The complete chloroplast genome sequence of Sorbus amabilis (Rosaceae)

Xingwang Zhang\textsuperscript{a,b}, Yanping Xie\textsuperscript{a}, Shuping Xiao\textsuperscript{c} and Xiaomin Wu\textsuperscript{a}

\textsuperscript{a}School of Life Sciences, Huaibei Normal University, Huaibei, China; \textsuperscript{b}School of Information, Huaibei Normal University, Huaibei, China; \textsuperscript{c}Forestry Bureau of Mingxi County, Sanming, China

ABSTRACT

Sorbus amabilis Cheng ex Yü, a small excellent ornamental tree species, is only distributed in Eastern China. In this study, we assembled and annotated the complete chloroplast (cp) genome of the species using the next-generation sequencing for the first time. The cp genome was 160,006 bp in size, consisting of two copies of inverted repeat (IR) regions of 26,405 bp, one large single-copy (LSC) region of 87,870 bp, and one small single-copy (SSC) region of 19,326 bp. The overall GC content of the genome was 36.55%. The genome was predicted to contain 128 genes, including 88 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Phylogenetic analysis of 25 chloroplast genomes in Rosaceae indicated that S. amabilis is most closely related to S. commixta. These findings may provide useful information to the phylogeny of the genus Sorbus.

Sorbus amabilis Cheng ex Yü is a small deciduous broad-leaved tree species belongs to the genus Sorbus of the family Rosaceae. It can only be found in eastern China and grows at an altitude ranging from 900 to 2100 m above sea level. As an excellent ornamental plant, it is very beautiful with pinnate leaves, white flowers, and red fruits. Furthermore, S. amabilis is also considered as an important economic tree species for producing timber and extracting medicine (Zhang et al. 2018). The taxonomy of Sorbus and its subgenus group classification is very complicated due to the existence of polyploidy gametophytic apomixis and natural hybridization within genus (Wang et al. 2020). The chloroplast (cp) genome has been extensively applied in species identification, uncovering plant phylogeny and evolution and in recent years, therefore the complete cp genome of S. amabilis was firstly determined based on the next-generation sequencing technology, which could provide more informatics data for the phylogeny of Sorbus.

Fresh leaves of S. amabilis were sampled from Shixin peak, Huangshan Mountain, Anhui Province, China (30.14'N, 118.18'E). The voucher specimens were preserved at the Key Laboratory of Plant Resource and Biology in Huaibei Normal University with the accession number of HS20190526007. Total DNA extraction and whole genome sequencing were conducted by Nanjing GenePioneer Biotechnologies Inc. (Nanjing, China) with the Illumina Hiseq X Ten platform. Raw data was filtered using the fatap 0.20.0 (Chen et al. 2018). A total of 18,903,785 clean reads were produced and assembled by NOVOplasty 2.7.2 (Dierckxsens et al. 2016).

Annotation was performed using the CpGAVAS pipeline (Liu et al. 2012) and BLAST searches and coupled with manual correction to remove errors and redundant annotations.

The complete cp genome of S. amabilis (GenBank accession number MT357029) was 160,006 bp in length and exhibited a typical quadripartite structure, which was composed of two copies of inverted repeat (IR, 26,405 bp) regions separated by a large single-copy (LSC, 87,870 bp) and a small single-copy (SSC, 19,326 bp) regions. The overall GC content was 36.55%, while the corresponding values of LSC, SSC, and IR regions were 33.59%, 30.24%, and 42.63%, respectively. A total of 128 genes were annotated in the cp genome, including 88 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Seven protein-coding genes, eight tRNA genes, and four rRNA genes were duplicated in the IR regions. Moreover, we also found that 17 genes contain introns, 14 of which contained one intron, the other 3 of genes (clpP, rps12, and ycf3) contain two introns.

To investigate phylogenetic position of S. amabilis, the sequence alignment was performed on the 25 cp genome sequences using MAFFT 7.307 (Kazutaka and Standley 2013), including 23 Maloideae species and two Prunnoideae species as outgroup. The maximum likelihood (ML) tree was constructed by online RAxML BlackBox software based on rapid bootstrap algorithm (Stamatakis et al. 2008). Our results indicated that S. amabilis is most related to S. commixta with very high bootstrap support values, and formed a clade sister with Pyrus, and the other Sorbus species are close to Malus species (Figure 1). The current phylogenetic relationship was
consistent with the results of Niu et al. (2019) based on cp genome sequences.

**Disclosure statement**

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

**Funding**

This work was supported by the Natural Science Foundation from Educational Commission of Anhui Province [KJ2017A373 and KJ2020A0037], and the Program for outstanding youth talents support plan in Universities of Anhui Province [gyyq2018163]. This work was also supported by the conservation project of wild animals and plants with extremely small populations in Mingxi County of Fujian province.

**Data availability statement**

The complete chloroplast genome sequence and annotation of *Sorbus amabilis* that support the findings of this study are openly available in Zenodo at [https://doi.org/10.5281/zenodo.3777235](https://doi.org/10.5281/zenodo.3777235).

**References**

Chen SF, Zhou YP, Chen YR, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34(17):i884–i890.

Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

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

Liu C, Shi LC, Zhu YJ, Chen HM, Zhang JH, Lin XH, Guan XJ. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics. 13:715.

Niu ZY, Xiong ZR, Xi LL, Chen X. 2019. The complete chloroplast genome of a Chinese endemic ornamental plant *Sorbus unguiculata* Koehne. Mitochondr DNA B. 4(2):2227–2228.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxMLweb-servers. Syst Boil. 57(5):758–771.

Wang Q, Niu ZY, Li J, Zhu KL, Chen X. 2020. The complete chloroplast genome sequence of the Chinese endemic species *Sorbus setschwanensis* (Rosaceae) and its phylogenetic analysis. Nord J Bot. 38:1–11.

Zhang XW, Li Y, Xie YP, Bao XM, Fang YM. 2018. Effect of climate change on potential geographical distribution of *Sorbus amabilis*. J Plant Resour Environ. 27:31–41.