Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complete Chloroplast Genome Sequence from Rosa lucieae and Its Characteristics

Weixiang Shen
Southwest Forestry University
Zhanghong Dong
Southwest Forestry University
Wenzhi Zhao
Southwest Forestry University
Luyao Ma
Southwest Forestry University
Fei Wang
Southwest Forestry University
Weiying Li
Southwest Forestry University
Peiyao Xin (xpytgyx@163.com)
Southwest Forestry University

Research Article

Keywords: Rosa, Plastid, Repeat sequence, Positive selection, Phylogenetic analysis

Posted Date: June 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1756651/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Rosa lucieae Franch. & Rochebr. ex Crép. is one of the famous wild ancestors of cultivated roses and plays a very important role in horticultural research, but there is still a lack of research on the R. lucieae chloroplast genome. In this study, we used the Illumina MiSeq sequencing platform for sequencing, assembly and annotation to obtain the sequence information for the R. lucieae chloroplast genome and compared genomics, selection 1 stress analysis, and phylogenetic analysis with 12 other chloroplast genomes of Rosa. The R. lucieae cpDNA sequence has a total length of 156,504 bp and 130 genes are annotated. The length of all 13 studied chloroplast genomes is 156,333~157,385 bp. Their gene content, gene sequence, GC content and IR boundary structure were highly similar. Five kinds of large repeats were detected that numbered 100~116, and SSR sequences ranged from 78 to 90 bp. Four highly differentiated regions were identified, which can be used as potential genetic markers for Rosa. Selection stress analysis showed that there was significant positive selection among the 18 genes. The phylogenetic analysis of R. lucieae and R. cymose, R. maximowicziana, R. multiflora, and R. pricei showed the closest relationship. Overall, our results provide a more comprehensive understanding of the systematic genomics and comparative genomics of Rosa.

Introduction

Rosa lucieae Franch. & Rochebr. ex Crép. is a perennial woody vine of *Rosa* in the family Rosaceae. *R. lucieae* is synonymous with *R. luciae* (Jeon et al. 2019). An additional synonym is *R. wichuriana* Crépin (http://www.floraofalabama.org), which is now revised to *R. wichurana* (http://www.iplant.cn), one of the most famous wild ancestors of cultivated roses (Debener et al. 2009). *R. lucieae* plays an important role in horticultural research, especially in breeding, because of its bright leaves, dense flowers, long flowering period and pleasant aroma, and many horticultural varieties have been cultivated (Lv 2013).

Rosa is a large genus in Rosaceae, with a large number of species, varieties and cultivars. There are approximately 256 species in the genus including 95 species in China, of which 65 species are endemic. It is the modern center of distribution for the genus *Rosa* (http://www.iplant.cn). Many *Rosa* species have strong stress resistance and can survive in harsh conditions. They are often used as constructive species for ecological restoration and vegetation restoration (Jin et al. 2020). At present, there are few reports on the classification and phylogenetic relationships of *Rosa* based on the chloroplast genome. The study of the phylogenetic relationships of *Rosa* plays an important role in the protection, introduction, development and utilization of *Rosa* resources. It also has certain significance for the classification, phylogeny and genetic diversity protection of *Rosa* (Wang et al. 2022). In future research, it will be necessary to gradually sequence the plastoid genome and nuclear genome of species in *Rosa* and build a more complete phylogenetic tree of *Rosa* to clarify the phylogenetic relationships between species in the genus.

Chloroplasts generally exist in some cells of mesophyll and young stems of higher plants, and are also found in algal cells. Chloroplasts have independent genetic information and can semiretain replication. They are very important organelles (Xing et al. 2008). The chloroplast genome consists of four regions: two inverted repeat regions (IRs), a large single-copy region (LSC) and a small single-copy region (SSC). The four regions are connected in the form of covalently closed circular double chains (Raubeson et al. 2005; Jansen et al. 2012). The chloroplast genome is involved in encoding many key proteins in photosynthesis and other metabolic processes (Daniell et al. 2016). Combined with its short genome length, small molecular weight, highly conserved sequence, easy extraction and purification, and many SSR sites, the study of chloroplast genome structure and sequence information is of great value in revealing species' origins, evolution and interspecific genetic relationships (Xing et al. 2008; Liang et al. 2021).

In recent years, the development and application of molecular technology have made rapid progress. Molecular methods have been widely used in plant evolution and phylogeny, for which chloroplast genome sequencing has attracted much attention (Day et al.2014). Researchers have analyzed an increasing number of chloroplast genome sequences. Li et al. (Li

et al. 2021) identified *Prunus sargentii* Rehder Chloroplast genome characteristics and codon usage preference. Dong et al. (Dong et al. 2019) and Qu et al. (Qu et al. 2021) analyzed the characteristics of the chloroplast genome and codon usage bias of *Eriobotrya fragrans* Champ. ex Benth., providing a reference for future research on the evolution and origin of *Eriobotrya* plant genes and the construction of vectors in the transformation system. Su et al. (Su et al. 2021) sequenced and analyzed the chloroplast genome characteristics and phylogenetic relationships of *Lactuca tatarica* (L.) These results provide new evidence and a material foundation for species identification, phylogeny and resource development and utilization of *Mulgedium*. In addition, similar results for *Rubus* (wang et al. 2021; Yu et al. 2022), *Geum* (Li et al. 2020; Zhang et al. 2022), Anacardiaceae (Xin et al. 2021), *Platanus* (Moore et al. 2006), Araceae (Bayly et al. 2013) and other related species have been reported.

The *R. lucieae* chloroplast genome has not been fully analyzed. Matsumoto et al. (Matsumotoa et al. 1998) constructed a maximum likelihood phylogenetic tree for *Rosa* using the *matK* sequence in 1998, and the molecular classification conformed closely to traditional botanical classification. However, the bootstrap confidence of the phylogenetic tree was relatively low, only 51% to 95%. Jeon et al. (Jeon et al. 2019) assembled the chloroplast genomes of *R. multiflora*, *R. maximowicziana* and *R. lucieae* to compare the genomic characteristics of Sect. Synstylae of *subgen. Rosa* and compared them with other subordinate groups. However, the phylogenetic relationships among the above three species have not been inferred because the branch lengths of the phylogenetic tree within the column group are short and the support value is low. The phylogenetic tree constructed by Gao et al. (Gao et al. 2020) using the maximum likelihood (ML) method shows that *R. lucieae* is closely related to *R. maximowicziana.* Zhao et al. (Zhao et al. 2020) also showed the same results.

Here, we use Illumina sequencing technology to show the complete sequence characteristics and codon usage of the *R. lucieae* chloroplast genome and plan to compare and analyze the repeat sequence and SSRs, IR boundary, nucleotide variability values and positive selection of the chloroplast genome of several *Rosa* species to provide a theoretical molecular basis for *R. lucieae* chloroplast genome research and genetic improvement, clarify the phylogenetic relationships between *R. lucieae* and other species of *Rosa* and provide genomic information for the study of the phylogeny and kinship of *Rosa* for further research and applications.

Methods

Taxon Sampling

Fresh young and healthy leaves of *R. lucieae* were collected from Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, wrapped in tin foil and quickly frozen in liquid nitrogen at -80 ℃ until use.

DNA Extraction and Sequencing

Total genomic DNA was extracted using the modified CTAB method (Doyle et al. 1987), and *R. lucieae* chloroplast genome sequencing was performed using the Illumina sequencing platform by Annoroad Gene Technology Co., Ltd., Beijing, China.

Chloroplast Genome Assembly, Gene Annotation and Relative Synonymous Codon Usage

The sequenced data were filtered and screened. The complete chloroplast genome was assembled using GetOrganelle software (Jin et al. 2020), and the chloroplast genome was checked and modified with Bandage (Wick et al. 2015). The *R. lucieae* chloroplast genome (GenBank Accession: MN689791) was downloaded from GenBank as a reference sequence, and Geneious R8.1.3 (Kearse et al. 2012) was used to annotate and manually correct the chloroplast genome of *R. lucieae*. Organellar Genomedra (OGDRAW) v1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw. HTML) (Greiner et al. 2019) was used to perform visual analysis of the genome to obtain the physical map. The assembled and annotated chloroplast genome of *R. lucieae* was uploaded to GenBank (Accession: OK938394). To reduce error, sequences and repetitive genes with sequence lengths less than 300 bp and internal termination codons were removed from 85 CDs (coding DNA

sequences). Finally, 53 gene sequences with AUG as the starting codon and UAA, UAG and UGA as the termination codon were selected for subsequent analysis using CodonW1.4.2 (http://codonw.sourceforge.net).

Repeat Sequence and SSR Analysis

The tandem repeat sequences and scattered repeat sequences of the *R. lucieae* chloroplast genome were analyzed using the online websites REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer)(Kurtz et al. 2001) and Tandem Repeats Finder (https://tandem.bu.edu/trf/trf.html)(Benson et al. 1999), with parameters set to the default values. SSRs were identified using the MISA (https://webblast.ipk-gatersleben.de/misa/) (Beier et al. 2017) online program, with parameters set as 1-10, 2-5, 3-4, 4-3, 5-3 and 6-3 (the first number represents the base number of repeats, and the second number represents the minimum number of repeats). The minimum interval between two SSRs was 100 bp.

Contraction and Expansion of IRs

Twelve *Rosa* species close to *R. lucieae* were selected for IR boundary contraction and expansion analysis. The IR boundary comparison map was drawn using the IRscope (Https://irscope.shinyapps.io/irapp/) online program (Amiryousefi et al. 2018). The parameter was set to the default value.

Sliding Window Analysis

The chloroplast genome sequence was calibrated using MAFFT v.7.129 (Katoh et al. 2013), and DanSP v6.12.03 (Rozas et al. 2017) was used to conduct sliding window analyses and determine the nucleotide diversity (Pi) of 13 chloroplast genome sequences closely related to *R. lucieae* and all 28 chloroplast genome sequences, with the following parameters: 200 bp step size and 600 bp window length.

Positive selection analysis

Twenty-eight chloroplast genome sequences in *Rosa* were used to detect positive selection sites in genes. Phylosuite v1.2.1 (Zhang et al. 2019) was used to extract the CDS in the sequence and align each CDS using the MAFFT plug-in. The aligned CDS must be checked one by one to manually adjust the small error. After all CDSs are adjusted correctly they are concatenated in series to form a supermatrix and export a FASTA format file. The BI tree was built using the CIPERS online website (https://www.phylo.org/portal2/login!input.action)(Miller et al. 2010) the tree file was exported in Newick format using FigTree v 1.4.3 (http://tree.bio.ed.ac.uk/publications/). EasyCodeml v1.21 (Gao et al. 2019) was used to perform positive selection analysis with the site model in the preset mode.

Phylogenetic Analyses

To reconstruct the phylogenetic relationships among *Rosa* species, a total of 27 plastid genome sequences were downloaded from GenBank, and 2 species of *Geum* were selected as outgroups (Table 1). Construction of the phylogenetic tree used maximum likelihood and Bayesian inference (BI) methods. After sequence alignment using MAFFT version 7 software (Katoh et al. 2013), BioEdit software (Hall et al. 1999) was used to correct the alignment results. ML analysis was performed using IQ-TREE v1.6.1 software (Nguyen et al. 2015). In ML interpretation, 70% and above support values are considered well supported and 50% and below are poorly supported values. MrBayes (version 3.2.6) was used for Bayesian inference (Ronquist et al. 2003). jModelTest (version 2.1.10) (Darriba et al. 2012) was used to select the most suitable replacement DNA model for phylogenetic reconstruction. The most suitable model was chosen as “TPM1uf+I+G” (freqA = 0.3143, freqC = 0.1841, freqG = 0.1784, freqT = 0.3233, R (a) [AC] = 1.0000, R(b) [AG] = 1.7321, R(c) [AT] = 0.5192, R (d) [CG] = 0.5192, R(e) [CT] = 1.7321, R(f) [GT] = 1.0000, p-inv = 0.7160, and gamma shape= 1.0510) to construct the phylogenetic tree. Similarly, all phylogenetic analyses were edited using FigTree v1.4.3.

Table1. Summary of complete chloroplast genomes of 28 *Rosa* sequences and 2 *Geum* sequences

Taxon	Accession number	Gene number				Length (bp)				GC(%)
		CDS	tRNA	rRNA	Genome	Genome	LSC	SSC	IR	
R. acicularis	MK714016	84	37	8	130	156527	85673	18748	26053	37.2%
R. banksiae	MK361034	84	37	8	130	156575	85792	18767	26008	37.2%
R. canina	MN661140	85	37	8	130	156501	85653	18742	26053	37.3%
R. chinensis	MH332770	85	37	8	130	156591	85737	18766	26044	37.2%
R. chinensis var. *spontanea*	MG523859	84	37	8	130	156590	85825	18677	26044	37.2%
R. cymosa	MT471268	92	39	8	140	156607	85722	18763	26061	37.2%
R. davurica	MW381769	85	37	8	131	156971	86032	18837	26051	37.2%
R. filipes	MT062883	90	37	8	137	156624	85754	18784	26043	37.2%
R. hybrid	MK947051	84	37	8	130	156989	86227	18816	25973	37.2%
R. kokanica	MW298478	85	37	8	131	156793	85890	18773	26065	37.2%
R. laevigata	MN661139	85	37	8	130	156333	85452	18785	26048	37.3%
R. laevigata var. *leiocarpa*	NC_047418	92	39	8	140	156373	85494	18785	26047	37.3%
R. lucidissima	MK782979	83	37	8	129	156588	85713	18779	26048	37.2%
R. lucieae	OK938394	85	37	8	130	156504	85660	18744	26050	37.2%
R. lucieae	MN689791	85	37	8	130	156504	85661	18743	26050	37.2%
R. lucieae	MH355580	85	37	8	130	156500	85651	18751	26049	37.2%
R. lucieae	MG727864	88	37	8	134	156506	85631	18759	26058	37.2%
R. maximowicziana	MG727865	88	37	8	134	156405	85529	18760	26058	37.2%
R. minutifolia	MT755634	86	39	8	135	157396	86547	18903	25973	37.2%
R. multiflora	MN435990	88	37	8	96	157385	86255	19014	26058	37.2%
R. odorata var. *gigantea*	KF753637	88	40	8	139	156634	85767	18761	26053	37.2%
R. odorata var. *pseudindica*	MK116518	85	37	8	133	156652	85785	18761	26053	37.2%
R. praelucens	MG450565	84	37	8	130	157186	86313	18743	26065	37.2%
R. pricei	MK613354	86	39	8	137	156599	85731	18750	26059	37.2%
R. roxburghii	KX768420	88	39	8	139	156749	85852	18791	26053	37.2%
R. rugosa	MK641521	85	37	8	135	157110	86215	18819	26038	37.2%
R. sterilis (nom. nud.)	NC_053909	84	37	8	130	156561	85701	18746	26057	37.2%
R. xanthina	MT547539	86	39	8	137	157214	86302	18800	26056	37.2%

Geum macrophyllum	MT774132	85	37	8	130	155940	85307	18329	26152	36.6%
Geum rupestre	MG262388	87	39	8	138	155479	85771	18550	25579	36.8%

Results And Discussion

Chloroplast genome characteristics of *R. lucieae*

The results of assembly annotation showed that the total length of the chloroplast genome of *R. lucieae* is 156,504 bp, and the GC content is 37.2%, including 85,660 bp in the LSC region, 26,050 bp in the IR region and 18,744 bp in the SSC region (Fig. 1). There are 130 genes, including 85 coding genes, 37 tRNA genes and 8 rRNA genes. There are 18 genes in the IR region, including six protein coding genes (*rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, *rps12*), eight tRNA genes (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*) and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*). In the *R. lucieae* chloroplast genome, 18 genes contain introns. Among these, eight protein coding genes and six tRNA genes contain one intron, and three protein coding genes (*ycf3*, *clpP and rpsl2*) contain two introns (Table 2).

Using CodonW1.4.2 and the online program CUSP, we analyzed the base composition of 53 CDSs in the chloroplast genome of *R. lucieae* and determined the codon content and termination codons of 20 amino acids from 53 coding genes (Figure 2). The total number of codons in the *R. lucieae* chloroplast genome is 21,371, and there are 30 codons with RSCU > 1. Among these, 29 ended with A and U, accounting for 97%, indicating that the *R. lucieae* chloroplast genome prefers to use synonymous codons ending with A or U.

Table 2. Genes present in the chloroplast genome of *R. lucieae*

Category	Gene group	Gene name	Number
Photosynthesis gene	Photosystem I gene	*psaA*, *psaB*,*psaC*,*psaI*,*psaJ*	5
	Photosystem II gene	*psbA*,*psbB*,*psbC*,*psbD*,*psbE*,*psbF*,*psbH*,*psbI*,*psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	15
	Cytochrome b/f complex gene	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*	6
	ATP synthase gene	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*	6
	NADH dehydrogenase gene	*ndhA*, *ndhB*[C], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	11
	Rubis CO large subunit gene	*rbcL*	1
Self-replication gene	RNA polymerase gene	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*	4
	Ribosomal proteins (SSU) gene	*rps2*, *rps3*, *rps4*, *rps7*[C], *rps8*, *rps11*, *rps12*[A,C], *rps14*, *rps15*, *rps16*, *rps18*, *rps19c*	12
	Ribosomal proteins (LSU) gene	*rpl2*[C], *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*[C], *rpl32*, *rpl33*, *rpl36*	9
	Ribosomal RNAs gene	*rrn4.5*[C], *rrn5*[C], *rrn16*[C], *rrn23*[C]	4
	Transfer RNAs gene	*trnA-UGC*[A,C], *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*[A], *trnH-GUG*, *trnI-CAU*[C], *trnI-GAU*[A,C], *trnK-UUU*[A], *trnL-CAA*[C], *trnL-UAA*[A], *trnL-UAG*, *trnM-CAU*, *trnN-GUU*[C], *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*[C], *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*[C], *trnV-UAC*[A], *trnW-CCA*, *trnY-GUA*	29
Other genes	Translational initiation factor gene	*infA*	1
	Maturase K gene	*matK*	1

Subunit of acetyl-Co A gene	*accD*	1
Envelop membrane protein gene	*cemA*	1
c-type cytochrome synthesis gene	*ccsA*	1
Protease gene	*clpP*	1
Hypothetical chloroplast reading frames （ycf）	*ycf1*^{C}, *ycf2*^{C}, *ycf3*, *ycf4*	4

Note：A and B indicate an intron and two introns in genes，respectively. C indicates two copies of genes.

Repeat sequence and SSR analysis

Six types of SSRs (mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats) were detected using MISA analysis of 13 closely related *Rosa* species (Fig. 3A), and 86 SSRs were found in *R. lucieae*. In the other 12 *Rosa* species the number of SSRs ranges from 78 to 90. The most abundant type of SSR are mononucleotide repeats, from 44 in *R. banksiae* to 56 in *R. sterilis*, followed by dinucleotide repeats, tetranucleotide repeats, trinucleotide repeats, hexanucleotide repeats and pentanucleotide repeats. Further study found that most SSRs are located in the LSC region, followed by the IR and SSC regions (Fig. 3B). Eighty-six SSRs are detected in *R. lucieae*, of which the number of A repeats and T repeats in mononucleotide repeats was the most frequent, accounting for 59.3%, followed by tetranucleotide repeats, accounting for 13.95%, dinucleotide repeats, accounting for 12.79%, and only one pentanucleotide repeat (Fig. 3C). The repeats of 13 *Rosa* species were analyzed. A total of 51 tandem repeats and 50 scattered repeats were found in *R. lucieae*. Among the other 12 *Rosa* species, 100-116 repeats were detected, except that *R. minutifolia* and *R. odorata* do not contain complementary repeat sequences, and all other species contain five types of repeats. Eighteen forward repeats (F), 15 reverse repeats (R), 16 palindromic repeats (P), and 1 complementary repeat (C) were detected (Fig. 3D). Among these, the number of tandem repeats is large, mainly distributed in the LSC region, followed by the IR region and SSC region (Fig. 3E). Among the 51 tandem repeats, six were located in the exon, 2 in the intron and 43 in the intergenic region, accounting for 11.8%, 3.9% and 84.3% of the total repeats, respectively (Fig. 3F), and 28 were located in the LSC region, four in the SSC region and 19 in the IR region, accounting for 54.9%, 7.8% and 37.3%, respectively (Fig. 3G).

Inverted Repeat Contraction and Expansion Analysis

By comparing the expansion and contraction of the IR/SC boundary of 13 *Rosa* chloroplast genomes, it can be seen that the chloroplast genomes of 13 Rosa plants have high similarity on the IR/SC boundary, and the boundary genes are consistent (Fig. 4). The boundary gene between IRb and LSC is *rpl2*, and the boundary gene between SSC and IRa and IRb is *ycf1*. Although the *ycf1* gene of *R. lucieae* did not pass through the IRb/SSC boundary, other species crossed the boundary. Overall, the length and structure of the IR region in the genomes of 13 *Rosa* species are similar.

Fig. 4. IR/SC boundary contraction and expansion of chloroplast genomes of 13 *Rosa* species.

Sliding Window Analysis

DnaSP 6.0 software was used to calculate the nucleotide variation value (π) within 600 bp of the chloroplast genome of *R. sterilis, R. roxburghii, R. lucidissima, R. laevigata, R. filipes, R. chinensis, R. banksiae, R. pricei, R. odorata, R. maximowicziana, R. cymosa,* and *R. minutifolia.* The differences between the thirteen *Rosa* species varied from 0 to 0.00936, with an average of 0.00181, suggesting that their genomic differences are small. However, four highly variable loci with much higher π values (π > 0.007), including *trnK* (*UUU)*, *rps16-trnQ (UUG)*, *trnT (UGU)-trnL (UAA)*, and *ycf1*, were precisely located (Fig. 5A). Among the twenty-eight *Rosa* sequences and the two *Geum* sequences, the π values varied from 0 to 0.01166 with a mean of 0.00284, indicating that the differences among Rosaceae species are larger than those between congeneric species. Four highly variable loci included *rps16-trnQ (UUG)*, *trnT (UGU)-trnL (UAA)*, *psbE-petL* and *ycf1*. (π > 0.010; Fig. 5B).

Positive selection analysis

The nonsynonymous (dN) and synonymous (dS) substitution rates of 78 protein-coding genes in 28 chloroplast genome sequences of *Rosa* were compared. After likelihood ratio test (M1a vs. M2a, M7 vs. M8). The results of the statistical neutrality test showed that 18 genes (*atpF*, *matK*, *ndhD*, *ndhH*, *ndhJ*, *ndhK*, *petB*, *psaA*, *psbA*, *psbB*, *psbC*, *rbcL*, *rpl20*, *rpl23*, *rpoA*, *ycf1*, *ycf2*, *and ycf4*) were in a significantly indigenous positive selection state (Table 3). According to the M8 model, *psaA*, *psbC*, *rbcL*, *rpoA*, *ycf1*, *ycf2*, *and ycf4* contain multiple sites under positive selection, and other genes contain only one site. Among these, the *rbcL* gene and *ycf2* gene reached 9 and 10 positive selection sites, respectively.

Table 3. Positive selected sites detected in the cp genome of the *Rosa*.

Gene Name	M8		Gene Name	M8	
	Selected site	score		Selected site	score
atpF	108L	0.989*	*rpl20*	72N	0.955*
matK	83F	1.000**	*rpl23*	24S	0.960*
ndhD	72R	1.000**	*rpoA*	271Y	0.958*
ndhH	269M	0.971*		326I	0.993**
ndhJ	93G	0.965*		328K	0.964*
ndhK	173N	0.967*		329H	0.951*
petB	2S	1.000**	*ycf1*	615K	0.965*
psaA	148G	0.988*		1460I	0.997**
	209G	0.989*		1768I	0.969*
psbA	155T	0.998**	*ycf2*	933L	0.983*
psbB	494T	1.000*		1997A	0.998**
psbC	280A	0.985		1999V	0.996**
	427A	0.999**		2001S	0.994**
rbcL	91A	0.956*		2006E	0.982*
	225I	1.000**		2007M	0.955*
	249D	0.974*		2009I	0.981*
	255V	0.975*		2010G	0.984*
	279T	0.989*		2011F	0.971*
	309M	0.977*		2012M	0.967*
	340E	0.973*	*ycf4*	141I	0.978*
	365T	0.959*			
	475L	1.000*			

*p < 0.05; **p < 0.01.

Phylogenetic Analysis

Two chloroplast genome sequences of *Geum* in Rosaceae were selected as outgroups, and 28 chloroplast genome sequences of *Rosa* were combined to construct phylogenetic trees using IQ-tree (Fig. 6). The phylogenetic relationships indicate that *R. lucieae* is closely related to *R. maximowicziana*, *R. multiflora*, *R.* cymosa, and *R. pricei*. They belong to Sect. Synstylae and the Sect. Banksianae, followed by a close relationship between *R. odorata* and its varieties. In addition, *R. roxburghii* and *R. banksiae* are independent branches, and *R. praelucens*, *R. davurica*, *R. acicularis, R. kokanica*, *R. hybrid*, *R. minutifolia* and *R. rugosa* are branches. *R. xanthina* is a separate branch. The molecular phylogenetic tree constructed using the maximum likelihood method was basically consistent with the topological complement structure of the BI tree, but the branch support value of the BI tree was high, and the molecular phylogenetic tree constructed by the BI method was selected as the main method (Supplementary Fig. S1). The molecular phylogenetic BI tree topology constructed by CDS with 28 sequences is also basically the same (Supplementary Fig. S2).

Discussion

Comparison of cp genomes in the Rosa species

This study describes the chloroplast genome of *R. lucieae,* an ancient vine ornamental plant. Its quantitative characteristics are similar to those of other reported plants in *Rosa* (Table 1). The largest number of annotated genes in the chloroplast genome of *Rosa* species was 140 (*R. cymosa*, MT471268; *R. laevigata* var. *leiocarpa*, NC_047418), with its CDS also reaching a maximum of 92. Of all annotated genes, the *ycf15* gene was only annotated in *R. multiflora* (NC039989), *R. filipes* (NC053856) and *R. cymose* (NC051550), and the *ycf68* gene was only annotated in *R. multiflora* (NC039989) and *R. cymose* (NC051550) (Jeon et al. 2019; Wang et al. 2021; Ding et al. 2020). Lu et al. (Lu et al. 2017) and Raubeson et al. (Raubeson et al. 2007), discussed whether the *ycf15* and *ycf68* genes are pseudogenes or protein coding genes. In *R. lucieae*, the length of these two genes is short, so they were not annotated. In the study of IR/SC boundaries, *ycf1* and *ycf2* genes are located at the junction of the IR region and LSC and SSC regions and have the same incomplete replication as observed in other studies (Li et al. 2013; Song et al. 2015).

These results are consistent with most other studies. The codons of each gene of the *R. lucieae* chloroplast genome mostly end with A or U, and there is a preference for use, such as in *Medicago truncatula* (Yang et al. 2015), *Pinus massoniana* (Ye et al. 2018), and *Dalbergia odorifera* (Yuan et al. 2021). This shows that there are some similarities in codon preference among different species.

Sliding Window Analysis

In addition to random genetic variation events, some mutations constitute highly variable regions in the genome, namely, mutational hotspots (Shaw et al. 2007). Four highly variable sites were detected in 13 closely related *Rosa* species. Five highly variable regions were detected in 28 chloroplast genome sequences of 22 *Rosa* species. Three regions of the same degree of variability were detected twice, namely, *rps16-trnQ (UUG)*, *trnT (UGU)-trnL (UAA)* and *ycf1.* Six highly variable regions were detected in Ji et al.'s (Jeon et al. 2019) study of chloroplast genome mutation hotspots in *Rosa* plants, two of which were consistent with the results of this study, namely, *rps16-trnQ (UUG)* and *ycf1*. The results of our study are similar to those of Jeon et al. (0.7% and 0.6%) in terms of nucleotide variation. These highly variable loci can be used for phylogenetic studies of the *Rosa* DNA barcode and at the species level.

Positive selection analysis

Nonsynonymous substitution (Ka) and synonymous substitution (Ks) and their ratio (Ka/Ks), similar to (dN/dS), have been used to assess the natural selection pressure and evolution rate of nucleotides (Ninio et al. 1984; Yang et al. 2000). In this study, the genes identified as positive selection sites were the ATP synthase gene (*atpF*), Maturase K gene (*matK*), NADH dehydrogenase gene (*ndhD, ndhH, ndhJ, ndhK*), Cytochrome b/f complex gene (*petB*), Photosystem I gene (*psaA*), Photosystem II gene (*psbA, psbB, psbC*), Rubiscolarge subunit gene (*rbcL*), Ribosomal proteins (LSU) gene (*rpl20, rpl23*), RNA polymerase gene (*rpoA*), and hypothetical chloroplast reading frames (*ycf1, ycf2, ycf4*). The amino acid changes from site mutation, caused by selection pressure, can drive evolution within a specific classification pedigree (Nawae et al. 2020). In the process of positive selection favorable amino acid changes increase plant adaptation to ecological habitats (Sen et al. 2011). Compared with other genus studies, positive selection of multiple loci was found in *Rosa* and many genes were involved (Rono et al. 2020; Sheng et al. 2021; Huang et al. 2020; Xie et al. 2018). It is speculated that the reason is that most *Rosa* plants are widely welcomed as ornamental plants. To obtain better characteristics, such as color and taste, *Rosa* plants have undergone many introductions and hybridizations. The occurrence of an abnormal increase in positive selection is a formal genetic change to adapt to diverse climate and environmental conditions (https://www.britannica.com). Many positive selection genes found in this study were also found to have positive selection in other plants and to be involved in the adaptive evolution of plants. These include *matK*, *atpF*, *psbA*, *ycf*, *ycf2*, and *rbcL* (Bock et al. 2014). For example, several studies have found that adaptive evolution of the *rbcL* gene is related to

photosynthetic performance under changes in temperature, drought and carbon dioxide concentrations (Sheng et al. 2021; Galmes et al. 2014; Kapralov et al. 2012). The findings in this study are consistent with previous studies, and nine positive selection sites were found in the *rbcL* gene. The other two genes with more positive selection sites, *ycf2* and *ycf1*, play a key role in cell viability (Drescher et al. 2000). Kikuchi et al. (Kikuchi et al. 2013) observed that the *ycf1* gene was involved in the synthesis of endometrial complexes for protein transport. In addition, the positive selection of the photosynthetic genes *rbcL*, *ndh* and *psb* was related to the adaptation of rice to different sunlight levels (Gao et al. 2019). It is speculated that the positive selection of the same gene in *Rosa* is also related to the level of sunlight. These results can provide a data reference for studying the adaptive evolution of *Rosa* plants.

Phylogenetic Analysis

According to the *Flora of China* (http://www.iplant.cn), *Rosa* is divided into nine groups (Sect. Pimpinellifoliae DC., Sect. Rosa, Sect. Cinnamomeae DC., Sect. Chinenses DC. ex Ser., Sect. Synstglae DC., Sect. Banksianae Lindl., Sect. Laevigatae Theory, Sect. Braeteatae Theory, Sect. Mierophyllae Crep.) and seven series (Ser. Spinosissimae Yu et Ku, Ser. Sericeae (Crep) Yu et Ku, Ser. Beggerianae Yu et Ku, Ser. Cinnamomeae Yu et Ku, Ser. Webbianae Yu et Ku, Ser. Multiflorae Yu et Ku, Ser. Brunoaianae Yu et Ku) according to their external morphology, internal anatomical characteristics, geographical distribution and paleontology. However, in this study, the inferred phylogenetic relationships were not consistent with the above groupings. For example, *R. cymosa* and *R. banksiae* belong to Sect. Banksianae, but their evolutionary relationship is distant. The evolutionary relationship between seedless *R. sterilis* and *R. chinensis* is close, but they belong to Sect. Chiinenses DC. and Sect. Microphyllae Crep., respectively, far from *Rosa roxburghii*, and both belong to Sect. Microphyllae Crep. This shows that the genetic relationships obtained from traditional plant classification and those based on DNA are different. The latter, by analyzing the genetic variation of plastid genome sequences, infers evolution among plant groups and explores their phylogenetic relationships, playing an important role in revealing plant systematics and evolution (Zhu 2014). The phylogenetic tree shows that *R. lucieae* (MG727864) is closely related to *R. maximowicziana*, which is consistent with the research results of Zhao et al. (Zhao et al. 2020) and Gao et al. (Gao et al. 2020).

Conclusions

In this study, the whole genome sequence of *R. lucieae* chloroplasts was sequenced and assembled, and a physical map of the *R. lucieae* chloroplast genome was obtained. The repetitive sequences, IR boundaries, codons and SNPs of the chloroplast genomes of 13 species with close genetic relationships in *Rosa* were compared and analyzed. Positive selection analysis of 28 chloroplast genome sequences in *Rosa* was carried out and a phylogenetic tree was constructed to clarify the genetic relationships of *R. lucieae* within *Rosa*. These studies provide new references for species identification, marker development and utilization, genetic breeding and phylogenetic evolution of *R. lucieae* and provide a more comprehensive understanding of the systematic genomics and comparative genomics of *Rosa*.

References

Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics (Oxford, England). 2018; 34(17):3030-3031 https://doi.org/10.1093/bioinformatics/bty220

Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, et al. Chloroplast genome analysis of Australian eucalypts—*Eucalyptus, Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). Molecular Phylogenetics and Evolution. 2013; 69(3):704–716. https://doi.org/10.1016/j.ympev.2013.07.006

Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017; 33(16):2583-2585. https://doi.org/10.1093/bioinformatics/btx198

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res.. 1999; 27(2):573-580. https://doi.org/10.1093/nar/27.2.573

Bock DG, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants. Mol.Ecol. 2014; 23:4899–4911. https://doi.org/10.1111/mec.12920

Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016; 17(1): 134. https://doi.org/10.1186/s13059-016-1004-2

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012; 9(8):772-772. https://doi.org/10.1038/nmeth.2109

Day PD, Madeleine B, Laurence H, Fay MF, Leitch AR, Leitch IJ, et al. Evolutionary relationships in the medicinally important genus *Fritillaria* L. (Liliaceae). Molecular Phylogenetics and Evolution. 2014;80:11-19. https://doi.org/10.1016/j.ympev.2014.07.024.

Debener T, Linde M. Exploring complex ornamental genomes: The Rose as a model plant. CRC Crit Rev Plant Sci. 2009; 28(4):267-280. https://doi.org/10.1080/07352680903035481

Ding MY, Liao M, Liu PQ, Tan GQ, Chen YQ, Shi S. The complete chloroplast genome of *Rosa cymosa* (Rosaceae), a traditional medicinal plant in South China. Mitochondrial DNA B Resour. 2020; 5(03):2571-2572. https://doi.org/10.1080/23802359.2020.1781563

Dong ZH, Qu SH, Liu C, Ye P, Xin PY. The complete chloroplast genome sequence of *Eriobotrya fragrans*. Mitochondrial DNA Part B. 2019; 4(2):3549-3550. https://doi.org/10.1080/23802359.2019.1676675

Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 1987; 19:11-15. https://doi.org/10.1016/0031-9422(80)85004-7

Drescher A, Ruf S, Calsa T, Jr Carrer H, Bock R, The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 2000; 22:97-104. https://doi.org/10.1046/j.1365-313x.2000.00722.x

Galmes J, Andralojc PJ, Kapralov MV, Flexas J, Keys AJ, Molins A, et al. Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus *Limonium* (Plumbaginaceae). New Phytologist, 2014; 203:989–999. https://doi.org/ 10.1111/nph.12858

Gao CW, Wu CH, Zhang Q, Wu MX, Chen RR, Zhao YL, et al. Sequence and phylogenetic analysis of the chloroplast genome for *Rosa xanthina*. Mitochondrial DNA Part B. 2020; 5(3): 2940-2941. https://doi.org/10.1080/23802359.2020.1792369

Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecology and Evolution. 2019; 9(7):3891-3898. https://doi.org/10.1002/ece3.5015

Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, et al. Evolution of Oryza chloroplast genomes promotedad aptation to diverse ecological habitats. Commun.Biol. 2019; 2:1-13. https://doi.org/10.1038/s42003-019-0531-2

Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research. 2019; 47(W1):59-64. https://doi.org/10.1101/545509

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 1999; 41:95-98. https://doi.org/10.1021/bk-1999-0734.ch008

Huang SN, Ge XJ, Cano A, Salazar BGM, Deng YF. Comparative analysis of chloroplast genomes for five *Dicliptera* species(Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. PeerJ. 2020; 8:e8450. http://doi.org/10.7717/peerj.8450

Jansen RK, Ruhlman TA. Plastid genomes of seed plants. genomics of chloroplasts and mitochondria. Dordrecht: Springer; 2012. https://doi.org/10.1007/978-94-007-2920-9_5

Ji-Hyeon Jeon, Seung-Chul Kim. Comparative analysis of the complete chloroplast genome sequences of three closely related east-asian wild roses (Rosa sect. Synstylae; Rosaceae). Genes. 2019; 10(1). https://doi.org/10.3390/genes10010023

Jin J, Jin P, Wu HE, Dong WP,Yang CH, Zhou HY. Investigation and Application of *Rosa* plant Resources in Guizhou. Seed. 2020; 39(8):61-65,69. https://doi.org/10.16590/j.cnki.1001-4705.2020.08.061

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology. 2020; 21(1):1-31. https://doi.org/10.1186/s13059-020-02154-5

Kapralov MV, Smith JAC, Filatov DA. Rubisco evolution in C4 eudicots: an analysis of Amaranthaceae sensu lato. PLoS One. 2012; 7: e52974. https://doi.org/10.1371/journal.pone.0052974

Katoh K, Standley D. MAFFT multiple sequence alignment software version improvements in performance and usability. Molecular Biology & Evolution. 2013; 30:772-780. https://doi.org/10.1093/molbev/mst010

Kearse M, Moir RD, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12):1647-1649. https://doi: 10.1093/bioinformatics/bts199

Kikuchi S, Bedard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, et al.. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 2013; 339:571-574. https://doi.org/10.1126/science.1229262

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res.. 2001; 29(22):4633-4642. https://doi.org/10.1093/nar/29.22.4633

Li M, Ye Q, Song YF, Wu SH, Yi XG, Wang XR. The Analysis of the Codon Usage Bias in the Chloroplast Genome of *Prunus sargentii*.Molecular Plant Breeding. 2021. https://kns.cnki.net/kcms/detail/46.1068.S.20210728.0854.002.html

Li QQ, Wen J. The complete chloroplast genome of *Geum macrophyllum* (Rosaceae: Colurieae). Mitochondrial DNA Part B. 2021; 6(02):297-298. https://doi.org/10.1080/23802359.2020.1861562

Li R, Ma PF, Wen J, Yi TS. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PLoS ONE. 2013;8(10):e78568. https://doi.org/10.1371/journal.pone.0078568.

Liang X, Li P, Xu DM, Jia XY, Wang WB. Characteristics Analysis of Whole Chloroplast Genome in *Perilla frustescens*. Journal of Shanxi Agricultural Sciences. 2021;49(3):265-272. https://doi.org/10.3969/j.issn.1002-2481.2021.03.02

Lu RS, Li P, Qiu YX. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. Frontiers in Plant Science. 2017; 10(7):2054. https://doi.org/ 10.3389/fpls.2016.02054

Lv JJ. Establishment of plant regeneration system and preliminary studies on genetic transformation system of *Rosa wichuriana* 'Basye's thornless'. M.Sc. Thesis, Huazhong Agricultural University. 2013.

Matsumotoa S, Kouchib M, Yabukib J, Kusunokia M, Uedac Y, Fukuib H. Phylogenetic analyses of the genus *Rosa* using the *matK* sequence: molecular evidence for the narrow genetic background of modern roses. ScientiaHorticulturae. 1998; 77(1-2):73-82. https://doi.org/10.1016/S0304-4238(98)00169-1

Miller MA, Pfeiffer W, Schwartz T. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, LA: Creating the CIPRES science gateway for inference of large phylogenetic trees; 2010. p. 1-8. https://doi.org/10.1109/GCE.2010.5676129

Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, et al. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biology. 2006; 6(17). https://doi.org/10.1186/1471-2229-6-17

Nawae W, Yundaeng C, Naktang C, Kongkachana W, Yoocha T, Sonthirod C, et al. The Genome and Transcriptome Analysis of the *Vigna mungo* Chloroplast. Plants. 2020;9(9):1247. https://doi.org/10.3390/plants9091247

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 2015; 32(1):268-274. https://doi.org/10.1093/molbev/msu300

Ninio J. The neutral theory of molecular evolution: edited by Mooto Kimura Cambridge University Press. Cambridge, 1983 366 pages. FEBS Lett. 1984; 170:210-211. https://doi.org/10.1016/0014-5793(84)81411-8

Qu YY, Xin J, Feng FY, Dong ZH, Qu SH,Wang HY. Codon Usage Bais in Chloroplast Genome of *Eriobotrya fragrans* Champ. Ex Benth. 2021; 36(4):138-144,158. https://doi.org/0.3969/j.issn.1001-7461.202.04.20

Raubeson LA, Jansen RK. Chloroplast genomes of plants, plant diversity and evolution: genotypic and phenotypic variation in higher plants. Cambridge: CABI Publishing; 2005. https://doi.org/10.1079/9780851999043.0045

Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Jeffrey L, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus.BMC Genomics. 2007; 8(174). https://doi.org/10.1186/1471-2164-8-174

Rono PC, Dong X, Yang JX, Mutie FM, Oulo MA, Malombe I. Initial complete chloroplast genomes of *Alchemilla*(Rosaceae): comparative analysis and phylogenetic relationships. Frontiers in Genetics. 2020;11. https://doi.org/10.3389/fgene.2020.560368

Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003; 19(12):1572-1574. https://doi.org/10.1093/bioinformatics/btg180

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 2017;34(12):3299–3302. https://doi.org/10.1093/molbev/msx248

Sen L, Fares MA, Liang B, Gao L, Wang B, Wang T, et al. Molecular evolution of *rbcL* in three gymnosperm families: Identifying adaptive and coevolutionary patterns. Biol. Direct. 2011; 6:29. https://doi.org/ 10.1186/1745-6150-6-29

Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94(3):275-288. https://doi.org/10.3732/ajb.94.3.275

Sheng JJ, Yan M, Wang J, Zhao LL, Zhou FS,Hu ZL, et al.. The complete chloroplast genome sequences of five *Miscanthus* species, and comparative analyses with other grass plastomes. Industrial Crops & Products. 2021; 162.

https://doi.org/10.1016/j.indcrop.2021.113248

Song Y, Dong WP, Liu B, Xu C, Yao X, Gao J, et al. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front. Plant Sci. 2015;6:662. https://doi.org/10.3389/fpls.2015.00662.

Su Y, Liu JJ, Wan B, Zhang PJ, Cheng ZG, Suzuki JJ, et al. Chloroplast Genome Structure Characteristic and Phylogenetic Analysis of *Mulgedium tataricum.* Journal of Agricultural Science and Technology. 2021; 23(6):33-42. https://doi.org/10.13304/J.NYKJDB.2020.1116

Wang QR, HuangZR, Gao CS, Ge YQ, Cheng RB. The complete chloroplast genome sequence of *Rubus hirsutus* Thunb. and a comparative analysis within Rubus species. Genetica. 2021; 149(5-6):299-311. https://doi.org/10.1007/s10709-021-00131-9

Wang SQ, Chen JF, Zhu ZM. The complete chloroplast genome sequence of *Rosa filipes*(Rosaceae). Mitochondrial DNA Part B. 202; 5(02):1376-1377. https://doi.org/10.1080/23802359.2020.1735958

Wang SQ, Zhu ZM. Relationship between species richness patterns of *Rosa* L. and environmental factors in China. Acta Ecologica Sinica. 2022; 42(1). https://doi.org/10.5846/stxb202002280363

Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015; 31(20):3350-3352. https://doi.org/10.1093/bioinformatics/btv383

Xie DF, Yu Y, Deng YQ, Li J, Liu HY, Zhou SD, et al.. Comparative analysis of the chloroplast genomes of the chinese endemic genus *Urophysa* and their contribution to chloroplast phylogeny and adaptive evolution.

Xin YX, Li RZ, LiX, Chen LQ, Tang JR, Qu YY, et al. Analysis on codon usage bias of chloroplast genome in *Mangifera indica.* Journal of Central South University of Forestry & Technology. 2021; 41(9):148-156,165. https://doi.org/10.14067/j.cnki.1673-923x.2021.09.016

Xing SC, Clarke JL. Progress in Chloroplast Genome Analysis. Progress in Biochemistry and Biophysics. 2008; 4(01):21-28.

Yang GF, Su KL, Zhao YR, Song ZB, Sun J. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*. Acta Prataculturae Sinica. 2015; 24(12):171-178. https://doi.org/10.11686/cyxb2015016

Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 2000; 17:32-43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Ye YJ, Ni ZX, Bai TD, Xu L. The analysis of chloroplast genome codon usage bais in *Pinus massoniana*. Genomics and Applied Biology. 2018; 37(10):4464-4471. https://doi.org/10.13417/j.gab.037.004464

Yu JJ, Fu J, Fang YP, Xiang J, Dong HJ. Complete chloroplast genomes of *Rubus* species (Rosaceae) and comparative analysis within the genus. BMC Genomics. 2022; 23. https://doi.org/10.1186/s12864-021-08225-6

Yuan XL, Li YQ, Zhang JF, Wang Y. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera*. Guihaia. 2021; 41(04):622-630. https://doi.org/10.11931/guihaia.gxzw201906012

Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources. 2019; 20(01):348-355 https://doi.org/10.1111/1755-0998.13096

Zhang PP, Wang L, Lu X. Complete chloroplast genome of Geum aleppicum (Rosaceae). Mitochondrial DNA B Resour. 2022; 7(01):234-235. https://doi.org/10.1080/23802359.2021.2024461

Zhao X, Gao CW. The complete chloroplast genome sequence of *Rosa minutifolia*. Mitochondrial DNA Part B. 2020; 5(3):3338-3339. https://doi.org/10.1080/23802359.2020.1817807

Zhu SX. Systematics of chaetoseris and stenoseris (Compositae-Lactuceae). M.Sc. Thesis, Graduate School of Chinese Academy of Sciences (Institute of Botany). 2014.

Figures

Figure 1

Gene map of the chloroplast genome of *R. lucieae*

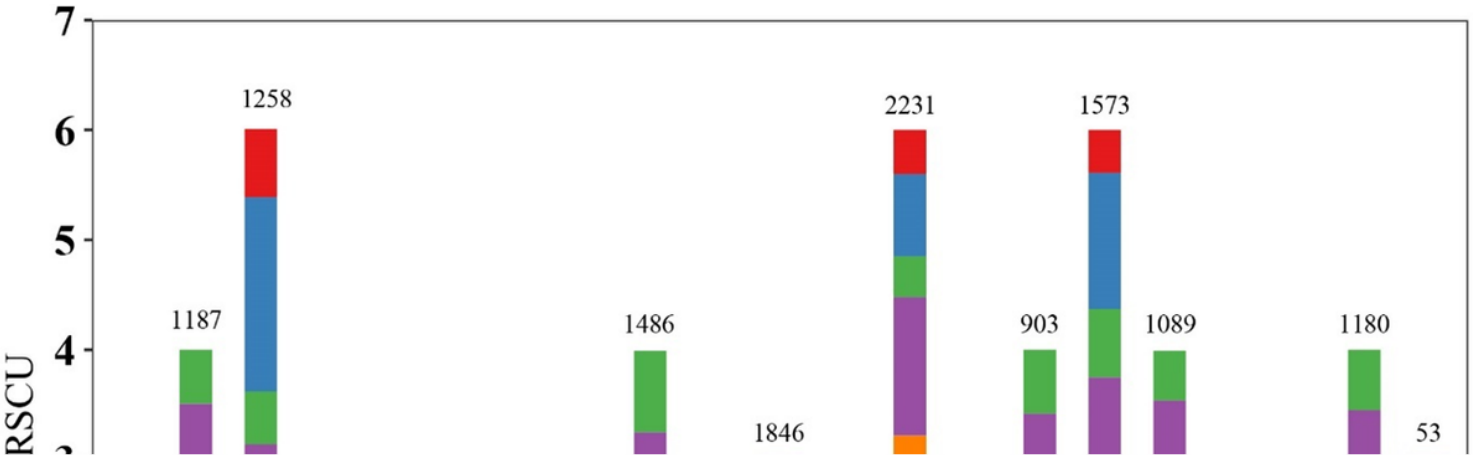

Figure 2

Codon content of 20 amino acids and stop codons in 53 coding genes of the *Rosa lucieae* chloroplast genome. The color of the histogram corresponds to the color of codons.

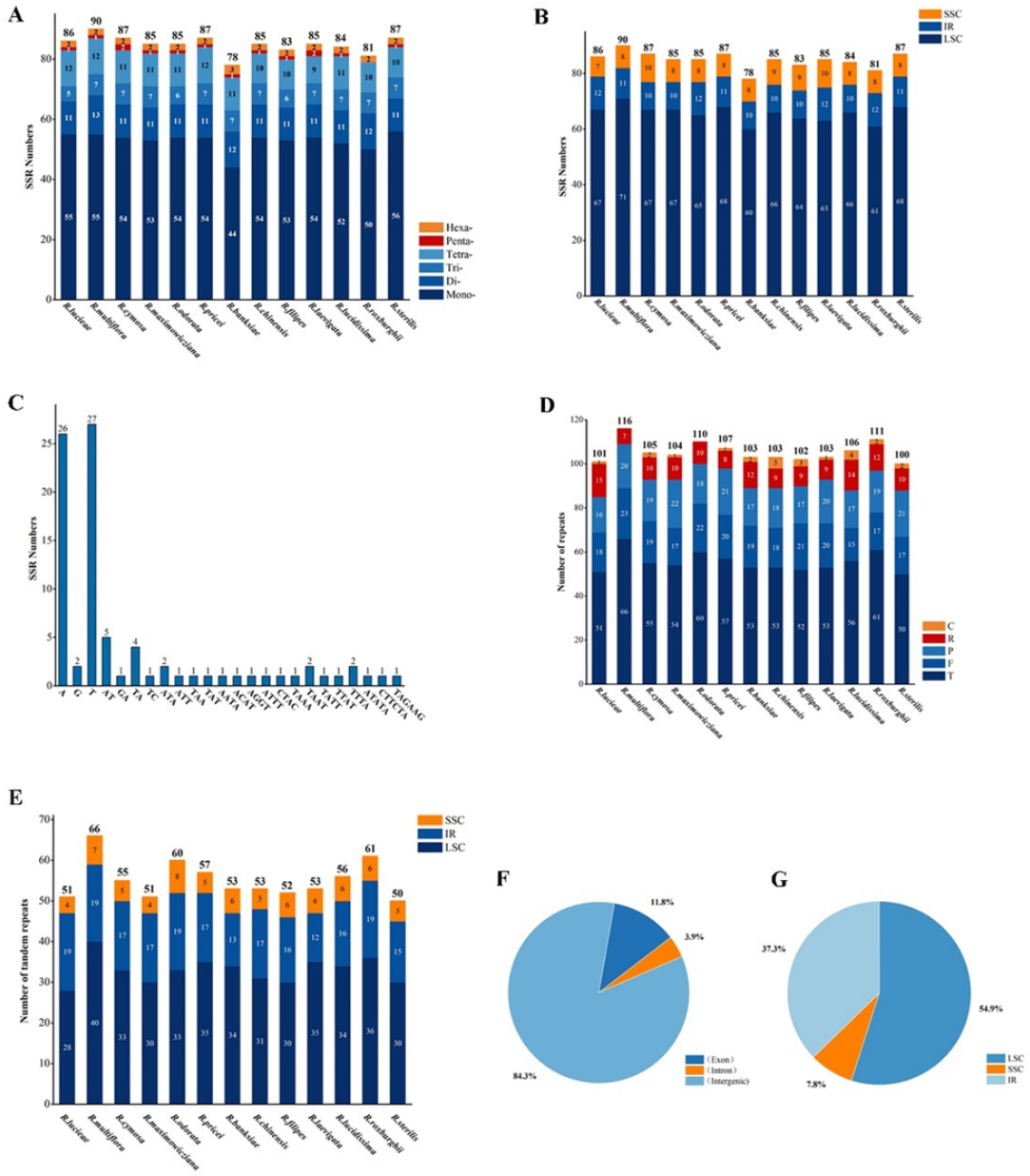

Figure 3

Analysis of sequence repeats in 13 *Rosa* chloroplast genomes. A. Different SSR types detected in 13 genomes. **B.** Distribution frequency of SSRs in the LSC, SSC and IR regions. **C.** Frequencies of SSR motifs of different repeat types in the chloroplast genome of *R. licieae.* **D.** Thirteen large repeat types were detected in the genome. **E.** Distribution frequency of tandem repeats in the LSC, SSC and IR regions. **F.** Distribution frequency of tandem repeats in exon, intron and intergenic regions. **G.** The distribution frequency of tandem repeats in the LSC, SSC and IR regions.

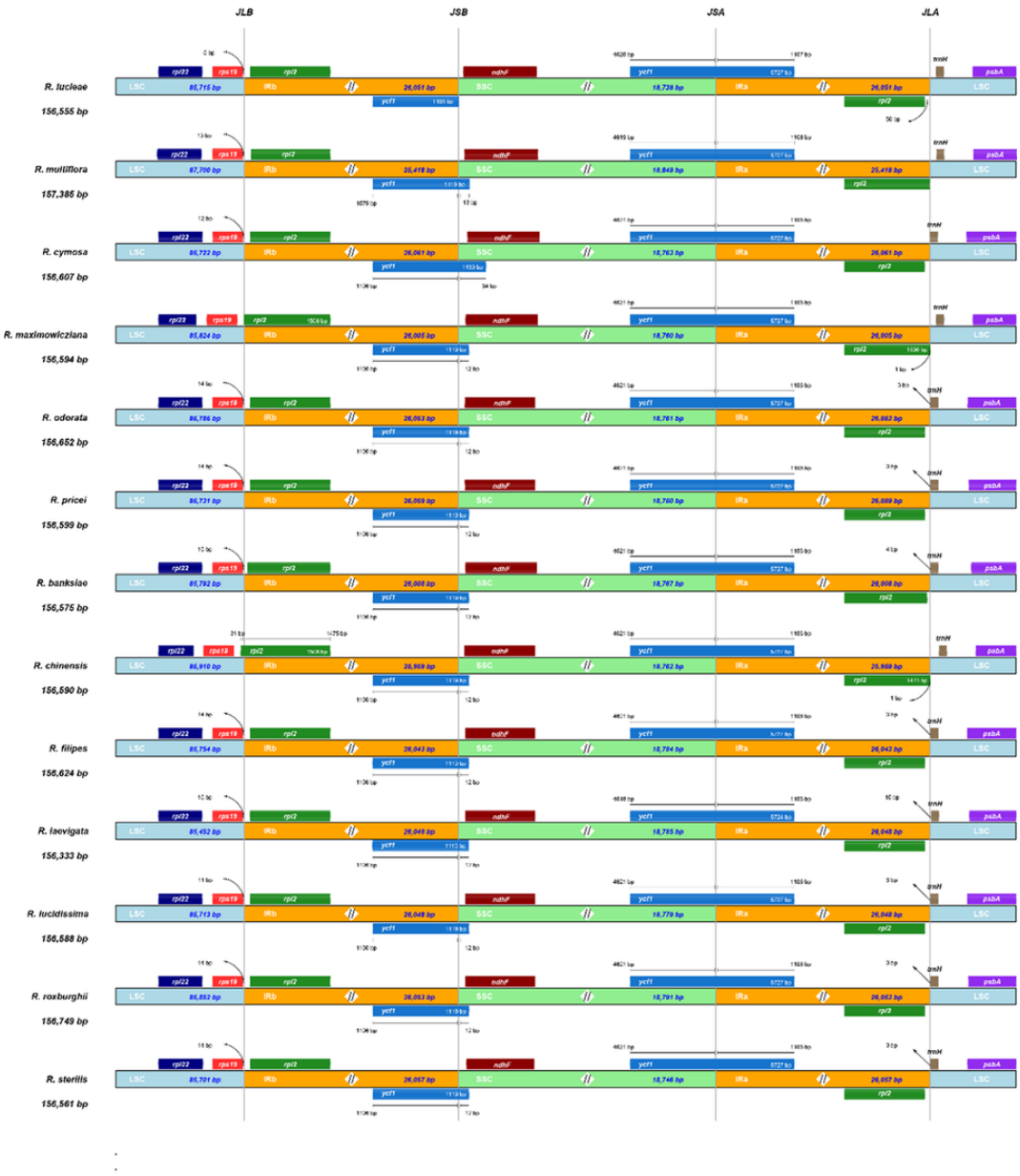

Figure 4

IR/SC boundary contraction and expansion of chloroplast genomes of 13 *Rosa* species.

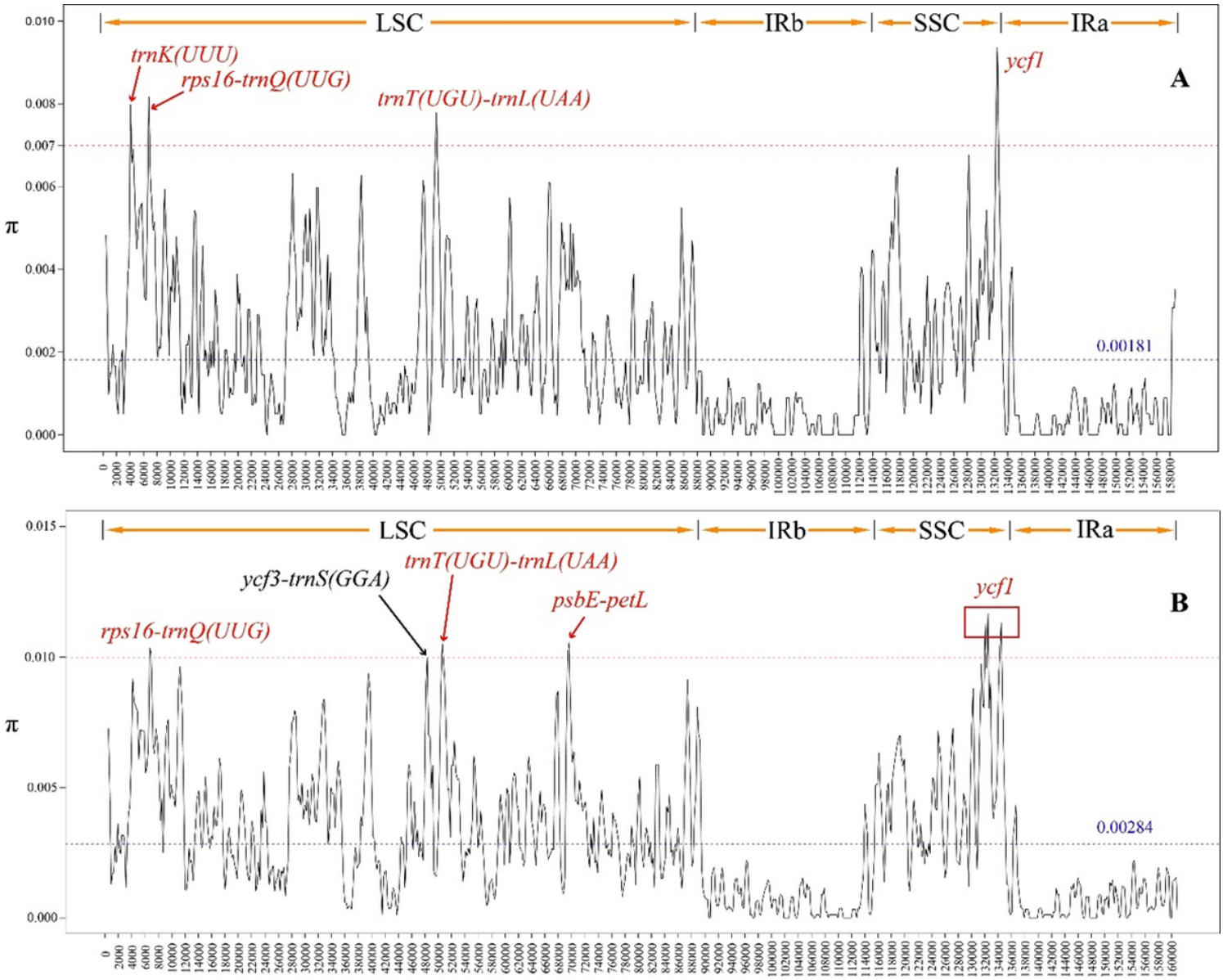

Figure 5

Gene nucleotide variability (pi) values. A. Gene nucleotide variability (pi) values of 13 *Rosa* species closely related to *Rosa lucieae.* **B.** Gene nucleotide variability (pi) values of 28 *Rosa* species and 2 Geum species.

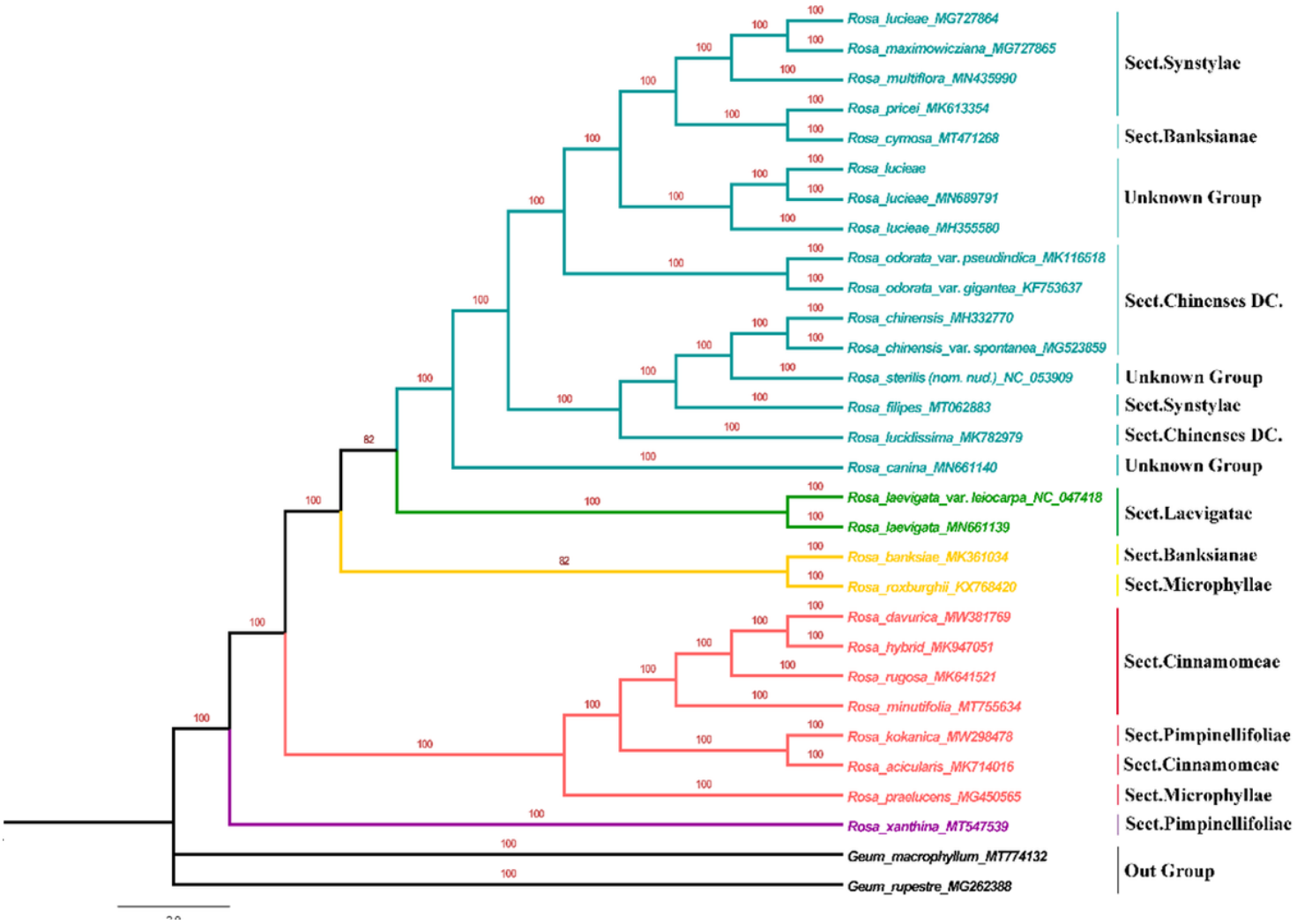

Figure 6

Molecular phylogenetic tree of *Rosa* based on 30 chloroplast genome sequences