Characterization of the complete chloroplast genome of Syringa oblata Lindl. from China and analysis of genome structures, comparative, phylogenetic relationships

Jing Li
Shaanxi University of Chinese Medicine
Ji-Qing Bai (baijiqing323@126.com)
Shaanxi University of Chinese Medicine
Shao-Bing Du
Shaanxi University of Chinese Medicine
Yi Meng
Shaanxi University of Chinese Medicine
Yi-Jun Zhao
Shaanxi University of Chinese Medicine
Hui Hui Zhou
Shaanxi University of Chinese Medicine
Zhi-Peng Xue
Shaanxi University of Chinese Medicine
Xiao-Ping Wang
Shaanxi University of Chinese Medicine
Ping Wang
Xianyang Normal University
Huai-Zhu Li
Xianyang Normal University

Research Article

Keywords: Syringa oblata Lindl., Oleaceae, Chloroplast genome, Repeat sequence, Phylogenetic analysis

Posted Date: January 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2428791/v1

Abstract

Syringa oblata is a flowering tree endemic to China, which has high medicinal and economic value. Comparative genomic analysis is necessary to elucidate the variation and phylogenetic relationships among *Syringa* species. An Illumina NovaSeq 6000 sequencing platform was used to sequence the whole genome of *Syringa oblata* plastomes in this study, and sequence assembly, annotation and characterization were performed by bioinformatics analysis, and a phylogenetic tree using Maximum Likelihood (ML) was constructed. A total of 155,668 bp was contained in the genome, which comprised two inverted repeats (IR) of 25,733 bp each, separated by a small single copy (SSC) region of 17,926 bp and a large single copy (LSC) region of 86 bp. There is a total GC content of 37.9% in the chloroplast genome, while the GC content in each region of SSC, LSC, and IR is 36.0%, 32.1%, and 43.2%. The genome contains 132 genes, including 88 protein-coding genes (CDS), 36 tRNA genes and 8 rRNA. There are seven protein-coding genes in the IR region, seven tRNA genes, and four rRNA genes. There was the most abundance of leucine and the least abundance of tryptophan among all amino acids. We detected 36 simple sequence repeats (SSRs) in the repeat structure analysis which included 32 mononucleotides (p1), 1 dinucleotide (p2), 1 trinucleotide (p3), 1 tetranucleotides (p4), and 1 complex nucleotide were explored, in the *S. oblata* chloroplast genome. In addition, we identified 50 long repeats, comprising 21 forward repeats, 28 palindromic repeats, and 1 complementary repeat. A phylogenetic analysis suggested that *S. oblata* (MW464119) is closely related to *Syringa vulgaris* (MG255768). It is anticipated that our results will be useful to further species identification, population genetic analysis, and biological research on *Syringa oblata* in the future.

Introduction

Syringa oblata Lindl (Lilac) is a perennial shrub or small tree of Oleaceae Syringa, with important medicinal and ornamental values. Lilac leaves contain a variety of chemical components such as flavonoids, phenylpropanoids, triterpenoids (Zhang SJ et al.,2018; Tian L et al.,2013; Li Q et al.,2009), and cyclic enol ether terpenoids, and have pharmacological activities such as anti-inflammatory, antioxidant, antiviral, and hepatoprotective (Gegen JL et al.,2022; Cui L et al.,2019). It has the functions of clearing away heat and detoxifying, removing dampness and relieving jaundice, and is used to treat acute diarrhea, jaundice hepatitis, fiery eyes, sores, etc (Cui L et al.,2019; Zhu WB et al.,2021). China is the natural distribution center of *Syringa*, and *S. oblata*, as a genus of *Syringa*, are mainly distributed in northeast, north and northwest China (except Xinjiang) to southwest China and northwest Sichuan, *S. oblata* is mainly grown in the slopes of the jungle, gully brooks, valley roads and other places at an altitude of 300–2400 meters. There are many cultivated products, and the tender leaves can replace tea. (Editorial Committee of the Flora of China,1992; Yi W et al.,2022). The author reviewed the literature and found that there are few reports on lilac leaves in recent years, Zhao M et al. used the Illumina HiSeq sequencing platform to sequence *S. oblata* (MT025818) and obtained a chloroplast genome of size 155648 bp, but the article only briefly mentioned the basic structure of the chloroplast genome and the type and number of genes of this species and constructed a phylogenetic tree by NJ method (Zhao M et

al,2020). However, it lacks systematic analysis of the predicted and bioanalytical analysis of the genetic structure of the chloroplast genome of *S. oblata*, such as codon preference analysis and simple sequence repeat analysis. Therefore, it is necessary to conduct a comparative analysis of the whole chloroplast genome of lilac.

A chloroplast (cp) Is a critical organelle in plant cells because it is the metabolic center of cellular activity (Chan KX et al.,2016). It is the organelle for photosynthesis in plant cells, which is the plastid that contains chlorophyll and other pigments. In angiosperms, the chloroplast genome is a typical tetrad formed by a small single-copy region (SSC), a large single-copy region (LSC), and two inverted repeat regions (IR) (Liu XF et al.,2019; Zhang Z et al.,2022). The chloroplast genomes of higher plants are highly conserved, showing a typical double-stranded closed-loop structure. At present, the conserved structure, length differences, and gene composition of chloroplast genomes have been reported (Alwadani KG et al.,2019). The chloroplast genome sequence is an important basis for the study of the phylogeny and evolutionary history of plant groups, revealing the variation within and among plant species, which is of great value for species identification, species phylogeny, and genetic engineering (CBOL Plant Working Group ,2009).

Plant chloroplast genome information has been analyzed with the rapid development of high-throughput sequencing technology, which also gives us the opportunity to obtain and analyze the complete chloroplast genome of S. oblata. The resource development and utilization of S. oblata plays a vital role (Zhang Z et al.,2022).

In the study of Syringa identification and phylogenetic relationship, Chen et al. (Chen XL et al.,1999) analyzed the genetic relationship of 15 Syringa varieties by random amplified polymorphic DNA markers and showed that *Syringa vulgaris* was closely related to *Syringa × hyacinthiflora*; Huang et al. (Huang G et al.,2007) discussed the cultivated species and variety origin of Xining clove and divided it into two subgenus of lilac and ligustrum lucidum; Lu et al. (Lu YC et al.,2022)obtained the complete chloroplast genome sequence of *Syringa villosa* subsp. *wolfii*, analyzed its structure and characteristic information, and constructed a phylogenetic tree. Yi Wang et al. (Yi W et al.,2022) revealed the formation of chromosomes in Oleaceae plants and the evolutionary history of lilacs by studying the chromosome-level genome of lilacs.

This study sequenced and assembled a complete cp genome sequence from *Syringa oblata* for analysis of phylogenetic relationships with other members of Oleaceae. We sequenced *S. oblata* using Illumina high-throughput sequencing platform NovaSeq 6000, followed by assembly, gene annotation and structural analysis, identifying simple sequence repeat (SSR) markers and reconstructing evolutionary relationships among Oleaceae species. In the future, these studies will contribute to a better understanding of the cp genome and provide a theoretical framework for studying *S. oblata*. It provides data basis for the research of genetic resources, phylogeny, molecular marker development and plastid genome engineering of *S. oblata* and even Oleaceae.

Materials And Methods

Plant material, DNA extraction and sequencing

Fresh leaves of *Syringa oblata* were collected from Yangjiazhai, jinsuoguan Town, Yijun County, Shaanxi Province, in China (109°03'05.849",35°14'08.216"), a specimen was deposited in the Shaanxi University of Chinese Medicine (Voucher Storage: Shaanxi University of Chinese Medicine, voucher number: ZDX-20200412, the person in charge of the collection: Jiqing Bai, email is baijiqing@126.com). Firstly, the leave tissues were dried rapidly with the silica gel. Next, the total genomic DNA of *S. oblata* from a single individual was extracted through a modified CTAB method. An agarose gel electrophoresis was used to analyze purity and integrity of DNA, and Nanodrop was used to detect purity. To construct the library, use a covariance sonicator to break immediately, followed by end repair, A-tail, sequencing adapters, purification, PCR amplification and other steps. After the library is qualified, it is sequenced by using Illumina high-throughput sequencing platform NovaSeq 6000, and the raw reads is filtered to remove the sequencing linker and low-quality sequences, so as to obtain clean reads (Yang F et al.,2019). The sequencing and filtering are performed by Beijing Nuohezhiyuan Technology Co., Ltd.

Chloroplast gene assembly and annotation

Using MIRA 4.0.2 software (Chevreux B et al.,2004), chloroplast genome of *Syringa oblata* (NCBI accession: MT025818.1) as reference sequence Column, set parameters to assemble the clean reads after sequencing, compare the obtained Contigs with the reference sequence. Use MITObim v1.7 script (Hahn C et al.,2013) to extend the Gap region between Contigs and adjust the overlap between different contigs to splice to obtain the complete chloroplast genome sequence of *Syringa oblata* (NCBI accession: MW464119.1) (Gao S et al.,2019).

Use GENEIOUS 8.0.2 to annotate the assembled files (Brankovics B et al.,2016). Consistent sequences were generated. By searching in the original assembly file, the MAFFT plug-in was used to compare with the reference sequence, and the fuzzy bases (non-A, T, G, C bases) were modified to obtain the accurate chloroplast genome sequence. Then the accurate sequence was annotated with the reference sequence, and the open reading frame (ORF) of the sequence was predicted by using the prokaryotic codon. According to the predicted ORF, the initiation codon and termination codon of each coding sequence (CDS) were checked. After manual adjustment, the accurate sequence was copied and changed to the reverse sequence, and the reverse sequence was compared with the accurate sequence. Partition (LSC, SSC, IRa and IRb) to complete chloroplast genome annotation. And the online software OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) was used to draw the tetrad structure of *S. oblata*. Submit the sequencing data and gene annotations of *S. oblata* to NCBI GenBank database (accession number: MW 464119.1).

Codon preference analysis

The frequency of codon usage (RSCU) in the process of gene translation varies among species, and this phenomenon is closely related to the genetic information carrier DNA and biomolecular proteins, and it is important to analyze the codon preference among different species (JIANG M et al.,2020). In this study, the codon preference analysis of the chloroplast genome of *S. oblata* was performed by the software CodonW 1.4.2 and Online Software EMBOSS explorer (https://www.bioinformatics.nl/emboss-explorer/).

SSRs and long repeat structure

We used Misa-web (https://webblast.ipk-gatersleben.de/misa/index.php)to analyze the SSR loci, and 100 bases was set as the minimum distance between two SSRs. To be statistically significant, the following conditions had to be met: There must be 10 or more repetitions of one base, 8 or more repetitions of two bases, 4 or more repetitions of three bases, 3 or more repetitions of four bases, 3 or more repetitions of five bases, and 3 or more repetitions of six bases. Moreover, when there were less than 100 bp between two microsatellites, they formed a composite microsatellite.

We used REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) to detect the long duplicate. The minimum sequence length was 30 bp, and the distance is edited by 3-bp. Four repetitive search methods were used: (1) F: forward, (2) R: reverse, (3) C: complementary, and (4) P: palindromic

Cp genome comparison

In order to show where the boundaries of the chloroplast genome's four regions are located, the IRScope (https://irscope.shinyapps.io/irapp/) was used to map the genes at the junctions of the eight syringa boundaries to show whether there is contraction and expansion of the IR region of the chloroplast genome.

The whole chloroplast genomes of *S. yunnanensis* (NC 042468), *S. wolfii* (NC 049090), *S. vulgaris* (MG 255768), *S. reticulata* subsp. *Amurensis* (MT 872640), *S. pinnatifolia* (NC 041119), *S. oblata* (MW 464119), *S. persica* (NC042280), *S. oblata* (MT025818) were compared by mVISTA (https://webblast.ipk-gatersleben.de/misa/index.php), with the annotation of *S. yunnanensis* (NC 042468.1) as the reference. MAFFT was performed on the above eight chloroplast genome sequences for comparison, and nucleotide polymorphisms were calculated using DNAsp6.12.03 software with parameters set to window length:600 and Step size:200.

Phylogenetic analysis

A total of 14 reference sequences were downloaded from the NCBI website. All the 15 plastid sequences were aligned using the program MAFFT (Katoh K et al.,2013) with the default molecular parameters. The programs Fast Tree and MEGA-X (GTR) software was used to construct a maximum likelihood tree with the iteration 1,000 times. BI phylogenetic tree is constructed using Bayesian method (default parameter). Finally, the phylogenetic tree was beautified using the online software iTOL (https://itol.embl.de/).

Results And Analysis

Chloroplast characteristics analysis of *Syringa oblata*

The newly sequenced genome (MW464119.1) was a typical quadripartite structure 15668 bp in length, containing a pair of two IR (IRa and IRb) region of 25733 bp each, separated by a LSC region of 86276 bp and an SSC region of 17926 bp. The GC contents of the whole chloroplast genome and the LSC, SSC, and IR regions is 37.9 %, 36.0 %, 32.1 %, 43.2% (Fig.1 and Table 1). It is very common in other plants for the IR and LSC regions to have a higher GC content than the SSC and LSC regions (Asaf S et al.,2017).

There are 132 genes in the genome, including 88 protein-coding genes (CDS), 36 tRNA genes, and eight rRNA genes. IR contains 7 protein-coding genes, 7 tRNA genes, and 4 rRNA genes. SSC contains 12 protein-coding genes and 1 tRNA gene, while LSC contains 62 protein-coding genes and 21 tRNA genes. There were 22 genes with introns, including 15 protein-coding genes and 7 tRNA genes. Three genes (*ycf3*, clpP and *rps12*) have two introns, while the other 19 genes have one intron (Table 2). TrnK-UUU gene intron (2492) is the largest intron. The *rps12* gene is a trans-spliced gene with the 5'end located in the LSC region and the 3'end duplicated in the IR.

It harbored 132 genes, 14 of which encoded the small ribosomal subunit (SSU), 11 the large ribosomal subunit (LSU), and 4 the DNA-directed RNA polymerase (Table 3). A total of 45 genes were associated with photosynthesis, these include five encoding photosystem I and sixteen encoding photosystem II, 11 encoding subunits of NADH dehydrogenase, six encoding cytochrome b/f complex, six for different subunits of ATP synthase, and one encoding rubisco. Thirty-two genes, including 7 tRNA genes, 2 *rrn4.5*, 2 *rrn5*, 2 *rrn16*, 2 *rrn23*, 2 *rps7*, 2 *rpl23*, 2 *ycf1*, 2 *ycf2* and 2 *ycf15* were duplicated in the IR regions. There were 20 gene species with two copies, and most genes were found as single copies, including 4 rRNA species (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*), 9 tRNA species (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnM-CAU*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), and 7 CDS species (*rps7, rps12, rpl2, rpl23, ycf1, ycf2, and ycf15*), in addition, one CDS species (*rps12*) occurred in two copies. Ten CDS species (*rps16, atpF, rpoC1, petB, petD, rpl16, rpl2, ndhB, ndhA and ycf1*) and five tRNA species (*trnK-UUU, trnG-GCC, trnL-UAA, trnI-GAU*, and *trnA-UGC*) contained a single intron, while three other CDS species (*ycf3, rps12*, and *clpP*) harbored two introns.

Table 1 Summary of cp genome of *S. oblata*

Features	Numerical value	Features	Numerical value
Genome size(bp)	155668	GC content in SSC region (%)	32.1
LSC length (bp)	86276	Gene number	132
SSC length (bp)	17926	Protein-coding gene number	88
IR length (bp)	25733	tRNA gene number	36
AT content (%)	62	rRNA gene number	8
GC content (%)	37.9	Gene number in LSC regions	83
GC content in IR region (%)	43.2	Gene number in SSC regions	13
GC content in LSC region (%)	36.0	Gene number in IR regions	36

Table 2 The length of exons and introns in genes with introns in the *S. oblata* chloroplast genome.

Gene	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
trnK-UUU	LSC	38	2492	37		
rps16	LSC	40	879	227		
trnG-GCC	LSC	23	679	48		
atpF	LSC	144	698	411		
rpoC1	LSC	453	711	1620		
ycf3	LSC	124	697	230	742	153
trnL-UAA	LSC	35	474	50		
rps12	LSC	114	-27416	232	541	26
clpP	LSC	71	808	292	643	228
petB	LSC	6	656	642		
petD	LSC	8	716	475		
rpl16	LSC	9	850	399		
rpl2	IRb	391	675	434		
ndhB	IRb	777	679	756		
trnI-GAU	IRb	37	947	35		
trnA-UGC	IRb	38	811	35		
ndhA	SSC	556	1110	530		
ycf1	SSC	2401	33	3245		
trnA-UGC	IRa	38	811	35		
trnI-GAU	IRa	37	947	35		
ndhB	IRa	777	679	756		
rpl2	IRa	391	675	434		

Table 3 Gene contents in the cp genome of *S. oblata*

trn	Group of genes	Gene names	Account
Self-replication	Ribosomal RNA genes	*rrn4.5*(×2), *rrn5*(×2), *rrn16*(×2), *rrn23*(×2)	8
	Transfer RNA genes	*trnA-UGC*(×2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC** (×2), *trnH-GUG*, *trnI-CAU*(×2), *trnI-GAU**(×2), *trnK-UUU**, *trnL-CAA*(×2), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*(×2), *trnN-GUU*(×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-UGU*, *trnV-GAC*(×2), *trnW-CCA*, *trnY-GUA*	36
	Small subunit ribosomal proteins (SSU)	*rps2a*, *rps3*, *rps4*, *rps7*(×2), *rps8*, *rps11*, *rps12***(×2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*	14
	Large subunit ribosomal proteins (LSU)	*rpl2**(×2), *rpl14*, *rpl16**, *rpl20*, *rpl22*, *rpl23*(×2), *rpl32*, *rpl33*, *rpl36*	11
	RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*	4
	Translational initiation factor	*infA*	1
Photosynthesis	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbG*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	16
	NADH dehydrogenase	*ndhA**, *ndhB*, *ndhB**, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*,	11
	Cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*,	6
	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*	6
	Large subunit of rubisco	*rbcL*	1
Other genes	Subunit of acetyl-CoA	*accD*	1
	Envelope membrane protein	*cemA*	1
	Maturase	*matK*	1
	Protease	*clpP***	1
	C-type	*ccsA*	1

	cytochrome synthesis Gene		
Genes of unknown function	Conserved hypothetical chloroplast ORF	*ycf1***(×2), *ycf2*(×2), *ycf3***, *ycf4*, *ycf15*(×2)	8

(×2) Indicates two copies of the gene.

*, ** The symbols indicate the gene with one intron and two introns, respectively.

Examination of codon usage frequency

Codon preference analysis of *S. oblata* chloroplasts was performed using Codon W 1.4.2 software. The results show that the effective number of codons (ENC) is 55.273, the codon adaptation index (CAI) is 0.159. In order to estimate relative synonymous codon usage frequency (RSCU) and codon usage frequency, we used the coding sequence (CDS) (Table4). 51,889 codons make up all protein-coding genes in the cp genome. There were three termination codons among these codons: UAA, UAG, and UGA. The RSCU value for AGA encoding methionine was the highest (1.94). As a result of the large number of codons, leucine was the most abundant amino acid in the protein-coding genes (5080 codons, approximately 9.79 % of the total), while tryptophan was the least abundant amino acid with only 759 codons (1.46 %).

According to the degree of synonymous codon usage (RSCU) of the chloroplast gene CDS sequence, the chloroplast genome CDS sequence of *S. oblata* encodes 88 genes, including 51889 codons. First, second, and third codon positions had GC contents of 38.11%, 38.41%, and 37.26%, respectively. Furthermore, 32 of the 64 codons had RSCU values >1, with most (14/14,87.5%) ending in base A or U, whereas 30 codons had RSCU values above 1, with most (14/12,93.3%) ending in base C or G. *Trp* was encoded by only a UGG codon while *Met* was encoded by only an AUG codon, indicating no biased usage (RSCU=1).

Table 4 Codon usage in *S. oblata*

Amino Acid	Codon	Number	RSCU	Amino Acid	Codon	Number	RSCU
Phe	UUU	2177	**1.21**	Ile	AUU	1834	**1.25**
Phe	UUC	1415	0.79	Ile	AUC	1113	0.76
Leu	UUA	1009	**1.19**	Ile	AUA	1459	0.99
Leu	UUG	1058	**1.25**	Met	AUG	822	**1.00**
Leu	CUU	1122	**1.33**	Val	GUU	800	**1.37**
Leu	CUC	664	0.78	Val	GUC	392	0.67
Leu	CUA	706	0.83	Val	GUA	740	**1.26**
Leu	CUG	521	0.62	Val	GUG	412	0.70
Ser	UCU	1216	**1.50**	Thr	ACU	726	**1.22**
Ser	UCC	909	**1.12**	Thr	ACC	601	**1.01**
Ser	UCA	974	**1.20**	Thr	ACA	694	**1.17**
Ser	UCG	599	0.74	Thr	ACG	355	0.60
Pro	CCU	738	**1.13**	Ala	GCU	519	**1.33**
Pro	CCC	636	0.97	Ala	GCC	352	0.90
Pro	CCA	838	**1.28**	Ala	GCA	456	**1.17**
Pro	CCG	409	0.62	Ala	GCG	234	0.60
Tyr	UAU	1285	**1.33**	Asn	AAU	1783	**1.42**
Tyr	UAC	651	0.67	Asn	AAC	734	0.58
TER	UAA	1091	**1.15**	Lys	AAA	2063	**1.33**
TER	UAG	753	0.80	Lys	AAG	1035	0.67
His	CAU	884	**1.41**	Asp	GAU	1117	**1.46**
His	CAC	372	0.59	Asp	GAC	415	0.54
Gln	CAA	1073	**1.41**	Glu	GAA	1395	**1.39**
Gln	CAG	446	0.59	Glu	GAG	618	0.61
Cys	UGU	683	**1.23**	Ser	AGU	679	0.84
Cys	UGC	427	0.77	Ser	AGC	476	0.59
TER	UGA	993	**1.05**	Arg	AGA	1101	**1.94**
Trp	UGG	759	**1.00**	Arg	AGG	639	**1.13**

Arg	CGU	402	0.71	Gly	GGU	575	**1.02**
Arg	CGC	250	0.44	Gly	GGC	331	0.59
Arg	CGA	595	**1.05**	Gly	GGA	828	**1.47**
Arg	CGG	412	0.73	Gly	GGG	524	0.93

Note: Red font indicates that it represents a start codon or a stop codon, Bolded font indicates that its RSCU value is equal to 1, Bolded and underlined font indicates that its RSCU value is greater than 1.

Analyses of repeat sequences and SSRs

A total of 50 repeat sequences, including 21 forward(F), 28 palindromic(P), and 1 complement(C), except for the first palindrome repetition, the other repetition lengths are generally between 30 and 58. Repeats were found in 22 regions in the LSC, 10 regions in the SSC, and 34 regions in the IR.

The SSR loci in the chloroplast genome of *S. oblata* were examined using MISA-Web, and a total of 42 loci were obtained, including 32 mononucleotides (p1), 2 trinucleotides (p3), 5 tetranucleotides (p4), 1 pentanucleotide (p5) and 2 complex nucleotides were explored. Most were distributed in the intergenic spaces (IGS) (27, 64.29%), with some in the CDS (15,35.71%) (Table 6).

Table 5 Repeat sequences in the *S. oblata* chloroplast genome.

ID	Repeat Start	Type	Size(bp)	Repeat Start2	Mismatch(bp)	E-Value	Gene	Region
1	86276	P	25733	129935	0	0.00e+00	rps19, ycf1	LSC, SSC
2	93445	F	58	93463	-3	6.83e-20	ycf2	IRb IRb
3	93445	P	58	148423	-3	6.83e-20	ycf2 ycf2	IRb, IRa
4	93463	P	58	148441	-3	6.83e-20	ycf2 ycf2	IRb, IRa
5	148423	F	58	148441	-3	6.83e-20	ycf2	IRa IRa
6	93455	F	52	93473	-3	2.01e-16	ycf2	IRb IRb
7	93455	P	52	148419	-3	2.01e-16	ycf2 ycf2	IRb, IRa
8	93473	P	52	148437	-3	2.01e-16	ycf2 ycf2	IRb, IRa
9	100441	F	41	121777	0	1.41e-15	IGS, ndhA	IRb, SSC
10	121777	P	41	141462	0	1.41e-15	ndhA, rps12	SSC, IRa
11	128556	F	41	128589	0	1.41e-15	ycf1	SSC SSC
12	76644	P	44	76644	-2	1.87e-13	IGS	LSC LSC
13	93463	F	44	93481	-2	1.87e-13	ycf2 ycf2	IRb IRb
14	93463	P	44	148419	-2	1.87e-13	ycf2 ycf2	IRb, IRa
15	93481	P	44	148437	-2	1.87e-13	ycf2 ycf2	IRb, IRa
16	93443	F	46	93479	-3	5.64e-13	ycf2	IRb IRb

17	93443	P	46	148419	-3	5.64e-13	ycf2 ycf2	IRb, IRa
18	93479	P	46	148455	-3	5.64e-13	ycf2 ycf2	IRb, IRa
19	148419	F	46	148455	-3	5.64e-13	ycf2	IRa IRa
20	114912	P	43	114912	-3	2.94e-11	IGS	SSC SSC
21	45367	F	39	100443	-2	1.50e-10	ycf3, IGS	LSC, IRb
22	45367	F	39	121779	-2	1.50e-10	ycf3, ndhA	LSC, SSC
23	45367	P	39	141462	-2	1.50e-10	ycf3, rps12	LSC, IRa
24	40367	F	41	42591	-3	4.06e-10	psaB-psaA	LSC LSC
25	148446	F	35	148464	-2	3.09e-08	ycf2	IRa IRa
26	93455	F	34	93491	-2	1.17e-07	ycf2	IRb IRb
27	93455	P	34	148419	-2	1.17e-07	ycf2 ycf2	IRb, IRa
28	93491	P	34	148455	-2	1.17e-07	ycf2 ycf2	IRb, IRa
29	9328	P	30	47020	-1	5.32e-07	IGS, trnS-GGA	LSC LSC
30	17395	F	35	17396	-3	1.02e-06	IGS	LSC LSC
31	115469	P	32	115469	-2	1.65e-06	IGS	SSC SSC
32	6726	P	33	6726	-3	1.36e-05	IGS	LSC LSC
33	14366	P	30	14366	-2	2.31e-05	IGS	LSC LSC
34	109291	F	30	109322	-2	2.31e-05	IGS	IRb IRb

35	109291	P	30	132592	-2	2.31e-05	IGS	IRb, IRa
36	109322	P	30	132623	-2	2.31e-05	IGS	IRb, IRa
37	115468	P	30	115468	-2	2.31e-05	IGS	SSC SSC
38	132592	F	30	132623	-2	2.31e-05	IGS	IRa IRa
39	40379	F	32	42603	-3	4.95e-05	psaB, psaA	LSC LSC
40	56591	P	31	56632	-3	1.79e-04	IGS	LSC LSC
41	115470	C	31	115472	-3	1.79e-04	IGS	SSC SSC
42	431	P	30	488	-3	6.48e-04	IGS	LSC LSC
43	10838	F	30	38164	-3	6.48e-04	trnG-GCC trnG-GCC	LSC LSC
44	37170	P	30	47020	-3	6.48e-04	psbC-trnS-UGA trnS-GGA	LSC LSC
45	45379	F	30	100455	-3	6.48e-04	ycf3 rps12-trnV-GAC	LSC IRb
46	45379	P	30	141459	-3	6.48e-04	ycf3 trnV-GAC-rps12	LSC IRa
47	91046	F	30	91088	-3	6.48e-04	ycf2	IRb
48	91046	P	30	150826	-3	6.48e-04	ycf2 ycf2	IRb IRa

49	91088	P	30	150868	-3	6.48e-04	ycf2	IRb
							ycf2	IRa
50	150826	F	30	150868	-3	6.48e-04	ycf2	IRa

Note: An inverted repeat is represented by IRa and IRb. A small single copy region and a large single copy region are represented by SSC and LSC, respectively.

Table 6 Distribution of SSRs in the *S. oblata* cp genome.

SSR nr.	SSR type	SSR	size	start	end	
1	p1	(T)10	10	508	517	IGS (trnH-GUG-psbA)
2	p1	(T)12	12	3309	3320	CDS (matK)
3	p1	(C)10	10	5537	5546	CDS (rps16)
4	p1	(T)10	10	5678	5687	CDS rps16
5	p1	(T)10	10	9058	9067	IGS psbK- psbI
6	p1	(T)10	10	10978	10987	IGS (trnG-GCC- trnR-UCU)
7	p1	(T)10	10	13324	13333	CDS (atpF)
8	p1	(T)12	12	14119	14130	IGS (atpF-atpH)
9	p1	(T)10	10	19235	19244	CDS (rpoC2)
10	p1	(T)11	11	19634	19644	CDS (rpoC2)
11	p1	(T)12	12	20172	20183	CDS (rpoC2)
12	p1	(A)10	10	31670	31679	IGS (psbM- trnD-GUC)
13	p1	(A)10	10	31808	31817	IGS (psbM- trnD-GUC)
14	p1	(A)12	12	34038	34049	IGS (trnM-CAU- psbD)
15	p1	(A)11	11	37162	37172	IGS (psbC-trnS-UGA)
16	p1	(A)10	10	37831	37840	IGS (psbZ-trnG-GCC)
17	p1	(A)11	11	46004	46014	CDS (ycf3)
18	p1	(A)12	12	46257	46268	IGS (ycf3-trnS-GGA)
19	p1	(A)11	11	48715	48725	IGS (trnT-UGU-trnL-UAA)
20	p1	(T)11	11	48898	48908	IGS (trnT-UGU-trnL-UAA)
21	p1	(T)11	11	50166	50176	IGS (trnF-GAA- ndhJ)
22	p1	(T)10	10	52717	52726	IGS (ndhC-trnM-CAU)
23	p1	(T)10	10	56370	56379	CDS (atpB)
24	p1	(T)12	12	61649	61660	IGS (psaI- ycf4)
25	p1	(T)13	13	65505	65517	IGS (petA- psbJ)
26	p1	(T)10	10	68428	68437	IGS (petL-petG)
27	p1	(A)10	10	71513	71522	IGS (rpl20-rps12)

28	p1	(T)10	10	72918	72927	CDS (clpP)
29	p1	(A)11	11	73066	73076	CDS (clpP)
30	p1	(T)10	10	73920	73929	CDS (clpP)
31	p1	(T)10	10	80396	80405	CDS (rpoA)
32	p1	(T)10	10	83955	83964	CDS (rpl16)
33	p3	(AAG)4	12	3033	3044	IGS (rps-4trnT-UGU)
34	p3	(AAT)5	15	43967	43981	IGS (psaA-ycf3)
35	p4	(CTTT)3	12	362	373	IGS (rpl32- trnL-UAG)
36	p4	(AATA)3	12	6621	6632	IGS (rps2-rpoC2)
37	p4	(AAAT)3	12	9301	9312	IGS (psbI-trnS-GCU)
38	p4	(AAAG)3	12	72594	72605	CDS (clpP)
39	p4	(ATTA)6	24	115471	115494	IGS (rpl32- trnL-UAG)
40	p5	(TAATC)3	15	125045	125059	IGS (rps15-ycf1)
41	c	(C)14 (A)12	36	17396	17431	IGS (rps2- rpoC2)
42	c	(ATAG)3 (TCTT)3	48	30964	31011	IGS (petN- psbM)

IR junction characteristics

There has been an expansion and contraction of the IR-LSC and IR-SSC boundaries of eight species, including *S. yunnanensis*, *S. wolfii*, *S. vulgaris*, *S. reticulata* subsp*. Amurensis*, *S. pinnatifolia*, *S. oblata* (MW464119), *S. persica*, *S.oblata* (MT025818) were compared (Fig.3). The figure shows the junctions of IRb/LSC, IRb/SSC, IRa/SSC, and IRa/SSC as indicated by the symbols JLB, JSB, JSA, and JLA. There was no significant difference in LSC, IR, or SSC size among the eight species, and there was a variation in IR length from 25362 bp in *S. wolfii* to 25844 bp in *S. reticulata* subsp*. Amurensis*. The JLB boundaries of the eight species mentioned above are located within the coding region of *rps19*, which varies in length from 1 to 6 bp in the IR region, *rpl22* gene is in the LSC region and *rpl2* gene is in the IR region. Except for *S. oblata* (MW464119), whose *trnH* is within the LSC region, the JLA boundaries of the remaining seven species are located within the *trnH*, mostly in the LSC region, and the length of the IR region is 13 or 14 bp. Five sequences had JSB boundaries across the *ndhF* gene, mostly in the SSC region, and 23 or 25 bp in length in the IR region. *ndhF* of *S.* persica was at the JSA boundary, mostly in the SSC region, and 25 bp in the IR region, and *psbA* was all distributed in the LSC region. The JSB and JSA boundaries of all seven species except *S. vulgaris* were across *ycf1*, with large differences in length.

Comparison of chloroplast genome sequences

In the region of cp genomes that is highly variable, markers can be exploited for identifying closely related species and for providing abundant phylogenetic information. Using DnaSP and mVISTA, we compared the differences between *S. oblata* cp genomes (Xu J et al.,2017).

An essential part of genomics is comparing chloroplast genomes (Chen S et al.,2012, Cui N et al.,2020). With *S. yunnanensis* as a reference, we used mVISTA to analyze the data, we plotted the overall sequence identities of the seven *Syringa* chloroplast genomes and observed approximately identical gene orders and organizational patterns (Fig.4). Across the seven data, plastome sequences were generally conserved, except for a few regions where variation occurred. The results showed that the variation between sequences mainly occurred in the non-coding region, and there were obvious differences in the *psbK*, *accD*, *ycf1*, *ycf2* genes in the coding region.

Polymorphism analysis using DnaSP6 software to detect hypervariable regions in the chloroplast genomes of eight species of *Syringa*. The sliding window analysis showed that the overall variation of nucleotide diversity (Pi) ranged from 0 to 0.19976, with a mean value of 0.02291. The gene fragments with Pi>0.1 were marked in the figure for the LSC region (*accD*, *psaI*), SSC region (*ccsA*, *trnL-UAG*, *ycf1*), and IR region (*ycf1*, *trnN-GUU*, *rrn23*) (Fig.5).

Phylogenetic analysis

Insight can be gained into plant evolutionary relationships, population genetics, and taxonomy through cpDNA-based phylogenetic analysis (Gu C et al.,2019). To explore the taxonomic status and evolutionary relationship of *S. oblata* plants, *Forsythia* and *Ligustrum* of Oleaceae were selected as outgroups in this study, and ML phylogenetic analysis was performed based on 15 complete cp genome sequences (Fig.6). The maximum likelihood (ML) method and Bayesian (BI) method were used to construct the phylogenetic tree of 15 species. Two different methods produced the same topological structure. The phylogenetic tree results indicate that *S. oblata* (MW464119) forms an evolutionary branch with *S. vulgaris* (MG255768), which is consistent with the classification in FRPS, both belonging to *Syringa S. Syringa*. Furthermore, the figure shows that the evolutionary tree has 12 nodes for 14 species, which can be identified as three major branches, *S. oblata* and *S. vulgaris* are clustered as a small branch with 100% support, indicating that they are closely related. Surprisingly, not all Syringa species are in the same branch, which may be related to their spatial distribution and range among different species.

Discussion

The chloroplast genome is an important part of herbaceous genomics. With the continuous development of bioinformatics and high-throughput sequencing technology, the research on chloroplast genome is getting deeper and deeper. The research on chloroplast genome is of great significance for revealing the structure and origin of chloroplast DNA, plant molecular markers, and species relationship (Wu LW et al.,2020, Xing SC et al.2008). In this study, the chloroplast of *S. oblata* is 155668 bp in length and has a typical circular tetrad structure, including LSC region (86276 bp), SSC region (17926 bp) and two IR regions (25733 bp). It has been shown that the lower the GC content, the worse the thermal stability of the

genome (Qiao YG et al.,2019), the GC content of *S. oblata* was 37.9%, which was essentially the same as that of closely related species in Oleaceae (Zhang JW,2019) , and the GC content was higher than the average GC content of chloroplast genomes of higher plants, which in summary suggests that the DNA sequence of *S. oblata* is relatively more stable(Wang J et al.,2018). In the IR expansion contraction analysis, the LSC, SSC, IR regions were highly similar among the eight species, and the gene fragments near the boundaries of *rps19*, *ycf1*, and *trnH* were in approximately the same intervals, also indicating that the chloroplast genome of *S. oblata* has a highly conserved character. The 132 genes in the genome include self-replicating genes, photosynthetic genes, other functional genes, and genes of unknown function. In addition, 18 genes contained one or two introns, and the intron of trnR-UKK is the largest. Several studies have shown that introns control gene expression in a significant way (Niu DK et al.,2011). Zhao et al. (Zhao M et al.,2020) sequenced the whole genome sequence of Syringa oblata (MT025818). The length of the chloroplast genome was 155648 bp, including LSC (86247 bp), SSC (17937 bp) and two IR regions (25732 bp). It is highly similar to the tetrad structure of this study.

The effective number of codons (ENC) and codon adaptation index (CAI) of *S. oblata*'s chloroplast genome codons, respectively, were determined by preference analysis. The ENC is 55.273 and the CAI is 0.159. It indicated that the *S. oblata* chloroplast genome has a weak codon bias and low expression of endogenous genes. Lower ENC values indicate stronger codon usage preference, and it is generally accepted that when ENC ≤ 35, the gene has significant codon preference. The GC content in the chloroplast genome is only 37.9 %, and the GC content at the first, second and third codon positions is 38.11 %, 38.41 % and 37.26 %, respectively, all less than 50%, indicating that the *S. oblata* chloroplast genome prefers to use A as a codon. and U bases. The factors that codon preference include natural selection, mutation pressure, etc (Suzuki Y,2010), Huo et al. (Huo X et al.,2021) have pointed out that the codons of the chloroplast genome are mainly affected by mutations, and base differences, natural selection, and gene expression levels will also affect them. Therefore, the codons of the chloroplast genome in this study are also affected by many factors. It provides a theoretical basis for genetic engineering modification and codon optimization of lilac and its related species and provides assistance for subsequent breeding research and variety optimization.

According to the repeat analysis of the genome of *S. oblata* cp, there are 21 forward, 28 palindromic, and 1 complement repeats, there were several repeats present in introns, intergenic spacers, and the *ycf* gene, but a few occurred also in CDS regions and tRNA sequences. SSRs are mostly found in the organelle genomes of eukaryotes and have abundant polymorphism, high reproducibility as well as good reliability and co-dominance (Wang SQ et al.,2022). A total of 42 SSR loci were tagged in the whole genome of *S. oblata* chloroplasts, with the highest number of single nucleotide repeats and most of them consisting of A/T bases, most were distributed in the intergenic spaces (IGS) (27, 64.29%), with some in the CDS (15,35.71%). The results of the analysis in this study showed that the SSRs in the whole genome sequence of *S. oblata* chloroplasts were mainly single nucleotide types (76.19% of the total), The repeat types of single nucleotide SSRs were mainly A/T (about 96.88 % of the total number of single nucleotides), which further verified the conclusion that SSRs in chloroplast genome sequences were mainly composed of polyA or polyT, and less C or G tandem repeats (Kuang DY et al.,2011). These SSRs

can be molecular remains of medicinal plants of *S. oblata* Genetic studies provide candidate molecular markers. In this study, we analyzed the characteristics of SSR loci in the chloroplast genome of *S. oblata*, and the findings can provide some reference for the studies on genetic diversity, genetic map construction, and molecular marker technology of *S. oblata*.

In this study, we analyzed the phylogenetic relationships of syringa oblata based on the complete chloroplast whole genome sequence using maximum likelihood and Bayesian methods, and the results showed that the genera clustered into branches with high support. The phylogenetic analysis of *S. oblata* with Oleaceae *Syringa* and other genera in this study shows that *S. oblata* belongs to sister taxa with *S. vulgaris*, which is consistent with the taxonomic system in Flora of China (Editorial Committee of the Flora of China,1992). From the phylogenetic tree, it can be seen that the sequence *S.oblata* (MT025818) and *S.oblata* (MW 464119) are clustered together with a support rate of 100, both of which are sister groups to *S. vulgaris* (MG255768). A new material basis was laid for subsequent resource development, species identification and genetic evolutionary studies of *S. oblata*. Since the chloroplast genome is a single-parent inheritance, its molecular fragments can only provide limited phylogenetic information. Therefore, in the future research, it is necessary to introduce a more comprehensive clove group and combine the chloroplast genome with the nuclear genome of parental genetic characteristics, and carry out morphological anatomy, micromorphology, and embryology analysis, in order to more fully understand its phylogenetic evolution (Bi LF et al., 2020, Lu YC et al.,2022).

This study obtained the basic information and phylogenetic relationship of the chloroplast genome of *S. oblata*, which provided data basis for the protection of genetic diversity, optimization of horticultural traits, phylogenetic evolution, and other related research of *S. oblata*, and provided new ideas for germplasm identification and DNA barcoding development.

Declarations

Data Availability Statement

All relevant data are within the manuscript and its Supporting Information files.

Availability of data and materials

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/nuccore/MW464119) under the accession no. MW464119. The associated Bio-Project, SRA, and Bio-Sample numbers are PRJNA816987, SRR18345671, and SAMN 26741405, respectively.

Author Contributions

Jing Li, Jiqing Bai, and Xiaoping Wang designed this study, Jing Li, Jiqing Bai and Zhi-Peng Xue collect samples, Jing Li, Shao-Bing Du, Yi Meng, Yi-Jun Zhao and Hui Hui Zhou retrieved databases and analyzed data, Ping Wang Huai-Zhu Li reviewed a draft of this manuscript.

Funding

This work was supported by grants from the Wei yang District 2019 Medical Service and Guarantee Capability Improvement Subsidy Fund Project (The Fourth National Traditional Chinese Medicine Resources Survey Project of the Inheritance and Development Section of Traditional Chinese Medicine Business) (Social Security Department of the Ministry of Finance [2019] No. 39).

Competing interests

The authors have declared that no competing interests exist.

References

1. Asaf S, Khan AL, Khan MA, Waqas M, Kang SM, Yun BW, et al. Chloroplast genomes of Arabidopsis halleri ssp. gemmifera and Arabidopsis lyrata ssp. petraea: Structures and comparative analysis[J]. Sci Rep. 2017 Aug 8,7(1):7556.
2. Alwadani KG, Janes JK, Andrew RL. Chloroplast genome analysis of box-ironbark Eucalyptus[J]. Mol Phylogenet Evol. 2019 Jul, 136:76-86.
3. Brankovics B, Zhang H, van Diepeningen AD, van der Lee TA, Waalwijk C, de Hoog GS. GRAbB: Selective Assembly of Genomic Regions, a New Niche for Genomic Research[J]. PLoS Comput Biol. 2016 Jun 16,12(6): e1004753.
4. Bi YF, Wen X, Pan YH, Cai HJ, Zhong H, Wang AK. Application and Research Progress of Chloroplast DNA Barcoding in Forest Trees. Molecular Plant Breeding[J]. 2020,18(16):5444-5452.
5. Chen XL, Chen ZF, Han J. Using random amplified polymorphic DNA (RAPD) markers for lilac genetic analysis and classification in lilac cultivars[J]. Acta Botanica Boreali-Occidentalia Sinica. 1999(02):3-10.
6. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WE, Wetter T, et al. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs[J]. Genome Res. 2004 Jun,14(6):1147-59. doi: 10.1101/gr.1917404. Epub 2004 May 12. PMID: 15140833, PMCID: PMC419793.
7. CBOL Plant Working Group. A DNA barcode for land plants[J]. Proc Natl Acad Sci U S A. 2009 Aug 4,106(31):12794-7.
8. Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, et al. Genome sequence of the model medicinal mushroom Ganoderma lucidum[J]. Nat. Commun. 2012, 3, 913.
9. Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ. Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond[J]. Annu Rev Plant Biol. 2016 Apr 29, 67:25-53.
10. Cui N, Liao BS, Liang CL, Li SF, Zhang H, Xu J, et al. Complete chloroplast genome of Salvia plebeia: organization, specific barcode, and phylogenetic analysis[J]. Chin J Nat Med. 2020 Aug,18(8):563-572.

11. Editorial Committee of the Flora of China, Chinese Academy of Sciences. Flora of China. Vol. 61 [M]. Beijing: Science Press,1992:71.
12. Gu C, Ma L, Wu Z, Chen K, Wang Y. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales[J]. BMC Plant Biol. 2019 Jun 26,19(1):281.
13. Cui L, Hu M, Cao P, Niu Y, Li C, Liu Z, et al. Chemical constituents, and coagulation activity of Syringa oblata Lindl flowers[J]. BMC Chem. 2019 Aug 14,13(1):108.
14. Gao S, Bai JQ, Liu ML, Wang PF, Li N, Yang L, et al. The complete chloroplast genome of Forsythia mira, an endemic medicinal shrub in China[J]. Mitochondrial DNA B Resour. 2019 Dec 9,5(1):57-58.
15. Gegen JL, Gao Y, Site GL, Alatan CLM, Tai BDLH, Tu Y. Mechanism of Syringa oblata in treating angina pectoris based on GC-MS and network pharmacology [J]. China Journal of Chinese Materia Medica,2022,47(03): 836-845.
16. Huang G, Liu Q., Ma WQ. Culture Species Names and Classification on Clove in Xining[J]. Science and Technology of Qinghai Agriculture and Forestry. 2007(03):82-83+85.
17. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach[J]. Nucleic Acids Res. 2013 Jul,41(13): e129. doi: 10.1093/nar/gkt371. Epub 2013 May 9. PMID: 23661685, PMCID: PMC3711436.
18. Huo X, Liu S, Li Y, Wei H, Gao J, Yan Y, et al. Analysis of synonymous codon usage of transcriptome database in Rheum palmatum. PeerJ. 2021 Jan 4,9: e10450.
19. JIANG M, WANG JF, YING MH, Yang RM, Ma JY. Assembly and sequence analysis of Tetrastigma hemsleyanum chloroplast genome [J]. Chinese Traditional and Herbal Drugs, 2020, 51(02): 461-468.
20. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. Mol Biol Evol. 2013 Apr,30(4):772-80.
21. Li Q, Xu QM, Hao LL, Lu BF, Yang SL, Li XR. Chemical constituents from leaf of Syringa oblata [J]. Chinese Traditional and Herbal Drugs,2009,40(03):369-371.
22. Liu XF, Zhu GF, Li DM, Wang XJ. Complete chloroplast genome sequence and phylogenetic analysis of Spathiphyllum 'Parrish'[J]. PLoS One. 2019 Oct 23,14(10): e0224038.
23. Lu YC, Zhang X, Zhang LY, Wang JL.Complete Chloroplast Genome Structure and Characterization of *Syringa villosa* subsp. *wolfii* [J/OL]. Bulletin of Botanical Research:1-11[2022-10-04]. http://kns.cnki.net/kcms/detail/23.1480.S.20220915.1317.002.html.
24. Niu DK, Yang YF. Why eukaryotic cells use introns to enhance gene expression: splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cut activity[J]. Biol Direct. 2011 May 18, 6:24.
25. Qiao YG, He JX, Wang YF, Cao YP, Jia MJ, Zhang XR, et al. Analysis of chloroplast genome and its characteristics of medicinal plant Sophora flavescens [J]. Acta Pharmaceutica Sinica, 2019,54(11):2106-2112.

26. Suzuki Y. Statistical methods for detecting natural selection from genomic data. Genes Genet Syst. 2010,85(6):359-76.
27. Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics[J]. Genome. 2011 Aug,54(8):663-73.
28. Tian L, Li YJ, Lv SW, Zhang L, Liu T. Chemical Constituents of Syringa oblata Leaves[J]. Chinese Journal of Experimental Traditional Medical Formulae,2013,19(01): 144-147.
29. Wang J, Dong P, Wu W, Pan X, Liang X. High-throughput thermal stability assessment of DNA hairpins based on high resolution melting[J]. J Biomol Struct Dyn. 2018 Jan,36(1):1-13.
30. Wu LW, Cui YX, Nie LP, Xu ZC, Wang Y, Song JY, et al. The characteristics of complete chloroplast genome sequence and phylogenetic analysis of Dendrobium moniliforme [J]. Acta Pharmaceutica Sinica,2020,55(05):1056-1066.
31. Wang SQ, Wang QQ, Xu XC, Zhao WB, Zhang WY. Chloroplast Genome structure and Characterization of Ixora chinensis [J/OL]. Molecular Plant Breeding:1-14[2022-07-13].http://kns.cnki.net/kcms/detail/46.1068.S.20220223.1836.019.html
32. Wang Y, Lu L, Li J, Li H, You Y, Zang S, et al. A chromosome-level genome of Syringa oblata provides new insights into chromosome formation in Oleaceae and evolutionary history of lilacs[J]. Plant J. 2022 Aug,111(3):836-848. doi: 10.1111/tpj.15858. Epub 2022 Jul 14. PMID: 35673966.
33. Xing SC, Clarke J.L. Progress in chloroplast genome analysis [J] Progress in Biochemistry and Biophysics,2008,35(1):21~28.
34. Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, et al. Panax ginseng genome examination for ginsenoside biosynthesis[J]. Gigascience 2017, 6, 1–15.
35. Yang F. Sequencing and Structural Analysis of Chloroplast Genome of Rosa banksias[J]. Genomics and Applied Biology,2019,38(08):3586-3594.
36. Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics[J]. New Phytol. 2017 May,214(3):1355-1367.
37. Zhang SJ, Shi ZC, Wang D, Wang JL, Zhao M, Li J. Chemical constituents from leaf of Syringa oblata [J]. Chinese Traditional and Herbal Drugs,2018,49(16):3747-3757.
38. Zhang JW. The complete chloroplast genome and phylogenetic analysis of the endangered species *Syringa pinnatifolia* Oleaceae [D]. Northwest Agriculture and Forestry University, 2019.
39. Zhao M, Zhang Y, Xin Z, Meng X, Wang W. The complete chloroplast genome of *Syringa oblata* (Oleaceae) [J]. Mitochondrial DNA B Resour. 2020 Jun 2,5(3):2278-2279.
40. Zhu WB, Su FZ, Sun YP, Yang BY, Wang QH, Kuang HX. Anti-pharyngitis Effects of Syringa oblata L. Ethanolic Extract in Acute Pharyngitis Rat Model and Anti-Inflammatory Effect of Iridoids in LPS-Induced RAW 264.7 Cells[J]. Evid Based Complement Alternat Med. 2021 Dec 10, 2021:5111752.
41. Zhang Z, Tao M, Shan X, Pan Y, Sun C, Song L, et al. Characterization of the complete chloroplast genome of Brassica oleracea var. italica and phylogenetic relationships in Brassicaceae[J]. PLoS

Figures

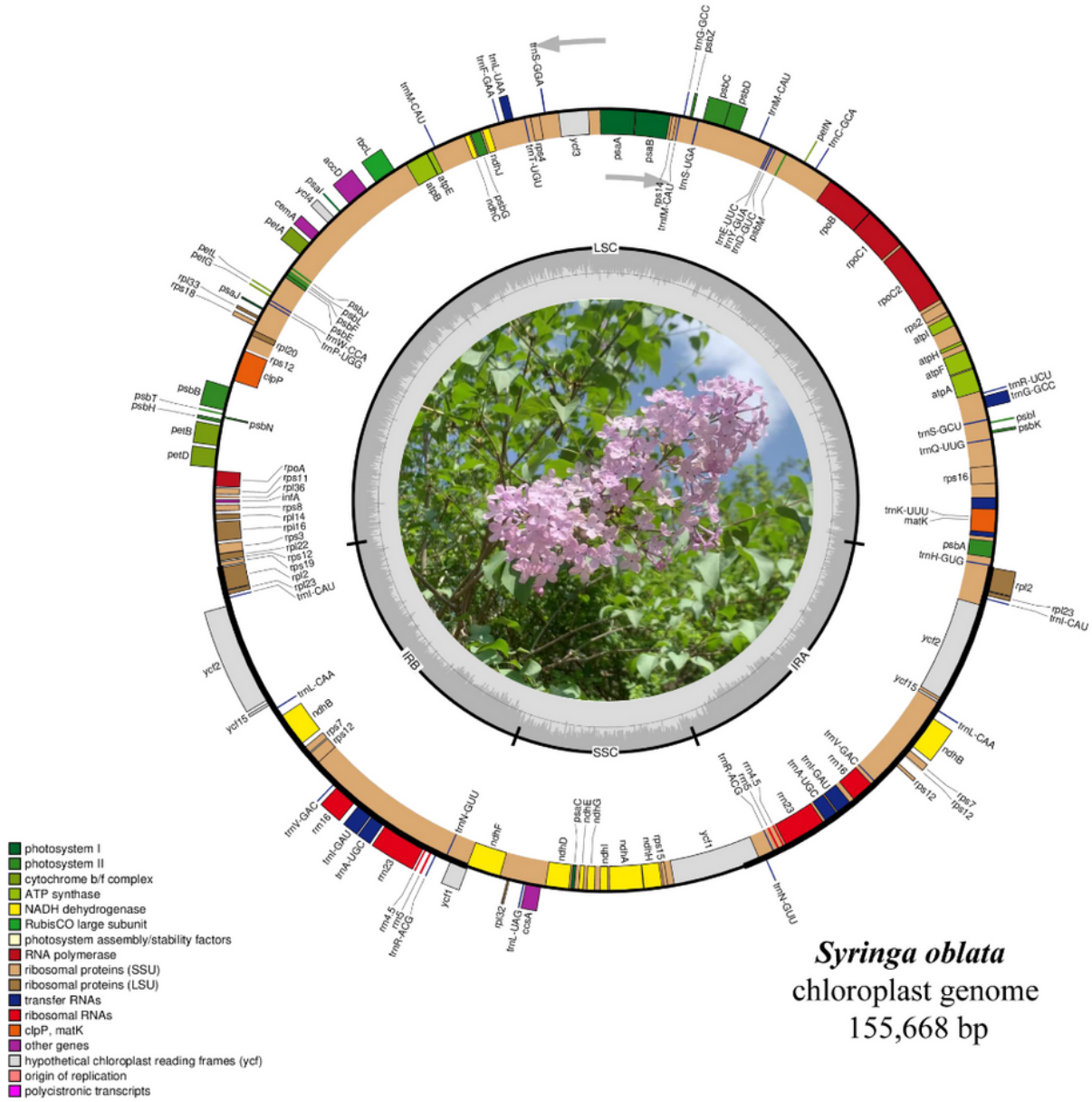


Figure 1

Gene map of *S. oblata*. (The genes inside the circle are transcribed clockwise and those outside the circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. Dark gray in the inner circle indicates GC content, light gray indicates AT content)

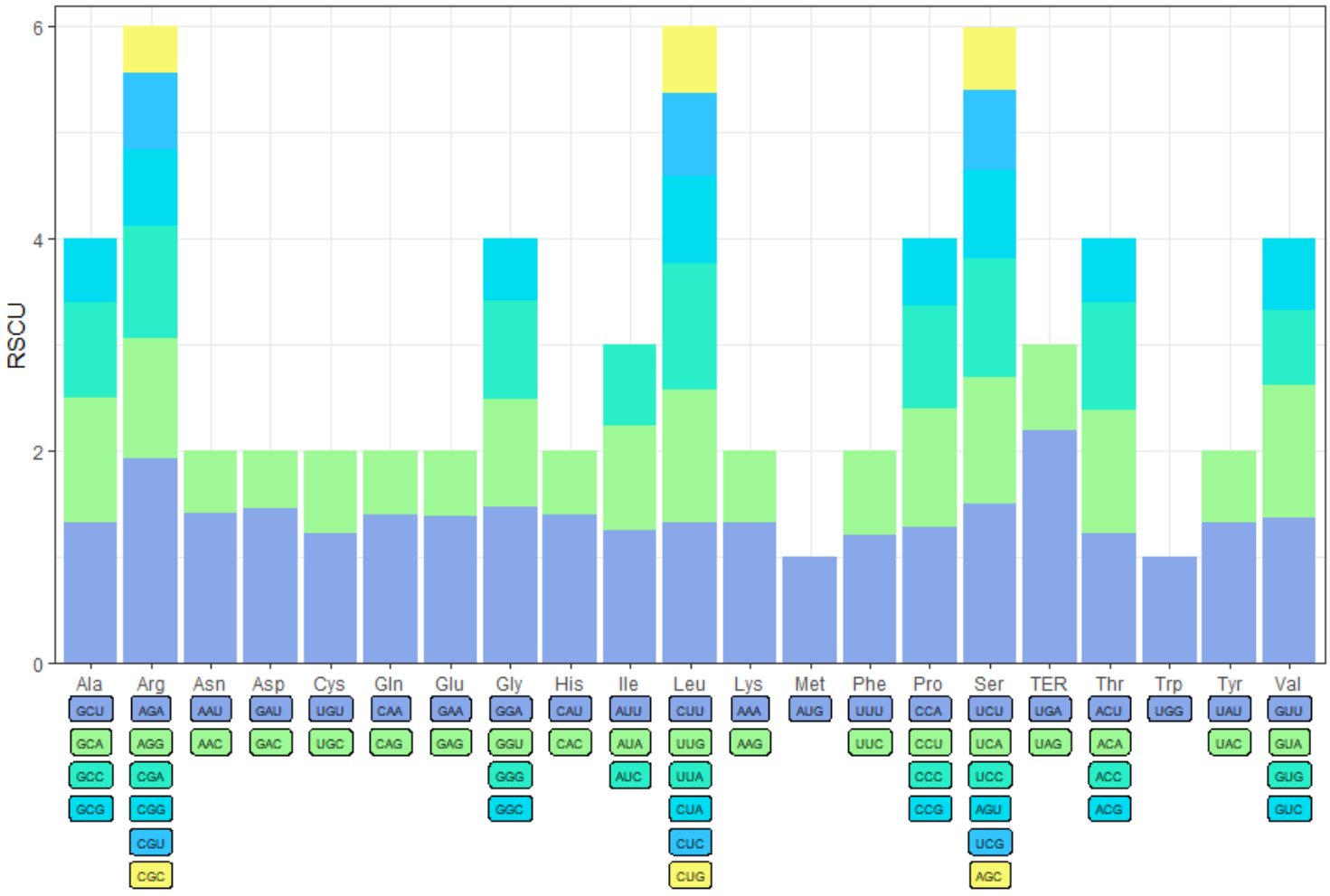


Figure 2

Codon contents of 20 amino acid and stop codons in all protein-coding genes of the *S. oblata*cp genome. (The corresponding codon is the same color in each column of the graph, and the height of the different colors in the bar graph indicates the number of its corresponding codon.)

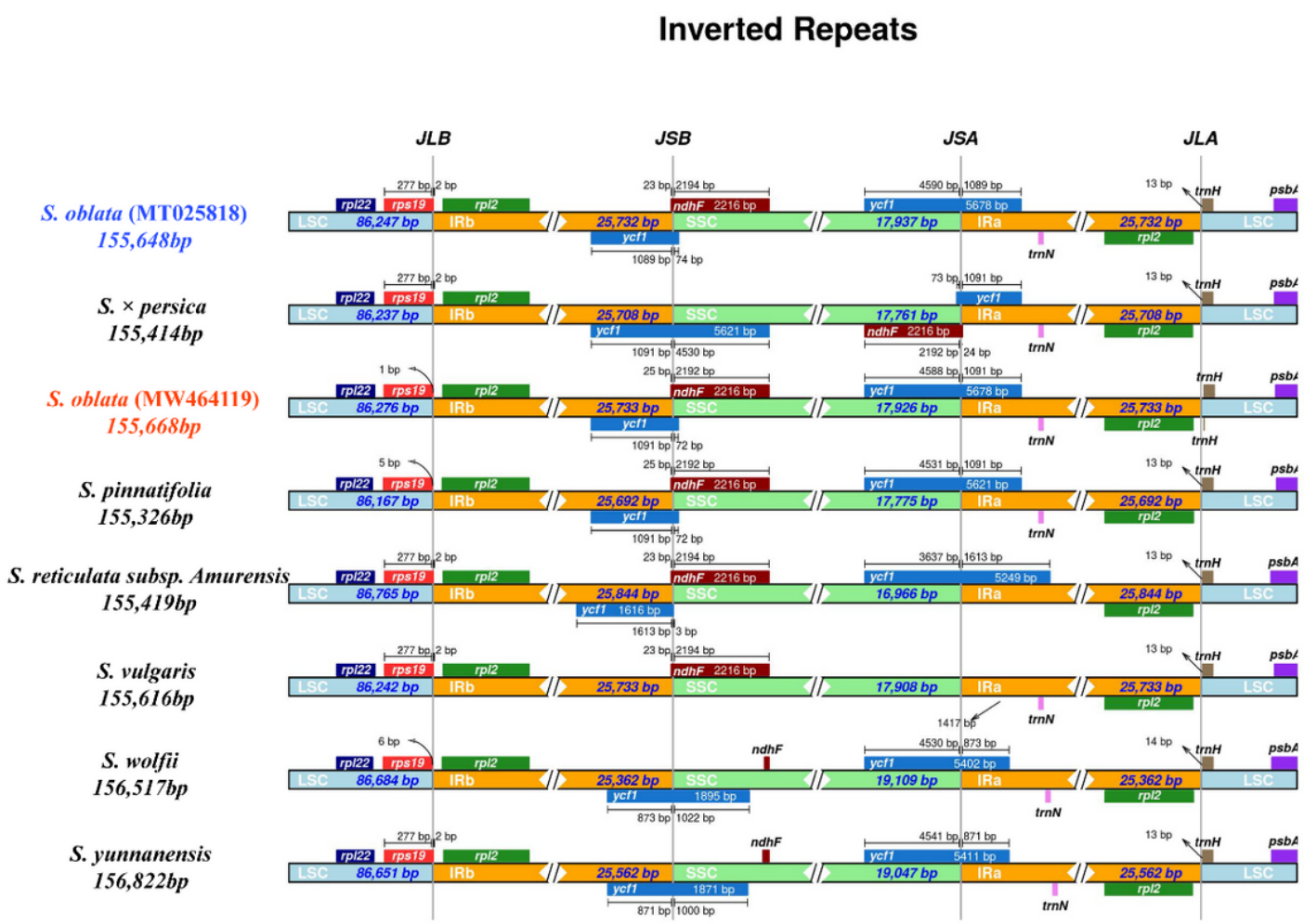


Figure 3

Comparison of boundaries between the LSC, IR, and SSC regions in chloroplast genomes of eight species. (Genes are depicted by colored boxes. Boxes above or below the main line indicate adjacent border genes. JLB, junction of LSC/IRb, JSB, junction of IRb/SSC, JSA, junction of SSC/IRa, JLA, junction of IRa/LSC. Sequence number S. oblata (MT025818) was sequenced by Zhao et al. in 2020, and sequence number S. oblata (MW464119) was sequenced in this study.)

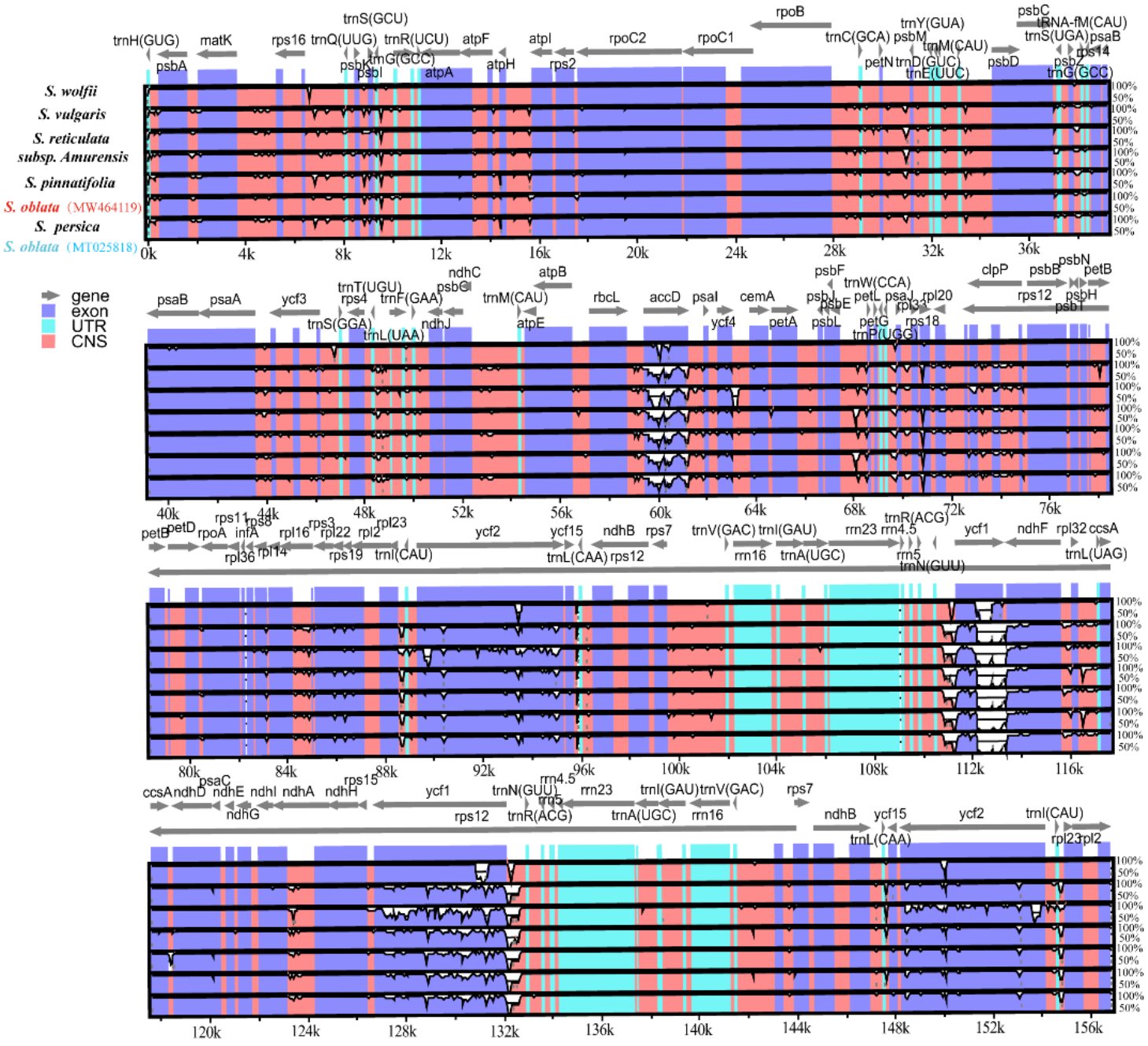


Figure 4

Comparison of the LSC, SSC, and IR regions among six chloroplast genomes. (The vertical scale indicates the percentage of identity, ranging from 50 to 100%, The horizontal axis indicates the coordinates within the cp genome.)

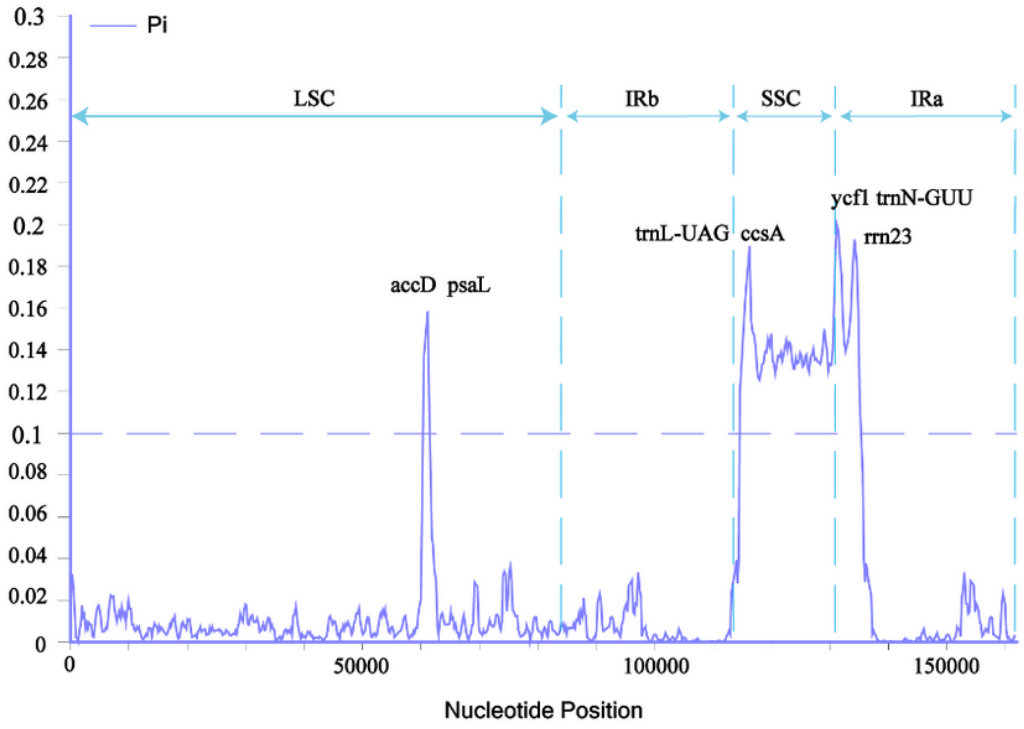
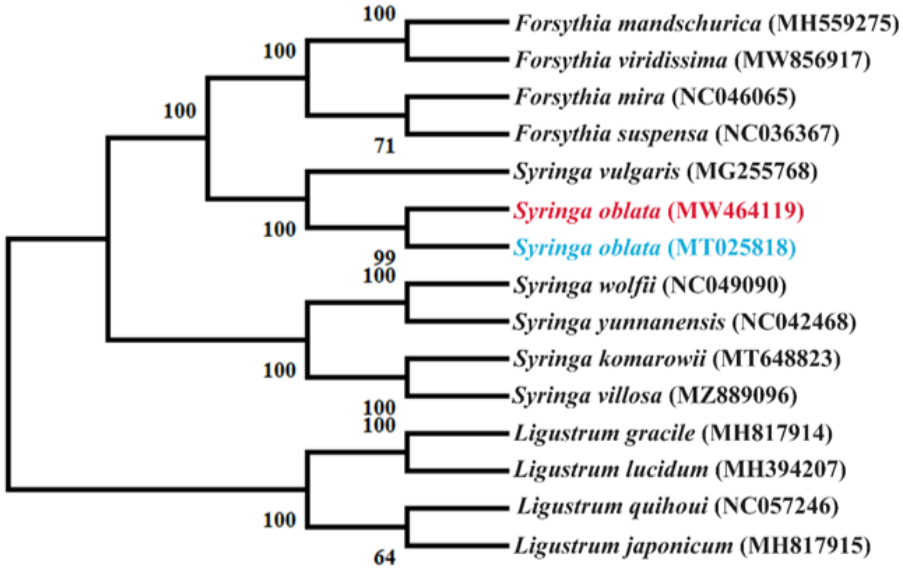
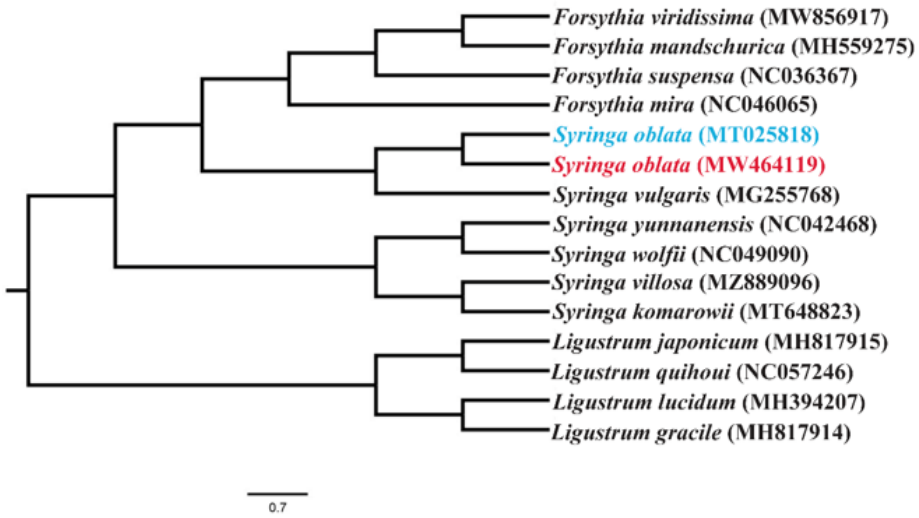


Figure 5

Comparative analysis of nucleotide diversity (Pi) values among the seven S. species cp genome sequences. (Window length: 600 bp, step size: 200 bp. X-axis: position of the midpoint of a window, Y-axis: nucleotide diversity of each window)



A: Phylogenetic analysis of chloroplast genomes of 15 species of Oleaceae by maximum likelihood (ML) method



B: Phylogenetic analysis of chloroplast genomes of 15 species of Oleaceae by Bayesian (BI) method

Figure 6

Maximum likelihood method (ML) phylogenetic tree was constructed based on the whole chloroplast genome sequence. (Red indicates species sequenced in this study, black indicates species downloaded from NCBI.)