

Characterization of complete chloroplast genome of *Malus sylvestris* L

Yunfang Xu^{a,b}, Yujie Zhao^{a,b}, Xueqing Zhao^{a,b}, Xuesen Chen^c and Zhaohe Yuan^{a,b}

^aCo-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China; ^bCollege of Forestry, Nanjing Forestry University, Nanjing, China; ^cState Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, Shandong, China

ABSTRACT

The European wild apple (*Malus sylvestris* L.) is an important economical fruit crop. In this present study, we characterized the complete chloroplast (cp) genome sequence of *Malus sylvestris* L. The complete cp genome is 159,926 bp in length with a typical quadripartite structure, containing a large single-copy region (88,064 bp), a small single-copy region (26,353 bp) and a pair of inverted repeat regions (19,157 bp each). A total of 110 unique genes were found in the newly sequenced genome, including 78 protein-coding genes, 28 tRNA genes, and 4 rRNA genes. Of these, 6 protein-coding genes, 7 tRNA genes, and all 4 rRNA genes are duplicated in the inverted regions. A phylogenetic tree was reconstructed using the neighbor-joining method based on the full length of cp genomes within genus *Malus*. The result showed that *M. sylvestris* L. was clustered together with the cultivated apple. The complete cp genome could provide valuable information for understanding the phylogenetic relationships within the genus *Malus*.

ARTICLE HISTORY

Received 20 May 2019
Accepted 31 May 2019

KEYWORDS

Malus sylvestris L.; whole genome resequencing; chloroplast genome; phylogenetic analysis

Malus sylvestris L., a small tree native to Europe, belongs to the genus *Malus* of the family Rosaceae. It has been identified as one of the main contributors to the domesticated apple *M. domestica* (Ruhsam et al. 2018). In 1986, the tobacco chloroplast genome was first reported (Shinozaki et al. 1986), and subsequently, the number of cp genome sequences has increased due to the rapid development of high-throughput sequencing technologies. The sequencing results are available in public sequence repositories, which are widely used for population genetics, taxonomic, and phylogenetic studies (Lu et al. 2018). Phylogenomic analysis using complete cp genome sequences provides new insight into the evolution and domestication of some important fruit trees (Nikiforova et al. 2013; Jose et al. 2015). In addition, comparing and analyzing cp genomes of related species helps to understand the relationship between the important traits under the control of plastid genome (Yan et al. 2019). In this study, we generated the complete chloroplast genome sequence of *M. sylvestris*, which will be helpful to elucidate the phylogenetic relationship between *M. sylvestris* and other species within genus *Malus*.

Total genomic DNA was extracted from fresh young leaf tissue sampled from Shandong Agricultural University using the DNeasy Plant Mini Kit (Qiagen, Venlo, the Netherlands). The voucher specimen was stored in Shandong Agricultural University (117.112 E, 36.194 N). Illumina paired-end (PE) library was prepared and sequenced using Illumina HiSeq

2500 platform (Illumina, San Diego, CA) using the fastp program to filter the raw paired-end reads (Chen et al. 2018).

Subsequently, a *de novo* assembly was performed using GetOrganelle and applied the reads which are of high quality (Jin et al. 2018). Annotation result was inspected by the online program GeSeq (Tillich et al. 2017) and adjusted manually as needed using Geneious (Kearse et al. 2012). The genome was deposited in the GenBank (MK434923).

The *M. sylvestris* cp genome is 159,888 bp in size. It shows a typical quadripartite structure, including a pair of inverted repeat regions (IRs, 26,347 bp), a large single copy (LSC, 88,035 bp), and a small single copy (SSC, 19,159 bp). In total, 110 unique genes were predicted on this cp genome, including 78 protein-coding, 28 tRNA, and 4 rRNA genes. Six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, *ycf2*), seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*), and all four rRNA genes (*rrn4.5S*, *rrn5S*, *rrn16S*, *rrn23S*) were found to be duplicated. 17 genes contain introns, of which three genes (*clpP*, *rps12*, *ycf3*) exhibit two introns. The overall GC content is 36.58%. The GC content of IRs (42.71%) was higher than LSC (34.26%) and SSC (30.43%).

To resolve the phylogenetic relationships within genus *Malus*, a neighbor-joining (NJ) tree was constructed based on complete cp genome sequences. *Prunus dulcis* and *Prunus serotina* were selected as outgroups. Multiple sequences alignment was performed by HomBlocks. The NJ tree was inferred using MEGA7 based on Kimura 2-parameter model.

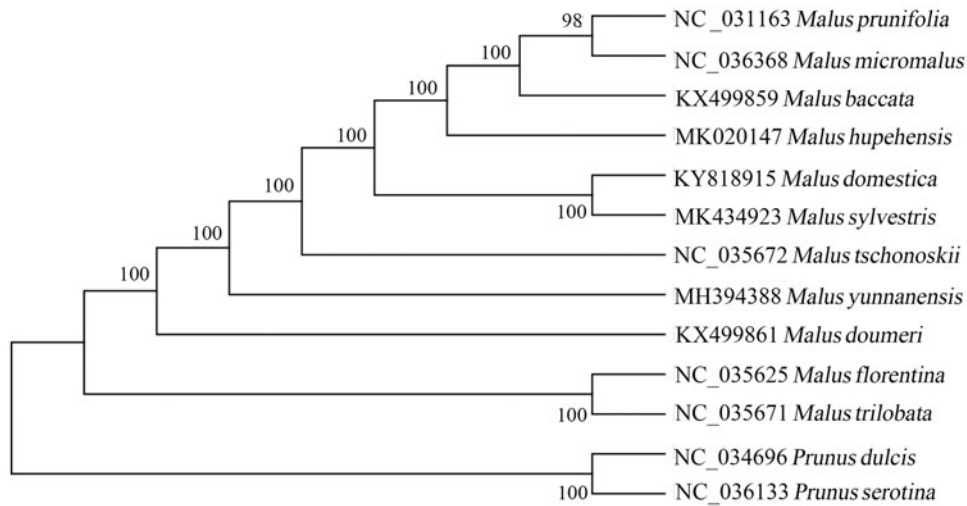


Figure 1. The phylogenetic tree was inferred using the neighbor-joining method. The numbers near each node were bootstrap values inferred from 1000 replicates.

The result (Figure 1) indicated that *M. sylvestris* was phylogenetically close to *M. domestica*.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Natural Science Foundation of Jiangsu Province [BK20180768], the Initiative Project for Talents of Nanjing Forestry University [GXL2014070, GXL2018032], the Priority Academic Program Development of Jiangsu High Education Institutions (PAPD).

References

- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34:i884–i890.
- Jin JJ, Yu WB, Jun B, Yang JB, Song Y, Yi TS, Li DZ. 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv*. DOI:10.1101/256479
- Jose CC, Roberto A, Victoria IE, Javier T, Manuel T, Joaquin D. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol*. 32:2015.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28: 1647–1649.
- Lu L, Li X, Hao ZD, Yang LM, Zhang JB, Peng Y, Xu HB, Lu Y, Zhang J, Shi JS, et al. 2018. Phylogenetic studies and comparative chloroplast genome analyses elucidate the basal position of halophyte *Nitraria sibirica* (Nitrariaceae) in the Sapindales. *Mitochondrial DNA Part A*. 29: 745–755.
- Nikiforova SV, Duccio C, Riccardo V, Vadim G. 2013. Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Mol Biol Evol*. 30:1751–1760.
- Ruhsam M, Jessop W, Cornille A, Renny J, Worrell R. 2018. Crop-to-wild introgression in the European wild apple *M. sylvestris* in Northern Britain. *Forestry*. 92:85–96.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J*. 5:2043–2049.
- 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:W6–W11.
- Yan M, Zhao X, Zhou J, Huo Y, Ding Y, Yuan Z. 2019. The complete chloroplast genome of cultivated apple (*Malus domestica* Cv. ‘Yantai Fuji 8’). *Mitochondrial DNA Part B*. 4:1213–1216.