

RESEARCH ARTICLE

The Complete Chloroplast Genome of Ye-Xing-Ba (*Scrophularia dentata*; Scrophulariaceae), an Alpine Tibetan Herb

Lianghong Ni¹, Zhili Zhao^{1*}, Gaawe Dorje², Mi Ma²

1 School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai, China, **2** Tibetan Traditional Medical College, Lhasa, China

* zhilzhao@163.com



OPEN ACCESS

Citation: Ni L, Zhao Z, Dorje G, Ma M (2016) The Complete Chloroplast Genome of Ye-Xing-Ba (*Scrophularia dentata*; Scrophulariaceae), an Alpine Tibetan Herb. PLoS ONE 11(7): e0158488. doi:10.1371/journal.pone.0158488

Editor: Shilin Chen, Chinese Academy of Medical Sciences, Peking Union Medical College, CHINA

Received: January 2, 2016

Accepted: June 16, 2016

Published: July 8, 2016

Copyright: © 2016 Ni et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by grants from the National Natural Science Foundation of China (No. 81173654). Professor Zhili Zhao received the funding in 2012 (<https://isisn.nsf.gov.cn/egrantindex/fundindex/prjsearch-list>). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Scrophularia dentata is an important Tibetan medicinal plant and traditionally used for the treatment of exanthema and fever in Traditional Tibetan Medicine (TTM). However, there is little sequence and genomic information available for *S. dentata*. In this paper, we report the complete chloroplast genome sequence of *S. dentata* and it is the first sequenced member of the Sect. Tomiophyllum within *Scrophularia* (Scrophulariaceae). The gene order and organization of the chloroplast genome of *S. dentata* are similar to other Lamiales chloroplast genomes. The plastome is 152,553 bp in length and includes a pair of inverted repeats (IRs) of 25,523 bp that separate a large single copy (LSC) region of 84,058 bp and a small single copy (SSC) region of 17,449 bp. It has 38.0% GC content and includes 114 unique genes, of which 80 are protein-coding, 30 are transfer RNA, and 4 are ribosomal RNA. Also, it contains 21 forward repeats, 19 palindrome repeats and 41 simple sequence repeats (SSRs). The repeats and SSRs within *S. dentata* were compared with those of *S. takesimensis* and present certain discrepancies. The chloroplast genome of *S. dentata* was compared with other five publicly available Lamiales genomes from different families. All the coding regions and non-coding regions (introns and intergenic spacers) within the six chloroplast genomes have been extracted and analysed. Furthermore, the genome divergent hotspot regions were identified. Our studies could provide basic data for the alpine medicinal species conservation and molecular phylogenetic researches of Scrophulariaceae and Lamiales.

Introduction

Ye-Xing-Ba (Tibetan name), a common Tibetan herb, is traditionally used for the treatment of exanthema and fever in Traditional Tibetan Medicine (TTM). As a medicinal alpine species [1], *Scrophularia dentata* Royle ex Benth. (Scrophulariaceae) is one of the original plants of the Tibetan herb [2]. In recent years, our research team has been concentrating on the ethnobotanical survey of Tibetan herbs since 1992 [3, 4]. Based on our specimen collection and taxonomic identification, some chemical and pharmacological studies on *S. dentata* have been conducted by our team [5, 6, 7]. Because of the harsh living environment and increasing demand for medicinal use of the alpine plant, it is of utmost urgency to develop a conservation strategy for

Abbreviations: cp, chloroplast; LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat; bp, base pair(s); rRNA, ribosomal RNA; tRNA, transfer RNA; IGS, intergenic spacer; CDS, coding sequence; SSR, simple sequence repeat.

the species. Also, selection of excellent germplasm requires more efficient genetic data and molecular markers [8]. However, there is little sequence and genomic information available for the species in the GenBank, except for nrDNA ITS sequences, some chloroplast and mitochondrion segment sequences submitted by our team (<http://www.ncbi.nlm.nih.gov/nuccore/?term=Scrophularia+dentata>).

Chloroplast contains its own independent genome encoding a specific set of proteins [9]. The plastid genomes typically are composed of a large single copy (LSC) region and a small single copy (SSC) region, which are separated by two copies of inverted repeats (IRs) [10]. The small size of the chloroplast genome, about 115 to 165 kb in length, makes it suitable for complete sequencing and the data can be further applied to DNA bar coding and phylogeny construction [11, 12]. As complete chloroplast genome sequences contain sufficient information, sequencing of whole chloroplast genomes is essential for the analyses of plant species [13, 14].

S. dentata belongs to Scrophulariaceae, which is a large family of Lamiales, consisting of over 3,000 species [15]. Also, the genus *Scrophularia* has ca. 200 species and is divided into two sections, i.e. Sect. *Scrophularia* and Sect. *Tomiophyllum* [16]. However, the complete chloroplast genome of *Scrophularia takesimensis* within Sect. *Scrophularia* is the first published whole sequence in *Scrophularia*. This is the first sequenced member of the Sect. *Tomiophyllum* in the paper. Details of the *S. dentata* chloroplast genome structure and organization are reported and compared with previously annotated chloroplast genomes of other Lamiales species. Our studies could expand understanding of the diversity of *Scrophularia* cp genomes and provide basic data for the alpine medicinal species conservation and molecular phylogenetic researches of Scrophulariaceae and Lamiales.

Materials and Methods

Plant material

Samples of *S. dentata* were collected in Lhasa, Tibet, China. The voucher specimens were deposited in the herbarium of Shanghai University of Traditional Chinese Medicine (field number: XZ201416). Also, the location of the specimens is not within any protected area.

DNA extraction, genome sequencing and validation

Total chloroplast DNA was extracted from 100 g of fresh leaves using a sucrose gradient centrifugation method improved by Li et al. [17]. The genome was sequenced and assembled on an Illumina MiSeq platform following Gogniashvili et al. [18]. The four junctions between the SSC/LSC and IRs were validated by PCR amplification and Sanger sequencing, and nine other fragments were selected to validate the genome sequence further (S1 Table) [19].

Genome annotation

The chloroplast genome was annotated using the annotator DOGMA [20], coupled with manual corrections. For all of the tRNA genes, the annotations were verified using tRNAscan-SE1.21 [21]. The genome map was drawn by GenomeVX [22].

Genome comparison

The complete cp genomes of *S. dentata* and five other Lamiales species were compared using the mVISTA program [23]. The boundaries between IR and SC regions of these species were also compared and analysed.

Repeat structure

The REPuter program [24] was used to identify repeats (forward, palindrome, complement and reverse sequences). The size and identity of the repeats were limited to no less than 30 bp and 90%, respectively, with Hamming distance equal to 3 [25, 26]. Meanwhile, simple sequence repeats (SSRs) were detected using MISA [27] by setting the minimum number of repeats to 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexanucleotides. The repeats and SSRs of *S. dentata* were compared with those of *S. takesimensis*, the only one complete chloroplast genome available in Scrophulariaceae.

Results and Discussion

Genome organization and features

The junction regions between IRs and SSC/LSC and nine additional regions were confirmed by PCR amplification and Sanger sequencing. We compared these sequences to the assembled genome and no mismatch or indel was observed, which validated the accuracy of genome sequencing and assembly. The chloroplast genome sequence of *S. dentata* has been submitted to GenBank (accession number: KT428154).

The complete chloroplast genome of *S. dentata* has a total length of 152,553 bp, with a pair of inverted repeats (IRs) of 25,523 bp that separate a large single copy (LSC) region of 84,058 bp and a small single copy (SSC) region of 17,449 bp (Fig 1). The total GC content is 38.0%, which is similar to the published asterids cp genomes [28, 29]. And the GC content is unevenly distributed in the genome. The GC content of IRs (43.1%) is higher than that of LSC and SSC region (36.0 and 32.29%). The high GC content in the IR regions is due to the increased presence of GC nucleotides in the four rRNA genes: *rrn5*, *rrn4.5*, *rrn23* and *rrn16*, which is congruent to what has been found in other chloroplast genomes [30, 31].

The chloroplast genome of *S. dentata* encodes a total of 114 unique genes, of which 18 are duplicated in IR regions. Among the 114 genes, there are 80 protein-coding genes (70.2%), 30 transfer genes (26.3%) and 4 rRNA genes (3.5%) (Table 1). 18 genes contain introns, 15 (nine protein-coding and six tRNA genes) of which contain one intron and three (*clpP*, *ycf3* and *rps12*) contain two introns (Table 2). The *rps12* gene is a trans-spliced gene, with the first exon located in LSC region and the other two exons duplicated in IR regions.

Protein-coding regions account for 59.6% of the whole genome, while tRNA and rRNA regions account for 4.6% and 5.9%, respectively. The remaining regions are noncoding sequences, including intergenic spacers, introns and pseudogenes. There are two pseudogenes identified: *ycf1* and *rps19*. They are located in the boundary regions between IRb/SSC and IRa/LSC, respectively. The lack of their protein-coding ability is due to the partial gene duplication.

Comparison to other Lamiales species

The chloroplast genome of *S. dentata* was compared with other five publicly available Lamiales genomes from different families, i.e., *Boea hygrometrica* (Gesneriaceae), *Hesperelaea palmeri* (Oleaceae), *Salvia miltiorrhiza* (Lamiaceae), *Andrographis paniculata* (Acanthaceae) and *Sesamum indicum* (Pedaliaceae) (Table 3). Organization of the Lamiales chloroplast genome is conserved; neither translocations nor inversions were detected in the analyses. However, there are differences in terms of genome size. The *A. paniculata* chloroplast genome is the shortest (150,249 bp), while that of *H. palmeri* (155,820 bp) is longer than other species. The genome size variation is attributed mainly to the difference in the length of the LSC region, similar to the features in Asteraceae chloroplast genomes [36]. Also, the average size of the six Lamiales chloroplast genomes is 152,794 bp.

Fig 1. Chloroplast genome map of *Scrophularia dentata*. Genes drawn inside the circle are transcribed clockwise, and those outside are counterclockwise. Genes belonging to different functional groups are shown in different colors. The asterisks indicate genes which contain intron (s).

doi:10.1371/journal.pone.0158488.g001

The overall sequence identity of the six Lamiales chloroplast genomes was plotted using the software mVISTA with *S. dentata* as reference (Fig 2). The results show that the chloroplast genomes within Lamiales are conservative, although divergent regions could be detected. As expected, the IR region is more conserved than the LSC and SSC regions in these species. Meanwhile, non-coding regions reveal a higher divergence than coding regions. In order to check divergent hotspot regions further, all the coding regions and non-coding regions (introns and intergenic spacers) within the six chloroplast genomes were extracted and analysed (S2 Table, Fig 3). The most divergent regions are localized in the intergenic spacers, including *ccsA-ndhD*, *rps16-trnQ-UUG*, *ndhJ-ndhK*, *ndhE-ndhG*, *ndhC-trnV-UAC*, *trnH-GUG-psbA* and *ndhG-ndhI*. These intergenic regions could be used in assessing phylogenetic relationships within Lamiales species. For the coding regions, the most divergent regions are *ycf1*, *matK*, *rpl22*, *rpl32*, *rps15* and *ccsA*. These genes are all located in single copy regions.

The expansion and contraction at the borders of IR/SC are common in chloroplast genomes, which is the main reason for the size variation within angiosperm chloroplast genomes [37]. The junctions of LSC/IRb/SSC/IRa of the six Lamiales chloroplast genomes were compared (Fig 4). Although the length of the IR regions of the six species is similar, from 25,141 bp to 25,712 bp, some differences in IR expansions and contractions were observed.

Table 1. Genes present within *Scrophularia dentata* chloroplast genome.

	Group of genes	Gene names
1	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psal</i> , <i>psaJ</i>
2	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> , <i>psbZ</i>
3	Cytochrome b/f complex	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
4	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
5	NADH dehydrogenase	<i>ndhA*</i> , <i>ndhB*(x2)</i> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
6	RubisCO large subunit	<i>rbcL</i>
7	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
8	Ribosomal proteins (SSU)	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7(x2)</i> , <i>rps8</i> , <i>rps11</i> , <i>rps12**(x2)</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> , <i>rps18</i> , <i>rps19</i>
9	Ribosomal proteins (LSU)	<i>rpl2*(x2)</i> , <i>rpl14</i> , <i>rpl16*</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23(x2)</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
10	Other genes	<i>clpP**</i> , <i>matK</i> , <i>accD</i> , <i>ccsA</i> , <i>infA</i> , <i>cemA</i>
11	Proteins of unknown function	<i>ycf1</i> , <i>ycf2(x2)</i> , <i>ycf3**</i> , <i>ycf4</i> , <i>ycf15(x2)</i>
12	Ribosomal RNAs	<i>rrn4.5(x2)</i> , <i>rrn5(x2)</i> , <i>rrn16(x2)</i> , <i>rrn23(x2)</i>
13	Transfer RNAs	<i>trnA-UGC*(x2)</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnG-UCC*</i> , <i>trnH-GUG</i> , <i>trnI-CAU(x2)</i> , <i>trnI-GAU*(x2)</i> , <i>trnK-UUU*</i> , <i>trnL-CAA(x2)</i> , <i>trnL-UAA*</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnM-CAU</i> , <i>trnN-GUU(x2)</i> , <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG(x2)</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC(x2)</i> , <i>trnV-UAC*</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>

*Gene containing one intron.

**Gene containing two introns.

(x2) Two gene copies in the IRs.

doi:10.1371/journal.pone.0158488.t001

Table 2. Genes with introns within *Scrophularia dentata* chloroplast genome and the length of exons and introns.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
<i>petB</i>	LSC	6	720	642		
<i>petD</i>	LSC	8	735	475		
<i>atpF</i>	LSC	144	723	411		
<i>ndhA</i>	SSC	552	1067	531		
<i>ndhB</i>	IR	777	679	756		
<i>rpoC1</i>	LSC	435	756	1626		
<i>rps12*</i>	LSC	114	-	232	537	26
<i>rps16</i>	LSC	42	848	225		
<i>rpl2</i>	IR	393	664	435		
<i>rpl16</i>	LSC	9	886	399		
<i>clpP</i>	LSC	69	727	291	626	228
<i>ycf3</i>	LSC	129	702	228	720	153
<i>trnA-UGC</i>	IR	38	813	35		
<i>trnG-UCC</i>	LSC	23	683	47		
<i>trnI-GAU</i>	IR	42	943	35		
<i>trnK-UUU</i>	LSC	37	2517	35		
<i>trnL-UAA</i>	LSC	37	488	50		
<i>trnV-UAC</i>	LSC	38	585	37		

* The *rps12* is a trans-spliced gene with the 5' end located in the LSC region and the duplicated 3' end in the IR regions.

doi:10.1371/journal.pone.0158488.t002

The *rps19* genes of *H. palmeri* and *A. paniculata* are located in the LSC region completely, and the IR region is expanded to *rps19* gene in the other four genomes, with *rps19* pseudogene created at the IRa/LSC border. In *S. dentata*, a short *rps19* pseudogene of 40 bp is created. The same pseudogene is 30 bp in *S. indicum*, 43 bp in *S. miltiorrhiza* and 37 bp in *B. hygrometrica*, respectively. The border between IRb and SSC extends into the *ycf1* genes, with *ycf1* pseudogenes created in all of the six species. The length of *ycf1* pseudogene is 910 bp in *S. dentata*, 1,012 bp in *S. indicum*, 1,091 bp in *H. palmeri*, 983 bp in *A. paniculata*, 1,056 bp in *S. miltiorrhiza* and 816 bp in *B. hygrometrica*, respectively. Overlaps are detected between the *ycf1* pseudogene and *ndhF* gene in *H. palmeri*, *A. paniculata*, *S. miltiorrhiza* and *B. hygrometrica*, with lengths of 25 bp, 40 bp, 32bp and 118 bp, respectively. The *trnH-GUG* genes are all located in the LSC region, with 4–57 bp apart from the LSC/IRa boundary.

Table 3. Comparison of the base composition of six Lamiales chloroplast genomes.

	<i>Scrophularia dentata</i>	<i>Boea hygrometrica</i>	<i>Hesperelaea palmeri</i>	<i>Salvia miltiorrhiza</i>	<i>Andrographis paniculata</i>	<i>Sesamum indicum</i>
LSC	84,058 bp	84,692 bp	86,616 bp	82,695 bp	82,459 bp	85,170 bp
SSC	17,449 bp	17,901 bp	17,780 bp	17,555 bp	17,190 bp	17,872 bp
IR	25,523 bp	25,450 bp	25,712 bp	25,539 bp	25,300 bp	25,141 bp
Total	152,553 bp	153,493 bp	155,820	151,328 bp	150,249 bp	153,324 bp
CG content	38.0%	37.6%	37.8%	36.2%	38.3%	38.2%
Coding gene	80	86	80	80	80	80
tRNA	30	29	30	30	30	30
rRNA	4	4	4	4	4	4
References	this study	[32]	[33]	[34]	NC_022451	[35]

doi:10.1371/journal.pone.0158488.t003

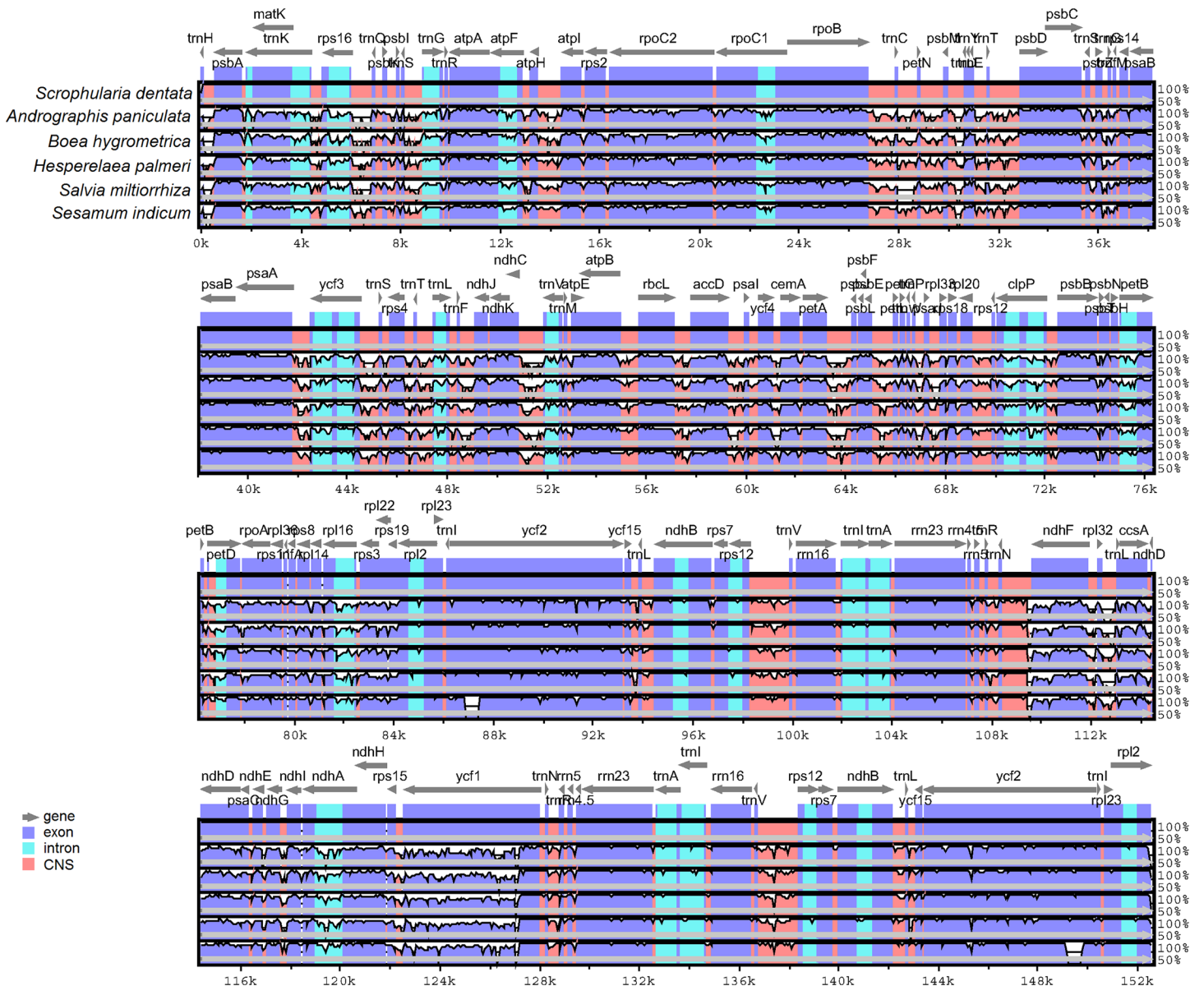


Fig 2. Sequence alignment of six chloroplast genomes in the Lamiales order using mVISTA program with *S. dentata* as reference. The top line shows genes in order (transcriptional direction indicated with arrow). A cut-off of 70% identity was used for the plots, and the Y-scale represents the percent identity between 50–100%. Genome regions are color-coded as exon, intron, and conserved non-coding sequences (CNS).

doi:10.1371/journal.pone.0158488.g002

Repeat and SSR analysis

Repeat sequences in the *S. dentata* chloroplast genome were analysed by REPuter and the results showed that 21 forward repeats and 19 palindrome repeats were at least 30 bp long per repeat unit with a sequence identity greater than 90% (Table 4). The complement repeats and reverse repeats had not been detected. Of all the repeats found, 33 repeats (82.5%) are 30 to 39 bp long, 7 repeats (17.5%) are 40 to 49 bp long, and the longest repeat is 44 bp. We present a comparative analysis of repeats between *S. dentata* and *S. takesimensis*, the only one complete chloroplast genome available in Scrophulariaceae [38] (S3 Table, Fig 5A and 5B). Most (92.5%)

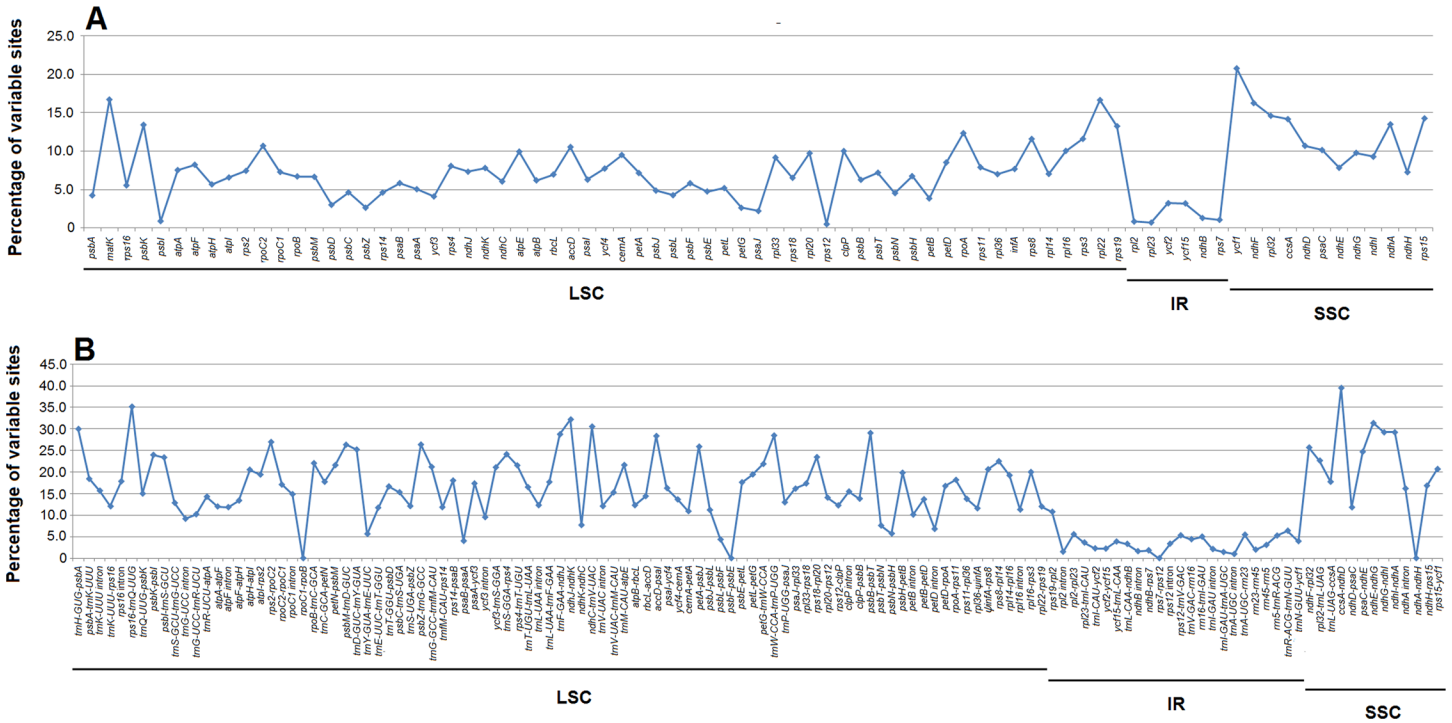


Fig 3. Percentages of variable sites in homologous regions across the six Lamiales chloroplast genomes. (A) coding regions; (B) introns and intergenic spacers (non-coding regions).

doi:10.1371/journal.pone.0158488.g003

of the repeats in *S. dentata* are conserved and could be found in *S. takesimensis*, and there are another 3 complement repeats and 2 reverse repeats found in *S. takesimensis*.

In general, repeats are mostly distributed in noncoding regions [39, 40]. However, most of the repeats (62.5%) in the *S. dentata* chloroplast genome are located in coding regions (CDS),

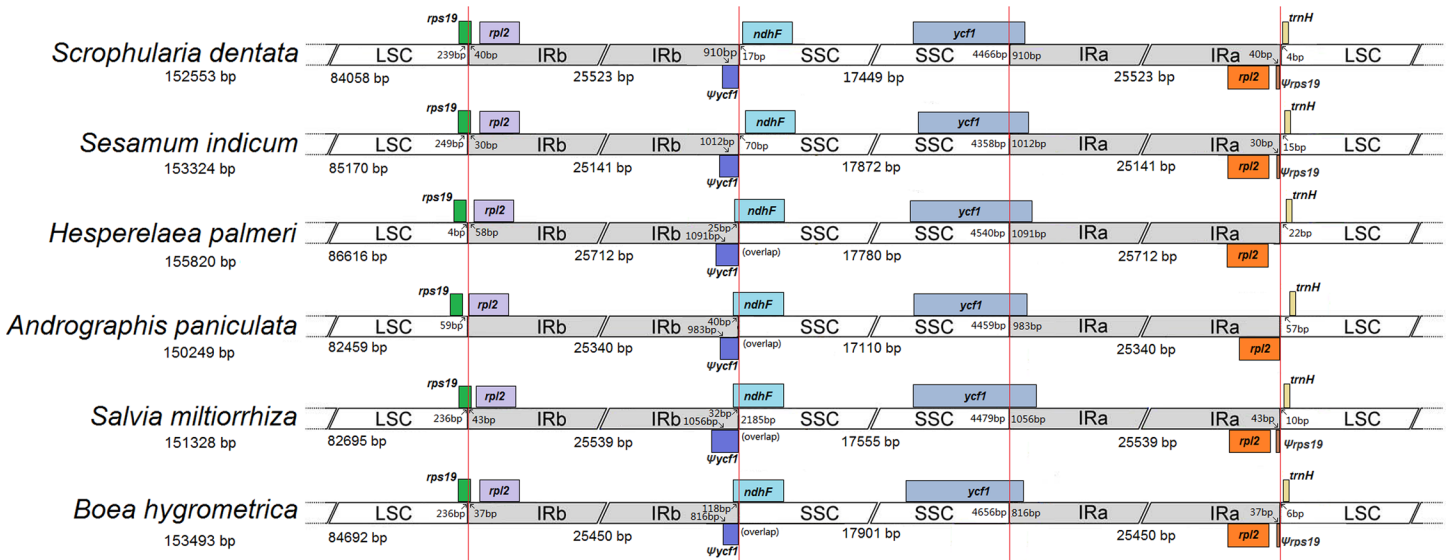


Fig 4. Comparison of the borders of LSC, SSC and IR regions in six Lamiales chloroplast genomes. Ψ indicates a pseudogene. This figure is not to scale.

doi:10.1371/journal.pone.0158488.g004

Table 4. Repeat sequences and their distribution within *Scrophularia dentata* chloroplast genome.

No.	Size(bp)	Type	Repeat 1 start (location)	Repeat 2 start (location)	Region
1	30	F	9597(<i>trnG-UCC* -trnR-UCU</i>)	36418(<i>trnG-GCC* -trnM-CAU</i>)	LSC
2	30	F	38641(<i>psaB</i>)	40865(<i>psaA</i>)	LSC
3	30	F	43669(<i>ycf3</i> intron1)	119061(<i>ndhA</i> intron)	LSC; SSC
4	30	F	88876(<i>ycf2</i>)	88918(<i>ycf2</i>)	IRb
5	30	F	91289(<i>ycf2</i>)	91343(<i>ycf2</i>)	IRb
6	30	F	91318(<i>ycf2</i>)	91336(<i>ycf2</i>)	IRb
7	30	F	145238(<i>ycf2</i>)	145292(<i>ycf2</i>)	IRa
8	30	F	145258(<i>ycf2</i>)	145294(<i>ycf2</i>)	IRa
9	30	F	147663(<i>ycf2</i>)	147705(<i>ycf2</i>)	IRa
10	30	F	43669(<i>ycf3</i> intron1)	98305(<i>rps12-trnV-GAC</i>)	LSC; IRb
11	31	F	12909(<i>atpF-atpH</i>)	69566(<i>rpl20-rps12</i>)	LSC
12	32	F	8053(<i>psbI-trnS-GCU*</i>)	35525(<i>psbC-trnS-UGA*</i>)	LSC
13	34	F	91308(<i>ycf2</i>)	91326(<i>ycf2</i>)	IRb
14	34	F	145251(<i>ycf2</i>)	145269(<i>ycf2</i>)	IRa
15	36	F	91288(<i>ycf2</i>)	91306(<i>ycf2</i>)	IRb
16	36	F	145269(<i>ycf2</i>)	145287(<i>ycf2</i>)	IRa
17	39	F	43657(<i>ycf3</i> intron1)	98293(<i>rps12-trnV-GAC</i>)	LSC; IRb
18	39	F	43657(<i>ycf3</i> intron1)	119049(<i>ndhA</i> intron)	LSC; SSC
19	40	F	91290(<i>ycf2</i>)	91326(<i>ycf2</i>)	IRb
20	40	F	145245(<i>ycf2</i>)	145281(<i>ycf2</i>)	IRa
21	41	F	98291(<i>trnR-UCU</i>)	119047(<i>ndhA</i> intron)	IRb; SSC
22	30	P	8055(<i>psbI-trnS-GCU*</i>)	45358(<i>trnS-GGA</i>)	LSC
23	30	P	43669(<i>ycf3</i> intron1)	138276(<i>trnV-GAC-rps12</i>)	LSC; IRa
24	30	P	75205(<i>petB</i> intron)	75205(<i>petB</i> intron)	LSC
25	30	P	35527(<i>psbC-trnS-UGA*</i>)	45358(<i>trnS-GGA</i>)	LSC
26	30	P	88876(<i>ycf2</i>)	147663(<i>ycf2</i>)	IRb; IRa
27	30	P	88918(<i>ycf2</i>)	147705(<i>ycf2</i>)	IRb; IRa
28	30	P	91289(<i>ycf2</i>)	145238(<i>ycf2</i>)	IRb; IRa
29	30	P	91318(<i>ycf2</i>)	145245(<i>ycf2</i>)	IRb; IRa
30	30	P	91336(<i>ycf2</i>)	145263(<i>ycf2</i>)	IRb; IRa
31	30	P	91343(<i>ycf2</i>)	145292(<i>ycf2</i>)	IRb; IRa
32	34	P	91308(<i>ycf2</i>)	145251(<i>ycf2</i>)	IRb; IRa
33	34	P	91326(<i>ycf2</i>)	145269(<i>ycf2</i>)	IRb; IRa
34	36	P	91288(<i>ycf2</i>)	145269(<i>ycf2</i>)	IRb; IRa
35	36	P	91306(<i>ycf2</i>)	145287(<i>ycf2</i>)	IRb; IRa
36	39	P	43657(<i>ycf3</i> intron1)	138279(<i>trnV-GAC-rps12</i>)	LSC; IRa
37	40	P	91290(<i>ycf2</i>)	145245(<i>ycf2</i>)	IRb; IRa
38	40	P	91326(<i>ycf2</i>)	145281(<i>ycf2</i>)	IRb; IRa
39	41	P	119047(<i>ndhA</i> intron)	138279(<i>trnV-GAC-rps12</i>)	SSC; IRa
40	44	P	74375(<i>psbT-psbN</i>)	74375(<i>psbT-psbN</i>)	LSC

* part in the gene.

Repeat sequences conserved in the *S. takesimensis* chloroplast genome are highlighted in bold.

F = forward, P = palindrome, IGS = intergenic spacer.

doi:10.1371/journal.pone.0158488.t004

mainly in *ycf2*; and it is similar to that of *S. takesimensis* (Fig 5B). Meanwhile, 30% repeats are located in intergenic spacers (IGS) and introns, and 7.5% repeats in parts of the IGS and CDS.

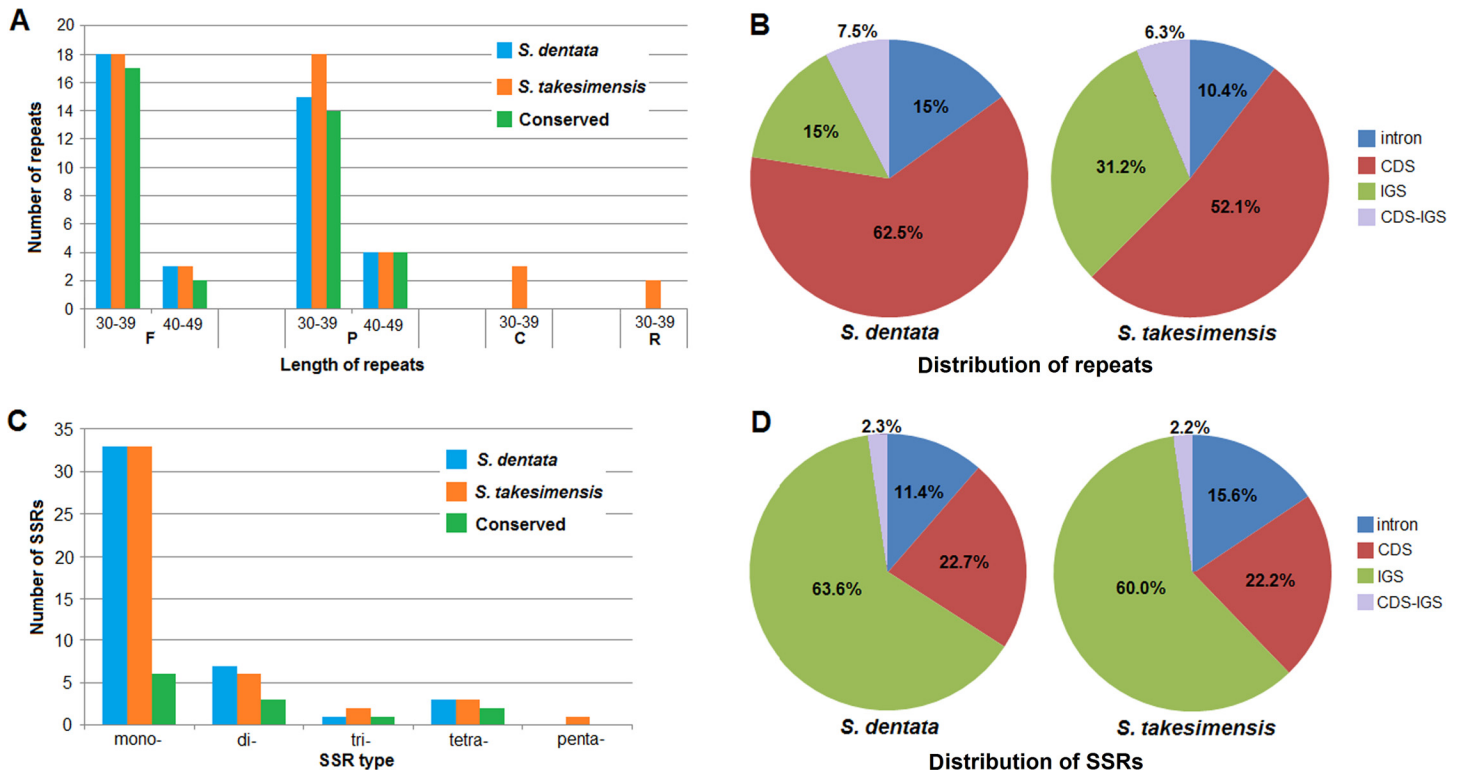


Fig 5. Repeat sequences and simple sequence repeats (SSRs) within *Scrophularia dentata* and *Scrophularia takesimensis* chloroplast genomes. (A) number of different types of repeats. F = forward repeats; P = palindrome repeats; C = complement repeats; R = reverse repeats. (B) distribution of repeats. (C) number of different types of SSRs. (D) distribution of SSRs. CDS = coding sequence; IGS = intergenic spacer; CDS-IGS = part in CDS and part in IGS.

doi:10.1371/journal.pone.0158488.g005

Also, simple sequence repeats (SSRs) exert significant influence on genome rearrangement and recombination [41]. A total of 44 SSRs were detected in the *S. dentata* chloroplast genome, accounting for 500 bp of the total sequence (ca.0.33%), and there were 33, 7, 1 and 3 mono-, di-, tri-, and tetra- nucleotide repeats, respectively (Table 5). No penta- or hexa- nucleotide repeats were found. Most of the SSRs are mononucleotide repeats. 43 SSRs (97.7%) are composed of A and T nucleotides, whereas only one is composed of “GTCT” repeat. The high content of AT in SSRs contributes to the AT richness of the chloroplast genome [42, 43]. Among the SSRs, 33 are located in IGS and introns, 10 are found in coding genes, including *ycf1*, *rpoC2*, *atpB*, *rpoA*, *atpA* and *ndhD*, and 1 is located in parts of the IGS and CDS. Compared with *S. takesimensis*, 12 SSRs were identity, 30 exhibited length polymorphisms and 2 had not been detected (S4 Table, Fig 5C and 5D). These repeat sequences may be useful in developing lineage-specific markers, which could be widely used in genetic diversity and evolutionary studies of *Scrophularia*.

Conclusions

In this paper, we reported the complete sequence of *S. dentata* cp genome, as the first whole cp genome in the Sect. Tomiophyllum of the genus *Scrophularia*. This cp genome sequence was compared to other five genomes from the Lamiales species (i.e., *B. hygrometrica*, *H. palmeri*, *S. miltiorrhiza*, *A. paniculata* and *S.indicum*). No significant structural changes were detected among the chloroplast genomes. All the coding regions and non-coding regions (introns and intergenic spacers) within the six chloroplast genomes were extracted and analysed. The most

Table 5. Simple sequence repeats within *Scrophularia dentata* chloroplast genome.

Unit	Length	No.	SSR start	SSR-containing region	Region		
A	16	1	44297	<i>ycf3</i> intron1	LSC		
	13	1	12158	<i>atpF</i> intron	LSC		
	12	2	8024	IGS (<i>psbI-trnS-GCU</i>)	LSC		
			16175	IGS (<i>rps2-rpoC2</i>)	LSC		
	11	1	75204	<i>petB</i> intron	LSC		
	10	8	22732	<i>rpoC1</i> intron	LSC		
			44557	IGS (<i>ycf3-trnS-GGA</i>)	LSC		
			59913	IGS (<i>accD-psaI</i>)	LSC		
			63314	IGS (<i>petA-psbJ</i>)	LSC		
			111914	IGS (<i>ndhF-rpl32</i>)	SSC		
			112458	IGS (<i>rpl32-trnL-UAG</i>)	SSC		
			137924	IGS (<i>trnV-GAC-rps12</i>)	IRa		
			152482	IGS (<i>rpl2-Ψrps19</i>)	IRa		
T	15	1	12921	IGS (<i>atpF-atpH</i>)	LSC		
	14	1	60099	IGS (<i>psaI-ycf4</i>)	LSC		
	13	1	69580	IGS (<i>rpl20-rps12</i>)	LSC		
	12	1	124137	<i>ycf1</i>	SSC		
	11	8	8632	IGS (<i>trnS-GCU-trnG-UCC</i>)	LSC		
			14323	IGS (<i>atpH-atpI</i>)	LSC		
			18378	<i>rpoC2</i>	LSC		
			57534	IGS (<i>rbcl-accD</i>)	LSC		
			80092	IGS (<i>infA-rps8</i>)	LSC		
			80606	IGS (<i>rps8-rpl14</i>)	LSC		
			112115	IGS (<i>ndhF-rpl32</i>)	SSC		
			124952	<i>ycf1</i>	SSC		
			10	8	52007	<i>trnV-UAC</i> intron	LSC
					54874	<i>atpB</i>	LSC
	54959	IGS (<i>atpB-rbcl</i>)			LSC		
	78182	<i>rpoA</i>			LSC		
	84121	IGS (<i>rps19-rpl2</i>)			IRb		
98679	IGS (<i>rps12-trnV-GAC</i>)	IRb					
124603	<i>ycf1</i>	SSC					
124880	<i>ycf1</i>	SSC					
AT	12	1	42432	IGS (<i>psaA-ycf3</i>)	LSC		
	10	2	19751	<i>rpoC2</i>	LSC		
			13822	IGS (<i>atpH-atpI</i>)	LSC		
TA	18	1	46524	IGS (<i>rps4-trnT-UGU</i>)	LSC		
	12	3	8212	IGS (<i>trnS-GCU-trnG-UCC</i>)	LSC		
			31456	IGS (<i>trnE-UUC-trnT-GGU</i>)	LSC		
			42433	IGS (<i>psaA-ycf3</i>)	LSC		
ATA	12	1	54928	IGS (<i>atpB*-rbcl</i>)	LSC		
GTCT	12	1	10993	<i>atpA</i>	LSC		
TTTA	12	1	49645	IGS (<i>ndhJ-ndhK</i>)	LSC		
AATA	12	1	114611	<i>ndhD</i>	SSC		

* part in the gene.

SSRs that are identical in the *S. takesimensis* chloroplast genome are highlighted in bold.

doi:10.1371/journal.pone.0158488.t005

divergent regions are localized in the intergenic spacers, including *ccsA-ndhD*, *rps16-trnQ-UUG*, *ndhJ-ndhK*, *ndhE-ndhG*, *ndhC-trnV-UAC*, *trnH-GUG-psbA* and *ndhG-ndhI*. For the coding regions, the most divergent regions are *ycf1*, *matK*, *rpl22*, *rpl32*, *rps15* and *ccsA*. These genes are all located in single copy regions. There were some differences in the genome size and IR expansion or contraction; and the divergent regions were analysed. Tandem repeats and SSRs within *S. dentata* were compared with those of *S. takesimensis*, which may provide markers for the analyses of infraspecific genetic differentiation within *Scrophularia*. In addition, our studies could provide basic data for the alpine medicinal species conservation and molecular phylogenetic researches of Scrophulariaceae and Lamiales.

Supporting Information

S1 Table. List of all pairs of primers used for assembly validation.

(DOC)

S2 Table. Percentages of variable sites across the six Lamiales chloroplast genomes.

(XLS)

S3 Table. Repeat sequences within *Scrophularia takesimensis* chloroplast genome.

(XLS)

S4 Table. Simple sequence repeats within *Scrophularia takesimensis* chloroplast genome.

(XLS)

Author Contributions

Conceived and designed the experiments: ZZ. Performed the experiments: LN. Analyzed the data: LN. Contributed reagents/materials/analysis tools: ZZ GD MM. Wrote the paper: LN ZZ.

References

1. Zhong BQ, Tang YC. *Scrophularia* L. In: Zhong BQ, Yang HB, editors. Flora Reipublicae Popularis Sinicae, Vol 67(2). Beijing, China: Science Press. 1979; pp. 46–51.
2. Wang WY, Lu SL, Liu SW, Yang YC, Wu ZL, He TN, et al. Ye-Xin-Ba. In: Yang YC, editor. Tibetan Medicines. Xining, China: Qinghai People's Publishing House. 1991; pp. 333–334.
3. Zhao ZL, Zhao RN. Original name of traditional Chinese medicinal herb Chuan Bu. Chinese Pharmaceutical Journal. 1992; 27(5): 269–270.
4. Zhao ZL, Gaawe D, Wang ZT. Identification of medicinal plants used as Tibetan Traditional Medicine Jie-Ji. Journal of Ethnopharmacology. 2010; 132(1): 122–126. doi: [10.1016/j.jep.2010.07.051](https://doi.org/10.1016/j.jep.2010.07.051) PMID: [20692329](https://pubmed.ncbi.nlm.nih.gov/20692329/)
5. Zhang L, Zhu T, Qian F, Xu J, Dorje G, Zhao Z, et al. Iridoid glycosides isolated from *Scrophularia dentata* Royle ex Benth. and their anti-inflammatory activity. Fitoterapia. 2014; 98: 84–90. doi: [10.1016/j.fitote.2014.07.005](https://doi.org/10.1016/j.fitote.2014.07.005) PMID: [25016952](https://pubmed.ncbi.nlm.nih.gov/25016952/)
6. Zhang L, Yang Z, Jia Q, Dorje G, Zhao Z, Guo F, et al. Two new phenylpropanoid glycosides with inter-esterification from *Scrophularia dentata* Royle ex Benth. Journal of Molecular Structure. 2013; 1049 (30): 299–302.
7. Zhang L, Zhang D, Jia Q, Wang R, Dorje G, Zhao Z, et al. 19(4→3)-abeo-abietane diterpenoids from *Scrophularia dentata* Royle ex Benth. Fitoterapia. 2015; 106: 72–77. doi: [10.1016/j.fitote.2015.08.005](https://doi.org/10.1016/j.fitote.2015.08.005) PMID: [26291646](https://pubmed.ncbi.nlm.nih.gov/26291646/)
8. Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA., et al. Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. PLoS ONE. 2013; 8(2): e57533. doi: [10.1371/journal.pone.0057533](https://doi.org/10.1371/journal.pone.0057533) PMID: [23460871](https://pubmed.ncbi.nlm.nih.gov/23460871/)
9. Olmstead RG, Palmer JD. Chloroplast DNA systematic: a review of methods and data analysis. American Journal of Botany. 1994; 81(9): 1205–1224.

10. Wicke S, Schneeweiss GM, Depamphilis CW, Kai FM, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*. 2011; 76(3–5): 273–97. doi: [10.1007/s11103-011-9762-4](https://doi.org/10.1007/s11103-011-9762-4) PMID: [21424877](https://pubmed.ncbi.nlm.nih.gov/21424877/)
11. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. *PLoS ONE*. 2011; 6(5): e19254. doi: [10.1371/journal.pone.0019254](https://doi.org/10.1371/journal.pone.0019254) PMID: [21637336](https://pubmed.ncbi.nlm.nih.gov/21637336/)
12. Rivas JDL, Lozano JJ, Ortiz AR. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Research*. 2002; 12(4): 567–583. PMID: [11932241](https://pubmed.ncbi.nlm.nih.gov/11932241/)
13. Williams AV, Boykin LM, Howell KA, Nevill PG, Small I. The Complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PLoS ONE*. 2015; 10(5): e0125768. doi: [10.1371/journal.pone.0125768](https://doi.org/10.1371/journal.pone.0125768) PMID: [25955637](https://pubmed.ncbi.nlm.nih.gov/25955637/)
14. Yap JY, Rohner T, Greenfield A, Van Der Merwe M, McPherson H, et al. Complete chloroplast genome of the Wollemi Pine (*Wollemia nobilis*): structure and evolution. *PLoS ONE*. 2015; 10(6): e0128126. doi: [10.1371/journal.pone.0128126](https://doi.org/10.1371/journal.pone.0128126) PMID: [26061691](https://pubmed.ncbi.nlm.nih.gov/26061691/)
15. Oxelman B, Bremer B. Further disintegration of Scrophulariaceae. *Taxon*. 2005; 54(2): 411–425.
16. Hong DY. The distribution of Scrophulariaceae in the Holarctic with special reference to the floristic relationships between eastern Asia and eastern North America. *Annals of the Missouri Botanical Garden*. 1983; 70(4): 701–712.
17. Li XW, Hu ZG, Lin XH, Li Q, Gao HH, Luo GA, et al. High-throughput pyrosequencing of the complete chloroplast genome of *Magnolia officinalis* and its application in species identification. *Acta Pharmaceutica Sinica*. 2012; 47(1): 124–130. PMID: [22493817](https://pubmed.ncbi.nlm.nih.gov/22493817/)
18. Gogniashvili M, Naskidashvili P, Bedoshvili D, Kotorashvili A, Kotaria N, Beridze T. Complete chloroplast DNA sequences of Zanduri wheat (*Triticum* spp.). *Genetic Resources and Crop Evolution*. 2015; 62: 1269–1277.
19. Ni L, Zhao Z, Xu H, Chen S, Dorje G. The complete chloroplast genome of *Gentiana straminea* (Gentianeaceae), an endemic species to the Sino-Himalayan subregion. *Gene*. 2015; 577(2): 281–288. doi: [10.1016/j.gene.2015.12.005](https://doi.org/10.1016/j.gene.2015.12.005) PMID: [26680100](https://pubmed.ncbi.nlm.nih.gov/26680100/)
20. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17): 3252–3255. PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
21. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*. 2005; 33(suppl 2): W686–W689.
22. Conant GC, Wolfe KH. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 2008; 24(6): 861–862. doi: [10.1093/bioinformatics/btm598](https://doi.org/10.1093/bioinformatics/btm598) PMID: [18227121](https://pubmed.ncbi.nlm.nih.gov/18227121/)
23. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Research*. 2004; 32(suppl 2): W273–W279.
24. Kurtz S, Schleiermacher C. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics*. 1999; 15(5): 426–427. PMID: [10366664](https://pubmed.ncbi.nlm.nih.gov/10366664/)
25. Chen J, Hao Z, Xu H, Yang L, Liu G, Sheng Y, et al. The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Frontiers in Plant Science*. 2015; 6: 447. doi: [10.3389/fpls.2015.00447](https://doi.org/10.3389/fpls.2015.00447) PMID: [26136762](https://pubmed.ncbi.nlm.nih.gov/26136762/)
26. Vieira Ldo N, Faoro H, Rogalski M, Fraga HP, Cardoso RL, de Souza EM, et al. The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. *PLoS ONE*. 2014; 9(3): e90618. doi: [10.1371/journal.pone.0090618](https://doi.org/10.1371/journal.pone.0090618) PMID: [24594889](https://pubmed.ncbi.nlm.nih.gov/24594889/)
27. 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.). *Theoretical and Applied Genetics*. 2003; 106(3): 411–422. PMID: [12589540](https://pubmed.ncbi.nlm.nih.gov/12589540/)
28. Zhang Y, Li L, Yan TL, Liu Q. Complete chloroplast genome sequences of *Praxelis (Eupatorium catarium* Veldkamp), an important invasive species. *Gene*. 2014; 549(1): 58–69. doi: [10.1016/j.gene.2014.07.041](https://doi.org/10.1016/j.gene.2014.07.041) PMID: [25042453](https://pubmed.ncbi.nlm.nih.gov/25042453/)
29. Ku C, Hu JM, Kuo CH. Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. *PLoS ONE*. 2013; 8(4): e62548. doi: [10.1371/journal.pone.0062548](https://doi.org/10.1371/journal.pone.0062548) PMID: [23638113](https://pubmed.ncbi.nlm.nih.gov/23638113/)
30. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae. *PLoS ONE*. 2015; 10(3): e0120589. doi: [10.1371/journal.pone.0120589](https://doi.org/10.1371/journal.pone.0120589) PMID: [25774672](https://pubmed.ncbi.nlm.nih.gov/25774672/)
31. Yang JB, Yang SX, Li HT, Yang J, Li DZ. Comparative chloroplast genomes of *Camellia* species. *PLoS ONE*. 2013; 8(8): e73053. doi: [10.1371/journal.pone.0073053](https://doi.org/10.1371/journal.pone.0073053) PMID: [24009730](https://pubmed.ncbi.nlm.nih.gov/24009730/)

32. Zhang T, Fang Y, Wang X, Deng X, Zhang X, Hu S, et al. The Complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. PLoS ONE. 2012; 7(1): e30531. doi: [10.1371/journal.pone.0030531](https://doi.org/10.1371/journal.pone.0030531) PMID: [22291979](https://pubmed.ncbi.nlm.nih.gov/22291979/)
33. Zedane L, Hong-Wa C, Murienne J, Jeziorski C, Baldwin BG, Besnard G. Museomics illuminate the history of an extinct, paleoendemic plant lineage (Hesperelaea, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. Biological Journal of the Linnean Society. 2015; 117: 44–57.
34. Qian J, Song JY, Gao HH, Zhu YJ, Xu J, Pang XH, et al. The Complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS ONE. 2013; 8(2): e57607. doi: [10.1371/journal.pone.0057607](https://doi.org/10.1371/journal.pone.0057607) PMID: [23460883](https://pubmed.ncbi.nlm.nih.gov/23460883/)
35. Dong-Keun Y, Ki-Joong K. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS ONE. 2012; 7(5): e35872. doi: [10.1371/journal.pone.0035872](https://doi.org/10.1371/journal.pone.0035872) PMID: [22606240](https://pubmed.ncbi.nlm.nih.gov/22606240/)
36. Choi KS, Park SJ. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. Gene. 2015; 572(2): 214–221. doi: [10.1016/j.gene.2015.07.020](https://doi.org/10.1016/j.gene.2015.07.020) PMID: [26164759](https://pubmed.ncbi.nlm.nih.gov/26164759/)
37. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Molecular Biology and Evolution. 2006; 23(11): 2175–2190. PMID: [16916942](https://pubmed.ncbi.nlm.nih.gov/16916942/)
38. Choi KS, Park S. The complete chloroplast genome sequence of the Korean endemic plant *Scrophularia takesimensis*. Mitochondrial DNA. 2014; 1–2.
39. Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The First complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. PLoS ONE. 2015; 10(6): e0129347. doi: [10.1371/journal.pone.0129347](https://doi.org/10.1371/journal.pone.0129347) PMID: [26046631](https://pubmed.ncbi.nlm.nih.gov/26046631/)
40. Nazareno AG, Carlsen M, Lohmann LG. Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. PLoS ONE. 2015; 10(6): e0129930. doi: [10.1371/journal.pone.0129930](https://doi.org/10.1371/journal.pone.0129930) PMID: [26103589](https://pubmed.ncbi.nlm.nih.gov/26103589/)
41. Smith TC. Chloroplast evolution: secondary symbiogenesis and multiple losses. Current Biology. 2002; 12: R62–R64. PMID: [11818081](https://pubmed.ncbi.nlm.nih.gov/11818081/)
42. Kong W, Yang J. The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. Current Genetics. 2015; 1–8.
43. Yang Y, Dang Y, Li Q, Lu J, Li X, Wang Y. Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: organizations and implications for genetic engineering. PLoS ONE. 2014; 9(11): e110656. doi: [10.1371/journal.pone.0110656](https://doi.org/10.1371/journal.pone.0110656) PMID: [25365514](https://pubmed.ncbi.nlm.nih.gov/25365514/)