



Research
Traditional Chinese Medicine—Article

A Comparative Analysis of the Chloroplast Genomes of Four *Salvia* Medicinal Plants



Conglian Liang^{a,b,c}, Lei Wang^{a,b,d}, Juan Lei^e, Baozhong Duan^f, Weisi Ma^g, Shuiming Xiao^b, Haijun Qi^c, Zhen Wang^c, Yaoqi Liu^g, Xiaofeng Shen^b, Shuai Guo^b, Haoyu Hu^b, Jiang Xu^{b,*}, Shilin Chen^{b,*}

^a College of Pharmacy, Shandong University of Traditional Chinese Medicine (TCM), Jinan 250355, China

^b Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

^c State Key Laboratory of Innovative Natural Medicine and TCM Injections, Jiangxi Qingfeng Pharmaceutical Co. Ltd., Ganzhou 341000, China

^d Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China

^e School of Biological Science, The University of Queensland, Brisbane, QLD 4072, Australia

^f College of Pharmacy, Dali University, Dali 671000, China

^g Institute of Medicinal Plants, Yunnan Academy of Agricultural Sciences, Kunming 650205, China

ARTICLE INFO

Article history:

Received 23 July 2018

Revised 27 November 2018

Accepted 7 January 2019

Available online 16 August 2019

Keywords:

Salvia

Chloroplast genome

Comparative analysis

ABSTRACT

Herbgenomics is an emerging field of traditional Chinese medicine (TCM) research and development. By combining TCM research with genomics, herbgenomics can help to establish the scientific validity of TCM and bring it into wider usage within the field of medicine. *Salvia* Linn. (*S. Linn.*) is a large genus of Labiatae that includes important medicinal plants. In this herbgenomics study, the complete chloroplast (cp) genomes of two *Salvia* spp.—namely, *S. przewalskii* and *S. bulleyana*, which are used as a surrogate for *S. miltiorrhiza*—were sequenced and compared with those of two other reported *Salvia* spp.—namely, *S. miltiorrhiza* and *S. japonica*. The genome organization, gene number, type, and repeat sequences were compared. The annotation results showed that both *Salvia* plants contain 114 unique genes, including 80 protein-coding, 30 transfer RNA (tRNA), and four ribosomal RNA (rRNA) genes. Repeat sequence analysis revealed 21 forward and 22 palindromic sequences in both *Salvia* cp genomes, and 17 and 21 tandem repeats in *S. przewalskii* and *S. bulleyana*, respectively. A synteny comparison of the *Salvia* spp. cp genomes showed a high degree of sequence similarity in the coding regions and a relatively high divergence of the intergenic spacers. Pairwise alignment and single-nucleotide polymorphism (SNP) analyses found some candidate fragments to identify *Salvia* spp., such as the intergenic region of the *trnV-ndhC*, *trnQ-rps16*, *atpI-atpH*, *psbA-ycf3*, *ycf1*, *rpoC2*, *ndhF*, *matK*, *rpoB*, *rpoA*, and *accD* genes. All of the results—including the repeat sequences and SNP sites, the inverted repeat (IR) region border, and the phylogenetic analysis—showed that *S. przewalskii* and *S. bulleyana* are extremely similar from a genetic standpoint. The cp genome sequences of the two *Salvia* spp. reported here will pave the way for breeding, species identification, phylogenetic evolution, and cp genetic engineering studies of *Salvia* medicinal plants.

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Salvia Linn. (*S. Linn.*) is a large genus of Labiatae that comprises approximately 1000 species worldwide; of these species, 84 are found in China—mainly in Southwest China [1]. Among these 84 species, nearly 30 have been recorded as medicinal plants [2].

Danshen is a representative traditional Chinese medicine (TCM) [3] that has a significant effect on the treatment of cardiovascular and cerebrovascular diseases [4,5]. The *Chinese Pharmacopoeia* indicates that *S. miltiorrhiza* Bunge is the only source of Danshen [6]. However, many other *Salvia* plants have similar chemical components and corresponding pharmacological effects to those of Danshen, and are used as Danshen in nongovernmental applications and local medication. For example, *S. przewalskii* is an herbaceous perennial plant that is widely distributed in Southwest China [7] and has been used as a substitute for *S. miltiorrhiza* for at least 300 years. This plant is also

* Corresponding authors.

E-mail addresses: jxu@icmm.ac.cn (J. Xu), sichen@icmm.ac.cn (S. Chen).

used as a substitute for *Gentiana macrophylla* in Sichuan Province. Many studies have shown that *S. przewalskii* has similar components to those of *S. miltiorrhiza*; however, the amount of fat-soluble components in the root of *S. przewalskii* is several times higher than in *S. miltiorrhiza*, giving *S. przewalskii* great application value [8–10]. *S. bulleyana*, also known as “purple Danshen” is mainly distributed in Dali in Yunnan Province and is often used as Danshen folk medicine [11]. However, the chloroplast (cp) genomes of most *Salvia* plants remain unknown.

Chloroplasts, which play a key role in autotrophic photosynthesis, are important organelles in green plants [12–14]. The cp genome is typically a circular multicopy DNA molecule in cells [15,16]. The structure of plant cp genome is conservative and can be divided into four segments—that is, two copies of inverted repeat (IR) regions (IRa and IRb), which are separated by a large single-copy (LSC) region and a small single-copy (SSC) region [17,18]. The sizes of the cp genomes of different species differ mainly due to IR contraction or expansion [19,20]. With the advent of new sequencing technologies, sequencing is becoming faster and cheaper; more than 2000 cp genomes have now been reported to the National Coalition Building Institute (NCBI) [21]. At present, the use of a DNA barcode to differentiate between closely related species and infraspecific taxa is hampered due to short gene fragments and a low number of phylogenetic informative sites. Therefore, some researchers have proposed using the whole cp genome as a “super barcode” for species identification [22]. Given its high expression efficiency, site-specific integration, and maternal inheritance, cp transformation techniques have shown considerable potential for genetic improvement [23–26].

Genomic data, including organelle genomic data, can provide a molecular basis for the study of the original species of TCM. In our previous work, we have studied many genome sequences [27–29], including the complete cp genome sequence of *S. miltiorrhiza* in 2013 [30] and the draft sequence and analysis of the *S. miltiorrhiza* genome in 2016 [31]. In the present study, we obtained the complete cp genome sequence of *S. przewalskii* and *S. bulleyana* using a next-generation sequencing platform. We then compared the two cp genomes with those of the two other *Salvia* spp. that have been reported, in terms of genome organization, repeat sequence, and IR length. Finally, we performed a phylogenetic analysis based on the whole cp genome of 16 angiosperms. These efforts provide additional information for constructing the cp genome library of *Salvia*, which will aid in identifying *Salvia* spp. and provide insights into its evolutionary origins.

2. Results

2.1. Cp genome organizations

Raw data (approximately 5.9×10^9 and 1.372×10^{10} , respectively) and trimmed data (approximately 5.2×10^9 and 1.343×10^{10} , respectively) were obtained from *S. przewalskii* and *S. bulleyana*. The resulting cp genomes of the two *Salvia* spp., *S. przewalskii* (MH603953) and *S. bulleyana* (MH603954), have been presented to the NCBI. The total cp genome sizes of the two *Salvia* spp. are 151 319 bp and 151 547 bp, respectively. Regarding the cp genomes of the two *Salvia* spp., both *S. przewalskii* and *S. bulleyana* have a typical quadripartite structure, like the majority of flowering plants, composed of a pair of IRs (50 982 and 51 098 bp) and two single-copy regions (LSC: 82 732 and 82 853 bp; SSC: 17 605 and 17 596 bp), as shown in Fig. 1 and Table 1.

The cp genome of *S. przewalskii* is the shortest of the four sequenced *Salvia* spp., while *S. bulleyana* ranks between *S. miltiorrhiza* and *S. japonica* in length. The cp genome lengths of the four species range from 151 319 to 153 995 bp, while the LSC region

lengths range from 82 695 to 84 573 bp. The IR region ranges from 50 982 to 51 832 bp, and the SSC region ranges from 17 555 to 17 605 bp. The GC content of the *S. przewalskii* and *S. bulleyana* cp genomes is 37.96% and 37.99%, respectively, while the IR regions possess a higher GC content (43.11% and 43.12%) than the LSC (36.08% and 36.12%) and SSC regions (31.88% and 31.89%). In general, the IR region has four ribosomal RNA (rRNA) genes enriched in GC [32]. The *S. przewalskii* and *S. bulleyana* cp genomes have more AT than GC; the same is true for the cp genomes of the two other *Salvia* spp., and for those of other land plants [33–36].

In this study, we annotated 134 genes in two *Salvia* cp genomes, of which 114 are unique, consisting of 80 protein-coding, 30 transfer RNA (tRNA), and four rRNA genes (Fig. 1, Table S1 in Supplementary data (SD)). Out of these, 18 genes, consisting of seven protein-coding, seven tRNA, and four rRNA genes, are repetitive in the IR regions. A total of 61 protein-coding and 22 tRNA genes are present in the LSC region, whereas 12 protein-coding and one tRNA genes are present in the SSC region. One *ycf1* pseudogene and one *rps19* pseudogene are located in two IR boundary regions (Fig. S1 in SD).

The quantity and type of genes in the two *Salvia* cp genomes are the same as those in the two other *Salvia* spp. A slight variation was observed in the nucleotide composition in the coding sequence (CDS) of the four *Salvia* cp genomes (Table 2). The GC contents are 45.75%/44.91%, 38.19%/38.49%, and 30.32%/30.89% at the first, second, and third codon positions in two *Salvia* CDS regions identified in this study, respectively (Table 2). The contents of the other two *Salvia* spp. are 45.77%/45.73%, 38.21%/38.13%, and 30.38%/30.17% at the first, second, and third codon positions, respectively. The trait of a preference for AT the third codon position that was observed in the cp genomes of these four *Salvia* species also appears in other angiosperms [37–40].

A total of 87 protein-coding genes have 26 439 and 26 432 codons in the *S. przewalskii* and *S. bulleyana* cp genomes, respectively, compared with 26 483 and 26 485 in *S. miltiorrhiza* and *S. japonica*, respectively (Table S2 in SD). Out of these codons, AUU (1096/1106) encoding isoleucine (Ile) and UGC (76/114) encoding cysteine are the most and least used in *S. przewalskii* and *S. bulleyana*, respectively. AUU and UGC are also the most and least used in the two other *Salvia* cp genomes (1100/1102 and 70/71). The relative synonymous codon usages (RSCUs) of the four cp genomes differ slightly from each other. Multiple codons are present for all the amino acids, in addition to methionine and tryptophan (Fig. 2). There are six synonymous codons for arginine (Arg), leucine (Leu), and serine (Ser); four synonymous codons for each of valine (Val), proline (Pro), threonine (Thr), alanine (Ala), and glycine (Gly); three synonymous codons for each of Ile and stop codons; and two synonymous codons for each of the remaining amino acids. The synonymous codon is usually mutated only at the third position, thereby reducing harmful mutations.

Introns play a crucial part in the regulation of gene expression and can enhance exogenous gene expression at plant sites to produce ideal agronomic traits [41,42]. A total of 18 genes containing introns were observed in the four *Salvia* cp genomes—15 containing one intron and three containing two introns. The intron of the *trnK-UUU* gene is the longest, and contains the *matK* gene. The 5' end of the *rps12* gene is located in the LSC region, while the 3' end is located in the IR regions, making it a trans-spliced gene. In general, the exon lengths are conserved in the four *Salvia* spp., except for the *ndhB* gene (Table S3 in SD).

The 18 genes of the four *Salvia* spp. were compared with those of four other Labiate species—namely, *Mentha longifolia*, *Ocimum basilicum*, *Perilla frutescens*, and *Scutellaria baicalensis*. Minor variations were observed in most exon lengths of the eight Labiate cp genomes; however, some of these genes are only conserved in *Salvia* spp. and *Mentha longifolia*. These genes include *trnV-UAC*, *rpoC1*,

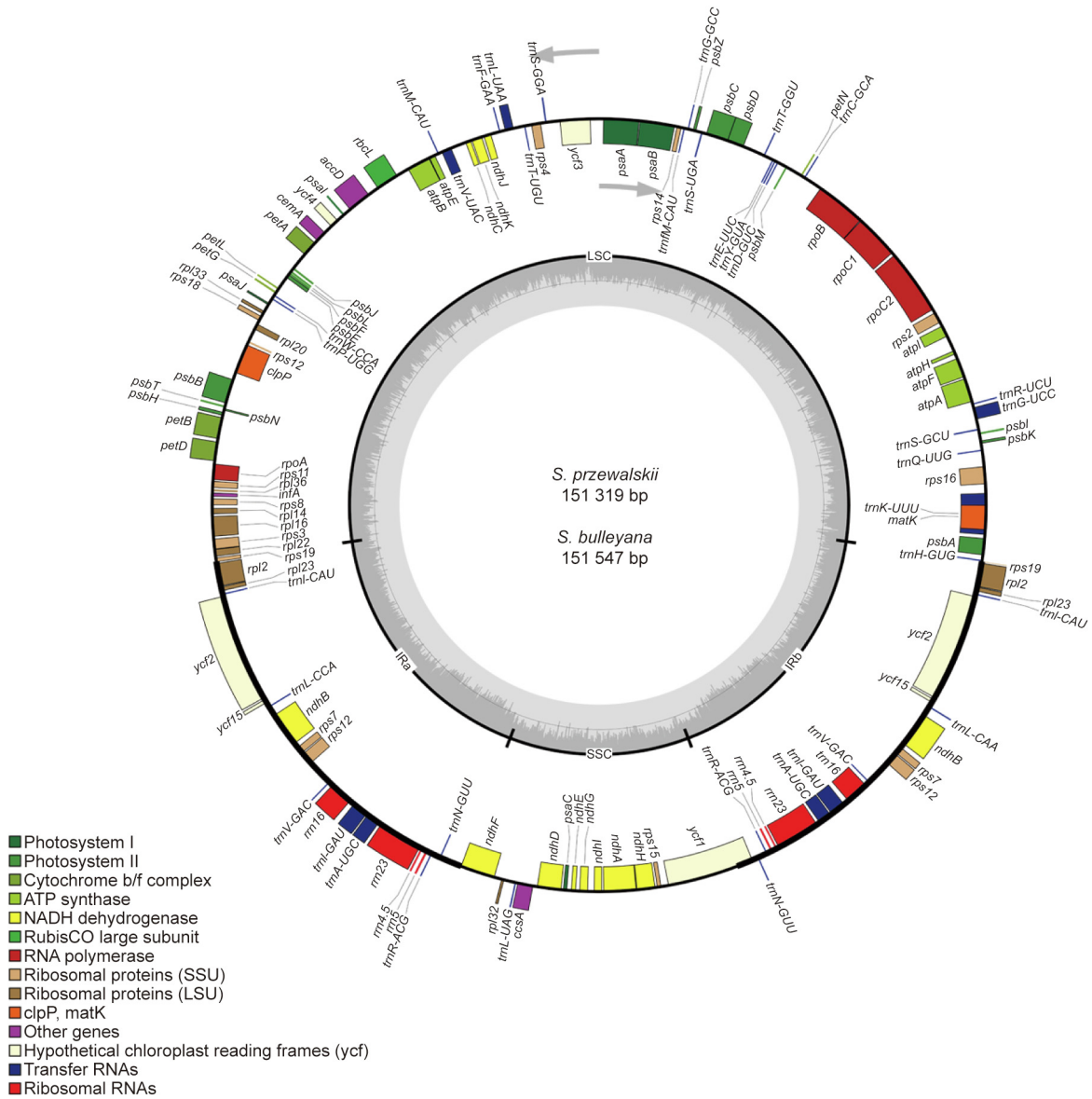


Fig. 1. Gene map of the *S. przewalskii* and *S. bulleyana* cp genomes. The genes inside and outside the outer circle are transcribed in the direction of the grey arrows inside and outside at the top. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. Within the inner circle, dark grey represents GC content and light grey represents AT content.

Table 1
Summary of the base composition of the cp genomes of four *Salvia* species.

Species	Total cp genome size (bp)	LSC region (bp)	IR region (bp)	SSC region (bp)	GC content (%)	GC content in LSC region (%)	GC content in IR region (%)	GC content in SSC region (%)
<i>S. przewalskii</i>	151 319	82 732	50 982	17 605	37.96	36.08	43.11	31.88
<i>S. bulleyana</i>	151 547	82 853	51 098	17 596	37.99	36.12	43.12	31.89
<i>S. miltiorrhiza</i>	151 328	82 695	51 078	17 555	36.20	36.20	43.10	32.00
<i>S. japonica</i>	153 995	84 573	51 832	17 590	36.24	36.24	43.03	31.67

and *ycf3*, where the intron phase are different in the eight Labiate cp genomes. The exon length of the *rps16* gene is specific for the four *Salvia* spp. Greater variation in the intron length was observed than in the exon length, although the intron lengths of some genes, such as *trnL-UAA* and *rpl2*, are the same in three of the four *Salvia* spp., with the exception of *S. japonica*.

RNA editing participates in plastid transcription regulation, which can enrich transcription and protein diversity [43–45]. In this study, 35 genes of the four *Salvia* cp genomes were predicted for their potential RNA editing sites. A total of 43 RNA editing sites

were predicted; of these, 37 are common sites of the four species (Table S4 in SD). Of the 35 genes, 16—including *atpA*, *atpB*, *clpP*, *petD*, *petG*, *petL*, *psaB*, *psal*, *psbB*, *psbE*, *psbF*, *psbL*, *rpl23*, *rpoC1*, *rps8*, and *ycf3*—were not measured for their potential RNA editing sites. The *rps16* gene was not measured for its potential RNA editing sites in *S. japonica*, but one potential RNA editing site was observed in the three other species. Of the 43 potential RNA editing sites, 11 were observed at the first position of the corresponding codon and 32 were observed at the second position. No potential RNA editing site was observed at the third position, and the base

Table 2
Gene number and CDS nucleotide composition of the cp genomes of the four *Salvia* species.

Species	Number of genes		CDS (bp)						ATCG contents in CDS (%)						ATCG contents in 1st position (%)						ATCG contents in 2nd position (%)						ATCG contents in 3rd position (%)					
	Unique	Protein-coding	tRNA	rRNA	T		C		A		G		T		C		A		G		T		C		A		G					
					31.33	17.80	30.58	20.29	23.71	18.98	30.54	26.77	32.56	20.38	29.24	17.81	37.73	14.03	31.96	16.29												
<i>S. przewalskii</i>	114	80	30	4	79 317	31.33	17.80	30.58	20.29	23.71	18.98	30.54	26.77	32.56	20.38	29.24	17.81	37.73	14.03	31.96	16.29											
<i>S. bulleyana</i>	114	80	30	4	79 296	31.34	17.80	30.56	20.30	24.25	18.84	30.83	26.07	32.26	20.22	29.25	18.27	37.51	14.34	31.59	16.55											
<i>S. miltiorrhiza</i>	114	80	30	4	79 449	31.35	17.82	30.54	20.30	23.69	19.01	30.54	26.76	32.56	20.40	29.23	17.81	37.79	14.05	31.84	16.33											
<i>S. japonica</i>	114	80	30	4	79 455	31.39	17.77	30.60	20.24	23.68	18.98	30.59	26.75	32.68	20.35	29.19	17.78	37.81	13.99	32.02	16.18											

conversion type is all C to T. This result is similar to those of other land plants [46,47]. The conversion of amino acids from Ser to Leu occurs most frequently, while the conversions from Pro to Ser and Thr to Ile occur least frequently.

2.2. Repeat and simple sequence repeat (SSR) analyses

In the repeat sequences analysis, 43 repeats—comprising 21 forward and 22 palindromic repeats—were found in both *Salvia* spp. by using REPuter (Table S5 in SD, Fig. 3). However, most repeats were found in the *ycf2* coding region; four were found between the *ycf3* intron region and the intergenic spacer (IGS) of *rps12* and *trnV-GAC*, and four were found in the IGS of *rrm4.5* and *rrn5*. In *S. przewalskii* and *S. bulleyana*, 17 and 21 tandem repeats were detected, respectively (Table S6 in SD, Fig. 3). Approximately half of these repeats are located in the *ycf1* and *ycf2* genes, while the other half of the repeats are located in the IGS regions. The two longest repeats, which are approximately 90 bp in length, are present in *ycf2* in the two *Salvia* spp.

A comparative analysis of the repeats in the eight Labiate cp genomes showed that *S. przewalskii* and *S. bulleyana* are resemblant with the other cp genomes in repeat type, while *S. japonica* possesses more long-segment repeats than the others. The lengths of the tandem repeats in the eight Labiate cp genomes range from 10 to 30 bp, while the lengths of the forward and palindromic repeats mostly range from 30 to 45 bp.

SSRs are widely distributed at different locations in the genome [48]. The cp genome has the characteristics of uniparental inheritance, and SSRs have a high variation level within the same species. Thus, cp SSRs have been widely used as molecular markers in the study of genetic map construction, target gene calibration, and mapping [49–51]. Here, a total of 178 SSRs, comprising 134 mononucleotide (mono), 35 dinucleotide (di), seven tetranucleotide (tetra), and two pentanucleotide (penta) SSRs were observed in the *S. przewalskii* cp genome; a total of 177 SSRs, comprising 136 mono, 32 di, and nine tetra SSRs were observed in the *S. bulleyana* cp genome (Fig. 4, Table S7 in SD). The major type of SSR is the mono SSR; the A/T type SSR (132/129) accounts for the vast majority of the mono SSRs.

A comparison of the SSRs in the eight cp genomes showed that all eight cp genomes are similar. Most of the SSRs in the eight cp genomes consist of mono and di repeat motifs. The mono repeats vary from 109 (*Ocimum basilicum*) to 136 (*S. bulleyana*), and the di repeats vary from 25 (*S. miltiorrhiza*) to 39 (*Perilla frutescens*). However, three or more oligonucleotide repeats are relatively low in number but rich in types (Table S7).

2.3. Synteny comparison and SNP analyses

The pairwise cp genomic alignment of *S. przewalskii* and *S. bulleyana* with the two other *Salvia* cp genomes was conducted with annotated *S. przewalskii* cp genomes as a reference, using mVISTA (Fig. S2 in SD). The variable sites of the cp genomes and the 80 unique genes of the four *Salvia* cp genomes were analyzed (Fig. 5 and Table S8 in SD). The comparison and single-nucleotide polymorphism (SNP) analyses of the alignment of the four cp genomes revealed that the IR regions are more conserved than the LSC and SSC regions, in which divergences are dispersed. The noncoding regions have more divergent regions than the coding regions. In these cp genomes, highly divergent regions appear in the IGSs, including *rps16-trnQ*, *trnG-trnS*, *atpH-atpI*, *psbA-ycf3*, *ycf4-cenA*, and *thrnV-ndhC* in the LSC region and *ycf1-rps15*, *rpl32-trnL*, and *ndhI-ndhG* in the SSC region. Some divergences were also observed in the coding regions of the *rpoC2*, *ndhF*, and *ycf1* genes in the four cp genomes. *S. przewalskii* and *S. bulleyana* are more similar to *S. miltiorrhiza* than to *S. japonica*.

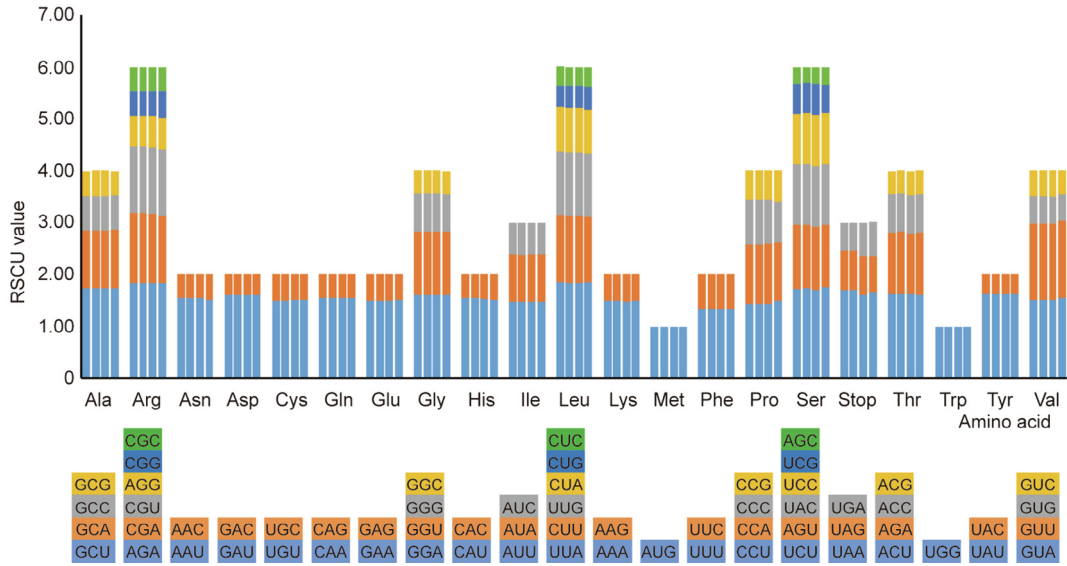


Fig. 2. Codon content for the CDS in the four *Salvia* cp genomes. The abscissa represents 20 amino acids and terminators, while the ordinate represents the RSCU value. For each amino acid, the corresponding species from left to right are *S. przewalskii*, *S. bulleyana*, *S. miltiorrhiza*, and *S. japonica*. The different colors of each amino acid correspond to the codon of the same color below. Asn: asparagine; Asp: asparticacid; Cys: cysteine; Gln: Glutamine; Glu: glutamicacid; His: histidine; Lys: lysine; Met: methionine; Phe: phenylalanine; Trp: tryptophan; Tyr: tyrosine.

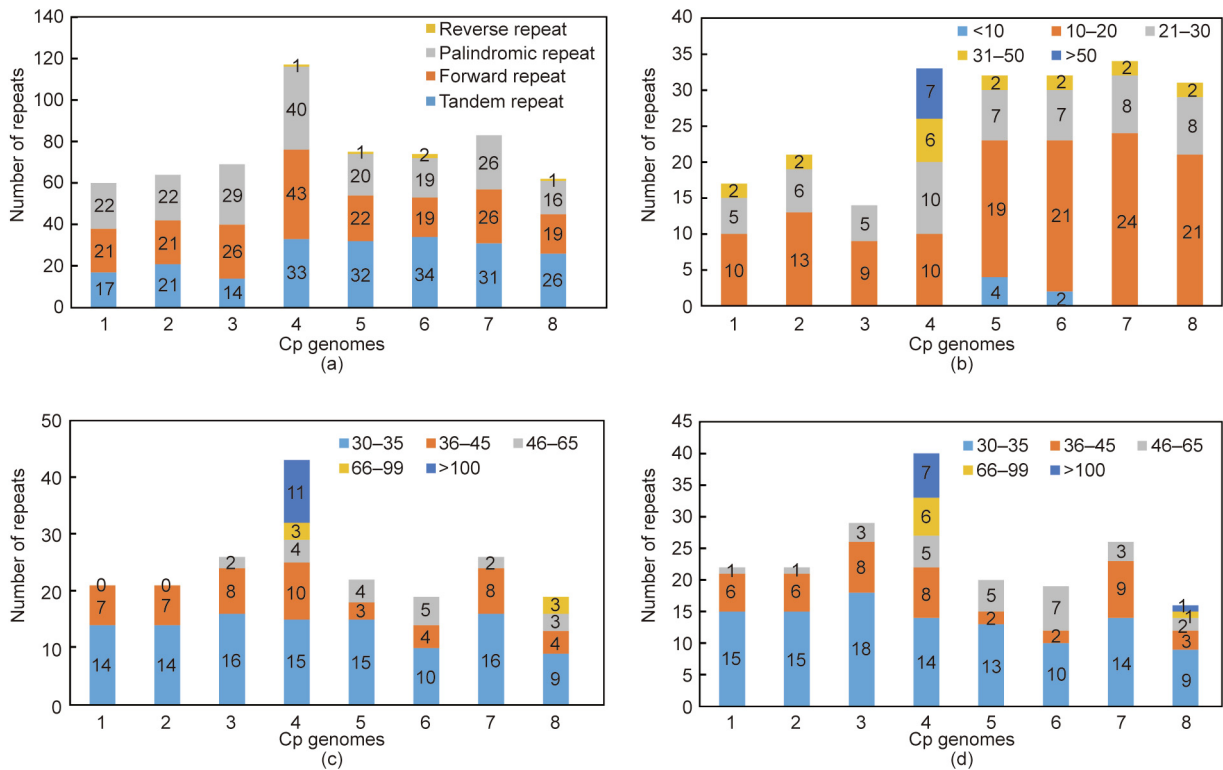


Fig. 3. Repeat sequences analysis of eight cp genomes. (a) Repeat types in eight cp genomes; (b) tandem repeats in eight cp genomes; (c) forward repeats in eight cp genomes; (d) palindromic repeats in eight cp genomes. In (a), different colors show different repeat types; in (b-d), different colors show different lengths. The ordinate represents the number of repeats; in the abscissa, numbers 1–8 represent the following: 1 for *S. przewalskii*, 2 for *S. bulleyana*, 3 for *S. miltiorrhiza*, 4 for *S. japonica*, 5 for *Mentha longifolia*, 6 for *Ocimum basilicum*, 7 for *Perilla frutescens*, and 8 for *Scutellaria baicalensis*.

2.4. Phylogenetic analysis

A phylogenetic tree was built based on eight Labiate, four Solanaceae, and three other Asterales cp genomes using the maximum likelihood (ML) method, with *Arabidopsis thaliana* as the outgroup (Fig. 6). All of these have a bootstrap value of 100%, and in the ML tree, all four *Salvia* spp. form a robust monophyletic branch. Two

Salvia spp.—namely, *S. przewalskii* and *S. bulleyana*—are clustered together in one terminal branch, thereby representing subgen. *Salvia*. All eight Labiate spp. are clustered together in one monophyletic group and nested in the branch of *Sesamum indicum* (Pedaliaceae), *Boea hygrometrica* (Gesneriaceae), and *Olea europaea* (Burseraceae). This result is consistent with the findings of a previous report [30].

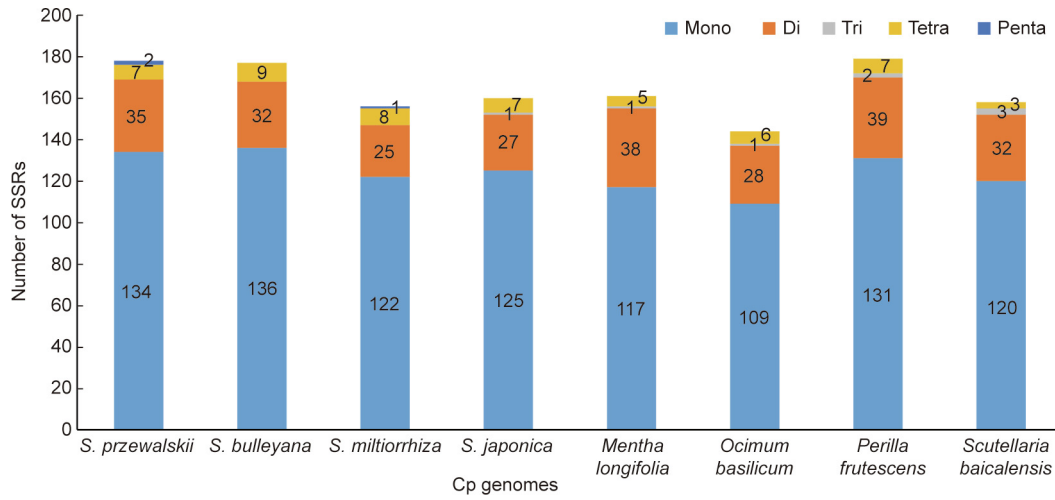


Fig. 4. SSR analysis of eight cp genomes. The ordinate represents the number of SSRs.

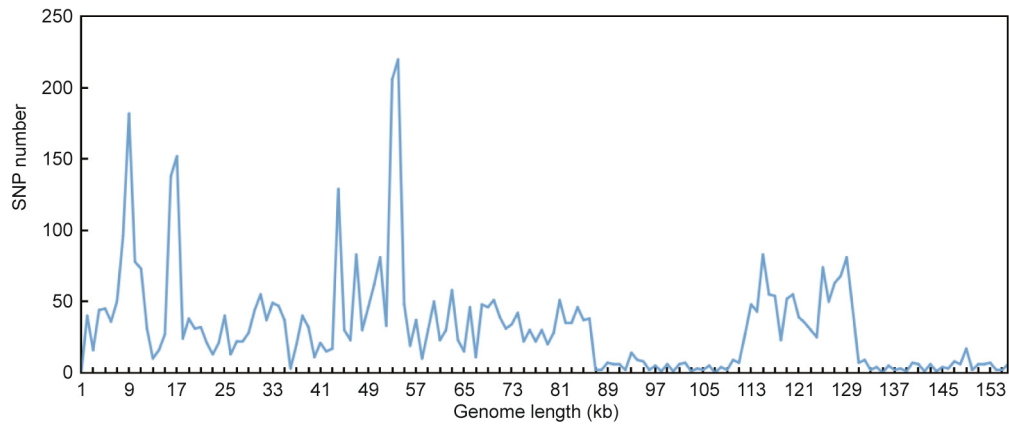


Fig. 5. Statistics of the SNPs in the four cp genomes.

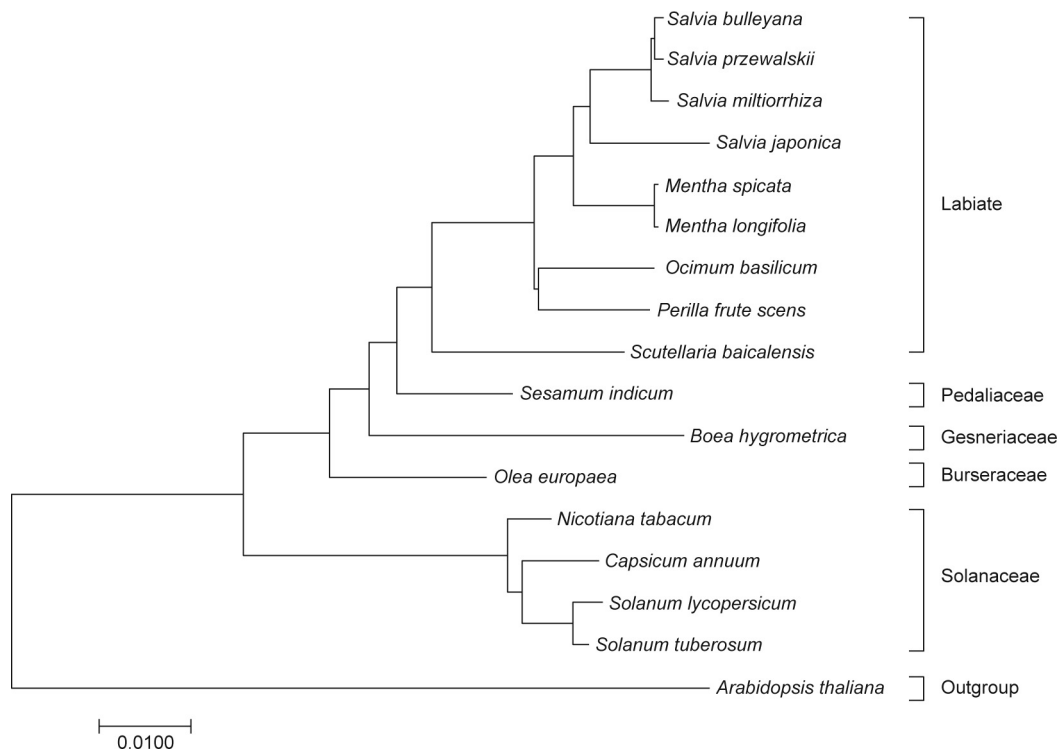


Fig. 6. ML phylogenetic tree reconstruction containing the cp genomes of 16 plants. Arabidopsis thaliana was set as the outgroup.

3. Methods

3.1. Plant material and DNA extraction

S. przewalskii was collected from Yulong County in the Lijiang City in Yunnan Province, and *S. bulleyana* was collected from Wulongba, Binchuan County in the Dali Bai Autonomous Prefecture in Yunnan Province. The total genomic DNA of *S. przewalskii* was extracted from 100 mg of fresh leaf using the DNeasy Plant Mini Kit (QIAGEN GmbH, Germany), while the total genomic DNA of *S. bulleyana* was extracted from 100 mg of fresh leaf using the modified cetyl trimethyl ammonium bromide (CTAB) method [52]. The quality and concentration of the genomic DNA were estimated using agarose gel electrophoresis and a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., USA). Qualified DNA was used for library construction in terms of the user guide. The Illumina HiSeq 1500 platform (Illumina Inc., USA) was used for sequencing.

3.2. Assembly and annotation of the two *Salvia* spp.

The whole-genome sequences were used to extract the cp genomes. First, raw reads were evaluated using FastQC and trimmed using Trimmomatic to remove low-quality bases ($Q < 30$, $Q = -10 \log_{10}(\text{error } P)$, $Q < 30$ means that limiting the error rate < 0.001 , length < 50) and adapter sequences [53,54]. Next, cp-like reads were extracted from trimmed reads using Basic Local Alignment Search Tool (BLAST) [55], with the cp genomes of *S. miltiorrhiza* (JX312195) being used as reference sequences. Finally, cp-like reads were used for genome assembly using SOAPdenovo [56]. Through a comparison of the contigs members and the lengths of different k -mer sizes (k -mer means that reads are divided into strings containing k bases and then assembled), k -mer sizes of 127 and 77, respectively, were found to provide the best results for *S. przewalskii* and *S. bulleyana*. The results from these two parameters were used to generate the final assembly. SSPACE and GapCloser were used to obtain scaffolds and fill gaps [56,57]. Cp genome annotation was performed using CPGAVAS with default parameters [58]. The result from CPGAVAS was corrected manually for start codons, stop codons, and intron/exon boundaries using Apollo [59]. The tRNA genes were identified using tRNAscanSE [60]. The circular maps of the two *Salvia* spp. were obtained using OGDRAW [61]. Codon usage and cp genomic characteristics were analyzed using MEGA6 [62]. The RNA editing sites of the gene-coding proteins in the two cp genomes were predicted using a predictive RNA editor, PREP suite [63], with a cut-off value set at 0.8.

3.3. Characterization of repeat sequences and SSRs

Repeat sequences containing the three types of forward, reverse, and palindromic were identified using REPuter with a Hamming distance set at 3 and a minimum repeat size set at 30 bp [64]. Tandem repeats were analyzed using the Tandem Repeats Finder with default parameters [65]. SSRs were detected using MISA [66] with the following thresholds: eight repeat units for mono SSRs, four repeat units for di- and trinucleotide repeat SSRs, and three repeat units for tetra-, penta-, and hexanucleotide repeat SSRs [15].

3.4. Genome comparison

The mVISTA [67] in the Shuffle-LAGAN mode was used to compare the cp genomes of *S. przewalskii* and *S. bulleyana* with the cp genomes of the other two reported *Salvia* spp. cp genomes (i.e., *S.*

miltiorrhiza [JX312195] and *S. japonica* [NC_035233]) by using the annotation of *S. przewalskii* as the reference.

3.5. Phylogenetic analysis

A total of 14 complete cp genome sequences were downloaded from the NCBI Organelle Genome and Nucleotide Resource Database for the phylogenetic analysis. The cp genomes were aligned using the MAFFT software, while MEGA6 [68] was used to construct the phylogenetic tree with the ML method. A bootstrap analysis was executed with 1000 replicates and tree bisection and reconnection (TBR) branch swapping, while *Arabidopsis thaliana* was set as the outgroup.

4. Discussion

4.1. Cp genome organizations

In this study, we are the first to report the complete cp genomes of *S. przewalskii* and *S. bulleyana*. The cp genome lengths of the two species range from 151 319 to 151 547 bp, with a typical quadripartite structure. These two cp genomes encode 114 unigenes, including 80 protein-coding genes, 30 tRNA genes, and four rRNA genes. Out of these genes, the CDS of the *ycf2* gene is the largest, while that of the *petN* gene is the smallest. The 5' end of the *rps12* gene is located in the LSC region, while its 3' end is located in the IR region, which is common in angiosperm cp genomes. In conclusion, the cp genome organization, GC content, gene number and type, and codon usage of *S. przewalskii* and *S. bulleyana* are similar to those of the two other *Salvia* cp genomes.

4.2. Gene comparison

Some genes, such as *matK-trnK*, *atpB-atpE*, *psbC-psbD*, and *rps3-rpl22*, overlap each other in the eight Labiate cp genomes. A comparison of the overlap length shows that the overlap lengths of *atpB-atpE*, *psbC-psbD*, and *rps3-rpl22* are the same, while the *rps3* and *rpl22* genes in *Scutellaria baicalensis* have no overlap. The length of the *matK* gene (1563 bp) is the same, but 12 variable sites were observed in *S. przewalskii*, *S. bulleyana*, and *S. miltiorrhiza*. The length of the *matK* gene (1530 bp) is the same in *S. japonica*, *Mentha longifolia*, and *Scutellaria baicalensis*. The *matK* gene was chosen by the Consortium for the Barcode of Life as the land plant barcode in 2009 [69]. A previous study has found that the *matK* gene can be used to identify different species in Labiatae better than the internal transcribed spacer (ITS) gene [70].

Here, we found two different *ndhB* gene types in the eight Labiate species; one encodes 511 amino acids, while the other encodes 493 amino acids. The length of exon 1 of the *ndhB* gene is 54 bp shorter in the two *Salvia* cp genomes than in *S. miltiorrhiza* (JX312195) when the length is annotated according to the cp genome of *S. miltiorrhiza*, where A substitutes for G in the eighth base position. This phenomenon causes a termination to appear, which is same as that in the *ndhB* gene in *Perilla frutescens* (NC_030756). This site was confirmed by polymerase chain reaction (PCR) with the forward primer AACAAACGAAAAGGAAACG and the reverse primer CTCATAGGAACAATAGGG. The two exons of the fern *ndhB* gene have a unique pattern of intragenic copy number variants [71]. Although the *rps16* gene is functionally lost in various legume lineages, reports show that the *rps16* gene is essential even under heterotrophic conditions and can be functionally replaced by a nuclear gene [72,73]. In the present study, we found that this gene length is conserved in the four *Salvia* cp genomes, and only three variable sites were observed in this gene CDS (Table S8).

4.3. Variance within genera

The technique of DNA barcoding, which was first proposed by Hebert et al. [74], can be used to identify species by means of DNA sequences such as *ITS2*, *matK*, *psbA-trnH*, and *rbcl*. However, the identification of proximal species—and particularly of morphologically confusing species in the same genus—still presents some difficulties. Therefore, finding a suitable DNA marker for such species is essential. The cp genomes have often been used for phylogenetic studies and species identification because of their slower evolution in comparison with nuclear genomes [75]. In the present study, an analysis of the pairwise cp genomic alignment and SNP in four *Salvia* cp genomes revealed an increased number of variable sites in the IGS of the *trnV-ndhC*, *trnQ-rps16*, *atpI-atpH*, *psbA-ycf3*, *ycf1*, *rpoC2*, *ndhF*, *matK*, *rpoB*, *rpoA*, and *accD* genes. Thus, these regions may be used as new candidate fragments to identify *Salvia* spp. In addition, *ycf1a*, or *ycf1b* is the most variable plastid genome region and can serve as a core barcode for land plants [76]. However, more *Salvia* cp sequence data support is required and should be addressed in future research.

4.4. Relationship among the four *Salvia* spp.

S. przewalskii and *S. bulleyana* both belong to subgen. *Salvia*, while *S. miltiorrhiza* belongs to subgen. *Sclarea* and *S. japonica* belongs to subgen. *Allagospadonopsis*. This genetic relationship is reflected in the phylogenetic tree, where *S. przewalskii* and *S. bulleyana* are clustered together in one terminal branch. In terms of appearance and characteristics, many studies show that *S. przewalskii* and *S. bulleyana* are similar [77,78]. In terms of ingredients, *S. przewalskii* and *S. bulleyana*, which are often used as *S. miltiorrhiza* substitutes, contain most of the chemical content of *S. miltiorrhiza*; in contrast, *S. japonica* contains only some ingredients that are the same as those of *S. miltiorrhiza* [79,80]. In the phylogenetic tree, *S. przewalskii* and *S. bulleyana* form a robust monophyletic branch with *S. miltiorrhiza* first, and then with *S. japonica*. A total of 297 SNP sites were observed between the *S. przewalskii* and *S. bulleyana* genomes; this number is 785 among the *S. przewalskii*, *S. bulleyana*, and *S. miltiorrhiza* genomes. A total of 4982 variable sites were observed among the four *Salvia* cp genomes. All the analyses—including the repeat number and type, IR region borders, and phylogenetic tree—showed that *S. przewalskii* and *S. bulleyana* are closely related and that they differ more from *S. japonica* than from *S. miltiorrhiza*.

5. Conclusion

In this paper, we reported the complete cp genomes of *S. przewalskii* and *S. bulleyana*, which have long been used as *S. miltiorrhiza* surrogates in Southwest China. We also compared these two cp genomes with those of two other *Salvia* spp. published in the NCBI. The gene order and genome organization of the two *Salvia* species studied here are similar to those of the two *Salvia* cp genomes that have already been published. Moreover, repeated sequences containing SSRs were compared with six other cp genomes, revealing fine distinction. The variable site, IR region border, and phylogenetic analysis results showed that *S. przewalskii* and *S. bulleyana* are more similar to each other than to the two other compared species. The data and analysis in this paper may provide an in-depth understanding of the phylogenetic relationships between the species within the *Salvia* genus, and the complete cp genomes may be useful for future breeding and further biological discovery.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (QFSL2018004, 2017YFC1702100, and 81741060)

and the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZXKT17004).

Author contributions

Shilin Chen and Jiang Xu conceived and designed the research framework; Baozhong Duan and Weisi Ma collected and identified the sample; Conglian Liang and Lei Wang performed the experiments; Conglian Liang, Xiaofeng Shen, and Shuai Guo analyzed the data; and Conglian Liang and Lei Wang wrote the paper. Shuiming Xiao, Haijun Qi, Zheng Wang, and Yaoqi Liu polished the language. Haoyu Hu, Juan Lei, and Jiang Xu made revisions to the final manuscript. All the authors have read and approved the final manuscript.

Compliance with ethics guidelines

Conglian Liang, Lei Wang, Juan Lei, Baozhong Duan, Weisi Ma, Shuiming Xiao, Haijun Qi, Zhen Wang, Yaoqi Liu, Xiaofeng Shen, Shuai Guo, Haoyu Hu, Jiang Xu, and Shilin Chen declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eng.2019.01.017>.

References

- [1] Xiao XH, Fang QM, Xia WJ, Yin GP, Su ZW, Qiao CZ. Numerical taxonomy of medicinal *Salvia* L. and the genuineness of Danshen. *J Plant Resour Environ* 1997;6(2):17–21. Chinese.
- [2] Wang Y, Li D, Zhang Y. Analysis of ITS sequences of some medicinal plants and their related species in *Salvia*. *Yao Xue Xue Bao* 2007;42(12):1309–13. Chinese.
- [3] Li ZM, Xu SW, Liu PQ. *Salvia miltiorrhiza* Burge (Danshen): a golden herbal medicine in cardiovascular therapeutics. *Acta Pharmacol Sin* 2018;39(5):802–24.
- [4] Wang L, Ma R, Liu C, Liu H, Zhu R, Guo S, et al. *Salvia miltiorrhiza*: a potential red light to the development of cardiovascular diseases. *Curr Pharm Des* 2017;23(7):1077–97.
- [5] Chen W, Chen GX. Danshen (*Salvia miltiorrhiza* Bunge): a prospective healing sage for cardiovascular diseases. *Curr Pharm Des* 2017;23(34):5125–35.
- [6] Chinese Pharmacopoeia Commission. Pharmacopoeia of the People's Republic of China. 2015 ed. Beijing: People's Medical Publishing; 2015. Appendix 76.
- [7] Ren H, Hu X, Liu Y, Dai D, Liu X, Wang Z, et al. *Salvia przewalskii* extract of total phenolic acids inhibit TLR4 signaling activation in podocyte injury induced by puromycin aminonucleoside *in vitro*. *Ren Fail* 2018;40(1):273–9.
- [8] Yang Y, Wang ZP, Gao SH, Ren HQ, Zhong RQ, Chen WS. The effects of *Salvia przewalskii* total phenolic acid extract on immune complex glomerulonephritis. *Pharm Biol* 2017;55(1):2153–60.
- [9] Wang HQ, Yang LX, Chen XY, Yang PF, Chen RY. Chemical constituents from *Salvia przewalskii* root. *J Chin Med Mater* 2015;38(6):1197–201. Chinese.
- [10] Liu X, Liu Y, Yang Y, Xu J, Dai D, Yan C, et al. Antioxidative stress effects of *Salvia przewalskii* extract in experimentally injured podocytes. *Nephron* 2016;134(4):253–71.
- [11] Li SF, Li JQ, Li XJ, Wang XY, Du PJ, Zhou KP, et al. Pharmacognostical studies on *Salvia bulleyana* Diels. *Chin J Ethnomed Ethnopharm* 2008;7:18–22. Chinese.
- [12] Brunkard JO, Runkel AM, Zambryski PC. Chloroplasts extend stromules independently and in response to internal redox signals. *Proc Natl Acad Sci USA* 2015;112(32):10044–9.
- [13] Poccai P, Hyvönen J. The complete chloroplast genome sequence of the CAM epiphyte Spanish moss (*Tillandsia usneoides*, Bromeliaceae) and its comparative analysis. *PLoS ONE* 2017;12(11):e0187199.
- [14] Douglas SE. Plastid evolution: origins, diversity, trends. *Curr Opin Genet Dev* 1998;8(6):655–61.
- [15] Lin CP, Wu CS, Huang YY, Chaw SM. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biol Evol* 2012;4(3):374–81.
- [16] Mo JS, Kim K, Lee MH, Lee JH, Yoon UH, Kim TH. The complete chloroplast genome sequence of *Perilla citriodora* (Makino) Nakai. *Mitochondrial DNA Part A* 2017;28(1):131–2.
- [17] Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 2011;76(3–5):273–97.

- [18] Wang MX, Cui LC, Feng KW, Deng PC, Du XH, Wan FH, et al. Comparative analysis of Asteraceae chloroplast genomes: structural organization, RNA editing and evolution. *Plant Mol Biol Rep* 2015;33(5):1526–38.
- [19] Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 2007;8(1):174.
- [20] Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 2008;8(1):36.
- [21] Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol* 2016;209(4):1747–56.
- [22] Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev Camb Philos Soc* 2015;90(1):157–66.
- [23] Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 2016;17(1):134.
- [24] Cheng L, Li HP, Qu B, Huang T, Tu JX, Fu TD, et al. Chloroplast transformation of rapeseed (*Brassica napus*) by particle bombardment of cotyledons. *Plant Cell Rep* 2010;29(4):371–81.
- [25] Svab Z, Hajdukiewicz P, Maliga P. Stable transformation of plastids in higher plants. *Proc Natl Acad Sci USA* 1990;87(21):8526–30.
- [26] Repková J. Potential of chloroplast genome in plant breeding. *Czech J Genet Plant Breed* 2010;46(3):103–13.
- [27] Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, et al. Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat Commun* 2012;3(1):913.
- [28] Huang ZH, Xu J, Xiao SM, Liao BS, Gao Y, Zhai CC, et al. Comparative optical genome analysis of two pangolin species: *Manis pentadactyla* and *Manis javanica*. *GigaScience* 2016;5(1):1–5.
- [29] Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, et al. *Panax ginseng* genome examination for ginsenoside biosynthesis. *GigaScience* 2017;6(11):1–15.
- [30] Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, et al. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE* 2013;8(2):e57607.
- [31] Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol Plant* 2016;9(6):949–52.
- [32] He Y, Xiao H, Deng C, Xiong L, Yang J, Peng C. The complete chloroplast genome sequences of the medicinal plant *Pogostemon cablin*. *Int J Mol Sci* 2016;17(6):820.
- [33] Li XW, Gao HH, Wang YT, Song JY, Henry R, Wu HZ, et al. Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. *Sci China Life Sci* 2013;56(2):189–98.
- [34] Shen X, Wu M, Liao B, Liu Z, Bai R, Xiao S, et al. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules* 2017;22(8):1330.
- [35] Jiang D, Zhao Z, Zhang T, Zhong W, Liu C, Yuan Q, et al. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* 2017;8(9):227.
- [36] Vining KJ, Johnson SR, Ahkami A, Lange I, Parrish AN, Trapp SC, et al. Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. *Mol Plant* 2017;10(2):323–39.
- [37] Guo H, Liu J, Luo L, Wei X, Zhang J, Qi Y, et al. Complete chloroplast genome sequences of *Schisandra chinensis*: genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms. *Sci China Life Sci* 2017;60(11):1286–90.
- [38] Mariotti R, Cultrera NG, Díez CM, Baldoni L, Rubini A. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biol* 2010;10(1):211.
- [39] Yi DK, Kim KJ. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE* 2012;7(5):e35872.
- [40] Lu C, Shen Q, Yang J, Wang B, Song C. The complete chloroplast genome sequence of Safflower (*Carthamus tinctorius* L.). *Mitochondrial DNA, Part A* 2016;27(5):3351–3.
- [41] Wu JY, Xiao JF, Wang LP, Zhong J, Yin HY, Wu SX, et al. Systematic analysis of intron size and abundance parameters in diverse lineages. *Sci China Life Sci* 2013;56(10):968–74.
- [42] Xu J, Feng D, Song G, Wei X, Chen L, Wu X, et al. The first intron of rice EPSP synthase enhances expression of foreign gene. *Sci China Life Sci* 2003;46(6):561.
- [43] Yan J, Zhang Q, Yin P. RNA editing machinery in plant organelles. *Sci China Life Sci* 2018;61(2):162–9.
- [44] Chen C, Bundschuh R. Systematic investigation of insertional and deletional RNA-DNA differences in the human transcriptome. *BMC Genomics* 2012;13(1):616.
- [45] Knoop V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* 2011;68(4):567–86.
- [46] Grennan AK. To thy proteins be true: RNA editing in plants. *Plant Physiol* 2011;156(2):453–4.
- [47] Lutz KA, Maliga P. Chapter 23—transformation of the plastid genome to study RNA editing. *Methods Enzymol* 2007;424:501–18.
- [48] Asaf S, Khan AL, Khan MA, Waqas M, Kang SM, Yun BW, et al. Chloroplast genomes of *Arabidopsis halleri* ssp. *gemma* and *Arabidopsis lyrata* ssp. *petraea*: structures and comparative analysis. *Sci Rep* 2017;7(1):7556.
- [49] Yin D, Wang Y, Zhang X, Ma X, He X, Zhang J. Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of *Arachis*. *Sci Rep* 2017;7(1):11649.
- [50] Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X. Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers. *Int J Mol Sci* 2018;19(7):1962.
- [51] Flannery ML, Mitchell FJ, Coyne S, Kavanagh TA, Burke JL, Salamin N, et al. Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. *Theor Appl Genet* 2006;113(7):1221–31.
- [52] Shi QH, Yao ZP, Zhang H, Xu L, Dai PH. Comparison of four methods of DNA extraction from chickpea. *J Xinjiang Agric Univ* 2009;1:64–7. Chinese.
- [53] Kaila T, Chaduvla PK, Rawal HC, Saxena S, Tyagi A, Mithra S, et al. Chloroplast genome sequence of clusterbean (*Cyamopsis tetragonoloba* L.): genome structure and comparative analysis. *Genes* 2017;8(9):212.
- [54] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–20.
- [55] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403–10.
- [56] Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 2015;4(1):30. Corrected and republished from *GigaScience* 2012;1(1):18.
- [57] Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 2011;27(4):578–9.
- [58] Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 2012;13(1):715.
- [59] Misra S, Harris N. Using Apollo to browse and edit genome annotations. *Curr Protoc Bioinformatics* 2006;12(1):9.5.1–28.
- [60] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005;33(Suppl 2):W686–9.
- [61] Lohse M, Drechsel O, Bock R. Organellar genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 2007;52(5–6):267–74.
- [62] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30(12):2725–9.
- [63] Jansen RK, Kaittani C, Saski C, Lee SB, Tomkins J, Alverson AJ, et al. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosid. *BMC Evol Biol* 2006;6(1):32.
- [64] Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 2001;29(22):4633–42.
- [65] Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999;27(2):573–80.
- [66] Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 2003;106(3):411–22.
- [67] Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 2004;32(Suppl 2):W273–9.
- [68] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28(10):2731–9.
- [69] Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. *Proc Natl Acad Sci USA* 2009;106(31):12794–7.
- [70] Yang P, Shen WH, Shi JM, Chen XY, Zhang KK, Guan P. Identification of DNA barcoding of common medicinal plants in Lamiaceae. *Chin Tradit Herb Drugs* 2017;7:1397–402. Chinese.
- [71] Gao L, Wang B, Wang ZW, Zhou Y, Su YJ, Wang T. Plastome sequences of *Lygodium japonicum* and *Marsilea crenata* reveal the genome organization transformation from basal ferns to core leptosporangiates. *Genome Biol Evol* 2013;5(7):1403–7.
- [72] Fleischmann TT, Scharff LB, Alkhatib S, Hasdorf S, Schöttler MA, Bock R. Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 2011;23(9):3137–55.
- [73] Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, et al. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Mol Biol Evol* 2008;25(8):1566–75.
- [74] Hebert PD, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. *Proc R Soc B* 2003;270(1512):313–21.
- [75] Song Y, Chen Y, Lv J, Xu J, Zhu S, Li MF, et al. Development of chloroplast genomic resources for species discrimination. *Front Plant Sci* 2017;8:1854.
- [76] Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, et al. *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep* 2015;5(1):8348.
- [77] Zhang L, Yang Z, Huang X, Li J, Wan D. Study on the leaf epidermal structural characters of *Salvia miltiorrhiza* and *Salvia* from Sichuan. *J Sichuan Univ Nat Sci* 2008;45(3):674–80. Chinese.
- [78] Wang T, Liu SY, Wang L, Wang HY, Zhang L. Anatomical characteristics of laminae and petioles of 11 species of *Salvia* and their taxonomic significance. *China J Chin Mater Med* 2014;39(14):2629–34. Chinese.
- [79] Wang Y, Jiang K, Wang L, Han D, Yin G, Wang J, et al. Identification of *Salvia* species using high-performance liquid chromatography combined with chemical pattern recognition analysis. *J Sep Sci* 2018;41(3):609–17.
- [80] Skala E, Wysocki H. Tanshinone production in roots of micropropagated *Salvia przewalskii* maxim. *Z Naturforsch C* 2005;60(7–8):583–6.