

The complete chloroplast genome sequence of *Malus prattii* (Rosaceae) and its phylogenetic analysis

Lu Fan^a, Ling Qin^a, Jie Yan^a, Chuanyuan Mo^a, Chunxiao Rong^a, Ying Meng^b and Manrang Zhang^a

^aCollege of Horticulture, Northwest A&F University, Yangling, China; ^bForestry Bureau of Luonan County, Shangluo, Shaanxi, China

ABSTRACT

Malus prattii (Rosaceae) is an endemic species to China. Here, the complete chloroplast genome of *M. prattii* was obtained by using Illumina pair-end sequencing. The *M. prattii* complete chloroplast genome was 160,239 bp in length, containing a large single-copy region (LSC) of 88,355 bp and a small single-copy region (SSC) of 19,176 bp, which were separated by a pair of inverted repeat regions (IRs) of 26,354 bp. The chloroplast genome of *Malus prattii* contained 112 unique functional genes, including 78 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. The overall C + G content of the whole genome was 36.5%. A neighbour-joining phylogenetic analysis suggested a close relationship between *M. prattii* and *M. micromalus*.

ARTICLE HISTORY

Received 7 May 2019
Accepted 17 May 2019

KEYWORDS

Malus prattii; chloroplast genome; Rosaceae; phylogenetic analysis

Malus prattii is a deciduous tree, grows to 10 m, belongs to a genus of plants in the family Rosaceae. It is endemic to China and is mainly distributed in western provinces, which occurs in mixed forests on slopes at elevations from 1400 m to 3500 m (Editorial Committee of Flora of China, Chinese Academy of Sciences 1974). *Malus prattii* is very popular among people for its charming flower. This plant has stronger cold-resistant capacity (Luby et al. 1998) and disease resistance (Abe et al. 2007), which have breeding potential especially for rootstocks and ornamentals. In this study, we report and characterize the complete chloroplast of *M. prattii* based on Illumina paired-end sequencing data, which are a resource for future studies on the taxonomy and utilization of Rosaceae. The annotated genomic sequences have been submitted to GenBank under accession number MH929090 for *M. prattii*, respectively.

The voucher specimen (FL-05-13) of *M. prattii* has been deposited in the Institute of College of Horticulture, Northwest A&F University, Yangling, China. Total genomic DNA was isolated from fresh leaves of a single individual (Yangling, Shaanxi, China; 34°30'N, 108°04'E) of *M. prattii* using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA), and was used for the subsequent shotgun library following the manufacturer's protocol for Novaseq6000 Sequencing System (Illumina, CA, USA). All of 37,918,958 of 150 bp raw paired-end reads were generated. The program MITObim v1.8 (Hahn et al. 2013) was employed for the assembly of chloroplast genome, with that of *Malus micromalus* (GenBank accession MF062434) as the reference. The complete cp genomes were annotated using the online program Dual Organellar Genome

Annotator (DOGMA) (Wyman et al. 2004). In addition, we used tRNAscan-SE v2.0 (Lowe and Chan 2016) to confirm the identified transfer RNAs (tRNAs).

The complete chloroplast genome of *M. prattii* was 160,239 bp long, consisting of a pair of inverted repeat regions (IRs, 26,354 bp each), separated by a large single-copy (LSC, 88,355 bp) region and one small single-copy (SSC, 19,176 bp) region. The cpDNA of *M. prattii* contained 112 functional genes, including 78 protein-coding genes (PCG), 30 transfer RNA genes (tRNA), and 4 ribosomal RNA genes (rRNA), of which 19 were duplicated in the IRs, including 8 protein-coding genes (*rpl2*, *rpl23*, *rps7*, *rps12*, *rps19*, *ndhB*, *ycf1*, and *ycf2*), 7 tRNA genes (*trnA-UGC*, *trnL-CAA*, *trnV-GAC*, *trnI-CAU*, *trnR-ACG*, *trnI-GAU*, and *trnN-GUU*), and 4 rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*). Among the 112 unique genes, 12 protein-coding genes (*atpF*, *rps12*, *rps16*, *rpoC1*, *ycf3*, *clpP*, *rpl16*, *rpl2*, *petB*, *petD*, *ndhA*, and *ndhB*) and six transfer RNA (*trnA-UGC*, *trnV-UAC*, *trnK-UUU*, *trnG-UCC*, *trnL-UAA*, and *trnI-GAU*) contained one or two introns. The nucleotide composition is asymmetric (31.3% A, 18.6% C, 17.9% G, 32.1% T) with an overall C + G content of 36.5%. The GC content of cp genome in the LSC (large single copy), SSC (small single copy), IRA (inverted repeat A region), and IRB (inverted repeat B region) is 34.2%, 30.4%, 42.7%, and 42.7%, respectively. The whole complete chloroplast genome sequence data of all 15 species were used to construct the phylogeny tree. A neighbor-joining (NJ) phylogenetic tree result showed that *M. prattii* is closely related to *M. micromalus* (Figure 1).

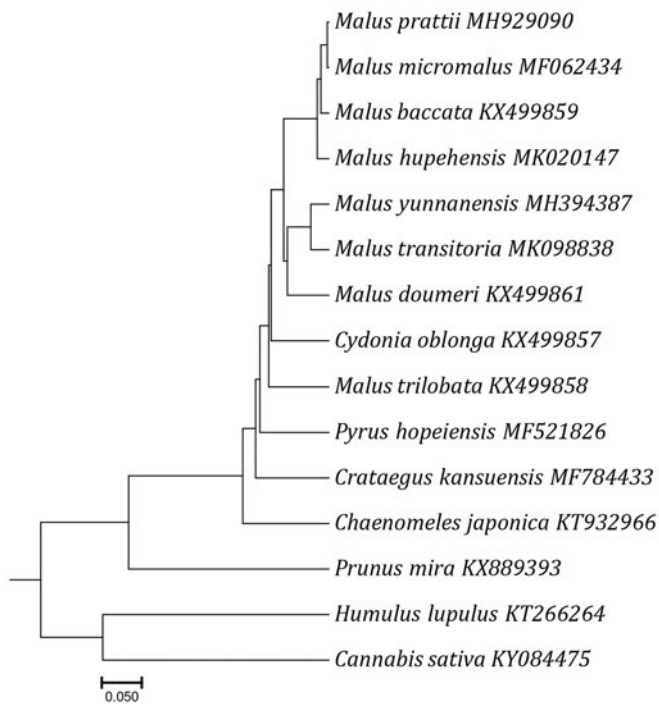


Figure 1. Neighbor-joining (NJ) phylogenetic tree reconstruction including 15 species based on all chloroplast genomes. *Humulus lupulus* and *Cannabis sativa* (Moraceae) were used as the outgroup.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the earmarked fund for China Agriculture Research System [CARS-27].

References

- Abe K, Kotoda N, Kato H, Soejima J. 2007. Resistance sources to Valsa canker (*Valsa ceratosperma*) in a germplasm of diverse *Malus* species. *Plant Breed.* 126:449–453.
- Editorial Committee of Flora of China, Chinese Academy of Sciences. 1974. *Flora of China*. Vol. 36. Beijing: Science Press.
- Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41:e129–e129.
- Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44: W54–W57.
- Luby J, Hoover E, Peterson M, Larson D, Bedford D. 1998. Cold hardiness in the USDA malus core germplasm collection. *Eucarpia Symp Fruit Breed Genet.* 484:109–114.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20:3252–3255.