

The complete chloroplast genome and phylogenetic analysis of *Rubus sumatranus* Miq 1861 (Rosaceae)

Wei Guo, Longyuan Wang and Wei Wu

College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong, China

ABSTRACT

The wild raspberry species *Rubus sumatranus* Miq 1861 is a promising resource for breeding thermotolerant cultivars. Its complete chloroplast genome spans 155,935 base pairs (bp), featuring the classic quadripartite structure: an 18,729 bp small single-copy region, an 85,662 bp large single-copy region, and two 25,772 bp inverted repeats. A total of 130 genes were identified, including 86 protein-coding, 36 tRNA genes, and 8 rRNA genes. A maximum likelihood phylogenetic tree based on chloroplast genomes shows that *R. sumatranus*, within the subgenus *Idaeobatus*, is sister to the subgenus *Batothamnus*. This confirms the polyphyletic nature of the subgenus *Idaeobatus*. The chloroplast genome assembly of *R. sumatranus* enhances our understanding of its evolutionary history.

ARTICLE HISTORY

Received 18 May 2024
Accepted 29 November 2024

KEYWORDS

Chloroplast genome;
next-generation sequencing;
phylogenomics; Rosaceae;
wild raspberry

Introduction

The *Rubus* L. genus, within the Rosoideae subfamily of the Rosaceae, is cosmopolitan, with a notable prevalence in temperate zones. Comprising around 740 species, the genus was traditionally divided into twelve subgenera (Focke 1910, 1911, 1914). However, recent morphological and molecular evidence suggests a reduction to ten subgenera (Huang et al. 2023). The subgenera *Idaeobatus*, *Rubus*, and *Malachobatus* contain the most species. *Rubus* is one of the most genetically complex genera due to interspecific hybridization, polyploidy, and apomixis (Thompson 1997; Alice et al. 2001). Significant morphological variation within species presents challenges for classification (Robertson 1974). Recent studies, restricted by inadequate sampling and a narrow selection of molecular markers, have left the phylogeny at subgeneric and sectional levels unresolved (Wang et al. 2016; Carter et al. 2019; Wang et al. 2019; Huang et al. 2023).

Rubus sumatranus, a shrub that grows erect or scrambles up to 2 meters high, is native to the warm temperate and subtropical regions of East Asia, including southern China, Japan, and extending to northeast India, Thailand, Laos, Vietnam, Malaysia, and Indonesia. Historically, it was considered synonymous with *Rubus pirifolius* Smith (Focke 1910). The plant features branches, rachises, petioles, pedicels, and inflorescences covered with purplish-red glandular hairs, soft hairs, and bristles (Lu and Boufford 2003) (Figure 1). Classified under the subgenus *Idaeobatus*, *R. sumatranus* is

notable for its abundant orange-red aggregate fruits. Its roots are used in traditional medicine for their heat-clearing, detoxifying and diuretic properties (Lu and Boufford 2003). Given its broad distribution, *R. sumatranus* is a valuable resource for adaptability breeding programs, particularly for developing thermotolerant cultivars. Its resilience across various climates makes it an excellent candidate for genetic studies aimed at enhancing the productivity of cultivated varieties. So far, no genetic resources are available for this species. As a first step toward utilizing this species, we characterized the chloroplast genome and determined the phylogeny placement for *R. sumatranus*.

Materials and methods

Plant material collection and DNA extraction

Fresh leaves from an individual of *R. sumatranus* were collected at Yuanbao Mountain National Natural Reserve in Rongshui County, Guangxi Province (25°22'36.8"N, 109°08'23.2"E). The voucher specimen is deposited at the SYSU Herbarium (Contact person: Wei Wu; e-mail: wuei06@zhku.edu.cn) under accession number GXYBS0021. Total genomic DNA was extracted using a modified cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle 1987) and quantified with a Qubit™ 4.0 Fluorometer (Fisher Scientific, USA).

CONTACT Wei Wu  wuei06@zhku.edu.cn  College of Horticulture and Landscape Architecture, Zhongkai University of Agriculture and Engineering, Guangzhou 510250, Guangdong, China.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2024.2438277>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



Figure 1. *Rubus sumatranus* growing in its natural habitat. A – the whole plant; B – aggregate fruit; C – stems with purplish-red glandular hairs, soft hairs, and bristles; D – infructescence. The photo was taken and redacted by GW & WW at Mount Yuanbao in Rongshui County, Guangxi Province (25°22'36.8"N, 109°08'23.2"E).

Library construction and sequencing

A DNA library with 350–550 bp insert size library was constructed using 1.0 µg DNA according to the Illumina's protocol and sequenced on an Illumina Novaseq 6000 platform (Illumina, USA) using paired-end 150 bp reads.

Sequence assembly and annotation

After trimming adaptors, raw Illumina reads were filtered using fastp v0.23.4 with default settings (Chen et al. 2018). Pair-end clean reads of *R. sumatranus* were processed through the GetOrganelle v1.7.7.0 pipeline for plastome assembly using the parameter '-R 30 -k 65,85,105,115 -F emb-plant_pt' (Jin et al. 2020). Chloroplast genes were annotated with two reliable software tools: CpGAVAS (Liu et al. 2012) and GeSeq v2.03 (Tillich et al. 2017), simultaneously. The annotated chloroplast genome was manually verified and corrected by comparison with the published plastome of *R. leucanthus* (GenBank accession number: MK105853.1; Guo et al. 2019). The chloroplast structure and gene order were visualized using CPGView (Liu et al. 2023).

Phylogenetic tree construction

To determine the phylogenetic position of *R. sumatranus*, we selected complete plastome sequences from 11 other *Rubus* species across six subgenera. Before reconstructing the phylogenetic tree, 82 shared unique CDS sequences from each species were extracted using BEDTools (Quinlan 2014), aligned with MAFFT v7.487 (Katoh and Standley 2013), and

concatenated using AMAS (Borowiec 2016). A maximum likelihood (ML) phylogenetic tree was constructed based on the best-fit model of TVM+F+R5, determined by BIC, with IQ-TREE v2.0 (Nguyen et al. 2015) and 1,000 bootstrap replicates. Two closely related species, *Fragaria vesca* subsp. *vesca* (Rosaceae; GenBank accession number: NC_015206.1; Shulaev et al. 2011) and *Geum aleppicum* (Rosaceae; GenBank accession number: NC_060733.1; Zhang et al. 2022), were included as outgroups.

Results

Out of 10.0 GB of raw Illumina pair-end reads, 9.85 GB of clear reads were obtained for the *R. sumatranus* chloroplast genome assembly. The raw reads were mapped to the final assembly, revealing substantial sequencing coverage. The base depth ranged from 1566× to 6578×, with a mean base depth of 4567×, ensuring high accuracy (see Supplementary material, Figure S1).

The *R. sumatranus* chloroplast genome, spanning 155,935 bp, exhibits the typical quadripartite structure found in other seed plants (Jansen and Ruhlman 2012). It includes two 25,772-bp inverted repeat regions, an 85,662-bp large single-copy region, and an 18,729-bp small single copy region (Figure 2). A total of 130 genes were predicted: 86 protein-coding, 36 tRNA, and 8 rRNA genes. Among these, 18 genes contain two exons and 4 genes contain three exons. The *rps12* genes undergo trans-splicing, as confirmed by CPGView (<https://www.1kmpgcn/cpgview/>) (see Supplementary Figures S2 and S3). The overall GC content is 37.03%, with 31.08% A, 18.87% C, 31.89% T and 18.16% G. The plastome sequence is

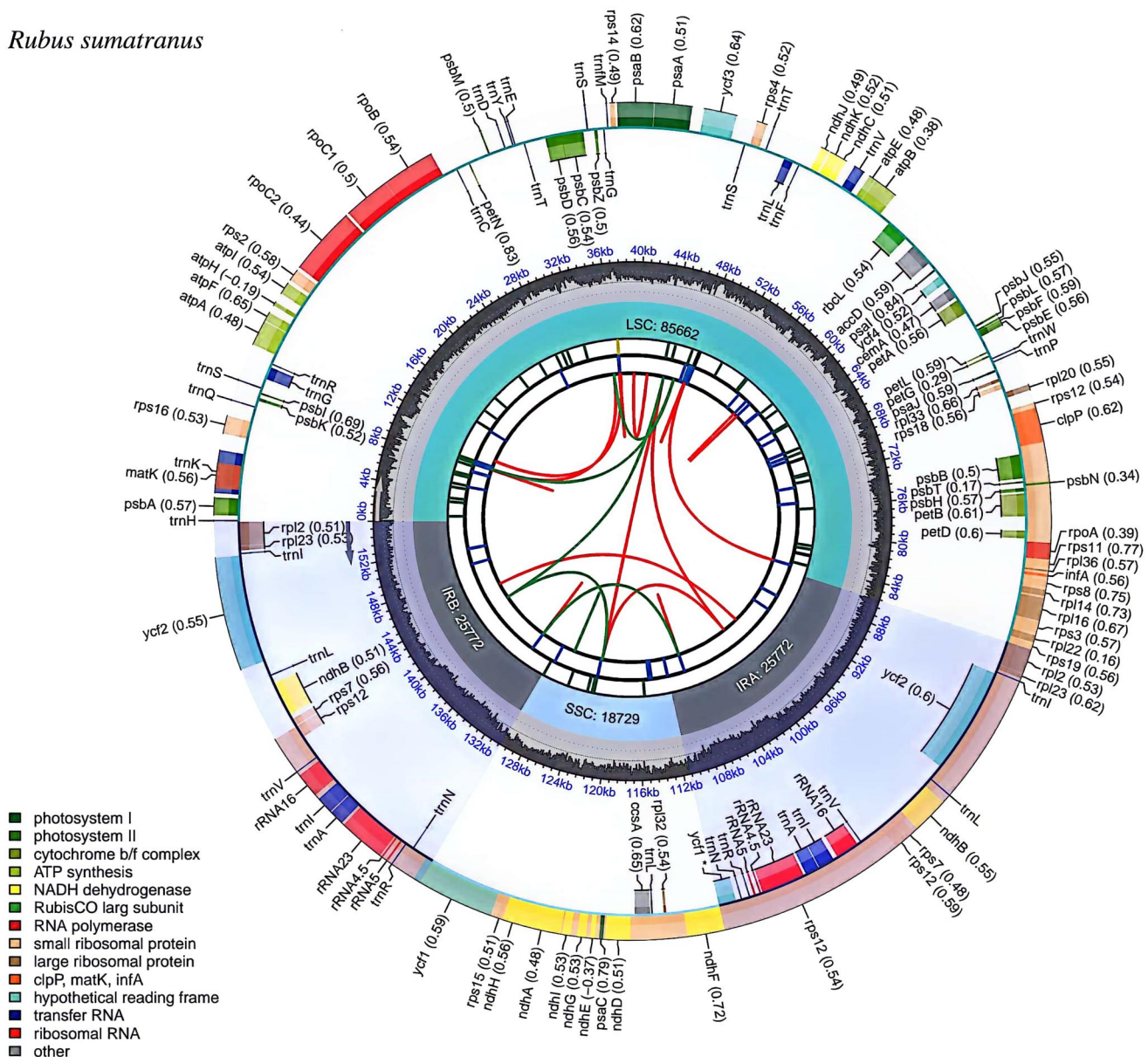
Rubus sumatranus

Figure 2. Circular chloroplast genome map of *Rubus sumatranus*. The innermost track displays dispersed repeats, including direct (red) and palindromic (green) repeats. The next track shows long tandem repeats as short blue bars. The third track depicts short tandem repeats/microsatellites using colored bars, where the color indicates the repeat unit size (black: complex, green: 1 bp, yellow: 2 bp, purple: 3 bp, blue: 4 bp, orange: 5 bp, red: 6 bp). The fourth track outlines the small single-copy (SSC), inverted repeat (IRA and IRB), and large single-copy (LSC) regions. The fifth track plots GC content, and the base frequency is shown between the fourth and fifth tracks. The outermost track displays the genes, color-coded by functional category, with transcription direction indicated (clockwise/counterclockwise). The functional classifications are provided in the bottom left corner.

deposited in the NCBI GenBank database under accession number PP737792.1.

Maximum likelihood phylogenetic analysis revealed that *R. sumatranus* is sister to the subg. *Batothamnus* (*R. leucanthus*, *R. ellipticus*, *R. wallichianus*) with strong support, and subg. *Idaeobatus* is not monophyletic (Figure 3).

Discussion and conclusion

Due to extensive morphological variations, *R. sumatranus* has been treated as synonyms with *Rubus croceacanthus* H. Lév 1912, *Rubus glandulosocalycinus* Hayata 1915, *Rubus dolichocephalus* Hayata 1913, respectively (Huang and Hu 2009). The phylogenetic position of *R. sumatranus* remains unresolved.

For example, it was placed in the sect. *Rosifolii* in subg. *Idaeobatus* (Kikuchi et al. 2022). A recent study, the phylogenetic position of *R. sumatranus* was inconsistent between chloroplast genome and nuclear genomes, suggesting incomplete lineage sorting or hybridizations associated with *R. sumatranus* (Gao et al. 2023). Our preliminary study shows *R. sumatranus* is sister to the subg. *Batothamnus*, with subg. *Idaeobatus* being non-monophyletic, confirming previous findings (Wang et al. 2016; Carter et al. 2019). Given its wide distribution and polymorphic traits, more extensive sampling and informative markers are needed to clarify its taxonomy and phylogenetic position.

Overall, this study characterized the chloroplast genome of *R. sumatranus* and determined its phylogenetic position,

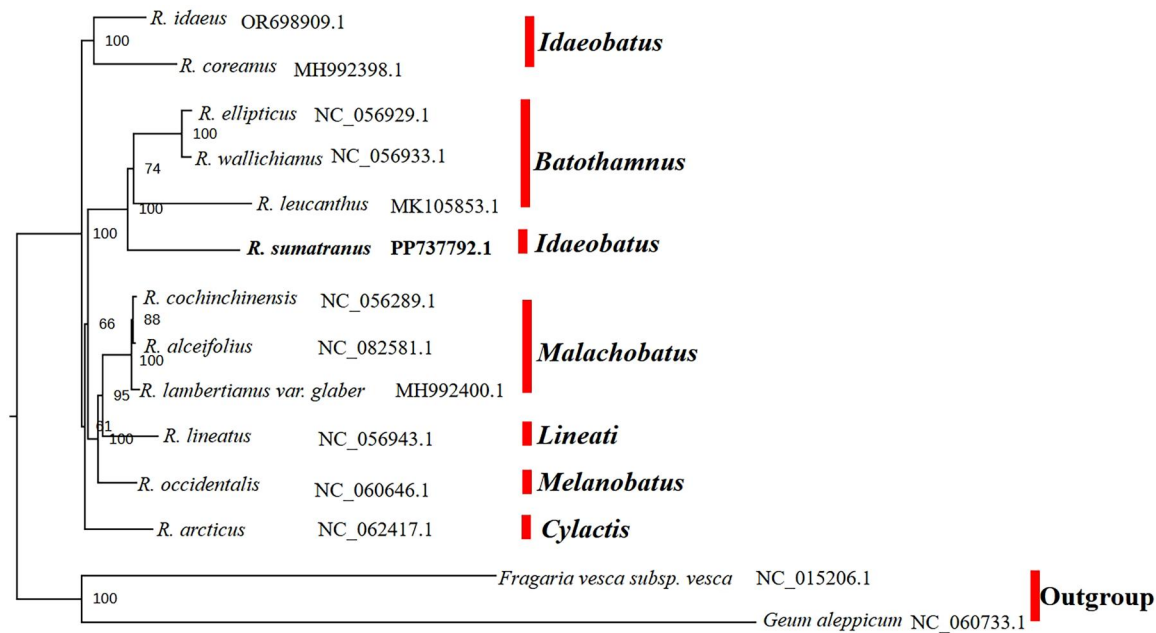


Figure 3. The maximum likelihood tree of 12 *Rubus* species and two outgroups based on chloroplast genome sequences. The 12 species were *R. idaeus* (OR698909.1), *R. coreanus* (MH992398.1) (Chen et al. 2019), *R. ellipticus* (NC_056929.1) (unpublished), *R. wallichianus* (NC_056933.1) (unpublished), *R. leucanthus* (MK105853.1) (Guo et al. 2019), *R. sumatranus* (PP737792.1), *R. cochinchinensis* (NC_056289.1) (Chen et al. 2020), *R. alceifolius* (NC_082581.1) (unpublished), *R. lambertianus* var. *glaber* (MH992400.1) (unpublished), *R. lineatus* (NC_056943.1) (unpublished), *R. occidentalis* (NC_060646.1) (unpublished), *R. arcticus* (NC_062417.1), two outgroups *Fragaria vesca* subsp. *vesca* (NC_015206.1) (Shulaev et al. 2011), *Geum aleppicum* (NC_060733.1) (Zhang et al. 2022). Bootstrap supports were calculated from 1000 replicates. The *R. sumatranus* was labeled in bold font in the phylogenetic tree, and the subgenus position of *Rubus* species were marked with red bar next to it.

providing valuable genomic resources for future breeding efforts involving this wild raspberry species.

Guangxi Provincial Management Bureau for Yuanbao Mountain National Natural Reserve.

Authors contributions

Conception and data acquisition: GW, WLY & WW. Data analysis: GW, WLY & WW. Manuscript – drafting: GW & WW. Manuscript – revision: GW & WW. Photo-taking: GW. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Funding

This study was supported by Science and Technology Program from Forestry Administration of Guangdong Province under Grant number 2021KJCX015, Guangdong Province University Innovative Team Project: Innovation and Development Application of Ornamental Plant Germplasm with Lingnan Characteristics under Grant number 2023KCXTD017.

Data availability statement

The chloroplast genome sequence data in this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov/> under the accession number PP737792. Raw sequencing reads used here have been deposited in the SRA database of NCBI under accession number SRR28810725. The associated 'BioProject', and 'Bio-Sample' numbers are PRJNA1104367 and SAMN41075590, respectively.

Disclosure statement

No potential competing interest was reported by the authors.

Ethics statement

This study was ethically approved and received permission for sample collection from the Guangxi Provincial Department of Forestry and

References

- Alice LA, Eriksson T, Eriksen B, Campbell CS. 2001. Hybridization and gene flow between distantly related species of *Rubus* (Rosaceae): evidence from nuclear ribosomal DNA internal transcribed spacer region sequences. *Syst. Bot.* 26:769–778.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ.* 4:e1660. doi:10.7717/peerj.1660.
- Carter KA, Liston A, Bassil N, Alice LA, Bushakra JM, Sutherland BL, Mockler TC, Bryant DW, Hummer KE. 2019. Target capture sequencing unravels *Rubus* evolution. *Front Plant Sci.* 10:e01615. doi:10.3389/fpls.2019.01615.
- Chen YZ, Guo W, Wang LY, Wu W, Li YQ. 2020. The complete chloroplast genome sequence of *Rubus cochinchinensis*. *Mitochondrial DNA Part B Resour.* 5(2):1636–1637. doi:10.1080/23802359.2020.1745710.
- Chen Q, Wang Y, Sun B, Chen T, Luo Y, Zhang Y, Wang XR, Tang HR. 2019. The complete chloroplast genome sequence of *Rubus coreanus*, an excellent diseases-resistant resource. *Mitochondrial DNA B Resour.* 4(1):216–217. doi:10.1080/23802359.2018.1546139.
- Chen SF, Zhou YQ, Chen YR, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Focke WO. 1910. *Species Ruborum*. Monographiae Generis Rubi Prodromus part I. New York (NY): E. Schweizerbart; p. 1–120.
- Focke WO. 1911. *Species Ruborum*. Monographiae Generis Rubi Prodromus part II. New York (NY): E. Schweizerbart; p. 121–223.
- Focke WO. 1914. *Species Ruborum*. Monographiae Generis Rubi Prodromus part III. New York (NY): E. Schweizerbart; p. 224–498.
- Gao XF, Xiong XH, Boufford DE, Gao YD, Xu B, Zhang C. 2023. Phylogeny of the diploid species of *Rubus* (Rosaceae). *Genes (Basel).* 14(6):1152. doi:10.3390/genes14061152.

- Guo W, Chen Y, Deng L, Wu W. 2019. The complete chloroplast genome sequence of *Rubus leucanthus* Hance (Rosaceae). *Mitochondrial DNA B Resour.* 4(1):376–377. doi:10.1080/23802359.2018.1547170.
- Huang TR, Chen JH, Hummer KE, Alice LA, Wang WH, He Y, Yu SX, Yang MF, Chai TY, Zhu XY, et al. 2023. Phylogeny of *Rubus* (Rosaceae): integrating molecular and morphological evidence into an infrageneric revision. *TAXON.* 72(2):278–306. doi:10.1002/tax.12885.
- Huang JY, Hu JM. 2009. Revision of *Rubus* (Rosaceae) in Taiwan. *Taiwania.* 54(4):285–310.
- Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. *Genomics of chloroplasts and mitochondria*. Dordrecht: Springer; p. 103–126.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21(1):241. doi:10.1186/s13059-020-02154-5.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. doi:10.1093/molbev/mst010.
- Kikuchi S, Mimura M, Naruhashi N, Setsuko S, Suzuki W. 2022. Phylogenetic inferences using nuclear ribosomal ITS and chloroplast sequences provide insights into the biogeographic origins, diversification timescales and trait evolution of *Rubus* in the Japanese Archipelago. *Plant Syst Evol.* 308(3):20. doi:10.1007/s00606-022-01810-6.
- Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour.* 23(3):694–704. doi:10.1111/1755-0998.13729.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* 13(1):715. doi:10.1186/1471-2164-13-715.
- Lu LT, Boufford DE. 2003. *Rubus* L. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China*, vol. 9. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press; pp. 195–285.
- Nguyen LT, Schmidt HA, Haeseler AV, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274. doi:10.1093/molbev/msu300.
- Quinlan AR. 2014. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinform.* 47(1):11.12.1–11.1234. doi:10.1002/0471250953.bi1112s47.
- Robertson RK. 1974. The genera of Rosaceae in the southeastern United States. *J. Arnold Arbor.* 55:352–360.
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet.* 43(2):109–116. doi:10.1038/ng.740.
- Thompson M. 1997. Survey of chromosome numbers in *Rubus* (Rosaceae: Rosoideae). *Ann. Miss. Bot. Gard.* 84(1):128–164. doi:10.2307/2399958.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45(W1):W6–W11. doi:10.1093/nar/gkx391.
- Wang Y, Chen Q, Chen T, Tang HR, Liu L, Wang XR. 2016. Phylogenetic insights into Chinese *Rubus* (Rosaceae) from multiple chloroplast and nuclear DNAs. *Front Plant Sci.* 7:e00968. doi:10.3389/fpls.2016.00968.
- Wang Y, Chen Q, Chen T, Zhang J, He W, Liu L, Luo Y, Sun B, Zhang Y, Tang HR, et al. 2019. Allopolyploid origin in *Rubus* (Rosaceae) inferred from nuclear granule bound starch synthase I (*GBSSI*) sequences. *B. M. C.pl. Biol.* 19:303.
- Zhang P, Wang L, Lu X. 2022. Complete chloroplast genome of *Geum aleppicum* (Rosaceae). *Mitochondrial DNA B Resour.* 7(1):234–235. doi:10.1080/23802359.2021.2024461.