horticulturae

MDPI

Article

Complete Chloroplast Genome Sequence of *Rosa lucieae* and Its Characteristics

Weixiang Shen [1,2], Zhanghong Dong [1,2], Wenzhi Zhao [1,2], Luyao Ma [1,2], Fei Wang [1,2], Weiying Li [1,2] and Peiyao Xin [1,2,*]

[1] Southwest Research Center for Landscape Architecture Engineering, National Forestry and Grassland Administration, Southwest Forestry University, Kunming 650224, China
[2] Key Laboratory of Forest Resources Conservation and Utilization in the Southwest Mountains of China Ministry of Education, Southwest Forestry University, Kunming 650224, China
* Correspondence: xpytgyx@163.com

Abstract: *Rosa lucieae* 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 platform for sequencing, assembly, and annotation to obtain the *R. lucieae* chloroplast genome sequencing information and compared genomics, selection 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*.

Keywords: *Rosa*; plastid; repeat sequence; positive selection; phylogenetic analysis

Citation: Shen, W.; Dong, Z.; Zhao, W.; Ma, L.; Wang, F.; Li, W.; Xin, P. Complete Chloroplast Genome Sequence of *Rosa lucieae* and Its Characteristics. *Horticulturae* **2022**, *8*, 788. https://doi.org/10.3390/horticulturae8090788

Academic Editors: Hamidou F. Sakhanokho, Nurul Islam-Faridi and Dilip R. Panthee

Received: 23 June 2022
Accepted: 26 August 2022
Published: 30 August 2022

1. 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* [1]. An additional synonym is *R. wichuriana* Crépin (http://www.floraofalabama.org, accessed on 15 March 2022), which is now revised to *R. wichurana* (http://www.iplant.cn, accessed on 15 March 2022), one of the most famous wild ancestors of cultivated roses [2]. *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 [3].

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, accessed on 15 March 2022). 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 [4]. 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* [5]. 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 semi retain replication. They are very important organelles [6]. 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 [7,8]. The chloroplast genome is involved in encoding many key proteins in photosynthesis and other metabolic processes [9]. 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 [6,10].

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 [11]. Researchers have analyzed an increasing number of chloroplast genome sequences. Li et al. [12] identified *Prunus sargentii* Rehder Chloroplast genome characteristics and codon usage preference. Dong et al. [13] and Qu et al. [14] 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. [15] 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* [16,17], *Geum* [18,19], Anacardiaceae [20], *Platanus* [21], Araceae [22], and other related species have been reported.

The *R. lucieae* chloroplast genome has not been fully analyzed. Matsumoto et al. [23] 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. [1] 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. Cui et al. [24] also reported the chloroplast genome of *R. wichuraiana*; however, except for molecular phylogenetic tree, no other relevant comparative analysis has been done. The phylogenetic tree constructed by Gao et al. [25] using the maximum likelihood (ML) method shows that *R. lucieae* is closely related to *R. maximowicziana*. Zhao et al. [26] 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, 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 more powerful theoretical and molecular basis for further research on *R. lucieae* chloroplast genome. Compared with the previous reports, our work increased the number of chloroplast genome sequences of *Rosa* included in the analysis. In terms of research content, we have added codon research, inverted repeat contraction and expansion analysis and positive selection analysis. In the analysis of repeat sequences, we also enriched the content. In addition, the new phylogenetic relationships between *R. lucieae* and other species of *Rosa* provides powerful evidence for the phylogeny and genetic relationship among various species of *Rosa*.

2. Materials and Methods

2.1. 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 °C until use.

2.2. DNA Extraction and Sequencing

Total genomic DNA was extracted using the modified CTAB method [27], and *R. lucieae* chloroplast genome sequencing was performed using the Illumina sequencing platform by Annoroad Gene Technology Co., Ltd., Beijing, China.

2.3. 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 v1.7.4 (Jin et al., New York, NY, USA) [28], and the chloroplast genome was checked and modified with Bandage [29]. The *R. lucieae* chloroplast genome (GenBank Accession: MN689791) was downloaded from GenBank as a reference sequence, and Geneious R8.1.3 (Biomatters Development Team, New York, NY, USA) [30] was used to annotate and manually correct the chloroplast genome of *R. lucieae*. Organellar Genomedra (OGDRAW) v1.3.1 (Greiner et al., Potsdam-Golm, Germany) [31] was used to perform a 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 coding DNA sequences (CDSs). 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 (John Peden, Nottingham, UK) (http://codonw.sourceforge.net, accessed on 25 March 2022).

2.4. 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 (Kurtz et al., Bielefeld, Germany) (https://bibiserv.cebitec.uni-bielefeld.de/reputer, accessed on 27 March 2022) [32] and Tandem Repeats Finder (Gary Benson, New York, NY, USA) (https://tandem.bu.edu/trf/trf.html, accessed on 27 March 2022) [33], with parameters set to the default values. SSRs were identified using the MISA-web (Beier et al., Gatersleben, Germany) (https://webblast.ipk-gatersleben.de/misa/, accessed on 29 March 2022) [34] 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 the two SSRs was 100 bp.

2.5. 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 (Amiryousefi et al., Helsinki, Finland) (https://irscope.shinyapps.io/irapp/, accessed on 8 April 2022) online program [35]. The parameter was set to the default value.

2.6. Sliding Window Analysis

The chloroplast genome sequence was calibrated using MAFFT v.7.129 (Kazutaka Katoh and Daron M. Standley, Osaka, Japan) [36], and DanSP v6.12.03 (Rozas wt al., Barcelona, Spain) [37] 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.

2.7. 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., California, CA, USA) [38] was used to extract the CDSs in the sequence and align each CDS using the MAFFT plug-in. The aligned CDSs 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 (Miller et al., San Diego, Chile) (https://www.phylo.org/portal2/login!input.action, accessed on 15 April 2022) [39], and the tree file was exported in Newick format using FigTree v1.4.3 (Andrew Rambaut, Edinburgh, United Kingdom) (http://tree.bio.ed.ac.uk/publications/, accessed on 15 April 2022). EasyCodeml v1.21 (Gao et al., Fuzhou, China) [40] was used to perform positive selection analysis with the site model in the preset mode.

2.8. Phylogenetic Analyses

To reconstruct the phylogenetic relationships among *Rosa* species, a total of 27 plastid genome sequences were downloaded from GenBank, and two 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 (Kazutaka Katoh and Daron M. Standley, Osaka, Japan) [36], BioEdit software (Thomas A. Hall, Washington, USA) [41] was used to correct the alignment results. ML analysis was performed using IQ-TREE v1.6.1 software (Nguyen et al., Vienna, Austria) [42]. In the ML interpretation, 70% and above support values are considered well-supported, and 50% and below are poorly supported values. MrBayes v3.2.6 (Ronquist et al., Uppsala, Sweden) was used for Bayesian inference [43]. jModelTest v2.1.10 (Darriba et al., Vigo, Spain) [44] 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 (Andrew Rambaut, Edinburgh, UK).

Table 1. 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	156,527	85,673	18,748	26,053	37.2%
R. banksiae	MK361034	84	37	8	130	156,575	85,792	18,767	26,008	37.2%
R. canina	MN661140	85	37	8	130	156,501	85,653	18,742	26,053	37.3%
R. chinensis	MH332770	85	37	8	130	156,591	85,737	18,766	26,044	37.2%
R. chinensis var. *spontanea*	MG523859	84	37	8	130	156,590	85,825	18,677	26,044	37.2%
R. cymosa	MT471268	92	39	8	140	156,607	85,722	18,763	26,061	37.2%
R. davurica	MW381769	85	37	8	131	156,971	86,032	18,837	26,051	37.2%
R. filipes	MT062883	90	37	8	137	156,624	85,754	18,784	26,043	37.2%
R. hybrid	MK947051	84	37	8	130	156,989	86,227	18,816	25,973	37.2%
R. kokanica	MW298478	85	37	8	131	156,793	85,890	18,773	26,065	37.2%
R. laevigata	MN661139	85	37	8	130	156,333	85,452	18,785	26,048	37.3%
R. laevigata var. *leiocarpa*	NC_047418	92	39	8	140	156,373	85,494	18,785	26,047	37.3%
R. lucidissima	MK782979	83	37	8	129	156,588	85,713	18,779	26,048	37.2%
R. lucieae	OK938394	85	37	8	130	156,504	85,660	18,744	26,050	37.2%
R. lucieae	MN689791	85	37	8	130	156,504	85,661	18,743	26,050	37.2%
R. lucieae	MH355580	85	37	8	130	156,500	85,651	18,751	26,049	37.2%
R. lucieae	MG727864	88	37	8	134	156,506	85,631	18,759	26,058	37.2%
R. maximowicziana	MG727865	88	37	8	134	156,405	85,529	18,760	26,058	37.2%
R. minutifolia	MT755634	86	39	8	135	157,396	86,547	18,903	25,973	37.2%
R. multiflora	MN435990	88	37	8	96	157,385	86,255	19,014	26,058	37.2%

Table 1. *Cont.*

Taxon	Accession Number	Gene Number				Length (bp)				GC (%)
		CDS	tRNA	rRNA	Genome	Genome	LSC	SSC	IR	
R. odorata var. *gigantea*	KF753637	88	40	8	139	156,634	85,767	18,761	26,053	37.2%
R. odorata var. *pseudindica*	MK116518	85	37	8	133	156,652	85,785	18,761	26,053	37.2%
R. praelucens	MG450565	84	37	8	130	157,186	86,313	18,743	26,065	37.2%
R. pricei	MK613354	86	39	8	137	156,599	85,731	18,750	26,059	37.2%
R. roxburghii	KX768420	88	39	8	139	156,749	85,852	18,791	26,053	37.2%
R. rugosa	MK641521	85	37	8	135	157,110	86,215	18,819	26,038	37.2%
R. sterilis (nom. nud.)	NC_053909	84	37	8	130	156,561	85,701	18,746	26,057	37.2%
R. xanthina	MT547539	86	39	8	137	157,214	86,302	18,800	26,056	37.2%
Geum macrophyllum	MT774132	85	37	8	130	155,940	85,307	18,329	26,152	36.6%
Geum rupestre	MG262388	87	39	8	138	155,479	85,771	18,550	25,579	36.8%

3. Results and Discussion

3.1. 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 (Figure 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 6 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 4 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).

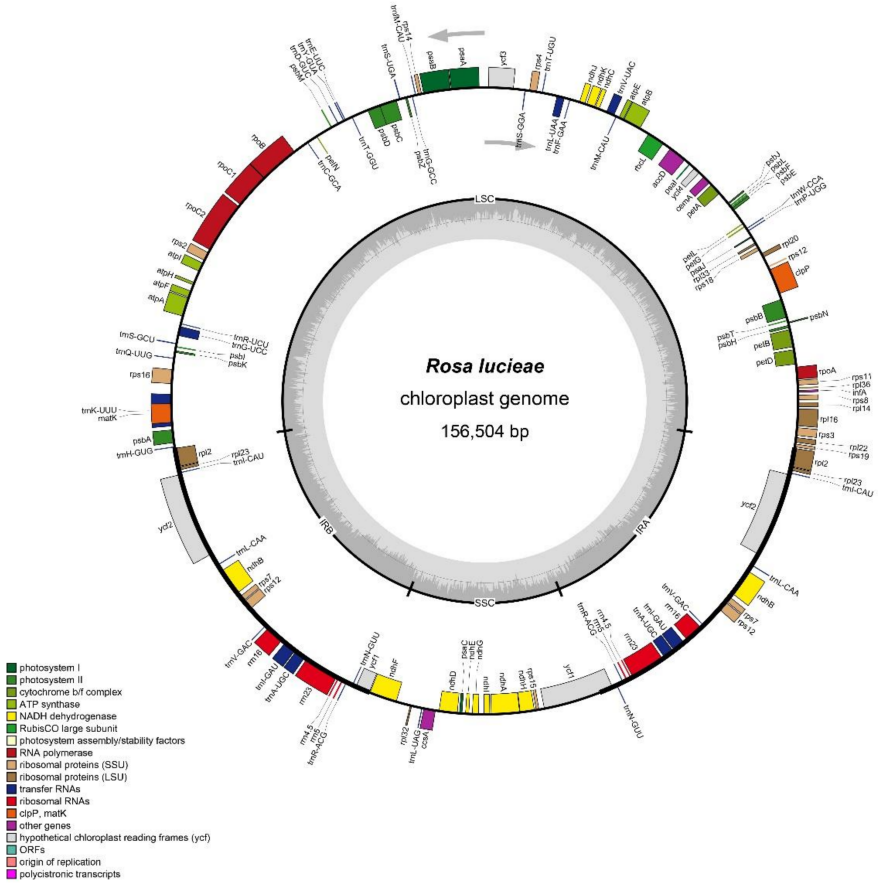

Figure 1. Gene map of the chloroplast genome of *R. lucieae*.

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.

Using CodonW1.4.2 ((John Peden, Nottingham, UK) 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 (Relative synonymous codon usage) > 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.

3.2. 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 (Figure 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 types of SSRs 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 (Figure 3B). Eighty-six SSRs are detected in *R. lucieae*, of which the number of A 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 (Figure 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 (Figure 3D). Among these, the number of tandem repeats is large, mainly distributed in the LSC region, followed by the IR and SSC regions (Figure 3E). Among the 51 tandem repeats, 6 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 (Figure 3F), and 28 were located in the LSC region, 4 in the SSC region, and 19 in the IR region, accounting for 54.9%, 7.8%, and 37.3%, respectively (Figure 3G).

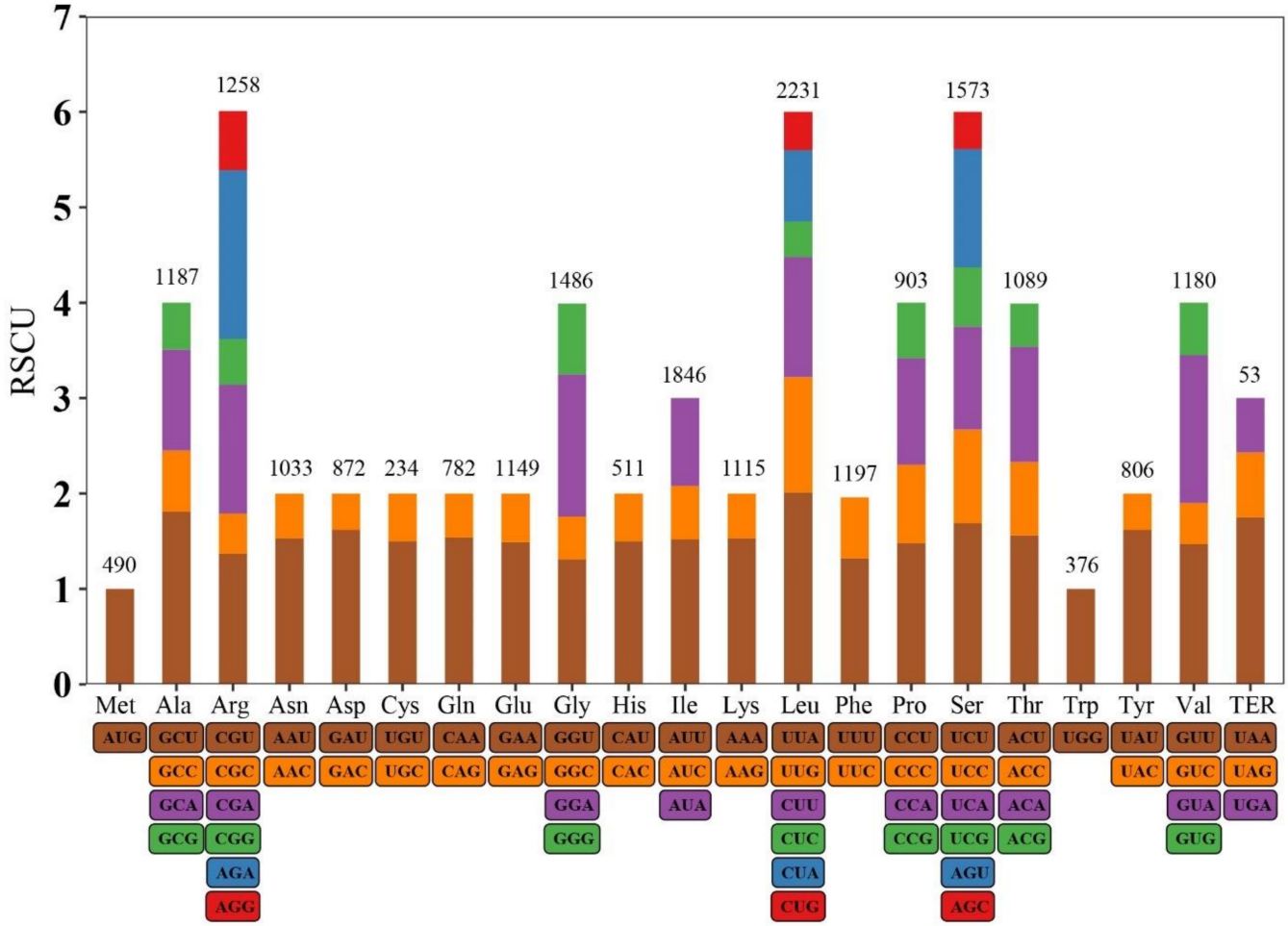

Figure 2. Codon content of 20 amino acids and stop codons in 53 coding genes of the *Rosa lucieae* chloroplast genome. The color of the histogram corresponds to the color of codons.

3.3. 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 (Figure 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.

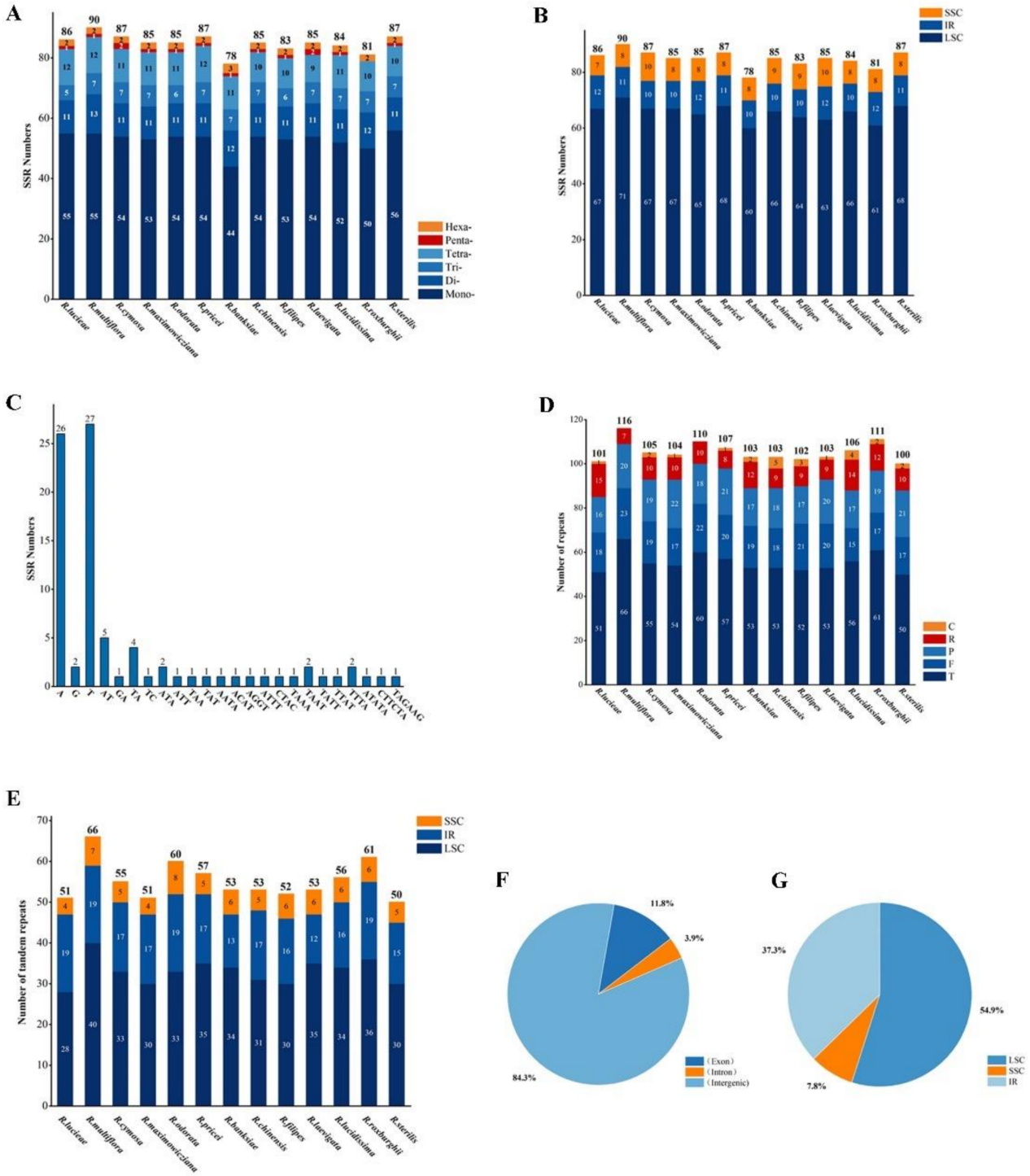

Figure 3. Analysis of sequence repeats in 13 *Rosa* chloroplast genomes: (**A**) different SSR types detected in 13 genomes; (**B**) distribution frequency of SSRs in the LSC, SSC, and IR regions; (**C**) frequencies of SSR motifs of different repeat types in the chloroplast genome of *R. lucieae*; (**D**) 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.

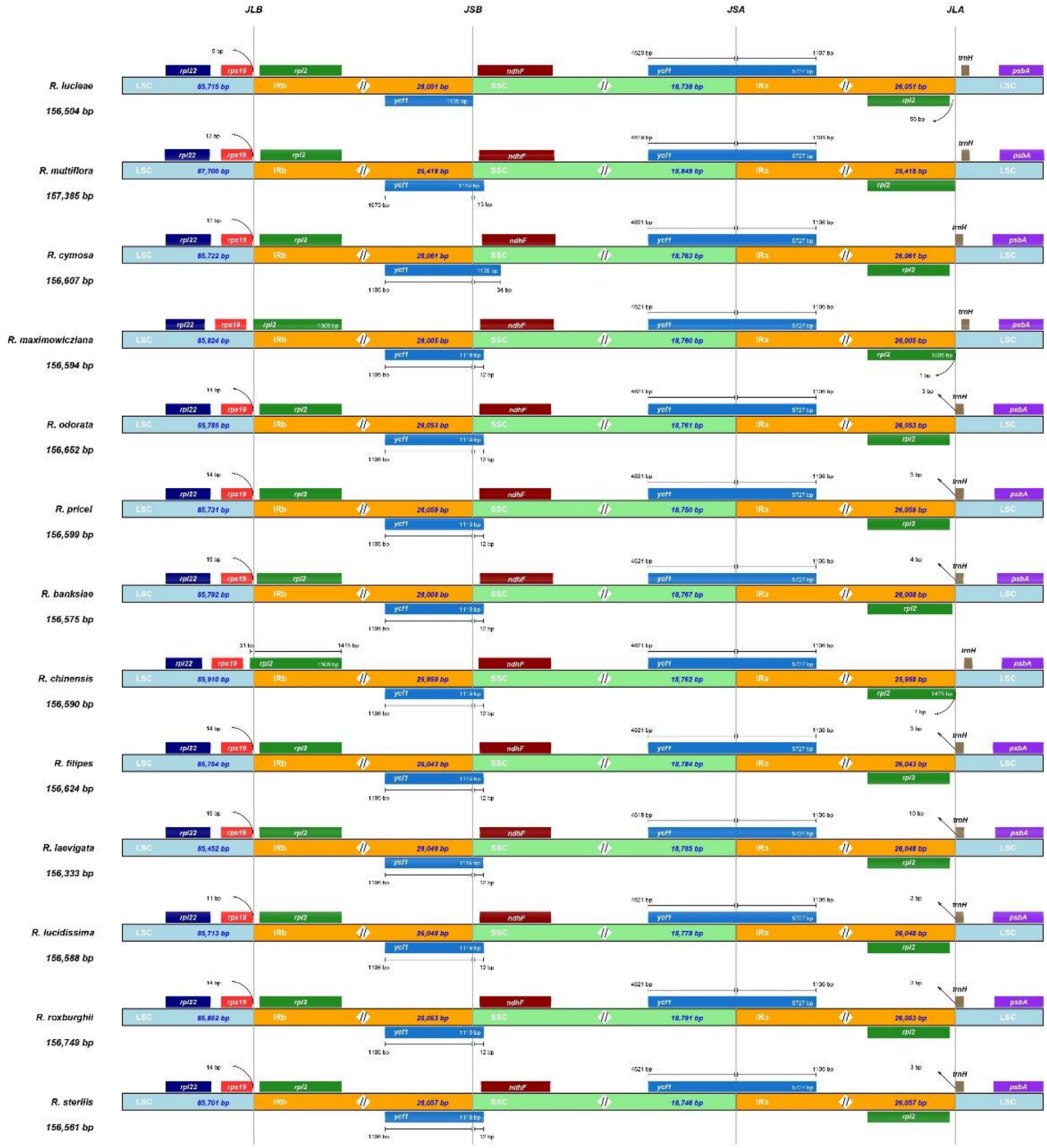

Figure 4. IR/SC boundary contraction and expansion of chloroplast genomes of 13 *Rosa* species.

3.4. Sliding Window Analysis

DanSP v6.12.03 (Rozas wt al., Barcelona, Spain) 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 13 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 ($\pi > 0.007$), including trnK (UUU), rps16-trnQ (UUG), trnT (UGU)-trnL (UAA), and ycf1, were precisely located (Figure 5A). Among the 28 Rosa cp genome sequences and the 2 Geum cp genome 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. ($\pi > 0.010$; Figure 5B).

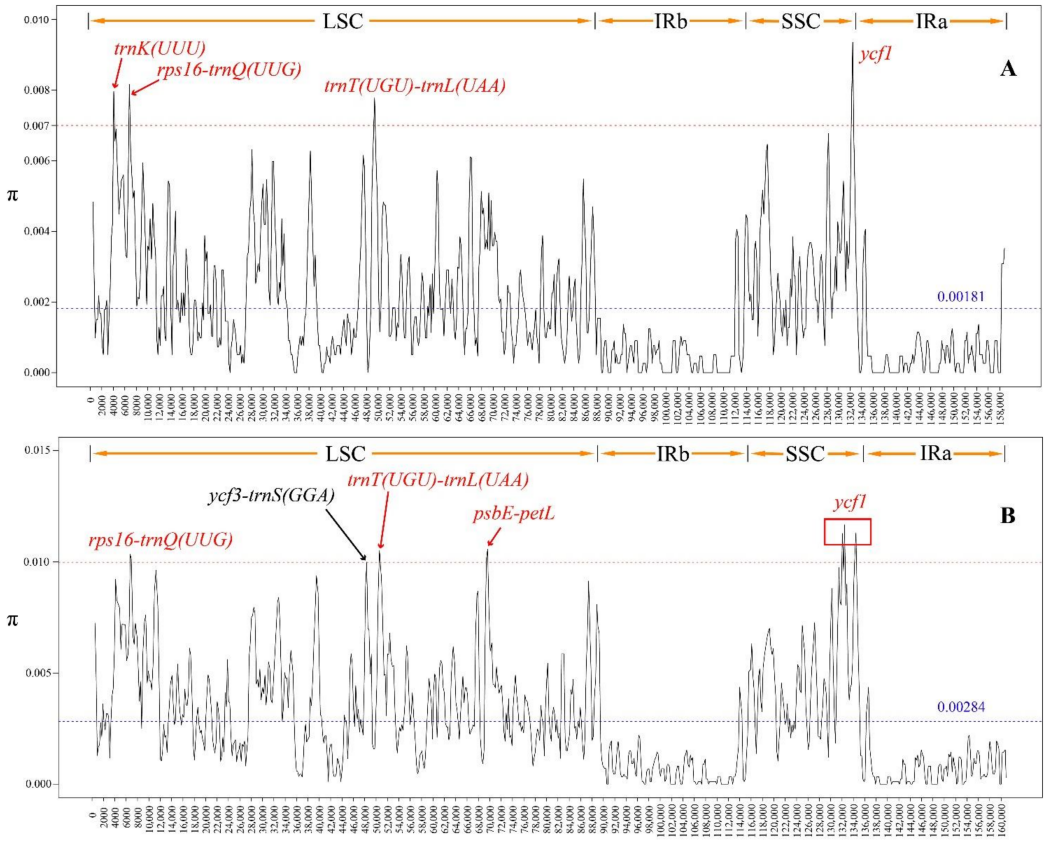

Figure 5. Gene nucleotide variability (pi) values: (**A**) gene nucleotide variability (pi) values of 13 *Rosa* species closely related to *Rosa lucieae*; (**B**) gene nucleotide variability (pi) values of 28 *Rosa* species and 2 Geum species.

3.5. 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 the 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.

3.6. Phylogenetic Analysis

Two chloroplast genome sequences of *Geum* in Rosaceae were selected as outgroups, and twenty-eight chloroplast genome sequences of *Rosa* were combined to construct phylogenetic trees using IQ-tree (Figure 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 (Appendix A Figure A1). The molecular phylogenetic BI tree topology constructed by CDS with 28 cp genome sequences is also basically the same (Appendix A Figure A2).

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 *			

* $p < 0.05$; ** $p < 0.01$.

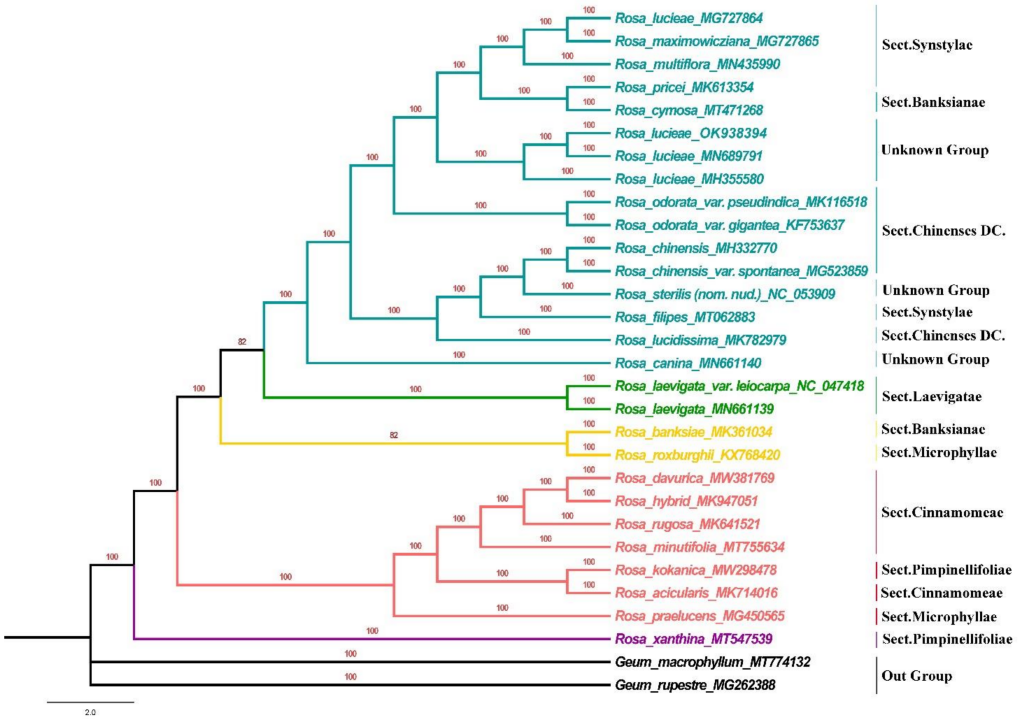

Figure 6. Molecular phylogenetic tree of *Rosa* based on 30 chloroplast genome sequences.

4. Discussion

4.1. 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* species (Table 1). The largest number of annotated genes in the chloroplast genome of *Rosa* species was 140 (*R. cymosa*, MT471268; *R. laevigata* var. *leiocarpa*, NC047418), 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. cymosa* (NC051550), and the *ycf68* gene was only annotated in *R. multiflora* (NC039989) and *R. cymosa* (NC051550) [1,45,46]. Lu et al. [47] and Raubeson et al. [48] 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 [49,50]. Kim et al. believe that the IR/SC boundary variation of the chloroplast genome is the main driving force of chloroplast genome structure variation [51]. We find that the IR/SC boundary of *Rosa* is relatively conservative, which is similar to the research results of *Rubus* [16]. The phylogenetic reconstructions based on the representative proteins of chloroplast genomes have illustrated robust and consistent relationships with high support, providing a reference to develop tools to study *Rosa* species in more detail.

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* [52], *Pinus massoniana* [53], and *Dalbergia odorifera* [54]. This shows that there are some similarities in codon preference among different species.

The A–T bond is a less hydrogen bond than the G–C bond, and it is easier to break than the G–C bond. Therefore, the probability of the A–T bond in the chloroplast genome SSR is greater [10]. A–T SSR has the highest proportion in *R. lucieae* chloroplast. Moreover, it contains G–C, which is consistent with the research results of *Rubus* [16]. It has more contributions to the genetic variation than the longer SSRs. In this study, it was found that *R. multiflora* detected the largest number of SSRs and repeat sequences, and it had the longest sequence length (157,385 bp). It is speculated that the number of sequence repeats may affect the sequence length. The SSRs identified in this study are of great significance for understanding the genetic diversity of *Rosa*, constructing a DNA fingerprint database, generating a genetic map, and providing a reference for the identification of *Rosa*.

4.2. Sliding Window Analysis

In addition to random genetic variation events, some mutations constitute highly variable regions in the genome, namely, mutational hotspots [55]. 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 Jeon et al.'s [1] 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. (0.007 and 0.006) in terms of nucleotide variation. Compared with The *Dendrobium* (0.2) [56] and *Yulania* (0.02) [57], the nucleotide variation value of *Rosa* (0.007) is relatively low, which shows that the chloroplast genome sequence of *Rosa* is conservative and not easy to produce mutations. These highly variable loci can be used for phylogenetic studies of the *Rosa* DNA barcode and at the species level.

4.3. 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 [58,59]. 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 [60]. In the process of positive selection, favorable amino acid changes increase plant adaptation to ecological habitats [61]. Compared with other studies, positive selection of multiple loci was found in *Rosa*, and many genes were involved [62–65]. 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, accessed on 15 March 2022). Many positive selection genes found in this study were also found to have the positive selection in other plants and to be involved in the adaptive evolution of plants. These include *matK*, *atpF*, *psbA*, *ycf*, *ycf2*, and *rbcL* [66]. For example, several studies have found that the adaptive evolution of the *rbcL* gene is related to photosynthetic performance under changes in temperature, drought, and carbon dioxide concentrations [63,67,68]. 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 [69]. Kikuchi et al. [70] 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 [71]. 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.

4.4. Phylogenetic Analysis

According to the Flora of China (http://www.iplant.cn, accessed on 15 March 2022), 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 Thory, Sect. Braeteatae Thory, Sect. Microphyllae 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 *R. sterilis* and *R. chinensis* is close, but they belong to Sect. Chinenses 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. In addition, molecular phylogenetic tree reconstruction shows that *R. lucieae* has a nonmonophyletic branch (MG727864), which is consistent with the research results of sequence submitters Jeon and Kim [1]. It is speculated that *R. lucieae* in this study has hybridization or chloroplast capture events or incomplete lineage sorting, which suggests a need for more data analysis. 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 [72].

5. 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 repeat sequences, IR/SC boundaries, codons, and a sliding window of the chloroplast genomes of 13 species with close genetic relationships in *Rosa* were compared and analyzed. Among the 13 chloroplast genomes, the IR/SC boundary is relatively conservative; the difference regions of SSRs, repeat sequences, and highly variable regions can be used to develop genetic markers for further population genetics research. 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 more 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*.

Author Contributions: Sampling, Z.D. and W.S.; DNA sequencing, W.S., W.Z. and F.W.; data analysis, W.S., L.M. and W.L.; thesis writing, W.S.; project fund support provider, P.X. All authors have read and agreed to the published version of the manuscript.

Funding: 1. Science and Technology Innovation Fund Project of Southwest Forestry University (No.KY21027). 2. Yunnan Science and Technology Talents and Platform Program (No. 202205AF150022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: *R. lucieae* sequence data used in the paper have been uploaded to GenBank (Accession:OK938394). Other sequence data involved in the analysis have been downloaded to GenBank.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

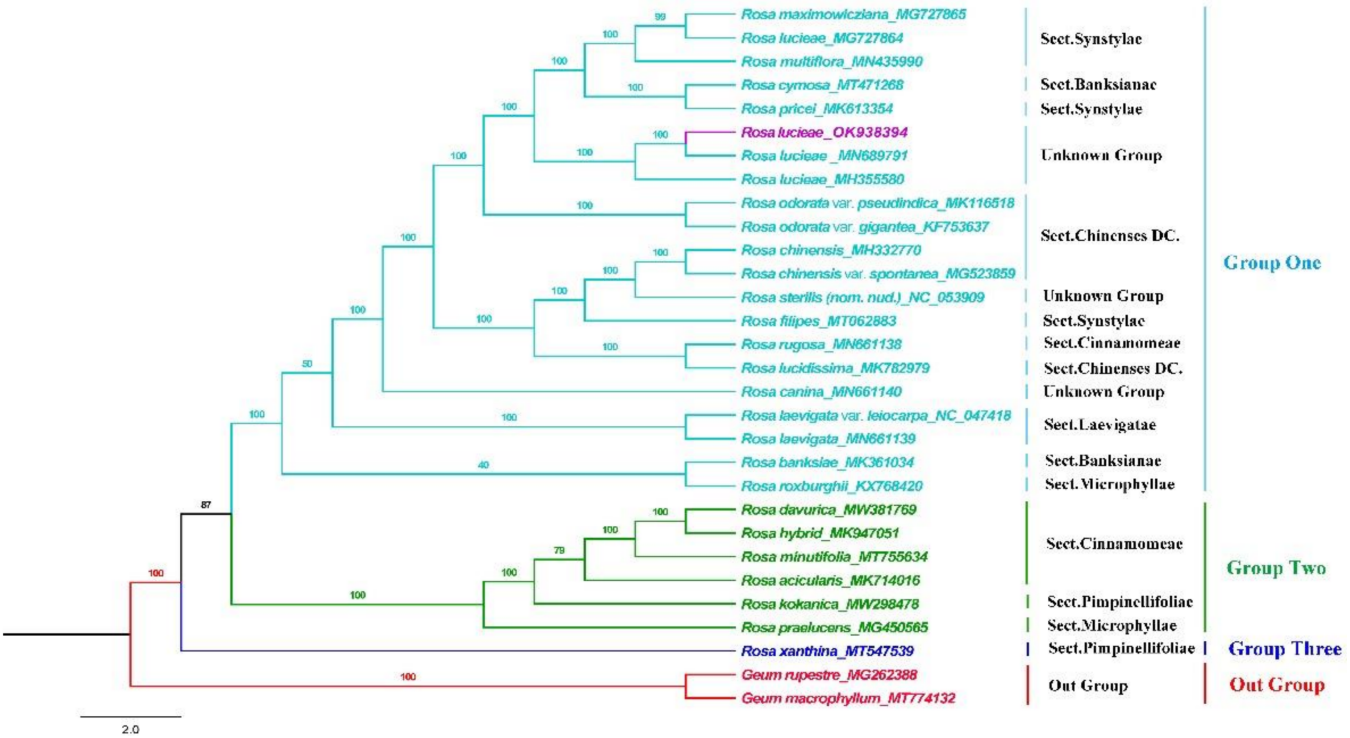

Figure A1. Molecular phylogenetic tree of *Rosa* based on 30 chloroplast genome sequences by maximum likelihood method.

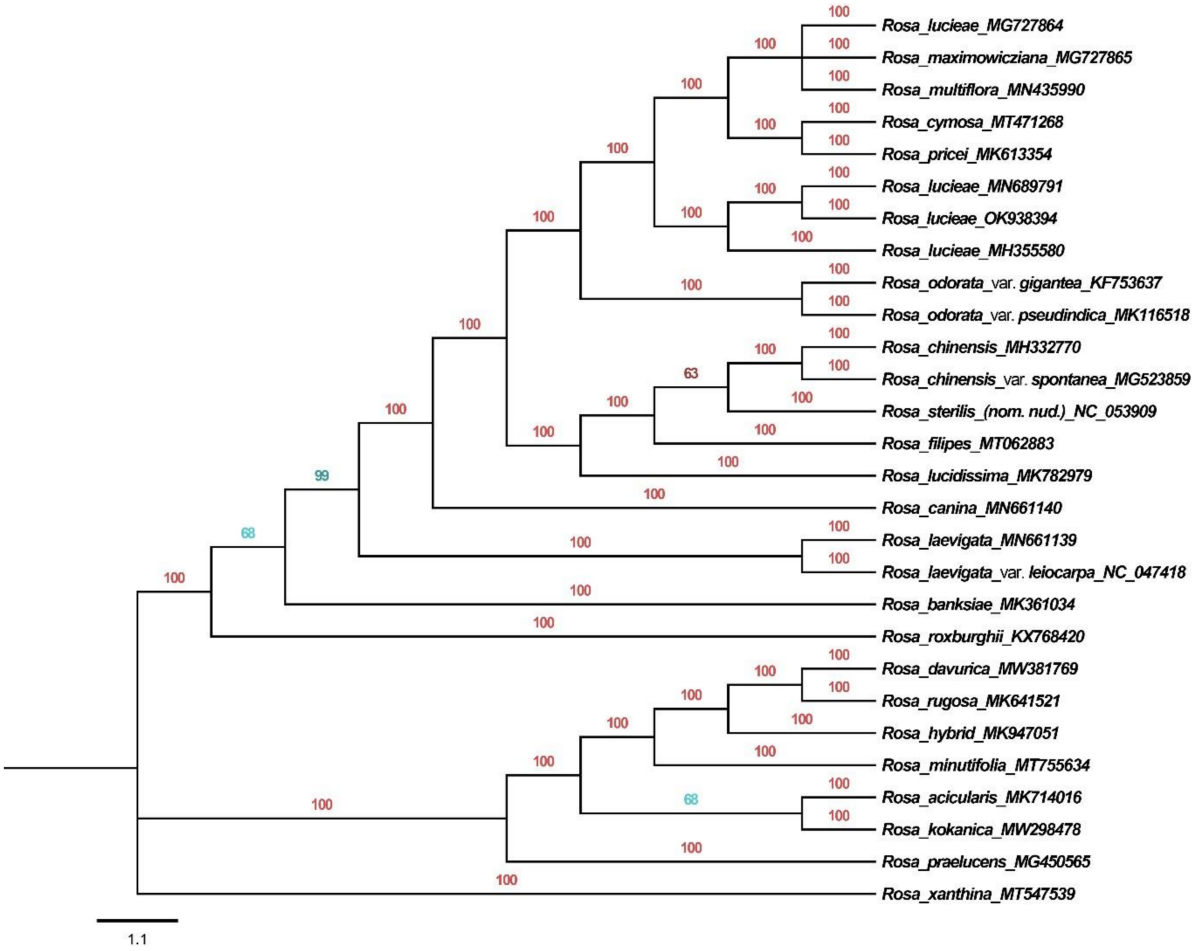

Figure A2. Molecular phylogenetic BI tree topology constructed by CDS with 28 sequences.

References

1. Jeon, J.H.; Kim, S.C. Comparative analysis of the complete chloroplast genome sequences of three closely related East-Asian wild roses (Rosa sect. Synstylae; Rosaceae). *Genes* **2019**, *10*, 23. [CrossRef] [PubMed]
2. Debener, T.; Linde, M. Exploring complex ornamental genomes: The Rose as a model plant. *CRC Crit. Rev. Plant Sci.* **2009**, *28*, 267–280. [CrossRef]
3. Lv, J.J. Establishment of plant regeneration system and preliminary studies on genetic transformation system of *Rosa wichuriana* 'Basye's thornless. Master's Thesis, Huazhong Agricultural University, Wuhan, China, 2013.
4. Jin, J.; Jin, P.; Wu, H.E.; Dong, W.P.; Yang, C.H.; Zhou, H.Y. Investigation and Application of *Rosa* plant Resources in Guizhou. *Seed* **2020**, *39*, 61–65,69. [CrossRef]
5. Wang, S.Q.; Zhu, Z.M. Relationship between species richness patterns of *Rosa* L. and environmental factors in China. *Acta Ecol. Sin.* **2022**, *42*, 209–219. [CrossRef]
6. Xing, S.C.; Clarke, J.L. Progress in Chloroplast Genome Analysis. *Prog. Biochem. Biophys.* **2008**, *4*, 21–28.
7. Raubeson, L.A.; Jansen, R.K. *Chloroplast Genomes of Plants, Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; CABI Publishing: Cambridge, UK, 2005. [CrossRef]
8. Jansen, R.K.; Ruhlman, T.A. Plastid Genomes of Seed Plants. In *Genomics of Chloroplasts and Mitochondria*; Springer: Dordrecht, The Netherlands, 2012. [CrossRef]
9. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef]
10. Liang, X.; Li, P.; Xu, D.M.; Jia, X.Y.; Wang, W.B. Characteristics Analysis of Whole Chloroplast Genome in *Perilla frustescens*. *J. Shanxi Agric. Sci.* **2021**, *49*, 265–272. [CrossRef]
11. Day, P.D.; Madeleine, B.; Laurence, H.; Fay, M.F.; Leitch, A.R.; Leitch, I.J.; Kelly, L.J. Evolutionary relationships in the medicinally important genus *Fritillaria* L. (Liliaceae). *Mol. Phylogenetics Evol.* **2014**, *80*, 11–19. [CrossRef]
12. Li, M.; Ye, Q.; Song, Y.F.; Wu, S.H.; Yi, X.G.; Wang, X.R. The Analysis of the Codon Usage Bias in the Chloroplast Genome of *Prunus sargentii*. *Mol. Plant Breed.* **2021**. Available online: https://kns.cnki.net/kcms/detail/46.1068.S.20210728.0854.002.html (accessed on 8 May 2022).

13. Dong, Z.H.; Qu, S.H.; Liu, C.; Ye, P.; Xin, P.Y. The complete chloroplast genome sequence of *Eriobotrya fragrans*. *Mitochondrial DNA Part B* **2019**, *4*, 3549–3550. [CrossRef] [PubMed]
14. Qu, Y.Y.; Xin, J.; Feng, F.Y.; Dong, Z.H.; Qu, S.H.; Wang, H.Y. Codon Usage Bais in Chloroplast Genome of *Eriobotrya fragrans* Champ. Ex Benth. *J. Northwest For. Univ.* **2021**, *36*, 138–144,158.
15. Su, Y.; Liu, J.J.; Wan, B.; Zhang, P.J.; Cheng, Z.G.; Suzuki, J.J.; Wang, C. Chloroplast Genome Structure Characteristic and Phylogenetic Analysis of *Mulgedium tataricum*. *J. Agric. Sci. Technol.* **2021**, *23*, 33–42. [CrossRef]
16. Wang, Q.R.; Huang, Z.R.; Gao, C.S.; Ge, Y.Q.; Cheng, R.B. The complete chloroplast genome sequence of *Rubus hirsutus* Thunb. and a comparative analysis within Rubus species. *Genetica* **2021**, *149*, 299–311. [CrossRef]
17. Yu, J.J.; Fu, J.; Fang, Y.P.; Xiang, J.; Dong, H.J. Complete chloroplast genomes of *Rubus* species (Rosaceae) and comparative analysis within the genus. *BMC Genom.* **2022**, *23*, 32. [CrossRef]
18. Li, Q.Q.; Wen, J. The complete chloroplast genome of *Geum macrophyllum* (Rosaceae: Colurieae). *Mitochondrial DNA Part B* **2021**, *6*, 297–298. [CrossRef] [PubMed]
19. Zhang, P.P.; Wang, L.; Lu, X. Complete chloroplast genome of *Geum aleppicum* (Rosaceae). *Mitochondrial DNA B* **2022**, *7*, 234–235. [CrossRef] [PubMed]
20. Xin, Y.X.; Li, R.Z.; Li, X.; Chen, L.Q.; Tang, J.R.; Qu, Y.Y.; Xin, P.; Li, Y. Analysis on codon usage bias of chloroplast genome in *Mangifera indica*. *J. Cent. South Univ. For. Technol.* **2021**, *41*, 148–156+165. [CrossRef]
21. Moore, M.J.; Dhingra, A.; Soltis, P.S.; Shaw, R.; Farmerie, W.G.; Folta, K.M.; Soltis, D.E. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol.* **2006**, *6*, 17. [CrossRef]
22. Bayly, M.J.; Rigault, P.; Spokevicius, A.; Ladiges, P.Y.; Ades, P.K.; Anderson, C.; Bossinger, G.; Merchant, A.; Udovicic, F.; Woodrow, I.E.; et al. Chloroplast genome analysis of Australian eucalypts—*Eucalyptus*, *Corymbia*, *Angophora*, *Allosyncarpia* and *Stockwellia* (Myrtaceae). *Mol. Phylogenetics Evol.* **2013**, *69*, 704–716. [CrossRef]
23. 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. *Sci. Hortic.* **1998**, *77*, 73–82. [CrossRef]
24. Cui, W.H.; Zhong, M.C.; Du, X.Y.; Qu, X.J.; Jiang, X.D.; Sun, Y.B.; Wang, D.; Chen, S.Y.; Hu, J.Y. The complete chloroplast genome sequence of a rambler rose, *Rosa wichuraiana* (Rosaceae). *Mitochondrial DNA Part B* **2020**, *5*, 252–253. [CrossRef] [PubMed]
25. Gao, C.W.; Wu, C.H.; Zhang, Q.; Wu, M.X.; Chen, R.R.; Zhao, Y.L.; Guo, A.; Li, Z. Sequence and phylogenetic analysis of the chloroplast genome for *Rosa xanthina*. *Mitochondrial DNA Part B* **2020**, *5*, 2940–2941. [CrossRef]
26. Zhao, X.; Gao, C.W. The complete chloroplast genome sequence of *Rosa minutifolia*. *Mitochondrial DNA Part B* **2020**, *5*, 3338–3339. [CrossRef] [PubMed]
27. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15. [CrossRef]
28. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef] [PubMed]
29. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [CrossRef]
30. Kearse, M.; Moir, R.D.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]
31. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, 59–64. [CrossRef]
32. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef]
33. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef]
34. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]
35. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef] [PubMed]
36. Katoh, K.; Standley, D. MAFFT multiple sequence alignment software version improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
37. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]
38. Zhang, D.; Gao, F.L.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2019**, *20*, 348–355. [CrossRef]
39. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8. [CrossRef]
40. Gao, F.; Chen, C.; Arab, D.A.; Du, Z.; He, Y.; Ho, S.Y.W. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evol.* **2019**, *9*, 3891–3898. [CrossRef]

41. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98. [CrossRef]
42. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
43. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [CrossRef]
44. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef]
45. Wang, S.Q.; Chen, J.F.; Zhu, Z.M. The complete chloroplast genome sequence of *Rosa filipes* (Rosaceae). *Mitochondrial DNA Part B* **2020**, *5*, 1376–1377. [CrossRef]
46. Ding, M.Y.; Liao, M.; Liu, P.Q.; Tan, G.Q.; Chen, Y.Q.; Shi, S. The complete chloroplast genome of *Rosa cymosa* (Rosaceae), a traditional medicinal plant in South China. *Mitochondrial DNA B* **2020**, *5*, 2571–2572. [CrossRef] [PubMed]
47. Lu, R.S.; Li, P.; Qiu, Y.X. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: Comparative genomic and phylogenetic analyses. *Front. Plant Sci.* **2017**, *10*, 2054. [CrossRef] [PubMed]
48. Raubeson, L.A.; Peery, R.; Chumley, T.W.; Dziubek, C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. *BMC Genom.* **2007**, *8*, 174. [CrossRef]
49. Li, R.; Ma, P.F.; Wen, J.; Yi, T.S. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. *PLoS ONE* **2013**, *8*, e78568. [CrossRef]
50. Song, Y.; Dong, W.; Liu, B.; Xu, C.; Yao, X.; Gao, J.; Corlett, R.T. 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. [CrossRef]
51. Kim, K.J.; Lee, H.L. Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* **2004**, *11*, 247–261. [CrossRef]
52. Yang, G.F.; Su, K.L.; Zhao, Y.R.; Song, Z.B.; Sun, J. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*. *Acta Prataculturae Sin.* **2015**, *24*, 171–178. [CrossRef]
53. Ye, Y.J.; Ni, Z.X.; Bai, T.D.; Xu, L. The analysis of chloroplast genome codon usage bais in *Pinus massoniana*. *Genom. Appl. Biol.* **2018**, *37*, 4464–4471. [CrossRef]
54. Yuan, X.L.; Li, Y.Q.; Zhang, J.F.; Wang, Y. Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera*. *Guihaia* **2021**, *41*, 622–630. [CrossRef]
55. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. 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*, 275–288. [CrossRef] [PubMed]
56. Du, Z.H.; Yang, L.; Zhang, C.J.; Xu, H.J.; Chen, Z.L. Characteristics of the complete chloroplast genome of *Dendrobium ochreatum* and its comparative analysis. *Chin. J. Trop. Crops* **2021**, *42*, 3111–3119. [CrossRef]
57. Sun, L.Y.; Jiang, Z.; Wan, X.X.; Zou, X.; Yao, X.Y.; Wang, Y.L.; Yin, Z.F. The complete chloroplast genome of *Magnolia polytepala*: Comparative analyses offer implication for genetics and phylogeny of *Yulania*. *Gene* **2020**, *736*, 100010. [CrossRef] [PubMed]
58. 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. [CrossRef]
59. Yang, Z.; Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [CrossRef]
60. Nawae, W.; Yundaeng, C.; Naktang, C.; Kongkachana, W.; Yoocha, T.; Sonthirod, C.; Narong, N.; Somta, P.; Laosatit, K.; Tangphatsornruang, S.; et al. The Genome and Transcriptome Analysis of the *Vigna mungo* Chloroplast. *Plants* **2020**, *9*, 1247. [CrossRef]
61. Sen, L.; Fares, M.A.; Liang, B.; Gao, L.; Wang, B.; Wang, T.; Su, Y.J. Molecular evolution of *rbcL* in three gymnosperm families: Identifying adaptive and coevolutionary patterns. *Biol. Direct* **2011**, *6*, 29. [CrossRef]
62. Rono, P.C.; Dong, X.; Yang, J.X.; Mutie, F.M.; Oulo, M.A.; Malombe, I. Initial complete chloroplast genomes of *Alchemilla* (Rosaceae): Comparative analysis and phylogenetic relationships. *Front. Genet.* **2020**, 11. [CrossRef]
63. Sheng, J.; Yan, M.; Wang, J.; Zhao, L.; Zhou, F.; Hu, Z.; Jin, S.; Diao, Y. The complete chloroplast genome sequences of five *Miscanthus* species, and comparative analyses with other grass plastomes. *Ind. Crops Prod.* **2021**, *162*, 113248. [CrossRef]
64. Huang, S.N.; Ge, X.J.; Cano, A.; Salazar, B.G.M.; Deng, Y.F. Comparative analysis of chloroplast genomes for five *Dicliptera* species (Acanthaceae): Molecular structure, phylogenetic relationships, and adaptive evolution. *PeerJ* **2020**, *8*, e8450. [CrossRef]
65. Xie, D.F.; Yu, Y.; Deng, Y.Q.; Li, J.; Liu, H.Y.; Zhou, S.D.; He, X.J. Comparative analysis of the chloroplast genomes of the chinese endemic genus *Urophysa* and their contribution to chloroplast phylogeny and adaptive evolution. *Int. J. Mol. Sci.* **2018**, *19*, 1847. [CrossRef] [PubMed]
66. Bock, D.G.; Andrew, R.L.; Rieseberg, L.H. On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* **2014**, *23*, 4899–4911. [CrossRef] [PubMed]
67. Galmes, J.; Andralojc, P.J.; Kapralov, M.V.; Flexas, J.; Keys, A.J.; Molins, A.; Conesa, M.À. Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus *Limonium* (Plumbaginaceae). *New Phytol.* **2014**, *203*, 989–999. [CrossRef] [PubMed]

68. Kapralov, M.V.; Smith, J.A.C.; Filatov, D.A. Rubisco evolution in C4 eudicots: An analysis of Amaranthaceae sensu lato. *PLoS ONE* **2012**, *7*, e52974. [CrossRef]
69. Drescher, A.; Ruf, S.; Calsa, T.; Carrer, H., Jr.; Bock, R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* **2000**, *22*, 97–104. [CrossRef]
70. Kikuchi, S.; Bedard, J.; Hirano, M.; Hirabayashi, Y.; Oishi, M.; Imai, M.; Takasetoru, M.; Ide, T.; Nakai, M. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **2013**, *339*, 571–574. [CrossRef]
71. Gao, L.Z.; Liu, Y.L.; Zhang, D.; Li, W.; Gao, J.; Liu, Y.; Li, K.; Shi, C.; Zhao, Y.; Zhao, Y.-J.; et al. Evolution of Oryza chloroplast genomes promotedad aptation to diverse ecological habitats. *Commun. Biol.* **2019**, *2*, 278. [CrossRef]
72. Zhu, S.X. Systematics of Chaetoseris and Stenoseris (Compositae-Lactuceae). Master's Thesis, Graduate School of Chinese Academy of Sciences (Institute of Botany), Beijing, China, 2014.