

Article

Comparative and Phylogenetic Analyses of Complete Chloroplast Genomes of *Scrophularia incisa* Complex (Scrophulariaceae)

Ruihong Wang ¹, Jing Gao ^{1,2}, Jieying Feng ¹, Zhaoping Yang ³, Zhechen Qi ^{1,*} , Pan Li ^{4,*}  and Chengxin Fu ⁴

- ¹ Zhejiang Province Key Laboratory of Plant Secondary Metabolism and Regulation, College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou 310018, China
 - ² CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Wuhan 430074, China
 - ³ Key Laboratory of Biological Resources and Conservation and Application, College of Life Sciences, Tarim University, Alaer 843300, China
 - ⁴ Laboratory of Systematic & Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058, China
- * Correspondence: zqi@zstu.edu.cn (Z.Q.); panli@zju.edu.cn (P.L.)

Abstract: The *Scrophularia incisa* complex is a group of closely related desert and steppe subshrubs that includes *S. incisa*, *S. kiriloviana* and *S. dentata*, which are the only *S. sect. Caninae* components found in Northwest China. Based on earlier molecular evidence, the species boundaries and phylogenetic relationships within this complex remain poorly resolved. Here, we characterized seven complete chloroplast genomes encompassing the representatives of the three taxa in the complex and one closely related species, *S. integrifolia*, as well as three other species of *Scrophularia*. Comparative genomic analyses indicated that the genomic structure, gene order and content were highly conserved among these eleven plastomes. Highly variable plastid regions and simple sequence repeats (SSRs) were identified. The robust and consistent phylogenetic relationships of the *S. incisa* complex were firstly constructed based on a total of 26 plastid genomes from Scrophulariaceae. Within the monophyletic complex, a *S. kiriloviana* individual from Pamirs Plateau was identified as the earliest diverging clade, followed by *S. dentata* from Tibet, while the remaining individuals of *S. kiriloviana* from the Tianshan Mountains and *S. incisa* from Qinghai–Gansu were clustered into sister clades. Our results evidently demonstrate the capability of plastid genomes to improve phylogenetic resolution and species delimitation, particularly among closely related species, and will promote the understanding of plastome evolution in *Scrophularia*.

Keywords: *Scrophularia*; desert and steppe subshrubs; chloroplast genome; comparative genomics; phylogenomics



Citation: Wang, R.; Gao, J.; Feng, J.; Yang, Z.; Qi, Z.; Li, P.; Fu, C.

Comparative and Phylogenetic Analyses of Complete Chloroplast Genomes of *Scrophularia incisa* Complex (Scrophulariaceae). *Genes* **2022**, *13*, 1691. <https://doi.org/10.3390/genes13101691>

Academic Editor: Zhonghu Li

Received: 30 August 2022

Accepted: 19 September 2022

Published: 21 September 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/>).

1. Introduction

The figwort genus *Scrophularia* L. (Scrophulariaceae, Lamiales) comprises 200–300 species worldwide, is a taxonomically challenging and complicated lineage with great morphological diversity and high levels of reticulation evolution [1–5]. The genus is widespread across the broad temperate climate zone in the Northern Hemisphere, forming four diversity centers in Southwest Asia, Central Asia, the Himalayas and the Iberian Peninsula, with Iran and Turkey having the highest species richness with 49 and 60 species, respectively. Transcaucasia and Central Asia (74 species), East Asia and the Himalayas (55 species), as well as the Iberian Peninsula in Europe and adjacent regions (23 species) are three other secondary diversity centers [5–7]. *Scrophularia* was formally divided into two morphologically distinct sections in Stiefel's (1910) [8] widely acknowledged taxonomic treatment of the genus: *Scrophularia* sect. *Scrophularia* is characterized by perennial herbs

with reticulate leaf veins, and *Scrophularia* sect. *Caninae* is characterized by perennial subshrubs and xerophytes without reticulate leaf veins. The latter primarily distributed in Central Asia and Northwest China [9]. Over the last decade, great advances and extensive explorations have been made in understanding and interpreting the reticulate phylogeny and evolution of *Scrophularia* [5,10–15]. The most comprehensively sampled phylogeny of *Scrophularia* reported to date was based on nuclear ribosomal ITS and two plastid DNA loci (*trnQ-rps16*, *trnL-trnF*) comprising 147 species [5]. However, many subclades and infrageneric relationships (e.g., species in *S.* sect. *Caninae*) were only weakly supported, and incongruences between cytoplasmic and nuclear data were discovered [5].

S. incisa complex (*S.* sect. *Caninae*) is a group of subshrubby plants that are primarily (but not exclusively) distributed in the deserts and steppes of Northwest China. It consists of three closely related species: *S. incisa* Weinm., *S. kiriloviana* Schischk and *S. dentata* Royle ex Benth [16,17]. They are widely used for medicinal purposes to treat fever and exanthema in Chinese Traditional Tibetan medicine (TTM) and Traditional Mongolian medicine (TMM) due to their secondary metabolites such as iridoids and phenylpropanoids [18,19]. Of these species, *S. incisa* displays a belt-like pattern of distribution, primarily in the northern Chinese provinces of Qinghai and Gansu as well as in Inner Mongolia, spreading to Central Asia in the west and to Siberia in the east [7,16]; *S. dentata* occurs in western and southern Tibet, northwest India and Pakistan; *S. kiriloviana* is found in Xinjiang Province and Central Asia [7]. However, the interspecific taxonomy is difficult due to similar morphological characters and habitat preferences (e.g., floodplains, grasslands, mountain valleys) [20], with primary differences being the calyx membrane and leaf shape, which are considered to be greatly influenced by environmental factors (Figure 1). Our preliminary phylogenetic analyses of the genus *Scrophularia* based on nrITS and cpDNA placed this species complex as a monophyletic group embedded within the *S.* sect. *Caninae* clade [21]. Population-level phylogeographic analyses focusing on the *S. incisa* complex based on plastid DNA (*trnL-trnF*, *psbA-trnH*, *trnQ-rps16*) indicated that the evolutionary relationships among the major clades of the *S. incisa* complex were poorly resolved with insufficient genetic informative sites [17,21]. Therefore, more efficient techniques were required to facilitate species phylogenetic identification of the *S. incisa* complex.

Plastid genomes with a relatively small size, and a conserved structure and gene content have greatly facilitated the study of comparative genomics and species evolution [22–24]. Compared to traditional plastid DNA markers, the complete plastid genome arrays can provide substantially more genetic variation information and significantly improve the phylogenetic resolution of the species complex [24–26]. This approach is highly effective and recommended at low taxonomic levels, especially for closely related species, as demonstrated in previous reports [27–29]. Hence, in the present study, plastomes of individuals representing all the major clades in the *S. incisa* complex were reported, specifically aiming to: (1) characterize the entire plastid genomes of the *S. incisa* complex and comparative analyses with other plastomes of the genus using publicly available data; (2) gain insights into the *Scrophularia* plastid genome's evolutionary pattern and resolve phylogenetic relationships within the *S. incisa* complex; (3) examine variations in simple sequence repeats (SSRs), repeat sequences and mutational hotspot regions for future studies of species identification and phylogeographic studies of *Scrophularia*.

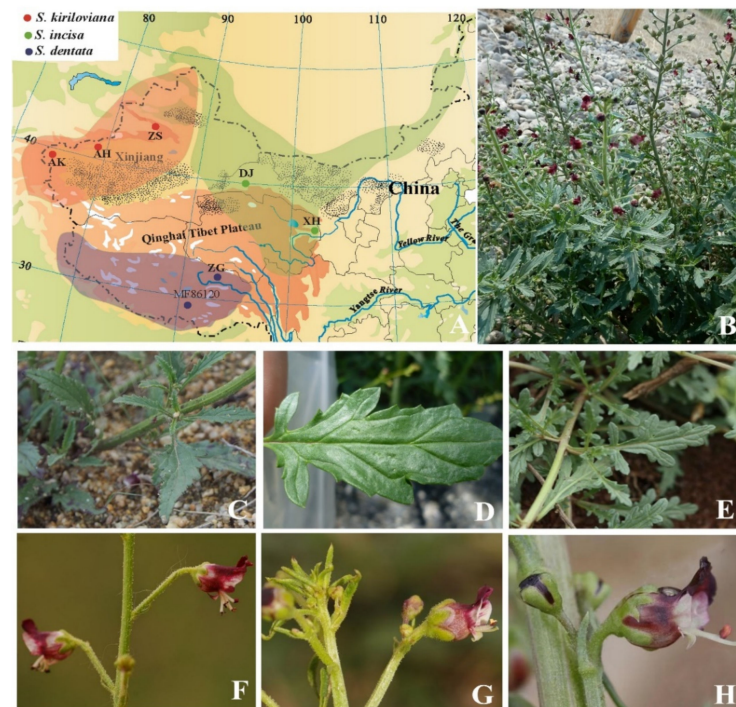


Figure 1. Distribution map and morphological characters of *S. incisa* complex. (A) Distribution map of *S. incisa* complex. (B) Morphology and habitat of *S. incisa*. (C–E) Leaves of *S. incisa*, *S. kiriloviana* and *S. dentata*. (F–H) Flowers of *S. incisa*, *S. kiriloviana* and *S. dentata* (Photo (B): by Chong-Lie Qu, (C): by Guo-Jun Hua, (D,G): by Pan Li, (E): by Xin-Xin Zhu, (F): Jian-Bin Pan, (H): by Po-Po Wu).

2. Materials and Methods

2.1. Plant Samples, Sequencing and Assembly

Leaves of six individuals of major clades of *S. incisa* complex, including three individuals of *S. kiriloviana* from Xinjiang, two individuals of *S. incisa* from Qinghai and one individual of *S. dentata* from Tibet, together with a closely related species *S. integrifolia* from Tajikistan, were collected (Figure 1, Table 1). The voucher specimens were deposited in herbarium of Zhejiang Sci-Tech University (HZSTU), and specimen numbers are listed in Table S1. Genomic DNA was extracted from silica gel dried leaf material using DNA Plantzol (Invitrogen, Carlsbad, CA, USA) and sequenced on the platform Illumina Hiseq2500 at Beijing Genomics Institute (BGI, Shenzhen, China). Complete chloroplast DNA sequences of *S. buergeriana* (KP718626), *S. dentata* (MF861202), *S. henryi* (NC036943) and *S. takeimensis* (KP718628) were downloaded from NCBI for comparative analyses with *S. incisa* complex.

The de novo assembly and reference-guided combined method was implemented to assemble these raw reads into complete plastid genomes [30]. First, for each *Scrophularia* individual, raw reads of the paired-end sequencing data were trimmed to get rid of adapters and the reads with the Phred value of base calling quality score less than 20 using CLC Genomics Workbench v10.1.1 (CLC Bio, Aarhus, Denmark; <http://www.clcbio.com> (accessed on 22 June 2017)). Second, the left reads with high quality were assembled into contigs applying the CLC assembler, under all optimally controlled parameters as: minimum contig length of 200 bp, mismatch cost = 2, deletion and insertion costs = 3, length fraction = 0.9 and similarity fraction = 0.8. Third, because of both organellar and nuclear DNA contained in the sequencing data, all contigs were aligned and directed according to the reference plastome of *S. buergeriana* (KP718626) to obtain the correct ones by BLAST (<http://blast.ncbi.nlm.nih.gov/> (accessed on 23 June 2008)). Finally, the filtered clean contigs were ordered and realigned with the reference chloroplast genome and complete chloroplast sequences of *Scrophularia* were generated in GENEIOUS v11.01 (<http://www.geneious.com> (accessed on 26 September 2017)) [23,26].

Table 1. The basic characteristics of *Scrophularia* chloroplast genomes.

Characteristics	<i>S. integrifolia</i>	<i>S. kiriloviana</i>	<i>S. kiriloviana</i>	<i>S. kiriloviana</i>	<i>S. incisa</i>	<i>S. incisa</i>	<i>S. dentata</i>	<i>S. dentata</i>	<i>S. buergeriana</i>	<i>S. takesimensis</i>	<i>S. henryi</i>
	(Tajikistan H8)	(Xinjiang AH18)	(Xinjiang AK1)	(Xinjiang ZS3)	(Gansu DJ1)	(Qinghai XH7)	(Tibet ZG4)				
Total cpDNA size (bp)	152,088	152,554	152,365	152,528	152,600	152,481	152,509	152,600	153,631	152,425	152,868
Latitude (°N)	36.617	40.746	38.762	43.197	39.633	35.662	29.490	30.653	—	—	31.470
Longitude (°E)	68.564	77.826	75.193	81.201	94.340	102.644	97.470	97.568	—	—	110.395
LSC length	83,589	84,009	83,892	84,057	84,168	83,945	84,071	84,091	84,454	83,531	84,020
SSC length	17,505	17,561	17,473	17,493	17,404	17,506	17,430	17,447	17,929	17,938	17,940
IR length	25,497	25,492	25,500	25,489	25,514	25,515	25,504	25,531	25,624	25,478	25,454
Total GC content (%)	38	38	38	38	37.9	38	38	38	38	38.1	38
LSC	36.1	36.1	36.1	36.1	36	36.1	36.1	36	43.2	36.2	36.1
SSC	32.1	32.1	32.1	32.2	32.1	32.1	32.2	32.2	32.2	32.2	32.2
IR	43.1	43.1	43.1	43.1	43.1	43.1	43.1	43.1	43.2	43.2	43.2
Total number of genes	132	132	132	132	132	132	132	132	132	132	132
Protein-coding genes	80	80	80	80	80	80	80	80	80	80	80
rRNA genes	4	4	4	4	4	4	4	4	4	4	4
tRNA genes	31	31	31	31	31	31	31	31	31	31	31
Duplicated genes	17	17	17	17	17	17	17	17	17	17	17
GenBank Acc. No.	OP018678	OP018676	OP036427	OP036428	OP036429	OP018675	OP018677	MF861202	KP718626	KP718628	MF861203

IR, inverted repeat region; LSC, large single-copy region; rRNA, ribosomal RNA; SSC, small single-copy region; tRNA, transfer RNA.

2.2. Chloroplast Genome Annotation, Comparison and Codon Usage Bias Analysis

Complete plastomes were annotated in the GENEIOUS v11.01 and DOGMA (Dual Organella GenoMe Annotator) database software [31–34]. The genes' information about start/stop codon positions and exon/intron boundaries were accurately identified and artificially modified by reference to homologous genes of *S. buergeriana* (KP718626). Additionally, all the determined tRNA genes were further confirmed by the predicted corresponding structures in tRNAscan-SE v1.21 [35]. The plastid genome maps were drawn by OrganellarGenome [36]. Finally, the seven annotated plastid genomes were submitted to GenBank. The codon usage bias parameters including effective number of codons (ENC), relative synonymous codon usage (RSCU) value [37,38], as well as the GC contents respectively corresponding to positions of the first, second, third codon of objective coding sequences (CDS) and the whole plastomes were estimated with CUSP and CHIPS plugins in EMBOSS (<http://vmbioinfo.toulouse.inra.fr/emboss> (accessed on 15 July 2007)) and CODONW v1.4.2 (<http://codonw.sourceforge.net/> (accessed on 12 May 2005)).

Chloroplast genome comparative analyses across these *Scrophularia* species mentioned above and the *S. incisa* complex plastid genomes were performed with mVISTA tool (genome.lbl.gov/vista/index.shtml (accessed on 1 July 2004)) in reference to the annotation of *S. buergeriana* [39]. For the purpose of discriminating hypervariable regions, all plastome sequences were aligned and subjected to diversity analysis using DNASP v5.1 [40] to evaluate the total number of mutations (E_{ta}) and nucleotide diversity (P_i) values of protein-coding and non-coding regions that were extracted following criterion of the aligned length larger than 200 bp and no less than one mutation site.

2.3. Characterization of Repeat Sequences and Simple Sequence Repeats

The program REPuter [41] was operated to recognize detailed position and size information of the repeat sequences, exactly involving direct (forward), inverted (palindromic), reverse and complement repeats. The constraint conditions for all the above repeat patterns were repeat size, gap size between repeats and sequence identity more than 30 bp, 3 bp and 90%, respectively. In addition, simple sequence repeats' (SSRs) loci across these eleven cp genomes of *Scrophularia* were searched by running MICROSATELLITE (MISA) perl script [42], according to the repeat numbers of mono-, di-, tri-, tetra-, penta- and hexanucleotide SSRs more than 10, 5, 5, 3, 3 and 3 repeat units, respectively. Generally, SSRs have significant influence on genome rearrangement and recombination [43,44].

2.4. Selective Pressure Analysis

In genetics, the ratio (ω) of non-synonymous (K_a) and synonymous (K_s) substitution rates is usually employed to ascertain whether the selection pressure acting on the protein-coding gene exists or not [45–47]. In this study, we used the program of MAFFT v7 and DnaSP v5.1 [40,48] to extract and align the CDS region, also calculate values of K_a , K_s and K_a/K_s for each plastid gene. Genes' evolutionary processes, namely, positive, purifying or neutral selection were identified as $K_a > K_s$ ($\omega > 1$), $K_a < K_s$ ($\omega < 1$) or $K_a = K_s$ ($\omega = 1$), respectively [49].

2.5. Phylogenetic Analysis

Altogether 26 chloroplast genomes representing 7 genera of Scrophulariaceae (Table S1) were directed at phylogenetic tree construction, including 13 representatives of *Scrophularia*, two of *Verbascum*, five of *Buddleja* and six individuals from four genera of tribe Myoporeae as the outgroup taxa based on previous studies [50–52]. These 26 plastome sequences were aligned with MAFFT v7 [48]. The maximum likelihood (ML) algorithm was performed in RAxML-HPC v8.2.20 [53,54], with a GTR + I + G substitution model ascertained under the Bayesian Information Criterion (BIC) in jModelTest v2.1.7 [55]. Bayesian inference (BI) of phylogeny was conducted in MrBayes v3.2.5 [56,57], with the Markov chain Monte Carlo (MCMC) algorithm executed for 2 million generations and trees sampled every

500 generations, discarding the first quarter of calculated trees as burn-in and constructing a consensus tree from the remaining trees to compute posterior probabilities (PPs).

3. Results

3.1. Genome Organization and Features

Plastomes of six individuals of the *S. incisa* complex and one individual of *S. integrifolia* were sequenced using the Illumina HiSeq 2500 system with totals from 14,276,152 to 20,707,954 clean reads produced. Seven complete plastid genomes from four species of *S. sect. Caninae* with no gaps were obtained by means of de novo and reference-guided assembly. Integrating previously published chloroplast genomes of *S. dentata* (MF861202), *S. buergeriana* Miq. (KP718626), *S. henryi* Hemsl. (NC036943) and *S. takeimensis* Nakai (KP718628), the nucleotide sequences of the eight *S. sect. Caninae* chloroplast genomes were determined to be 152,088–152,688 bp, and the three *S. sect. Scrophularia* species ranged limitedly from 152,425 bp to 153,631 bp (Figure 2, Table 1). All of those plastomes presented a typical quadripartite structure as with other angiosperms, consisting of a pair of inverted repeats (IR) (50,908–51,248 bp) partitioned by the large single-copy (LSC, 83,531–84,454 bp) and small single-copy (SSC, 17,431–17,941 bp) regions [58–60]. The overall GC content among these cp genomes was practically identical (37.9–38.1%), whereas the regionalized GC content in the LSC, SSC and IR were 36.0–36.2%, 32.1–32.2% and 43.1–43.2%, respectively (Table 1).

The eleven *Scrophularia* plastomes coded the same series of 132 genes, including 80 protein-coding genes, 31 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes, thereinto 115 unique and 17 duplicated genes (Tables 1 and 2). Among the 115 unique genes, 9 protein-coding genes and 5 tRNA genes contained one single intron, while 3 genes (*clpP*, *rps12*, and *ycf3*) had two introns. The region *ycf1* gene appeared as a pseudogene (Table 2), as has been shown in almost all plant chloroplast genomes sequenced at present, and it has high variability [23,58] because certain internal stop codons were found in the coding sequence. Furthermore, the partial duplicate of *rps19* at the LSC/IRa joined region lost the function of protein-coding due to the incomplete duplication [61,62].

Codon usages were fairly similar across these *Scrophularia* individuals. The codon numbers encoded in the eleven *Scrophularia* chloroplast genomes ranged from 50,696 to 51,210. The effective number of codon (ENC) values ranged from 55.423 (*S. dentata* Tibet ZG4) to 56.141 (*S. kiriloviana* Xinjiang AK1) (Table S2). The most and least prevalent amino acids were leucine (approx. 9.9%) and tryptophane (approx. 1.3%), respectively (Table S2). Except for *Met* and *Trp* amino acids encoded by a particular codon, the rest showed obvious codon usage bias and that partial codons appeared much more frequently. Among a set of synonymous codons, codons ending in A or T have a relatively high number and RSCU, while codons ending in C or G have a relatively low number and RSCU. In all codons, the total GC content was approximately 37.9%, and the GC content at the third position of the synonym codons was approximately 37.9% (Table S2), which fully indicated that the codons, especially those at the third position, had a strong property of A/T, as well as being consistent with many previous studies [63,64].

Altogether, our newly obtained results generally correspond with other previously reported *Scrophularia* species [50,65], and the eleven *Scrophularia* cp genomes were highly conserved throughout evolution in various aspects of genome features concerning gene content, arrangement and GC content, as well as codon usage.



Figure 2. Gene maps of the seven *S. sect. Caninae* individuals including three *S. kiriloviana*, two *S. dentata*, two *S. incisa* and one *S. integrifolia*. Genes shown on the outside of the circle are transcribed clockwise, and genes inside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner corresponds to GC content, and the lighter gray corresponds to AT content.

Table 2. Gene composition of *Scrophularia* chloroplast genomes.

Groups of Gene	Name of Gene
Ribosomal RNAs	<i>rrn16</i> (×2), <i>rrn23</i> (×2), <i>rrn4.5</i> (×2), <i>rrn5</i> (×2)
Transfer RNAs	^a <i>trnA</i> -UGC(×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -CAU, ^a <i>trnG</i> -GCC, <i>trnG</i> -UCC, <i>trnH</i> -CAU, <i>trnH</i> -GUG, <i>trnI</i> -CAU, ^a <i>trnI</i> -GAU(×2), <i>trnK</i> -UUU, <i>trnL</i> -CAA(×2), ^a <i>trnLUAA</i> , <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU(×2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG(×2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC(×2), ^a <i>trnV</i> -UAC, <i>trnW</i> -CCA, <i>trnY</i> -GUA
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i>
Cytochrome	<i>petN</i> , <i>petA</i> , <i>petL</i> , <i>petG</i> , ^a <i>petB</i> , ^a <i>petD</i>
ATP synthase	<i>atpA</i> , ^a <i>atpF</i> , <i>atpH</i> , <i>atpI</i> , <i>atpE</i> , <i>atpB</i>
Rubisco	<i>rbcL</i>
NADH dehydrogenase	<i>ndhJ</i> , <i>ndhK</i> , <i>ndhC</i> , ^a <i>ndhB</i> (×2), <i>ndhF</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhI</i> , ^a <i>ndhA</i> , <i>ndhH</i>
ATP-dependent protease subunit P	^b <i>clpP</i>
Chloroplast translational initiation factor	<i>infA</i>
Chloroplast envelope membrane protein	<i>cemA</i>
Large units	<i>rpl33</i> , <i>rpl20</i> , <i>rpl36</i> , <i>rpl14</i> , ^a <i>rpl16</i> , ^a <i>rpl2</i> (×2), <i>rpl23</i> (×2), <i>rpl32</i>
Small units	^a <i>rps16</i> , <i>rps2</i> , <i>rps14</i> , <i>rps4</i> , <i>rps18</i> , ^b <i>rps12</i> (×2), <i>rps11</i> , <i>rps8</i> , ^Ψ <i>rps19</i> , <i>rps3</i> , <i>rps7</i> (×2), <i>rps15</i>
RNA polymerase	<i>rpoC2</i> , ^a <i>rpoC1</i> , <i>rpoB</i> , <i>rpoA</i>
Miscellaneous proteins	<i>matK</i> , <i>accD</i> , <i>ccsA</i>
Hypothetical proteins and conserved reading frame	^b <i>ycf3</i> , <i>ycf4</i> , <i>ycf2</i> (×2), ^Ψ <i>ycf1</i> , <i>ycf15</i> (×2)

^a Indicates the genes containing a single intron. ^b Indicates the genes containing two introns. (×2) Indicates genes duplicated in the IR regions. Pseudogene is represented by ^Ψ.

3.2. Contraction and Expansion of Inverted Repeats (IRs)

Generally, the expansion and contraction of boundary regions between the IR and SC always bring about change in plant chloroplast genomes' length [27]. The IR/SC region boundaries were analyzed and compared across these eleven *Scrophularia* chloroplast genomes (Figure 3). The overall genomic structure was basically the same, and the slight expansion of the IRs might be caused by gene translocation. Specifically, the IR_a region extended 40 and 41 bp towards the *rps19* gene, resulting in the *rps19* pseudogene fragment being situated at the margin between the IR_a and LSC. The location of fragment *ycf1* with 877–910 bp was at the IR_b section as the SSC–IR_a borders stretched into the *ycf1* gene. For the *S. henryi* chloroplast genome, the *ycf1* fragment passed the IR_a to the SSC region by 1 bp. Moreover, the *ycf1* genes were entirely located in the SSC region for other individuals (Figure 3).

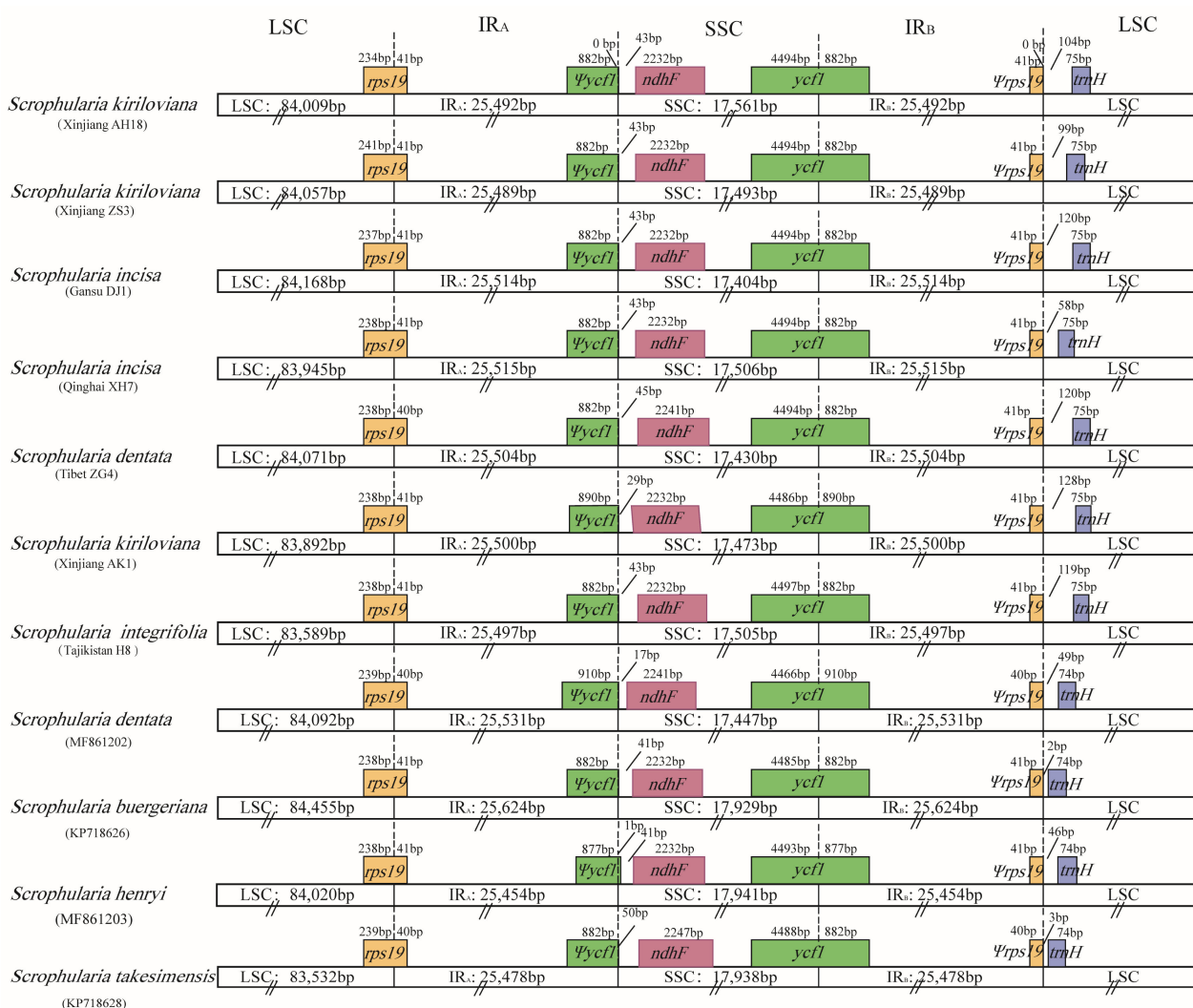


Figure 3. Comparison of LSC, IR and SSC junction positions among *Scrophularia* plastomes.

3.3. Genome Comparative Analysis and Molecular Marker Identification

The comprehensive comparison of sequence divergence was examined methodically by the mVISTA package for the eleven *Scrophularia* plastomes in reference to the annotation of *S. buergeriana*. The sequence comparison and alignment showing high levels of similarity with the sequence identity of a few regions less than 90%, suggested that the *Scrophularia* chloroplast genomes' organization are structurally conservative (Figure 4).

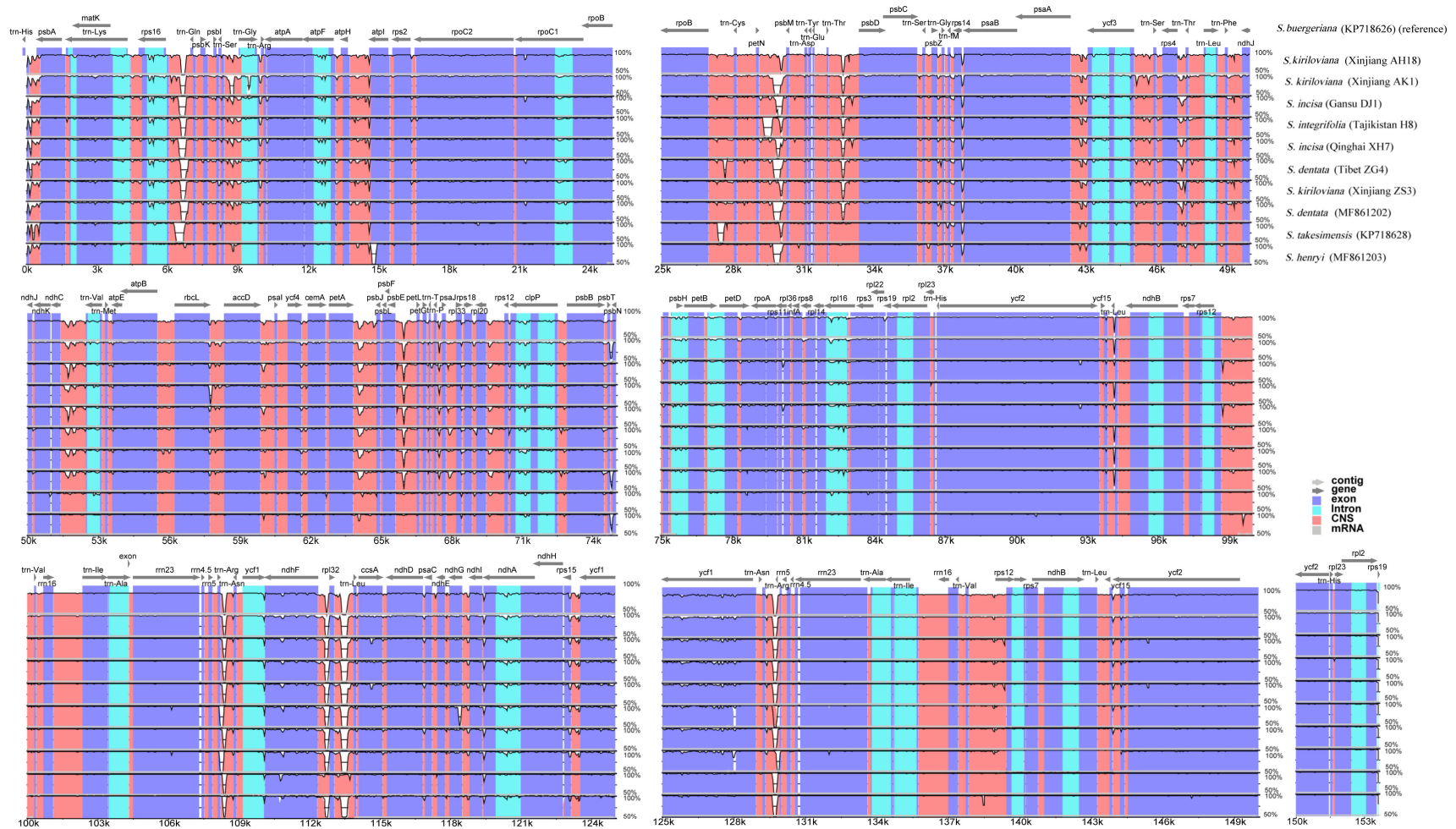


Figure 4. Sequence identity plots among the eleven *Scrophularia* chloroplast genomes, with *S. buergeriana* as a reference. Annotated genes are displayed along the top. The vertical scale represents the percent identity between 50 and 100%. Genome regions are color-coded as exon, intron and conserved non-coding sequences (CNS).

In conformity to most studies [66,67], IRs displayed much less variation than polytropic LSC and SSC. Comparison of the coding genes, and non-coding and intron regions generated 129 loci (60 coding genes, 59 intergenic spacers, and 10 introns) from those individuals (Figure 5; Table S3). The values of nucleotide variability (P_i) for these loci in the coding regions ranged from 0.00158 (*ycf2*) to 0.02438 (*rps15*); *rps15*, *ndhG*, *ycf1*, *ccsA*, *ndhF* and *rps11* with a high P_i value (>0.01) were considered as divergence hotspots. The P_i values of non-coding regions ranged from 0.00112 to 0.04866; in addition, 11 of the most divergent and variable regions with P_i values greater than 0.015 included *trnH-psbA*, *rpl32-trnL*, *ndhF-rpl32*, *trnL-rps16*, *ycf3-trnS*, *petA-psbJ*, *ndhE-ndhG*, *ndhC-ndhV*, *ccsA-ndhD*, *ycf4-cemA* and *rps4-trnT*. The hotspot and variable regions could be profitable in the development of genetic markers for the species phylogenetic identification analysis of *Scrophularia*.

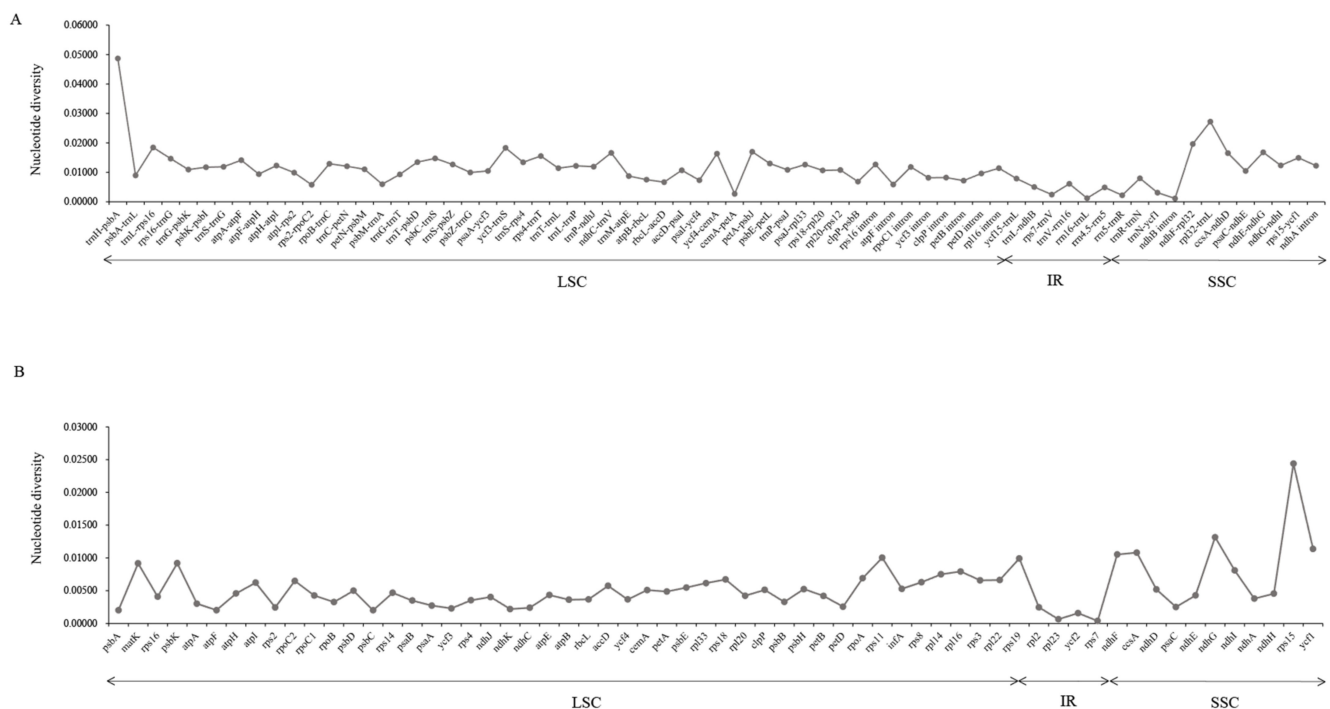


Figure 5. Comparison of nucleotide variability (P_i) values in *Scrophularia* plastomes. (A) P_i values among intergenic spacer (IGS) and intron regions. (B) P_i values among protein-coding genes (CDS).

3.4. Characterization of Repeat Sequences and SSRs

Altogether there were 473 repeats detected in the eleven *Scrophularia* cp genomes with REPuter program [41], including 238 forward, 223 palindromic, 8 reverse and 4 complement repeats. *S. incisa* (Qinghai XH7) comprised the maximum number of repeats (52), whereas *S. dentata* possessed the least (38) (Figure 6a). The percentage of the repeats' size ranging between 30 and 39 bp reached 84.8% for all the individuals of *Scrophularia* (Figure 6b; Table S4). According to the principle that the fragments of the same length discovered in homologous DNA regions were considered as shared repeats [27,28], 32 investigated repeats were shared by the eleven *Scrophularia* cp genomes. Additionally, *S. incisa* (Qinghai XH7) and *S. takesimensis* exhibited the most distinct repeats (8), while *S. henryi* and *S. buergeriana* contained the fewest number (2), respectively (Table S5).

SSR is a repetitive sequence consisting of simple repeating units in series and has been extensively used as the effective molecular marker for phylogenetic research and genetic diversity analysis [68,69]. A total of 493 SSRs were recognized across the 11 *Scrophularia* plastomes through MISA analysis identification. The amount of SSRs per single plastome ranged between 40 (*S. integrifolia* Tajikistan H8) and 53 (*S. incisa* Qinghai XH7), with 11 SSRs shared among all plastomes (Table S6). Among these SSRs, more than three-quarters (76.2%) were constituted by A/T base components (Figure 7a). Maybe owing to the reason that the intergenic spacers (IGS) possessed a relatively higher level of mutation rates [27,29],

most SSR loci lay in IGS (62.3%), followed by CDS (24.7%) and introns (13.0%) (Figure 7b). Generally, the cp genome SSRs revealed abundant genetic variability, and therefore can be applied to the study of the population genetics of *Scrophularia* species in the future.

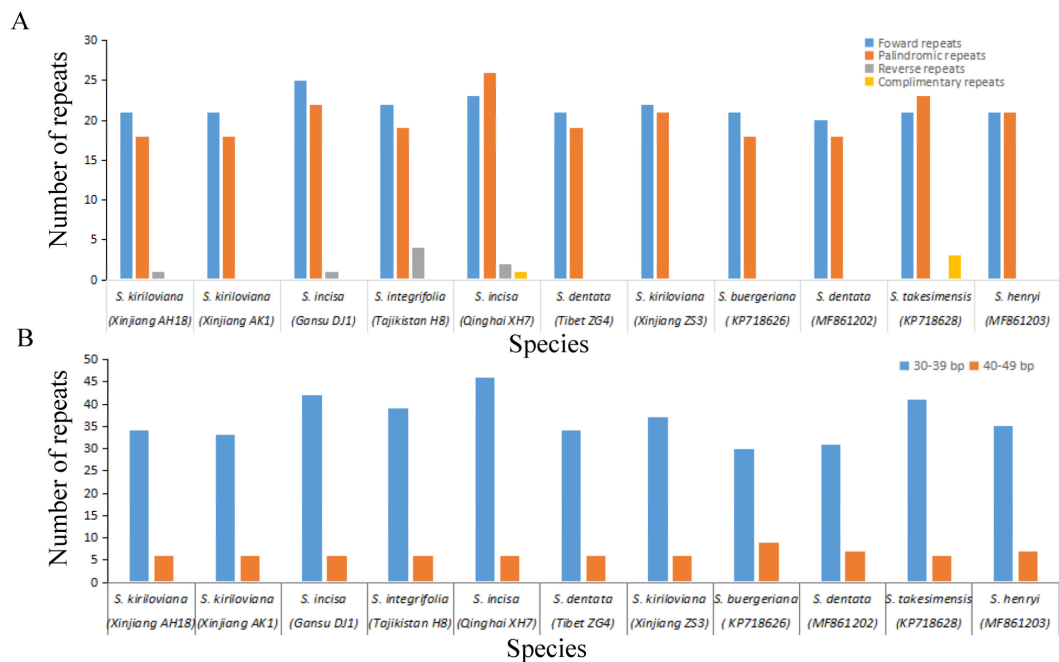


Figure 6. Repeat analyses in eleven chloroplast genomes of *Scrophularia*. (A) Frequency of repeat types. (B) Frequency of repeats by length.

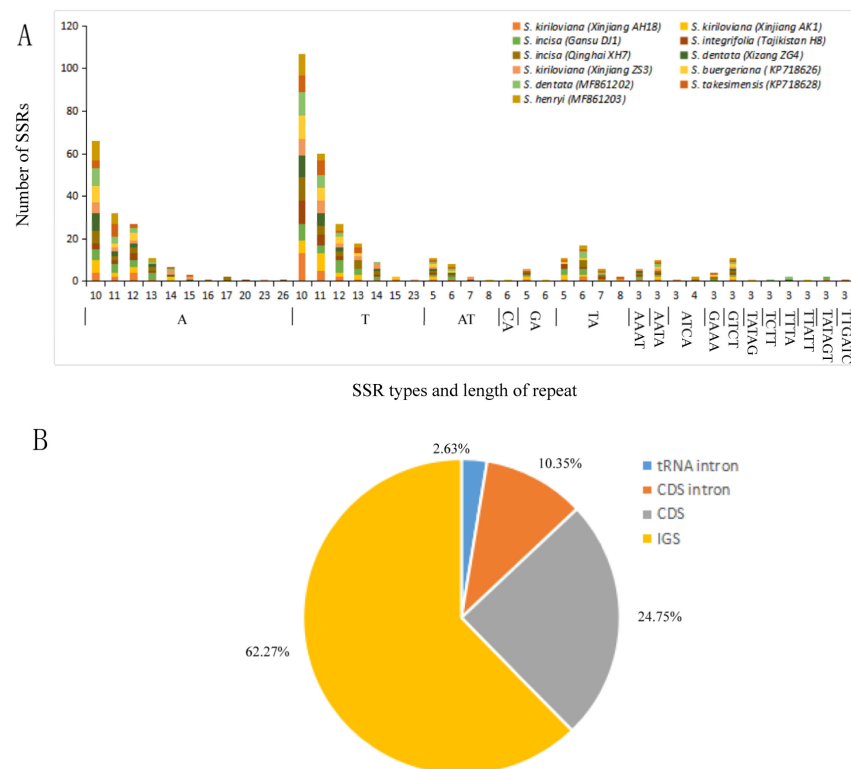


Figure 7. Simple sequence repeats (SSRs) in the eleven *Scrophularia* chloroplast genomes. (A) Number of SSRs by length. (B) Distribution of SSR loci. CDS, coding DNA sequence; IGS, intergenic spacer region.

3.5. Selective Pressure Analysis

We calculated the non-synonymous (Ka) and synonymous (Ks) substitution rates together with the Ka/Ks (ω) value to detect the selective pressure on the 80 cp genes of eleven *Scrophularia* individuals. The average Ka/Ks values among the *S. incisa* complex ranged from 0.0001 (*S. dentata* Tibet ZG4) to 2.65591 (*S. incisa* Gansu DJ1), and the highest Ka/Ks value was 3.2702 (*S. buergeriana*) among these 11 *Scrophularia* individuals (Table S7, Figure 8). In fact, positive selection for the eleven individuals was exclusively observed in *S. incisa* (Gansu DJ1), and other genes had higher synonymous than non-synonymous changes, suggesting high evolutionary constraints in the *Scrophularia* genome species. The Ka/Ks of most genes were less than 1, indicating that they may be subjected to purification options, i.e., mutations were detrimental and eliminated in the population (Table S7). Among them, 21 groups had Ka/Ks values greater than 1: the *rpl22* gene of *S. takesimensis* (KP718628) (Ka/Ks = 1.01824, $p > 0.05$), the *rps15* gene of all the individuals except *S. buergeriana* (KP718626) and *S. henryi* (MF861203), and the *ycf15* gene of all the individuals (average Ka/Ks = 2.1879, $p > 0.05$) (Figure 8).

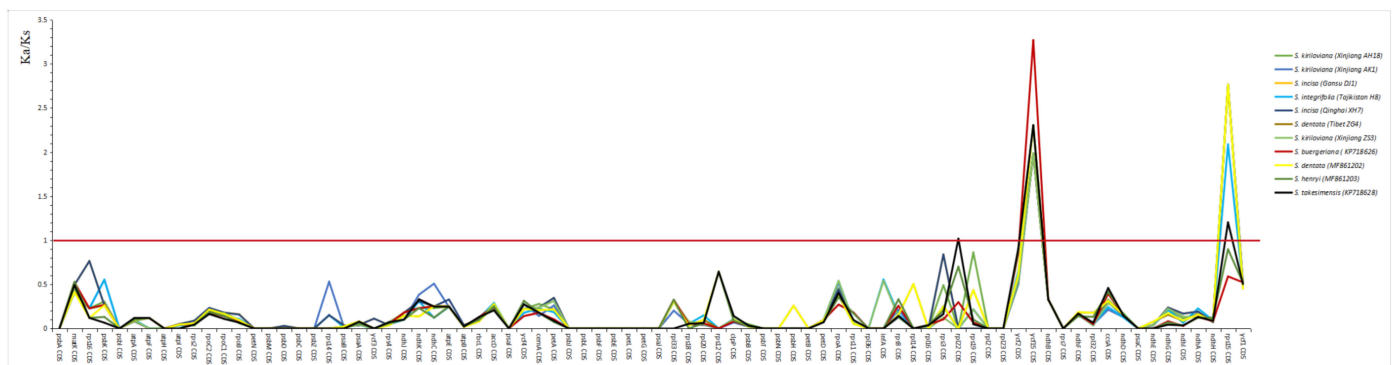


Figure 8. The Ka/Ks analysis of eleven individuals of *Scrophularia*.

3.6. Phylogenetic Analyses

Both ML and BI phylogenetic reconstruction analyses of the 26 complete cp genomes from Scrophulariaceae (Table S1) generated identical tree topologies and achieved high support values in each branch. The phylogenetic tree nodes nearly received fully strong support (PP/BS = 1/100), with three monophyly groups manifested involving tribe Scrophularieae, tribe Buddlejaceae and tribe (Myoporeae + Leucophylleae) as outgroup taxa (Figure 8). *Buddleja* was sister to *Verbascum* and *Scrophularia*, and *S. sect. Caninae* was resolved as a robust monophyletic clade. Within *S. sect. Caninae*, *S. integrifolia* (Tajikistan H8) was sister to the clade of the *S. incisa* complex, and within the complex *S. kiriloviana*, AK1 was identified as the earliest diverging clade, followed by *S. dentata* from Tibet, while the remaining individuals of *S. kiriloviana* (ZS3, AH18) and *S. incisa* (XH7, DJ1) were clustered into sister clades (Figure 9). Here, the phylogenetic relationship pattern within the *S. incisa* complex was consistent with our previous phylogeographic study based on chloroplast DNA fragments with relatively weak support [17,21]. In conjunction with previous molecular dating of the *S. incisa* complexes' Plio-Pleistocene divergence [17,21], we speculated that speciation and population divergence within the *S. incisa* complex may be due to the vicariance and fragmentation distribution triggered by aridification increase and sand deserts' enlargement, which was brought by monsoon-dominated climate patterns and geological changes in the QTP *sensu lato* in Northwest China [70–73]. We also suggested that the application of a whole cp genome as a super-barcode significantly improves the effectiveness and brings a better resolution within the closely related taxa than the traditional markers of universal or specific DNA barcodes, which provide too little variable genetic information in a certain group of species [25,58].

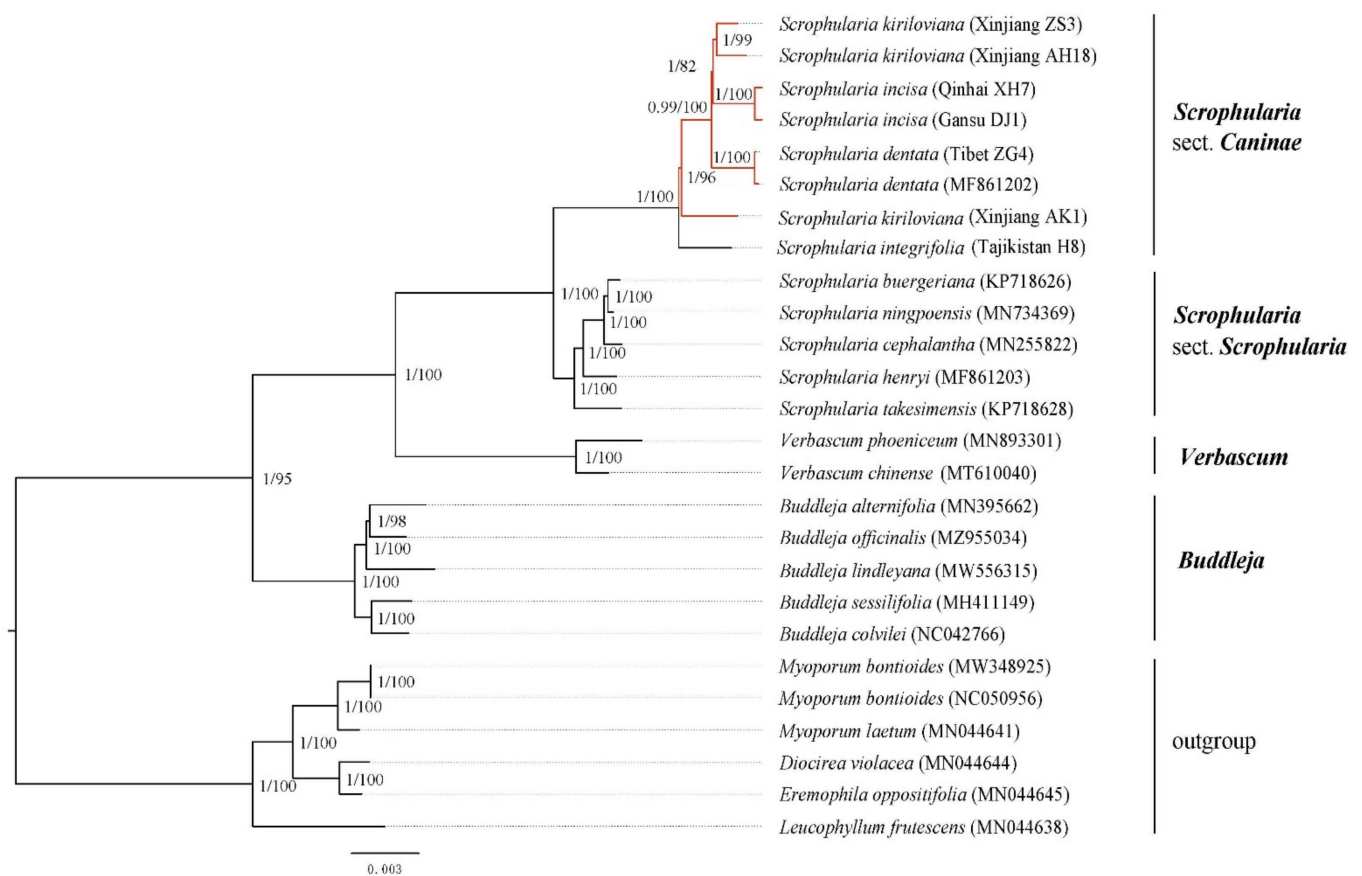


Figure 9. Phylogenetic relationships of *Scrophularia* species inferred from Bayesian inference (BI) and maximum likelihood (ML) based on the complete cp genome sequence dataset. Support values marked above the branches follow the order PP (posterior probability)/BS (bootstrap support).

4. Discussion

4.1. Comparative Genomics

Our comparative results revealed that the genome structure and features including gene content, gene arrangement and GC content, as well as codon usage within both the eight *S. sect. Caninae* individuals and the eleven *Scrophularia* cp genomes were highly conserved, which was also consistent with the general nature of the highly conserved chloroplast genomes in angiosperms. The determined nucleotide sequences of the eight *S. sect. Caninae* chloroplast genomes ranged narrowly from 152,088 bp to 152,868 bp, and the three *S. sect. Scrophularia* species ranged from 152,425 bp to 153,631 bp (Figure 2, Table 1). The amount of protein-coding, rRNA, and tRNA genes in these individuals were identical. The overall GC contents among these plastomes were almost equal (37.9–38.1%), with 36.0–36.2%, 32.1–32.2% and 43.1–43.2% corresponding to the LSC, SSC and IR, respectively (Table 1). This conservative property of *S. sect. Caninae* plastid genomes was probably due to the morphological character and habitat preference similarities as well as the close affinities between them, while the cp genome conservative property of the genus *Scrophularia* was suggested due to the low chloroplast substitution rate and relatively recent evolutionary divergence time [17,21,50]. There are pseudogenes, *ycf1* and *rps19*, which deserve closer scrutiny. Comparative genomics have shown that the *ycf1* gene is found in almost all plant chloroplast genomes sequenced up to now and has high variability [74–76]. The IR/LSC boundaries of *Scrophularia* stretched into *rps19*, which was congruent with many standard angiosperms' chloroplast genome feature [61]. IR contraction and expansion is a common evolutionary phenomenon [77–81] and may cause variation in the lengths of plastid genomes [80]. The slight differences in the IR/SC boundaries in *Scrophularia* may be caused by IR contraction/expansion. We performed full sequences' alignment and

nucleotide polymorphism analysis on 11 individuals, and the results indicated that the IR regions were more conserved than the LSC and SSC (Figure 4). Maybe due to the consensus when a mutation occurs in IRs, the repetitive gene is corrected by gene conversion, thus reducing the frequency of variation in the remodeling [82]. In addition, because of this, the IRs play a significant role in stabilizing the plastid structure.

SSRs are often used as molecular markers for population genetic studies. We identified 40 to 53 chloroplast SSRs (Table S5) in the eleven individuals that can be applied to conduct population genetic research of *Scrophularia* in the future. We performed a selective pressures analysis on the genes, and the results showed that the Ka/Ks of most genes was fewer than 1, and they may have received purifying selection. The Ka/Ks of *rpl22* in *S. takesimensis* (KP718628) was 1.01824, but it was less than 1 in other individuals, so we thought that this group of data does not indicate that the gene has received positive selection.

4.2. Phylogenetic Relationships within *S. incisa* Complex and Scrophulariaceae

Phylogenetic analysis can ordinarily provide important insights into the organism's origin and evolution [83–85]. Scrophulariaceae is now known as being composed of the eight tribes: Aptosimeae, Buddlejaceae, Hemimerideae, Leucophylleae, Limoselleae, Myoporeae, Scrophularieae and Teedieae [52,86,87]. The phylogeny and relationship of the genus and tribes in Scrophulariaceae were not well resolved because of the deficiency of the evidence and genetic variation information [86,87]. Our objective phylogenetic relationships within the *S. incisa* complex were reconstructed in the framework of Scrophulariaceae based on the complete chloroplast genomes. The phylogenomic analysis of Scrophulariaceae, based on current species coverage of 26 cp genomes representing four tribes, provided full support for the monophyly of the three groups of tribe Scrophularieae, tribe Buddlejaceae and tribe Myoporeae + Leucophylleae (Figure 9). Meanwhile the tree also strongly supported sister relationships of the monophyly tribe Myoporeae with *Leucophyllum frutescens* from tribe Leucophylleae, and *Buddleja* with *Verbascum* + *Scrophularia*. Monophyly of the genus *Verbascum* with respect to the genus *Scrophularia* was sufficiently supported, which was consistent with available results [5,12,14].

Our present study contributes a phylogenetic framework to the chloroplast genome evolution of Scrophulariaceae, especially for the *S. incisa* complex within genus *Scrophularia*. As we obtained the results here by the ML and BI analyses, *Scrophularia* can be divided into two monophyletic groups: *S. sect. Scrophularia* and *S. sect. Caninae* (Figure 9). Within the *S. sect. Caninae*, *S. integrifolia* (Tajikistan H8) from the west of the Tianshan Mountains in Tajikistan was at the basal position, followed by an *S. kiriloviana* individual (Xinjiang AK1) from the eastern foothills of the Pamirs Plateau in Southwest Xinjiang, which was consistent with our previous phylogeographic study of the *S. incisa* complex based on *psbA-trnH*, *trnL-trnF* and *trnQ-rps16* markers [17,21]. Within the core *S. incisa* complex, a topology of (*S. dentata* + (*S. incisa* + *S. kiriloviana*)) was strongly supported. This result solves the polytomy of the *S. incisa* complex in the previous study [17,21]. In addition, although *S. kiriloviana* from Xinjiang does not form a single monophyletic cluster, maybe due to introgression and/or incomplete lineage sorting between *S. kiriloviana* AK1 and neighboring *S. integrifolia*, the *S. incisa* complex is subdivided into three distinct gene pools and all the *S. kiriloviana* individuals show dominant red genetic clusters from the Xinjiang gene pool based on previous nSSR population data [17,21]. Therefore, we suggest that there are three recognized species in the *S. incisa* complex. In combination with our previous ancestral distribution reconstruction analysis and species distribution range of *S. sect. Caninae* [5,9,17,21], the robust phylogenetic evidence here firmly support what we have been suggesting about the Central Asian origin of the *S. incisa* complex with subsequent diversification on the Qinghai–Tibet Plateau (QTP), the Tianshan Mountains, the Mongolian Plateau and surrounding regions. This pattern has been explored across various genera, such as *Solms-laubachia*, *Incarvillea* and *Myricaria* [88–91]. However, multi-marker time-calibrated phylogeny based on a more comprehensive sampling and both the

chloroplast and nuclear genomes should be built to investigate this biogeographic scenario more rigorously.

5. Conclusions

The present study was designed to improve the phylogenetic resolution within the desert and steppe subshrubs *S. incisa* complex by comparing the complete chloroplast genome sequences of the main representatives covering all the lineages within the *S. incisa* complex. This analysis revealed that their chloroplast genomes are highly conserved in terms of length, genomic structure, gene number and arrangement. Close relationships between these three taxa were inferred and confirmed depending on high chloroplast genome similarities and the topology of phylogenetic trees in which *S. incisa*, *S. kiriloviana* and *S. dentata* form a strong supported distinct cluster within the *S. sect. Caninae* clade of genus *Scrophularia*. Recognition of highly variable regions and distinct microsatellite loci patterns in the chloroplast genomes of the *S. incisa* complex and three other species of *Scrophularia* could potentially be used in the future for identification and discrimination of these taxa. These analyses of chloroplast genomes consequently contributed to our understanding of the plastid phylogeny of the *S. incisa* complex and provide valued implications for the evolutionary history and conservation of these medicinal plants, as well as the genus *Scrophularia* in general.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13101691/s1>, Table S1: The list of 26 chloroplast genomes from Scrophulariaceae in phylogenetic analyses. Table S2: Summary of average relative synonymous codon usage (RSCU) of the codon usage in the *Scrophularia* cp genomes. Table S3: Nucleotide variability (P_i) values and total number of mutation (E_{ta}) in *Scrophularia*; Table S4: Repeat types and lengths of *Scrophularia* chloroplast genomes; Table S5: Analyses of repeat sequences in eleven *Scrophularia* chloroplast genomes; Table S6: Simple sequence repeat (SSR) polymorphism in eleven *Scrophularia* chloroplast genomes. Table S7: The K_a/K_s of the 80 protein-coding genes of the eleven *Scrophularia* individuals.

Author Contributions: R.W. conceived the idea. P.L., Z.Y. and C.F. contributed to the sampling. J.G. analyzed the data and wrote drafts of the manuscript. J.F. analyzed the data. R.W. helped improve the final manuscript. Z.Q. and P.L. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Nature Science Foundation of China [31600183], the Natural Science Foundation of Zhejiang Province [LGH22H280005, LY21C030008], the open project of Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin [BRZD1303], the Science Foundation of Zhejiang Sci-Tech University [15042170-Y], the Fundamental Research Funds of Zhejiang Sci-Tech University [2020Q031].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data contained within the article are openly available in GeneBank.

Acknowledgments: We thank Q. Wang for help with the submission of the chloroplast genome to NCBI, and L. Guo for help in organizing the format of the references.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Hong, D.-Y. The Distribution of Scrophulariaceae in the Holarctic with Special Reference to the Floristic Relationships Between Eastern Asia and Eastern North America. *Ann. Mo. Bot. Gard.* **1983**, *70*, 701. [\[CrossRef\]](#)
- Willis, J.C. A Dictionary of the Flowering Plants and Ferns. *Q. Rev. Biol.* **1974**, *49*, 436. [\[CrossRef\]](#)
- Mabberley, D.J. *The Plant-Book: A Portable Dictionary of the Vascular Plants*; Cambridge University Press: Cambridge, UK, 1997; p. 169.
- Fischer, E. Scrophulariaceae. In *The Families and Genera of Vascular Plants*; Kadereit, J.W., Ed.; Springer: Berlin/Heidelberg, Germany, 2004; Volume 7, pp. 333–432.
- Scheunert, A.; Heubl, G. Against all odds: Reconstructing the evolutionary history of *Scrophularia* (Scrophulariaceae) despite high levels of incongruence and reticulate evolution. *Org. Divers. Evol.* **2017**, *17*, 323–349. [\[CrossRef\]](#)
- Tsoong, P.C.; Tang, Y.C. *Scrophularia*. In *Flora Reipublicae Popularis Sinicae, Tomus*; Tsoong, P.C., Yang, H.B., Eds.; Science Press: Beijing, China, 1979; Volume 67, pp. 46–85.
- Hong, D.Y.; Yang, H.B.; Jin, C.L.; Holmgren, N.H. Scrophulariaceae. In *Flora of China*; Wu, Z.-Y., Ed.; Missouri Botanical Garden: St. Louis, MI, USA, 1998; Volume 18, pp. 1–212.
- Stiefel, H. Systematische und Pflanzengeographische Studien zur Kenntnis der Gattung *Scrophularia*. *Bot. Jahrbücher Syst. Pflanzengesch. Pflanzengeogr.* **1910**, *44*, 406–496.
- Wang, R.H.; Xia, M.Q.; Tan, J.B.; Chen, C.; Jin, X.J.; Li, P.; Fu, C.X. A new species of *Scrophularia* (Scrophulariaceae) from Hubei, China. *Phytotaxa* **2018**, *350*, 1–14. [\[CrossRef\]](#)
- Scheunert, A.; Heubl, G. Phylogenetic relationships among New World *Scrophularia* L. (Scrophulariaceae): New insights inferred from DNA sequence data. *Plant Syst. Evol.* **2011**, *291*, 69–89. [\[CrossRef\]](#)
- Faride, A.; Mehrshid, R.; Fatemeh, D.; Fatemeh, A.; Fatemeh, A. Preliminary molecular phylogeny of Eurasian *Scrophularia* (Scrophulariaceae) based on DNA sequence data from *trnS-trnG* and ITS regions. *G. Bot. Ital.* **2011**, *145*, 857–865. [\[CrossRef\]](#)
- Navarro-Pérez, M.L.; López, L.; Fernández-Mazuecos, M.; Rodríguez-Riaño, T.; Ortega-Olivencia, A. The role of birds and insects in pollination shifts of *Scrophularia* (Scrophulariaceae). *Mol. Phylogenet. Evol.* **2013**, *69*, 239–254. [\[CrossRef\]](#)
- Scheunert, A.; Heubl, G. Diversification of *Scrophularia* (Scrophulariaceae) in the Western Mediterranean and Macaronesia—Phylogenetic relationships, reticulate evolution and biogeographic patterns. *Mol. Phylogenet. Evol.* **2014**, *70*, 296–313. [\[CrossRef\]](#)
- Chen, C.; Cai, M.Q.; Xu, B.; Jin, X.J.; Wang, R.H.; Li, P.; Zhao, Y.P.; Fu, C.X. Systematic position of *Oreosolen* (tribe Scrophularieae, Scrophulariaceae) based on nuclear and plastid sequences. *J. Syst. Evol.* **2017**, *55*, 446–452. [\[CrossRef\]](#)
- Wang, R.H.; Wang, D.D.; Li, P.; Qi, Z.C.; Fu, C.X. *Scrophularia koraiensis*, a new synonym to *Scrophularia kakudensis* (Lamiales: Scrophulariaceae). *Phytotaxa* **2015**, *202*, 228–230. [\[CrossRef\]](#)
- Wang, R.H.; Chen, C.; Ma, Q.; Li, P.; Fu, C.X. Development of microsatellite loci in *Scrophularia incisa* (Scrophulariaceae) and cross-amplification in congeneric species. *Appl. Plant Sci.* **2014**, *2*, 1300077. [\[CrossRef\]](#)
- Wang, R.H.; Yang, Z.P.; Zhang, Z.C.; Comes, H.P.; Qi, Z.C.; Li, P.; Fu, C.X. Plio-Pleistocene climatic change drives allopatric speciation and population divergence within *Scrophularia incisa* complex (Scrophulariaceae) of the desert and steppe subshrubs in Northwest China. *Front. Plant Sci.* **2022**, *13*, 985372. [\[CrossRef\]](#)
- Zhang, L.Q.; Yang, Z.; Jia, Q.; Dorje, G.; Zhao, Z.; Guo, F.J.; Li, Y.M. Two new phenylpropanoid glycosides with interesterification from *Scrophularia dentata* Royle ex Benth. *J. Mol. Struct.* **2013**, *1049*, 299–302. [\[CrossRef\]](#)
- Zhang, L.Q.; Zhu, T.T.; Qian, F.; Xu, J.W.; Dorje, G.; Zhao, Z.L.; Guo, F.J.; Li, Y.M. Iridoid glycosides isolated from *Scrophularia dentata* Royle ex Benth. and their anti-inflammatory activity. *Fitoterapia* **2014**, *98*, 84–90. [\[CrossRef\]](#)
- Ullah, F.; Gao, Y.; Sari, I.; Jiao, R.F.; Saqib, S.; Gao, X.F. Macro-Morphological and Ecological Variation in *Rosa sericea* Complex. *Agronomy* **2022**, *12*, 1078. [\[CrossRef\]](#)
- Wang, R.H. *Phylogeny and Biogeography of Scrophularia and Phylogeography of S. incisa Complex*; Zhejiang University: Hangzhou, China, 2015. [\[CrossRef\]](#)
- Liu, L.X.; Li, R.; Worth, J.R.P.; Li, X.; Li, P.; Cameron, K.M.; Fu, C.X. The Complete Chloroplast Genome of Chinese Bayberry (*Morella rubra*, Myricaceae): Implications for Understanding the Evolution of Fagales. *Front. Plant Sci.* **2017**, *8*, 968. [\[CrossRef\]](#)
- Ran, H.; Liu, Y.Y.; Wu, C.; Cao, Y.N. Phylogenetic and comparative analyses of complete chloroplast genomes of Chinese *Viburnum* and *Sambucus* (Adoxaceae). *Plants* **2020**, *9*, 1143. [\[CrossRef\]](#)
- Burke, S.V.; Grennan, C.P.; Duvall, M.R. Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—Extend phylogenomic understanding of Bambusoideae. *Am. J. Bot.* **2012**, *99*, 1951–1961. [\[CrossRef\]](#)
- Sokolowska, J.; Fuchs, H.; Celinski, K. New insight into taxonomy of European Mountain Pines, *Pinus mugo* complex, based on complete chloroplast genomes sequencing. *Plants* **2021**, *10*, 1331. [\[CrossRef\]](#)
- Zhang, Y.; Wang, Z.; Guo, Y.; Chen, S.; Wang, R. Complete chloroplast genomes of *Leptodermis scabrida* complex: Comparative genomic analyses and phylogenetic relationships. *Gene* **2021**, *791*, 145715. [\[CrossRef\]](#)
- Lu, R.S.; Li, P.; Qiu, Y.X. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: Comparative genomic and phylogenetic analyses. *Front. Plant Sci.* **2017**, *7*, 2054. [\[CrossRef\]](#)
- Li, P.; Lu, R.S.; Xu, W.Q.; Ohi-Toma, T.; Cai, M.Q.; Qiu, Y.X.; Cameron, K.M.; Fu, C.X. Comparative Genomics and Phylogenomics of East Asian Tulips (*Amana*, Liliaceae). *Front. Plant Sci.* **2017**, *8*, 451. [\[CrossRef\]](#)
- Ye, W.Q.; Yap, Z.Y.; Li, P.; Comes, H.P.; Qiu, Y.X. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). *Mol. Phylogenet. Evol.* **2018**, *127*, 978–987. [\[CrossRef\]](#)

30. 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. [\[CrossRef\]](#)
31. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [\[CrossRef\]](#)
32. Feng, J.Y.; Wu, Y.Z.; Wang, R.R.; Xiao, X.F.; Wang, R.H.; Qi, Z.C.; Yan, X.L. The complete chloroplast genome of balsam aster (*Aster ageratoides* Turcz. var. *scaberulus* (Miq.) Ling., Asteraceae). *Mitochondrial DNA Part B* **2021**, *6*, 2464–2465. [\[CrossRef\]](#)
33. Zhang, C.J.; Xia, P.G.; Wu, R.; Mans, D. The complete chloroplast genome of *Scutellaria meehanioides* (Lamiaceae) from Shaanxi Province, China. *Mitochondrial DNA Part B* **2021**, *6*, 1685–1686. [\[CrossRef\]](#)
34. Zhang, M.; Chen, M.M.; Zhang, X.M.; Chen, S.N.; Liang, Z.S. The complete chloroplast genome sequence of traditional Chinese medicine *Uncaria macrophylla* (Rubiaceae). *Mitochondrial DNA B Resour.* **2022**, *7*, 694–695. [\[CrossRef\]](#)
35. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, 686–689. [\[CrossRef\]](#)
36. Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, 575–581. [\[CrossRef\]](#)
37. Sharp, P.M.; Li, W.H. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [\[CrossRef\]](#)
38. Liu, X.S. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virol. J.* **2012**, *9*, 68. [\[CrossRef\]](#)
39. 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. [\[CrossRef\]](#)
40. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [\[CrossRef\]](#)
41. Kurtz, S.; Schleiermacher, C. REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* **1999**, *15*, 426–427. [\[CrossRef\]](#)
42. Thiel, T.; Michalek, W.; Varshney, R.; 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*, 411–422. [\[CrossRef\]](#)
43. Ni, L.H.; Zhao, Z.L.; Dorje, G.; Ma, M. The Complete Chloroplast Genome of Ye-Xing-Ba (*Scrophularia dentata*; Scrophulariaceae), an Alpine Tibetan Herb. *PLoS ONE* **2016**, *11*, e0158488. [\[CrossRef\]](#)
44. Cavalier-Smith, T. Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr. Biol.* **2002**, *12*, R62–R64. [\[CrossRef\]](#)
45. Kim, P.M.; Korb, J.O.; Gerstein, M.B. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20274–20279. [\[CrossRef\]](#)
46. Alvarez-Ponce, D.; Aguade, M.; Rozas, J. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. *Genome Res.* **2008**, *19*, 234–242. [\[CrossRef\]](#)
47. Kosiol, C.; Vinař, T.; Fonseca, R.R.; Hubisz, M.J.; Bustamante, C.D.; Nielsen, R.; Siepel, A. Patterns of positive selection in six Mammalian genomes. *PLoS Genet.* **2008**, *4*, e1000144. [\[CrossRef\]](#)
48. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
49. Yang, Z.; Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [\[CrossRef\]](#)
50. Xu, W.Q.; Losh, J.; Chen, C.; Li, P.; Wang, R.H.; Zhao, Y.P.; Qiu, Y.X.; Fu, C.X. Comparative genomics of figworts (*Scrophularia*, Scrophulariaceae), with implications for the evolution of *Scrophularia* and Lamiales. *J. Syst. Evol.* **2019**, *57*, 55–65. [\[CrossRef\]](#)
51. Bi, Y.Q.; Deng, P.; Liu, L.X. The complete chloroplast genome sequence of purple mullein (*Verbascum phoeniceum* L.). *Mitochondrial DNA Part B* **2020**, *5*, 819–820. [\[CrossRef\]](#)
52. Fowler, R.M.; Mclay, T.G.B.; Schuster, T.M.; Buirchell, B.J.; Murphy, D.J.; Bayly, M.J. Plastid phylogenomic analysis of tribe Myoporeae (Scrophulariaceae). *Plant Syst. Evol.* **2020**, *306*, 52. [\[CrossRef\]](#)
53. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [\[CrossRef\]](#)
54. Alexandros, S. RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [\[CrossRef\]](#)
55. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [\[CrossRef\]](#)
56. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Li, X.; Li, Y.; Zang, M.; Li, M.; Fang, Y. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. *Int. J. Mol. Sci.* **2018**, *19*, 2443. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Lin, M.; Qi, X.; Chen, J.; Sun, L.; Zhong, Y.; Fang, J.; Hu, C. The complete chloroplast genome sequence of *Actinidia arguta* using the PacBio RS II platform. *PLoS ONE* **2018**, *13*, e0197393. [\[CrossRef\]](#)

60. Timme, R.E.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **2007**, *94*, 302–312. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Yang, J.B.; Yang, S.X.; Li, H.T.; Yang, J.; Li, D.Z. Comparative Chloroplast Genomes of *Camellia* Species. *PLoS ONE* **2013**, *8*, e73053. [\[CrossRef\]](#)
62. Song, Y.; Yao, X.; Tan, Y.H.; Gan, Y.; Corlett, R.T. Complete chloroplast genome sequence of the avocado: Gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae. *Can. J. For. Res.* **2016**, *46*, 1293–1301. [\[CrossRef\]](#)
63. Nie, X.; Lv, S.; Zhang, Y.; Du, X.; Wang, L.; Biradar, S.S.; Tan, X.; Wan, F.; Weining, S. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* **2012**, *7*, e36869. [\[CrossRef\]](#)
64. Sithichoke, T.; Pichahpuk, U.; Duangjai, S.; Juntima, C.; Thippawan, Y.; Nukoon, J.; Somvong, T. Characterization of the complete chloroplast genome of *Hevea brasiliensis* reveals genome rearrangement, RNA editing sites and phylogenetic relationships. *Gene* **2011**, *475*, 104–112. [\[CrossRef\]](#)
65. Yi, D.K.; Kim, K.J. The two complete plastomes from *Scrophularia* (Scrophulariaceae): *Scrophularia buergeriana* and *S. takesimensis*. *Mitochondrial DNA Part B* **2016**, *1*, 710–712. [\[CrossRef\]](#)
66. Nazareno, A.G.; Carlsen, M.; Lohmann, L.G. Complete Chloroplast Genome of *Tanaecium tetragonolobum*: The First Bignoniaceae Plastome. *PLoS ONE* **2015**, *10*, e0129930. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Zhang, Y.; Du, L.; Liu, A.; Chen, J.J.; Wu, L.; Hu, W.M.; Zhang, W.; Kim, K.; Lee, S.C.; Yang, T.J.; et al. The Complete Chloroplast Genome Sequences of Five Epi medium Species: Lights into Phylogenetic and Taxonomic Analyses. *Front. Plant Sci.* **2016**, *7*, 306. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Bodin, S.S.; Kim, J.S.; Kim, J.H. Complete Chloroplast Genome of *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae): Comparative Genomics and Evaluation of Universal Primers for Liliales. *Plant Mol. Biol. Rep.* **2013**, *31*, 1407–1421. [\[CrossRef\]](#)
69. Zhao, Y.; Yin, J.; Guo, H.; Zhang, Y.; Xiao, W.; Sun, C.; Wu, J.; Qu, X.; Yu, J.; Wang, X. The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Front. Plant Sci.* **2015**, *5*, 696. [\[CrossRef\]](#)
70. Mao, K.S.; Wang, Y.; Liu, J.Q. Evolutionary origin of species diversity on the Qinghai-Tibet Plateau. *J. Syst. Evol.* **2021**, *59*, 1142–1158. [\[CrossRef\]](#)
71. Spicer, R.A. Tibet, the Himalaya, Asian monsoons and biodiversity—in what ways are they related? *Plant Divers.* **2017**, *39*, 5–16. [\[CrossRef\]](#)
72. Miao, Y.F.; Herrmann, M.; Wu, F.L.; Yan, X.L.; Yang, S.L. What controlled Mid-Late Miocene long-term aridification in Central Asia—Global cooling or Tibetan Plateau uplift: A review. *Earth Sci. Rev.* **2012**, *112*, 155–172. [\[CrossRef\]](#)
73. Xu, L.; Cao, M.T.; Wang, Q.C.; Xu, J.H.; Liu, C.L.; Ullah, N.; Li, J.J.; Hou, Z.N.; Liang, Z.S.; Zhou, W.J.; et al. Insights into the plateau adaptation of *Salvia castanea* by comparative genomic and WGCNA analyses. *J. Adv. Res.* **2022**. [\[CrossRef\]](#)
74. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; Henry, R.J., Ed.; CABI Publishing: Wallingford, UK, 2005; pp. 45–68.
75. Szczecińska, M.; Sawicki, J. Genomic Resources of Three *Pulsatilla* Species Reveal Evolutionary Hotspots, Species-Specific Sites and Variable Plastid Structure in the Family Ranunculaceae. *Int. J. Mol. Sci.* **2015**, *16*, 22258–22279. [\[CrossRef\]](#)
76. Ma, J.; Yang, B.; Zhu, W.; Sun, L.; Tian, J.; Wang, X. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* **2013**, *528*, 120–131. [\[CrossRef\]](#)
77. Davis, J.I.; Soreng, R.J. Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae). *Am. J. Bot.* **2010**, *97*, 874–892. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Hansen, D.R.; Dastidar, S.G.; Cai, Z.; Penaflor, C.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol. Phylogenet. Evol.* **2007**, *45*, 547–563. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Huang, H.; Shi, C.; Liu, Y.; Mao, S.Y.; Gao, L.Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **2014**, *14*, 151. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Kim, K.; Lee, H. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [\[CrossRef\]](#)
81. Wang, R.J.; Cheng, C.L.; Chang, C.C.; Wu, C.L.; Su, T.M.; Chaw, S.M. Dynamics and evolution of the inverted repeat/large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **2008**, *8*, 36. [\[CrossRef\]](#)
82. Khakhlova, O.; Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **2010**, *46*, 85–94. [\[CrossRef\]](#)
83. Abbott, R.J.; Smith, L.C.; Milne, R.I.; Crawford, R.M.M.; Wolff, K.; Balfour, J. Molecular analysis of plant migration and refugia in the Arctic. *Science* **2000**, *289*, 1343–1346. [\[CrossRef\]](#)
84. Avise, J.C. *Phylogeography: The History and Formation of Species*; Harvard University Press: London, UK, 2000.
85. Zaman, W.; Ye, J.; Ahmad, M.; Saqib, S. Phylogenetic exploration of traditional Chinese medicinal plants: A case study on Lamiaceae. *Pak. J. Bot.* **2022**, *54*, 1033–1040. [\[CrossRef\]](#)
86. Oxelman, B.; Kornhall, P.; Olmstead, R.G.; Bremer, B. Further disintegration of Scrophulariaceae. *Taxon* **2005**, *54*, 411–425. [\[CrossRef\]](#)

87. Tank, D.C.; Beardsley, P.M.; Kelchner, S.A.; Olmstead, R.G. Review of the systematics of Scrophulariaceae s. l. and their current disposition. *Austral. Syst. Bot.* **2006**, *19*, 289–307. [[CrossRef](#)]
88. Chen, S.; Guan, K.; Zhou, Z.; Olmstead, R.; Cronk, Q. Molecular phylogeny of *Incarvillea* (Bignoniaceae) based on ITS and *trnL-F* sequences. *Am. J. Bot.* **2005**, *92*, 625–633. [[CrossRef](#)] [[PubMed](#)]
89. Wang, Y.; Liu, Y.F.; Liu, S.B.; Huang, H.W. Molecular phylogeny of *Myricaria* (Tamaricaceae): Implications for taxonomy and conservation in China. *Bot. Aquat.* **2009**, *50*, 343–352. [[CrossRef](#)]
90. Wen, J.; Zhang, J.Q.; Nie, Z.L.; Zhong, Y.; Sun, H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* **2014**, *5*, 4. [[CrossRef](#)]
91. Yue, J.P.; Sun, H.; David, B.A.; Li, J.H.; Al-Shehbaz, I.A.; Ree, R. Molecular phylogeny of *Solms-laubachia* (Brassicaceae) s.l., based on multiple nuclear and plastid DNA sequences, and its biogeographic implications. *J. Syst. Evol.* **2009**, *47*, 402–415. [[CrossRef](#)]