Chloroplast Genome Characteristics and Intraspecific Variation in Rosa zhongdianensis (Postprint)

Authors: Wang Qigang, Cao Shirui, Wang Huichun, Ma Changle, Yan Huijun, Xianqin Qiu, Jing Weikun, Jian Hongying

Date: 2024-03-26T00:00:00+00:00

Abstract

Rosa praelucens Byhouwer is an endemic “critically endangered”plant in Shangri-La County, Yunnan Province, a nationally second-class key protected plant, a renowned alpine ornamental flower, and an important decaploid rose germplasm resource that exhibits rich phenotypic diversity within the species. To elucidate the genetic background of intraspecific phenotypic variation in Rosa praelucens, this study employed next-generation sequencing technology to sequence, assemble, and comparatively analyze the chloroplast genomes of 40 representative individuals with distinct phenotypes. The results demonstrated that: (1) The chloroplast genome sequence of Rosa praelucens ranged from 157,173 to 157,261 bp in length, with only an 88-bp difference among individuals, encoding a total of 132 functional genes primarily associated with photosynthesis and self-replication. All genes were encoded by 27,155 codons, with codons terminating in A and U being relatively prevalent. (2) The chloroplast genome of Rosa praelucens contained 36 repeat sequences and 73 simple sequence repeats (SSRs), the majority of which were mononucleotide SSRs predominantly located in the intergenic spacers of the Large Single-Copy (LSC) region. (3) The haplotype diversity of the complete chloroplast genome within Rosa praelucens was 0.928$±$0.027, with a nucleotide polymorphism of 0.00012. Intergenic spacers such as petN-trnD and psaA-ycf3 in the LSC region, as well as genes including rps16 and ycf1, exhibited relatively high nucleotide polymorphism. No large-scale fragment or gene inversion/loss was detected in structural comparisons among chloroplast genomes of representative individuals with different phenotypes. These findings indicate that Rosa praelucens is highly conserved in chloroplast genome size, sequence, and gene structure, and that the rich intraspecific phenotypic diversity is not attributable to variation in the chloroplast genome.

Full Text

Preamble

ChinaXiv DOI: 10.11931/guihaia.gxzw202312023

Chloroplast Genome Features and Intraspecific Variation of *Rosa praelucens*

WANG Qigang[1], CAO Shirui[1,2], WANG Huichun[1], MA Changle[2], YAN Huijun[1], QIU Xianqin[1], JING Weikun[1], JIAN Hongying[1]*

[1] Flower Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
[2] College of Landscape and Horticulture, Southwest Forestry University, Kunming 650224, China

Abstract

Rosa praelucens Byhouwer is a critically endangered alpine wildflower endemic to Shangri-La County, Yunnan Province, and represents an important decaploid rose germplasm resource with rich phenotypic diversity. To clarify the genetic basis of intraspecific phenotypic variation, we sequenced, assembled, and compared the chloroplast genomes of 40 representative individuals exhibiting different phenotypes. The results revealed: (1) Chloroplast genomes of *R. praelucens* ranged from 157,173 to 157,261 bp, with only 88 bp variation among individuals. These genomes encoded 132 functional genes, primarily related to photosynthesis and self-replication, comprising 27,155 codons that showed preference for A- or U-ending codons. (2) Thirty-six repeat sequences and 73 simple sequence repeats (SSRs) were identified, with most cpSSRs being mononucleotide repeats located in the intergenic regions of the large single-copy (LSC) region. (3) Intraspecific haplotype diversity (Hd) was $0.928\pm0.027, with nucleotide diversity($) of 0.00012. Relatively higher nucleotide polymorphism was detected in the intergenic regions *petN-trnD* and *psaA-ycf3* in the LSC region, as well as in the *rps16* and *ycf1* genes. No large-scale inversions or gene losses were observed among individuals. These findings indicate that *R. praelucens* chloroplast genomes are highly conserved in size, sequence, and structure, suggesting that the rich intraspecific phenotypic diversity is not directly caused by chloroplast genome variation.

Keywords: *Rosa praelucens* Byhouwer, chloroplast genome, comparative genomics, simple sequence repeats (SSRs), nucleotide diversity (π), codon preference
CLC number: Q943
Document code: A

Introduction

Chloroplasts play vital roles throughout the plant life cycle (Wicke et al., 2011). Most vascular plant chloroplast genomes are approximately 150 kb in length and maintain a conserved quadripartite structure comprising a large single-copy region (LSC), a small single-copy region (SSC), and two inverted repeat regions (IRs) (Wicke et al., 2011; Shetty et al., 2016; Zhu et al., 2016). Although chloroplast genomes of higher plants are generally highly conserved, certain lineages exhibit large-scale inversions (Sun et al., 2017), extensive repeat sequences (Guisinger et al., 2011), gene loss or pseudogenization (Ye et al., 2018), and IR region expansion or contraction (Li et al., 2017; Liu et al., 2017). In most angiosperms, chloroplast genomes are maternally inherited (Neale & Sederoff, 1989; Daniell et al., 2016). Compared with nuclear genomes, chloroplast genomes offer advantages including lower molecular weight, simpler structure, and higher conservation, while containing abundant repeat sequences including simple sequence repeats (SSRs) (Cavalier, 2002). Consequently, they are widely applied in phylogenetic studies, DNA barcoding, genetic engineering, and kinship analysis (Dong et al., 2018). With the development of next-generation sequencing (NGS) technology, an increasing number of complete chloroplast genome sequences have been reported, with over 8,500 plant chloroplast genomes currently deposited in the NCBI database.

Rosa praelucens is a critically endangered plant species endemic to Shangri-La County, Yunnan (Qin et al., 2017; Ku & Robertson, 2003) and is listed as a National Second-Class Protected Plant. As a renowned alpine ornamental flower and important rose germplasm resource, it exhibits rich phenotypic diversity (Li & Zhou, 2005). Since Jian et al. (2010) discovered that it is the only decaploid wild rose species (2n=10x=70), numerous studies have investigated its habitat and community characteristics (Guan et al., 2012), breeding system (Wu et al., 2014), population status (Zhou et al., 2016), phylogenetic position (Wang et al., 2018), karyotypic features based on fluorescence in situ hybridization (Fang et al., 2020), and genetic diversity and structure (Jian et al., 2018a). These studies have revealed substantial intraspecific phenotypic diversity, particularly in flower color and form (Li et al., 2013; Jian et al., 2018a).

Understanding how polyploidy modifies phenotypic traits represents a major focus of evolutionary biology (Balao et al., 2011). Numerous studies have demonstrated that naturally occurring or artificially induced polyploid plants undergo genetic and epigenetic changes that alter gene expression, leading to differentiation in genetics, physiology, and morphology, ultimately generating novel phenotypes (Ramsey & Schemske, 2002). However, the mechanisms underlying the rich phenotypic variation in decaploid *R. praelucens* remain unclear. The high ploidy level has limited the application of many molecular techniques for studying its genetic background. Jian et al. (2017) reported basic information on the *R. praelucens* chloroplast genome, including its size, partition lengths, GC content, and gene number, revealing a genome length of 157,186 bp—larger than related species such as *Rosa chinensis* var. *spontanea*—with a 505 bp insertion

between *psbM* and *trnD* in the LSC region. Building upon this foundation, the present study sequenced, assembled, and compared chloroplast genomes from 40 representative individuals with different phenotypes to address two key questions: (1) What are the sequence characteristics and codon usage preferences of the *R. praelucens* chloroplast genome? (2) Do individuals with different phenotypes exhibit substantial chloroplast genome variation? These findings will provide essential genetic information for understanding speciation and conservation in *R. praelucens* and establish a chloroplast genomic foundation for investigating the molecular mechanisms of intraspecific phenotypic variation.

Materials and Methods

Plant Materials

The *R. praelucens* individual used for chloroplast genome characterization was collected from Tang' anpei, Xiaozhongdian Town, Shangri-La County (99°49 38.1 E, 27°32 16.68 N, 3,248 m). Its sequence has been deposited in NCBI under accession number MG450565.1. Information for the remaining 40 representative individuals with different phenotypes is provided in Table 1 . Fresh, healthy leaves were collected in late June 2021, immediately dried with silica gel, and stored at -4°C for subsequent experiments.

DNA Extraction, Sequencing, Assembly, and Annotation

Total DNA was extracted from leaf tissue using a modified CTAB method. DNA meeting library construction requirements was sent to Beijing Novogene Bioinformatics Technology Co., Ltd. for Illumina HiSeq 2000 sequencing. Approximately 3.5 Gb of 150 bp paired-end raw reads were generated per sample. Raw data were filtered using NGS QC Toolkit v2.3.3 (Patel & Jain, 2012) with default parameters to obtain high-quality clean reads. Chloroplast genomes were *de novo* assembled using GetOrganelle (https://github.com/Kinggerm/GetOrganelle). Assembled genomes were automatically annotated with CpGAVAS (Liu et al., 2012), followed by manual correction of gene boundaries using Geneious 9.1 (Kearse et al., 2012). Physical maps were generated using OGDRAW (Lohse et al., 2013).

Chloroplast Genome Structure Analysis

Gene composition was analyzed for the *R. praelucens* chloroplast genome (NCBI accession MG450565.1) using Geneious software. Codon usage bias was analyzed using MEGA 6.06 (Tamura et al., 2013) to calculate relative synonymous codon usage (RSCU) values and AT content. REPuter software (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (Kurtz et al., 2001) was used to identify forward and reverse repeats with parameters set to minimum length $ $20 bp and sequence identity >85%. Simple sequence repeats were identified using MISA (Beier et al., 2017) with search thresholds of $ $10, $ $5, $ $4, $ $3,

$ $3, and $ $3 repeats for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs, respectively.

Comparative Analysis of Intraspecific Chloroplast Genomes

Mauve (Darling et al., 2004) was implemented in Geneious to align chloroplast genomes from 40 representative individuals and detect large-scale inversions or losses. DnaSP v5.10 (Librado et al., 2009) was used to calculate haplotype diversity (Hd) and nucleotide polymorphism (π) across intraspecific chloroplast genomes and to identify highly variable regions.

Results

Chloroplast Genome Structure of *Rosa praelucens*

Basic characteristics of chloroplast genomes from 40 representative individuals are summarized in Figure 1 [Figure 1: see original paper] and Table 2 . Genome length ranged from 157,173 to 157,261 bp, with only 88 bp variation among individuals. The largest genome (157,261 bp) belonged to individual 7-1, while the smallest (157,173 bp) was from individual 2-5. The LSC region varied from 86,300 to 86,353 bp (53 bp difference), with individual 7-1 having the longest and 2-5 the shortest. The SSC region ranged from 18,765 to 18,803 bp (38 bp difference). IR region length was consistently 26,054 bp across all individuals, indicating that genome size variation primarily originated from the LSC and SSC regions. GC content showed no significant differences among individuals, with whole-genome GC content at 37.2%, IR regions at 42.7%, LSC regions at 35.2%, and SSC regions at 31.2%.

Gene Composition of the *Rosa praelucens* Chloroplast Genome

The chloroplast genome encoded 132 functional genes, including 87 protein-coding genes, 37 tRNA genes, and 8 rRNA genes (Table 3). Among these, 45 genes were associated with photosynthesis, 76 with self-replication, and 11 had unknown functions. Six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *ycf2*), seven tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), and four rRNAs (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*) were completely duplicated in the IR regions. Thirteen genes contained one intron (*petB*, *petD*, *ndhA*, *ndhB*, *rps16*, *rpl2*, *rpl16*, *rpoC1*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*), while *ycf3* and *clpP* each contained two introns. The *rps12* gene was trans-spliced, with its 5 end in the LSC region and 3 end duplicated in the IR regions.

Codon Usage Bias in the *Rosa praelucens* Chloroplast Genome

Codon usage analysis revealed 27,155 codons encoding all genes, with leucine being the most frequent amino acid (2,765 codons, 10.85%) and histidine the

least frequent (530 codons, 1.95%). A- and U-ending codons were predominant. Except for *trnL-CAA*, *trnS-GGA*, Arg-AGG, and Gly-GGG, all preferred synonymous codons (RSCU > 1) ended with A or U (Table 4).

Repeat Sequences in the *Rosa praelucens* Chloroplast Genome

Thirty-three forward repeats and three reverse repeats were identified, mostly 20–30 bp in length (Table 5). The longest repeats were located in the *rps12-trnV*(GAC) intergenic region and the *ndhA* intron. Most repeats were distributed in the LSC and IR regions, with nine spanning different regions (e.g., repeat 1 started in both the IRB and SSC regions).

Simple Sequence Repeats in the *Rosa praelucens* Chloroplast Genome

MISA analysis identified 73 SSRs, including 50 mononucleotide repeats, 13 dinucleotide (AG/AT/TA/TC), 5 trinucleotide, 11 tetranucleotide, and 2 hexanucleotide repeats; no pentanucleotide SSRs were found. Most SSRs (65) were simple types, with only eight compound types and no interrupted types (Table 6). Fifty-eight SSRs (79.5%) were located in the LSC region, five in the SSC region, and five each in the IRA and IRB regions. Only 23 SSRs resided within genes, with the remainder in intergenic regions. Seventy-four percent of mononucleotide SSRs consisted of A/T repeats, consistent with the hypothesis that cpSSRs primarily comprise short adenine or thymine repeats rather than guanine or cytosine repeats.

Intraspecific Chloroplast Genome Sequence Variation

Intraspecific chloroplast genome variation was minimal across the 40 individuals. Whole-genome alignment detected 58 variable sites defining 22 haplotypes, with haplotype diversity of 0.928$±$0.027 and nucleotide diversity of 0.00012. Both genic and intergenic regions showed low polymorphism, with relatively higher diversity in the LSC intergenic regions *psbI-trnS*(GCU), *trnS*(GCU)-*trnG*(UCC), *trnG*(UCC)-*trnfM*(CAU), *petN-trnD*(GUC), *petA-psbJ*, and *psaA-ycf3*, as well as in the *rps16* and *ycf1* genes (Figure 2 [Figure 2: see original paper]). Mauve alignment revealed no large-scale rearrangements, inversions, or gene losses among individuals (Figure 3 [Figure 3: see original paper]).

Discussion and Conclusion

Although chloroplast genomes are generally conserved in gene content and order, long-term adaptation to different environments can cause size variation, structural rearrangements, and IR region expansion or contraction among congeneric species (Daniell et al., 2016). The 40 representative individuals of *R. praelucens* exhibited chloroplast genome sizes of 157,173–157,261 bp encoding 132 genes primarily involved in photosynthesis and self-replication. Compared with related species such as *Rosa chinensis* var. *spontanea* (Jian et al., 2018b),

R. lucidissima (Zhao et al., 2019), *R. banksiae* (Yang, 2019), *R. odorata* var. *gigantea* (Yang et al., 2014), *R. laevigata* (Yin et al., 2020), and other *Rosa* species (Chen et al., 2019; Cui et al., 2022), the *R. praelucens* chloroplast genome is approximately 500 bp larger but similar in GC content, gene composition, and arrangement, indicating that *Rosa* chloroplast genomes are relatively conserved with minor interspecific differences primarily in non-coding region length.

Codons serve as the link between nucleic acids and proteins. Analyzing codon usage bias and identifying optimal codons can inform the design of expression vectors to improve target gene expression, with important applications in crop breeding and improvement (Qi et al., 2015). The *R. praelucens* chloroplast genome showed preference for A- and U-ending codons (RSCU > 1), with leucine as the most frequently used amino acid and histidine the least. This pattern is consistent with codon usage in *Rosa chinensis* var. *spontanea* (Jian et al., 2018b), providing a foundation for studying molecular evolution and heterologous expression of relevant genes.

Chloroplast SSRs (cp-SSRs) are uniparentally inherited and exhibit high intraspecific polymorphism, serving as important genetic markers for evolutionary and population genetic studies (Cavalier, 2002; Provan, 2000; Flannery et al., 2006) and for constructing genetic maps in crops (Powell et al., 1995; Xue et al., 2012). Poly-A and poly-T repeats likely possess greater structural stability than poly-C and poly-G repeats (Gragg et al., 2002), which explains why most plant chloroplast SSRs consist of adenine or thymine repeats. The 73 cpSSRs identified in *R. praelucens* were predominantly mononucleotide repeats composed of short A/T repeats located mainly in LSC intergenic regions, consistent with patterns in *Rosa chinensis* var. *spontanea* (Jian et al., 2018b) and other plants such as *Camellia* (Ding et al., 2022; Deng et al., 2024) and *Seriphidium* (Jin et al., 2023).

Previous studies have identified highly variable regions among *Rosa* species primarily in LSC intergenic regions (*trnK-rps16*, *ps16-trnQ*, *trnS-trnG*, *atpF-atpH*, *rps2-rpoC2*), coding regions of *rps19* and *ycf1*, and intron regions of *rpl2*, *rps16*, and *ndhA* (Jian et al., 2018b). In contrast, intraspecific nucleotide polymorphism in *R. praelucens* was low, with relatively higher diversity limited to a few LSC intergenic regions (*petN-trnD*(GUC), *petA-psbJ*, *psaA-ycf3*) and genes (*rps16*, *ycf1*). Combined with Mauve alignment results, these findings demonstrate that *R. praelucens* chloroplast genomes are highly conserved in sequence and structure, with no large-scale inversions or gene losses, indicating that intraspecific phenotypic variation is not attributable to chloroplast genome variation.

In summary, this study provides a detailed analysis of *R. praelucens* chloroplast genome characteristics, including gene composition, codon usage, and SSR features, along with comparative genomic analysis of individuals with different phenotypes. The results demonstrate high conservation of chloroplast genome size, sequence, and structure within the species, with no evidence of large-scale rearrangements or gene losses. These findings provide fundamental chloroplast

genomic data for the conservation and utilization of *R. praelucens* and indicate that the rich intraspecific phenotypic variation is not caused by chloroplast genome variation. Given the decaploid nature of *R. praelucens*, further systematic investigation should focus on chromosome number and structure, gene expression, and epigenetic modifications.

References

BALAO F, HERRERA J, TALAVERA S, 2011. Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: a multivariate morphological approach [J]. *New Phytol*, 192(1): 256-265.

BEIER S, THIEL T, MÜNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 33(16): 2583-2585.

CAVALIER ST, 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses [J]. *Curr Biol*, 12(2): 62-64.

CHAO YE, CHANG Y, WANG MF, et al., 2012. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species [J]. *Acta Agric Boreal-Sin*, 27(4): 60-64.

CHEN MR, ZHANG C, GAO XF. 2019. The complete chloroplast genome sequence of *Rosa pricei* (Rosaceae) [J]. *Mitochondrial DNA B*, 4(1): 1918–1919.

CUI WH, DU XY, ZHONG MC, et al., 2022. Complex and reticulate origin of edible roses (*Rosa*, Rosaceae) in China [J]. *Hortic Res*, 9: 51.

DANIELL H, LIN C S, YU M, ET AL., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J]. *Genome Biol*, 17(1): 1-29.

DARLING ACE, MAU B, BLATTNER FR, et al., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements [J]. *Genome Res*, 14(7): 1394-1403.

DENG JQ, JIAN HY, LI SB, et al., 2013. Cold tolerance of several wild *Rosa* resources endemic of Yunnan [J]. *SW Chin J Agric Sci*, 26(2): 273-277.

DENG YB, ZHANG J, LAN LL, et al., 2024. Analysis of chloroplast genome features of endangered and rare plant *Camellia minima* [J]. *Guihaia*, 44(1):30-42.

DING XQ, LI WF, WU JL, et al., 2022. Chloroplast genome characteristics and genetic relationship of yellow *Camellia* [J]. *J Fujian Agric For Univ (Nat Sci Ed)*, 52(3): 1-11.

DONG W, XU C, WU P, et al., 2018. Resolving the systematic positions of enigmatic taxa: Manipulating the chloroplast genome data of Saxifragales [J]. *Mol Phylogenet Evol*, 126(4): 321-329.

FAN YL, CHEN YC, JIAN HY, et al., 2021. Screening of *Rosa* germplasm resources with resistance to aphids [J]. *J Yunnan Univ (Nat Sci Ed)*, 43(3):

619-628.

FANG Q, TIAN M, ZHANG T, et al., 2020. Karyotype analysis of *Rosa praelucens* and its closely related congeneric species based on FISH [J]. *Acta Hortic Sin*, 47(3): 503-516.

FLANNERY ML, MITCHELL FJ, COYNE S, et al., 2006. Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs [J]. *Theor Appl Genet*, 113(7): 1023-1031.

GRAGG H, HARFE BD, JINKS-ROBERTSON S, 2002. Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frame shift intermediates by mismatch repair in *Saccharomyces cerevisiae* [J]. *Mol Cell Biol*, 22(24):8756–8762.

GUAN WL, LI SF, SONG J, et al., 2012. Study on geographic distribution of *Rosa praelucens* endemic to Yunnan [J]. *J W Chin For Sci*, 41(1): 88-93.

GUISINGER MM, KUEHL JV, BOORE JL, et al, 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage [J]. *Mol Biol Evol*, 28(1): 1543.

JIAN HY, LI SF, GUO JL, et al., 2018a. High genetic diversity and differentiation of an extremely narrowly distributed and critically endangered decaploid rose (*Rosa praelucens*): implications for its conservation [J]. *Conserv Genet*, 19(4):761-776.

JIAN HY, ZHANG H, TANG KX, et al., 2010. Decaploidy in *Rosa praelucens* Byhouwer (Rosaceae) endemic to zhongdian plateau, Yunnan, China [J]. *Caryologia*, 63(2): 162-167.

JIAN HY, ZHANG SD, ZHANG T, et al., 2017. Characterization of the complete chloroplast genome of a critically endangered decaploid rose species, *Rosa praelucens* (Rosaceae) [J]. *Conserv Genet Resour*, 10: 851–854.

JIAN HY, ZHANG YH, YAN HJ, et al., 2018b. The complete chloroplast genome of a key ancestor of modern roses, *Rosa chinensis* var. *spontanea*, and a comparison with congeneric species [J]. *Molecules*, 23: 389.

JIN GZ, LI WJ, SONG F, et al., 2023. Comparative analysis of complete *Artemisia* subgenus *Seriphidium* (Asteraceae: Anthemideae) chloroplast genomes: insights into structural divergence and phylogenetic relationships [J]. *BMC Plant Biol*, 136: 1-23.

KEARSE M, MOIR R; WILSON A, et al., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data [J]. *Bioinformatics*, 28(12): 1647-1649.

KU TC, ROBERTSON KR, 2003. *Rosa* (Rosaceae) [M] // WU ZY, RAVEN PH. *Flora of China*, Vol. 9. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press: 339-381.

KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: The manifold applications of repeat analysis on a genomic scale [J]. *Nucleic Acids Res*, 29(22): 4633-4642.

LI P, LU RS, XU WQ, et al., 2017. Comparative genomics and phylogenomics of East Asian tulips (*Amana*, Liliaceae) [J]. *Front Plant Sci*, 8: 451.

LI SF, LI CJ, JIAN HY, et al., 2013. Studies on phenotypic diversity of vulnerable *Rosa praelucens* endemic to Shangrila, Yunnan [J]. *Acta Hortic Sin*, 40(5): 924-932.

LI XX, ZHOU ZK, 2005. Endemic wild ornamental plants from North Western Yunnan [J]. *HortScience*, 40(6):1612-1619.

LIBRADO P, ROZAS J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data [J]. *Bioinformatics*, 25(11):1451-1452.

LIU C, SHI L, ZHU Y, et al, 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences [J]. *BMC Genom*, 13: 715.

LIU LX, LI R, WORTH JRP, et al., 2017. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): implications for understanding the evolution of Fagales [J]. *Front Plant Sci*, 8: 968.

LOHSE M, DRECHSEL O, KAHLAU S, et al, 2013. Organellar GenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets [J]. *Nucleic Acids Res*, 41(W1): 575-581.

NEALE DB, SEDEROFF RR, 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine [J]. *Theor Appl Genet*, 77(2): 212-216.

PATEL RK, JAIN M, 2017. NGS QC toolkit: a toolkit for quality control of next generation sequencing data [J]. *PloS ONE*, 7: e30619.

POWELL W, MORGANTE M, MCDEVITT R, et al, 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines [J]. *Proc Natl Acad Sci USA*, 92(17): 7759-7763.

PROVAN J, 2000. Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana* [J]. *Mol Ecol*, 9(12): 2183-2185.

QI YY, XU WJ, XING T, et al., 2015. Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity [J]. *Evol Bio Online*, 11: 239-249.

QIN HN, YANG Y, DONG SY, et al., 2017. List of threatened species of higher plants in China [J]. *Biodivers Sci*, 25(7): 696-744.

RAMSEY J, SCHEMSKE DW, 2002. Neopolyploidy in flowering plants [J]. *Annu Rev Ecol Syst*, 33(1): 589-639.

SHETTY SM, SHAH MUM, MAKALE K, et al., 2016. Complete chloroplast genome sequence of *Musa balbisiana* corroborates structural heterogeneity of inverted repeats in wild progenitors of cultivated bananas and plantains [J]. *Plant Genome*, 9(2): 1–14.

SUN YX, MOORE MJ, LIN N, et al., 2017. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family [J]. *BMC Genomics*, 18: 592.

TAMURA K, STECHER G, PETERSON D, et al, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 [J]. *Mol Biol Evol*, 30(12): 2725-2733.

WANG KJ, ZHANG T, WANG QG, et al., 2018. The phylogenetic position and hybrid origination of *Rosa Praelucens* Byhouwer [J]. *J Plant Genet Resour*, 19(5): 1006-1015.

WICKE S, SCHNEEWEISS GM, DE PAMPHILIS CW, et al, 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function [J]. *Plant Mol Biol*, 76(3-5): 273–297.

WU XY, CHEN M, WANG QG, et al., 2014. Comparative study on the breeding systems of *Rosa praelucens* and *Rosa soulieana* [J]. *Acta Hortic Sin*, 41(10): 2075–2084.

XUE J, WANG S, ZHOU SL, 2012. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae) [J]. *Am J Bot*, 99(6): 240–244.

YANG F, 2019. Sequencing and structural analysis of chloroplast genome in *Rosa banksiae* [J]. *Genom Appl Biol*, 38(8): 3586-3594.

YANG JB, LI DZ, LI HT, 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs [J]. *Mol Ecol Resour*, 14(5): 1024–1031.

YE WQ, YAP ZY, LI P, et al., 2018. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae) [J]. *Mol Phylogenet Evol*, 127: 978-987.

YIN XM, LIAO BS, GUO S, et al., 2020. The chloroplasts genomic analyses of *Rosa laevigata*, *R. rugosa* and *R. canina* [J]. *Chin Med-UK*, 15: 18.

ZHAO L, ZHANG H, WANG QG, et al, 2019. The complete chloroplast genome of *Rosa lucidissima*, a critically endangered wild rose endemic to China [J]. *Mitochondrial DNA B*, 4(1): 1826-1827.

ZHOU YQ, SHU Q, ZHANG H, et al., 2016. Distribution and population quantitative dynamics of critically risked *Rosa praelucens* Byhouwer [J]. *J Plant Genet Resour*, 17(4):649-654.

ZHU A, GUO W, GUPTA S, et al., 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution

rates [J]. *New Phytol*, 209(4): 1747-1756.

Note: Figure translations are in progress. See original paper for figures.