

The complete chloroplast genome of *Rubus rosifolius* (Rosaceae), an ornamental and medicinal plant

Xuelian Yang, Xia Wang, Yongfei Wu, Li Yan, Xiaojing Hu and Wanping Zhang

College of Agriculture, Guizhou University, Guiyang, Guizhou Province, China

ABSTRACT

Rubus rosifolius belongs to the genus *Rubus* in the family Rosaceae and is widely distributed globally. It has white flowers and red fruits. Moreover, it has medicinal value for diseases of the stomach and other areas. However, the complete chloroplast (cp) genome of *R. rosifolius* remains unclear. In the present study, we sequenced the complete cp genome of *R. rosifolius* (GenBank accession no. OL435124), which had a typical quadripartite structure and a size of 155,650 bp. Fifteen genes (*trnK-UUU*, *rps16*, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnI-GAU*, *trnA-UGC*, and *ndhA*) contained an intron, and two genes (*clpP* and *ycf3*) contained two introns. The gene *rps12* showed trans-splicing. Phylogenetic analysis showed that *R. rosifolius* was closely related to *Rubus taiwanicola*, *Rubus rubroangustifolius*, and *Rubus glandulosopunctatus*.

ARTICLE HISTORY

Received 27 February 2022
Accepted 19 June 2022

KEYWORDS

Rubus rosifolius; chloroplast genome; Rosaceae; Illumina sequencing

Rubus rosifolius Sm. 2010 belongs to the genus *Rubus* in the family Rosaceae and is widely distributed globally. This species is native to the Solomon Islands, New Caledonia, Vanuatu (eastern coastal Australia), and Mauritius (southern Indonesia), Indochina, and China (Quadros et al. 2020). *R. Rosifolius* produces beautiful ornamental white flowers and red fruit. *R. rosifolius* is used in traditional medicine to treat diarrhea and stomach diseases and has analgesic, antimicrobial, antihypertensive, and other pharmacological properties (Quadros et al. 2020). Thus, genetic and genomic information on *R. rosifolius* is vital for systematic research and conservation. However, the complete chloroplast (cp) genome of *Rubus rosifolius* remains unclear. In this study, we sequenced the complete cp genome of *R. rosifolius* to promote further study and conservation.

Samples of *R. rosifolius* were collected from the Guizhou Botanical Garden, Guiyang, Guizhou Province, China (26°37'20"N, 106°43'29"E). The College of Agriculture approved this study, which was performed under the National Wild Plant Protective Regulations. The voucher specimen (KXP20210701YX) was deposited at the Laboratory of the College of Agriculture at Guizhou University, Guiyang (contact person: Xuelian Yang, email: yxl1299927812@outlook.com). Total genomic DNA was extracted from 300 mg of fresh leaves using the CTAB method. Libraries with an average length of 350 bp were constructed using a NexteraXT DNA library preparation kit. The Illumina Novaseq 6000 platform was used to sequence these libraries and create raw sequence data. After editing using the NGS QC Tool Kit v2.3.3 (Patel and Jain 2012), we assembled the complete cp genome from 3.82 G of high-quality data using the *de novo*

assembler SPAdes v3.11.0 (Bankevich et al. 2012). Finally, the complete cp genome was annotated using the PGA software (Qu et al. 2019).

The complete cp genome of *R. rosifolius* (GenBank accession no. OL435124) had a typical quadripartite structure and a size of 155,650 bp, including two inverted repeats (IRs; 25,748 bp each), a large single-copy region (85,443 bp), and a small single-copy region (18,711 bp). In this study, 131 genes in the cp genome were annotated, including 86 protein-coding genes (PCGs), 37 tRNA genes, and eight rRNA genes. Among them, we found that five PCGs (*rpl2*, *rpl23*, *ycf2*, *ndhB*, and *rps7*), four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*), and seven tRNA genes (*trnL-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG* and *trnN-GUU*) were duplicated in the IR regions. The overall GC content was 36.9%. The total sizes of the PCGs and RNA genes (tRNAs and rRNAs) were 80,334 and 11,839 bp, respectively; their GC contents were 37.71 and 54.91%, respectively. Thus, the GC content of the RNA genes was much higher than that of the PCGs and the complete cp genome. Furthermore, 15 genes (*trnK-UUU*, *rps16*, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnI-GAU*, *trnA-UGC*, and *ndhA*) contained an intron, and two genes (*clpP* and *ycf3*) contained two introns. Additionally, the gene *rps12* had trans-splicing.

To analyze the phylogenetic relationships of *R. rosifolius* within the genus *Rubus*, we aligned the 74 homologous protein-coding genes in each of the 34 complete cp genomes from NCBI using the MAFFT 7.037 software (Katoh and Standley 2013) and the FFT-NS-2 strategy. Subsequently, we used the model-finder var 1.6 to select the TVM + F + I + G4 model (Kalyaanamoorthy et al. 2017). The phylogenetic tree

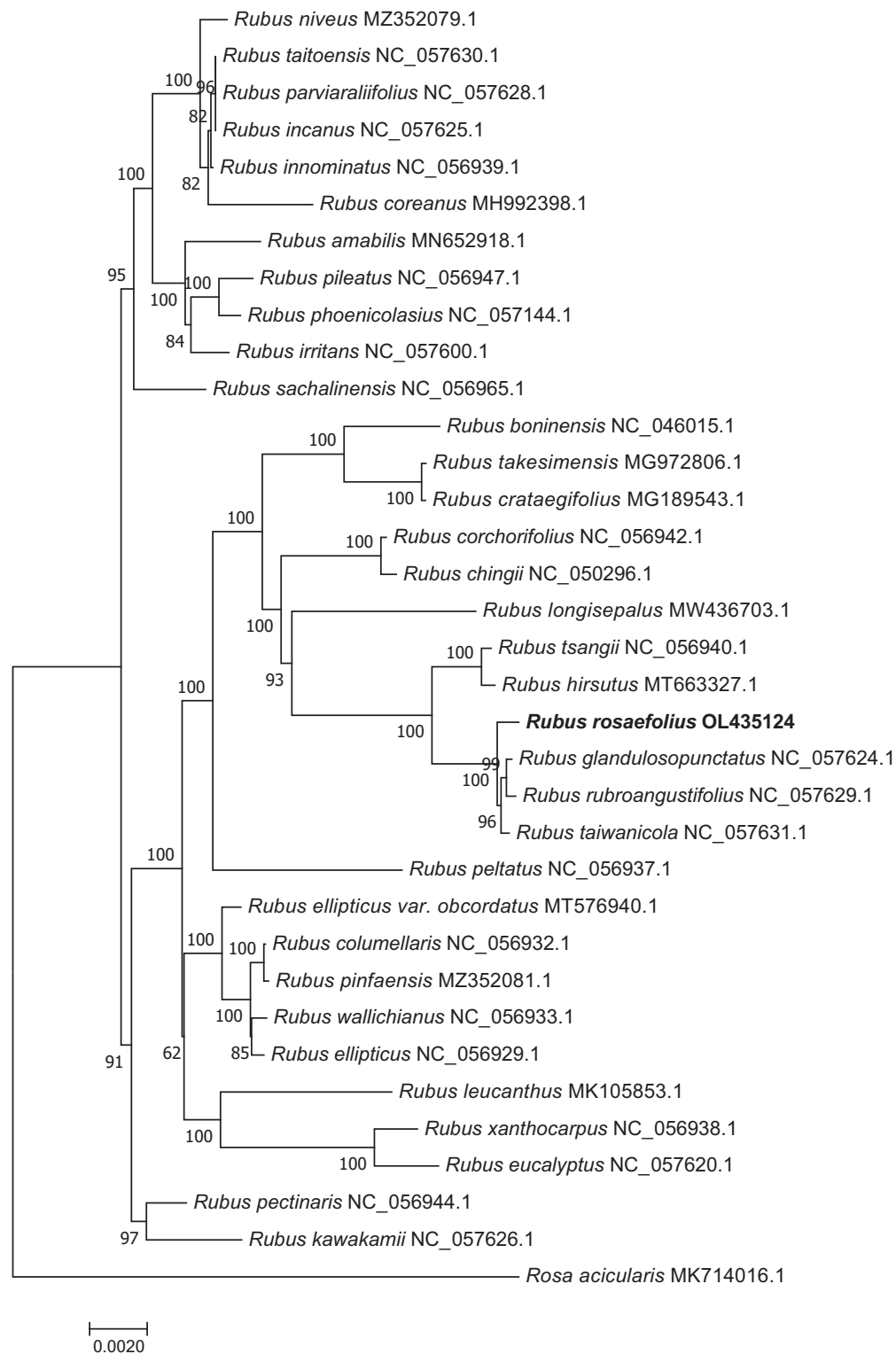


Figure 1. The maximum likelihood phylogenetic tree is based on 74 homologous protein-coding genes in 34 species. *Note:* Numbers to the right of the nodes represent the bootstrap value for 1000 replicates.

was constructed using the RAxML var 8.2.9 software (Stamatakis 2014) and 1000 bootstrap replicates, based on the maximum likelihood method. The results showed that *R. rosaefolius* is closely related to *Rubus taiwanicola* (NC_057631.1), *Rubus glandulosopunctatus* (NC_057624.1), and *Rubus rubroangustifolius* (NC_057629.1), with high bootstrap values (>90) (Figure 1).

Author contributions

Xue-Lian Yang and Xia Wang conceived of and designed the research. Xue-Lian Yang, Xia Wang, Wanping Zhang, and Li Yan collected the samples and performed the experiments. Xue-Lian Yang, Xia Wang, Wanping Zhang, Yong-Fei Wu, Xiang-Jing Hu, and Li Yan analyzed the data. Wanping Zhang and Xiang-Jing Hu drafted and revised the manuscript. All authors approved the final manuscript and agreed to be accountable for all aspects of this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The genome sequence data supporting the findings of this study are available in the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/>) under accession no. OL435124. The associated BioProject, SRA, and Bio-Sample numbers were PRJNA786327, SRR17183907, and SAMN23667426, respectively.

Funding

This work was supported by the National Fund under [Grant 31860225]; the Qian Kehe Foundation (2019) under [Grant 1408]; and the Qian Kehe Platform Personnel (2018) under [Grant 5781].

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a

new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLOS One.* 7(2):e30619.

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 15(1):1–12.

Quadros APOD, Almeida LM, Petreanu M, Niero R, Rosa PCP, Sawaya A, Mantovani MS, Gaivão I, Maistro EL. 2020. Risk assessment via genotoxicity, metabolism, apoptosis, and cell growth effects in a HepG2/C3A cell line upon treatment with *Rubus rosifolius* (Rosaceae) leaves extract. *J Toxicol Environ Health A.* 83(13–14): 495–508.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9): 1312–1313.