Complete chloroplast genome of *Exochorda serratifolia* (Rosaceae)

Namsu Je a, Serim Kim a, b, *, Enkhtsetseg Yeruult a, Jin Hee Park c, Jungho Lee d and Yi Lee a

a Department of Industrial Plant Science & Technology, Chungbuk National University, Cheongju, Republic of Korea; b Institute for Plant Resources, Memoreal Co., Ltd., Cheongju, Republic of Korea; c Nakdonggang National Institute of Biological Resources, Sangju, South Korea; d Green Plant Institute, Yongin, South Korea

**ABSTRACT**

*Exochorda serratifolia* (pearlbush) is a rosid shrub found in northeast Asia, including the Korean peninsula. This ornamental plant has white inflorescences and strong insect resistance; however, its genetic diversity is poorly understood and a complete plastid genome is unavailable. Here, we determined the complete chloroplast genome of *E. serratifolia* through de novo assembly using next-generation sequencing. The *E. serratifolia* chloroplast genome was 160,558 bp, comprising a large single-copy (LSC) region of 88,514 bp, a small single-copy (SSC) region of 19,308 bp, and two inverted repeat regions (IRs) of 26,368 bp each. We annotated 112 genes: 78 protein-coding genes, 30 tRNA genes, and four rRNA genes. This reference plastid genome increases our understanding of the phylogenetic relationships of *E. serratifolia* among Rosaceae plants.

**Keywords**

Chloroplast genome; next-generation sequencing; *Exochorda serratifolia*

---

*Exochorda serratifolia* S. Moore 1877 is a deciduous shrub distributed in northeast Asia, including Liaoning (Qian mountains) in China (Cuizhi and Alexander 2003). It occurs sporadically on the Korean peninsula (Kim et al. 2014). This ornamental shrub has white inflorescences and possesses shade tolerance, freezing tolerance, and insect resistance (Lee et al. 1987). The species is an important genetic resource in Korea, being designated as a biological resource subject to approval for export abroad by the Korean Government. Despite the important economic value of *E. serratifolia* as a genetic resource, studies on its genetic diversity and cultivar identification are very limited. Phenotypic variation among habitats (Song et al. 2016) and ISSR variation among populations (Hong et al. 2013) have been investigated, but not fully resolved. Partial plastid genomic information for this species is available (Zhang et al. 2017); however, characterizing the complete chloroplast genome as a reference is necessary.

Leaf samples were collected from the native habitat in Cheongju, Korea (36°39’ 03.9” N, 127° 31’ 16.4” E), which is designated and protected as a Forest Genetic Resource Reserve (FGRR). A dried voucher specimen was deposited in the Herbarium of the National Institute of Horticulture and Herbal Science, Eumsung, South Korea (IN; http://www.nihhs.go.kr/; voucher number: MPS006522-1; contact: Yoongee Lee, yoong0625@korea.kr). Genomic DNA was extracted using a DNeasy Plant Mini kit (Qiagen, Valencia, CA). A genomic library was constructed using the Illumina HiSeq platform. Raw data obtained through next-generation sequencing were trimmed using the CLC quality trim program (ver. 4.21) in the CLC Assembly Cell package (version 4.2.1, https://www.qiagenbioinformatics.com/products/clc-assembly-cell/), and the chloroplast genome contig was secured and its structure determined through de novo assembly. Loci were annotated using the GeSeq program (https://chlorobox.mpimp-golm.mpg.de/geseq-.app.html) (Tillich et al. 2017), with manual cur- ation based on BLAST search for final review. The complete chloroplast genome sequence of *E. serratifolia* was deposited in GenBank under accession no. MZ981786.

The complete *E. serratifolia* chloroplast genome was 160,558 bp in length with GC content of 36.5%. The circular structure consisted of a large single-copy (LSC) region of 88,514 bp, a small single-copy (SSC) region of 19,308 bp, and two inverted repeat regions (IRs) of 26,368 bp each. We annotated 112 genes, representing 78 protein-coding genes, 30 tRNA genes, and four rRNA genes.

We performed phylogenetic analysis of 16 species belonging to the Rosaceae, including *E. serratifolia*, using the time-reversal model of the maximum-likelihood algorithm with 1000 bootstrap replications in the QIAGEN CLC Genomics Workbench 21.0 (Hilden, Germany) (https://digitalinsights.qiagen.com/). The phylogenetic tree placed *E. serratifolia* together with the Exochordeae *Prinsepia uniflora* (MZ270554)
and close to *Sorbaria sorbifolia* (MN026875) and *Sorbaria arborea* (MN901450) of the Sorbarieae (Figure 1).

**Author contributions**

S.K., J.L., and Y.L. conceived and designed the study. S.K., E.Y., J.H.P., and J.L. analyzed the data. N.J. and S.K. performed the experiments. N.J., J.L., and Y.L. wrote the manuscript. All authors read and approved the final manuscript.

**Ethical approval**

*Exochorda serratifolia* samples were collected by Serim Kim, a researcher of Memoreal Co., Ltd., with permission from Jinsoon Oh, the Hwajangsa Temple Governor who is responsible for the Forest Genetic Resource Reserve. The plant is freely accessible to Serim Kim with noncommercial resource purpose. The authors comply with relevant institutional, national, and international guidelines and legislation for plant study.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by a grant from the ‘Leaders in Industry-university Cooperation +’ Project, supported by the Ministry of Education and National Research Foundation of Korea and R&D Program for Forest Science Technology [Project No. 2020203810-2122-BA01] provided by the Korea Forestry Service [Korea Forestry Promotion Institute].

**ORCID**

Jin Hee Park [http://orcid.org/0000-0002-5976-1467](http://orcid.org/0000-0002-5976-1467)
Jungho Lee [http://orcid.org/0000-0002-4392-9480](http://orcid.org/0000-0002-4392-9480)
Yi Lee [http://orcid.org/0000-0001-7547-6087](http://orcid.org/0000-0001-7547-6087)

**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](https://www.ncbi.nlm.nih.gov) under the accession no. MZ981786. The associated ‘BioProject’, ‘SRA’, and ‘Bio-Sample’ numbers are PRJNA759290, SRR15681606, and SAMN21162962, respectively.

**References**

Cuizhi G, Alexander C. 2003. *Exochorda* Lindl. 8. EXOCHORDA Lindley, Gard. Chron. 1858: 925. 1858. Flora of China. 9:82–83.
Hong KN, Lee JW, Kang JT. 2013. Genetic diversity and population genetic structure of *Exochorda serratifolia* in South Korea. J Korean Soc Forest Sci. 102(1):122–128.
Kim KA, Cheon KS, Yoo KO. 2014. Environmental characteristics of *Exochorda serratifolia* S. Moore habitats. Korean J Plant Resour. 27(2):155–173.
Lee KE, Han KP, Park WG, Kim IS. 1987. A study on the wild *Exochorda serratifolia* for landscape horticultural cultivation. J Korean Inst Landsc Architect. 5:139–148.
Song JH, Kong MJ, Oak MK, Hong SP. 2016. A study on the distribution, external morphological characteristics and soil condition of *Exochorda serratifolia* S. Moore. Korean J Environ Ecol. 30(6):929–938.
Tillich M, Lehnhard 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.
Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, Yang JB, Li DZ, Yi TS. 2017. Diversification of Rosaceae since the late Cretaceous based on plastid phylogenomics. New Phytol. 214(3):1355–1367.

**Figure 1.** Maximum-likelihood phylogenetic tree based on entire chloroplast genome sequences of *E. serratifolia* and 15 plants belonging to the Rosaceae family. Numbers at nodes represent bootstrap values for 1000 replicates.