



Article

De Novo Hybrid Assembly of the *Salvia miltiorrhiza* Mitochondrial Genome Provides the First Evidence of the Multi-Chromosomal Mitochondrial DNA Structure of *Salvia* Species

Heyu Yang ^{1,2}, Haimei Chen ², Yang Ni ², Jingling Li ² , Yisha Cai ², Binxin Ma ², Jing Yu ², Jiehua Wang ^{1,*} and Chang Liu ^{2,*}

¹ School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China

² Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China

* Correspondence: jiehuaawang@tju.edu.cn (J.W.); cliu@implad.ac.cn (C.L.); Tel.: +86-022-87402072 (J.W.); +86-10-57833111 (C.L.); Fax: +86-022-27407956 (J.W.); +86-10- 62899715 (C.L.)



Citation: Yang, H.; Chen, H.; Ni, Y.; Li, J.; Cai, Y.; Ma, B.; Yu, J.; Wang, J.; Liu, C. *De Novo* Hybrid Assembly of the *Salvia miltiorrhiza* Mitochondrial Genome Provides the First Evidence of the Multi-Chromosomal Mitochondrial DNA Structure of *Salvia* Species. *Int. J. Mol. Sci.* **2022**, *23*, 14267. <https://doi.org/10.3390/ijms232214267>

Academic Editor: Frank M. You

Received: 29 September 2022

Accepted: 15 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Salvia miltiorrhiza* has been an economically important medicinal plant. Previously, an *S. miltiorrhiza* mitochondrial genome (mitogenome) assembled from Illumina short reads, appearing to be a single circular molecule, has been published. Based on the recent reports on the plant mitogenome structure, we suspected that this conformation does not accurately represent the complexity of the *S. miltiorrhiza* mitogenome. In the current study, we assembled the mitogenome of *S. miltiorrhiza* using the PacBio and Illumina sequencing technologies. The primary structure of the mitogenome contained two mitochondrial chromosomes (MC1 and MC2), which corresponded to two major conformations, namely, Mac1 and Mac2, respectively. Using two approaches, including (1) long reads mapping and (2) polymerase chain reaction amplification followed by Sanger sequencing, we observed nine repeats that can mediate recombination. We predicted 55 genes, including 33 mitochondrial protein-coding genes (PCGs), 3 rRNA genes, and 19 tRNA genes. Repeat analysis identified 112 microsatellite repeats and 3 long-tandem repeats. Phylogenetic analysis using the 26 shared PCGs resulted in a tree that was congruent with the phylogeny of Lamiales species in the APG IV system. The analysis of mitochondrial plastid DNA (MTPT) identified 16 MTPTs in the mitogenome. Moreover, the analysis of nucleotide substitution rates in Lamiales showed that the genes *atp4*, *ccmB*, *ccmFc*, and *mttB* might have been positively selected. The results lay the foundation for future studies on the evolution of the *Salvia* mitogenome and the molecular breeding of *S. miltiorrhiza*.

Keywords: *Salvia miltiorrhiza*; lamiales; mitogenome; multi-chromosomal structure; MTPT

1. Introduction

Salvia miltiorrhiza (Danshen) is among the most economically important medicinal plants, and its products have been widely used for centuries to treat various human diseases, such as cardiovascular disease, dysmenorrhea, and amenorrhea [1]. The annual production of this species reaches 20,000 tons [2], and there have been 869 formulas containing *S. miltiorrhiza* that were stored in the Encyclopedia of Traditional Chinese Medicine (ETCM) database [3]. As a result, molecular breeding and synthetic biology studies have been conducted extensively to improve the yield of *S. miltiorrhiza* materials. Several nuclear genome sequences of *S. miltiorrhiza* have been reported [4–7]. In addition, a mitochondrial genome (mitogenome) has been reported before [8,9]. The mitogenome is a single circular molecule based on short reads generated using the next-generation DNA sequencing technologies. In numerous plant mitochondrial genomes, one or more pairs of repeats can act as inter- or intramolecular recombination sites and generate multiple alternative

arrangements (isoforms) [10]. As a result, a previously reported mitogenome cannot capture the complete spectrum of isoforms resulting from repeat-mediated recombination.

By September 2022, 499 complete plant mitogenomes have been released in the GenBank [11]. These plant mitogenomes vary considerably in terms of genome size and architecture, gene composition, RNA editing potential, mutation rates of the protein-coding genes (PCGs), and the rate of recombination across different types of repeats [12–17]. The available mitogenomes will aid in the reconstruction of the ancestral angiosperm mitogenome and the understanding of its subsequent evolutionary changes [18,19], which lead to the currently extraordinary diversity of mitogenomes [20,21].

Analyses of the available mitogenomes revealed significant variations in the genes, introns, and RNA-editing capacities of plant mitogenomes [22,23]. Although the mitogenome of *Viscum* contains only 19 PCGs, the mitogenome of the liverwort genera comprises 39–42 PCGs [13,24]. For RNA-editing events, 427 and 441 C-to-U RNA-editing sites were found in the open reading frames of the mitogenomes of rapeseed and *Arabidopsis thaliana*, respectively, and 225 C-to-U RNA-editing sites were identified in the PCG regions of the *S. miltiorrhiza* mitogenome [9,23].

Research of the plant mitogenomes also focused on the intricate structure and revealed significant inter-specific variations in the structure of plant mitochondria [25]. The mitogenome of *Arabidopsis thaliana* was assembled as a standard and single circular structure [26], whereas, in another species (*Silene conica*), the mitogenome showed complex multichromosomal structures [15]. Multichromosomal mitochondrial genomes have also been reported in numerous plants, including cucumber (*Cucumis sativus*) [17], which harbors three circular chromosomes (1556, 84, and 45 Kb). Recently, the multichromosomal structure has also been observed in the mitochondrial genome of onion, and two circular chromosomes have been obtained (173,131 and 143,157 bp) [27].

S. miltiorrhiza is a member of the order Lamiales. This enormous angiosperm order is a member of the asterid clade and contains nearly 23,810 species, 1059 genera, and 24 families. Most species of this order, such as sesame, olive, jasmine, psyllium, and lavender, are known for their essential economical value [28,29]. However, the available mitogenomes from Lamiales plant lineages are limited, preventing the in-depth understanding of the mitogenome evolution in this group. The mitogenomes of only nine Lamiales plants, namely, *Ajuga reptans* (NC_023103.1) [30], *Castilleja paramensis* (NC_031806.1), *Boea hygrometrica* (NC_016741.1) [31], *Erythranthe lutea* (NC_018041.1) [20], *Hesperelaea palmeri* (NC_031323.1) [32], *Salvia miltiorrhiza* (NC_023209.1), *Utricularia reniformis* [33], *Rothea serrata* (NC_049064.1), and *Scutellaria tsinyunensis* (MW553042.1) [34], have been reported.

The *S. miltiorrhiza* mitogenome has been reported as a single circular chromosome using the Illumina sequencing reads alone [9]. Here, we assembled the *S. miltiorrhiza* mitogenome based on the combination of short reads generated from Illumina technology and long reads from PacBio technology. The newly assembled mitogenome showed significantly different arrangements. The results suggest that the dominant form of the *S. miltiorrhiza* mitogenome contains two subgenomic chromosomes, and nine pairs of repeat sequences can mediate recombination, leading to a large collection of minor conformations.

2. Results

2.1. Structure Analysis of the *S. miltiorrhiza* Mitogenome

In total, 19 and 16 Gb of raw sequence data were generated from the PacBio RS (Menlo Park, CA, USA) and Illumina (San Diego, CA, USA) platforms, respectively (Table S1). We extracted the mitochondrial short reads using GetOrganelle and conducted de novo assembly of the extracted reads using the Unicycler software [35], resulting in a unitig graph (Figure 1A). The unitig graph contained seven double bifurcating structures (DBS) (bs01–07). Each DBS has conformations 1, 2, 3, and 4 (C1, C2, C3, and C4). We used Unicycler software to resolve the DBSs. Unicycler mapped the long reads to the DBSs' four conformations and identified those that were supported by more reads. These conformations were then used in the final assembly. The results of the Unicycler analysis were then loaded into the Bandage

software. By using the “Merge all possible nodes” module of the Bandage software, we finally obtained two chromosomes of the mitogenome of *S. miltiorrhiza* (Figure 1B).

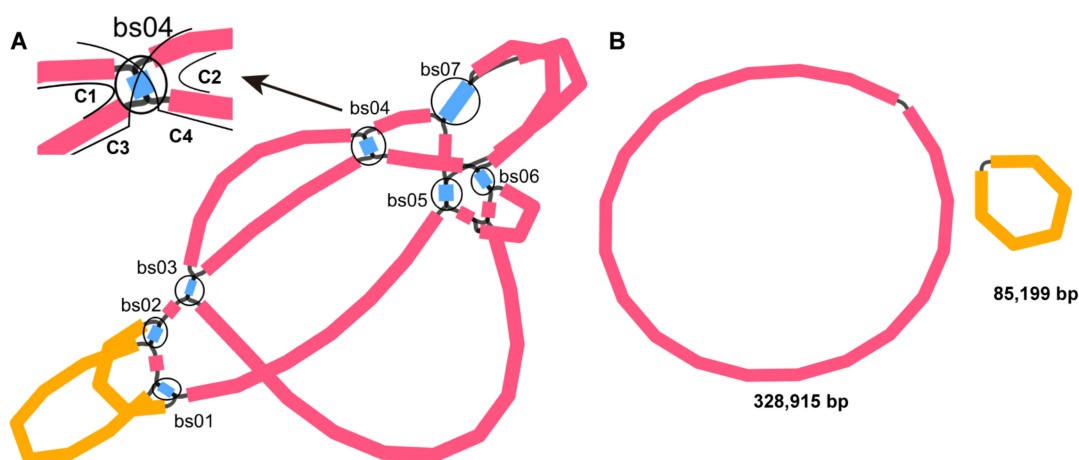


Figure 1. Schematic of the assembly steps of the *S. miltiorrhiza* mitogenome. (A) Unitig graph of the *S. miltiorrhiza* mitogenome was obtained from de novo assembly of Illumina reads with Unicycler. The unitig graph contained seven contigs (blue) that formed DBSs (bs01–07, black circle). Each DBS has four conformations (C1, C2, C3, and C4), as shown in the top left corner with bs04, as an example. (B) Schematic graph of MC1 (pink circle) and MC2 (orange circle) of *S. miltiorrhiza* after the DBSs were resolved by long reads. The contigs shown in pink and yellow correspond to chromosomes 1 and 2, respectively.

In parallel, we assembled the plastome sequence of *S. miltiorrhiza* using GetOrganelle. The size of the plastome was 151,394 bp, which is close to the size of the published chloroplast genome (151,328 bp) [36]. We compared the plastome sequences obtained here and the one published before and discovered their high similarity (Figure S1). The plastome sequence assembled from this study is provided in Supplementary File S1.

To confirm that the DBS conformations selected by Unicycler were supported by most long reads, we constructed sequences corresponding to the four conformations of bs01–07 (Supplementary File S2). We mapped the long PacBio reads to these sequences. Figure S2a–g show the mapping results. We also counted the reads mapped to each of the four conformations, and the results are shown in Table 1. We denoted the conformations found in the mitogenome assembly as major conformations (Mac1 and Mac2) and those not found in the mitogenome assembly as minor conformations (Mic1 and Mic2).

To validate the assembly result, we first mapped the long reads to the mitogenome sequences and obtained average coverage depths of 334.04 and 288.20 for mitogenome chromosomes 1 and 2 (MC1 and MC2), respectively (Figure S3a,b). The entire chromosomes were covered well by the reads, with the lowest coverage being 257. In parallel, we mapped the short reads to the mitogenome sequences, obtaining average coverage depths of 965.88 and 1153.56 for MC1 and MC2, respectively (Figure S3c,d).

Previously, an *S. miltiorrhiza* mitogenome assembled from Illumina short reads was published (NC_023209) [9]. We compared the sequences of the MC1 and MC2 obtained in this study with the published ones. The two assemblies of the *S. miltiorrhiza* mitogenome differed significantly (Figure S4). Numerous rearrangements were observed between the two assemblies, and the largest collinear block was 66,778 bp in length. To determine whether the PacBio reads supported NC_023209, we mapped the PacBio reads to the sequence of NC_023209 and discovered numerous regions not covered by PacBio reads, indicating the potential misassembly of this sequence (Figure S5).

Table 1. Results for mapping PacBio long reads to the four possible conformations associated with eleven HSPs (r01–r11) and their corresponding seven DBSs (bs01–bs07). “MC1/2”: mitogenome chromosome 1/2. ^a: Validated successfully with PCR; ^b: all positions are based on MC1, except those marked with MC2. The recombination frequency was calculated as the number of reads mapped to Mic1 and Mic2 divided by that mapped to all found conformations. NA: not applicable.

ID of the HSP	ID of the DBS	Identity (%)	Alignment Length	Numbers of Mis-matches	Numbers of Gap Openings	Positions of Repeat Copy 1 ^b		Positions of Repeat Copy 2		E-Value	Type	Numbers of Long Reads Mapped to Each Conformation				Recombination Frequency (%)
						Start	End	Start	End			Mac1	Mac2	Mic1	Mic2	
r01 ^a	bs02	100	127	1	1	328,831	328,915	85,115 (MC2)	85,199 (MC2)	3.51×10^{-37}	direct	70	46	5	3	6.45%
r02 ^a	bs05	99.853	682	1	0	279,954	280,635	235,468	234,787	0	inverted	31	28	5	11	21.33%
r03 ^a	bs04	100	369	0	0	223,308	223,676	27,513	27,145	0	inverted	28	22	5	3	13.79%
r04 ^a	bs06	95.312	192	5	4	232,877	233,068	78,659	78,846	1.53×10^{-80}	direct	24	17	1	1	4.65%
r05 ^a	bs03	100	87	0	0	185,367	185,453	1433	1519	2.72×10^{-38}	direct	65	60	2	1	2.34%
r06 ^a	NA	95.139	144	5	2	191,145	191,287	141,672	141,814	9.49×10^{-58}	direct	46	59	6	5	9.48%
r07 ^a	NA	91.176	136	11	1	175,216	175,351	209,605	209,471	5.80×10^{-45}	inverted	30	37	3	2	6.94%
r08 ^a	NA	98.649	74	1	0	300,772	300,845	19,641	19,714	2.13×10^{-29}	direct	33	29	11	3	18.42%
r09 ^a	NA	98.305	59	0	1	311,788	311,846	252,783	252,840	1.67×10^{-20}	direct	41	48	9	3	11.88%
r10	bs07	99.966	5835	1	1	269,937	275,770	42,128	36,294	0	inverted	15	10	15	12	51.92%
r11	bs01	94.828	116	5	1	27,842	27,957	326,567 (MC2)	326,681 (MC2)	1.54×10^{-44}	direct	61	65	0	1	0.79%

2.2. Gene Content of *S. miltiorrhiza* Mitogenome

The length of the mitogenome chromosome 1 and 2 (MC1 and MC2) was 328,915 bp and 85,199 bp, respectively, and the GC content was 44.62% and 44.43%, respectively. The mitogenome of *S. miltiorrhiza* contained a total of 33 PCGs, 3 rRNA genes (*rrn5*, *rrn18*, and *rrn26*), and 19 tRNA genes (Figure 2, Table 2). In angiosperm mitogenomes, there are a set of 24 core protein-coding genes mostly coding respiratory proteins and 17 variable protein genes coding ribosomal proteins [37]. The *S. miltiorrhiza* mitogenome included the entire set of 24 core PCGs and 9 out of 17 variable PCGs (Figure 3 and Table 2). The DNA sequences of two chromosomes of the mitogenome of *S. miltiorrhiza* are provided as fasta files, along with the annotation information provided as Genbank files (MN585275.1 and MN585276.1) and also deposited in the Figshare repository (<https://doi.org/10.6084/m9.figshare.21195841> (accessed on 28 June 2022)).

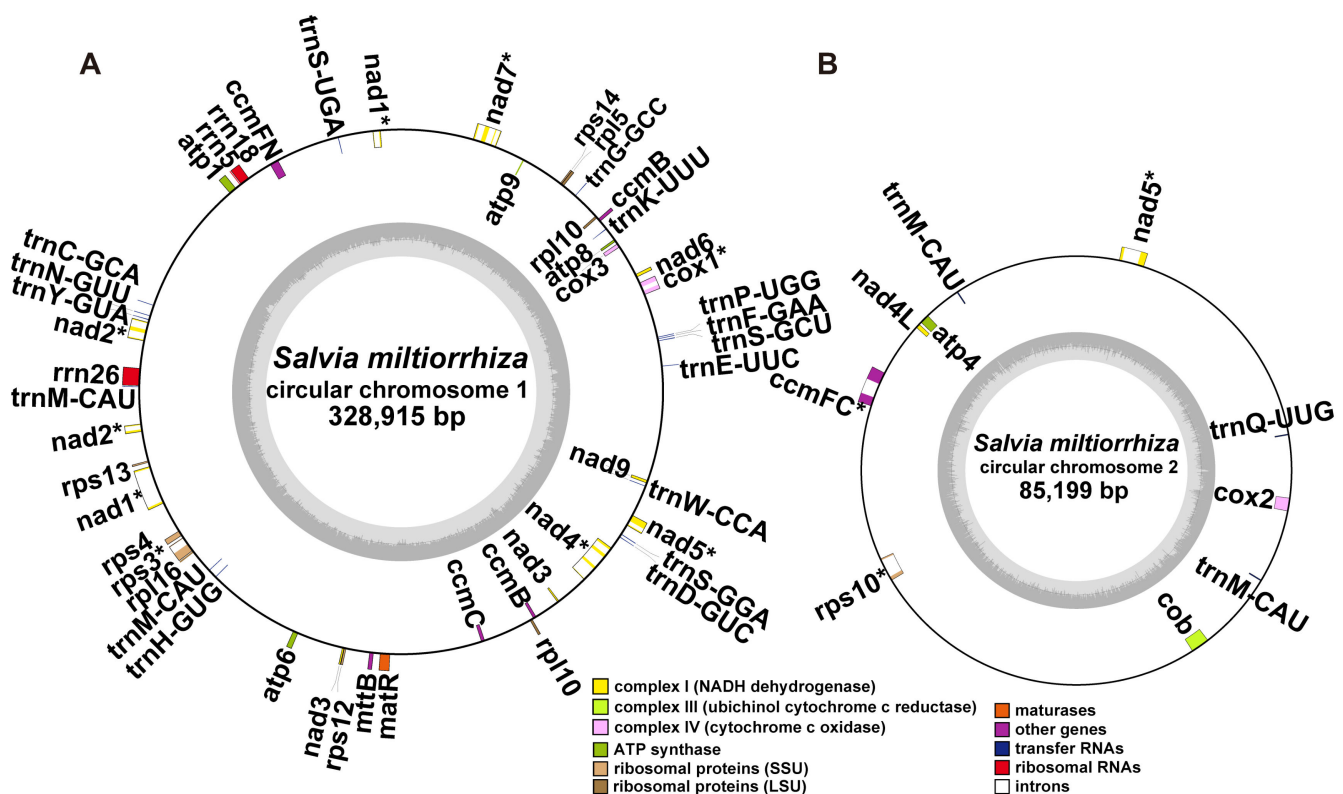


Figure 2. Schematic of the circular MC1 (A) and MC2 (B) of *S. miltiorrhiza*. The graph was drawn using PMGView (<http://www.1kmpg.cn/pmgview> (accessed on 26 June 2022)). Genes shown on the inside were on the negative strand, whereas those on the outside were on the positive strand. Genes with introns were highlighted using “*”. The gray circle represents the GC contents. The circle inside the GC content graph marks the 50% threshold. The colors indicate different functional categories shown in the legend.

Five (*cox1*, *rps3*, *cox2*, *ccmFc*, and *rps10*), two (*nad4* and *nad7*), and three (*nad1*, *nad2*, and *nad5*) PCGs contained two, four, and five exons, respectively. The remaining 23 PCGs had no intron. We counted the gene contents from 10 Lamiaceae mitogenomes and observed the absence of *sdh3* in the mitogenomes of Lamiaceae (Figure 3). The implication of this observation remains to be determined.

Table 2. Genes predicted in the mitogenome of *S. miltiorrhiza*.

Group of Genes		Name of Genes
Core genes	ATP synthase	<i>atp1, atp4, atp6, atp8, atp9</i>
	Cytochrome c biogenesis	<i>ccmB, ccmC, ccmFc^a, ccmFn</i>
	Ubichinol cytochrome c reductase	<i>cob</i>
	Cytochrome c oxidase	<i>cox1^a, cox2^a, cox3</i>
	Maturases	<i>matR</i>
	Transport membrane protein	<i>mttB</i>
	NADH dehydrogenase	<i>nad1^c, nad2^c, nad3, nad4^b, nad4L, nad5^c, nad6, nad7^b, nad9</i>
Variable genes	Ribosomal protein large subunit	<i>rpl5, rpl10, rpl16</i>
	Ribosomal protein small subunit	<i>rps3^a, rps4, rps10^a, rps12, rps13, rps14</i>
rRNA genes	Ribosomal RNAs	<i>rrn5, rrn18, rrn26</i> <i>trnC-GCA, trnD-GUC, trnE-UUUC, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnM-CAU, trnM-CAU, trnN-GUU, trnP-UGG, trnS-GCU, trnS-GGA, trnS-UGA, trnW-CCA, trnY-GUA, trnQ-UUG, trnM-CAU, trnM-CAU</i>
tRNA genes	Transfer RNA	

Note: “a”, “b”, and “c”: genes with two, four, and five exons, respectively.

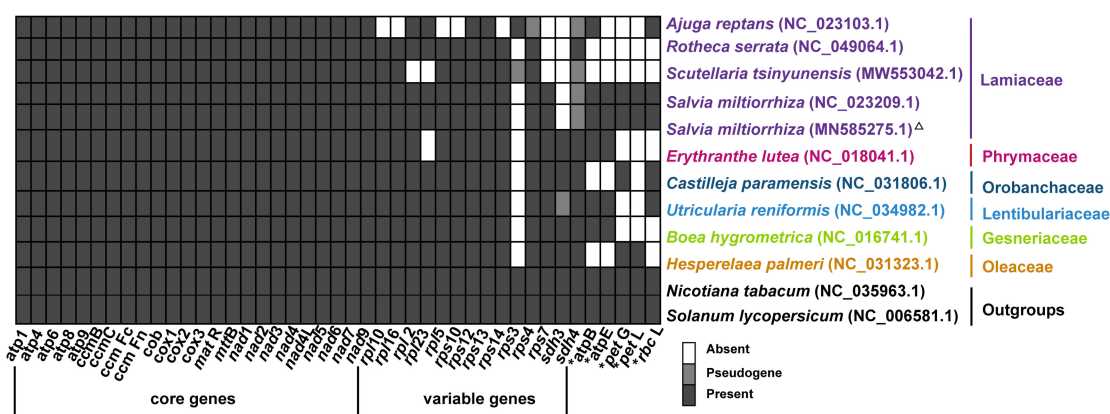


Figure 3. PCG contents in nine sequenced Lamiales mitogenomes. The X-axis shows the name of core and variable genes. Five PCGs, namely, *atpB*, *atpE*, *petG*, *petL*, and *rbcl*, were similar to the sequences from plastomes (labeled with “*”). The mitogenome generated in this study was labeled with the triangle. The species belonging to *Lamiaceae*, *Phrymaceae*, *Lentibulariaceae*, *Gesneriaceae*, and *Oleaceae* were labeled with the colors of purple, peach, dark green, blue, green, and orange, respectively.

2.3. Repeat Sequences in the Lamiales Mitogenomes

Repeat sequences make up a large proportion of eukaryotic genomes; they play important roles in genome evolution and have been used widely as molecular markers for discrimination at the subspecies levels [38]. Simple sequence repeats (SSRs), also called microsatellite repeats, have been considered effective molecular markers due to their high variability in the whole plasmid. Thus, they can provide useful information for phylogenetic and population genetic studies [39]. By contrast, tandem repeats consist of multiple copies of repeat units (≥ 5 nucleotides) that are adjacent to one another.

To explore the potential roles of repeat sequences, we analyzed two types of repeat sequences in the mitogenome of *S. miltiorrhiza* and another nine species of Lamiales. We detected 112 SSRs, including 99 SSRs on MC1 and 13 on MC2. Among them, 35 (35.35%) SSRs in MC1 were tetra-nucleotide repeats, and 27 (27.27%), 16 (16.16%), 13 (13.13%), 4 (4.04%), and 4 (4.04%) were di-, tri-, mono-, penta-, and hexanucleotide repeats, respectively

(Tables S2 and S3). In MC2, tetra- and tri-nucleotide repeats were the most abundant, with a ratio of 30.77% (4). The SSRs of tetranucleotide dominated *S. miltiorrhiza* (Table S3).

By contrast, one and two tandem repeat sequences were found in MC1 and MC2 of the *S. miltiorrhiza* mitogenome, respectively. These three repeats had lengths between 15 and 30 bp (Table S4). The two types of repeats had also been detected in the other nine Lamiales mitogenomes (Tables S3 and S4). In addition, the numbers of tandem repeats were not correlated with the genome size of the nine Lamiales mitogenomes and differed from the number of SSRs repeats that were positively correlated with the genome size in most of the nine Lamiales species (Tables S3 and S4).

2.4. Homologous Recombination Mediated by Repeats

To explore the potential subgenomic structures of the *S. miltiorrhiza* mitogenome, we identified the repeats in the MC1 and MC2 using the BLASTN program with a cutoff of E-value = 10^{-6} and a word size = 7 [40]. We identified 72 high scoring pairs (HSPs) in total. For simplicity, we used HSPs and repeats interchangeably in the following text. To determine whether these repeats can mediate recombination, we extracted the sequences containing the repeats themselves and 500 bp sequences upstream and downstream of the repeats. We then switched the flanking regions to generate alternative conformations. These created sequences corresponded to four conformations, namely, C1, C2, C3, and C4 (Figure 4A). The sequences corresponding to these conformations are provided in Supplementary File S2. Here, C1 and C2 were found in the genome assembly. C3 and C4 were created by switching the flanking sequences of C1 and C2, respectively. C1 and C2 sequences were reverse-complementary to each other, and the same condition applied to C3 and C4 sequences. We mapped the long reads to the sequences of the four conformations. Nine pairs of repeats were likely associated with the homologous recombination based on the mapping result of the long reads (Table 1, highlighted with “a”). Figure S6a–i shows the mapping results.

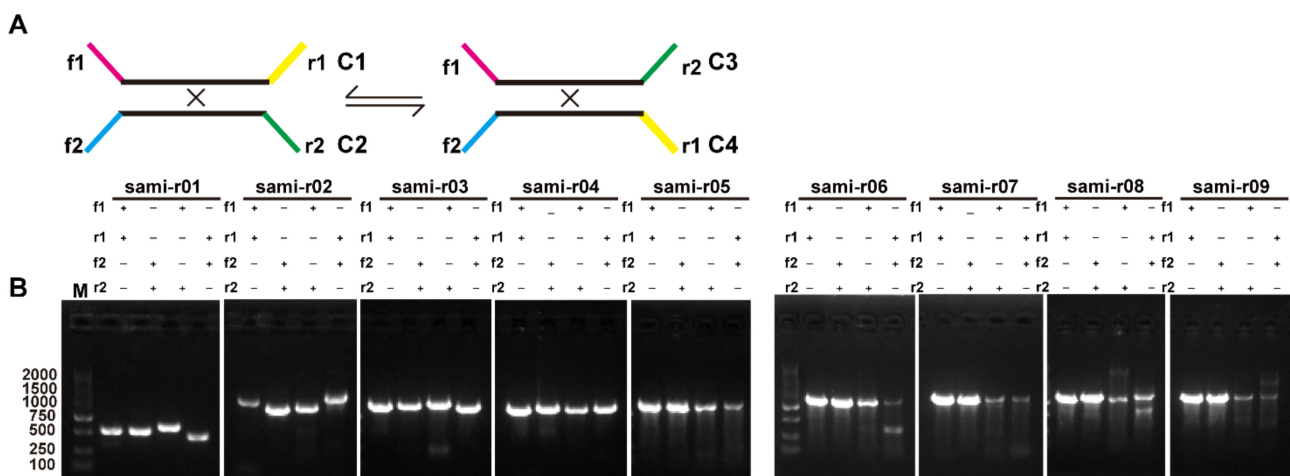


Figure 4. Verification of recombination products associated with the repeats on MC1 and MC2. (A) Schematic of the DBSs and their corresponding Mac and alternative conformations. The blank lines represent the two repeat units. The flanking sequences are shown in purple, yellow, green, and blue. The Mac were those observed in our genome assembly. (B) PCR amplification of products representing the Mac and Mic conformations. The primer combinations for r01–09 are shown. The Sanger sequencing results of the PCR products are displayed in Figure S6.

Among the nine repeats, eight had both repeat subunits on MC1, and one repeat had one repeat unit on MC1 and another repeat unit on MC2. The length of these nine repeats, including six direct and three inverted repeats, ranged from 59 bp (r09) to 682 bp (r02). The recombination frequency was calculated by the numbers of long reads mapped to the Mic divided by those mapped to all conformations. The percentages of recombination products

associated with the nine repeat sequences were as follows: r05 (2.34%), r04 (4.65%), r07 (6.94%), r01 (7.54%), r06 (9.48%), r09 (11.8%), r03 (13.79%), r08 (18.42%), and r02 (21.33%), in ascending order (Table 1).

To further validate the presence of homologous recombination products identified by long-read mapping, we designed primers (Table S5) to amplify the sequences corresponding to the four conformations. The polymerase chain reaction (PCR) products of the primer pairs F1 + R1 and F2 + R2 supported the presence of C1 and C2 conformations. By contrast, the PCR products of the primer pairs F1 + R2 and F2 + R1 supported the presence of C3 and C4 conformations, respectively (Figure 4A). Figure 4B shows the gel electrophoresis results of the PCR products. The observed band patterns of PCR products were consistent with the expected results. The PCR products were then subjected to Sanger sequencing. The Sanger sequencing results of the PCR products were aligned with the expected sequences, and they were mostly identical (Figure S7a–i).

We compared the contig sequences corresponding to the shared nodes of DBSs in the unitig graph. The contig sequences of the shared nodes of bs02, bs03, bs04, bs05, and bs06 were the same as r01, r05, r03, r02, and r04, respectively (Table 1). The HSPs r11 and r10 were the same as the shared contig sequences of bs01 and bs07, respectively. However, the four conformations were not verified successfully by PCR and Sanger sequencing experiments. The examination of their sequences showed that r11 was associated with very low recombination frequencies, and r10 was too long to be amplified by PCR (Table 1).

To determine the number of conformations possibly contained by the mitogenome, we defined MC1 and MC2 as Mac1 and Mac2, respectively. We then inferred the potential Mic generated by nine r01–r09 (Figure 5). Mic2, Mic3, and Mic7 resulted from the recombination mediated by r02, r03, and r07 from Mac1, respectively. Their recombination frequencies were 21.33% (Mic2), 13.79% (Mic3), and 6.94% (Mic7), and they contained the rearranged structure of Mac1. The r04, r05, r06, r08, and r09 can split Mac1 into two Mic (Mic4-1, Mic4-2, Mic5-1, Mic5-2, Mic6-1, Mic6-2, Mic8-1, Mic8-2, Mic9-1, and Mic9-2), with the recombination frequencies of 4.65%, 2.34%, 9.48%, 18.42%, and 11.88%, respectively (Figure 5).

2.5. Identification of Mitochondrial Plastid Sequences (MTPTs)

Mitochondrial plastid DNAs (MTPTs) are plastid-derived DNA fragments in mitochondrial genomes [41]. Sixteen fragments similar to the plastome were identified in the mitogenome of *S. miltiorrhiza*. Figure 6 shows their locations on the plastome and mitogenome. Table S6 provides detailed information for these MTPT fragments in terms of the percentage of match, start and end positions, gene contents, etc.

To determine whether these MTPTs were present in the mitogenome rather than the assembly artifact, we searched for PacBio long reads that can be mapped to these MTPT fragments. We identified numerous PacBio long reads that can cover the 16 MTPT fragments and their flanking of 2,000 bp regions in the mitogenome (Figure S8a–p). We obtained screenshots in Tablet to visualize the alignment of long PacBio reads to sami-mtpt-001 to sami-mtpt-016 (Figure S8a–p, respectively). These long reads confirmed the occurrence of MTPT events in the mitogenome of *S. miltiorrhiza*.

These MTPTs had a total size of 12,583 bp, representing approximately 3.04% of the total length of the *S. miltiorrhiza* mitogenome and 8.31% of the plastome. The longest MTPT fragment (sami-mtpt-012) was observed from positions 73,481 to 78,467 on MC2, and it contained the full coding sequences of *rbcl*, *atpB*, *atpE*, and *trnM-CAU*. The second-longest MTPT (sami-mtpt-001) was from base 70,229 to base 72,166 of MC1 and contained a fragment of the chloroplast gene *psbB*. The third-longest MTPT (sami-mtpt-002) was from base 308,153 to base 309,889 and included the full coding sequences of the chloroplast genes *petG*, *petL*, and *trnW-CCA*, which were identified in the MC1 region (Figure 6 and Table S6).

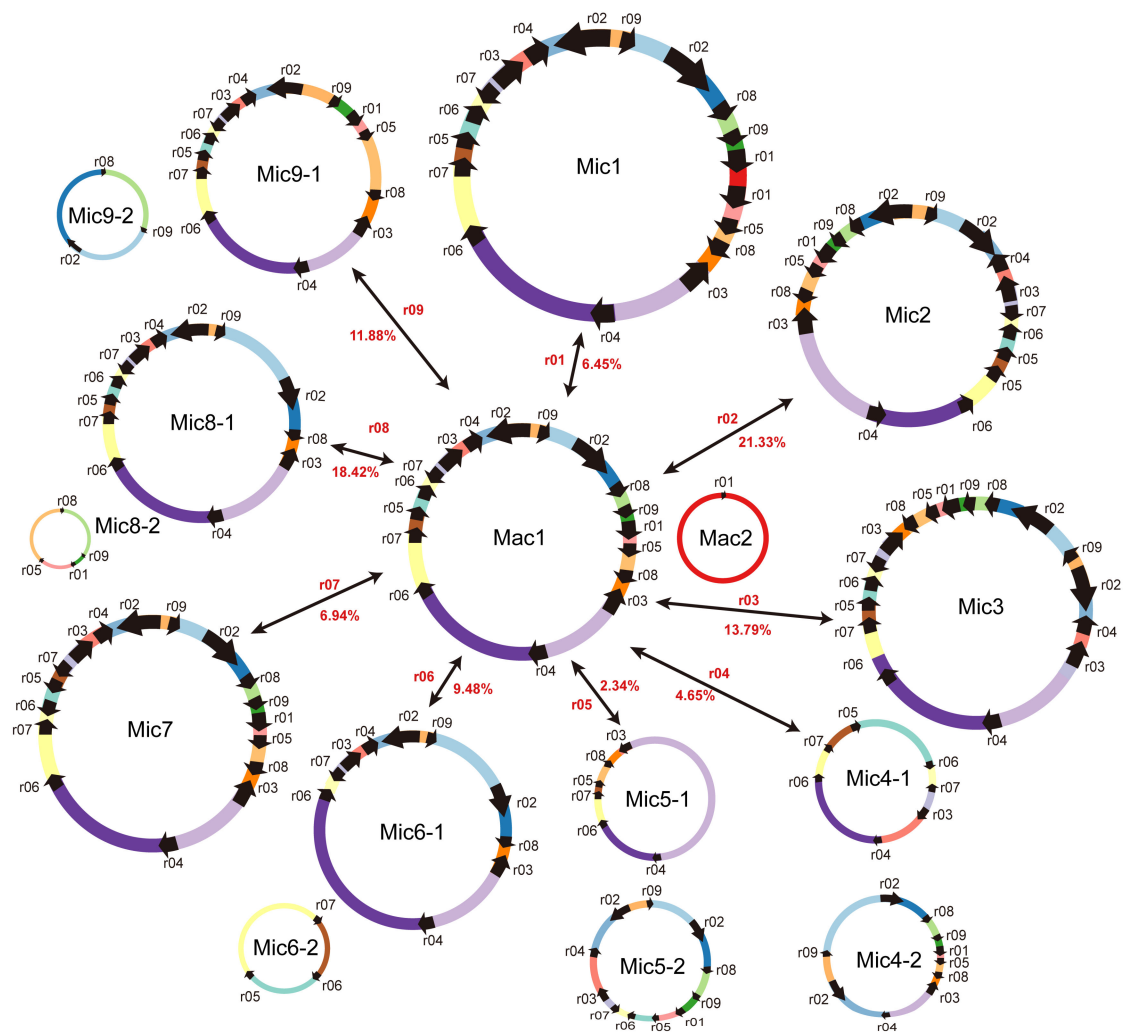


Figure 5. Potential conformations resulting from the recombination mediated by nine repeats (sami-r01 to sami-r09) in the *S. miltiorrhiza* mitogenome. The black arrows on the circular molecules represent the repeat unit sequences, and the colored lines denote the DNA fragments between repeats. Mac and Mic represent the major and minor conformations resulting from the rearrangement, respectively. The dominant conformations included Mac1 and Mac2. Mic1 resulted from the sami-r01-mediated recombination with a recombination frequency of 6.45% and contained one circular molecule. Mic2, 3, and 7 resulted from r02, r03, and r07-mediated recombination from Mac1, respectively, with recombination frequencies of 21.33%, 13.79%, and 6.94%. They all contained one circular molecule. Mic4, 5, 6, 8, and 9 resulted from r04, r05, r06, sami-r08, and sami-r09-mediated recombination from Mac1, respectively, and their recombination frequencies were 4.65%, 2.34%, 9.48%, 18.42%, and 11.88%. They all split Mac1 into two Mic. r01 is an inter-molecular repeat and causes inter-chromosomal recombination. On the contrary, r02, r03, r04, r05, r06, and r07 lead to intra-chromosomal recombination.

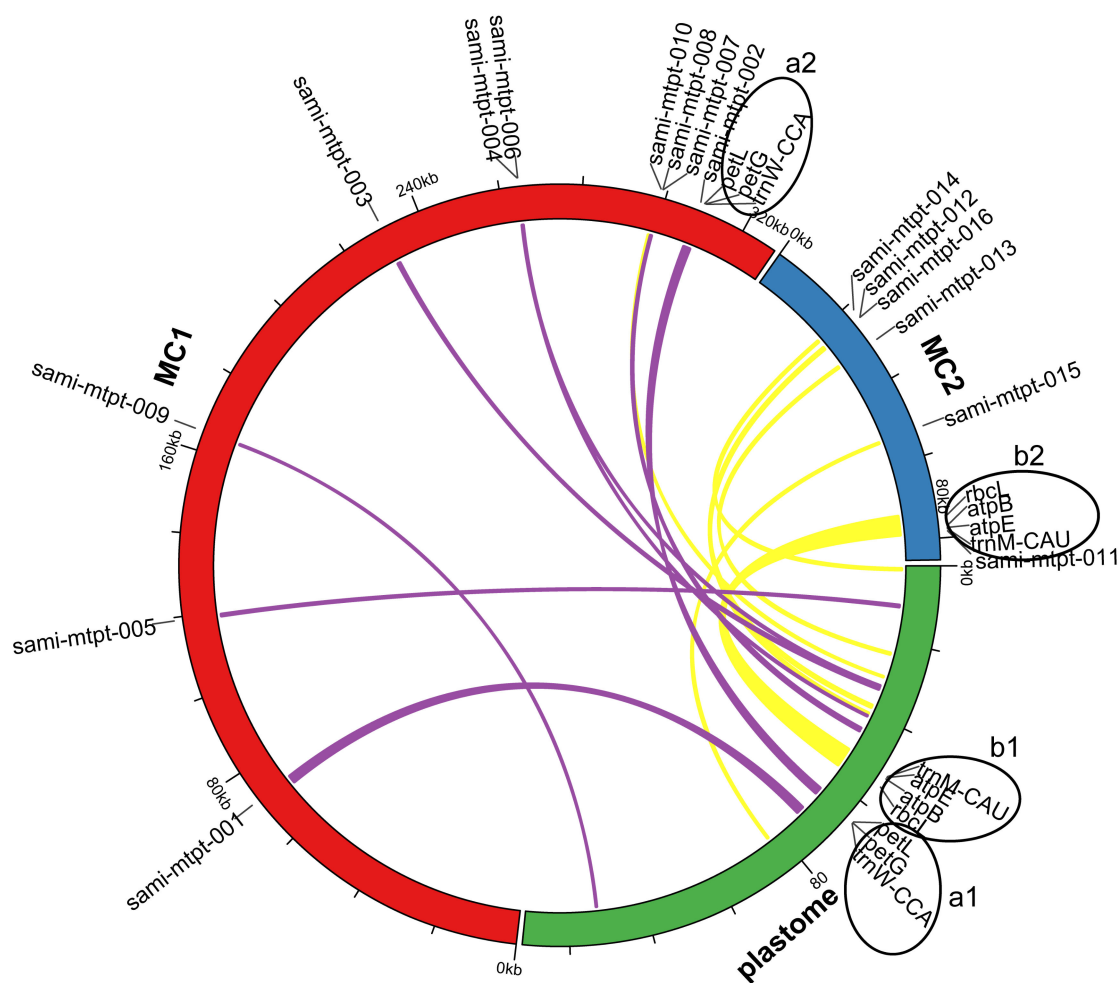


Figure 6. Circular map showing the distribution of MTPT on MC1 and MC2 of the *S. miltiorrhiza* mitogenome. The red and blue lines of the outer circle represent the MC1 and MC2 of the *S. miltiorrhiza* mitogenome, respectively. The green line of the outer circle is the plastome of *S. miltiorrhiza*. The purple and yellow links in the inner circle represent MTPT on MC1 and MC2 of the *S. miltiorrhiza* mitogenome, respectively. a1 and a2 denote the gene cluster of the MTPT fragment sami-mtpt-002 located on MC1 and plastome, respectively. b1 and b2 represent the gene cluster of the MTPT fragment of sami-mtpt-011 located on MC2 and plastome, respectively.

2.6. Phylogenetic Analysis

To determine the phylogenetic relationship of ten Lamiales mitogenomes, we constructed a phylogenetic tree using 24 core mitochondrial PCGs (*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*) and 2 variable genes (*rps12* and *rps13*) from these 10 mitogenome sequences. The mitogenomes of *Solanum lycopersicum* and *Nicotiana tabacum* were used as outgroups. The ten mitogenome sequences of Lamiales were from the following families: Lamiales, Phrymaceae, Orobanchaceae, Lentibulariaceae, Gesneriaceae, and Oleaceae. Phylogenetic trees were constructed using the maximum likelihood (ML) and Bayesian inference (BI) methods. Most nodes on the phylogenetic tree had bootstrap support values >90 and posterior probabilities = 1, indicating the strong reliability of the phylogenetic relationship of the nine Lamiales species (Figure 7). The topological structure of the tree is identical to the phylogeny of Lamiales species in the APG IV system [42] and the results from a previous study [34]. In particular, although the organization of our genome assembly and the previously reported one (NC_023209.1) differed significantly (Figure S4), their protein sequences are highly conserved.

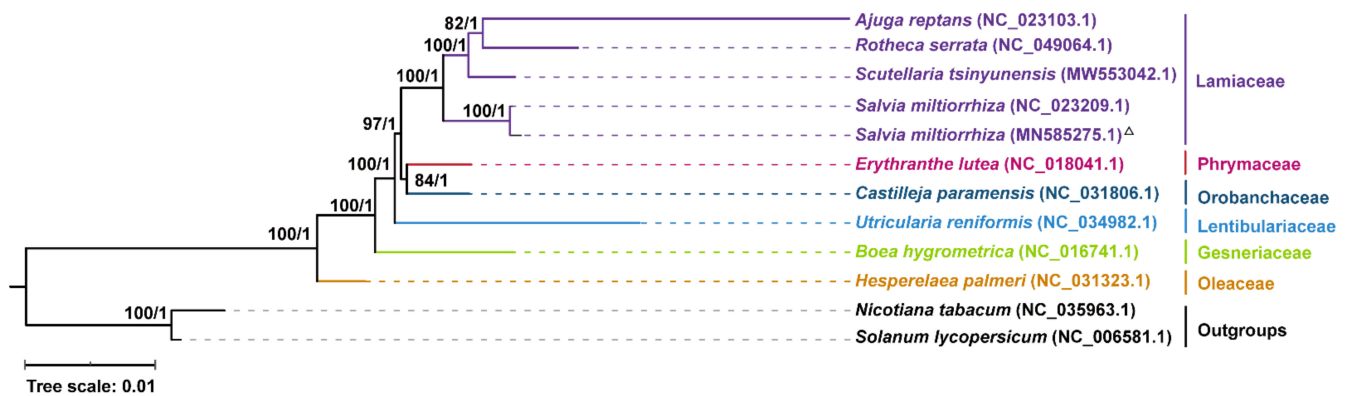


Figure 7. Molecular phylogenetic analysis of mitogenomes in Lamiales. The tree was constructed using concatenated conserved protein sequences from the mitogenomes of nine species via the ML and BI methods. The bootstrap score was obtained using 1000 replicates. The ML bootstrap support values and BI posterior probabilities were labeled at the corresponding nodes. Two species from Solanaceae (*Nicotiana tabacum* and *Solanum lycopersicum*) were used as outgroups. The mitogenome assembled in this study was labeled with a triangle. The species belonging to *Lamiaceae*, *Phrymaceae*, *Lentibulariaceae*, *Gesneriaceae*, *Oleaceae* were labeled with the colors of purple, peach, dark green, blue, green, and orange, respectively.

2.7. Identification of Genes under Selection

To determine variations in nucleotide substitution rates in the mitogenome of *S. miltiorrhiza* and the other nine mitogenome sequences of Lamiales, we calculated the pairwise non-synonymous substitution rate (dN), the synonymous substitution rate (dS), and the ratio of dN to dS of the 26 shared mitochondrial genes using the yn00 module of PAML (v4.9) [43]. The genes *atp4*, *ccmB*, *ccmFc*, and *mttB* showed that the dN/dS ratios were over 1.0 in most species, indicating a possible positive selection (Figure 8 and Figures S9 and S10, Supplementary File S3). Most of the genes, such as *atp1*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad4L*, *nad5*, *nad6*, and *rps12*, showed low dN/dS ratios, implying a possible negative selection. The *atp1* and *atp9* genes had a prominently low dN/dS ratio compared with those of other PCGs, suggesting that they may be functionally highly conserved.

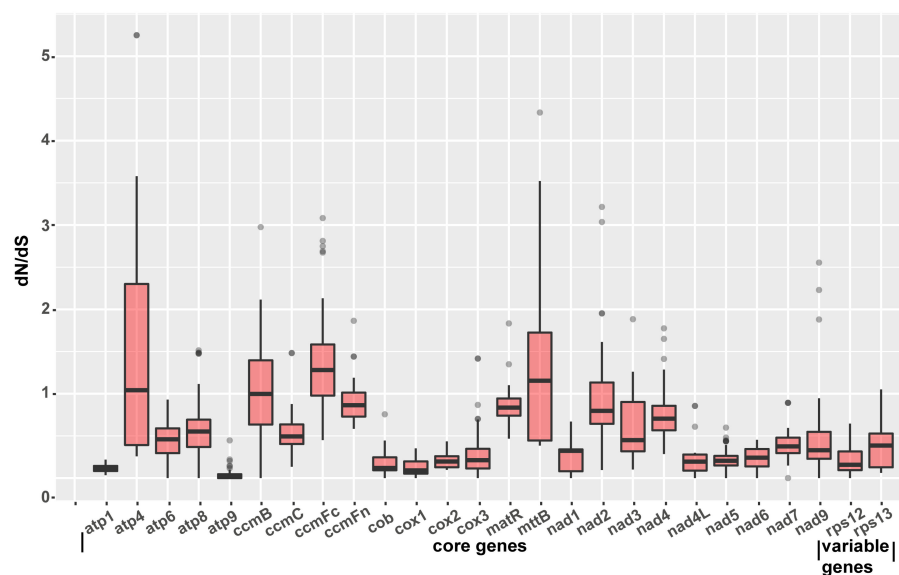


Figure 8. Molecular phylogenetic analysis of mitogenomes in Lamiales. The highest and lowest dot represented the maximum and minimum of the pairwise dN/dS values of each gene. The upper, medium, and lower lines in the box represented the upper quartile, median, and lower quartile of the pairwise dN/dS values of each gene, respectively.

3. Discussion

3.1. Overview of the *S. miltiorrhiza* Mitogenome

In this study, we achieved the following: (1) obtained a high-quality mitogenome of *S. miltiorrhiza* using a hybrid assembly strategy; (2) annotated the *S. miltiorrhiza* mitogenome and predicted its gene contents; (3) analyzed repeat elements; (4) predicted and validated the homologous recombination mediated by repeats; (5) identified the MTPTs between the plastome and mitogenome; (6) constructed phylogenetic trees with the PCG sequences; (7) calculated the substitution rates of mitochondrial PCGs. The detailed characterization of the high-quality assembly of the *S. miltiorrhiza* mitogenome may serve as the foundation for future studies on the genomic evolution of this important medicinal plant.

We compared the sequences of the *S. miltiorrhiza* mitogenome obtained in this study with the published one (NC_023209) [9] and observed that the two assemblies significantly differed in terms of structure with a number of rearrangements. The mapping of long reads strongly supported the assembly of this study. The structural differences between these assemblies may be due to the following reasons. First, large intra-specific variations may exist in mitogenome structures. Second, the use of different sequencing technologies and assembly algorithms may generate various mitogenome structures.

3.2. Repeat Mediated Homologous Recombination in Lamiaceae

Plant mitogenomes are a complex and dynamic mixture of forms rather than a single circle [44]. Previously, a single circular *S. miltiorrhiza* mitogenome was released. In the present study, we used Illumina and PacBio reads to investigate the diverse mitogenome structures. We observed that the *S. miltiorrhiza* mitogenome consists of two major conformations, Mac1 and Mac2, and multiple minor conformations, which resulted from the recombination mediated by nine repeats.

These findings were similar to those found in the mitogenomes of *Silene* [15], cucumber [17], sugarcane [45], *Lactuca* [10], rice [46], onion [27], and *Solanum tuberosum* [47]. In a previous report, the mitogenome of *Scutellaria tsinyunensis* of the Lamiaceae family consisted of a 175 bp direct repeat shared by two minor circular conformations. Similar to *S. tsinyunensis*, the *S. miltiorrhiza* mitogenome is divided into two direct circular structures by a 127 bp-long forward repeat. In *S. tsinyunensis*, the major confirmation of the mitogenome is a single circle. However, two circular molecules form the major confirmation in the *S. miltiorrhiza* mitogenome. By comparing the two repeat sequences mediating recombination in *S. tsinyunensis* and *S. miltiorrhiza*, we found no sequence similarity between them. The mechanism of repeats mediating recombination in other Lamiaceae species requires support from further experimental evidence.

3.3. Current Limitation of the Plant Mitogenome Assembly Method

Several technical limitations may affect the quality of mitogenome assembly. This study assembled mitogenome from total DNAs with a hybrid assembly strategy, combining the unitig graph assembled from short reads and contigs assembled from long reads. This strategy can avoid the false positive caused by the Polish strategy. In addition, the presence of MTPTs and nuclear mitochondrial sequences (NUMTs) in total DNAs may affect genome assembly. We have carefully checked the 16 MTPTs in this study based on the mapping results of MTPTs and regions flanking the MTPTs.

In addition, given their complex structure, mitogenomes may have multiple configurations as a combination of linear, circular, and branched molecules [23]. Several studies of plant mitogenomes were accustomed to using a separate ring molecule to represent the plant mitogenome. Such a representation is inadequate to describe the dynamic and complex structure of mitogenomes. Through cryo-electron microscopy, complex physical structures (circular, linear, and branched) of mtDNA molecules were observed in *Lactuca sativa* [10]. In addition, researchers of plant mitogenomes should identify as many of the Mac of the genome as possible and explore possible forms of recombination based on

sequencing results. Bioinformatic predictions must be further validated by quantitative PCR experiments, Sanger sequencing, Southern blot, and electronic microscopy.

3.4. Future Studies of the *S. miltiorrhiza* Mitogenome

Several directions can be pursued to further analyze the structure of the *S. miltiorrhiza* mitogenome. We can analyze the mitogenome structure by isolating mitochondria and DNAs for subsequent analyses with long-read DNA sequencing technology, the diversity and dynamics of *S. miltiorrhiza* mitogenomes at the population level, and the various levels of recombination among mitogenomes from different plants, particularly those that may affect the physical properties of *S. miltiorrhiza*. Finally, the structure of mitogenome DNAs can be examined via electron microscopy to visualize the actual mitogenome structures and confirm their structural diversity. The results obtained from these studies will lay a solid foundation for understanding mitogenome evolution and facilitate mitogenome-based breeding.

4. Materials and Methods

4.1. Plant Materials and DNA Extraction

Fresh leaves were collected from an inbred line (named sami-il01) of *S. miltiorrhiza* from the Research Center of Medicinal Plants, Shandong Academy of Agricultural Sciences, Shandong, China. *S. miltiorrhiza* is not an endangered or protected species, and, thus, no specific permissions were required for its collection. A voucher specimen was deposited with the accession number sami-001 at the institute. The leaves were stored at -80°C until the total DNA was extracted using a plant genomic DNA kit (Tiangen Biotech, Beijing, Co., Ltd., Beijing, China). DNA purity was evaluated with electrophoresis using a 1.0% agarose gel, and the DNA concentration was measured using a Nanodrop spectrophotometer 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

4.2. DNA Sequencing and Mitogenome Assembly

The DNA samples from sami-il01 were subjected to library construction using the SMRTbell Template kit 1.0 (Pacific Biosciences, Menlo Park, CA, USA). In particular, 10 μg DNA was sheared with a Covaris gTube at 4500 rpm for 2 min. DNA fragments (12–50 kb) were selected using the BluePippin cassette 0.75% DF Marker S1 high pass (15–20 Kb). A total of 15 cells were sequenced using the PacBio RSII sequencer with the DNA Sequencing Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA).

The same DNA sample used above was fragmented to 300–500 bp in length, barcoded, and subjected to library construction using the NEBNext[®] library construction kit [48] (NEB, Ipswich, MA, USA), following the manufacturer's instructions. The library was sequenced in the PE100 setting using an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) sequencer.

The plastome of *S. miltiorrhiza* (sami-il01) was de novo assembled from Illumina reads using GetOrganelle (v1.6.4) with the parameter “-R 15 -k 21,45,65,85,105 -F embplant_pt”. The mitogenome of *S. miltiorrhiza* (sami-il01) was assembled with a hybrid assembly strategy. First, the mitochondrial short reads were extended and achieved by GetOrganelle (v1.6.4). Then, we de novo assembled the GetOrganelle-extended reads into a unitig graph by the SPAdes software embedded in Unicycler (Pacific Biosciences, Menlo Park, CA, USA). Finally, we resolved the DBSs in the unitig graph with PacBio long reads using Unicycler.

To examine the accuracy of the unicycler in resolving the DBSs, we extracted the sequences containing the repeats in DBSs and 1000 bp sequences upstream and downstream of the repeats. We then switched the flanking regions, and these created sequences that corresponded to the four conformations. We mapped the long reads to the sequences using BWA (v0.7.12-r1039) [49] with default parameters.

To further examine the mitogenome assembly, we mapped the PacBio long reads to the plastome and mitogenome sequences using minimap2 (2.17-r941) [50] with the parameter “minimap2 -ax map-pb.” The Illumina short reads were mapped to the mitogenome by

BWA (v0.7.12-r1039) [49] with default parameters. The coverage depth was calculated using samtools (v1.3.1) [51]. The collinearities of the *S. miltiorrhiza* mitogenome obtained in this study and the one published before (NC_023209) [9] were examined using Gepard (v1.40). For further examination, we mapped the PacBio long reads to the mitochondrial genome available in GenBank (NC_023209).

4.3. Mitogenome Annotation

The Geseq web server [52] and the custom program MGA (<http://www.1kmpg.cn/mga> (accessed on 18 June 2022)) were used to annotate the mitogenome. The tRNA genes were identified using tRNAscan-SE [53]. The positions of the start and stop codons and intron/exon boundaries were manually corrected using the Apollo program [54]. The circular mitogenome map was visualized using PMGView (<http://www.1kmpg.cn/pmgview> (accessed on 22 June 2022)). The mitogenome sequences were submitted to GenBank with the accession numbers MN585275.1 and MN585276.1 and were also available at Figshare (<https://doi.org/10.6084/m9.figshare.21195841> (accessed on 28 June 2022)).

4.4. Repeat Sequence Analysis

SSRs were detected using the MISA web service [55], with the following thresholds: 10 for the number of mononucleotide repeat units, 5 for the number of dinucleotide repeat units, 4 for the number of trinucleotide repeat units, and 3 for the number of tetra-, penta-, and hexanucleotide repeat units. Tandem repeats were analyzed using the Tandem Repeats Finder [56] with parameter settings of 2 for matches and 7 for mismatches and indels. The minimum alignment score and maximum period size were set at 50 and 500, respectively.

4.5. Analysis of the Recombination Products

To identify the inter- and intra-molecular recombination mediated by the repeat sequences, we searched for the repeat sequences of MC1 and MC2 with BLASTN using the following parameters: E-value: 1E-6 and word size: 7 [40]. To examine the presence of possible recombination products around the repeats, we first extracted the 500 bp-long sequence segments around the repeats based on the expected sequences before and after recombination. We then mapped PacBio long reads to the extracted sequence segments of the four conformations and counted the repeat-spanning reads.

For the hypothetical recombination products identified by mapping the PacBio long reads, we designed PCR primers using the Primer 3 web service [57]. PCR reactions were performed in 50 μ L volumes with 23 μ L water, 25 μ L 2 \times Taq PCR Master Mix, 1 μ L of each primer, and 1 μ L DNA. PCR reactions were carried out on a Pro-Flex PCR system (Applied Biosystems, Waltham, MA, USA) under the following conditions: denaturation at 94 $^{\circ}$ C for 2 min, followed by 35 cycles of 94 $^{\circ}$ C for 30 s, 57 $^{\circ}$ C for 30 s, 72 $^{\circ}$ C for 60 s, and 72 $^{\circ}$ C for 2 min as the final extension. We separated and visualized the PCR products on 1.0% agarose gels. Finally, the PCR amplicons were sequenced using the Sanger method.

4.6. Identification of Mitochondrial Plastid Sequences (MTPTs)

MTPTs were identified by a reciprocal comparison strategy using BLASTN (v 2.2.30+), as described previously [58]. The plastome was assembled from Illumina reads using GetOrganelle [59]. The plastome was compared with the mitogenome using BLASTn through the following parameters: e-value: 1e-6 and word size: 7 [40]. BLASTn hits that were shorter than 100 bp were ignored. The MTPT gene cluster was identified in accordance with the shared boundary of continuous genes, as described by Wang et al. [41].

Then, the MTPT gene clusters on the mitogenome were identified, which were defined as a cluster of continuous genes on the plastome without the insertion of any mitochondrial genes. MTPT gene clusters were depicted by a circular map obtained from TBtools (v1.076) [60]. Moreover, to confirm the MTPT events, we extracted the MTPT fragments including 2000 bp sequences on its 5' and 3' ends. The PacBio long reads were aligned to these

long fragments using BWA-MEM [61]. Tablet software (Zebra Technologies, Lincolnshire, IL, USA) [62] was used to visualize the mapping results.

4.7. Phylogenetic Analysis of the Nine Lamiales Species

Nine complete mitochondrial DNA sequences belonging to the order Lamiales were obtained from GenBank. These sequences specifically belonged to the following species: *Ajuga reptans* (NC_023103.1), *Rothea serrata* (NC_049064.1), *Scutellaria tsinyunensis* (MW553042.1), *Salvia miltiorrhiza* (NC_023209.1), *Erythranthe lutea* (NC_018041.1), *Castilleja paramensis* (NC_031806.1), *Utricularia reniformis* (NC_034982.1), *Boea hygrometrica* (NC_016741.1), and *Hesperelaea palmeri* (NC_031323.1). The mitogenomes of *Nicotiana tabacum* (NC_006581.1) and *Solanum lycopersicum* (NC_035963.1) were used as outgroup taxa.

For phylogenetic analysis, the DNA sequences of the 26 PCGs shared among these ten mitogenomes (two for *S. miltiorrhiza*) were extracted using PhyloSuite (v1.2.1) [63]. The 26 conserved PCGs were *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rps12*, and *rps13*. These sequences were aligned using MAFFT (v7.450) [64], and a phylogenetic tree was constructed using the alignment and ML method implemented in RAxML (v8.2.4) [65]. The detailed parameters were “raxmlHPC-PTHREADS-SSE3 -f a -N 1000 -m PROTGAM-MACPREV -x 551314260 -p 551314260 -o *Nicotiana tabacum*, *Solanum lycopersicum* -T 20.” The significance level for the phylogenetic tree was assessed using bootstrap testing with 1000 replications. We also performed the BI analysis using MrBayes (v3.2.7a) [66] on CIPRES Science Gateway (v3.3) [67]. The best model for the BI analysis was obtained by the jModelTest (v2.1.0) [68]. The resulting tree was visualized by iTOL [69].

4.8. Estimation of Nucleotide Substitution Rates

We estimated the pairwise non-synonymous substitution rate (dN), the synonymous substitution rate (dS), and the ratio of dN to dS of sequences of the 26 mitogenome protein-coding genes used in the previous analysis. The yn00 module in PAML (v4.9) [43] was used to conduct the estimation with the following parameters: verbose, icode, weighting, and common $3 \times 4 = 0$, and ndata = 1. The pairwise dN, dS, and dN/dS values were shown by the boxplot, which was plotted by R-package (ggplot2) [70].

5. Conclusions

We showed that the *S. miltiorrhiza* mitogenome consists of two circular chromosomes. Recombination mediated by nine repeats can result in a large number of various conformations. The results obtained from this study suggest that multiple chromosomal structures may be more prevalent than previously thought. They can be present in plants whose primary form of mitogenome is a “master circle”. In the future, obtaining complete mitogenome sequences of additional *Salvia* plants will provide insights into the role that homologous recombination plays in the diversification and evolution of mitogenomes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232214267/s1>, Figure S1: Comparison of the sequences of the plastome assembled in this study and NC_020341.1 of *S. miltiorrhiza*. The X-axis shows the nucleotide sequence of plastome assembled in this study, and the Y-axis displays the nucleotide sequence of NC_020341.1. Figure S2: Alignments of long PacBio reads to the four conformations of the seven DBSs (bs01–bs07) found in the unitig graph. The unitig graph was generated using Unicycler from Illumina reads, which were filtered with GetOrganelle for mitochondrial reads. Mac corresponds to the connection of shared contigs and their flanking contigs found in MC1 and MC2 (Figure 1B). Mic corresponds to the alternative connection of the shared contigs and their flanking contigs that are not found in MC1 and MC2. Panels a–g show Mac1 and Mac2 and Mic1 and Mic2 of bs01–bs07, respectively. The figure in each panel can be divided into the top and bottom parts. The top part shows a bird’s eye view, whereas the bottom part reveals a base-level view. At the bottom part of the figure, the DBS ID, the length of the shared contig, and the conformation name are shown. The boundaries of the shared contig are indicated with red vertical lines. The contigs in the DBS are

shown as the red line with arrows at each end. It was labeled as the “repeat region.” Figure S3: Mapping results of PacBio and Illumina reads to the MC1 and MC2 of *S. miltiorrhiza*. Panels a and b show the mapping results of PacBio reads to MC1 and MC2, respectively. Panels c and d show the mapping results of all Illumina reads to MC1 and MC2, respectively. The X- and Y-axis show the nucleotide position and corresponding coverage depth, respectively. Figure S4: Comparison of the two assemblies of the mitogenome of *S. miltiorrhiza*. Comparison of the sequences of MC1 (A), MC2 (B), and NC_023209.1. The X-axis shows the nucleotide sequence of MC1, and the Y-axis presents the nucleotide sequence of NC_023209.1. The largest collinear block between them was 66,778 bp in length and is highlighted with a red circle. Figure S5: Mapping results of PacBio long reads to the mitogenome of *S. miltiorrhiza* in GenBank (NC_023209.1). The X-axis shows the nucleotide position, and the Y-axis reveals the corresponding coverage depth. Three large regions not supported by the long reads in this study are highlighted with red squares. Figure S6: Alignments of long PacBio reads to the Mac and Mic resulted from nine pairs of HSPs (sami-r01 to sami-r09). The nine HSPs were identified using BLASTn with the E-value = 1E-6. The alignment of two HSPs generated a DBS similar to the ones observed in the unitig graph (Figure 1). In the presence of homologous recombination, the DBSs formed by the alignment of HSPs and their flanking sequences will also have four conformations, similar to those shown in Figure 1B. We constructed the sequences corresponding to the four conformations of each DBS for each HSP. We then mapped the long PacBio reads to these sequences. The results are shown in panels a–i for sami-r01 to sami-r09, respectively. We named the conformations with the most supporting reads as Mac1 and Mac2. The reference sequences of Mac1 and Mac2 are reverse-complementary to each other. The alternative conformations were named as Mic1 and Mic2. The figure in each panel can be divided into the top and bottom parts. The top part shows a bird’s eye view, whereas the bottom part reveals a base-level view. At the bottom part of the figure, the HSP ID, the length of HSP, and the conformation name are shown. The boundaries of HSPs are indicated with red vertical lines. The HSPs are shown as the red line with arrows at each end. They were labeled as “repeat region.” Figure S7: Validation of four conformations corresponding to the recombination products mediated by the nine repeats. PCR primers were designed based on the four conformations. The genomic DNAs were then amplified by PCR, and PCR products were subjected to Sanger sequencing. The sequencing chromatogram (top), Sanger sequencing results (labeled with “PCR” and conformation number), expected sequences (labeled with repeat ID and conformation number), and consensus sequences are shown from top to bottom. Panels a–i correspond to repeats of the r01 to r09 results, respectively. The sequences of r01–r09 are highlighted in yellow. Figure S8: Alignments of the long PacBio reads to the MTPT fragments in the mitogenome of *S. miltiorrhiza*. Panels a–p show the alignments of long PacBio reads to the ten MTPT fragments in MC1 (sami-mtpt-001 to sami-mtpt-010) and six MTPT fragments in MC2 (sami-mtpt-011 to sami-mtpt-016). The MTPT regions are indicated with the red lines with arrow heads at each end. The 2000 bp-long flanking sequences are indicated with a red line without arrow heads. Figure S9: Boxplots of pairwise dN values for mitochondrial genes among the ten Lamiales plants. Figure S10: Boxplots of pairwise dS values for mitochondrial genes among the ten Lamiales plants. Table S1: Summary of sequence data generated by the PacBio RS and Illumina platforms. Table S2: SSRs in the *S. miltiorrhiza* mitogenome. “MC1/2”: mitogenome chromosome 1/2. Table S3: Summary of SSRs in the mitogenome of Lamiales. Table S4: Tandem repeats in the mitogenome of Lamiales. Table S5: PCR primers used to validate the recombination products of nine repeats in the *S. miltiorrhiza* mitogenome. Table S6: Chloroplast DNA insertions in the *S. miltiorrhiza* mitogenome. The three longest regions (sami-mtpt-001, sami-mtpt-002, and sami-mtpt-012) similar to the mitogenome sequence are indicated with “*”. “MC1/2”: mitogenome chromosome 1/2, frag: fragment. File S1: DNA sequence of the plastome of *S. miltiorrhiza* assembled using the GetOrganelle toolkit. File S2: DNA sequences of the four possible conformations associated with eleven HSPs (r01–r11) and their corresponding seven DBSs (bs01–bs07) in the mitogenome of *S. miltiorrhiza*. File S3: Alignment of the sequences of 26 PCGs of ten Lamiales mitogenomes and the results of pairwise dN, dS, and dN/dS values of the two species.

Author Contributions: Conceptualization, C.L.; methodology, C.L., H.Y., H.C. and Y.N.; software, H.C. and J.L.; validation, J.L.; formal analysis, H.Y., H.C. and Y.N.; investigation, Y.C. and B.M.; resources, J.Y.; data curation, J.Y., Y.C. and B.M.; writing—original draft preparation, H.Y., H.C. and Y.N.; writing—review and editing, H.Y., J.W. and C.L.; visualization, H.Y., H.C. and Y.N.; supervision,

J.W. and C.L.; project administration, J.W. and C.L.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Chinese Academy of Medical Sciences, the Innovation Funds for Medical Sciences (CIFMS) [2021-I2M-1-022], the National Science Foundation Funds [81872966], the National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], and the National Mega-Project for Innovative Drugs of China [2019ZX09735-002]. The funders were not involved in the study design, data collection, analysis, decision to publish, or manuscript preparation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw sequencing data from the Illumina and PacBio platforms generated during the current study are available in GenBank. The associated BioProject, BioSample, and SRA numbers and the associated link are PRJNA782861, SAMN23402358, and SRR17041866 for Illumina sequencing reads and SRR17042314 for PacBio sequencing reads. The mitogenome sequences have been released in GenBank (<https://www.ncbi.nlm.nih.gov/>, (accessed on 28 September 2022)) with the following accession numbers: MN585275.1 and MN585276.1. The DNA sequences of two chromosomes of the mitogenome of *S. miltiorrhiza* are provided as fasta files, along with the annotation information, which is provided as Genbank files and is also available at Figshare (<https://doi.org/10.6084/m9.figshare.21195841>, (accessed on 28 September 2022)). The plant sample has been stored at the Herbarium of the Institute of Medicinal Plant Development, Beijing, China (voucher numbers: Implad20181026).

Acknowledgments: We would like to thank Mei Jiang for the assistance in the identification of the mitochondrial plastid sequences and Jingwen Yue for the assistance in the DNA preparation.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

mitogenome	Mitochondrial genome
NCBI	National Center for Biotechnology Information
PCG	Protein-coding genes
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
C-to-U	Cytidine-to-Uridine
PacBio	Pacific Biosciences
Gb	Giga base pair
DBS	Double bifurcating structures
MC1/2	Mitochondrial genome chromosomes
SSR	Simple sequence repeat
HSP	High-scoring pair
Mac	Major conformation
Mic	Minor conformation
MTPT	Mitochondrial plastid DNA
PCR	Polymerase chain reaction
sami	<i>Salvia miltiorrhiza</i>
NUMT	Nuclear mitochondrial sequence

References

1. Cheng, T.O. Danshen: A popular Chinese cardiac herbal drug. *J. Am. Coll. Cardiol.* **2006**, *47*, 1498. [[CrossRef](#)] [[PubMed](#)]
2. Sui, C. *Salvia miltiorrhiza* Resources, Cultivation, and Breeding. In *The Salvia miltiorrhiza Genome. Compendium of Plant Genomes*; Lu, S., Ed.; Springer: Cham, Switzerland, 2019; pp. 17–32.
3. Xu, H.-Y.; Zhang, Y.-Q.; Liu, Z.-M.; Chen, T.; Lv, C.-Y.; Tang, S.-H.; Zhang, X.-B.; Zhang, W.; Li, Z.-Y.; Zhou, R.-R. ETCM: An encyclopaedia of traditional Chinese medicine. *Nucleic Acids Res.* **2019**, *47*, D976–D982. [[CrossRef](#)] [[PubMed](#)]
4. Xu, H.; Song, J.; Luo, H.; Zhang, Y.; Li, Q.; Zhu, Y.; Xu, J.; Li, Y.; Song, C.; Wang, B.; et al. Analysis of the Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *Mol. Plant* **2016**, *9*, 949–952. [[CrossRef](#)] [[PubMed](#)]

5. Zhang, G.; Tian, Y.; Zhang, J.; Shu, L.; Yang, S.; Wang, W.; Sheng, J.; Dong, Y.; Chen, W. Hybrid de novo genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). *Gigascience* **2015**, *4*, s13742-015. [[CrossRef](#)] [[PubMed](#)]
6. Song, Z.; Lin, C.; Xing, P.; Fen, Y.; Jin, H.; Zhou, C.; Gu, Y.Q.; Wang, J.; Li, X. A high-quality reference genome sequence of *Salvia miltiorrhiza* provides insights into tanshinone synthesis in its red rhizomes. *Plant Genome* **2020**, *13*, e20041. [[CrossRef](#)]
7. Ma, Y.; Cui, G.; Chen, T.; Ma, X.; Wang, R.; Jin, B.; Yang, J.; Kang, L.; Tang, J.; Lai, C.; et al. Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. *Nat. Commun.* **2021**, *12*, 685. [[CrossRef](#)]
8. Chen, H.; Liu, C. The Chloroplast and Mitochondrial Genomes of *Salvia miltiorrhiza*. In *The Salvia miltiorrhiza Genome*; Springer: Cham, Switzerland, 2019; pp. 55–68.
9. Wu, B.; Chen, H.; Shao, J.; Zhang, H.; Wu, K.; Liu, C. Identification of Symmetrical RNA Editing Events in the Mitochondria of *Salvia miltiorrhiza* by Strand-specific RNA Sequencing. *Sci Rep.* **2017**, *7*, 1–11. [[CrossRef](#)]
10. Kozik, A.; Rowan, B.A.; Lavelle, D.; Berke, L.; Schranz, M.E.; Michelmore, R.W.; Christensen, A.C. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet.* **2019**, *15*, e1008373. [[CrossRef](#)]
11. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)]
12. Rice, D.W.; Alverson, A.J.; Richardson, A.O.; Young, G.J.; Sanchez-Puerta, M.V.; Munzinger, J.; Barry, K.; Boore, J.L.; Zhang, Y.; DePamphilis, C.W. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **2013**, *342*, 1468–1473. [[CrossRef](#)]
13. 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. USA* **2015**, *112*, E3515–E3524. [[CrossRef](#)]
14. Guo, W.; Grewe, F.; Fan, W.; Young, G.J.; Knoop, V.; Palmer, J.D.; Mower, J.P. *Ginkgo* and *Welwitschia* Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. *Mol. Biol. Evol.* **2016**, *33*, 1448–1460. [[CrossRef](#)]
15. Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* **2012**, *10*, e1001241. [[CrossRef](#)]
16. Montier, L.L.C.; Deng, J.J.; Bai, Y. Number matters: Control of mammalian mitochondrial DNA copy number. *J. Genet. Genom.* **2009**, *36*, 125–131. [[CrossRef](#)]
17. Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* **2011**, *23*, 2499–2513. [[CrossRef](#)]
18. Richardson, A.O.; Rice, D.W.; Young, G.J.; Alverson, A.J.; Palmer, J.D. The “fossilized” mitochondrial genome of *Liriodendron tulipifera* ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* **2013**, *11*, 1–17. [[CrossRef](#)]
19. Sloan, D.B.; Alverson, A.J.; Štorchová, H.; Palmer, J.D.; Taylor, D.R. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evol. Biol.* **2010**, *10*, 1–15. [[CrossRef](#)]
20. Mower, J.P.; Case, A.L.; Floro, E.R.; Willis, J.H. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biol. Evol.* **2012**, *4*, 670–686. [[CrossRef](#)]
21. Burger, G.; Gray, M.W.; Lang, B.F. Mitochondrial genomes: Anything goes. *Trends Genet.* **2003**, *19*, 709–716. [[CrossRef](#)]
22. Mower, J.P. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* **2020**, *53*, 203–213. [[CrossRef](#)]
23. Handa, H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* **2003**, *31*, 5907–5916. [[CrossRef](#)] [[PubMed](#)]
24. Dong, S.; Zhao, C.; Zhang, S.; Zhang, L.; Wu, H.; Liu, H.; Zhu, R.; Jia, Y.; Goffinet, B.; Liu, Y. Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): Conserved genome structure, and ongoing low frequency recombination. *BMC Genom.* **2019**, *20*, 1–14. [[CrossRef](#)]
25. Wu, Z.Q.; Liao, X.Z.; Zhang, X.N.; Tembrock, L.R.; Broz, A. Genomic architectural variation of plant mitochondria—A review of multichromosomal structuring. *J. Syst. Evol.* **2020**, *60*, 160–168. [[CrossRef](#)]
26. Sloan, D.B.; Wu, Z.; Sharbrough, J. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *Plant. Cell* **2018**, *30*, 525–527. [[CrossRef](#)]
27. Tsujimura, M.; Kaneko, T.; Sakamoto, T.; Kimura, S.; Shigyo, M.; Yamagishi, H.; Terachi, T. Multichromosomal structure of the onion mitochondrial genome and a transcript analysis. *Mitochondrion* **2019**, *46*, 179–186. [[CrossRef](#)]
28. Cossetin, L.F.; Santi, E.M.T.; Cossetin, J.F.; Dillmann, J.B.; Baldissera, M.D.; Garlet, Q.I.; de Souza, T.P.; Loebens, L.; Heinzmann, B.M.; Machado, M.M. In vitro safety and efficacy of lavender essential oil (Lamiales: Lamiaceae) as an insecticide against houseflies (Diptera: Muscidae) and blowflies (Diptera: Calliphoridae). *J. Econ. Entomol.* **2018**, *111*, 1974–1982. [[CrossRef](#)]
29. Pathak, N.; Rai, A.; Kumari, R.; Bhat, K. Value addition in sesame: A perspective on bioactive components for enhancing utility and profitability. *Pharmacogn. Rev.* **2014**, *8*, 147. [[PubMed](#)]
30. Zhu, A.; Guo, W.; Jain, K.; Mower, J.P. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Mol. Biol. Evol.* **2014**, *31*, 1228–1236. [[CrossRef](#)]

31. Zhang, T.; Zhang, X.; Hu, S.; Yu, J. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods* **2011**, *7*, 1–8. [[CrossRef](#)] [[PubMed](#)]
32. Van de Paer, C.; Hong-Wa, C.; Jeziorski, C.; Besnard, G. Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae. *Gene* **2016**, *594*, 197–202. [[CrossRef](#)] [[PubMed](#)]
33. Silva, S.R.; Alvarenga, D.O.; Aranguren, Y.; Penha, H.A.; Fernandes, C.C.; Pinheiro, D.G.; Oliveira, M.T.; Michael, T.P.; Miranda, V.F.O.; Varani, A.M. The mitochondrial genome of the terrestrial carnivorous plant *Utricularia reniformis* (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. *PLoS ONE* **2017**, *12*, e0180484. [[CrossRef](#)]
34. Li, J.; Xu, Y.; Shan, Y.; Pei, X.; Yong, S.; Liu, C.; Yu, J. Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. *Planta* **2021**, *254*, 1–16. [[CrossRef](#)]
35. 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.* **2017**, *13*, e1005595. [[CrossRef](#)]
36. Qian, J.; Song, J.; Gao, H.; Zhu, Y.; Xu, J.; Pang, X.; Yao, H.; Sun, C.; Li, X.; Li, C.; et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* **2013**, *8*, e57607. [[CrossRef](#)]
37. Adams, K.L.; Qiu, Y.-L.; Stoutemyer, M.; Palmer, J.D. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9905–9912. [[CrossRef](#)]
38. Sperisen, C.; Büchler, U.; Gugerli, F.; Mátyás, G.; Geburek, T.; Vendramin, G. Tandem repeats in plant mitochondrial genomes: Application to the analysis of population differentiation in the conifer Norway spruce. *Mol. Ecol.* **2001**, *10*, 257–263. [[CrossRef](#)]
39. Nishikawa, T.; Vaughan, D.A.; Kadowaki, K.-I. Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. *Theor. Appl. Genet.* **2005**, *110*, 696–705. [[CrossRef](#)]
40. Dong, S.; Zhao, C.; Chen, F.; Liu, Y.; Zhang, S.; Wu, H.; Zhang, L.; Liu, Y. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics* **2018**, *19*, 1–12. [[CrossRef](#)]
41. Wang, X.-C.; Chen, H.; Yang, D.; Liu, C. Diversity of mitochondrial plastid DNAs (MTPTs) in seed plants. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **2018**, *29*, 635–642. [[CrossRef](#)]
42. Group, A.P.; Chase, M.W.; Christenhusz, M.J.; Fay, M.F.; Byng, J.; Judd, W.; Soltis, D.; Mabberley, D.; Sennikov, A.; Soltis, P. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J. Linn. Soc.* **2016**, *181*, 1–20.
43. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
44. Sloan, D.B. One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure. *New Phytol.* **2013**, *200*, 978–985. [[CrossRef](#)] [[PubMed](#)]
45. Shearman, J.R.; Sonthirod, C.; Naktang, C.; Pootakham, W.; Yoocha, T.; Sangsrakru, D.; Jomchai, N.; Tragoonrung, S.; Tangphat-sornruang, S. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: Assembly and recombination analysis using long PacBio reads. *Sci. Rep.* **2016**, *6*, 1–7. [[CrossRef](#)] [[PubMed](#)]
46. Kazama, T.; Toriyama, K. Whole Mitochondrial Genome Sequencing and Re-Examination of a Cytoplasmic Male Sterility-Associated Gene in Boro-Taichung-Type Cytoplasmic Male Sterile Rice. *PLoS ONE* **2016**, *11*, e0159379. [[CrossRef](#)]
47. Varré, J.-S.; d’Agostino, N.; Touzet, P.; Gallina, S.; Tamburino, R.; Cantarella, C.; Ubrig, E.; Cardi, T.; Drouard, L.; Gualberto, J.M. Complete sequence, multichromosomal architecture and transcriptome analysis of the *Solanum tuberosum* mitochondrial genome. *Int. J. Mol. Sci.* **2019**, *20*, 4788. [[CrossRef](#)] [[PubMed](#)]
48. Emerman, A.B.; Bowman, S.K.; Barry, A.; Henig, N.; Patel, K.M.; Gardner, A.F.; Hendrickson, C.L. NEBNext Direct: A Novel, Rapid, Hybridization-Based Approach for the Capture and Library Conversion of Genomic Regions of Interest. *Curr. Protoc. Mol. Biol.* **2017**, *119*, 7–30. [[CrossRef](#)]
49. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
50. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
51. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
52. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [[CrossRef](#)]
53. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)]
54. Lee, E.; Harris, N.; Gibson, M.; Chetty, R.; Lewis, S. Apollo: A community resource for genome annotation editing. *Bioinformatics* **2009**, *25*, 1836–1837. [[CrossRef](#)]
55. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)]
56. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
57. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)]

58. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
59. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 1–31. [[CrossRef](#)]
60. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* **2020**, *13*, 1194–1202. [[CrossRef](#)]
61. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
62. Milne, I.; Bayer, M.; Cardle, L.; Shaw, P.; Stephen, G.; Wright, F.; Marshall, D. Tablet—next generation sequence assembly visualization. *Bioinformatics* **2010**, *26*, 401–402. [[CrossRef](#)]
63. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [[CrossRef](#)] [[PubMed](#)]
64. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)] [[PubMed](#)]
65. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
66. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)] [[PubMed](#)]
67. Miller, M.A.; Pfeiffer, W.; Schwartz, T. The CIPRES science gateway: A community resource for phylogenetic analyses. In *Proceedings of the 2011 TeraGrid Conference: Extreme digital discovery, Salt Lake City, UT, USA 18–21 July 2011*; Association for Computing Machinery: New York, NY, USA, 2011; pp. 1–8.
68. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)]
69. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)]
70. Wickham, H. ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]