The complete chloroplast genome sequence of *Malus toringo* (Rosaceae)

Yanan Lia, Yanlei Liub, Chao Xub, Fengyi Lia, Ling Wang and Shiliang Zhoub,c

aCollege of Landscape Architecture, Northeast Forestry University, Harbin, China; bState Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China; cCollege of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

*Malus toringo* (Siebold) Siebold ex de Vriese is a worthy species in the genus Malus (family Rosaceae) distributing from North America and Asia, especially in China. In addition to its ornamental value, the medicinal value of its fruits has also been developed to treat food stagnation diseases. Herein, we obtained the complete chloroplast genome of *M. toringo* using next-generation sequencing technology. The complete chloroplast genome is 160,039 base pair (bp) in length, including a large single copy (LSC) region of 88,142 bp, a small single copy (SSC) region of 19,183 bp, and a pair of inverted repeat regions (IRs, 26,357 bp). And, the overall GC content of the chloroplast genome was 36.6%. Besides, a total of 128 unique genes were found in the chloroplast genome, namely 83 protein-coding genes (PCGs), 37 tRNA genes (tRNA), and eight rRNA genes (rRNA). A maximum-likelihood phylogenetic tree was reconstructed using the full length of chloroplast genome to show the phylogenetic relationships among species in genus Malus. It was concluded that *M. toringo* was phylogenetically close to *M. angustifolia*.

Introduction

*Malus toringo* (Siebold) Siebold ex de Vriese is a worthy species in the genus Malus (family Rosaceae) mainly distributed in North America and Asia in a mixed forest or shrublands, with an elevation of 100–2000 m (Editorial Committee of Flora of North America 2015). In recent years, in addition to its ornamental value, the medicinal value of its fruits has also been developed to treat food stagnation diseases (Cui et al. 1993). However, its chloroplast genome has not been reported. Therefore, we examined the structure of complete chloroplast (CP) genome of *M. toringo* to explore the phylogenetic relationship in genus Malus by comparing it with other species’ chloroplast genome sequences within this genus.

The voucher specimen (BOP040556) of *M. toringo* was preserved in the herbarium, Institute of Botany, Chinese Academy of Sciences (PE). Molecular material was collected from the Royal Botanic Gardens, Kew (England; N 51° 28’ 43.57”, W 0° 17’ 44.12”). Total genomic DNA was isolated via modified cetyltrimethylammonium bromide (mCTAB) method (Li et al. 2013). An Illumina DNA library was constructed using total DNA and the DNA library was sequenced on Illumina HiSeq2500 platform for paired-end 150 bp reads (PE150). The genome was assembled de novo using SPades 3.9 (Bankevich et al. 2012) and annotated on Plann with reference to *M. angustifolia* (NC045410). Finally the complete chloroplast genome was corrected with Sequin (Huang and Cronk 2015) and submitted to GenBank under accession number MT593044.

The complete chloroplast genome of *M. toringo* is 160,039 bp in length with a typical quadripartite structure. It includes two copies of inverted repeats (IRs, 26,357 bp), a large single copy (LSC, 88,142 bp), and a small single copy (SSC, 19,183 bp). Overall GC content was 36.6% and those in LSC, SSC, and IR regions were 34.2%, 30.4%, and 42.7%, respectively. A total of 128 genes were annotated, containing 83 protein-coding genes (PCGs), 37 transfer RNA genes (tRNA), and eight ribosomal RNA genes (rRNA).

To resolve the phylogenetic relationships, the complete chloroplast genome sequences of eighteen species from genus Malus and two species from genus Prunus as out-groups were downloaded from the GenBank. Multiple sequences alignment was executed by MAFFT (Katoh and Standley 2013). The maximum-likelihood (ML) phylogenetic tree with 1000 bootstrap replicates was constructed using the IQ-TREE with the best-fit model determined using ModelFinder (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017). The phylogenetic tree indicated that *M. toringo* was phylogenetically close to *M. angustifolia* (100% ultrafast boot-
strap support, Figure 1) (Hoang et al. 2018). Our findings provide valuable information on the chloroplast genome of the genus and related group in Rosaceae.

**Disclosure statement**

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

**Funding**

This study was supported by the National Nature Fund Project [31670344], the National Fund Project [31872679] and the Strategic Priority Research Program of the Chinese Academy of Sciences [XDA 23080204].

**Data availability statement**

The data that support the findings of this study are openly available in NCBI GenBank database at [https://www.ncbi.nlm.nih.gov] with the accession number is MT593044, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**References**

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pjihelski AD, Pyshkin AV, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Cui Y, Ran X, Hua H, Zhang L, Wu J, Wu H. 1993. Chinese medicine sea. Harbin: Harbin Press.

Editorial Committee of Flora of North America. 2015. Flora of North America North of Mexico. Vol. 9. England: Oxford University Press.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 35(2):518–522.

Huang DI, Cronk Q. 2015. Plann: a command-line application for annotating plastome sequences. Appl Plant Sci. 3(8):1500026.

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

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

Li J, Wang S, Yu J, Wang L, Zhou S. 2013. A modified CTAB protocol for plant DNA extraction. Chinese Bulletin of Botany. 48:72–78.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.