

# Complete mitochondrial genome of *Perilla frutescens* (Lamiaceae) insights into structure, Codon usage, repeats, and RNA editing

Ru Wang

Hubei University for Nationalities

Yongjian Luo

Hubei University for Nationalities

Zheng Lan

Heilongjiang Bayi Agricultural University

Daoshou Qiu

qiudaoshou@gdaas.cn

Guangdong Academy of Agricultural Sciences

---

## Article

**Keywords:** Lamioideae, Sequence characteristics, Phylogenetic analysis, chloroplasts, RSCU, RNA editing

**Posted Date:** March 18th, 2024

**DOI:** <https://doi.org/10.21203/rs.3.rs-4003097/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

**Version of Record:** A version of this preprint was published at Scientific Reports on June 17th, 2024. See the published version at <https://doi.org/10.1038/s41598-024-64509-3>.

# Complete mitochondrial genome of *Perilla frutescens* (Lamiaceae) insights into structure, Codon usage, repeats, and RNA editing

Ru Wang<sup>2,+</sup>, Yongjian Luo<sup>2,+</sup>, Zheng Lan<sup>3</sup>, and Daoshou Qiu<sup>1,\*</sup>

<sup>1</sup>Crop Research Institute of Guangdong Academy of Agricultural Sciences/Guangdong Key Laboratory of Crop Genetic Improvement, Guangzhou 510640, China

<sup>2</sup>Hubei Minzu University, School of Forestry and Horticulture, Enshi, 445000, China

<sup>3</sup>Heilongjiang Bayi Agricultural University, Daqing, 163319, China

\*[qiudaoshou@gdaas.cn](mailto:qiudaoshou@gdaas.cn)

<sup>+</sup>these authors contributed equally to this work

## ABSTRACT

*Perilla frutescens* (L.) Britton, a member of the Lamiaceae family, is a versatile plant highly valued for its unique aroma and medicinal properties. Additionally, *P. frutescens* seeds are rich in  $\gamma$ -linolenic acid, holding significant economic importance. While the nuclear and chloroplast genomes of *P. frutescens* have already been documented, the complete Mitochondrial genome sequence has yet to be reported. In this investigation, we conducted the sequencing, annotation, and assembly of the entire Mitochondrial genome of *P. frutescens* using a combination of Illumina and PacBio data. The resulting assembled *P. frutescens* Mitochondrial genome spans 299,551 bp and exhibits a typical circular structure, with 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, with 18 introns present in 8 protein-coding genes. Notably, the codons of the *P. frutescens* Mitochondrial genome display a notable A/T bias. Our analysis also revealed 293 dispersed repeat sequences, 77 simple sequence repeats (SSRs), and 6 tandem repeat sequences. Additionally, RNA editing sites exhibited a preference for the formation of leucine at amino acid editing sites. Furthermore, we identified 70 sequence fragments (12,680 bp) that have been transferred from the chloroplast to the Mitochondrial genome, 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*. Inter-species Ka/Ks results suggested that Ka/Ks <1 for 28 PCGs, indicating that these genes will continue to evolve under purifying selection pressure. The findings of this study will contribute to the enrichment of Mitochondrial genome data for *P. frutescens* and provide a theoretical foundation for future molecular breeding research on *P. frutescens*.

**keywords:** Lamioideae; Sequence characteristics; Phylogenetic analysis; chloroplasts; RSCU; RNA editing;

## Introduction

The Mitochondrial genome is relatively small, highly conserved, and densely packed with genes, while also containing highly variable non-coding regions<sup>1</sup>. In plant research, in addition to its role in energy metabolism, the Mitochondrial genome is often related to plant fertility, as well as playing an important role in plant development and stress resistance<sup>2</sup>. Comparative analyses of Mitochondrial genome sequences, structures, and functions can help reveal the complexity of plant Mitochondrial genomes, providing a basis for deciphering their classification, evolution, and exploring cytoplasmic male sterility mechanisms<sup>3,4</sup>. However, research on complete plant Mitochondrial genomes lags far behind that of complete plastid genomes. While there are nearly 13,000 complete plastid genomes in the NCBI database, there are only 673 complete plant Mitochondrial genomes<sup>5</sup>. This is due to the complex structure of plant Mitochondrial, which are difficult to purify and are frequently interfered with by chloroplasts and other organelles, making their genome assembly challenging. In addition, there are significant differences in the structure and content of plant Mitochondrial genomes, as well as in nucleotide substitution rates and repeat sequences<sup>6,7</sup>. Therefore, a complete description of the plant Mitochondrial genome remains a bottleneck in evolutionary biology, and most plant systematic studies focus on nuclear and plastid genomes<sup>8</sup>. For example, although the nuclear and plastid genomes of *Perilla frutescens* have been sequenced and assembled, its Mitochondrial genome remains unknown<sup>9-11</sup>. 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 GetOrganelle<sup>12</sup>, Mitofiner<sup>13</sup>, GSAT<sup>14</sup>, and PMAT<sup>15</sup> (<https://github.com/bichangwei/PMAT>), have been developed, making Mitochondrial genome sequencing and assembly more accurate and efficient<sup>16</sup>.

In addition, in land plants, Mitochondrial genomes typically contain exogenous genes or fragments due to horizontal gene transfer or intracellular gene transfer<sup>17,18</sup>. As the Mitochondrial genome undergoes continuous recombination, this sequence transfer between genomes also occurs<sup>19</sup>. This can result in cytoplasmic male sterility (CMS), where plants are unable to produce viable pollen grains due to a mismatch between the nuclear and cytoplasmic genomes, but maintain female fertility, as observed in Brassica, rice, and other plants<sup>19&21</sup>. Additionally, CMS is often inherited maternally and is commonly associated with abnormal open reading frames (ORFs) and RNA editing in the Mitochondrial genome<sup>22</sup>. These ORFs are often located near functional Mitochondrial genes and are co-transcribed with them, affecting Mitochondrial function<sup>23</sup>. Hybridization techniques based on CMS have been used to breed offspring with significantly improved yield, stress resistance, and adaptability, making this approach a promising way to maintain crop productivity<sup>24</sup>. Therefore, Mitochondrial genomes are a valuable source of genetic information for plant systematics and necessary cellular process research, with important implications for species evolution research, species identification, and genetic transformation<sup>25</sup>.

*Perilla frutescens* (L.) Britt. is a plant of the Lamiaceae family, and is an annual upright herb that is a traditional dual-purpose medicinal and edible plant with a unique aroma<sup>26</sup>. Its leaves, stems, and fruits can all be used for medicine. *P. frutescens* is widely cultivated in China and Southeast Asia, from Indonesia in the south to Japan in the east.<sup>27</sup> Current research on *P. frutescens* mainly focuses on its pharmacological effects and chemical composition of its compound perillaldehyde<sup>28</sup>. Currently, 400 compounds have been isolated from *P. frutescens* leaves, including terpenoids, flavonoids, alkaloids, steroids, quinines, and phenolic compounds, which have tremendous potential applications<sup>29&31</sup>. People have discovered the pharmacological functions of *P. frutescens*, such as anti-allergic, anti-depressive, lipid-lowering, hepatoprotective, neuroprotective, anti-inflammatory, anti-cancer, antioxidant and antibacterial activities, which are attributed to secondary metabolites present in different parts of the plant<sup>32&35</sup>. In addition, common oil crops such as soybean, rapeseed, peanut, and olive generally have an alpha-linolenic acid content of less than 5%<sup>36,37</sup>. *P. frutescens* seeds have 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, of which  $\alpha$ -linolenic acid has the highest content, reaching 55% 65%<sup>38</sup>. Therefore, *P. frutescens* is considered a natural resource with significant economic and medicinal value. Despite being a plant of significant medicinal and economic value, *P. frutescens* has not yet undergone the assembly and analysis of its mitochondrial genome, a crucial step in understanding its genetic makeup and unlocking its full potential for various applications.

In this study, we conducted sequencing and annotation of the *P. frutescens* mitochondrial genome, and performed a comprehensive analysis of mitochondrial genome features, RNA editing, codon bias, and conducted systematic evolutionary analysis to provide essential background information for further understanding the genetics of this plant. This research lays the groundwork for future studies on molecular breeding strategies for *Perilla frutescens*.

## Results

### Features of the *P. frutescens* mitogenome

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 had an N50 value of 9,597 bp and an N90 value of 4,224 bp. The longest subread had 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 (Figure 1). The nucleotide composition of *P. frutescens* Mitochondrial DNA is 27.3% A, 27.5% T, 22.5% G, and 22.7% C, with a GC content of 45.23%. The *P. frutescens* Mitochondrial genome sequence has been archived in CNGB with the accession number AA059311.1.

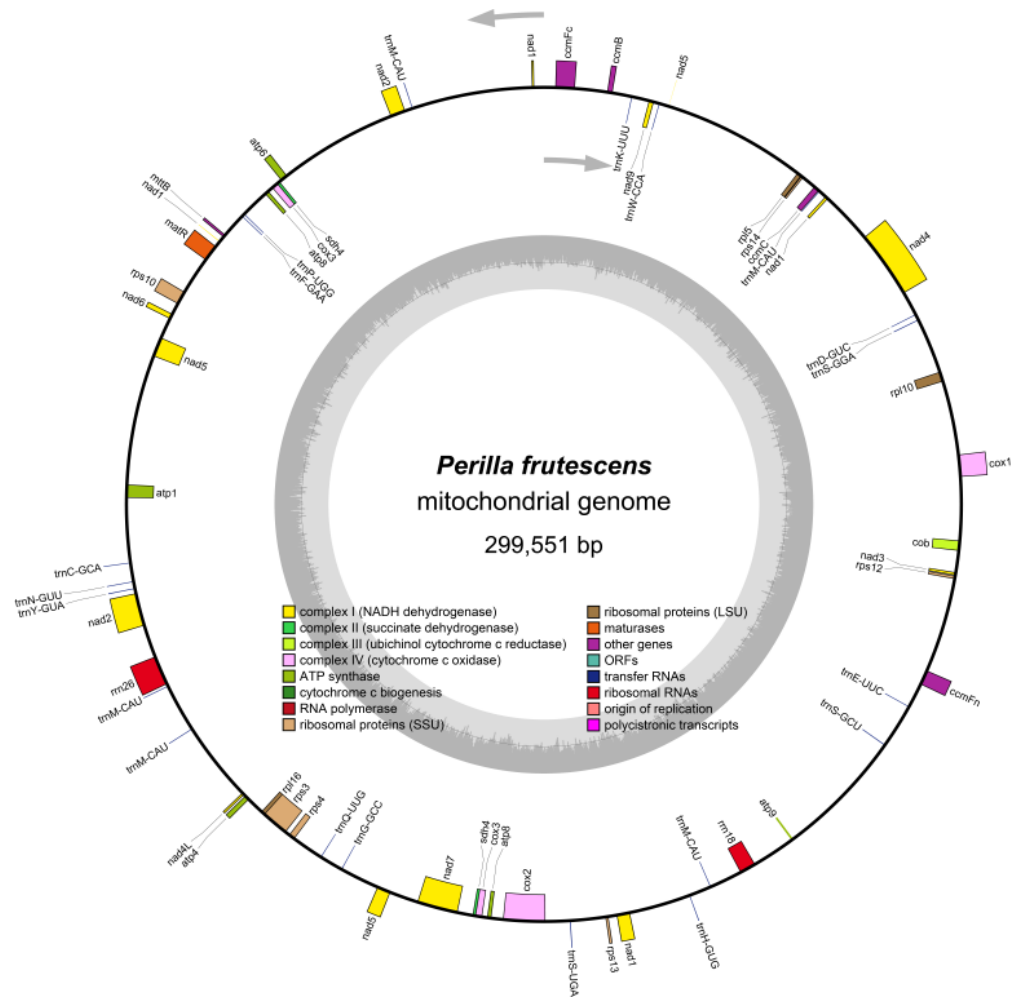


**Figure 1.** *P. frutescens* Mitochondrial genome circular map.

The Mitochondrial genome of *P. frutescens* encompasses a total of 59 genes, consisting of 37 protein-coding genes (PCGs), 20 transfer RNAs (*trnA*s), and 2 ribosomal RNAs (rRNAs) (refer to Table 1). The essential genes include six ATP synthase genes (*atp1*, *atp4*, *atp6*, *atp8*, and *atp9*), nine NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*), four cytochrome c biogenesis genes (*ccmB*, *ccmC*, *ccmFC*, and *ccmFN*), four cytochrome c oxidase genes (*cox1*, *cox2*, and *cox3*), one Maturases gene (*matR*), one membrane transport protein gene (*mtrB*), and one ubiquinol cytochrome c reductase gene (*cob*). The variable genes include three large subunits of ribosomal protein (*rpl16*, *rpl10*, *rpl5*), six small subunits of ribosomal protein (*rps10*, *rps12*, *rps13*, *rps14*, *rps3*, *rps4*), and two succinate dehydrogenases (*sdh4*). Notably, the genes *atp8*, *cox3*, and *sdh4* are present 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* is duplicated five times, while the remaining *trnA* genes exist 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 repeats<sup>51</sup>. In the *P. frutescens* Mitochondrial genome, two different types of dispersed repeats, 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, revealing that these repeats are widely distributed in the intergenic regions of the entire genome (Figure 2A and Supplementary Table S1). A total of 293 dispersed repeat sequences were identified, with a combined 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 (Figure 2B). The longest forward repeat sequence was 5,545 bp, while the longest palindromic repeat sequence was 222 bp. Further analysis showed that 20-29 bp repeats were the most common type, accounting for 51.53% of the

**Table 1.** Classification of genes in the *P. frutescens* Mitochondrial genome

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

Notes: (\*): intron number; (#): Pseudo gene; Gene (2): Number of copies of multi-copy genes;

total repeat occurrences. The presence of these dispersed repeat sequences may 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 genomes<sup>52</sup>. In the mitogenome of *P. frutescens*, a total of 77 SSRs have been identified (Figure 2C). Among these SSRs, tetrameric repeats dominate, 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). Within the monomeric SSRs, adenine (A) repeats have the highest proportion (90.91%), and AG repeats have the highest proportion among dimeric repeats (65.00%) (Figure 2D). None of the repeat types have sequences with a repeat length greater than 20 base pairs. These widely distributed SSRs provide 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 repeats<sup>53</sup>. Six tandem repeat sequences with lengths between 34 and 60 base pairs have been identified in the *P. frutescens* Mitochondrial genome, all of which are located in intergenic regions (Supplementary Table S2).

### Homologous sequence analysis of organellar and nuclear 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 (Figure 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 (Figure 3). These inserted segments are 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 of the Mitochondrial genes *nad2* and *atp6*, while the second longest sequence (820 bp) was transferred from the cp genome's *petA* gene to the intergenic regions of the Mitochondrial genes *nad2* and *atp6*. These transferred sequences are predominantly located in the Mitochondrial genome's intergenic spacers (27), ribosomal RNA genes (20), and transfer RNAs (19). In the *P. frutescens* chloroplast genome, 22 *rrn16S* and *rrn23S* sequences have been inserted into the Mitochondrial genome, with most being transferred to *rrn18* or *rrn26*, except for four segments that have been transferred to the IGS region. In the remaining chloroplast-like sequences, 15 *trnA* genes have been completely transferred (*trnP*-UGG (2), *trnF*-GAA, *trnS*-GGA, *trnT*-GGU, *trnD*-GUC, *trnS*-GCU (2), *trnQ*-UUG, *trnH*-GUG, *trnM*-CAU (2), *trnN*-GUU, *trnN*-GUU, and *trnW*-CCA). Apart from *rps11*, which has been fully transferred to the Mitochondrial genome *cox1* and *rps10*'s IGS, the rest are partial sequences from the chloroplast genome (*psbN*, *psbB*, *clpP*, *psbE*, *petA*, *accD*, *rps4*, *ycf3*, *Psa*, *PsaB*, *psbC*, *rpoB*, *atpA*, *psbA*, *rpl2*, *rpl23*, *ycf2*, *nadF*, and *nadH*) transferred to partial genes or IGS of the Mitochondrial genome.

### Condon usage and RSCU analysis

The total length of 34 PCGs in *P. frutescens* is 34,059 bp, which contains 9,729 codons. The number of encoded codons varies from 8 to 389, with a total of 64 codons encoding 20 amino acids, including the stop codons (UAG (\*), UAA (\*), and UGA (\*)). Among the 64 codons, excluding the three stop codons and the unbiased methionine (Met) and threonine (Thr), 31 codons displayed a preference with RSCU values exceeding 1, indicating a higher priority for these codons. The GCU codon for Alanine (Ala) had the highest frequency of occurrence, with an average RSCU value of 1.61. The AUG codon, which coded for Methionine (Met), exhibited the highest frequency of occurrence, as indicated by an average RSCU value of 1.89. Among

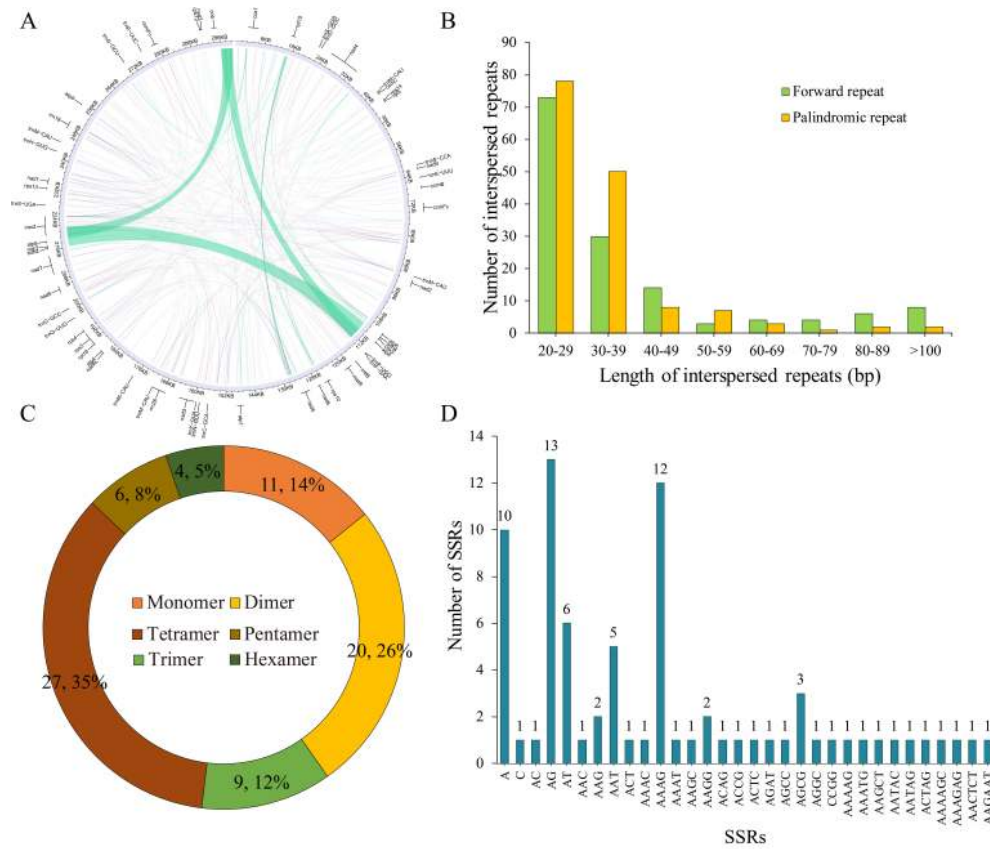


**Figure 2.** Mitochondrial genomic repeat sequences of *P. frutescen*. (A) Distribution of repeat sequences on the genome. (B) Allocation of the lengths of dispersed repeats in the *P. frutescen* mitogenome. (C) Statistical histogram of the *P. frutescen* Mitochondrial genome with repeat sequences of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

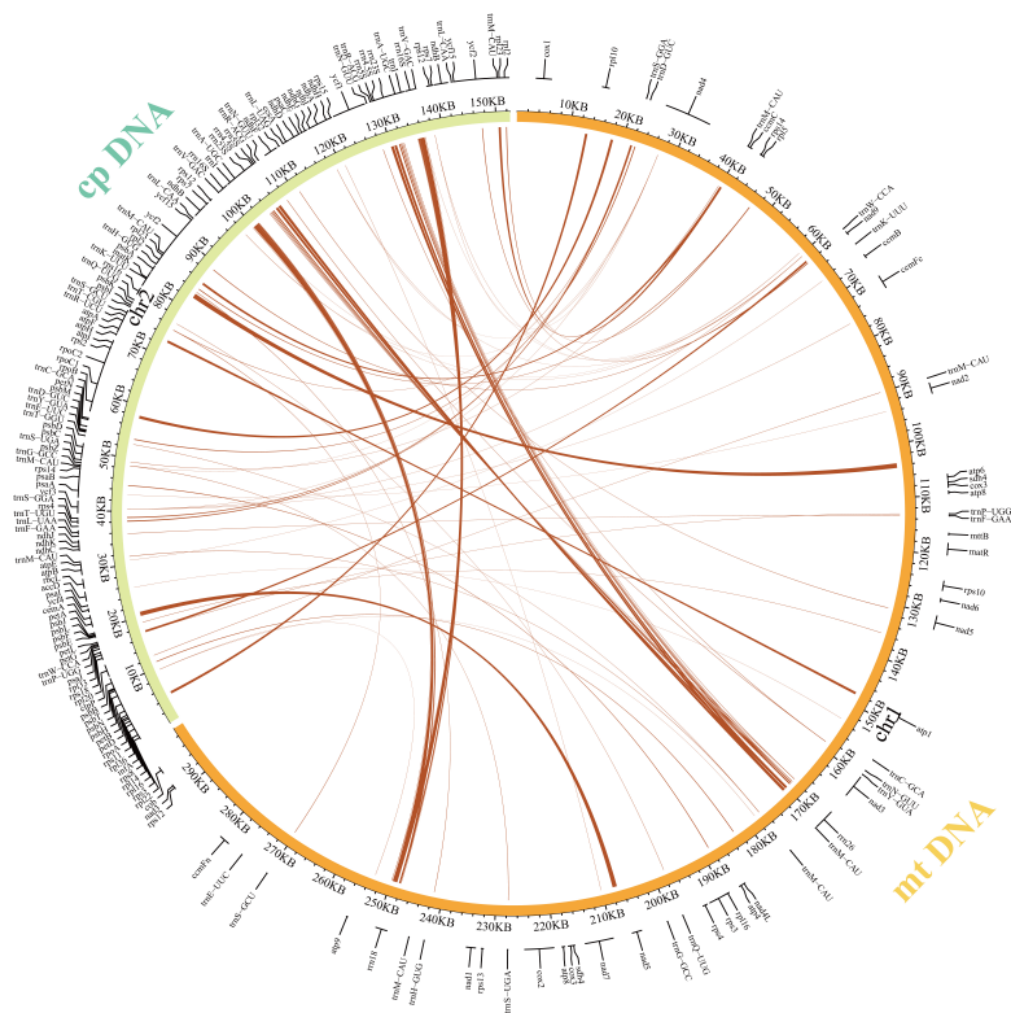


**Figure 3.** Fragments transferred from chloroplasts to Mitochondrial in the *P. frutescens* genome.

the analyzed codons, the remaining 31 codons displayed a relatively low bias, as indicated by their RSCU values less than 1 (Figure 4). Results showed that codons ending with A or U had RSCU values mostly greater than 1.0, while codons ending with C or G had RSCU values mostly less than 1. 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.

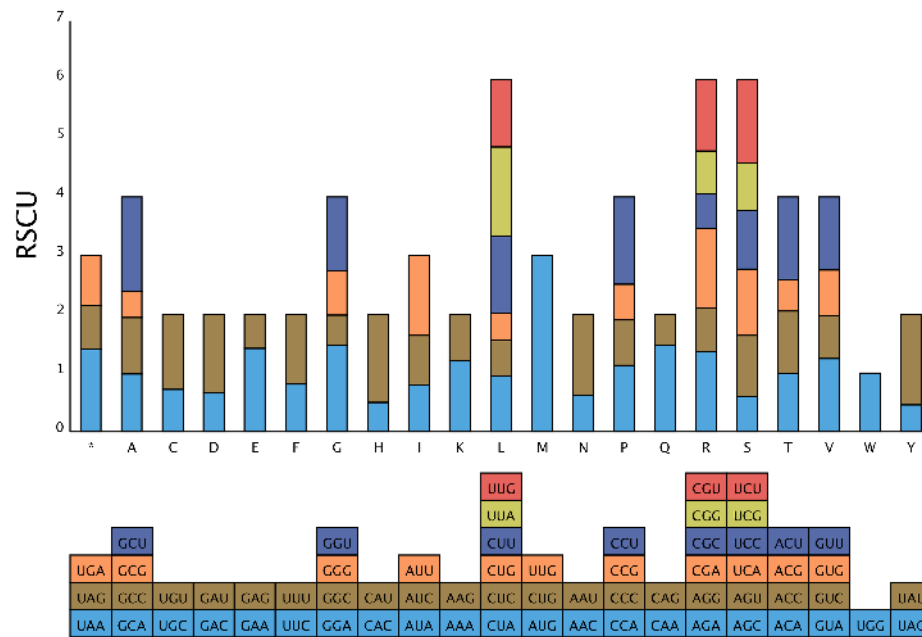


**Figure 4.** Analysis of relative synonymous codon usage (RSCU) in the *P. frutescens* mitogenome.

### Selective pressure analyses

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 species<sup>54</sup>. In genetics, the Ka/Ks ratio serves as a key indicator to determine whether specific protein-coding genes are subject to selection pressure during the course of evolution. When Ks equals Ka, or the Ka/Ks ratio equals 1, it 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, if Ks surpasses Ka or the Ka/Ks ratio is less than 1, it signifies negative selection<sup>55</sup>. The Ka/Ks values of common protein-coding genes in the *P. frutescens* Mitochondrial genome were calculated and compared with nine Lamiaceae species (*Ajuga reptans*, *Lavandula angustifolia*, *Platostoma chinense*, *Pogostemon heyneanus*, *Prunella vulgaris*, *Rotheca serrata*, *Scutellaria barbata*, *Scutellaria franchetiana*, *Vitex trifolia*) (Figure 5). The average Ka/Ks value for 32 identical protein-coding genes was 0.33. Further analysis of specific genes' selective pressure revealed that, except for *ccmFn*, *mttB*, *rps10*, and *matR*, which showed Ka/Ks ratios > 1 compared to the other nine plants, indicating they may be under positive selection. Most genes exhibited a negative selection effect (Ka/Ks < 1) compared to the other nine plants, suggesting that most protein-coding genes in the *P. frutescens* Mitochondrial genome are 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 eukaryotes<sup>17</sup>. In our study, we predicted a total of 559 RNA editing sites within the Mitochondrial genome's 34 protein-coding genes (PCGs) of *P. frutescen*. 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 (Figure 6). Additionally, mttB, ccmC, nad7, nad2, ccmFN, and nad5 each exhibited over 30 editing sites. Conversely, genes like rps14, rpl2, and rps7 had the fewest editing sites, with only two C-to-U edits observed. A total of 56 types of codon transitions were identified, resulting in 25 different amino acid changes. Among these codon transition types, the UCA→UUA transition was the most frequent, 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

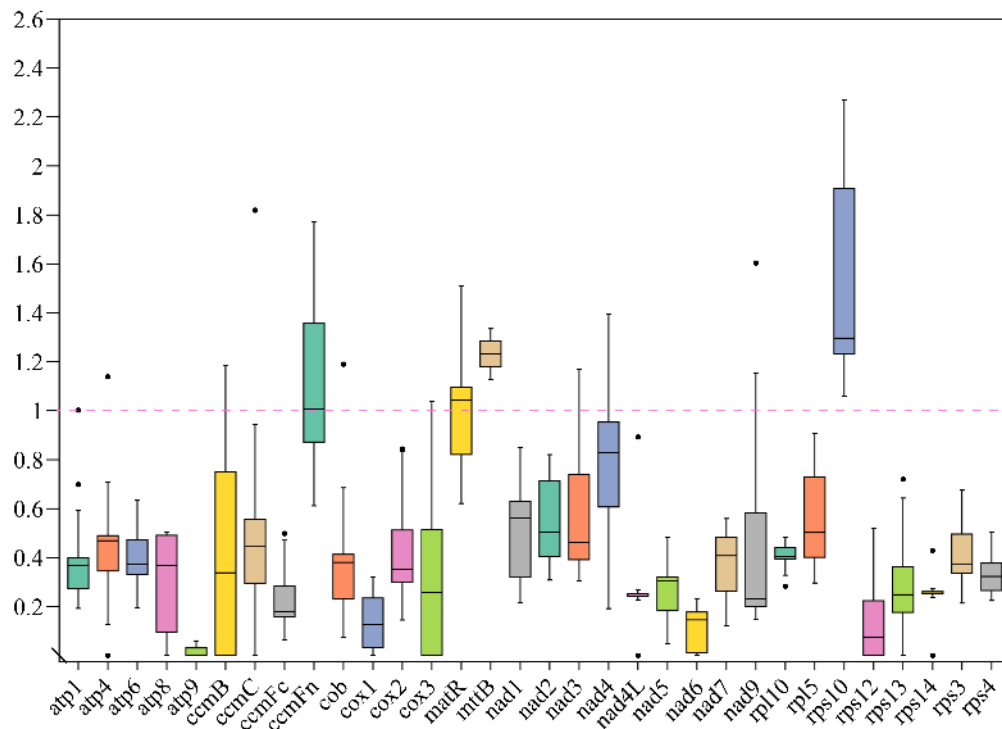


**Figure 5.** Boxplots of the Ka/Ks ratio of *P. frutescens* and 9 other plant species.

acids, accounting for 99.83% of the total changes. 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 play a crucial role in protein translation.

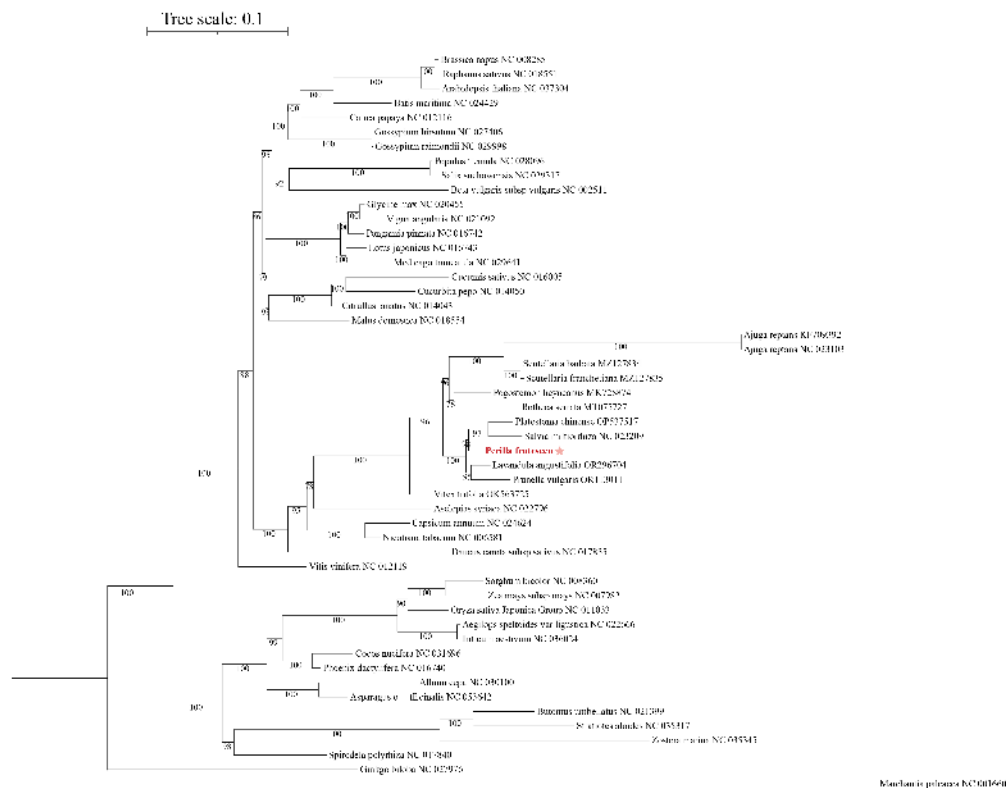
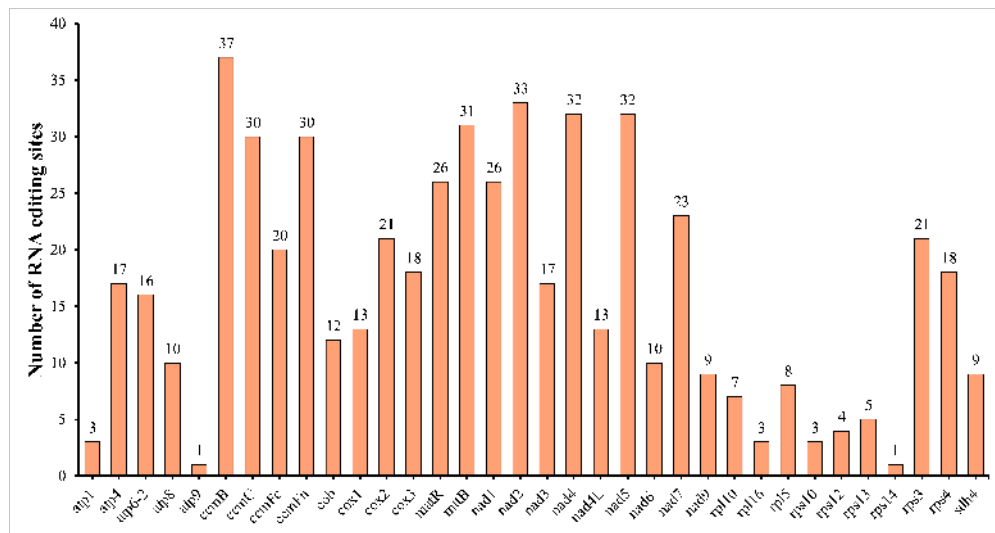
## Phylogenetic analysis

Using conservative Mitochondrial protein-coding genes reconstruction of the phylogenetic tree helps to determine the molecular evolutionary relationships of green plants. To determine the classification status of the *P. frutescens* Mitochondrial genome, *P. frutescens* was analyzed phylogenetically with 50 species, including 35 dicotyledonous plants (including 9 species of Lamiaceae), 13 monocotyledonous plants, 1 gymnosperm, and 1 bryophyte (outgroup), with 23 protein-coding genes corresponding to each species being analyzed.

The results indicate that the phylogenetic tree strongly supports the separation of dicots and monocots, as well as the separation of angiosperms and gymnosperms (Figure 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, Arecales are well resolved in the phylogenetic tree. The relationships in the phylogenetic tree are consistent with the traditional classification relationships of these species, indicating the congruence of traditional and molecular classifications at the family level. Phylogenetic analysis reveals that *P. frutescens* is most closely related to *S. miltiorrhiza* and *Platostoma chinense* genetically. Further acquisition of Mitochondrial genome information from smaller taxa is necessary to determine the phylogenetic relationship of *P. frutescens*.

## Discussion

Mitochondrial generate energy through aerobic respiration in eukaryotic cells and play a key role in regulating important cellular activities such as differentiation, apoptosis, growth, and division<sup>56,57</sup>. Despite their similar functions, the Mitochondrial genomes of different plants exhibit significant differences in size and structure due to the presence of multiple types of repeat sequences and relatively conserved coding sequences, adding to their mystery with linear, fragmented, circular, branched, and multipartite structures<sup>5,17,58,59</sup>. In this study, we combined Illumina and PacBio sequencing technologies to first describe the basic characteristics of the *P. frutescens* genome. We found that the Mitochondrial genome of *P. frutescens* is a circular structure with a total length of 299,551 bp. Its genome size is similar to the published Mitochondrial genome of *Prunella*



*vulgaris* (274,779 bp) in the Lamiaceae family<sup>60</sup>. Compared to other Lamiaceae plant Mitochondrial genomes, the size of *P. frutescens* Mitochondrial genome is relatively small. However, the Mitochondrial genome of *P. vulgaris* is a double-ring structure, indicating significant structural variation and species specificity in Mitochondrial genomes. The GC content is an important factor for evaluating species<sup>61</sup>. The GC content of the mitotic genome of *P. frutescens* is 45.23%, which is highest in the rRNA genes. This result is similar to other Lamiaceae species such as *Scutellaria tsinyunensis* (45.09%)<sup>62</sup>, *Ajuga reptans* (45.26%)<sup>63</sup>, *Rothea serrata* (45.53%)<sup>64</sup>, and *P. vulgaris* (43.92%)<sup>60</sup>. It suggests that the GC content of the mitotic genome is relatively conserved during the evolutionary process in plants. A total of 59 genes were identified in the Mitochondrial genome of *P. frutescens*, including 37 protein-coding genes, 20 *trnA* genes, and 2 rRNA genes. The types and numbers of genes are consistent with most reported circular Mitochondrial genomes in terrestrial plants<sup>63,65,69,70</sup>. The coding region of the *P. frutescens* Mitochondrial genome accounts for only 19.23% of its total length, with over 80% being non-coding regions. This may be the result of the gradual increase of repeat sequences in the Mitochondrial genome during evolution<sup>66</sup>. Genomic comparisons and Ka/Ks analysis of protein-coding genes provide a good theoretical basis for exploring the evolution of the Mitochondrial genome<sup>67,68</sup>. Generally, few protein-coding genes in this study were affected by positive selection, which is similar to existing research reports<sup>6,62,65,69,70</sup>. The Ka/Ks analysis of the Mitochondrial genome indicates that widely present PCGs (protein-coding genes) in Mitochondrial are retained. However, we also found genes with Ka/Ks > 1, such as *ccmFn*, *mttB*, *rps10*, and *matR*. This suggests that positive selection has played a role in the evolutionary history of these coding genes, and high Ka/Ks genes are important in the study of gene selection and evolution in Lamiaceae plants. The phylogenetic relationship of the *P. frutescens* Mitochondrial genome is consistent with traditional classification relationships. However, to obtain the Mitochondrial genome phylogeny of smaller groups, more Mitochondrial genomes of Lamiaceae plants need to be published.

The transfer of genes between chloroplast and Mitochondrial genomes is an important feature of plant Mitochondrial genome evolution<sup>71</sup>. When exogenous genes are inserted into the mitotic genome, they are preferentially inserted into intergenic regions<sup>72,73</sup>. 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 is crucial for 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 sequence<sup>74</sup>, such as in *Garcinia mangostana* L. (1.7%)<sup>75</sup> and *cucurbita pepo* (11.6%)<sup>76</sup>. In this study, 70 migrating segments with a total length of 12,680 bp, accounting for 4.23% of the entire Mitochondrial genome, were identified. In addition, 15 *trnA* genes and one *rps11* from the chloroplast genome were completely transferred. The transfer of *trnA* genes from the chloroplast to the Mitochondrial DNA is common in angiosperms. Interestingly, we also observed that DNA migration often occurs in the inverted repeat regions of the *P. frutescens* chloroplast genome.

RNA editing is an important source of variation in angiosperm Mitochondrial genomes<sup>77</sup>. 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 plays a pivotal role in regulating mitochondrial gene expression and function by inducing alterations in RNA sequences, thereby influencing the translated protein products<sup>78,79,80</sup>. Additionally, the mechanism of crop male sterility has been associated with RNA editing processes, which facilitate RNA splicing by cleaving intron structures. These mechanisms are essential for plant evolution, adaptation, and development<sup>81</sup>, as RNA editing can promote RNA splicing by cleaving intron structures, which is critical for plant evolution, adaptation, and development<sup>82</sup>. In the Mitochondrial genome of *P. frutescens*, 559 RNA editing sites have been 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 may be attributed to their high binding energy, which plays a role in maintaining translation accuracy<sup>82</sup>. By identifying RNA editing sites, researchers can gain valuable insights into the functional implications of predicting newly encoded codons. It is worth noting that RNA editing events produce stop codons in the *rps10* gene, and start codons in three genes (*cox2*, *nad4L*, and *rps10*), which are often related to the production of highly conserved and homologous proteins found in other species, thereby promoting efficient gene expression in Mitochondrial. Furthermore, RNA editing events are least common in ribosomal protein genes like *rps14*, while genes like *nad5*, *mttB*, and *ccmB* exhibit higher frequencies of RNA editing events. These findings suggest that RNA editing plays a crucial role in plant adaptation to environmental changes and signal transduction<sup>83,84</sup>. Codon usage bias is an important factor reflecting the evolution of Mitochondrial genomes<sup>85</sup>. Generally, factors such as mutation, natural selection, and phylogenetic relationships may lead to differences in codon usage preferences<sup>86</sup>. In the Mitochondrial genome of *P. frutescens*, 30 codons with bias (RSCU > 1) were identified, and these codons used more A/T bases. Therefore, in future studies, consideration should be given to codon usage preferences based on Mitochondrial RNA editing sites when creating male sterile materials to promote genetic breeding processes.

Repeat sequences are widely present in Mitochondrial genomes and are usually crucial for molecular recombination, structural variation, and extreme differences in mitotic genome size<sup>87</sup>. Repeat sequences also serve as important sources of information for population development and evolutionary analysis markers<sup>88</sup>. In this study, a total of 293 repeats were observed,

including 141 pairs of reverse complementary repeats and 152 pairs of forward repeats, which may indicate 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) are a special type of tandem repeat sequence that is highly repeatable, polymorphic, commonly co-dominant, and abundant in plant genomes<sup>52</sup>. SSR markers provide useful information for analyzing genetic diversity, genetic relationships, and population structures in crop species germplasm. In our study, 77 SSRs were distributed in different genomic regions of *P. frutescens* mitotic genomes. The most abundant SSR type was tetrameric repeats (27.35%, 27), followed by dimeric repeats (20%), which is consistent with previous studies on Mitochondrial genomes. We also observed a nearly universal A/T bias in the Mitochondrial genome of *P. frutescens*. These SSRs consist of motifs rich in A and T, which is consistent with previous observations and confirms the correlation between the AT content of the complete mitotic genome and SSRs. The collection of genetic resources in *P. frutescens* has been extensive<sup>89</sup>. Although 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 hinder the successful utilization of the genetic potential of plant genetic resources<sup>89,92,91</sup>. Currently, efforts have shifted from the collection of plant genetic resources to the identification of genetic diversity and core germplasm in *P. frutescens*<sup>89,92,95</sup>. 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

### Condon usage and RSCU analysis

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 six months. The 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. The extracted DNA samples were then shipped in dry ice to Wuhan BMK Tech Solutions Company Limited (Qingdao, China) for SMRT and Illumina sequencing.

### Genome Assembly and Annotation

We used Pacbio RS II sequencing technology and filtered and corrected the raw read data. The SMRT Analysis (v2.3.0) with default settings was used to filter out sequencing adapters and low-quality sequences, resulting in clean reads. From the filtered reads containing chloroplast and Mitochondrial genomes, we extracted the mt genome sequence of *P. frutescens*. The NCBI chloroplast genome data was used for BLAST<sup>39</sup> filtering of reads containing chloroplast genomes, and reads with a match percentage greater than 90% were removed. Then, we used the third-generation assembly software Canu<sup>40</sup> to correct the obtained third-generation data. Bowtie2 (v2.3.5.1) was employed to align the second-generation data with the corrected sequences. Using default parameters, Unicycler (v0.4.8)<sup>41</sup> concatenated the aforementioned second-generation data and the corrected third-generation data, resulting in a circular Mitochondrial genome of *P. frutescens*. The average depth of the assembled Mitochondrial genome was 181.87×. 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/>)<sup>42</sup>. ORFs were annotated using Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The circular Mitochondrial genome map was visualized using the online software Draw Cellel Genome Map<sup>43</sup>. By following these steps, the assembly and annotation of the *P. frutescens* Mitochondrial genome can 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/>)<sup>44</sup>. In this analysis, we identified 10, 5, 4, 3, 3, and 3 repeat sequences with 1, 2, 3, 4, 5, and 6 base pairs, respectively. Tandem repeat sequences with a length greater than 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.options.html>)<sup>45</sup> with the following parameters: 2 7 7 80 10 50 2000 -f -d -m. Dispersed repeat sequences were identified using BLASTN (v2.10.1) with parameters: word size 7, e-value 1e-5, remove redundancy, and remove tandem repeats. Circos v0.69-5 (<http://circos.ca/software/download/>)<sup>46</sup> was used for visualizing these repeat sequences. By conducting thorough repeat sequence analyses, valuable insights into the repeat elements present in the Mitochondrial genome of *P. frutescens* can be obtained, shedding light on its genomic organization and evolution.

## Exploration of Chloroplast to Mitochondrion DNA Transformation, RNA Editing, Codon Usage Patterns, and Selection Pressure

We utilized an online cloud platform 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, we identified homologous fragments between the mitochondrial and chloroplast genomes, with screening criteria set at match rate  $\geq 70\%$ , E-value  $\leq 1e-5$ , and length  $\geq 30$  bp. The results were visualized using Circos (v0.69-5). Furthermore, we imported genbank files of 9 Lamiaceae species into the Ka/Ks cloud tool to calculate Ka/Ks values for 32 shared proteins. Subsequently, we conducted one-way analysis of variance on the Ka/Ks ratios of the 32 protein-coding genes using the R programming language.

## Phylogenetic Inference

A phylogenetic tree was constructed using the Mitochondrial genomes of 50 species downloaded from NCBI, with *Marchantia paleacea* used as the outgroup. To ensure comparability, these Mitochondrial genomes were re-annotated using the previously described tools. PhyloSuite (v.1.2.1)47 was utilized to identify and extract 30 orthologous Mitochondrial genes from the analyzed species. The corresponding nucleotide sequences were aligned using MAFFT (v.7.450)48. Subsequently, these aligned sequences were concatenated and used to construct the phylogenetic tree. ModelFinder was used to build the best model with default parameters, and the maximum likelihood (ML) analysis with 1,000 bootstrap replicates was performed in RAxML (v.8.2.4)49. Bayesian inference (BI) analysis was conducted using MrBayes (v.3.2.6)50 with the Markov chain Monte Carlo method for 200,000 generations, sampling the trees every 1000 generations. The first 20% of trees were discarded as burn-in, and the remaining trees were used to generate the consensus tree.

## Data availability

The complete mitochondrial genome of *P. frutescens* generated in this study was submitted to the NGDC database (<https://ngdc.cncb.ac.cn/genbank>) with GeneBase accession number C\_AA059311.1.

## References

1. Li, K. *et al.* The Phylogenetic Relationship of Lamiinae (Coleoptera: Cerambycidae) Using Mitochondrial Genomes. *Genes* **15**, 13 (2024).
2. 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 (2022).
3. MarŹchal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol* **186**, 299–317 (2010).
4. Gualberto, J. M. *et al.* The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**, 107–120 (2014).
5. Wang, J. *et al.* Plant organellar genomes: much done, much more to do. *Trends in Plant Science* (2024) doi:10.1016/j.tplants.2023.12.014.
6. Gualberto, J. M. *et al.* The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**, 107–120 (2014).
7. Day, D. A. Mitochondrial Structure and Function in Plants. in *Plant Mitochondria: From Genome to Function* (eds. Day, D. A., Millar, A. H. & Whelan, J.) vol. 17 1–11 (Springer Netherlands, Dordrecht, 2004).
8. MŹyler, I. M., Rasmusson, A. G. & Van Aken, O. Plant mitochondria – past, present and future. *The Plant Journal* **108**, 912–959 (2021).
9. 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 (2020).
10. Zhang, Y. *et al.* Incipient diploidization of the medicinal plant *Perilla* within 10,000 years. *Nat Commun* **12**, 5508 (2021).
11. Tamura, K. *et al.* A highly contiguous genome assembly of red perilla (*Perilla frutescens*) domesticated in Japan. *DNA Res* **30**, dsac044 (2023).
12. Jin, J.-J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241 (2020).
13. Allio, R. *et al.* MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour* **20**, 892–905 (2020).
14. 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 (2023).
15. Bi, C. *et al.* PMAT: an efficient plant mitogenome assembly toolkit using low coverage HiFi sequencing data. *Horticulture Research* uhae023 (2024) doi:10.1093/hr/uhae023.

16. Kuang, W. M. & Yu, L. [Mitogenome assembly strategies and software applications in the genome era]. *Yi Chuan* **41**, 979–993 (2019).
17. 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 (2022).
18. Roulet, M. E. *et al.* Multichromosomal structure and foreign tracts in the *Ombrophytum subterraneum* (Balanophoraceae) mitochondrial genome. *Plant Mol Biol* **103**, 623–638 (2020).
19. Wang, M. *et al.* Mitochondrial genome comparison and phylogenetic analysis of *Dendrobium* (Orchidaceae) based on whole mitogenomes. *BMC Plant Biology* **23**, 586 (2023).
20. 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 (2021).
21. Wu, Z. *et al.* Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in *Brassica juncea*. *BMC Genomics* **20**, 348 (2019).
22. 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 (2013).
23. Cheng, Q. *et al.* Complete Mitochondrial Genome Sequence and Identification of a Candidate Gene Responsible for Cytoplasmic Male Sterility in Celery (*Apium graveolens* L.). *International Journal of Molecular Sciences* **22**, 8584 (2021).
24. Melonek, J. *et al.* The genetic basis of cytoplasmic male sterility and fertility restoration in wheat. *Nat Commun* **12**, 1036 (2021).
25. Wang, Y. *et al.* Characterization and phylogenetic analysis of the complete mitochondrial genome sequence of *Photinia serratifolia*. *Sci Rep* **13**, 770 (2023).
26. Yu, H. *et al.* Phytochemical and phytopharmacological review of *Perilla frutescens* L. (Labiatae), a traditional edible-medicinal herb in China. *Food and Chemical Toxicology* **108**, 375–391 (2017).
27. 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 (2019).
28. Liu, S. *et al.* A comprehensive review of the botany, ethnopharmacology, phytochemistry, pharmacology, toxicity and quality control of *Perillae Fructus*. *Journal of Ethnopharmacology* **304**, 116022 (2023).
29. Wu, X. *et al.* *Perilla frutescens*: A traditional medicine and food homologous plant. *Chinese Herbal Medicines* **15**, 369–375 (2023).
30. Hou, T. *et al.* *Perilla frutescens*: A Rich Source of Pharmacological Active Compounds. *Molecules* **27**, 3578 (2022).
31. Aochen, C. *et al.* *Perilla frutescens* L.: a dynamic food crop worthy of future challenges. *Frontiers in Nutrition* **10**, (2023).
32. Ahmed, H. M. Ethnomedicinal, Phytochemical and Pharmacological Investigations of *Perilla frutescens* (L.) Britt. *Molecules* **24**, 102 (2018).
33. 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 (2014).
34. Ha, T. J. *et al.* Isolation and identification of phenolic compounds from the seeds of *Perilla frutescens* (L.) and their inhibitory activities against  $\alpha$ -glucosidase and aldose reductase. *Food Chem* **135**, 1397–1403 (2012).
35. Meng, L., Lozano, Y. F., Gaydou, E. M. & Li, B. Antioxidant activities of polyphenols extracted from *Perilla frutescens* varieties. *Molecules* **14**, 133–140 (2008).
36. Lee, D.-S., Noh, B.-S., Bae, S.-Y. & Kim, K. Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Analytica Chimica Acta* **358**, 163–175 (1998).
37. Li, S.-S. *et al.* Fatty acid composition of developing tree peony (*Paeonia* section *Moutan* DC.) seeds and transcriptome analysis during seed development. *BMC Genomics* **16**, 208 (2015).
38. Huang, J. *et al.* Effect of microwave pretreatment of perilla seeds on minor bioactive components content and oxidative stability of oil. *Food Chemistry* **388**, 133010 (2022).
39. Zaru, R., Orchard, S., & UniProt Consortium. UniProt Tools: BLAST, Align, Peptide Search, and ID Mapping. *Curr Protoc* **3**, e697 (2023).
40. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).
41. 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 (2017).
42. Zou, Q. *et al.* Improving tRNAscan-SE Annotation Results via Ensemble Classifiers. *Mol Inform* **34**, 761–770 (2015).

43. 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**, W59â–W64 (2019).
44. Beier, S., Thiel, T., MÃijnc, T., Scholz, U. & Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **33**, 2583â–2585 (2017).
45. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573â–580 (1999).
46. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639â–1645 (2009).
47. 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**, 348â–355 (2020).
48. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**, 1160â–1166 (2019).
49. HÃ¼hler, D., Pfeiffer, W., Ioannidis, V., Stockinger, H. & Stamatakis, A. RAXML Grove: an empirical phylogenetic tree database. *Bioinformatics* **38**, 1741â–1742 (2022).
50. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572â–1574 (2003).
51. Li, J. *et al.* Complete mitochondrial genome of *Agrostis stolonifera*: insights into structure, Codon usage, repeats, and RNA editing. *BMC Genomics* **24**, 466 (2023).
52. Srivastava, S., Avvaru, A. K., Sowpati, D. T. & Mishra, R. K. Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genomics* **20**, 153 (2019).
53. Fang, L. *et al.* DeepRepeat: direct quantification of short tandem repeats on signal data from nanopore sequencing. *Genome Biology* **23**, 108 (2022).
54. Zhang, Z. *et al.* KaKs\_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genomics Proteomics Bioinformatics* **4**, 259â–263 (2006).
55. Li, X. *et al.* Comparative Analysis of the Complete Mitochondrial Genomes of *Apium graveolens* and *Apium leptophyllum* Provide Insights into Evolution and Phylogeny Relationships. *IJMS* **24**, 14615 (2023).
56. Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335â–343 (2014).
57. Alberts, B. *et al.* The Mitochondrion. in *Molecular Biology of the Cell. 4th edition* (Garland Science, 2002).
58. 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. U.S.A.* **112**, (2015).
59. 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 Genomics* **21**, 1â–12 (2020).
60. Sun, Z. *et al.* Assembly and analysis of the mitochondrial genome of *Prunella vulgaris*. *Frontiers in Plant Science* **14**, (2023).
61. 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**, 73â–80 (2021).
62. 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 (2021).
63. Liu, F. *et al.* Episodic and guanine-cytosine-biased bursts of intragenomic and interspecific synonymous divergence in *Ajugoideae* (Lamiaceae) mitogenomes. *New Phytol* **228**, 1107â–1114 (2020).
64. Fang, X. & Zhang, S. The complete chloroplast genome sequence of *Rotheca myricoides* (Hochst.) Steane & Mabb., a traditional medicinal plant. *Mitochondrial DNA Part B* **6**, 2699â–2700 (2021).
65. Sun, Z. *et al.* Assembly and analysis of the mitochondrial genome of *Prunella vulgaris*. *Front. Plant Sci.* **14**, 1237822 (2023).
66. Nardi, F., Carapelli, A. & Frati, F. Repeated regions in mitochondrial genomes: Distribution, origin and evolutionary significance. *Mitochondrion* **12**, 483â–491 (2012).
67. Fay, J. C. & Wu, C.-I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet* **4**, 213â–235 (2003).
68. Pazos, F. & Valencia, A. Protein co-evolution, co-adaptation and interactions. *EMBO J* **27**, 2648â–2655 (2008).
69. Qiao, Y., Zhang, X., Li, Z., Song, Y. & Sun, Z. Assembly and comparative analysis of the complete mitochondrial genome of *Bupleurum chinense* DC. *BMC Genomics* **23**, 664 (2022).
70. 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 (2021).
71. Wang, D. *et al.* Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Mol Biol Evol* **24**, 2040â–2048 (2007).

72. 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 (2014).
73. Turmel, M., Otis, C. & Lemieux, C. Mitochondrion-to-Chloroplast DNA Transfers and Intragenomic Proliferation of Chloroplast Group II Introns in Gloeotilopsis Green Algae (Ulotrichales, Ulvophyceae). *Genome Biol Evol* **8**, 2789–2805 (2016).
74. 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., Greilhuber, J., Dolezel, J. & Leitch, I. J.) 123–144 (Springer, Vienna, 2012). doi:10.1007/978-3-7091-1130-7\_9.
75. Wee, C.-C. *et al.* Mitochondrial genome of *Garcinia mangostana* L. variety Mesta. *Sci Rep* **12**, 9480 (2022).
76. 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 (2010).
77. Kubo, T. & Newton, K. J. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* **8**, 5–14 (2008).
78. 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 (2019).
79. 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 (2023).
80. 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 (2020).
81. Chen, Z. *et al.* Plant Mitochondrial Genome Evolution and Cytoplasmic Male Sterility. *Critical Reviews in Plant Sciences* **36**, 55–69 (2017).
82. Yan, J., Zhang, Q. & Yin, P. RNA editing machinery in plant organelles. *Sci. China Life Sci.* **61**, 162–169 (2018).
83. Hao, W. *et al.* RNA Editing and Its Roles in Plant Organelles. *Front. Genet.* **12**, 757109 (2021).
84. Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B. & Brennicke, A. RNA editing in plants and its evolution. *Annu Rev Genet* **47**, 335–352 (2013).
85. Parvathy, S. T., Udayasuriyan, V. & Bhadana, V. Codon usage bias. *Mol Biol Rep* **49**, 539–565 (2022).
86. Zhao, S. *et al.* Comparative Analysis of Chloroplast Genome of *Meconopsis* (Papaveraceae) Provides Insights into Their Genomic Evolution and Adaptation to High Elevation. *International Journal of Molecular Sciences* **25**, 2193 (2024).
87. Chevigny, N., Schatz-Daas, D., Lotfi, F. & Gualberto, J. M. DNA Repair and the Stability of the Plant Mitochondrial Genome. *International Journal of Molecular Sciences* **21**, 328 (2020).
88. Novák, P. *et al.* Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants* **6**, 1325–1329 (2020).
89. 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 (2021).
90. Zhang, H. *et al.* Genetic diversity, phylogenetic structure and development of core collections in *Melilotus* accessions from a Chinese gene bank. *Sci Rep* **9**, 13017 (2019).
91. Salgotra, R. K. & Chauhan, B. S. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes (Basel)* **14**, 174 (2023).
92. 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 Genomics* **41**, 1329–1340 (2019).
93. 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* **16**, (2017).
94. 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 (2021).
95. 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 (Basel)* **10**, 1404 (2021).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (31860073, 81303169) and the Project of opening Fund for the Year 2022 of Hubei Key Laboratory for the Protection and Utilization of Biological Resources (PT012212).

## Author contributions statement

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

500 sequences. Z.L.,(Zheng Lan) and R.W., participated in the technical guidance and language revision of the paper. R.W., Z.L.,  
501 and Q.D (Daoshou Qiu) interpreted the results and co-wrote the manuscript.

## 502 **Competing interests**

503 The authors declare no competing interests.

## 504 **Statement of plant collection**

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

## 511 **Additional information**

512 **Supplementary Information** Te online version contains supplementary material available at :Supplementary Table S1:  
513 Dispersed repeats Statistics.; Supplementary Table S2: Tandem repeat sequences Statistics.; Supplementary Table S3: The  
514 distribution of RNA editing sites in the mt PCGs of *P. frutescens*.; Supplementary Table S4: RNA editor Statistics.

515 **Correspondence** and requests for materials should be addressed to Y.L.

516 **Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints)

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

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable.xlsx](#)