



Analysis of the chloroplast genome and phylogenetic evolution of three species of *Syringa*

Chengjun Yang^{1,2} · Kai Wang^{1,2} · Hang Zhang^{1,2} · Qingjie Guan¹ · Jian Shen^{1,3}

Received: 9 March 2022 / Accepted: 3 October 2022 / Published online: 12 November 2022
 © The Author(s) 2022

Abstract

Background By the time our study was completed, the chloroplast genomes of *Syringa oblata*, *S. pubescens* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis* had not been sequenced, and their genetic background was not clear.

The research content In this study, the chloroplast genomes of *Syringa oblata*, *S. pubescens* subsp. *Microphylla*, *S. reticulata* subsp. *Amurensis*, and five other species of *Syringa* were sequenced for a comparative genomics analysis, inverted repeat (IR) boundary analysis, collinearity analysis, codon preference analysis and a nucleotide variability analysis. Differences in the complete chloroplast genomes of 30 species of *Oleaceae* were compared with that of *S. oblata* as the reference species, and *Ginkgo biloba* was used as the out group to construct the phylogenetic tree.

Results The results showed that the chloroplast genomes of *S. oblata*, *S. pubescens* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis* were similar to those of other angiosperms and showed a typical four-segment structure, with full lengths of 155,569, 160,491, 155,419, and protein codes of 88, 95, and 87, respectively. Because the IR boundary of *S. pubescens* subsp. *Microphylla* was significantly expanded to the large single copy (LSC) region, resulting in complete replication of some genes in the IR region, the LSC region of *S. pubescens* subsp. *Microphylla* was significantly shorter than those of *S. oblata* and *S. reticulata* subsp. *Amurensis*. Similar to most higher plants, these three species have a preference for their codons ending with A/T.

Conclusions We consider the genus *Syringa* to be a synphyetic group. The nucleotide variability and phylogenetic analyses showed that *Syringa* differentiated before *Ligustrum* and *Ligustrum* developed from *Syringa*. We propose removing the existing section division and directly dividing *Syringa* into five series.

Keywords *Syringa* · Phylogeny · Chloroplast genome

Abbreviations

LSC	Large single copy regions
SSC	Small single copy regions
IR	Inverted repeat regions
CDS	Coding sequence
SSR	Simple sequence repeat
cpSSR	Chloroplast simple sequence repeat

mono	Mononucleotide repeats
di	Dinucleotide repeats
tri	Trinucleotide repeats
tetra	Tetranucleotide repeats
penta	Pentanucleotide repeats
hexa	Hexanucleotides repeats
complex	Complexnucleotide repeats
SNP	Single nucleotide polymorphism;
InDel	Insertion or deletion of small DNA sequences < 50 bp
Ts	Number of SNPS converted;
Tv	Number of switched SNPS

✉ Jian Shen
 myshenjian@163.com

¹ Key Laboratory of Saline-Alkali Vegetation Ecology Restoration, Northeast Forestry University, Harbin 150040, China

² Northeast Asia Biodiversity Research Center, Northeast Forestry University, Harbin 150040, China

³ China-Ukraine Agriculture & Forestry Technology Development and Application International Cooperation Joint Lab, Jiamusi University, Jiamusi 154007, China

Background

Syringa plants occupy an important position in Chinese gardens. *Syringa oblata*, *Syringa pubescentis* subsp. *Microphylla*, and *Syringa reticulata* subsp. *Amurensis* belong to *Syringa* of Oleaceae [1]. According to the relative length of corolla tubes and sepals, *Syringa* is classified into *Sect. Syringa* and *Sect. Ligustrina*. Accord to the different leaf shape and inflorescence, *Sect. Syringa* be divided into *Ser. Pinnatifoliae* Rehd, *Ser. Pubescentes* (Schneid.) Lingelsh, *Ser. Syringa* and *Ser. Villosae* (Schneid.) Rehd. However, this classification is controversial, and whether *Sect. Syringa* should be an independent section or a parallel series with the *Ser. Syringa* is the main point of controversy.

Chloroplasts are the photosynthetic organelles of green plants, and have functions of synthesizing starch, fatty acids, pigments, and proteins [2]. Chloroplasts are generally considered the origin of endosymbiosis. Photosynthesizing prokaryotic cyanobacteria were absorbed into cells by primitive eukaryotes with phagocytic ability, and gradually transitioned from parasitism to symbiosis, and became a eukaryote organelle that allowed solar energy to be used autonomously, and the transformation from heterotrophic to autotrophic occurred. The chloroplast has a complete genetic system called the chloroplast genome. The chloroplast genome of gymnosperms is generally inherited from the paternal line [3], while the chloroplast genome of most angiosperms is inherited from the maternal line. About 20% of the genome may be inherited from both parents or from the paternal line [4–6].

The sequence and structure of the chloroplast genome are relatively conserved. The chloroplast genome of most plants has a double-stranded ring structure, including four regions, as two inverted repeat regions (IRa/IRb), one large single copy region (LSC) and one small single copy region (SSC) in four parts. Genome length is relatively fixed at 12–18 kb [7]. As a higher organelle of plants, the chloroplast genome has a stable structure that is highly conserved, with a large number of copies that can be easily extracted and purified. A series of characteristics determine the important position and significance of the chloroplast genome in plant phylogeny, species identification, and genetic transformation. The chloroplast genome can be used as important research material to study species phylogeny, which plays an important role in determining the genetic relationships among related species. Applying the chloroplast genome at the population level provides clues to study the timing of differentiation and intraspecific intensity. Analyses with nuclear genome data provide a more comprehensive understanding of plant pollen transmission, migration pathways, and species evolutionary dynamics.

In this study, we sequenced the complete chloroplast genomes of *S. oblata*, *S. pubescentis* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis* and conducted comparative genomics of eight species of Oleaceae. We analyzed and combined the complete chloroplast genomes, LSC regions, SSC regions, IR regions, and introns of 30 species of Oleaceae published at the NCBI to construct a phylogenetic tree, and explored the phylogenetic relationships among the three species and their phylogeny in Oleaceae.

Methods

Plant materials

The plant materials required for this study were collected from the campus of Northeast Forestry University (45°43′ 16.83″ N, 126°38′ 2.04″ E) in July 2020. The fresh leaves of the three plants were quickly frozen in liquid nitrogen after collection, and then stored in a refrigerator at −40 °C. DNA extraction was carried out, and the extracted DNA was sent to Nanjing Jisihuiyuan Company for subsequent sequencing. The sequencing was submitted to NCBI database. The serial numbers of *Syringa oblata*, *S. reticulata* subsp. *Amurensis* and *S. pubescentis* subsp. *Microphylla* is MT872639.1 MT872640.1 MT872641.1. The complete chloroplast genome sequences of the other five *Syringa* species were obtained by NCBI database.

Genome assembly and annotation

This study adopted the whole genome shotgun strategy to construct libraries with different insert fragments [8] using second-generation sequencing technology (next-generation sequencing), based on the HiSeq sequencing platform, to construct the libraries using paired-end (Paired-end, PE) sequencing [9]. After DNA extraction, purification, library construction and sequencing. The off-machine data were saved in the Paired-end FASTQ format, and the off-machine data were further filtered to remove the reads with connectors and the low quality in the sequencing data to ensure the quality of the subsequent information analysis [10]. Spades 3.12 was used to de-assemble the sequencing data with the adapter sequence removed to obtain the final assembly sequence [11]. OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) was used to make a chloroplast genome map [12].

Genome feature analysis and comparative genome analytical method

Repetitive sequences are important genetic markers and are closely related to the origin and evolution of a species.

Repetitive sequences have been divided into scattered repetition and tandem repetition types. Scattered repetitive sequences are repetitive sequences that are scattered in the genome. The scattered repetitions were analyzed using Reputer software [13], there is a special form of tandem repetition, which is also called SSR. SSRs often have natural polymorphisms, and cpSSR is the key component of a chloroplast genome repeat sequence analysis. The simple repetitive sequences were identified using MISA tools (parameters: 1-10 2-5 3-4 4-3 5-3 6-3) [14]. Codon preference was analyzed using the EMBOSS [15] (The European Molecular Biology Open Software Suite) and CodonW online programs, and the loss of genetic diversity reflected in the chloroplast genome was enhanced and showed a lower level of genetic diversity. Nucleotide diversity (π) was used to evaluate the level of population genetic diversity, using DnaSP software [16] to determine the local collinearity between genomes through multiple sequence alignment, and to perform similarity and a reanalysis of row, inversion, and other phenomena. The Mauve tool was used for the collinearity analysis. RAxML software was used to construct the maximum likelihood tree for the chloroplast genome [17].

Results

Genetic analysis

Similar to other angiosperms, the chloroplast genome structures of the eight species of *Syringa* listed were all closed circular double-stranded DNA, each containing four partitions, one LSC region, and one SSC region with two IRs (Fig. 1). The length of seven genomes, except that of *S. pubescens* subsp. *Microphylla*, was about 15.5 kb, and GC content was 37.9%. The lengths of the LSC, SSC, and IR regions were 8.6, 1.7, and 2.5 kb, respectively, and GC contents were 36%, 32%, and 43% respectively. The total number of encoded genes was 130–132. *S. oblata* and *S. reticulata* subsp. *Amurensis* showed highly consistent chloroplast genome characteristics. The total length of the chloroplast genome of the two species was about 15.5 kb, the length of each partition was similar, and the numbers of tRNAs and rRNAs were 37 and 8, respectively. Only one protein-coding gene difference was detected in the chloroplast genomes of these two species (Table S1). The chloroplast genome of *S. oblata* contained 88 protein-coding genes, which had one more acetyl-CoA carboxylase gene (*accD*) than *S. reticulata* subsp. *Amurensis*. The chloroplast genome of *S. pubescens* subsp. *Microphylla* was significantly longer than that of the other seven species at 160,491 bp, as well as the length of the four partitions. The largest difference was found in the length of the LSC region of *S. pubescens* subsp. *Microphylla*, which was 82,890 bp, while the length of the LSC

region of the other seven species was about 86 kb, which was significantly shorter than the other listed species. There was little difference in the length of the SSC regions of these species. The IR region of *S. pubescens* subsp. *Microphylla* was 29,958 bp, which was longer than that of the other seven species. The IR regions of the other seven species were 25–26 kb. The shortening of the LSC region and expansion of the IR region of *S. pubescens* subsp. *Microphylla* caused *rpl14*, *rpl16*, *rpl22*, *rps19*, *rps3*, *rps8*, *infA*, and other genes originally located in the LSC region to fall into the IR region. Thus, resulting in 95 protein-coding genes and 140 total genes. The gene types in the chloroplast genome of *S. pubescens* subsp. *Microphylla* were the same as those in *S. oblata*. Only quantitative differences in the three types of genes were detected, such as large ribosomal subunits, small ribosomal subunits, and translational initiation factors. The chloroplast genome of *S. oblata* contained only one copy of the *rpl14*, *Rpl16*, *rpl22*, *rps19*, *rps3*, *rps8*, and *infA* genes, whereas *S. pubescens* subsp. *Microphylla* contained two. These differences appeared between *trnH*-GUG and *ycf2*, and no other differences were detected in genes at other positions (Table S2).

Expansion and contraction of the IR zone

Although the IR region of the chloroplast genome is the most conserved region, the contraction and expansion of its border region is a common phenomenon in the evolution of a chloroplast genome. This is also one of the most important factors for the length of the chloroplast genome. Among the eight species of *Syringa*, the borders of the LSC and IRb regions of seven species, except *S. pubescens* subsp. *Microphylla*, fell between the two genes, *rps19* and *rpl2*. The *rps11*, *rpl36*, and *infA* genes were found near the boundary of the LSC and IRb regions in *S. pubescens* subsp. *Microphylla*, which was different from the other seven species. The *rps19* gene of *S. oblata* and *S. reticulata* subsp. *Amurensis* was located on the boundary between the LSC and IRb regions, and both extended 2 bp into the IRb region. The *ycf1* and *ndhF* genes were found near the borders of the SSC and IRb regions of *S. oblata*, *S. pubescens* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis*. The lengths of the *ycf1* gene extending into the SSC region were 74, 305, and 99 bp, respectively. The *ndhF* gene of *S. pubescens* subsp. *Microphylla* and *S. reticulata* subsp. *Amurensis* fell completely in the SSC region, and the *ndhF* gene of *S. oblata* extended 23 bp into the IRb region, becoming a pseudogene. The *rpl2* and *trnH* genes were found near the boundary between the LSC and IRa regions of the remaining seven species, except *S. pubescens* subsp. *Microphylla*, and they were close to the boundary between the LSC and IRa regions. The *trnH* gene of *Syringa wolfii* was 14 bp from the boundary between the LSC and IRa

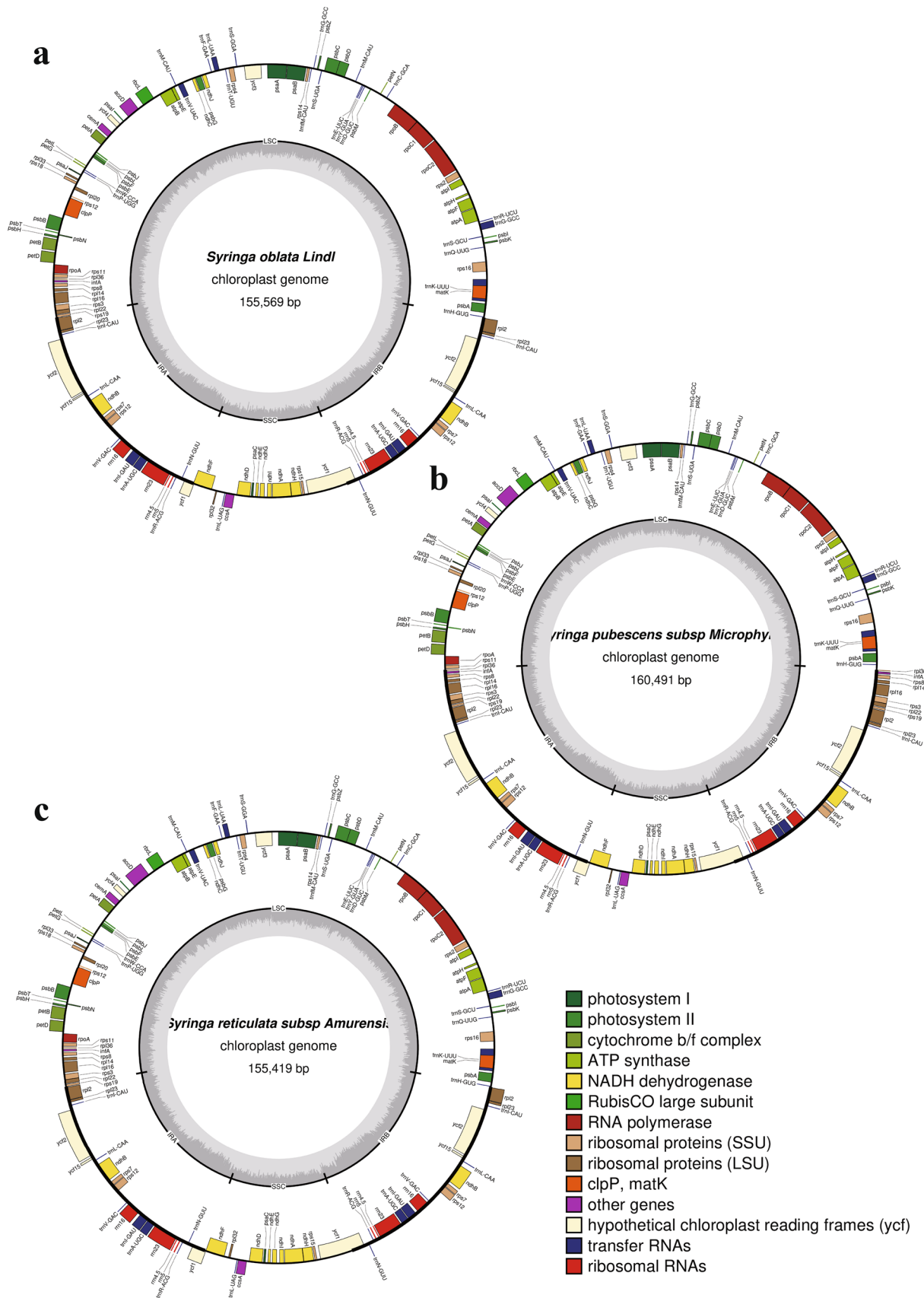


Fig. 1 Gene map of the complete of three species of *Syringa*

regions, and the remaining six species exhibited a distance of 13 bp. Three genes, such as *rps8*, *infA*, and *rpl36*, were detected at the boundary between the LSC and IRa regions of *S. pubescens* subsp. *Microphylla*. All three genes were located in the IRa region, and *rpl36* was 1 bp from the boundary. The *trnH* gene was also found in the LSC region of *S. pubescens* subsp. *Microphylla* but it was distant from the boundary between the LSC and IRa regions (Fig. 2).

Codon preference analysis

26,620, 27,428, and 25,723 codons were found in the complete chloroplast genomes of *S. oblata*, *S. pubescens* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis*, respectively. The most frequently used amino acid was Leu, and the smallest amino acid was Cys. All other amino acids have two or more codons, except Try, which has only one codon, and Arg, Leu and Ser have six codons. The codons of the eight species tended to end in A/U, which is the same as the codon preference of other angiosperm chloroplast genomes [18] (Fig. 3).

Repeat sequence analysis

Repetitive sequences are important genetic markers and are closely related to the origin and evolution of a species. Interspersed and tandem repetitive sequences are the two main types of repetitive sequences. Interspersed repetitive sequences generally include four types, namely forward repeats, reverse repeats, palindrome sequences, and complement sequences. There are differences in the numbers and types of scattered repeating sequences in different species. The total number of interspersed repetitive sequences of *S. pubescens* subsp. *Microphylla* ($n = 164$) and *S. reticulata* subsp. *Amurensis* ($n = 163$) was far more than the other listed species, and only 46 were found in *S. oblata*. Among the eight species of *Syringa*, only complementary sequences were found in the chloroplast genomes of *S. microphylla* and *S. yunnanensis*. The chloroplast genome of *S. reticulata* subsp. *Amurensis* contained only forward ($n = 135$) and palindromic sequences ($n = 29$), excluding IRs and complementary sequences. All four scattered repeats (111, 32, 10, and 10) were found in *S. pubescens* subsp. *Microphylla*. No complementary sequences were detected in the chloroplast genome of *S. oblata*. Twenty forward repeats and 25

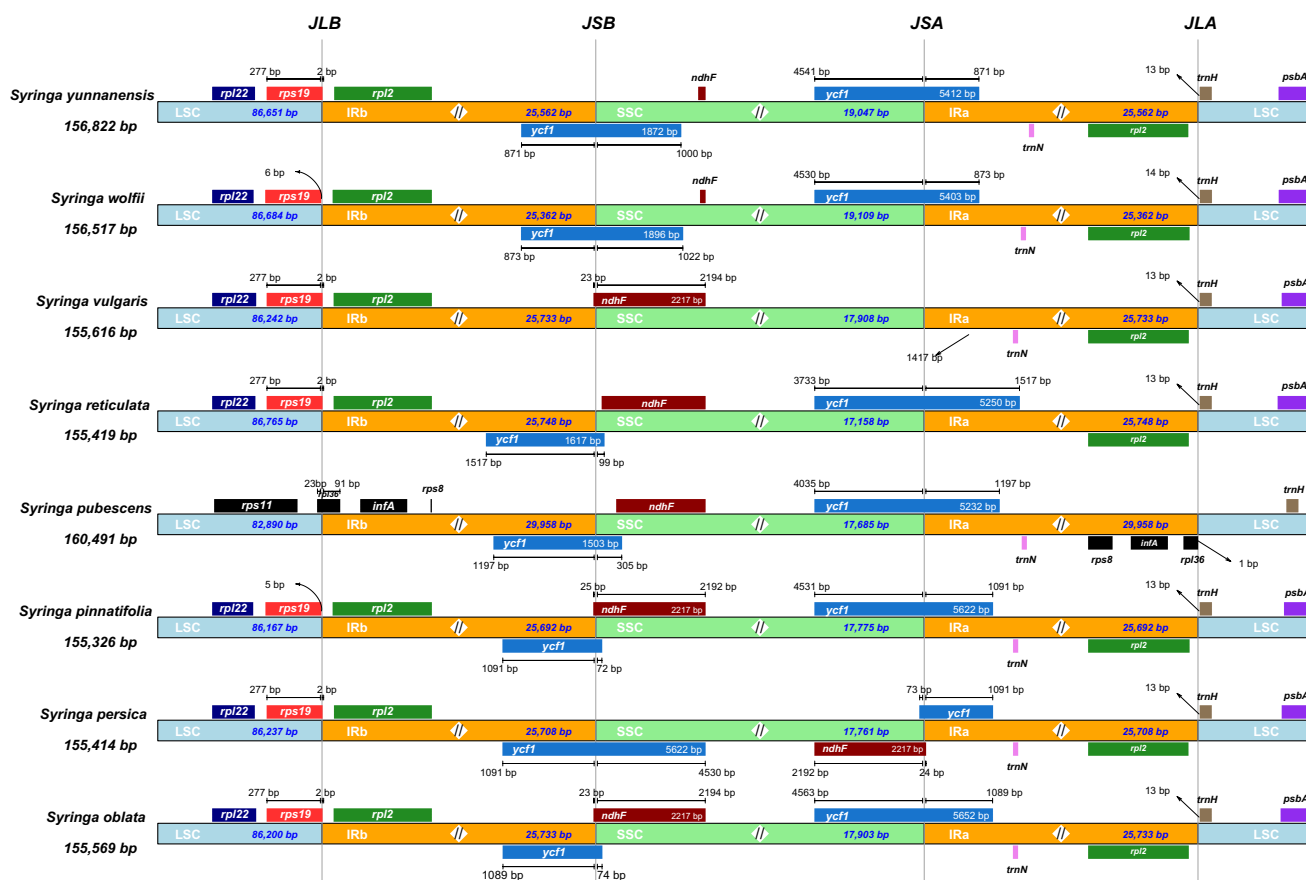


Fig. 2 Expansion and contraction of the inverted repeat (IR) region in eight species of *Syringa*

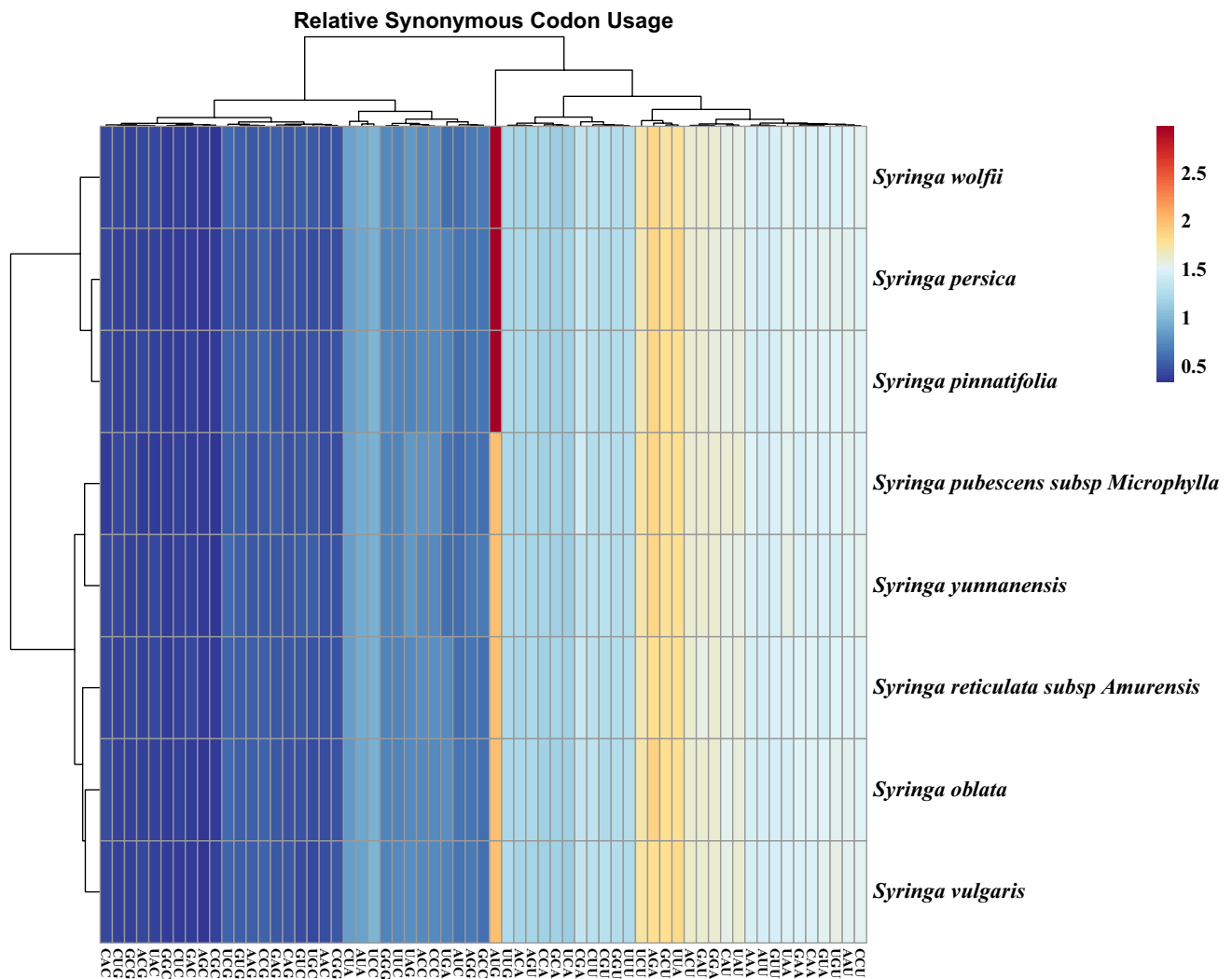
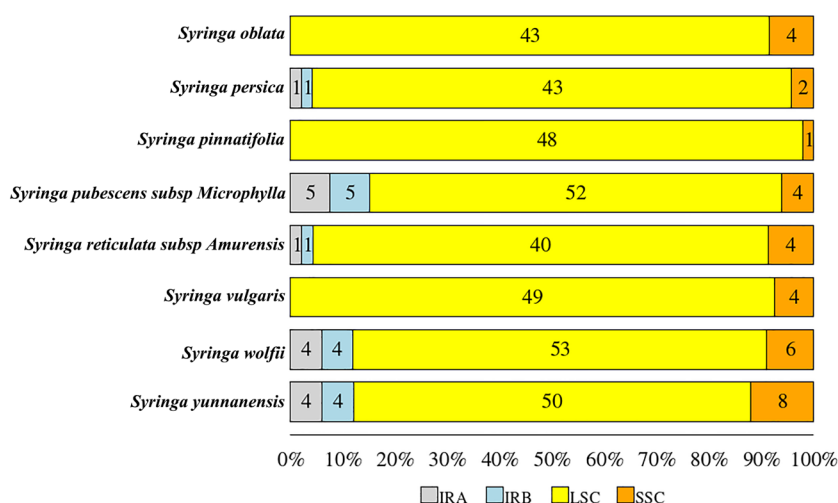


Fig. 3 Heatmap of codon clustering for eight species of *Syringa*

palindromic sequences were observed, which contained only one IR (Table S3).

A special form of tandem repeat is a simple tandem repeat, also called a simple sequence repeat (SSR). SSRs often have natural polymorphisms. cpSSR is the key content of a chloroplast genome repeat sequence analysis. The types and numbers of SSR sites in *S. oblata*, *S. pubescens subsp. Microphylla*, and *S. reticulata subsp. Amurensis* were studied. The results showed that the number of SSRs varied among the different species. Mononucleotide repeats account for the majority of SSRs, and these SSRs all appear in non-coding regions or protein-coding genes, and generally do not appear in tRNA or rRNA (Table S4). The numbers of mononucleotide SSRs in *S. oblata*, *S. pubescens subsp. Microphylla*, *S. reticulata subsp. Amurensis* were 35, 36 and 47, respectively. None of these three species contained pentanucleotide SSRs, and only *S. pubescens subsp. Microphylla* contained a hexanucleotides SSR. The SSRs of the

three species were located in the protein coding or non-coding regions, and no SSR sites were found on tRNA or rRNA (Table S5). Among the eight species of *Syringa*, the SSRs were primarily located in the LSC region, with fewer distributed in the SSC and IR regions. Some species, such as *S. oblata*, *S. pinnatifolia*, and *S. vulgaris*, have no SSRs in the IR region. Totals of 43, 40, 52 SSRs were found in the LSC region, and four SSR sites were found in the SSC regions of *S. oblata*, *S. pubescens subsp. Microphylla*, and *S. reticulata subsp. Amurensis*. No SSR was found in the IR of *S. oblata*, and one and five SSRs were found in the IRs of *S. pubescens subsp. Microphylla* and *S. reticulata subsp. Amurensis*, respectively (Fig. 4). The complex nucleotide SSRs of the three species were all located in the non-coding region of the LSC region, and two trinucleotide ((TTA)₄(ATA)₄) SSRs of *S. reticulata subsp. Amurensis* only appeared in the IRa/IRb region. The trinucleotide SSRs of *S. oblata* and *S. pubescens subsp. Microphylla* occurred in the LSC region,

Fig. 4 cpSSR distribution statistics of eight species of *Syringa*

and the four hexanucleotide SSRs of *S. pubescens* subsp. *Microphylla* all appeared in the IRA/IRb region. No hexanucleotides were detected in the chloroplast genomes of the other two species. SSRs mainly appear in several genes, such as *matK*, *rps16*, *atpF*, *rpoC2*, *ycf3*, *atpB*, *rps12*, *ycf1*, *ycf2*, and *rpl16*. No SSRs were found in the *ycf1*, *ycf2*, and *rpl16* genomes of lilac. No SSRs were found in the *ycf2* and *rpl16* genes of *S. reticulata* subsp. *Amurensis*.

Analysis of variations in the chloroplast genome

Many variations are found in chloroplast genome sequences. The variations in the genome are mainly divided into three categories: 1. single nucleotide variations, (usually called single nucleotide polymorphisms, that is, the difference in a single DNA base, referred to as a SNP); SNPs only involve the variation of a single base, which can be caused by transition or transversion of a single base, or by the insertion or deletion of bases; 2. small indels (insertion and deletion); that is, insertion or deletion of a small sequence < 50 bp at a certain position in the genome; 3. large structural variations (SVs), including those with lengths ≥ 50 bp, insertion or deletion of long fragments, shift inversion, copy number variations, and some more complicated forms of variation. The second and third types of variation are also called SVs to distinguish them from SNP variation. The analytical results are shown in Table S6 using *S. oblata* as the reference species. The maximum number of SNPs in *S. reticulata* subsp. *Amurensis* was 1,642, with 311 indels. In total, 372 indels were present in *S. pubescens* subsp. *Microphylla*, with 1588 SNPs, indicating that the two are quite different from the chloroplast genome of *S. oblata*, and their genetic relationship is relatively distant; the numbers of SNPs and indels in *S. vulgaris* were 171 and 49 respectively. In traditional taxonomy, *S. reticulata* subsp. *Amurensis* belongs to Sect. *Ligustrina*, whereas the other listed species belong to Sect.

Syringa. *S. vulgaris* and *S. oblata* belong to Ser. *Syringa*, and *S. pubescens* subsp. *Microphylla* belongs to Ser. *Pubescentes* (Schneid.) Lingelsh. The results of variations in the chloroplast genome verified this traditional classification result (Table S6).

Collinearity analysis

A collinearity analysis is performed to determine the local collinearity blocks between genomes through multiple sequence comparisons, and analyze the similarities, rearrangements, and inversions of the collinearity blocks to illustrate the events that occur during species evolution. Blocks of the same color connected by the same line represent similar gene segments of different species. Organisms increase the length of gene segments to follow changes in the evolutionary process. The collinearity analysis of eight species of *Syringa* is shown in the Fig. 4, considering *S. oblata* as the benchmark species. Only one gene segment of *S. persica* was rearranged, and the gene segments of the other seven *Syringa* species increased or decreased significantly compared with those of *S. oblata* (Fig. 5).

Chloroplast variability analysis

Using *S. oblata* as the benchmark species, the Vista tool was used to conduct a genome-wide comparative analysis of 28 Oleaceae species. The results showed that the variation in the non-coding region was greater than that in the coding region, the SC region was larger than the IR region, and the RNA part of the gene was the most conserved. The genes were divided into three categories based on variations in the different genes from different species. The first parts of the genes were quite different among some species. For example, the *accD*, *ycf1*, and *ycf2* genes of some species were quite different from those of lilac. The second parts of

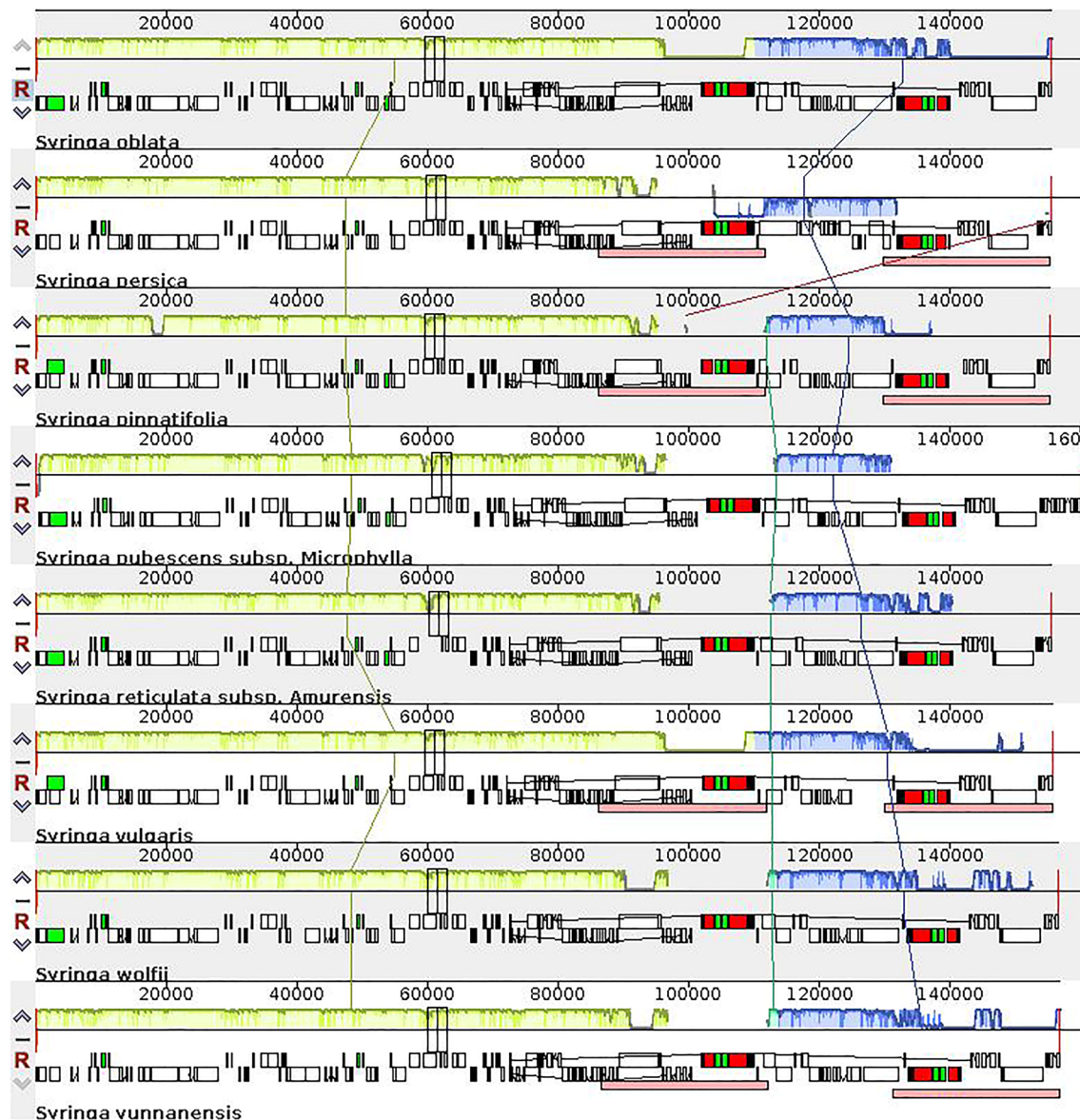


Fig. 5 Collinear analysis of eight species of *Syringa*

the genes were also quite different in all of the listed species, such as the *ndhD* and *psaC* genes. The third part was the more conserved genes, such as *pasA*, *pasB*, and other genes, which differed highly among the listed species. The first and second types of genes can be used for the development of genetic barcodes (Fig. 6).

Phylogenetic analysis

We use *Ginkgo biloba* as the out group and constructed a phylogenetic tree with the chloroplast genomes, CDS regions, introns, LSC regions, and IR region of *S. oblata*, *S. pubescens* subsp. *Microphylla*, *S. reticulata* subsp. *Amurensis* and 26 other species of Oleaceae published at the

NCBI. The five phylogenetic trees all showed that *Syringa* differentiated before *Ligustrum*, indicating that *Ligustrum* was derived from *Syringa* (Fig. 7).

The phylogenetic tree of the CDS region was similar to the phylogenetic tree of the complete chloroplast genome, as both show that *S. persica* and *S. pinnatifolia* were sequentially separated from the whole of Oleaceae, indicating that the two are evolutionarily primitive. *Ligustrum* and some *Syringa* aggregated into one large branch, and the lower branches belonged to *Ligustrum* and *Ser. Villosae* (Schneid.) Rehd of Sect. *Syringa*, including *Syringa wolfii*. *Syringa yunnanensis* and *Ser. Pubescentes* (Schneid.) Lingelsh, juxtaposed with *S. reticulata* subsp. *Amurensis*, which belonged to Sect. *Ligustrina* Rupr. In other branches, *Ser. Syringa*

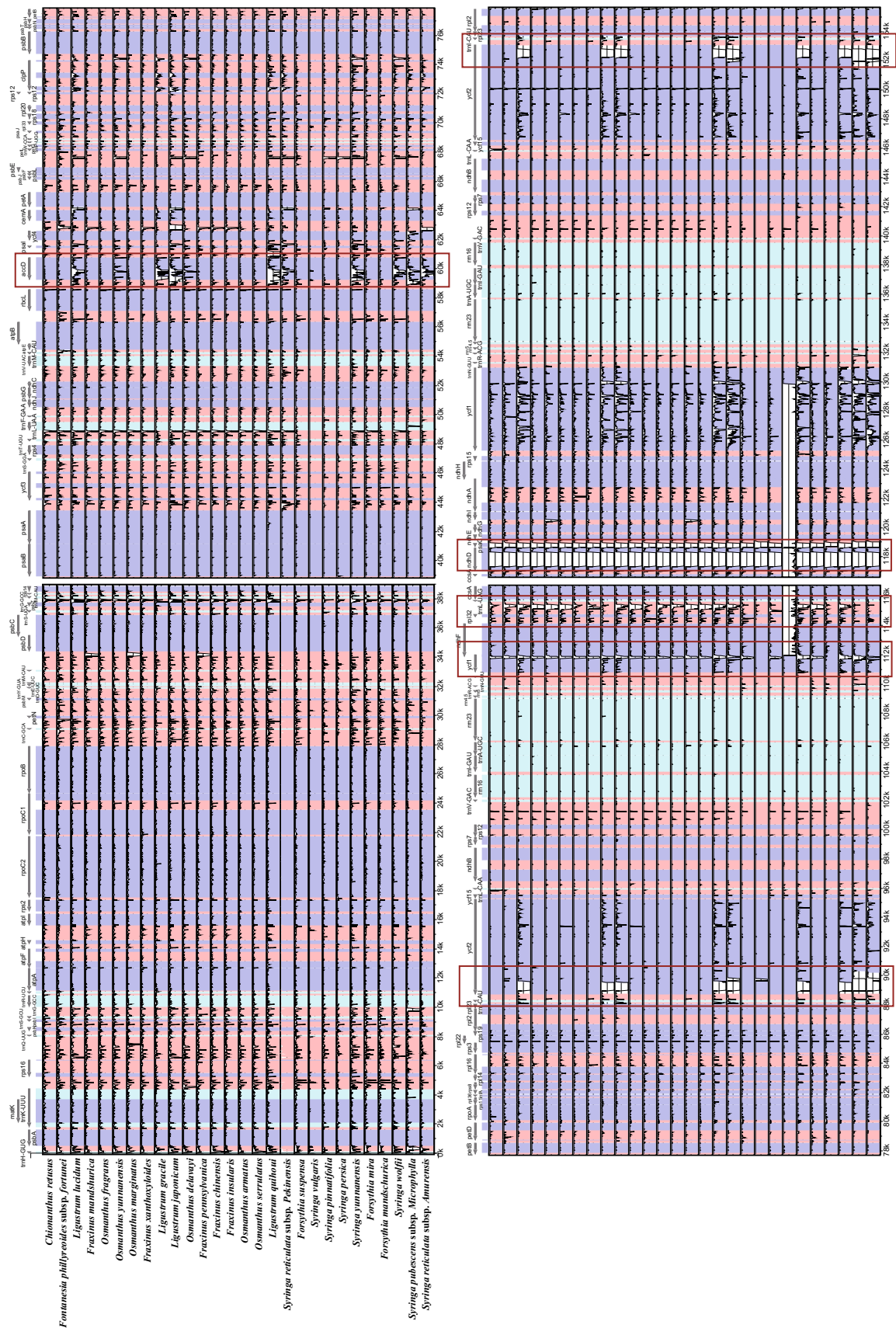


Fig. 6 Comparative analysis of the chloroplast genomes of 29 species of Oleaceae

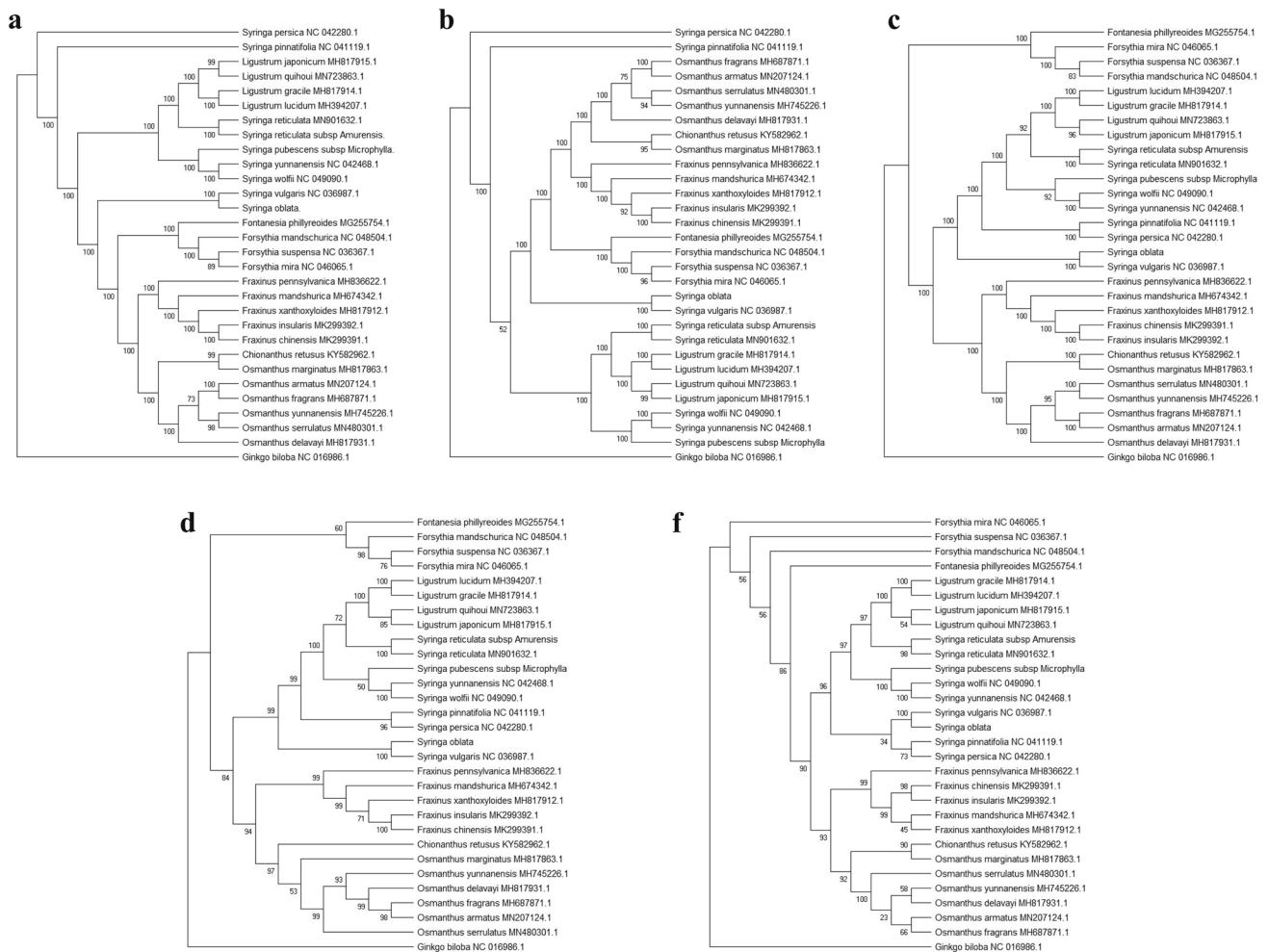


Fig. 7 A phylogenetic tree with *Ginkgo biloba* as the out group was constructed based on five parts of the chloroplast genome of 29 species of Oleaceae

represented by *S. vulgaris* and *S. oblata* was on the same branch with *Fraxinus*, *Forsythia*, and *Osmanthus*, and the relationship was distant from *Ser. Pubescentes* (Schneid.) Lingelsh represented by *S. pubescens* subsp. *Microphylla*, although the two both belonged to *Sect. Syringa* in the existing classification of *Syringa*.

The results of the evolutionary tree of introns and the LSC regions were similar. *Forsythia* was the first to differentiate. *Syringa* and *Ligustrum* aggregated on one branch, and several other genera aggregated on a large branch. Among the *Syringa* and *Ligustrum* branches, *Syringa* differentiated earlier than *Ligustrum*. The sections and series in *Syringa* were aggregated separately. *Sect. Syringa* was divided into groups before *Sect. Ligustrina*. The differentiation order of *Sect. Syringa* is *Ser. Syringa*, *Ser. Pinnatifoliae* Rehd, and *Ser. Villosae* (Schneid.) Rehd. *Ser. Pubescentes* (Schneid.) Lingelsh. In the IR zone of the evolutionary tree, *Forsythia* first differentiated and *Syringa* and *Ligustrum* converged into one branch. *Ser. Pinnatifoliae* Rehd and *Ser. Syringa*

converged into one branch juxtaposed with other lines of *Syringa* and *Ligustrum*. In the lower branch, *Ligustrum* and *Sect. Ligustrina* were juxtaposed with *Ser. Villosae* (Schneid.) Rehd and *Ser. Pubescentes* (Schneid.) Lingelsh on one branch.

The five-part phylogenetic tree shows that *Ligustrum* has the closest relationship with *Syringa* in Oleaceae, and it maybe evolved from *Syringa*, which is older than *Ligustrum*. The existing systematic classification system within *Syringa* has certain problems, and the results support converting the genus *Syringa* into five series.

Discussion

The chloroplast genomes of different species differed in length and the number of encoded proteins, but the range of variation was rather small. The chloroplast genome of most angiosperms is relatively fixed at 12–17 kb in length and

is divided into four partitions. The IR and SSC regions are relatively conserved, and the LSC region is the most variable. The number of genes encoding proteins was relatively fixed at about 85; four rRNAs were present, and GC% content was relatively fixed at about 38%. Studies have shown that some species of Amaryllidaceae [19] and Rosaceae [20] have similar genetic characteristics, such as the length of the chloroplast genome and the number of encoded proteins. The IR region is the most conserved region in the plant chloroplast genome. The length, structure, and boundary between the IR and SC regions are all highly conserved. Expansion and contraction of the IR region are important factors that change the length of the plant chloroplast genome, and they are generally divided into two types. Expansion and contraction of the IR region of most species are reflected in the small deviation in the IR/SC boundary within a few fixed genes. This deviation may cause some genes to become pseudogenes. *Labiateae* [21] and *Vitis* [22, 23] have *ψycf1* and *ψrps19* pseudogenes in the IR/SC boundary, and the IR/SSC boundary was more conserved than the IR/LSC boundary, indicating that the main direction of expansion of the IR region is towards the LSC area. The LSC/IRb boundary generally fell on or near *rps19*, *rpl2*, or *rpl22*. IRb/SSC generally fell on *ycf1* or between *ycf1* and *ndhF*, SSC/IRa generally fell on *ycf1*, and the IRa/LSC boundary generally fell on or near *rps19*, *rpl2*, *rpl12*, and *trnH*. The IRb/SSC boundary of *Ricinus* [24], *Gossypium* [25], *Gossypium* [22] and other species falls on or near the *ycf1* gene, and the SSC/IRa boundary falls on the *ycf1* gene. The IR/LSC boundaries of the chloroplast genomes of the above are different. The lcs/IRb boundary of *Ricinus* falls on *rpl22*, and the IRa/LSC boundary falls on or near *rpl22* and *trnH*. The *Gossypium* and *Vitis* lcs/IRb boundary falls on *rps19*, and the IRa/LSC boundary falls between *rps19* and *trnH*. In this case, expansion of the IR region usually only changes from a few to dozens of base pairs, and does not cause a change in the number of genes. The other is the IR boundary crossing to other genes. For example, expansion and contraction of the IR region of *S. oblate* and *S. reticulata* subsp. *Amurensis* allow the IR boundary to change tens or even a few bps on a few genes, while this region in *S. pubescentis* subsp. *Microphylla* expanded to the LSC region into the *rpl36* gene. Several genes originally located in the LSC region were included in the IR region, resulting in complete replication of the gene, and the total number of genes in the chloroplast genome changed accordingly.

cpSSR molecular markers were first developed by Powell in 1995 as a phylogenetic hotspot [26]. The density of the cpSSRs contained in different regions of the chloroplast genome is different because of the uneven distribution of molecular markers among different taxa. Other characteristics have become excellent material for phylogenetic research. SSRs contain multiple types of repetitive

sequences, including single nucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide, and complex nucleotide repetitive sequences. The results of published plant chloroplast genome analyses show that cpSSRs with a single nucleotide repeat sequence of more than eight bases are more common and more numerous in the non-coding regions of most plants. Jakobsson and Ebert reported that CpSSRs with a single-base repeat sequence of fewer than seven bases cannot be used to distinguish between species [27, 28]. The number of cpSSRs contained in different groups of species was also quite different. Studies have shown that some species of Nymphaeaceae and *Cuscuta* contain fewer cpSSRs, while some species of Cruciferae contain more cpSSRs [29].

Traditional systematic classification is based on the morphological characteristics of *Syringa*. Rehder is divided into two subgenera under *Syringa* (*Sect. Syringa* and *Sect. Ligustrina*) [30]. The current classification divides the genus *Syringa* into two subgenera, namely *Sect. Syringa* and *Sect. Ligustrina*, among which *Sect. Syringa* consists of *Ser. Pinnatifoliae* Rehd., *Ser. Pubescentes* (Schneid.) Lingelsh., *Ser. Syringa*, and *Ser. Villosae* (Schneid.) Rehd. Chen Jinyong [31] reported shape and multivariate analyses that the genus *Syringa* is divided into two sections, two series, 12 species, and 13 subspecies, among which the *Sect. Ligustrina* contains one species, three subspecies, and *Sect. Syringa* contains a terminal inflorescence system and a lateral inflorescence system and includes five species and five subspecies and six species and 5 subspecies, respectively. He Miao and Zhuo Lihuan selected 11 species of *Syringa* from northeastern China. After a phylogenetic analysis based on 25 morphological characteristics, *Sect. Ligustrina* belonged to *Sect. Syringa*, the division of the section was removed, and *Syringa* was directly divided into five series, namely *Ser. Pinnatifoliae* Rehd., *Ser. Pubescentes* (Schneid.) Lingelsh., *Ser. Syringa*, *Ser. Villosae* (Schneid.) Rehd., and *Ser. Ligustrina* Rupr. [32]. Gao Yan et al. studied the relationship between the leaf epidermis characteristics of seven wild species of *Syringa* and the environmental adaptation mechanism and systematics, and inferred that the relationships among *Ser. Pinnatifoliae* Rehd., *Ser. Pubescentes* (Schneid.) Lingelsh., *Ser. Villosae* (Schneid.) Rehd. are closer [33]. The results of Kim's restriction fragment length polymorphism technology using cpDNA and Qin Xiangkun's isozyme technology support division of the genus into two sections [34, 35]; Gao Hongxiao et al. used amplified fragment length polymorphism analysis to remove the division of the sections, and reported that *Syringa* should be directly divided into four series [36]. Li et al. reached a similar conclusion by studying the ITS and ETS of rDNA [37]. In the present study, the phylogenetic results were similar to the result of He Miao and Zhuo Li huan, and directly support dividing *Syringa* into five series.

Conclusion

In this study, the complete chloroplast genome sequences and annotations of *S. oblata*, *S. pubescentis* subsp. *Microphylla*, and *S. reticulata* subsp. *Amurensis* were described and compared with five other species of *Syringa*. A comparative genomics analysis, IR region boundary analysis, collinearity analysis, and codon preference analysis were performed. Using *S. oblata* as the reference species, 30 species of Oleaceae were used as materials to analyze variability in the entire chloroplast genome. A phylogenetic tree was constructed using *G. biloba* as the out group, as well as the CDS regions, IR regions, LSC regions, and introns. The results support the existing classification of Oleaceae, sort out the phylogenetic relationships among the genera in Oleaceae, and sort out the group divisions within *Syringa*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11033-022-08004-w>.

Author contributions This study was designed by CY and JS. Collection and identification of field material were performed by CY and QG. Paper writing and Data analysis were performed by KW and HZ. All the authors read and approved the final manuscript.

Funding We gratefully acknowledge financial support for this research from Key Laboratory of Saline-alkali Vegetation Ecology Restoration (Northeast Forestry University), Ministry of Education (0924210201).

Data accessibility The authors declare that we have submitted the sequence information to NCBI with the numbers MT872639.1 MT872640.1 MT872641.1. We have released the above data.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This experiment does not involve human and animal experiments. The plant materials used in this study were collected from the campus of NORTHEAST Forestry University, and the collection process was in accordance with the regulations of the university and relevant local laws and regulations.

Consent to participate This experiment does not involve human and animal experiments.

Consent for publication All data and manuscripts are published with the consent of all authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Editorial Board of Flora of China (1993) Chinese Academy of Sciences Flora of China. Science Press, Beijing 61:63–81
2. Hagemann R (2004) The sexual inheritance of plant organelles. In: Daniell H, Chase C (eds) Molecular biology and biotechnology of plant organelles. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-3166-3_4
3. Hu YC, Zhang Q, Rao GY (2008) Occurrence of plastids in the sperm cells of caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. Plant Cell Physiol 49:6:958–968
4. Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? J Plant Res 123(2):201–206
5. Corriveau J (1998) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperms species. Am J Bot 75(10):1443–58
6. Lilly JW, Havey MJ, Jiang JJ (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. Plant Cell 13(2):245–254
7. Zhang T, Fang Y, Wang X et al (2012) The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. PLoS ONE 7(1):e30531
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
9. Blake JD, Cohen FE (2001) PaIRwise sequence alignment below the twilight zone. J Mol Biol 307(2):721–735
10. Coil D, Jospin G, Darling AE (2014) A5-MiSeq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31(4):587–589
11. Mikkil S, Stinus L, Ludovic O (2016) Adapterremoval v2.2016: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88
12. Liu C et al (2012) An integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics 13:715
13. Kurtz S, Schleiermacher C (1999) REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 5:426–427
14. Langmead B, Trapnell C, Pop M, Salzberg SL (2019) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
15. Olson SA (2002) Emboss opens up sequence analysis. Brief Bioinform 1:87–91
16. Librado P, Rozas J (2009) DnaSPv5: software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
17. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313
18. Clegg MT, Gaut BS, Learn GH et al (1994) Rates and patterns of chloroplast DNA evolution. Proc Nat Acad Sci 91(15):6795–6801
19. Zhang F, Wang T, Shu X, Wang N, Zhuang W, Wang Z (2020) Complete chloroplast genomes and comparative analyses of *L. chinensis*, *L. anhuiensis*, and *L. aurea* (Amaryllidaceae). Int J Mol Sci 21(16):5729

20. Xue S, Shi T, Luo W, Ni X, Gao Z (2019) Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. Horticult Res. <https://doi.org/10.1038/s41438-019-0171-1>
21. Shen LQ (2018) Complete chloroplast genomes of three Lamiaceae medicinal plants: comparative and evolutionary analysis in Lamiaceae. Zhejiang University, Hangzhou
22. Yang Y, Jiao J, Fan X, Zhang Y, Liu C (2019) Complete chloroplast genome sequence and characteristics analysis of *Vitis ficifolia*. Acta Horticult Sin 46(4):635–48
23. Wang YH (2018) Chloroplast genome sequences of *Vitis* structural organization and phylogeny analysis. Huazhong Agricultural University, Wuhan
24. Muraguri S, Xu W, Chapman M et al (2020) Intraspecific variation within castor bean (*Ricinus communis* L.) based on chloroplast genomes. Ind Crops Prod 155:179
25. Feng K (2013) Chloroplast genome sequence of ten species of *Gossypium*: structural organization and Phylogeny analysis. Chinese Academy of Agricultural Sciences, Beijing
26. Powell W, Morgante M, Mcdevitt R, Vendramin J (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci 92:7759–7763
27. Jakobsson M, Sill T, Lind-Halldén C, Halldén C (2007) Evolution of chloroplast mononucleotide microsatellites in *Arabidopsis thaliana*. Theor Appl Genet 114(2):223–235
28. Ebert D, Peakall R (2009) A new set of universal de novo sequencing primers for extensive coverage of non-coding chloroplast DNA: new opportunities for phylogenetic studies and cpSSR discovery. Mol Ecol Resour 9:777–783
29. Li B, Liu HX (2012) Research advances in chloroplast simple sequence repeat (cpSSR). J Anhui Agricult 13:32–3343
30. Rehder (1945) Notes on some cultivated trees and shrubs. J Arnold Arbor 26:67–78
31. Chen JY, Zhang ZS, Hong DY (2008) Taxonomic revision of *Syringa pinetorum* complex (Oleaceae). J Syst Evolut 46(1):93–95
32. He M, Zhuo LH (2007) A cladistic analysis of *Syringa* from Northeastern China based on the morphologic characters. For Sci Technol 32(2):60–64
33. Gao Y, Cui HX, Shi L, Qu YY (2008) Characteristics of leaf epidermis of *Syringa* and the association with their environmental adaptation and systematics. Acta Bot Boreal-Occident Sin 03:475–484
34. Ki-Joong K, Robert KJ (1998) A chloroplast DNA phylogeny of lilacs (*Syringa*, Oleaceae): *Plastome* groups show a strong correlation with crossing groups. Am J Bot 85(9):1338–1351
35. Qin XK (1996) The use of peroxidases in the systematics of Oleaceae. Acta Botaulca Yunnanica 18(2):159–166
36. Gao HX, Yang K, Liu JB (2001) Analysis of the phylogenetic relationship of *Syringa* by AFLP technique. J China Agricult Univ 16(05):50–54
37. Li J, Alexander JH, Zhang D (2002) Paraphyletic *Syringa* (Oleaceae): evidence from sequences of nuclear ribosomal DNA ITS and ETS regions. Syst Bot 27(3):592–597

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.