

The complete chloroplast genome of *Prunus phaeosticta* (Hance) Maxim. (Rosaceae) and its phylogenetic implications

Jia-Qi Wu^a, Yi Wang^a, Ping Sun^a, Zhong-Shuai Sun^b and Jian-Sheng Shen^a

^aFlower, Forestry and Fruit Institute, Jinhua Academy of Agricultural Sciences, Jinhua, China; ^bZhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China

ABSTRACT

The complete chloroplast (cp) genome of *Prunus phaeosticta* (Hance) Maxim. has been characterized by reference-based assembly using Illumina paired-end data. The circular complete cp genome is 158,752 bp in length, comprising a large single-copy (LSC) region of 87,085 bp, a small single-copy (SSC) region of 18,923 bp, and a pair of inverted repeats (IRs) of 26,372 bp. A total of 129 functional genes were identified, including 84 protein-coding genes, 37 tRNA genes, and 8 ribosomal RNA genes. The phylogenetic analysis showed that *P. phaeosticta* displayed a kinship to *Prunus zippeliana*.

ARTICLE HISTORY

Received 24 August 2022
Accepted 26 December 2022

KEYWORDS

Prunus phaeosticta;
Laurocerasus; chloroplast
genome; phylogenomics

1. Introduction

Prunus L. (*Prunus sensu lato*) is well known for its taxonomic complexity, the phylogenetic relationships among subgenera, sections, and species are still unclear (Wen et al. 2008; Shi et al. 2013; Chin et al. 2014). The most recent classification of the genus recognized only three subgenera: *Prunus*, *Cerasus* (Mill.) A. Gray, and *Padus* (Mill.) Peterm., with a broader concept of subg. *Padus* that included *Laurocerasus* Duhamel and the former genera *Maddenia* Hook. f. & Thoms and *Pygeum* Gaertn. (Shi et al. 2013). Recent phylogenetic studies revealed *Prunus* subg. *Padus* to be polyphyletic, taxa of subg. *Padus* and subg. *Laurocerasus* are highly intermixed in phylogenetic trees reconstructed using a limited set of short DNA sequence (Liu et al. 2013; Chin et al. 2014; Zhao et al. 2016). The chloroplast genome is a robust and appropriate tool that could provide much better ability than the universal DNA markers on revealing phylogeny and evolutionary history of plants (Gitzendanner et al. 2018; Do et al. 2020; Su et al. 2021). However, genomic resources of *Laurocerasus* species are extremely scarce (Zou et al. 2019). In order to better understand the molecular phylogenetic relationship among the subg. *Padus*, we assembled and characterized the first complete chloroplast genome of *Prunus phaeosticta* (Hance) Maxim. 1883 using the next-generation sequencing technology. Since *P. phaeosticta* is one of the most widespread species, which ranges from India and Thailand all the way to East China and contains seven forms, we chose to sequence it to contribute genetic resources and enhance understanding of this polymorphic species. Furthermore, a phylogenomic analysis of 14 *Prunus* species was also presented. The results will lay a solid foundation for the future phylogenetic researches of *P. sensu lato*.

2. Materials and methods

2.1. Plant materials and DNA extraction

Fresh young leaves of *P. phaeosticta* (Figure 1) were collected from Mt. Jiulong, Suichang County, Zhejiang province, China (28.3850°N, 118.8927°E). A voucher specimen and its DNA sample were deposited at the Herbarium of Zhejiang University (HZU, <https://www.zju.edu.cn/>; collector: Pan Li, panli_zju@126.com) under the voucher number LP208067. Genomic DNA was extracted using a modified CTAB protocol (Doyle and Doyle 1987).

2.2. Plastome sequencing, assembly, and annotation

The next-generation sequencing was performed with an Illumina NovaSeq platform (Illumina, San Diego, CA). The cp genome was assembled via NOVOPlasty 2.6.3 (Dierckxsens et al. 2017), using *Prunus zippeliana* (NC043926, Zou et al. 2019) as reference genome. Genome annotation was conducted using the online software GeSeq (Tillich et al. 2017) by comparing with the *P. zippeliana* cp genome (NC043926, Zou et al. 2019). Geneious R11 (Biomatters Ltd., Auckland, New Zealand) was used for inspecting the genome structure. The circular gene map of was drawn by the cpGav2 web server (<http://47.96.249.172:16019/analyzer/view>).

2.3. Phylogenetic analysis

To reveal the phylogenetic position of *P. phaeosticta*, a phylogenetic analysis intra subg. *Padus* was evaluated based on protein coding regions (CDS) of cp genome of *P. phaeosticta* and 13 reported cp genomes from *Prunus* (NC054253, Li

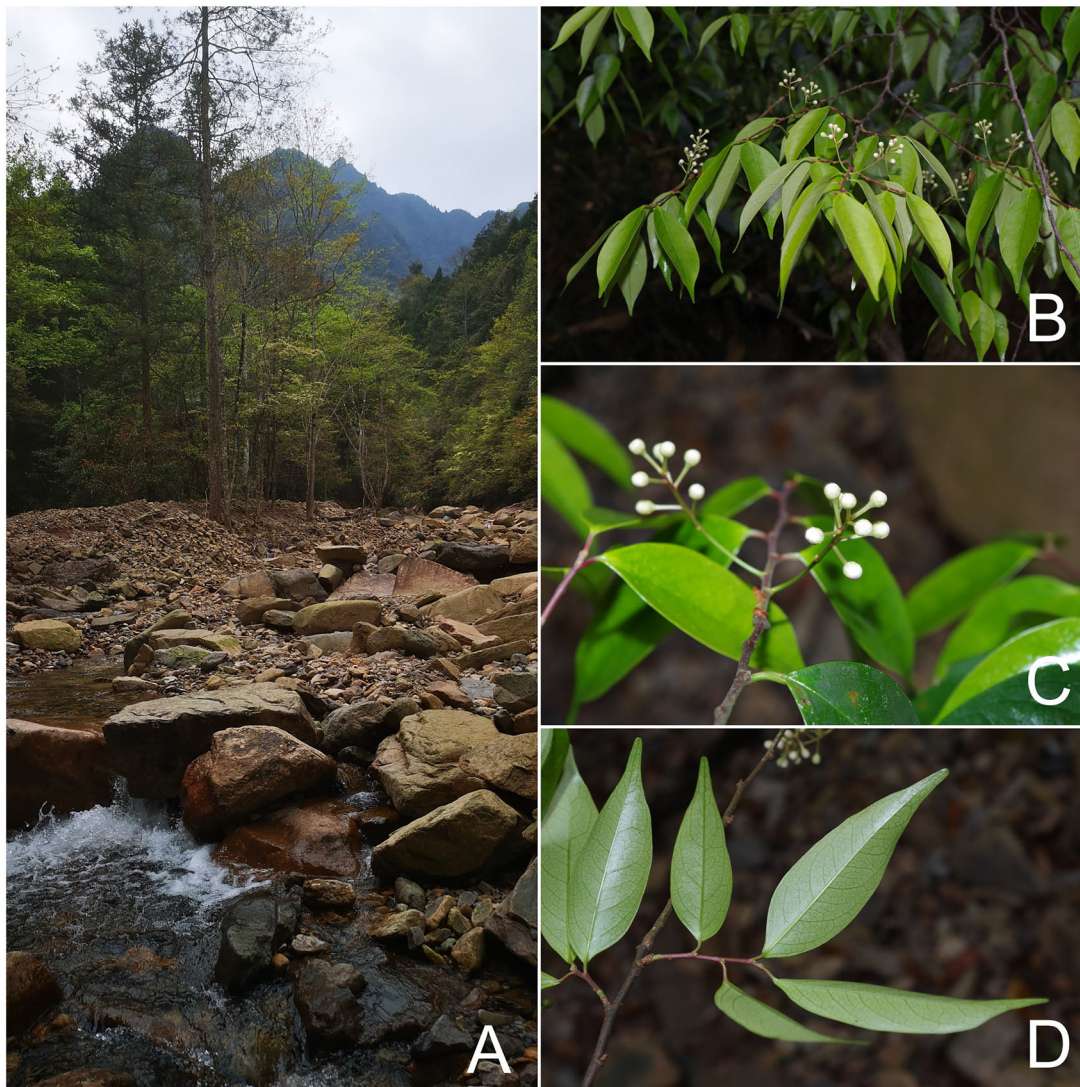


Figure 1. Pictures of *P. phaeosticta* taken on April 14th 2020, at Mt. Jiulong, Suichang County, Zhejiang Province, China. (A) Natural habitat; (B) flowering branches; (C) inflorescences and leaves; (D) abaxial side of leaves.

et al. 2021; NC035891, Xu et al. 2018; NC043926, Zou et al. 2019; NC026982, NC036133, Luan et al. 2018; MW801161 Jin et al. 2021; NC059022, NC059023, NC059024, NC059025, NC059026, NC059027, Su et al. 2021; NC053881, Wang et al. 2021). Eighty-four shared CDS were extracted and aligned by MAFFT (Katoh et al. 2005). *Prunus fasciculata* (NC054253) and *Prunus cerasoides* (NC035891) were selected as outgroup. We reconstructed a phylogeny employing the GTR+G model and 1000 bootstrap replicates under the maximum likelihood (ML) inference in RAxML-HPC v.8.2.10 on the CIPRES cluster (Miller et al. 2010).

3. Results

3.1. Genome organization and compositions

The chloroplast genome of *P. phaeosticta* presented a typical circular DNA molecule with a total length of 158,752 bp (Figure 2). It has a characteristic quadripartite structure with a large single-copy (LSC) region of 87,085 bp, a small single-copy (SSC) region of 18,923 bp, and a pair of inverted repeats (IRs) of 26,372 bp. The total GC content in the chloroplast

genome of *P. phaeosticta* was 38.6%. The GC content of the IR region (43.4%) was higher than those of the LSC (37.0%) and SSC (33.4%) regions.

A total of 129 functional genes were identified, including 84 protein-coding genes (PCGs), 37 tRNA genes, and 8 rRNA genes. Of these genes, 15 genes (*rps16*, *atpF*, *rpoC1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *ndhA*, *trnK-UUU*, *trnG-GCC*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, and *trnA-UGC*) had one intron and 3 genes (*rps12*, *clpP*, and *ycf3*) had two introns. Most genes occurred in a single copy, while six PCGs (*rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*, and *rps12*), seven tRNA genes (*trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnV-GAC*, *trnA-UGC*, *trnR-ACG*, and *trnN-GUU*), and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) in IR regions are duplicated.

3.2. Phylogenetic analysis

A robust phylogeny of subg. *Padus* was obtained based on the CDS data (Figure 3). In our phylogenetic tree, the 14 species were clustered into two main clades: 1) *Laurocerasus* clade; 2) *Maddenia* & *Padus* clade. *Maddenia* species formed a

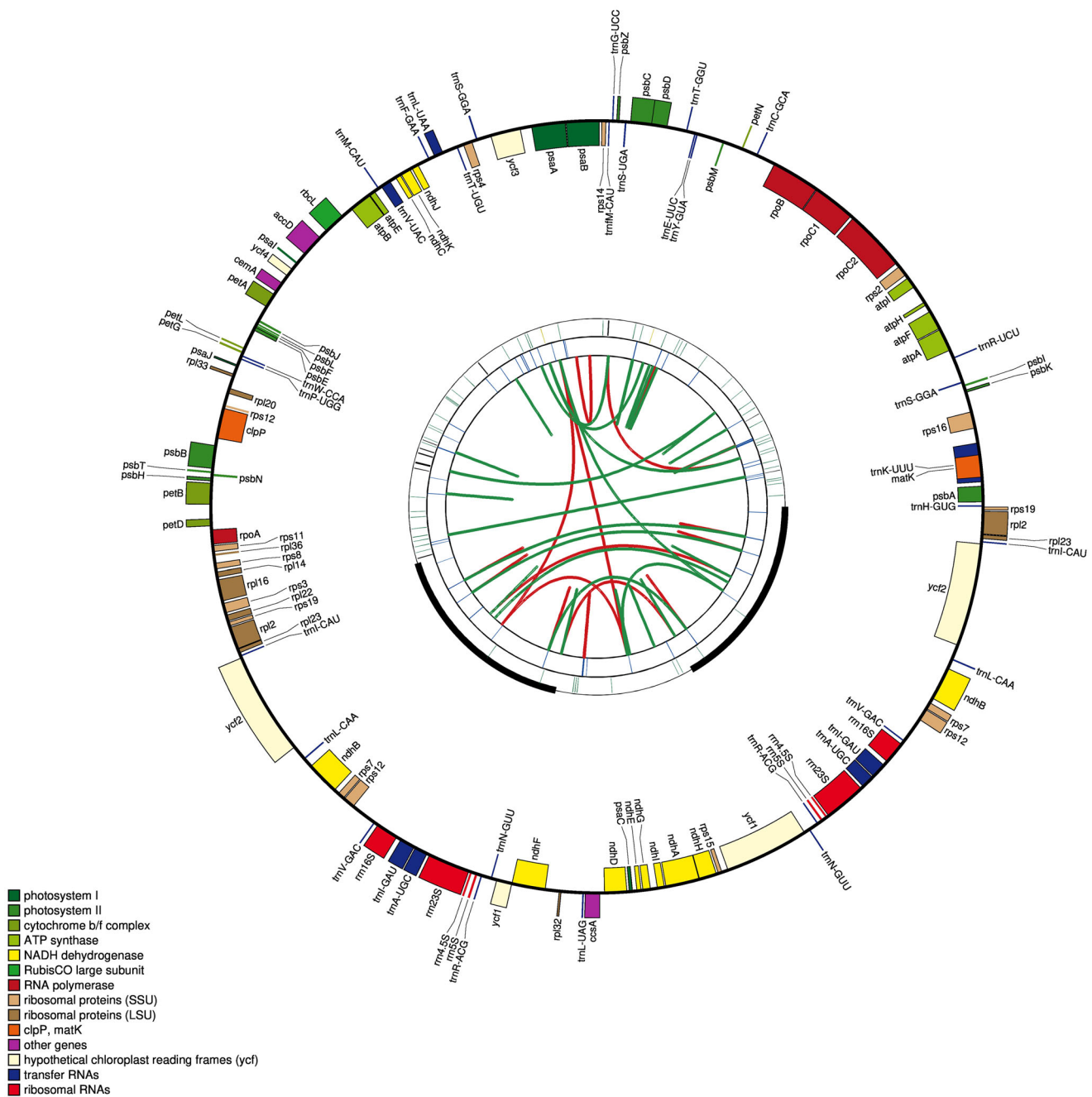


Figure 2. Graphic representation of features identified in *P. phaeosticta* chloroplast genome using CPGAVAS2. The functional classification is shown at the bottom left.

monophyletic clade, but *Padus* species were revealed as a paraphyletic group (Figure 3). According to the phylogenetic tree, *P. phaeosticta* and *P. zippeliana* clustered within the *Laurocerasus* clade and the sister relationships between *P. phaeosticta* and *P. zippeliana* were observed.

4. Discussion

As described previously, *Prunus* L. (*P. sensu lato*) is well known for its taxonomic complexity. Although some comprehensive studies on the phylogenetic relationships of *P. sensu lato* has significantly promoted the understanding of the interspecies relationships in this genus (Wen et al. 2008; Liu

et al. 2013; Shi et al. 2013; Chin et al. 2014; Zhao et al. 2016), but limited data of genome level has become an impediment to resolve the phylogeny and the interspecies relationships of *P. sensu lato*. In this study, we sequenced the chloroplast genome of *P. phaeosticta* and conducted a systematic analysis of the chloroplast genomes with 13 other *Prunus* chloroplast genomes. Our systematic analysis result was consistent with the recent phylogenetic study on *P. sensu lato*, the monophyly of *Maddenia* was supported and the *Padus* species were paraphyletic (Shi et al. 2013; Chin et al. 2014; Su et al. 2021). According to our result, *P. phaeosticta* clustered within the *Laurocerasus* group and exhibited the closest relationship with *P. zippeliana*. We expect that the cp genome of

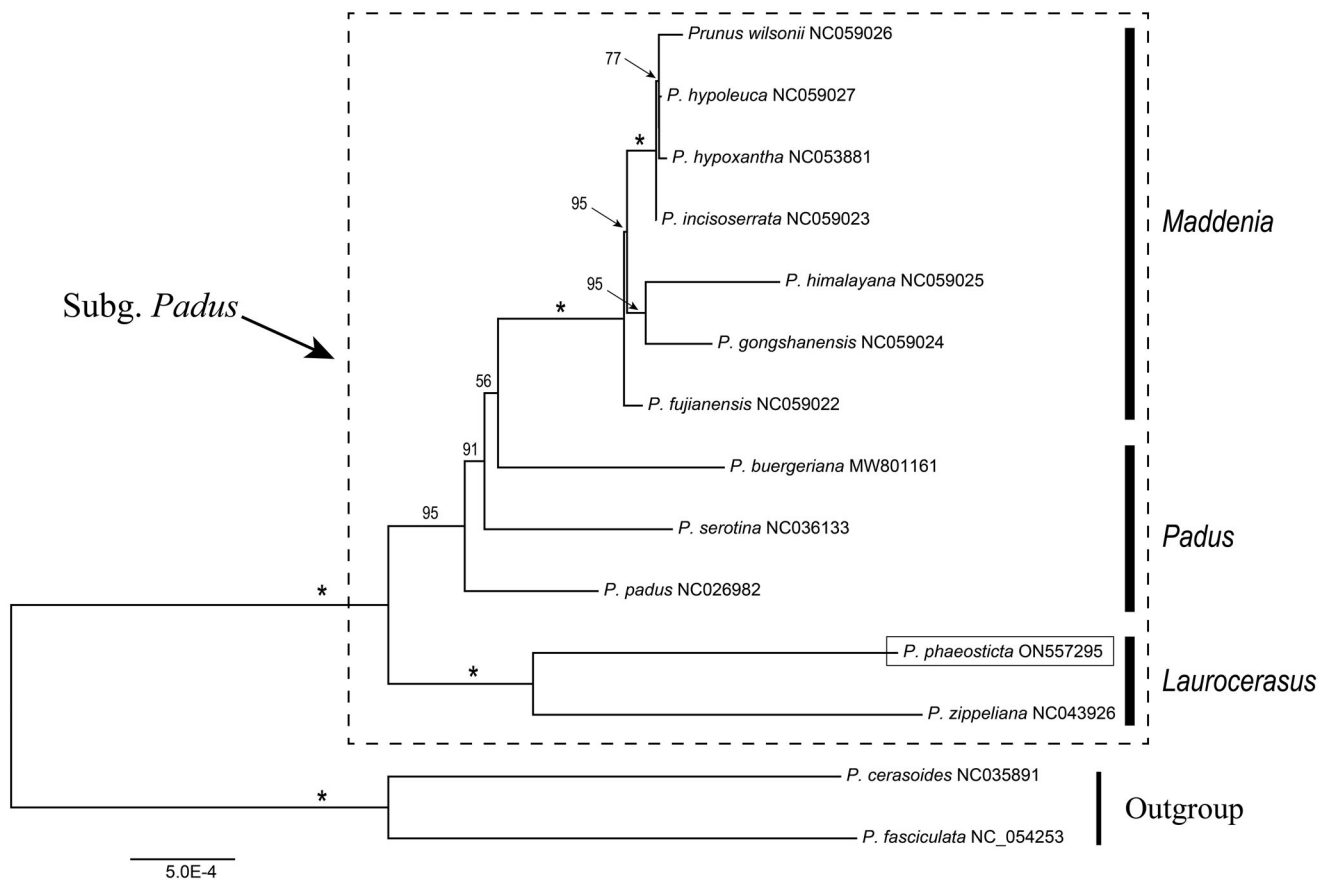


Figure 3. Maximum likelihood (ML) tree reconstruction of 14 taxa from *Prunus* based on 84 shared CDS in the chloroplast genomes. Relative branch lengths are indicated. Support values above the branches are ML bootstrap support; '*' indicates 100% support values.

P. phaeosticta will be a valuable genomic resource for future taxonomy, phylogeny studies on *P. sensu lato* and new *Prunus* cultivar breeding.

Funding

This work was supported by the Biological Seed Industry and Jinhua Modern Agricultural Machinery Laboratory Cooperation Project [ydhz2020ky03].

Ethical approval

The material involved in the article does not involve ethical conflicts. This study was permitted by Jinhua Academy of Agricultural Sciences and Taizhou University, China. All collection and sequencing work was strictly executed under local legislation and related laboratory regulations to protect wild resources.

Author contributions

The article was designed and conceived by Jai-Qi Wu and Jian-Sheng Shen; Yi Wang and Zhong-Shuai Sun assembled and annotated the cp genome; Jai-Qi Wu contributed significantly to phylogenetic analysis and manuscript preparation; Jian-Sheng Shen was involved in interpretation of the data and revised the manuscript critically for intellectual content. All authors approved the final version to be published and agreed to be accountable for all aspects of the work.

Disclosure statement

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

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (<https://www.ncbi.nlm.nih.gov/>) under the accession no. ON557295. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA841240, SRR19348248 and SAMN28578094, respectively.

References

- Chin SW, Shaw J, Haberle R, Wen J, Potter D. 2014. Diversification of almonds, peaches, plums and cherries – molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Mol Phylogenet Evol.* 76: 34–48.
- Dierckx N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
- Do HDK, Kim C, Chase MW, Kim JH. 2020. Implications of plastome evolution in the true lilies (monocot order Liliales). *Mol Phylogenet Evol.* 148:106818.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. 2018. Plastome phylogenetics: 30 years of inferences into plant evolution. *Adv Bot Res.* 85: 293–313.

- Jin L, Liu JJ, Xiao TW, Li QM, Lin L, Shao XN, Ma CX, Li BH, Mi XC, Qiao XJ, et al. 2021. Community phylogenetics require phylogenies reconstructed from plastid genomes. *Authorea*. doi: [10.22541/au.161834751.14170237/v1](https://doi.org/10.22541/au.161834751.14170237/v1).
- Katoh K, Kuma KI, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33(2): 511–518.
- Li YL, Reed J, Xu YQ, Qu DH, Sun ZS. 2021. Complete chloroplast genome of *Prunus fasciculata* (Rosaceae), a species native to western North America. *Mitochondrial DNA B Resour.* 6(2):480–482.
- Liu XL, Wen J, Nie ZL, Johnson G, Liang ZS, Chang ZY. 2013. Polyphyly of the *Padus* group of *Prunus* (Rosaceae) and the evolution of biogeographic disjunctions between eastern Asia and eastern North America. *J Plant Res.* 126(3):351–361.
- Luan A, Gao A, He J, Bi G, He Y. 2018. Characterization of the complete chloroplast genome of black cherry (*Prunus serotina* Ehrh.). *Conservation Genet Resour.* 10(3):367–370.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Comput Environ Workshop.* 14:1–8.
- Shi S, Li J, Sun J, Yu J, Zhou S. 2013. Phylogeny and classification of *Prunus sensu lato* (Rosaceae). *J Integr Plant Biol.* 55(11):1069–1079.
- Su N, Liu BB, Wang JR, Tong RC, Ren C, Chang ZY, Zhao L, Potter D, Wen J. 2021. On the species delimitation of the *Maddenia* group of *Prunus* (Rosaceae): evidence from plastome and nuclear sequences and morphology. *Front Plant Sci.* 12:743643.
- 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.
- Wang JR, Guo YH, Yang JQ, Tong RC, Cao Z, Zhao L. 2021. Complete plastome sequence of *Prunus hypoxantha* (Rosaceae), a species endemic in west of China. *Mitochondrial DNA B Resour.* 6(1):234–235.
- Wen J, Berggren ST, Lee CH, Ickert-Bond S, Yi TS, Yoo KO, Xie L, Shaw J, Potter D. 2008. Phylogenetic inferences in *Prunus* (Rosaceae) using chloroplast *ndhF* and nuclear ribosomal ITS sequences. *J Syst Evol.* 46(3):322–332.
- Xu X, Wen J, Wang W, Zheng W. 2018. The complete chloroplast genome of the threatened *Prunus cerasoides*, a rare winter blooming cherry in the Himalayan region. *Conserv Genet Resour.* 10(3):499–502.
- Zhao L, Jiang XW, Zuo Y, Liu XL, Chin SW, Haberle R, Potter D, Chang ZY, Wen J. 2016. Multiple events of allopolyploidy in the evolution of the racemose lineages in *Prunus* (Rosaceae) based on integrated evidence from nuclear and plastid data. *PLoS One.* 11(6):e0157123.
- Zou P, Dai S, Wang W, Liu G. 2019. The complete chloroplast genome of *Laurocerasus zippeliana* (Rosaceae). *Mitochondrial DNA B Resour.* 4(2): 3264–3265.