC Genom.* **16**, 208. <https://doi.org/10.1186/s12864-015-1429-0> (2015).
36. Huang, J. *et al.* Effect of microwave pretreatment of perilla seeds on minor bioactive components content and oxidative stability of oil. *Food Chem.* **388**, 133010. <https://doi.org/10.1016/j.foodchem.2022.133010> (2022).
37. Li, J. *et al.* Complete mitochondrial genome of agrostis stolonifera: Insights into structure, codon usage, repeats, and rna editing. *BMC Genom.* **24**, 466. <https://doi.org/10.1186/s12864-023-09573-1> (2023).
38. Srivastava, S., Avvaru, A. K., Sowpati, D. T. & Mishra, R. K. Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genom.* **20**, 153. <https://doi.org/10.1186/s12864-019-5516-5> (2019).
39. Fang, L. *et al.* Deeprepeat: Direct quantification of short tandem repeats on signal data from nanopore sequencing. *Genome Biol.* **23**, 108. <https://doi.org/10.1186/s13059-022-02670-6> (2022).
40. Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335–343. <https://doi.org/10.1038/nature12985> (2014).
41. Alberts, B. *et al.* The mitochondrion. In *Molecular Biology of the Cell*, 4th edn. (Garland Science, 2002).
42. Skippington, E., Barkman, T. J., Rice, D. W. & Palmer, J. D. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.1504491112> (2015).
43. Putintseva, Y. A. *et al.* Siberian larch (*Larix sibirica ledeb*) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genom.* **21**, 1–12. <https://doi.org/10.1186/s12864-020-07061-4> (2020).
44. Sun, Z. *et al.* Assembly and analysis of the mitochondrial genome of *Prunella vulgaris*. *Front. Plant Sci.* **14**, 1237822. <https://doi.org/10.3389/fpls.2023.1237822> (2023).
45. Nosek, J. & Tomáška, A. Mitochondrial genome diversity: Evolution of the molecular architecture and replication strategy. *Curr. Genet.* **44**, 73–84. <https://doi.org/10.1007/s00294-003-0426-z>.
46. Liu, F. *et al.* Episodic and guanine–cytosine-biased bursts of intragenomic and interspecific synonymous divergence in ajugoideae (lamiaceae) mitogenomes. *New Phytol.* **228**, 1107–1114. <https://doi.org/10.1111/nph.16753> (2020) (PMID: 32558926).
47. Fang, X. & Zhang, S. The complete chloroplast genome sequence of *Rothea myricoides* (hochst.) steane mabb., a traditional medicinal plant. *Mitochondrial DNA Part B* **6**, 2699–2700. <https://doi.org/10.1080/23802359.2021.1966340> (2021).
48. Li, J. *et al.* Assembly of the complete mitochondrial genome of an endemic plant, scutellaria tsinyunensis, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta* **254**, 36. <https://doi.org/10.1007/s00425-021-03684-3> (2021) (PMID: 34302538).
49. Yu, Y., Li, H.-T., Wu, Y.-H. & Li, D.-Z. Correlation analysis reveals an important role of gc content in accumulation of deletion mutations in the coding region of angiosperm plastomes. *J. Mol. Evol.* **89**, 73–80. <https://doi.org/10.1007/s00239-020-09987-5> (2021) (PMID: 33433638).
50. Nardi, F., Carapelli, A. & Frati, F. Repeated regions in mitochondrial genomes: Distribution, origin and evolutionary significance. *Mitochondrion* **12**, 483–491. <https://doi.org/10.1016/j.mito.2012.07.105> (2012).
51. Fay, J. C. & Wu, C.-I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genom. Hum. Genet.* **4**, 213–235. <https://doi.org/10.1146/annurev.genom.4.020303.162528> (2003) (PMID: 14527302).
52. Pazos, F. & Valencia, A. Protein co-evolution, co-adaptation and interactions. *EMBO J.* **27**, 2648–2655. <https://doi.org/10.1038/emboj.2008.189> (2008).

53. Qiao, Y., Zhang, X., Li, Z., Song, Y. & Sun, Z. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* dc. *BMC Genom.* **23**, 664. <https://doi.org/10.1186/s12864-022-08892-z> (2022).
54. Zhang, T. *et al.* Characterization and comparative analyses of mitochondrial genomes in single-celled eukaryotes to shed light on the diversity and evolution of linear molecular architecture. *Int. J. Mol. Sci.* **22**, 2546. <https://doi.org/10.3390/ijms22052546> (2021).
55. Wang, D. *et al.* Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 mya. *Mol. Biol. Evol.* **24**, 2040–2048. <https://doi.org/10.1093/molbev/msm133> (2007) (PMID: 17609537).
56. Sloan, D. B. & Wu, Z. History of plastid DNA insertions reveals weak deletion and at mutation biases in angiosperm mitochondrial genomes. *Genome Biol. Evol.* **6**, 3210–3221. <https://doi.org/10.1093/gbe/evu253> (2014).
57. Turmel, M., Otis, C. & Lemieux, C. Mitochondrion-to-chloroplast DNA transfers and intragenomic proliferation of chloroplast group ii introns in glaucotilopsis green algae (ulotrichales, ulvophyceae). *Genome Biol. Evol.* **8**, 2789–2805. <https://doi.org/10.1093/gbe/evw190> (2016).
58. Mower, J. P., Sloan, D. B. & Alverson, A. J. Plant mitochondrial genome diversity: The genomics revolution. In *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics* (eds. Wendel, J. F. *et al.*) 123–144 (Springer, 2012). <https://doi.org/10.1007/978-3-7091-1130-79>.
59. Wee, C.-C. *et al.* Mitochondrial genome of *Garcinia mangostana* l. variety mesta. *Sci. Rep.* **12**, 9480. <https://doi.org/10.1038/s41598-022-13706-z> (2022).
60. Alverson, A. J. *et al.* Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (cucurbitaceae). *Mol. Biol. Evol.* **27**, 1436–1448. <https://doi.org/10.1093/molbev/msq029> (2010).
61. Kubo, T. & Newton, K. J. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* **8**, 5–14. <https://doi.org/10.1016/j.mito.2007.10.006> (2008) (PMID: 18065297).
62. Wang, Y. *et al.* Empty pericarp21 encodes a novel PPR-DYW protein that is required for mitochondrial RNA editing at multiple sites, complexes I and V biogenesis, and seed development in maize. *PLoS Genet.* **15**, e1008305. <https://doi.org/10.1371/journal.pgen.1008305> (2019).
63. Wang, Y. *et al.* Maize PPR-E proteins mediate RNA C-to-U editing in mitochondria by recruiting the trans deaminase pcw1. *Plant Cell* **35**, 529–551. <https://doi.org/10.1093/plcell/koac298> (2023).
64. Liu, R. *et al.* The DYW-subgroup pentatricopeptide repeat protein ppr27 interacts with zmmorf1 to facilitate mitochondrial RNA editing and seed development in maize. *J. Exp. Bot.* **71**, 5495–5505. <https://doi.org/10.1093/jxb/eraa273> (2020) (PMID: 32531050).
65. Chen, Z. *et al.* Plant mitochondrial genome evolution and cytoplasmic male sterility. *Crit. Rev. Plant Sci.* **36**, 55–69. <https://doi.org/10.1080/07352689.2017.1327762> (2017).
66. Yan, J., Zhang, Q. & Yin, P. RNA editing machinery in plant organelles. *Sci. China Life Sci.* **61**, 162–169. <https://doi.org/10.1007/s11427-017-9170-3> (2018).
67. Hao, W. *et al.* RNA editing and its roles in plant organelles. *Front. Genet.* **12**, 757109. <https://doi.org/10.3389/fgene.2021.757109> (2021).
68. Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B. & Brennicke, A. RNA editing in plants and its evolution. *Annu. Rev. Genet.* **47**, 335–352. <https://doi.org/10.1146/annurev-genet-111212-133519> (2013) (PMID: 24274753).
69. Parvathy, S. T., Udayasuriyan, V. & Bhadana, V. Codon usage bias. *Mol. Biol. Rep.* **49**, 539–565. <https://doi.org/10.1007/s11033-021-06749-4> (2022).
70. Zhao, S. *et al.* Comparative analysis of chloroplast genome of *meconopsis* (papaveraceae) provides insights into their genomic evolution and adaptation to high elevation. *Int. J. Mol. Sci.* **25**, 2193. <https://doi.org/10.3390/ijms25042193> (2024).
71. Chevigny, N., Schatz-Daas, D., Lotfi, F. & Gualberto, J. M. DNA repair and the stability of the plant mitochondrial genome. *Int. J. Mol. Sci.* **21**, 328. <https://doi.org/10.3390/ijms21010328> (2020).
72. Novák, P. *et al.* Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants* **6**, 1325–1329. <https://doi.org/10.1038/s41477-020-00785-x> (2020).
73. Sa, K. J. *et al.* Construction of a core collection of native *Perilla* germplasm collected from South Korea based on SSR markers and morphological characteristics. *Sci. Rep.* **11**, 23891. <https://doi.org/10.1038/s41598-021-03362-0> (2021).
74. Zhang, H. *et al.* Genetic diversity, phylogenetic structure and development of core collections in *meliolus* accessions from a Chinese gene bank. *Sci. Rep.* **9**, 13017. <https://doi.org/10.1038/s41598-019-49355-y> (2019).
75. Salgotra, R. K. & Chauhan, B. S. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* **14**, 174. <https://doi.org/10.3390/genes14010174> (2023).
76. Ma, S. J., Sa, K. J., Hong, T. K. & Lee, J. K. Genetic diversity and population structure analysis in *Perilla frutescens* from northern areas of China based on simple sequence repeats. *Genet. Mol. Res. GMR* <https://doi.org/10.4238/gmr16039746> (2017).
77. Ha, Y. J., Sa, K. J. & Lee, J. K. Identifying SSR markers associated with seed characteristics in *perilla* (*Perilla frutescens* L.). *Physiol. Mol. Biol. Plants* **27**, 93–105. <https://doi.org/10.1007/s12298-021-00933-3> (2021).
78. Park, H., Sa, K. J., Hyun, D. Y., Lee, S. & Lee, J. K. Identifying SSR markers related to seed fatty acid content in *Perilla* crop (*Perilla frutescens* L.). *Plants* **10**, 1404. <https://doi.org/10.3390/plants10071404> (2021).
79. Zaru, R., Orchard, S. & Consortium, U. Uniprot tools: Blast, align, peptide search, and id mapping. *Curr. Protoc.* **3**, e697. <https://doi.org/10.1002/cpz1.697> (2023).
80. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736. <https://doi.org/10.1101/gr.215087.116> (2017).
81. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595. <https://doi.org/10.1371/journal.pcbi.1005595> (2017).
82. Zou, Q. *et al.* Improving trnscan-se annotation results via ensemble classifiers. *Mol. Inform.* **34**, 761–770. <https://doi.org/10.1002/minf.201500031> (2015) (PMID: 27491037).
83. Greiner, S., Lehwark, P. & Bock, R. Organellargenomedraw (ogdraw) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64. <https://doi.org/10.1093/nar/gkz238> (2019).
84. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580. <https://doi.org/10.1093/nar/27.2.573> (1999).
85. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645. <https://doi.org/10.1101/gr.092759.109> (2009).
86. Zhang, D. *et al.* Phylosuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **20**, 348–355. <https://doi.org/10.1111/1755-0998.13096> (2020) (PMID: 31599058).
87. Katoh, K., Rozewicki, J. & Yamada, K. D. Mafft online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166. <https://doi.org/10.1093/bib/bbx108> (2019).
88. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180> (2003).
89. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572–1574. <https://doi.org/10.1093/bioinformatics/btg180> (2003) (PMID: 12912839).
90. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153. <https://doi.org/10.1093/bioinformatics/btu462> (2018).

Acknowledgements

This work was supported by the Guangdong Provincial Science and Technology Plan Project (2019B030316008), Guangdong Provincial Science and Technology Plan Project (2023B1212060038), and Guangdong Provincial Agricultural Research and Technology Promotion Demonstration Project (Yue Cai Nong [2022] No. 16).

Author contributions

R.W., (Ru Wang) and Y.L., (Yongjian Luo) designed the study, defined sampling, obtained samples, and obtained funding. R.W and Y.L., (Yongjian Luo) annotated plastomes and performed comparative and phylogenetic analyses. R.W assembled Illumina sequences. Z.L.,(Zheng Lan) and R.W., participated in the technical guidance and language revision of the paper. R.W., Z.L., and Q.D (Daoshou Qiu) interpreted the results and co-wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-64509-3>.

Correspondence and requests for materials should be addressed to D.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024