



Article

Comparative Genomics and Phylogenetic Analysis of the Chloroplast Genomes in Three Medicinal *Salvia* Species for Bioexploration

Qing Du ^{1,2,†} , Heyu Yang ^{1,†}, Jing Zeng ^{1,3}, Zhuoer Chen ^{1,3} , Junchen Zhou ^{1,3}, Sihui Sun ^{1,3}, Bin Wang ³ and Chang Liu ^{1,*}

- ¹ Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China
² Key Laboratory of Medicinal Plant Resources of Qinghai-Tibetan Plateau in Qinghai Province, College of Pharmacy, Qinghai Minzu University, Xining 810007, China
³ College of Pharmacy, Xiangnan University, Chenzhou 423000, China
* Correspondence: cliu@implad.ac.cn; Tel.: +86-10-57833114; Fax: +86-10-62899715
† These authors contributed equally to this work.

Abstract: To systematically determine their phylogenetic relationships and develop molecular markers for species discrimination of *Salvia bowleyana*, *S. splendens*, and *S. officinalis*, we sequenced their chloroplast genomes using the Illumina Hiseq 2500 platform. The chloroplast genomes length of *S. bowleyana*, *S. splendens*, and *S. officinalis* were 151,387 bp, 150,604 bp, and 151,163 bp, respectively. The six genes *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, and *ycf2* were present in the IR regions. The chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis* contain 29 tandem repeats; 35, 29, 24 simple-sequence repeats, and 47, 49, 40 interspersed repeats, respectively. The three specific intergenic sequences (IGS) of *rps16-trnQ-UUG*, *trnL-UAA-trnF-GAA*, and *trnM-CAU-atpE* were found to discriminate the 23 *Salvia* species. A total of 91 intergenic spacer sequences were identified through genetic distance analysis. The two specific IGS regions (*trnG-GCC-trnM-CAU* and *ycf3-trnS-GGA*) have the highest K2p value identified in the three studied *Salvia* species. Furthermore, the phylogenetic tree showed that the 23 *Salvia* species formed a monophyletic group. Two pairs of genus-specific DNA barcode primers were found. The results will provide a solid foundation to understand the phylogenetic classification of the three *Salvia* species. Moreover, the specific intergenic regions can provide the probability to discriminate the *Salvia* species between the phenotype and the distinction of gene fragments.



Citation: Du, Q.; Yang, H.; Zeng, J.; Chen, Z.; Zhou, J.; Sun, S.; Wang, B.; Liu, C. Comparative Genomics and Phylogenetic Analysis of the Chloroplast Genomes in Three Medicinal *Salvia* Species for Bioexploration. *Int. J. Mol. Sci.* **2022**, *23*, 12080. <https://doi.org/10.3390/ijms232012080>

Academic Editor: Frank M. You

Received: 13 August 2022

Accepted: 26 September 2022

Published: 11 October 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/>).

Keywords: *Salvia bowleyana*; *Salvia splendens*; *Salvia officinalis*; chloroplast genome; comparative genomics; repeat analysis; hypervariable regions; DNA barcode; phylogenetic analysis

1. Introduction

The Lamiaceae family is the sixth-largest family of flowering plants. It includes 10 subfamilies, 220 genera, and 3500 species [1]. Most of the species are mainly distributed in Asia, Europe, and Africa. In historical evolution, the family of Lamiaceae is most closely related to the family of Verbenaceae and Violaceae [2]. In China, 99 genera and more than 800 species in the Lamiaceae family are found, which include about 1050 *Salvia* species. Among them, 78 varied species and 32 variants mostly grow in tropical or temperate areas [3]. Regarding the classification development of the *Salvia* genus, Bentham [4] once divided it into four subgenera and 12 groups and Briquet [5] divided it into 8 subgenera and 17 groups. There are also different taxonomic studies about the *Salvia* genus in the different regions, for example, subgenus *Calosphaea* was divided into 91 groups by the American scientist Carl Epling [6], and the genus increased to 102 within the subsequent 20 years. In Europe and Africa, the *Salvia* genus is divided into four subgenera and eight

groups in the flora of the USSR [7], whereas in the flora of Europe, the *Salvia* genus is divided into five groups [8]. The botanist of academician Wu Zhengyi in East Asia [9] divided the Chinese *Salvia* genus into five subgenus groups and 18 subbranches. Based on the molecular systems and data of *rbcL* and *trnL-F*, the genus *Salvia* is not monophyletic; it has the relationship of sister taxa embedded with the genera of *Rosmarinus*, *Perovskia*, *Dorystaechas*, *Meriandra*, and *Zhumeria* [10]. Meanwhile, through molecular systematics and morphological evidence, 15 species from the 5 genera of *Rosmarinus*, *Perovskia*, *Spear*, *Meriandra*, and *Zhumeria* were formally merged into the generalized *Salvia* genus with 10 identified independent clades [11]. The 11 *Salvia* species in Japan were clustered in one branch based on the comparative data of *rbcL*, *trnL-F*, and ITS sequences [12]. The molecular systems of 38 *Salvia* species in China were classified using the ITS, *rbcL*, *psbA-trnH*, and *matK* sequences in China, showing that the *Salvia* genus was clustered into one clade from China and Japan, except for the species of *Salvia* *deserta*, and the three subgenera defined in Chinese plants are not the monophyletic groups [13]. Using a study on the divergence of ITS, ETS, *psbA-trnH*, *ycf1-rps15*, *trnL-trnF*, and *rbcL* sequences, the phylogenetic tree containing 78 species and 10 variants confirmed that the *Salvia* of East Asian is a monophyletic group, formally naming the clade IV (S. Glutinaria Clade) as East Asian *Salvia* with eight groups [14]. More interestingly and meaningfully, the 345 species belonging to 77 Lamiaceae genera have been classified and clustered into phylogenetic groups based on the aspects of phytochemical constituents and treatment of the various disorders through the analysis of NRI and NTI metrics. The results showed that the *Salvia boweyara* had an effect on the treatment of reproductive and hepatic disorders [15]. Therefore, there are certain differences in the *Salvia* species from the aspects of their morphological characters, chemical composition, treatment effects on diseases, and molecular markers. We are looking forward to carrying out the integration of taxonomic research from various aspects to elucidate the classification status of the *Salvia* genus in the family.

The chloroplast is the essential organelle in plants. The chloroplast genome contains a variety of genes closely related to photosynthesis [16], evolution [17], and applications in genetic engineering [18]. In general, the chloroplast genome encodes more than 120 genes. These genes can be divided into three types [19] related to transcription and translation, photosynthesis, and the biosynthesis of amino acids and fatty acids. The genes distributed in the large-single copy (LSC) and small-single copy (SSC) regions are mainly related to photosynthetic systems I (PSA) and systems II (PSB). They also include large subunits of Rubisco (encoded by *rbcL*) [20], the *tRNA* gene (*tRNA*), the ATP enzyme gene (ATP), the NADH plastid-masking oxidoreductase gene (NADH), and the RNA polymerase gene (RPO) [21]. The genes distributed in the IRs region are mainly the genes encoding rRNA (RPS), including 16S and 23S genes, the intermediate genes being separated by encoding 4.5S rRNA, and 5S rRNA and 2tRNA genes, and some genes with unknown gene function [22].

The genes from chloroplast genomes can be used in species identification [23], phylogenetic evolution [24], genetic transformation [25], and molecular breeding of medicinal plants [26], providing basic data for resource identification and conservation. The sequences in the chloroplast genomes of medicinal plants, such as *psbA-trnH*, *matK*, and *rbcL*, have been widely used for DNA molecular identification, and have now been developed for the analysis of polymorphic locus combinations of multiple genes and gene spacers [27]. To date, the chloroplast genomes of the 14 *Salvia* species in the Lamiaceae family have been reported [28–30].

Compared with the diversification of nuclear and mitochondrial genomes, the comprehensive development of chloroplast genomes could provide a basic database for further exploration regarding structural variation, characteristics, genetic evolution, and chemicals. Therefore, we sequenced and analyzed the chloroplast genomes of three *Salvia* species for the first time to identify divergence hotspots of phylogenetic genome regions and detect the applicability of phylogenomics for further resolving the evolutionary and systematic relationship in the *Salvia* genus of the Lamiaceae family.

2. Results

2.1. Morphological Characteristics of the Three *Salvia* Species

The three *Salvia* species have the common specifications of the Lamiaceae family: quadrangular stem, opposite leaves, corolla flower lip, and four nutlets. However, they have the obvious distinction from the phenotype of flower colors (1) varying from pink and purple (*S. bowleyana* and *S. officinalis*) to red (*S. splendens*). Moreover, the three *Salvia* species are perennial herbs with oblong or oval leaves (2), cymose inflorescences (3), and nutlets. Nevertheless, for *S. bowleyana*, the leaves are glabrous on both sides, only the veins are slightly pilose, and the top of the fruit is hairy (Figure 1a). For *S. splendens*, the stems, leaves on both sides, and petioles are not glabrous with glandular spots below. The fruits have irregular folds at the top, and narrow wings at the edge (Figure 1b). For *Salvia officinalis*, the stems, many branches, leaf surfaces, and petioles are covered with white short villi. The fruits are smooth and hairless (4) (Figure 1c) [1].

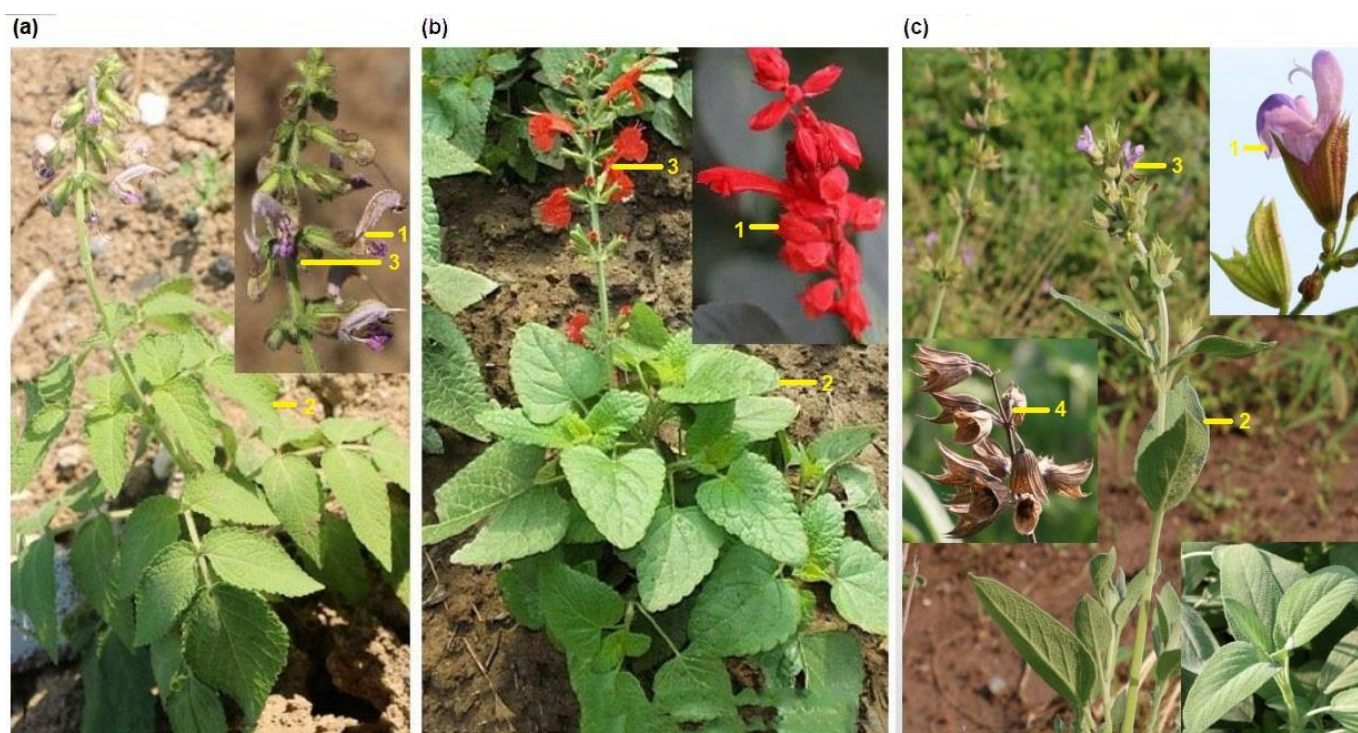


Figure 1. Three *Salvia* species of the Lamiaceae family. *S. bowleyana* (a), *S. splendens* (b), and *S. officinalis* (c). The numbers 1–4 shown in yellow refer to the four different characteristics among the three species, which include the colors of flowers, shape of leaves, type of inflorescences, and appearance of fruits. 1: flower; 2: leaf; 3: inflorescences; 4: fruit.

2.2. Gene Compositions Comparison of 23 *Salvia* Species

Schematic representations of *S. bowleyana*, *S. splendens*, and *S. officinalis* chloroplast genomes are shown in Figure 2, respectively. The total assembled length of them was 151,387 bp, 150,604 bp, and 151,163 bp, respectively. The lengths of LSC, SSC, and dual inverted repeat (IR) regions in the three chloroplast genomes were 82,772 bp, 17,573 bp, and 51,042 bp for *S. bowleyana*; 82,181 bp, 17,857 bp, and 50,566 bp for *S. splendens*; 82,429 bp, 17,510 bp, and 51,224 bp for *S. officinalis*. The GC contents of the three chloroplast genomes were 38.01%, 38.04%, and 38.04%, respectively (Table 1, Table S1).

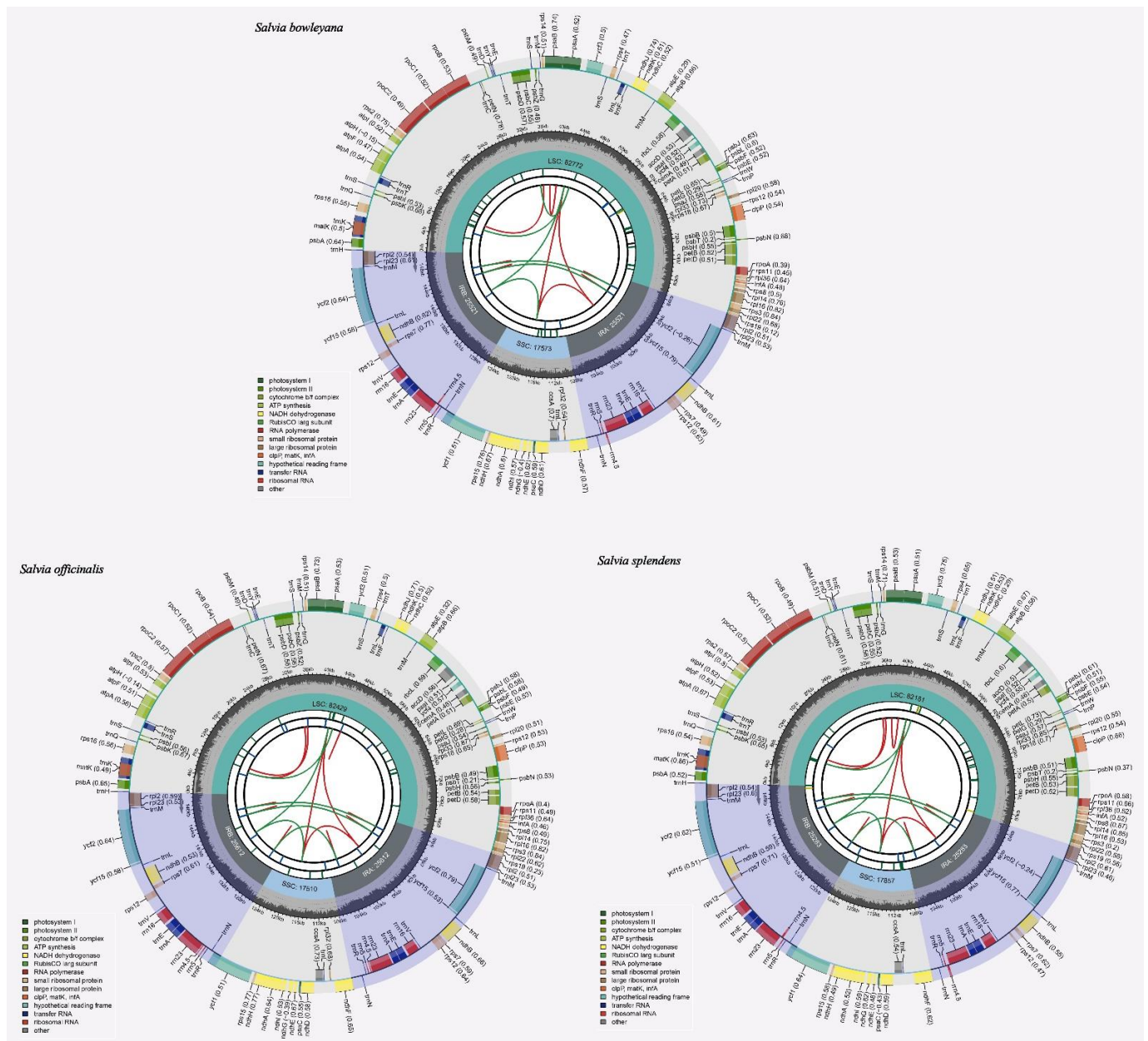


Figure 2. Graphic representation of features identified in the *S. bowleyana* (a), *S. splendens* (b), and *S. officinalis* (c) chloroplast genomes. Each map contains seven circles. From the center going outward, the first circle shows the distributed repeats connected with red (the forward direction) and green (the reverse direction) arcs. The next circle shows the tandem repeats marked with short bars. The third circle shows the microsatellite sequences as short bars. The fourth circle shows the size of the LSC and SSC. The fifth circle shows the IRA and IRB. The sixth circle shows the GC contents along the plastome. The seventh circle shows the genes having different colors based on their functional groups.

The chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis* contained 131, 130, and 131 genes, respectively, including 80, 79, and 80 protein-coding genes, 36 tRNA genes, and 8 rRNA genes (Table S1). There are 14 PCGs (*rps12* (×2), *rps7* (×2), *rpl2* (×2), *rpl23* (×2), *ndhB* (×2), *ycf2* (×2), and *ycf15* (×2)), 14 tRNA genes (*trnA*-UGC (×2), *trnE*-UUC (×2), *trnM*-CAU (×2), *trnL*-CAA (×2), *trnN*-GUU (×2), *trnR*-ACG (×2), *trnV*-GAC (×2)), and 8 rRNA genes (*rnn16S* (×2), *rnn23S* (×2), *rnn4.5S* (×2), and *rnn5S* (×2)) located in the both IRA and IRb regions (Table 1), respectively. Among the three genomes, twenty-two genes commonly exhibited introns, of which seven tRNA genes (*trnK*-UUU,

trnL-UAA, *trnC*-ACA, *trnE*-UUC ($\times 2$), and *trnA*-UGC ($\times 2$)), and twelve cis-splicing CDS genes (*rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petD*, *rpl16*, *rpl2* ($\times 2$), *ndhB* ($\times 2$), and *ndhA*) had a single intron. In particular, the three genes had one intron in the special species, of which both genes *trnT*-CGU and *petB* are identified in the species of *S. bowleyana* and *S. splendens*. IN contrast, the protein-coding gene *petB* was only shown in *S. officinalis*. Notably, two CDS genes of *ycf3* and *clpP* displayed two introns and three exons (Table 2, Figure S1). Additionally, those containing the intron gene *trnK*-UUU, making up the *matK*, had the largest intron in the three chloroplast genomes of *Salvia* species (2522 bp, 2494 bp, and 2517 bp, respectively). Except for the plants of Pteridophyta and parasitic species, the chloroplasts of land plants commonly contain the *matK* mature enzyme gene in the intron of the lysine tRNA-K (UUU) gene, for instance, species of *Cuscuta* genus [31–33], which acts as a splicing factor for introns of the highly structured ribozyme group II [34,35]. Furthermore, the three segments of *rps12* genes were located in the region of LSC, IRa, and IRb of the chloroplast genomes, respectively. The *rps12* gene was split into two introns; one intron between exon 2 and 3 was 528 bp in length, and another intron between exon 1 and 2 was about 28 kb in length (Table 2, Figure S2). The latter intron is trans-spliced to produce mature *rps12* mRNA (Figure S2) [36]. The exon 1 and the two copies of exons are trans-spliced together to form two transcripts. The arrows indicate the sense direction of the genes (Figures S1 and S2).

Among the 23 *Salvia* species, the lengths of the total genome, LSC, SSC, and IR varied from 150,604 bp to 153,995 bp, from 82,129 bp to 84,775 bp, from 17,464 bp to 17,875 bp, and from 25,283 bp to 25,815 bp, respectively. The percentage of GC contents for the total genome, LSC, SSC, and IRs regions varied from 37.94% to 38.05%, from 36.07% to 36.23%, from 31.63% to 32.07%, and from 43.06% to 43.20%. The gene numbers of the total genes, protein-encoding genes, and tRNA genes ranged from 130 to 133, from 85 to 88, and from 36 to 37, respectively. The chloroplast genomes in all 23 *Salvia* species encoded two copies of *rrn16S*, *rrn23S*, *rrn4.5S*, and *rrn5S* (Table S1).

Table 1. Comparison of the gene contents in the chloroplast genomes of *Salvia bowleyana*, *Salvia splendens*, and *Salvia officinalis*.

Species/Items		<i>S. bowleyana</i>	<i>S. splendens</i>	<i>S. officinalis</i>
Gene Function	Gene Type	Gene Name		
tRNA	tRNA genes	36 <i>trn</i> genes (include one intron in 8 genes)	36 <i>trn</i> genes (include one intron in 8 genes)	36 <i>trn</i> genes (include one intron in 8 genes)
Photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>		
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>		
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> , <i>ycf3</i>		
	Ribosomal RNAs	<i>rrn16s^a</i> , <i>rrn16s^b</i> , <i>rrn23s^a</i> , <i>rrn23s^b</i> , <i>rrn4.5s^a</i> , <i>rrn4.5s^b</i> , <i>rrn5s^a</i> , <i>rrn5s^b</i>		
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i>		
Gene expression	Small subunit of ribosome	<i>rps11</i> , <i>rps12^L</i> , <i>rps12^a</i> , <i>rps12^b</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i> , <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7^a</i> , <i>rps7^b</i> , <i>rps8</i>		
	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> , <i>rpl2^a</i> , <i>rpl2^b</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23^a</i> , <i>rpl23^b</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>		
	Subunits of NADH-dehydrogenase	<i>ndhA</i> , <i>ndhB^a</i> , <i>ndhB^b</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>		
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>		
	Ribulose diphosphate carboxylase subunit	<i>rbcL</i>		
	Subunit of acetyl-CoA-carboxylase	<i>accD</i>		
	C-type cytochrome synthase	<i>ccsA</i>		
Other genes	Protease	<i>clpP</i>		
	Translation initiation factor	<i>infA</i>		
	Mature enzyme	<i>matK</i>		
	Envelope membrane protein	<i>cemA</i>		
Unknown functions	Conservative open reading frame	<i>ycf1^{s-b}</i> , <i>ycf2^a</i> , <i>ycf2^b</i> , <i>ycf15^a</i> , <i>ycf15^b</i> , <i>ycf4</i>		

L: LSC region; a: IRa region; b: IRb region; s-b: Across the SSC and IRb regions.

Table 2. The lengths of introns and exons for the splitting genes in the chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis*.

Gene Name	Strand	Initial Position–Final Position			Length (bp)														
		<i>S. bowleyana</i> A	<i>S. splendens</i> B	<i>S. officinalis</i> C	The First Exon			The First Intron			The Second Exon			The Second Intron			The Third Exon		
					A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
<i>trnK-UUU</i>	-	1672–4266	1684–4250	1703–4292	37	37	37	2522	2494	2517	36	36	36						
<i>rps16</i>	-	4835–5945	4819–5917	4863–5972	40	40	40	874	862	873	197	197	197						
<i>trnT-CGU</i>	+	9001–9755	8765–9528	/	35	35	/	677	686	/	43	43	/						
<i>trnS-CGA</i>	+	/	/	8621–9377	/	/	32	/	/	665	/	/	60						
<i>atpF</i>	-	11,742–12,989	11,506–12,764	11,353–12,606	145	145	145	693	704	699	410	410	410						
<i>rpoC1</i>	-	20,712–23,525	20,528–23,339	20,399–23,215	430	430	430	759	757	762	1625	1625	1625						
<i>ycf3</i>	-	41,963–43,894	41,526–43,464	41,641–43,591	129	129	129	696	702	706	228	228	228	726	727	735	153	153	153
<i>trnL-UAA</i>	+	46,799–47,338	46,350–46,917	46,202–46,773	35	35	35	455	483	487	50	50	50						
<i>trnC-ACA</i>	-	50,870–51,518	50,236–50,881	50,440–51,087	38	38	38	555	552	554	56	56	56						
<i>rps12^L</i>		68,691–68,804	68,105–68,218	68,355–68,468	114	114	114												
<i>clpP</i>	-	68,928–70,839	68,342–70,250	68,591–70,509	71	71	71	692	703	711	294	294	294	629	615	617	226	226	226
<i>petB</i>	+	73,746–75,096	73,171–74,533	/	6	6	/	703	715	/	642	642	/						
<i>petD</i>	-	75,290–76,492	74,721–75,904	74,979–76,169	8	8	8	720	701	708	475	475	475						
<i>rpl16</i>	-	79,937–81,217	79,325–80,600	79,599–80,867	9	9	9	873	868	861	399	399	399						
<i>rpl2</i>	-	82,875–84,357	82,266–83,757	82,532–84,019	391	391	391	658	667	663	434	434	434						
<i>ndhB</i>	+	93,058–95,211	92,464–94,617	92,711–94,918	721	721	775	675	675	675	758	758	758						
<i>rps12^b</i>		96,061–96,844	96,018–96,260	95,714–96,507	114	114	114	/	/	/	232	243	232	528	/	538	26	/	26
<i>trnE-UUC</i>	+	100,535–101,546	99,979–100,997	100,210–101,229	32	32	32	940	947	948	40	40	40						
<i>trnA-UGC</i>	+	101,611–102,478	101,062–101,938	101,294–102,171	37	37	37	795	804	805	36	36	36						
<i>ndhA</i>	-	117,349–119,425	116,488–118,588	117,038–119,137	553	553	553	985	1009	1008	539	539	539						
<i>trnA-UGC</i>	-	131,682–132,549	130,848–131,724	131,422–132,299	37	37	37	795	804	805	36	36	36						
<i>trnE-UUC</i>	-	132,614–133,625	131,789–132,807	132,364–133,383	32	32	32	940	947	948	40	40	40						
<i>rps12^a</i>		137,316–138,099	136,526–136,768	137,086–137,879	114	114	114	/	/	/	232	241	232	528	/	528	26	/	26
<i>ndhB</i>	+	138,949–141,102	138,169–140,322	138,675–140,882	721	721	775	675	675	675	758	758	758						
<i>rpl2</i>	+	149,803–151,285	149,029–150,520	149,574–151,061	391	391	391	658	667	663	434	434	434						

"+" indicates a positive chain; "-" indicates a negative chain; A: *S. bowleyana*; B: *S. splendens*; C: *S. officinalis*. L: LSC region; a: IRa region; b: IRb region.

2.3. Gene Loss Analysis of the Chloroplast Genomes from 41 Species in the Lamiaceae Family

The gene losses of chloroplast genomes were analyzed in the 41 species of the Lamiaceae family that originated from the phylogenetic tree (Table 3). These species originated from 8 genera (*Salvia*, *Rosmarinus*, *Agastache*, *Dracocephalum*, *Ajuga*, *Leonurus*, *Elsholtzia*, and *Caryopteris*) of the Lamiaceae family. In the dual IR regions of chloroplast genomes, one of the *rpl20* genes was stable and found in all 41 species; however, another one was found only in *D. heterophyllum*. Therefore, the intact *rpl20* gene often can be used as the molecular signature gene in the angiosperm [37]. In addition, one of the *ycf1* genes was across the SSC and IRb regions, the other pseudogene was across the SSC and IRa regions. Loss of the first *ycf1* gene was observed in five chloroplast genomes of *A. campylanthoides*, *A. ciliata*, *A. decumbens*, *A. lupulina*, and *A. nipponensis*. Loss of the second one was not found in the chloroplast genomes of twenty-eight species except in the twelve chloroplast genomes from the six *Salvia* genus (*S. digitaloides*, *S. daiguii*, *S. meiliensis*, *S. chanryoenica*, *S. yangii*, and *S. nilotica*), *A. rugosa*, the four *Dracocephalum* genus (*D. heterophyllum*, *D. taliense*, *D. tanguticum*, and *D. moldavica*), and *L. japonicas*. As reported, in a total of 420 species, 357 species could be distinguished using *ycf1* by means of specific primers designed for the amplification of these regions [38]. Moreover, the losses of the *ycf15* genes occurred in five chloroplast genomes (*S. hispanica*, *S. tiliifolia*, *S. chanryoenica*, *A. forrestii*, and *E. densa*). Although the gene function of *ycf15* genes is unknown, the transcriptome analyses of the *Camellia* genus revealed that the *ycf15* gene was transcribed as a precursor polycistronic transcript which contained *ycf2*, *ycf15*, and antisense *trnL-CAA* [39]. Furthermore, the six genes in the LSC region, e.g., *petN*, *accD*, *rps2*, *rps16*, *rps18*, and *rps19* were absent in the chloroplast genomes of *C. trichosphaera*, *R. officinalis*, *D. moldavica*, *E. densa*, *D. heterophyllum*, and *L. japonicus*, respectively. In contrast, in the SSC region, loss of the *rpl32* and *ndhD* genes was found only in *S. splendens* and *C. mongholica* chloroplast genomes, respectively. Surprisingly, loss of the *rpl32* gene can be transferred to the nucleus from the chloroplast genome of *Euphorbia schimperi* and this can be verified through the method of being sequenced in the nuclear transcriptome of *E. schimperi* (Table 3) [40]. The type of gene loss was mostly affirmed to be consistent with the topology of the evolutionary tree.

2.4. Analysis of Simple Sequence Repeats Polymorphism in the 23 *Salvia* Chloroplast Genomes

Repeat sequences have been commonly used as genetic markers to understand the evolution of the genus in the same family. Scattered (interspersed) repetition and tandem repetition sequences consisting of simple sequence repeats (SSRs) were analyzed in the 23 *Salvia* chloroplast genomes (Table S2, Figure 3). We analyzed the content and percentage of SSR sequences in the 23 *Salvia* species. The results showed that 16, 12, and 10 SSR contained "A" as the repeat unit and 18, 14, and 14 SSR contained "T" as the repeat unit among the total 34, 26, and 24 mononucleotide repeats (Table S2) in the chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis*, respectively. Moreover, the mononucleotide numbers of "A" and "T" as the repeat unit have an obvious difference. From the statistical results, the number of Poly A and Poly T repeats varied from 6 (*S. yangii*) to 16 (*S. bowleyana* and *S. miltiorrhiza f. alba*), from 9 (*S. plebeia*) to 21 (*S. prattii*). Rare numbers of Poly C and Poly G repeats were found only in the chloroplast genomes of *S. hispanica*, *S. plebeia*, and *S. meiliensis* [41]. One SSR with "AT" as the repeat unit was found in the eight *Salvia* chloroplast genomes of *S. splendens*, *S. digitaloides*, *S. daiguii*, *S. hispanica*, *S. tiliifolia*, *S. chanryoenica*, *S. prattii*, and *S. roborowskii*. Di-nucleotide SSR contained "TA" as the repeat unit in twelve chloroplast genomes of *S. bowleyana*, *S. bulleyana*, *S. przewalskii*, *S. yunnanensis*, *S. miltiorrhiza f. alba*, *S. chanryoenica*, *S. prattii*, *S. roborowskii*, *S. splendens*, *S. daiguii*, *S. hispanica*, and *S. tiliifolia*, respectively. Nevertheless, one trinucleotide SSR with "AAT" as the repeat unit was found in the chloroplast genome of *S. yunnanensis* (Table S2). The mononucleotide repeat unit is the most abundant type of the SSR repeats and it accounted for the proportion from 88% to 100% through comprehensive statistics of chloroplast genomes in the 23 *Salvia* species.

Table 3. Gene losses in the different regions of the 41 chloroplast genomes from the Lamiaceae family.

Genus	Name of Species	The Genes in the IR Region				The Genes in the LSC Region						The Genes in the SSC Region	
		<i>rpl20_copy</i>	<i>ycf1</i>	<i>ycf1_copy</i>	<i>ycf15</i>	<i>petN</i>	<i>accD</i>	<i>rps2</i>	<i>rps16</i>	<i>rps18</i>	<i>rps19 *</i>	<i>rpl32</i>	<i>ndhD</i>
Salvia	<i>S. bowleyana</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. splendens</i>	-	+	-	+	+	+	+	+	+	+	-	+
	<i>S. officinalis</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. bulleyana</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. digitaloides</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>S. japonica</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. plebeia</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. przewalskii</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. yunnanensis</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. miltiorrhiza</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. daiguii</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>S. miltiorrhiza</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>f.alba</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. meiliensis</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. hispanica</i>	-	+	-	-	+	+	+	+	+	+	+	+
	<i>S. merjamie</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>S. sclarea</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. petrophila</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>S. tiliifolia</i>	-	+	-	-	+	+	+	+	+	+	+	+
	<i>S. chanryoenica</i>	-	+	+	-	+	+	+	+	+	+	+	+
<i>S. yangii</i>	-	+	+	+	+	+	+	+	+	+	+	+	
<i>S. Prattii Hemsl.</i>	-	+	-	+	+	+	+	+	+	+	+	+	
<i>S. roborowskii</i>	-	+	-	+	+	+	+	+	+	+	+	+	
<i>S. nilotica</i>	-	+	+	+	+	+	+	+	+	+	+	+	
Rosmarinus	<i>R. officinalis</i>	-	+	-	+	+	-	+	+	+	+	+	+
Agastache	<i>A. rugosa</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>D. heterophyllum</i>	+	+	+	+	+	+	+	+	-	+	+	+
Dracocephalum	<i>D. taliense</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>D. tanguticum</i>	-	+	+	+	+	+	+	+	+	+	+	+
	<i>D. moldavica</i>	-	+	+	+	+	+	-	+	+	+	+	+

Table 3. Cont.

Genus	Name of Species	The Genes in the IR Region				The Genes in the LSC Region					The Genes in the SSC Region		
		<i>rpl20_copy</i>	<i>ycf1</i>	<i>ycf1_copy</i>	<i>ycf15</i>	<i>petN</i>	<i>accD</i>	<i>rps2</i>	<i>rps16</i>	<i>rps18</i>	<i>rps19</i> *	<i>rpl32</i>	<i>ndhD</i>
<i>Ajuga</i>	<i>A. forrestii</i>	-	+	-	-	+	+	+	+	+	+	+	+
	<i>A. campylanthoides</i>	-	-	-	+	+	+	+	+	+	+	+	+
	<i>A. ciliata</i>	-	-	-	+	+	+	+	+	+	+	+	+
	<i>A. decumbens</i>	-	-	-	+	+	+	+	+	+	+	+	+
	<i>A. lupulina</i>	-	-	-	+	+	+	+	+	+	+	+	+
	<i>A. nipponensis</i>	-	-	-	+	+	+	+	+	+	+	+	+
<i>Leonurus</i>	<i>L. japonicus</i>	-	+	+	+	+	+	+	+	+	-	+	+
<i>Elsholtzia</i>	<i>E. densa</i>	-	+	-	-	+	+	+	-	+	+	+	+
<i>Caryopteris</i>	<i>C. trichosphaera</i>	-	+	-	+	-	+	+	+	+	+	+	+
	<i>C. mongholica</i>	-	+	-	+	+	+	+	+	+	+	+	-
	<i>C. incana</i>	-	+	-	+	+	+	+	+	+	+	+	+
	<i>C. forrestii</i>	-	+	-	+	+	+	+	+	+	+	+	+

The +/- refers to the presence/absence of a gene in each species that does not have the gene. "+": presence; "-" absence; * *rps19* is across the area of LSC and IRb (add family and order information).

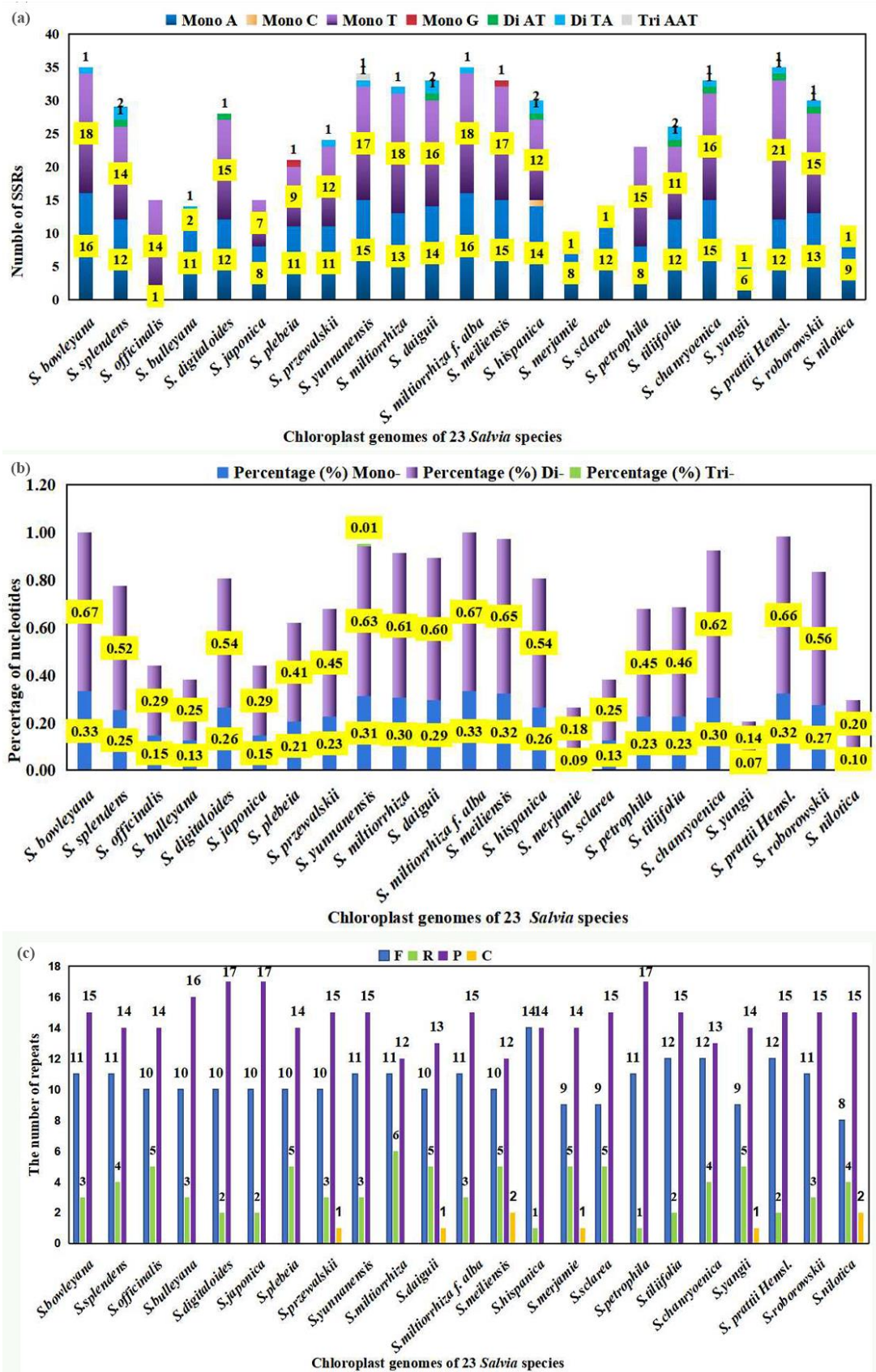


Figure 3. The repeats analysis in the 23 *Salvia* species. The number of diverse repeats has been marked on the strips in different colors. The abscissa represents the chloroplast genomes of 23 *Salvia* species; the ordinates represent the number of SSRs (a), the percentage of nucleotides (b), and the number of

repeats (c). In (a), the different types of SSRs are filled in blue (mono A), orange (mono C), purple (mono T), red (mono G), green (di AT), blue (di TA), and gray (Tri AAT) together marked with the detailed quantum in yellow and black within the diverse columns. In (b), the percentage of mononucleotides, dinucleotides, and trinucleotides is filled in blue, purple, and green together marked with the detailed quantum in yellow and black within the diverse columns. In (c), the number of repeats in the types of forward repeats (F), reverse repeats (R), palindromic repeats (P), and complement repeats (C) is filled in blue, green, purple, and orange together marked with the detailed quantum in black above the diverse columns. Mono: mononucleotide; Di: dinucleotide; Tri: trinucleotide; F: forward repeats, R: reverse repeats; P: palindromic repeats, and C: complement repeats.

2.5. Repeat Sequences Analysis in the Chloroplast Genomes of 23 *Salvia* Species

Except for in the SSR analysis of the 23 *Salvia* chloroplast genome, 29 tandem repeats by each species were identified for all the four kinds of tandem repeats, including the forward repeats, reverse repeats, palindromic repeats, and complement repeats in the chloroplast genomes of *S. bowleyana* (11 forward repeats, 3 reverse repeats, and 15 palindromic repeats), *S. splendens* (11 forward repeats, 4 reverse repeats, and 14 palindromic repeats) and *S. officinalis* (10 forward repeats, 5 reverse repeats, and 14 palindromic repeats), respectively. The greatest numbers of repeat types were forward repeats and palindromic repeats, while the numbers of reverse repeats and complement repeats were less and the latter were found only in the six chloroplast genomes, including *S. przewalskii*, *S. daiguii*, *S. meiliensis*, *S. merjamie*, *S. yangii*, and *S. nilotica*. The comparison of the number of predicted tandem repeats is shown in Tables S3–S6, and Figure 3c.

Among the 23 *Salvia* chloroplast genomes of the interspersed repeats, the number of palindromic and direct repeats varied from 14 (*S. merjamie*, *S. sclarea*, and *S. daiguii*) to 26 (*S. miltiorrhiza*, *S. petrophila*, *S. prattii*, *S. roborowskii*, and *S. splendens*). The number of tandem repeats will be reduced by more than half and diversified from 6 (*S. bowleyana*, *S. splendens*, *S. plebeia*, *S. miltiorrhiza*, and *S. miltiorrhiza* f. *alba*) to 24 (*S. japonica*) while the similarity among the repeat unit sequences $\geq 90\%$. The e-values of interspersed repeats varied from 7.65×10^{-23} to 6.07×10^{-4} . In this study, 47 interspersed repeats (25 palindromic repeats and 22 direct repeats), 49 interspersed repeats (23 palindromic repeats and 26 direct repeats), and 40 interspersed repeats (20 palindromic repeats and 20 direct repeats) were identified in the chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis*, respectively, with the length of repeat units 1, 2 being between 30 bp and 63 bp (Figure 3, Tables S7–S9).

2.6. Structures of the IR Boundaries and Gene Features from 23 *Salvia* Species

The IR boundaries' structure was analyzed in the 23 *Salvia* chloroplast genomes of the Lamiaceae family. From the analysis, six distinct genes, *rpl22*, *rps19*, *rpl2* ($\times 2$), *ycf1*, *ndhF*, and *psbA*, were most explicitly found in the diverse regions or at the border regions of 23 chloroplast genomes (Figure 4). Furthermore, the variation range of these gene lengths was similar and did not exceed 2%. The genes of *rpl22* and *psbA* were located in the LSC region, whereas *rpl2* genes were located in the two IR regions in these species. One of the *rps19* genes was located at the border area of LSC and IRb in all species. In addition, small fragments of the *rps19* genes (*rps19* pseudogene) were found at the border regions of the LSC and IRa in the fourteen chloroplast genomes of *S. bulleyana*, *S. digitaloides*, *S. japonica*, *S. plebeia*, *S. przewalskii*, *S. miltiorrhiza*, *S. daiguii*, *S. miltiorrhiza* f. *alba*, *S. meiliensis*, *S. petrophila*, *S. yangii*, *S. nilotica*, *S. prattii*, *S. roborowskii*. In contrast, the *ycf1* genes traversed the border regions of SSC and IRb in all 23 *Salvia* species, while *ycf1* gene fragments (*ycf1* pseudogene) were found at the border regions of SSC and IRa in six *Salvia* chloroplast genomes (*S. merjamie*, *S. digitaloides*, *S. daiguii*, *S. chanryoenica*, *S. nilotica*, and *S. yangii*). Besides, *ndhF* genes were located at the border regions of IRa and SSC in all 23 species. The IRa/LSC boundary positions were located on the *trnH* genes in the five chloroplast genomes of

S. chanryoenica, *S. splendens*, *S. nilotica*, *S. yangii*, and *S. tiliifolia*. Notably, a fragment of the *trnN* gene located in the IRb region of the *Salvia splendens* chloroplast genome (Figure 4) is often found in the *Cymbidium* genus among the photosynthetic orchids [42].

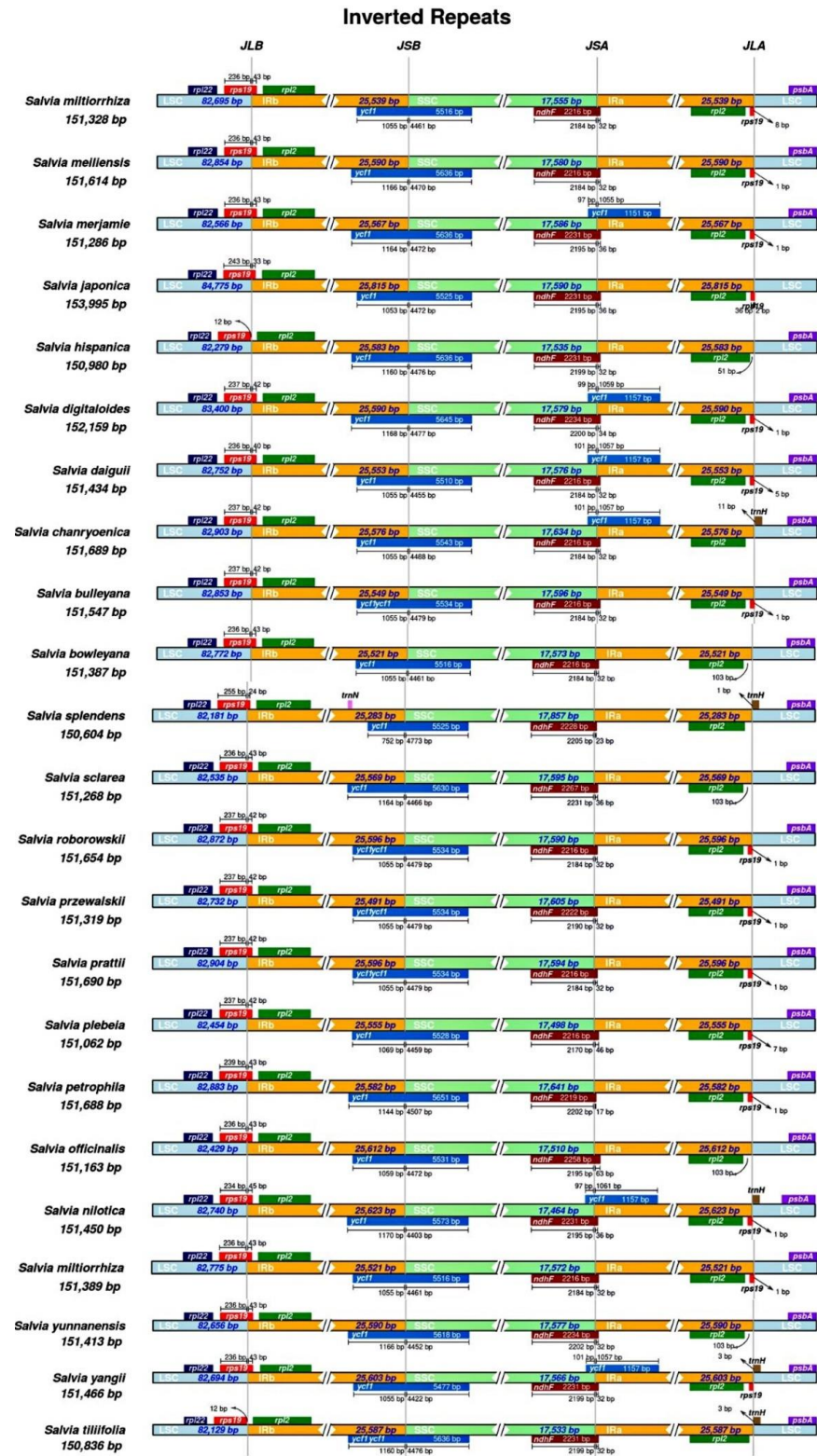


Figure 4. Comparison of the border areas among the LSC, SSC, and IR regions in the 23 *Salvia* chloroplast genomes. The genes are denoted by colored boxes. The gaps between the genes and the boundaries are indicated by the base lengths (bp). The thin lines represent the connection

points of each area, and the information of the genes near the connection points is shown in the figures. The species' Latin names and the length of the plastomes are shown on the left. The JLB, JSB, JSA, and JLA represent junction sites of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC, respectively. The distance from the start and end positions of different genes across junction sites is shown above or below the corresponding genes.

2.7. The Discrepancy of the 23 *Salvia* Chloroplast Genomes

The structures of chloroplast genomes are highly conserved. The medicinal plants can be accurately identified and distinguished by the comparison of barcodes from the whole chloroplast genome. The sequences of chloroplast genomes in the 23 *Salvia* species were analyzed using mVISTA, and the alignments were visualized with the *Salvia bowleyana* chloroplast genome as the reference genome (Figure S3). We found the sequences of 23 *Salvia* chloroplast genomes were mostly identically conserved except for the three variable areas located in the intergenic regions of the LSC region. The first one was the IGS region (*rps16-trnQ-UUG*) found in the nine *Salvia* chloroplast genomes (*S. officinalis*, *S. japonica*, *S. sclarea*, *S. meliensis*, *S. hispanica*, *S. tiliifolia*, *S. yangii*, *S. splendens*, *S. nilotica*) (Figure S3 (A)). The second one was the IGS region (*trnL-UAA-trnF-GAA*) varied in the chloroplast genome of *S. chanryoenica* (Figure S3 (B)). The last one was the IGS region (*trnM (cau)-atpE*) diversified in the three chloroplast genomes of *S. chanryoenica*, *S. hispanica*, and *S. japonica* (Figure S3 (C)).

2.8. Identification and Cloning of Hypervariable Regions

It is significant to develop molecular markers in the chloroplast genomes of plants by identifying the highly variable sites. In general, the large K2p distances indicate a high degree of sequence divergences. We analyzed the genetic distance among the IGS regions in the chloroplast genomes of 23 *Salvia* species. The results showed that K2p distances of 91 IGS regions ranged from 0.00 to 21.03 (Table S10). Among them, 30 IGS regions had K2p distances varying from 3.52 to 21.03 (Figure 5a). Particularly, five IGS regions had higher K2p values diversified from 5.80 to 21.03, which were the regions of *trnL-UAG-ccsA* (21.03), *rps16-trnQ-UUG* (13.19), *ccsA-ndhD* (7.68), *rps15-ycf1* (6.40), and *ndhE-ndhG* (5.80). Thus, these five regions of IGS can be suitable candidates for developing molecular markers in the 23 *Salvia* species. Meanwhile, the five IGS regions with higher K2p values were identified in the three studied *Salvia* species including *rps16-trnQ-UUG* (21.35), *trnG-GCC-trnM-CAU* (12.91), *ccsA-ndhD* (12.14), *ycf3-trnS-GGA* (10.92), and *rps15-ycf1* (9.67) (Figure 5b, Table S11).

Interestingly, the two IGS regions of *trnG-GCC-trnM-CAU* (Figure S4a, M1) and *ycf3-trnS-GGA* (Figure S4b, M2) were specific in the three studied species. We cloned the two regions and acquired the sequences of M1 (~300 bp) and M2 (~800bp) using Sanger sequencing (Table S12). Then, we comparatively analyzed the two molecular markers (MMs) among the three studied *Salvia* species to determine the variations, including indels and single nucleotide polymorphisms (SNP) (Table 4, M1 and M2). The amplification products of the two IGS were checked and the strips were clearly shown on the agarose gel (Figure S4). From the peak map (up) and sequencing results (down) of the three studied *Salvia* species with the pairs of primers from M1 and M2 (Figure 6), four variant loci of SNP or indels were found among them and marked A, B, C, and D, respectively, at Figure 6. Therefore, the three *Salvia* species can be successfully discriminated based on these SNP and indel loci by separately or unitedly using the two M1 and M2 molecular markers. The intergenic region's SNP (iSNP) has the potential to directly affect the protein structures or expression levels in accordance with the particular localization; therefore, it may affect the plant traits or genetic mechanisms [43]. In contrast to markers of the *Salvia* genus, two markers derived from the IGS regions of *petN-psbM* and *psaJ-rpl33* can be successfully used to distinguish the five *Alpinia* species [44].

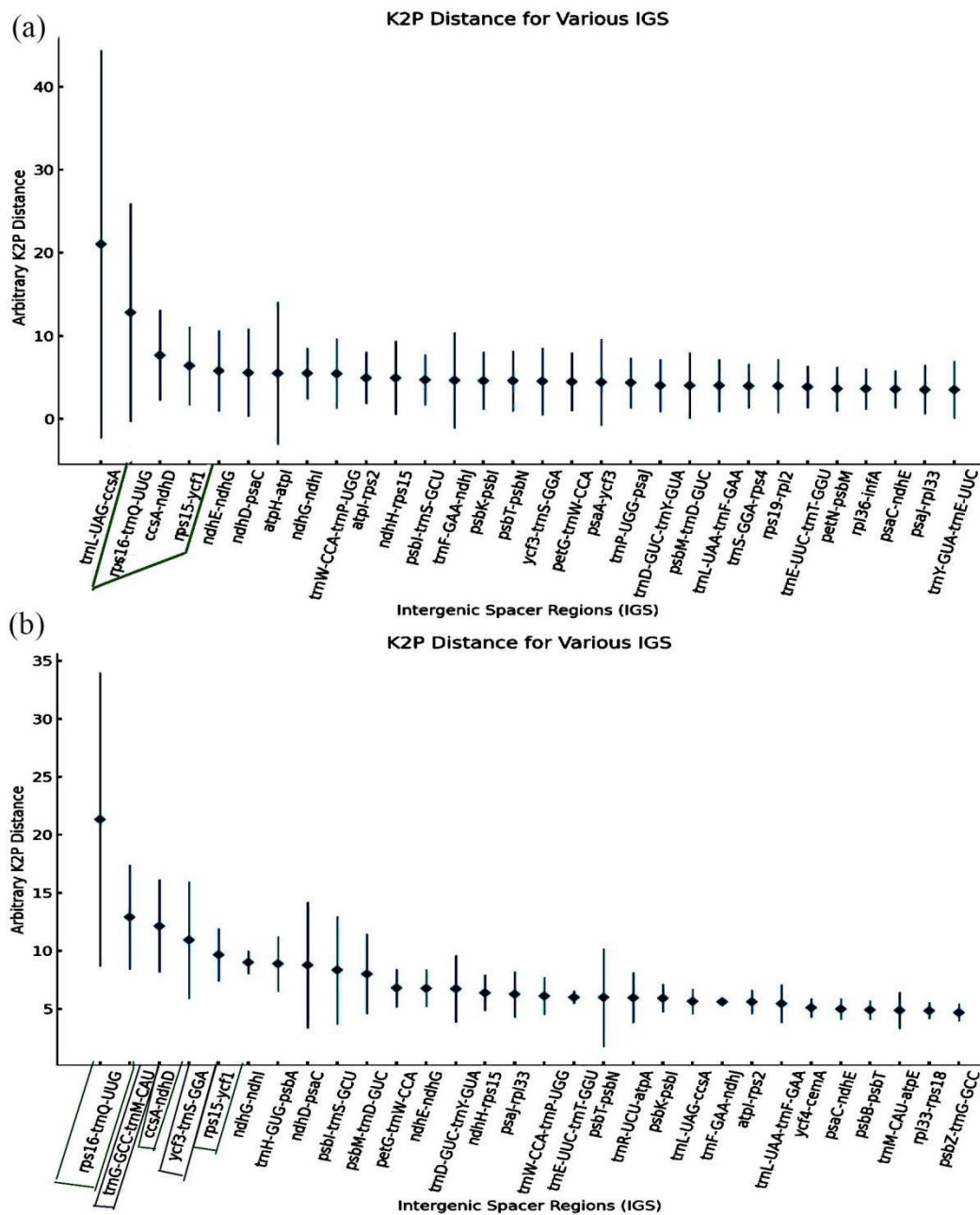


Figure 5. Average K2p distances for intergenic spacer regions in the chloroplast genomes of 23 *Salvia* species (a) and the three studies species (b) from the Lamiaceae family. The K2p distances were calculated among 23 *Salvia* chloroplast genomes in pairs. The black dots represent the average value of the three pairs. The error bars represent the standard error among the three pairs. Among the five IGSs with the highest K2p values, the IGSs marked in the green frame are common in the chloroplast genomes between 23 *Salvia* species and the three studies species, while the marked in purple are the specific IGSs in the chloroplast genomes of the three studies *Salvia* species.

Table 4. Primers for amplifying DNA barcodes to distinguish *Salvia* species in the Lamiaceae family.

No	Species	Conserved Sequences for Designing Forward Primers	Conserved Sequences for Designing Reverse Primers	IGS
M1	<i>S. bowleyana</i> ,	GCGGATATGGTCCGAATGGTAAA	GCAGTTTGGTAGCTCGCAAG	<i>trnG</i> -GCC- <i>trnM</i> -CAU
M2	<i>S. splendens</i> , <i>S. officinalis</i>	TGAAGTTGTCGGAATTATTGCA	AATGCTACGCCTTGAACCAC	<i>ycf3</i> - <i>trnS</i> -GGA
M3	23 <i>Salvia</i> species	TTTTCCCTTCTACCCC	AAAAAAAGATGTTGCGGAGACAG- GATTTGAACCCGTGAC- CTCAAGGTTATGAGCCTT- GCGAGCTACCAAACCTCTACCCC- GCGCTGAAGAGAAGAA	<i>trnM</i> -CAU- <i>atpE</i>
M4		TTACATAGTTATGGTTCATTTACA- TTAACATCTAATTAAT	TTTTTTCATTGTACAACGAAC	<i>ccsA</i> - <i>ndhD</i>

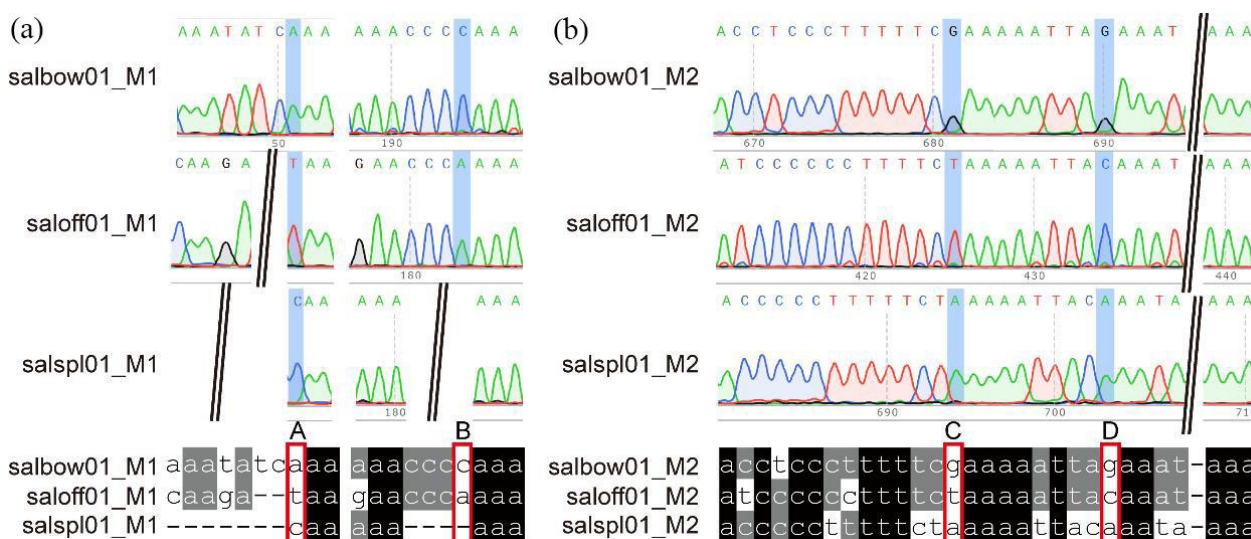


Figure 6. The peak map (up) and sequencing results (down) of the three studied *Salvia* species with the pairs of primers M1 (a) and M2 (b). The symbols of *salbow01_M1* (a) and *salbow01_M2* (b) are the sequencing results and peak map from one sample of *Salvia bowleyana*; the symbols of the *saloff01_M1* and *saloff01_M2* are the one sample of *Salvia officinalis*, and the symbols of *salspl01_M1* and *salspl01_M2* are the one sample of *Salvia splendens*. The variant bases have been marked A, B, C, and D in a red frame of the sequences.

2.9. Identification and Comparison of the Genus-Specific DNA Barcodes Primer and Sequences

Primers can be designed from highly variable intergenic spacer sequences for PCR amplification. Then, we can distinguish the 23 *Salvia* species in the Lamiaceae family by sequence alignment and analysis using *ecoPrimers* software. After comparison, the two conservative intervals can be amplified through the designed PCR amplification primers to distinguish the 23 *salvia* genus. The primer sequences are shown in Table 4 (M3 and M4). Surprisingly, the two pairs of primers can be used to amplify the sequences of *trnM*-CAU-*atpE* and *ccsA*-*ndhD* after comparison between the *Salvia* chloroplast genomes and the BlastN database. Furthermore, the alignment results based on the blast database indicate that the two pair primers can also especially suit other distinct species, e.g., *Scutellaria* genus (Lamiaceae), *Camellia* genus (Theaceae), *Styrax* genus (Styracaceae), *Melissa* genus (Lamiaceae), *Eucalyptus* genus (Myrtaceae), etc.

2.10. Phylogenetic Analysis

The sequences of chloroplast genomes are a valuable database for the research of the evolutionary relationship in plants. To determine the phylogenetic positions of the three *Salvia* species in the Lamiaceae family, 80 protein sequences were extracted using the *PhyloSuite* software from the 43 chloroplast genomes in the species (Table S1). Among

them, 25 shared CDS proteins sequences were found present in 43 species, including *rpl14*, *rpl33*, *rpl36*, *rps7*, *rps14*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbN*, *psaB*, *psaC*, *psaI*, *petA*, *petG*, *petL*, *ndhC*, *ndhG*, *cemA*, *atpA*, *atpB*, *atpH*, *atpI*, and *ycf4* genes. We identified 29 proteins shared by 37 Lamiaceae species. However, there were only 25 proteins commonly shared in the studied 43 species. The other four proteins, including *atpE*, *psbA*, *psbJ*, and *psbM*, were only shared in 37 Lamiaceae species. The multiple sequence alignments of the 29 proteins are shown in Figure S5. Using *L. chuanxiong* (Apiaceae family) and *P. notoginseng* (Araliaceae family) as the outgroups, the phylogenetic tree was generated by three methods of maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ) based on the above-described data of whole chloroplast genomes. The three phylogenetic trees showed the same evolutionary relationship, in which 41 species including 37 species of the Lamiaceae family and four species of the Verbenaceae family were clustered together with 6 obvious clades. Among them, five species including *Dracocephalum* species (*D. heterophyllum*, *D. Taliense*, *D. tanguticum*, and *D. moldavica*) and *A. rugose* were clustered into one branch; in contrast, 23 *Salvia* species and one *Rosmarinus* species (*R. officinalis*) were clustered into one branch with six subbranches (Figure 7). In addition, six species from the *Ajuga* genus and four species from the *Caryopteris* genus were clustered into the other two branches, respectively. Single species of *L. japonicus* and *Elsholtzia densa* were gathered into one branch, partly, whereas the species of outgroups were more distantly related to other species. The ML bootstrap showed strong support with bootstrap values of 100% for eight nodes. The phylogenetic results resolved 26 nodes with bootstrap support values of 54–100 and that of 17 nodes were $\geq 74\%$ (Figure 7).

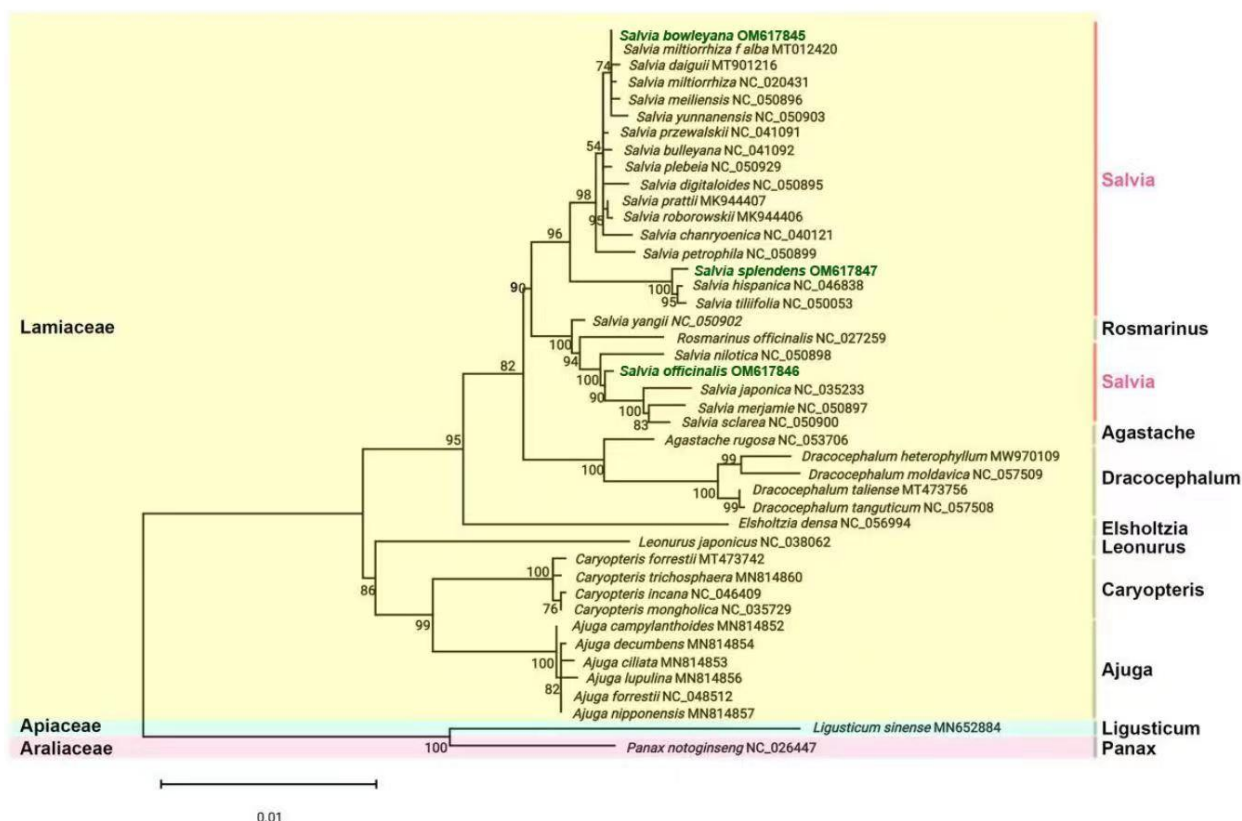


Figure 7. The phylogenetic relationships of the 43 species. These include 37 Lamiales species and 4 species of the Verbenaceae family, while the two species *Ligusticum chuanxiong* and *Panax notoginseng* were used as the outgroup. The tree was constructed with the sequences of 80 CDSs shared among all 43 species by using the three methods of maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ). Bootstrap supports were calculated from 1000 replicates.

3. Discussion

3.1. The Characteristics of Chloroplast Genomes and Genes in the *Salvia* Genus

In the chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis*, the total numbers of protein-coding genes were identical except that of *S. splendens* was one less. The total numbers of tRNA and rRNA genes were the same as those of other *Salvia* species. These results indicated that the chloroplast genomes of the *Salvia* species were highly conserved. The selected 41 species from the Lamiaceae family and the two outgroup species (*L. chuanxiong* and *P. notoginseng*) possessed similar pharmacological effects, such as promoting blood circulation for removing blood stasis, increasing coronary flow, improving microcirculation, protecting the heart, improving the body hypoxia resistance, and having anti-hepatitis, antitumor, and antiviral effects [45]. Chloroplasts play an irreplaceable role in the formation of chemicals and the development of phenotypes due to the genes from nuclear, and mitochondrial genomes. However, the variability of the nuclear genome was found to be higher than that of the chloroplast genome and mitochondrial genome, as reported from the average genetic distance among all the strains of CWR and cultivated rice [46]. Therefore, it is indispensable to analyze the genetic divergence in the chloroplast genomes of *Salvia* species.

3.2. The Divergence between IGS Regions of the *Salvia* Genus Compared to Other Plants

It makes sense that the DNA sequences of the hypervariable regions and comparison of chloroplast genomes in three IGS regions of *rps16-trnQ-UUG*, *trnL-UAA-trnF-GAA*, and *trnM(cau)-atpE* can be used to distinguish the ten *Salvia* species (*S. officinalis*, *S. japonica*, *S. sclarea*, *S. meiliensis*, *S. hispanica*, *S. tiliifolia*, *S. yangii*, *S. splendens*, *S. nilotica*, and *S. chanryoenica*). The first IGS region has been found in the species of *Zingiber officinale* and *Cofeae* alliance [47,48]. The second one commonly occurs in the angiosperm [49]. The last one has diversified and some parts of the oldest mtDNAs of *trnV(uac)-trnM(cau)-atpE-atpB-rbcL* were transferred from cpDNA to mtDNA since they have a common ancestor in extant gymnosperms and angiosperms [50]. As reported, the phylogenetic relationships in the *Eurystachys* clade were reconstructed utilizing nuclear ribosomal DNA sequences (nrETS, 5S-NTS) from 148 accessions into 12 well-supported genera, including widely recognized and well-defined segregates such as *Prasium* and *Sideritis* [51].

In contrast, the special IGS regions of the two iSNPs, namely *trnG-GCC-trnM-CAU* and *ycf3-trnS-GGA*, were used to discriminate the three studied *Salvia* species. In previous studies, most of the SNPs were found in intergenic sequences, and the *trnG-GCC-trnM-CAU* was one of the maximum number of SNPs found four times to distinguish the six *Saccharum* species [52]. Meantime, the variable hotspot regions of *ycf3-trnS-GGA* also can be useful as the candidate DNA barcodes for Adoxaceae and Caprifoliaceae species, and also for assessing interspecific divergence in Dipsacales species [53]. In addition, research has shown that the *rps14* gene can be used as a DNA barcode for the identification of 34 Lamiaceae species collected from plants in the Pakistan area [54].

Therefore, the DNA barcode primers identified in the study can be potentially developed for the identification and phytotaxonomy of genus *Salvia* species through the divergence IGS regions.

3.3. The Functional Features of IR Regions and Genes of the *Salvia* Genus together with Other Plants

The sequences of IR can complement a certain segment of the upstream sequence downstream of the same DNA strand. They can then form a hairpin structure with a double helix stem and a single-stranded ring with a DNA double helix. The sequence between two reverse repeat units forms a single chain loop. Two copies are separated by a sequence or no interval sequence, which is in reverse series, and will form a specific palindrome sequence (P) [55]. Compared to the IRLC between the Papilionoideae subfamily [56] and the Lamiaceae family, they have the four common genes of *ndhB*, *rpl23*, *ycf1*, and *ycf15*.

In the IR regions, the genes of *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, and *ycf2* were present in the chloroplast genomes of 41 species, and these genes have a special function in the area of gene expressions. There are the five hypothetical coding regions genes of *ycf1*, *ycf2*, *ycf4*, *ycf15*, and two open reading frames (ORF42 and ORF56), which are also found in the chloroplast genomes of the other species, such as *Clerodendranthus spicatus* [57]. Both genes *ycf3* and *ycf4* were present in the LSC region of the 41 species' chloroplast genomes. The sequence of *ycf3* is conserved in plants and contains three tetratricopeptide repeats (TPR), which can act as the functions essential for the accumulation of the photosystem I (PSI) complex through a post-translational level [58,59]. The *ycf4* gene forms modules that mediate PSI assembly and facilitate the integration of peripheral PSI subunits and LHCl into the PSI reaction center subcomplex [60].

4. Materials and Methods

4.1. Plant Photos and Materials

Salvia bowleyana, *S. splendens*, and *S. officinalis* are the three characteristic plants from the *Salvia* genus of the Lamiaceae family. The photos of *Salvia bowleyana* and *S. splendens* were provided by the Jiangsu Nanjing Botanical Garden and the Civic Park of Guangdong, and identified by Professor Peng LQ (Chuzhou Hospital of Integrated Traditional Chinese and Western medicine, Anhui Province). In addition, the *S. officinalis* photo is from Dr. Qi YD's team (Dr. Zhao Xinlei, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, zhaoxinlei2009@sina.com) (Figure 1). Furthermore, we collected the young leaves of *S. bowleyana*, *S. splendens*, and *S. officinalis* from the Guangxi Medical Botanical Garden, Nanning, Guangxi, China (Geospatial coordinates: 22°51'35.9" N, 108°23'00.5" E) and dried them with silica gel immediately for total genomic DNA isolation and sequencing of the chloroplast genome. The voucher specimens were deposited at the Institute of Medicinal Plant Development under the voucher number: implad201910237, implad201808155, and implad20170492, respectively (contact person: HM Chen; email: hmchen@implad.ac.cn). Moreover, the fresh leaves of three plants were used to clone the DNA barcode sequences from Jiujiang city, Jiangxi province (29°11'36.6" N, 114°47'52.9" E), Songjiang, Shanghai city (30°56'49.5" N, 121°15'23.3" E), and Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing city (116°25' E, 39°47' N), and Shunyi Dist., Beijing city (116°46'56" E, 40°5'41" N).

4.2. DNA Extraction, Determination of DNA Quality, and PCR Amplification Products

Total genomic DNA was extracted from the 20 dried leaves for sequencing of chloroplast genome and fresh leaves were taken from the three single plants for cloning the DNA barcode sequences using a plant genomic DNA kit (Tiangen Biotech, Beijing, China). The extraction of DNA is a universal technology as the flowchart shows in Dr. Li's research [61]. We firstly ground plant tissues with liquid nitrogen. Then, we added the GPS buffer, RNase A, GPA buffer, absolute ethyl alcohol, deprotein fluid RD, bleach solution PW, and elution buffer TB. Next, we loaded the collection solutions on a column. During the courses, we mixed the solution and centrifuged each step. Lastly, the DNA bound to the column was eluted with an elution buffer. The DNA purity and amplification products were detected by 1.0% agarose gel electrophoresis stained with ethidium bromide alongside a 100 bp ladder (New England Biolabs, Ipswich, MA, USA) using the DNA marker as the reference to determine the size of the amplified fragments (Takara) [62]. Otherwise, DNA concentration was determined using the Nanodrop spectrophotometer 2000 (Thermo, Waltham, Massachusetts, USA). Furthermore, the extraction of chloroplast DNA (cpDNA) for whole plastid genome sequencing should undergo three stages: separation of chloroplasts from cells, purification of chloroplasts, and isolation of cpDNA [63].

4.3. Chloroplast Genome Sequencing, Assembly, Annotation, and Manual Curation

DNA extracts containing the DNA concentration of 500 ng were applied to construct a library with lengths of short-insert fragments of 500 bps. The library was sequenced in a pair-end model with a read length of 150 bp on an Illumina HiSeq 2500 platform in accordance with the MiSeq platform provided by the manufacturer's directions [64]. The sequencing raw data were acquired from *S. bowleyana*, *S. splendens*, and *S. officinalis* with sizes of 7.1 Gbs, 6.8 Gbs, and 7.02 Gbs and 250bps pair-end read lengths, respectively. The raw data were submitted to the NCBI database and assigned the Sequence Read Archive (SRA) accession numbers SRR14415377, SRR17843445, and SRR17853381, respectively. The raw reads were filtered using Trimmomatic 0.35 with default parameters to remove adapters and low-quality bases [65]. The three chloroplast genomes were assembled using the NOVOPlasty (v 4.2) software [66] with the default parameters and the rbcL sequences as the seed. After that, we annotated these genomes using the CpGAVAS2 web service (<http://www.herbalgenomics.org/cpgavas2/>, accessed on 1 May 2022) [67]. The annotation errors were manually corrected using the Apollo software [68]. The assembly and the annotation results of *S. bowleyana*, *S. splendens*, and *S. officinalis* were submitted to GenBank with the accession numbers OM617845, OM617847, and OM617846, respectively.

4.4. Visualization and Analysis of Genome Content, cis- and Trans-Splicing genes

The chloroplast genome structure, cis-splicing genes, and trans-splicing PCGs were visualized using CPGview-RSG software (<http://www.1kmpg.cn/cpgview/>, accessed on 1 May 2022) [69]. The gene contents of 41 studied species (Table S1) were analyzed including the length of the complete genome sequences and the four regions, all genes, CDS, tRNAs, and rRNAs.

4.5. Repeat Analysis

We annotated the repeat sequences using the CPGAVAS2 for the chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis*. The SSRs of 23 *Salvia* species were identified using MISA software (<http://pgrc.ipk-gatersleben.de/misa/>, accessed on 1 May 2022) [70], also called the microsatellite sequence. The minimum numbers of repeat units for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide, and hexagenucleotide were set as 10, 5, 4, 3, 3, 3 and 3, respectively. The minimum distance between the 2 SSRs was set to 100 bp. If the distance was less than 100 bp, the two SSRs were treated as a composite microsatellite. The tandem repeats sequence (TRS) of the 23 *Salvia* chloroplast genomes was predicted using the Tandem Repeats Finder (TRF) software [71]. The interspersed repeats sequence (IRS) was predicted using the REPuter program (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>, accessed on 1 May 2022), with the parameters as follows: maximum computed repeats = 30 and minimal repeat size = 8) [72]. The comparison of the chloroplast genomes was conducted using VMATCH software (Professor Stefan Kurtz, Computer Science at the Center for Bioinformatics, University of Hamburg, Germany) [73].

4.6. Comparative Genomic Analysis

We downloaded 40 chloroplast genomes sequences from the GenBank database including 38 species from the Lamiaceae family and two outgroups (*Ligusticum chuanxiong* from the Apiaceae family and *Panax notoginseng* from the Araliaceae family, for further analysis. The boundaries of the LSC, SSC, and IR regions boundary of chloroplast genomes from 23 *Salvia* species were visualized using the IR scope software (<https://irscope.shinyapps.io/irapp/>) [74] and the characteristic genes including the diverse areas were analyzed. The chloroplast genome sequences of 23 species from *Salvia* genera were compared with the annotated *S. bowleyana* chloroplast as the reference using the mVISTA program in a Shuffle-LAGAN mode with default parameters (Rank VISTA probability threshold = 0.5) [75,76]. The genetic distances of IGS regions from the

chloroplast genomes of 23 *Salvia* species were calculated using the distmat program from EMBOSS (v6.3.1) [77] with the Kimura 2-parameters (K2p) evolutionary model [78].

4.7. Primer Identification and Design, PCR Amplification, Sequencing, and Analysis of Genus-Specific DNA Barcode Sequences

To discover DNA barcode sequences that can distinguish the 23 *Salvia* species, especially the three studied species, we analyzed the PCR amplification primers from their chloroplast genome sequences using ecoPrimers software [79]. Moreover, the sequences of two pairs of primers were compared to the other species through the CBI Multiple Sequence Alignment Viewer (Version 1.21.0, Max Seq Difference = 0.75) from the BLASTN website (<https://blast.ncbi.nlm.nih.gov/>) [80]. The two pairs of specific primers were designed to differently amplify the specific IGS regions identified in the three studied *Salvia* species by the Primer 3 software [81]. The PCR amplification system for genus-specific DNA barcode sequences of each reaction included 12.5 µL of 2 Taq PCR Master Mix (TransGen Biotech), 1.0 µL of each primer (0.4 µM), 2.0 µL of extracted template DNA, and ddH₂O added to a final volume of 25 µL [82]. A negative control (Milli-Q water in place of DNA template) was included in each PCR to ensure there was no contamination. All the amplifications were performed on a Pro-Flex PCR system (Applied Biosystems, Waltham, MA, USA) instrument with the amplification procedures: degeneration 94 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 60 s, and a final extension step at 72 °C for 2 min. The amplification products were saved at 4 °C and sequenced at SinoGenoMax Co., Ltd. using the Sanger sequencing platform with the same cloning primers on the ABI Prism 3730 Genetic Analyzer (Applied Biosystems, USA). The sequences were spliced and analyzed by the GeneDoc software (3.2) [83].

4.8. Phylogenetic Analysis

We developed phylogenetic analysis using the concatenated coding sequences (CDS) of the chloroplast genomes from 43 species. These include 37 Lamiaceae species (*S. bowleyana*, *S. splendens*, *S. officinalis*, *S. bulleyana*, *S. digitaloides*, *S. japonica*, *S. plebeia*, *S. przewalskii*, *S. yunnanensis*, *S. miltiorrhiza*, *S. daiguii*, *S. sclarea*, *S. meiliensis*, *S. miltiorrhiza f.alba*, *S. hispanica*, *S. merjamie*, *S. petrophila*, *S. tiliifolia*, *S. chanryoenica*, *S. yangii*, *S. prattii*, *S. roborowskii*, *S. nilotica*, *R. officinalis*, *A. rugosa*, *D. heterophyllum*, *D. taliense*, *D. tanguticum*, *D. moldavica*, *A. forrestii*, *A. campylanthoides*, *A. ciliata*, *A. decumbens*, *A. lupulina*, *A. nipponensis*, *L. japonicus*, and *Elsholtzia densa*) and 4 species of the Verbenaceae family (*C. trichosphaera*, *C. mongholica*, *C. incana*, and *C. forrestii*), while the two species *Ligusticum chuanxiong* from the Apiaceae family and *Panax notoginseng* from the Araliaceae family were used as the outgroup. The chloroplast genome sequences were downloaded from GenBank (Table S1). The shared CDSs were extracted, concatenated using PhyloSuite (v1.2.2) [84], and aligned using MAFFT (v7.313) [85]. Moreover, the sequences of 29 CDSs with small variations among the 37 chloroplast genomes from the Lamiaceae family were compared using the GeneDoc (3.2) [83]. Phylogenetic analysis was conducted based on three methods of maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ) implemented in IQ-TREE (v1.6.8) [86] under the TVM+F+I+G4 nucleotide substitution model. The reliability of the phylogenetic tree was assessed by bootstrap analysis with 1000 replications and was visualized using MEGA-X [87].

5. Conclusions

The complete chloroplast genomes of *S. bowleyana*, *S. splendens*, and *S. officinalis* were acquired using Illumina sequencing technology. These three species can be easily discriminated from the phenotype. Phylogenetic analysis showed that 23 *Salvia* species and one *Rosmarinus* genus were clustered into one branch with six subbranches, of which the three studied species were included in the diverse branches. The sequence divergence found seven sites of IGS regions: *rps16-trnQ-UUG*, *trnL-UAA-trnF-GAA*, *trnM-CAU-atpE*, *trnL-UAG-ccsA*, *ccsA-ndhD*, *rps15-ycf1*, and *ndhE-ndhG*. Notably, the two IGS regions of

trnG-GCC-trnM-CAU and *ycf3-trnS-GGA* were identified in the three studied *Salvia* species. The sequences' divergence had a high variability and indicates they can be developed as DNA markers for further identification and phytotaxonomy of the *Salvia* genus. Overall, the data obtained will contribute to further development of the authentication, diversity, ecology, taxonomy, phylogenetic evolution, and conservation of the *Salvia* genus in China.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232012080/s1>.

Author Contributions: Conceptualization, B.W. and C.L.; methodology, Q.D. and H.Y.; software, Q.D., H.Y., J.Z. (Jing Zeng), J.Z. (Junchen Zhou), S.S. and Z.C.; validation, Q.D., H.Y. and C.L.; formal analysis, Q.D., H.Y., J.Z. (Jing Zeng), Z.C., S.S. and Z.C.; data curation, Q.D. and H.Y.; writing—original draft preparation, Q.D.; writing—review and editing, C.L. and B.W.; visualization, Q.D.; project administration, Q.D. and C.L.; funding acquisition, C.L. and B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funds from the Chinese Academy of Medical Sciences, Innovation Funds for Medical Sciences (CIFMS) [2021-I2M-1-022], National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], National Science Foundation [81872966], Qinghai Provincial Key Laboratory of Phytochemistry of Qinghai Tibet Plateau [2020-ZJ-Y20], Hunan Technological Innovation Guidance Project (2018SK52001). 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 chloroplast genome sequence data of *S. bowleyana*, *S. splendens*, and *S. officinalis* are openly available in the GenBank database with accession numbers OM617845, OM617847, and OM617846 (<https://www.ncbi.nlm.nih.gov>). The associated BioProject, SRA, and BioSample numbers are PRJNA726222, PRJNA769231, and PRJNA769230; SAMN18926173, SAMN22106482, and SAMN22106467; SRR14415377, SRR17843445, and SRR17853381, respectively.

Acknowledgments: We would like to thank Liqiang Wang, Mei Jiang, Haimei Chen, Xinlei Zhao, Haodong Chen, Rongjun Fan, Xiaoying Pei, Jing Li, and Yufang Ma who provided support for data analysis.

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

References

- Li, H.W.; Ian, C.H. *Lamiaceae. Editorial board, Chinese Academy of Sciences. Flora of China*, 17th ed.; Science Press: Beijing, China, 1994.
- Ratray, R.D.; Van Wyk, B.E. The Botanical, Chemical and Ethnobotanical Diversity of Southern African Lamiaceae. *Molecules* **2021**, *26*, 3712. [[CrossRef](#)] [[PubMed](#)]
- Li, B.; Cantino, P.D.; Olmstead, R.G.; Bramley, G.L.; Xiang, C.L.; Ma, Z.H.; Tan, Y.H.; Zhang, D.X. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Sci. Rep.* **2016**, *6*, 34343. [[CrossRef](#)] [[PubMed](#)]
- Bentham, G.; Labiatae Bentham, G.; Hooker, J.D. *Genera Plantarum*; Reeve and Co: London, UK, 1876; Volume 2, pp. 1160–1223.
- Briquet, J. Labiatae. In *Engler & Prantl, Die natürlichen Pflanzenfamilien IV, 3a*; W. Engelmann: Leipzig, Germany, 1895–1897; Volume 4, pp. 183–375.
- Valdés, L.J., III; Díaz, J.; Paul, A.G. Ethnopharmacology of ska *María Pastora* (*Salvia divinorum*, Epling AND Játiva-M.). *J. Ethnopharmacol.* **1983**, *7*, 287–312. [[CrossRef](#)]
- Pobedimova, E.G. Rod Shalfei-*Salvia*, L. In *Flora SSSR*; The McGraw-Hill Companies, Inc.: Moscow, Russia, 1954; Volume 21.
- Hedge, I.C. *Salvia* L. In *Flora Europaea*; Tutin, T.G., Ed.; Cambridge University Press: Cambridge, UK, 1972; p. 188.
- Wu, Z.Y.; Sun, X.C. *Salvia* Genus. In *Flora of China*; Wu, Z.Y., Li, X.W., Eds.; Science Press: Beijing, China, 1977; pp. 70–196.
- Walker, J.B.; Sytsma, K.J. Staminal evolution in the genus *Salvia* (Lamiaceae): Molecular phylogenetic evidence for multiple origins of the staminal lever. *Ann. Bot.* **2007**, *100*, 375–391. [[CrossRef](#)]
- Drew, B.T.; González-Gallegos, J.G.; Xiang, C.L.; Kriebel, R.; Drummond, C.P.; Walker, J.B.; Sytsma, K.J. *Salvia* united: The greatest good for the greatest number. *Taxon* **2017**, *66*, 133–145. [[CrossRef](#)]
- Takano, A.; Okada, H. Phylogenetic relationships among subgenera, species, and varieties of Japanese *Salvia* L. (Lamiaceae). *J. Plant. Res.* **2011**, *124*, 245–252. [[CrossRef](#)]
- Li, M.H.; Li, Q.Q.; Liu, Y.Z.; Cui, Z.H.; Zhang, N.; Huang, L.Q.; Xiao, P.G. Pharmacophylogenetic study on plants of genus *Salvia* L. from China. *China Herb. Med.* **2013**, *5*, 164–181.

14. Hu, G.X.; Takano, A.; Drew, B.T.; Liu, E.D.; Soltis, D.E.; Soltis, P.S.; Peng, H.; Xiang, C.L. Phylogeny and staminal evolution of *Salvia* (Lamiaceae, Nepetoideae) in East Asia. *Ann. Bot.* **2018**, *122*, 649–668. [[CrossRef](#)]
15. Zaman, W.; Ye, J.; Hmad, M.; Saqib, S.; Shinwari, Z.K.; Chen, Z.D. Phylogenetic exploration of traditional chinese medicinal plants: A case study on lamiaceae (angiosperms). *Pak. J. Bot.* **2022**, *54*, 1033–1040. [[CrossRef](#)]
16. Green, B.R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **2019**, *66*, 34–44. [[CrossRef](#)]
17. Xiao-Ming, Z.; Junrui, W.; Li, F.; Sha, L.; Hongbo, P.; Lan, Q.; Jing, L.; Yan, S.; Weihua, Q.; Lifang, Z.; et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* **2019**, *7*, 1555. [[CrossRef](#)]
18. Lo'pez, E.-J. Plastid biogenesis, between light and shadows. *J. Exp. Bot.* **2007**, *58*, 11–26. [[CrossRef](#)]
19. Glynn, J.M.; Miyagishima, S.; Yoder, D.W.; Osteryoung, K.W.; Vitha, S. Chloroplast Division. *Traffic* **2007**, *8*, 451–461. [[CrossRef](#)]
20. Ichikawa, K.; Miyake, C.; Iwano, M.; Sekine, M.; Shinmyo, A.; Kato, K. Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit translation is regulated in a small subunit-independent manner in the expanded leaves of tobacco. *Plant Cell Physiol.* **2008**, *49*, 214–225. [[CrossRef](#)]
21. Palmer, J.D. Comparative organization of chloroplast genomes. *Ann. Rev. Genet.* **1985**, *19*, 325–354. [[CrossRef](#)]
22. Zhang, R.; Ge, F.; Li, H.; Chen, Y.; Zhao, Y.; Gao, Y.; Liu, Z.; Yang, L. PCIR: A database of Plant Chloroplast Inverted Repeats. *Database J. Biol. Databases Curation* **2019**, *2019*, baz127. [[CrossRef](#)]
23. Nock, C.J.; Waters, D.L.; Edwards, M.A.; Bowen, S.G.; Rice, N.; Cordeiro, G.M.; Henry, R.J. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* **2011**, *9*, 328–333. [[CrossRef](#)]
24. Yang, Y.C.; Kung, T.L.; Hu, C.Y.; Lin, S.F. Development of primer pairs from diverse chloroplast genomes for use in plant phylogenetic research. *Genet. Mol. Res.* **2015**, *14*, 14857–14870. [[CrossRef](#)]
25. Adem, M.; Beyene, D.; Feyissa, T. Recent achievements obtained by chloroplast transformation. *Plant Methods* **2017**, *13*, 30. [[CrossRef](#)]
26. Wu, W.G.; Dong, L.L.; Chen, S.L. Development direction of molecular breeding of medicinal plants. *Chin. J. Chin. Mater. Med.* **2020**, *45*, 2714–2719.
27. Santos, C.; Pereira, F. Identification of plant species using variable length chloroplast DNA sequences. *Forensic Sci. Int. Genet.* **2018**, *36*, 1–12. [[CrossRef](#)]
28. Qian, J.; Song, J.Y.; Gao, H.H.; Zhu, Y.J.; Xu, J.; Pang, X.H. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *PLoS ONE* **2013**, *8*, e57607. [[CrossRef](#)]
29. Liang, C.L.; Wang, L.; Lei, J.; Duan, B.Z.; Ma, W.S.; Xiao, S.M. Comparative Analysis of the Chloroplast Genomes of Four *Salvia* Medicinal Plants. *Engineering* **2019**, *5*, 907–915. [[CrossRef](#)]
30. Gao, C.W.; Wu, C.H.; Zhang, Q.; Zhao, X.; Wu, M.X.; Chen, R.R. Characterization of Chloroplast Genomes From Two *Salvia* Medicinal Plants and Gene Transfer Among Their Mitochondrial and Chloroplast Genomes. *Front Genet.* **2020**, *11*, 574962. [[CrossRef](#)]
31. Moriguchi, Y.; Kang, K.S.; Lee, K.Y. Genetic variation of *Picea jezoensis* populations in South Korea revealed by chloroplast, mitochondrial, and nuclear DNA markers. *J. Plant Res.* **2009**, *122*, 153–160. [[CrossRef](#)]
32. Funk, H.T.; Berg, S.; Krupinska, K.; Maier, U.G.; Krause, K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* **2007**, *7*, 45. [[CrossRef](#)]
33. McNeal, J.R.; Kuehl, J.V.; Boore, J.L.; Leebens-Mack, J.; dePamphilis, C.W. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS ONE* **2009**, *4*, e5982. [[CrossRef](#)]
34. Barthet, M.M.; Pierpont, C.L.; Tavernier, E.-K. Unraveling the role of the enigmatic *MatK* maturase in chloroplast group IIA intron excision. *Plant Direct.* **2020**, *4*, 1–17. [[CrossRef](#)]
35. Zoschke, R.; Nakamura, M.; Liere, K.; Sugiura, M.; Börner, T.; Schmitz-Linneweber, C. An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. USA.* **2010**, *107*, 3245–3250. [[CrossRef](#)]
36. Leeder, W.M.; Voskuhl, S.; Göringer, H.U. The 2D Structure of the *T. brucei* Preedited *RPS12* mRNA Is Not Affected by Macromolecular Crowding. *J. Nucleic Acids* **2017**, *2017*, 6067345. [[CrossRef](#)]
37. Weglöhner, W.; Subramanian, A.R. Nucleotide sequence of a region of maize chloroplast DNA containing the 3' end of *clpP*, exon 1 of *rps12* and *rpl20* and their cotranscription. *Plant. Mol. Biol.* **1992**, *18*, 415–418. [[CrossRef](#)] [[PubMed](#)]
38. Dong, W.P.; Xu, C.; Li, C.H.; Sun, J.H.; Zuo, Y.J.; Shi, S. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [[CrossRef](#)] [[PubMed](#)]
39. Shi, C.; Liu, Y.; Huang, H.; Xia, E.H.; Zhang, H.B.; Gao, L.Z. Contradiction between Plastid Gene Transcription and Function Due to Complex Posttranscriptional Splicing: An Exemplary Study of *ycf15* Function and Evolution in Angiosperms. *PLoS ONE* **2013**, *8*, e59620. [[CrossRef](#)] [[PubMed](#)]
40. Alqahtani, A.A.; Jansen, R.K. The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperii* (Euphorbiaceae) plastome. *Sci. Rep.* **2021**, *11*, 7466. [[CrossRef](#)]
41. Cheatham, T.E.; Srinivasan, J.; Case, D.A.; Kollman, P.A. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265–280. [[CrossRef](#)]
42. Niu, Z.; Pan, J.; Zhu, S.; Li, L.; Xue, Q.; Liu, W.; Ding, X. Comparative Analysis of the Complete Plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) Reveals Different Evolutionary Dynamics of IR/SSC Boundary among Photosynthetic Orchids. *Front. Plant Sci.* **2017**, *8*, 1713. [[CrossRef](#)]

43. Ferreira, A.O.; Cardoso, H.G.; Macedo, E.S.; Breviario, D.; Arnholdt-Schmitt, B. Intron polymorphism pattern in AOX1 of wild St John's wort (*Hypericum perforatum*) allows discrimination between individual plants. *Physiol. Plant.* **2009**, *137*, 520–531. [[CrossRef](#)]
44. Yang, H.Y.; Wang, L.Q.; Chen, H.M.; Jiang, M.; Wu, W.W.; Liu, S.Y. Phylogenetic analysis and development of molecular markers for five medicinal *Alpinia* species based on complete plastome sequences. *BMC Plant Biol.* **2021**, *21*, 431. [[CrossRef](#)]
45. Fisher, V.L. Indigenous *Salvia* Species—An Investigation of the Antimicrobial Activity, Antioxidant Activity and Chemical Composition of Leaf Extracts. Ph.D. Thesis, University of the Witwatersrand, Johannesburg, South Africa, 2006.
46. Sun, Q.; Wang, K.; Yoshimura, A.; Doi, K. Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **2002**, *104*, 1335–1345. [[CrossRef](#)]
47. Cui, Y.X.; Nie, L.P.; Sun, W.; Xu, Z.C.; Wang, Y.; Yu, J. Comparative and Phylogenetic Analyses of Ginger (*Zingiber officinale*) in the Family Zingiberaceae Based on the Complete Chloroplast Genome. *Plants* **2019**, *8*, 283. [[CrossRef](#)]
48. Amenu, S.G.; Wei, N.; Wu, L.; Oyetola, O.; Hu, G.W.; Zhou, Y.D. Phylogenomic and comparative analyses of *Coffeae alliance* (Rubiaceae): Deep insights into phylogenetic relationships and plastome evolution. *BMC Plant Biol.* **2022**, *22*, 88. [[CrossRef](#)]
49. Bakker, R.T.; Culham, A.; Gmez-Martinez, R.; Carvalho, J.; Compton, J.; Dawtrey, R. Patterns of Nucleotide Substitution in Angiosperm cpDNA *trnL* (UAA)-*trnF*(GAA) Regions. *Mol. Biol. Evol.* **2000**, *17*, 1146–1155. [[CrossRef](#)]
50. Wang, D.Y.; Wu, Y.W.; Shih, A.C.C.; Wu, C.S.; Wang, Y.N.; Chaw, S.M. Transfer of Chloroplast Genomic DNA to Mitochondrial Genome Occurred At Least 300 MYA. *Mol. Biol. Evol.* **2007**, *24*, 2040–2048. [[CrossRef](#)]
51. Salmaki, Y.; Heubl, G.; Weigend, M. Towards a new classification of tribe Stachydeae (Lamiaceae): Naming clades using molecular evidence. *Bot. J. Linn. Soc.* **2019**, *190*, 345–359. [[CrossRef](#)]
52. Li, S.; Duan, W.; Zhao, J.; Jing, Y.; Feng, M.; Kuang, B.; Wei, N.; Chen, B.; Yang, X. Comparative Analysis of Chloroplast Genome in *Saccharum* spp. and Related Members of 'Saccharum Complex'. *Int. J. Mol. Sci.* **2022**, *23*, 7661. [[CrossRef](#)]
53. Li, P.; Lou, G.; Cai, X.; Zhang, B.; Cheng, Y.; Wang, H. Comparison of the complete plastomes and the phylogenetic analysis of *Paulownia* species. *Sci. Rep.* **2020**, *10*, 2225. [[CrossRef](#)]
54. Ayaz, A.; Zaman, W.; Saqib, S.; Ullah, F.; Mahmood, T. Phylogeny and Diversity of Lamiaceae based on *rps14* gene in Pakistan. *Genetika.* **2020**, *52*, 435–452. [[CrossRef](#)]
55. Dong, W.P.; Liu, J.; Yu, J.; Wang, L.; Zhou, S.L. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e35071. [[CrossRef](#)]
56. Duan, L.; Li, S.J.; Su, C.; Sirichamorn, Y.; Han, L.N.; Ye, W.; Lôc, P.K.; Wen, J.; Compton, J.A.; Schrire, B.; et al. Phylogenomic framework of the IRLC legumes (Leguminosae subfamily Papilionoideae) and intercontinental biogeography of tribe Wisterieae. *Mol. Phylogenet. Evol.* **2021**, *163*, 107235. [[CrossRef](#)]
57. Du, Q.; Jiang, M.; Sun, S.S.; Wang, L.Q.; Liu, S.Y.; Jiang, C.B. The complete chloroplast genome sequence of *Clerodendranthus spicatus*, a medicinal plant for preventing and treating kidney diseases from Lamiaceae family. *Mol. Biol. Rep.* **2022**, *49*, 3073–3083. [[CrossRef](#)]
58. Boudreau, E.; Takahashi, Y.; Lemieux, C.; Turmel, M.; Rochaix, J.D. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.* **1997**, *16*, 6095–6104. [[CrossRef](#)]
59. Naver, H.; Boudreau, E.; Rochaix, J.D. Functional studies of *Ycf3*: Its role in assembly of photosystem I and interactions with some of its subunits. *Plant Cell.* **2001**, *13*, 2731–2745. [[CrossRef](#)]
60. Krech, K.; Ruf, S.; Masduki, F.F.; Thiele, W.; Bednarczyk, D.; Albus, C.A.; Tiller, N.; Hasse, C.; Schöttler, M.A.; Bock, R. The plastid genome-encoded *Ycf4* protein functions as a nonessential assembly factor for photosystem I in higher plants. *Plant Physiol.* **2012**, *159*, 579–591. [[CrossRef](#)]
61. Li, J.F.; Li, L.; Sheen, J. Protocol: A rapid and economical procedure for purification of plasmid or plant DNA with diverse applications in plant biology. *Plant Methods* **2010**, *6*, 1–8. [[CrossRef](#)]
62. Lee, S.B.; McCord, B.; Buel, E. Advances in forensic DNA quantification: A review. *Electrophoresis* **2014**, *35*, 3044–3052. [[CrossRef](#)]
63. Diekmann, K.; Hodkinson, T.R.; Fricke, E.; Barth, S. An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. *PLoS ONE* **2008**, *3*, e2813. [[CrossRef](#)]
64. Cronn, R.; Liston, A.; Parks, M.; Gernandt, D.S.; Shen, R.; Mockler, T. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* **2008**, *36*, e122. [[CrossRef](#)]
65. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
66. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2017**, *45*, e18.
67. Shi, L.C.; Chen, H.M.; Jiang, M.; Wang, L.Q.; Wu, X.; Huang, L.F.; Liu, C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* **2019**, *47*, W65–W73. [[CrossRef](#)]
68. Firtina, C.; Kim, J.S.; Alser, M.; Senol Cali, D.; Cicek, A.E.; Alkan, C.; Mutlu, O. Apollo: A sequencing-technology-independent, scalable and accurate assembly polishing algorithm. *Bioinformatics* **2020**, *36*, 3669–3679. [[CrossRef](#)] [[PubMed](#)]
69. Stothard, P.; Grant, J.R.; Van Domselaar, G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief Bioinform.* **2019**, *20*, 1576–1582. [[CrossRef](#)] [[PubMed](#)]
70. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]

71. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
72. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
73. van Melle Guy, D. *VMATCH: Stata Module to Match Variables between Subjects. Statistical Software Components S350801*; Boston College Department of Economics: Boston, MA, USA, 1998.
74. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [[CrossRef](#)] [[PubMed](#)]
75. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, W273–W279. [[CrossRef](#)]
76. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *1*, i54–i62. [[CrossRef](#)]
77. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [[CrossRef](#)]
78. Mahadani, A.K.; Awasthi, S.; Sanyal, G.; Bhattacharjee, P.; Pippal, S. Indel-K2P: A modified Kimura 2 Parameters (K2P) model to incorporate insertion and deletion (Indel) information in phylogenetic analysis. *Cyber-Phys. Syst.* **2021**, *7*, 1–13. [[CrossRef](#)]
79. Riaz, T.; Shehzad, W.; Viari, A.; Pompanon, F.; Taberlet, P.; Coissac, E. ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* **2011**, *39*, e145. [[CrossRef](#)]
80. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
81. Wang, K.; Li, H.; Xu, Y.; Shao, Q.; Yi, J.; Wang, R.; Cai, W.; Hang, X.; Zhang, C.; Cai, H.; et al. MFEprimer-3.0: Quality control for PCR primers. *Nucleic Acids Res.* **2019**, *47*, W610–W613. [[CrossRef](#)]
82. Lee, D.J.; Kim, J.D.; Kim, Y.S.; Song, H.J.; Park, C.Y. Evaluation-independent system for DNA section amplification. *Biomed. Eng. Online* **2018**, *17*, 150. [[CrossRef](#)]
83. Nicholas, K.B.; Nicholas, H.B., Jr.; Deerfield, I.I. GeneDoc: A tool for editing and annotating multiple sequence alignments. *Embnew. News.* **1997**, *4*, 1–4.
84. 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](#)]
85. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
86. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
87. Hall, B.G. Building phylogenetic trees from molecular data with MEGA. *Mol. Biol. Evol.* **2013**, *30*, 1229–1235. [[CrossRef](#)]