Characterization of the complete chloroplast genome of *Scrophularia cephalantha* endemic to Korea

Hyun-Do Jang\(^a\), Gi-Heum Nam\(^a\), Min-Su Park\(^b\) and Jumin Jun\(^c\)

\(^a\)Plant Resources Division, National Institute of Biological Resources, Incheon, South Korea; \(^b\)Department of Biology Education, Kongju National University, Gongju, South Korea; \(^c\)Biological Resources and Utilization Division, National Institute of Biological Resources, Incheon, South Korea

**ABSTRACT**

*Scrophularia* species are highly valued and widely used traditional medicinal plants in East Asia. In this article, the complete chloroplast genome of *Scrophularia cephalantha*, a species endemic to South Korea, is reported for the first time. The genome is 153,016 bp long, and it is composed of a pair of 25,485 bp inverted repeats (IRs), separated by a large single copy (LSC) region of 84,124 bp, and a small single copy (SSC) region of 17,922 bp. There are 133 predicted genes in the genome, comprising 88 protein-coding genes, 37 tRNAs, and 8 rRNAs, with an overall GC content of 38%. Phylogenetic analysis based on the chloroplast genome data showed that *S. cephalantha* is a sister species to *S. buergeriana* and *S. ningpoensis*. The data provide useful molecular information for phylogenetic and evolutionary studies of the genus *Scrophularia* and its related species.

The genus *Scrophularia* L. is one of the largest genera in the Scrophulariaceae and it consists of approximately 200 species mostly distributed in temperate regions of the Northern Hemisphere, with a few species distributed in the Southern Hemisphere (Stiefelhagen 1910; Hong 1983; Hong et al. 1998). *Scrophularia* species are popular traditional medicinal plants in East Asia used to treat inflammation, high blood pressure, and fever (Hwang et al. 1998; National Herbal Medicine Information 2021). Six species are recognized in South Korea, three of which, *Scrophularia cephalantha*, *Scrophularia koraiensis*, and *Scrophularia takesimensis*, are endemic to the country (Jang and Oh 2013).

*S. cephalantha* Nakai is remarkably well distinguished from its related species, *S. kakudensis* Franch. and *S. koraiensis* Nakai, by broadly ovate-shaped leaves, earlier flowering season, and a lower number of flowers on shorter inflorescences (Jang and Oh 2013). This endemic plant is distributed only in the Nakdong Mountains from Gangwon-do to Gyeongsangnam-do in South Korea (Jang 2016). Therefore, it is necessary to establish a strategy for the conservation of this species. However, the biological and genetic information of the species is lacking. In this study, we sequenced the chloroplast genome of *S. cephalantha* for the first time.

The specimen of *S. cephalantha* was collected from Mireuksan Mt., Tongyeong-si, Gyeongsangnam-do, Korea (34° 48' 49"N, 128° 24' 58"E, 249 m a.s.l.), and deposited at the National Science Museum (Seongjin Ji, jsj76@korea.kr) in Daejeon, South Korea under the voucher number H. D. Jang MN255822. Genomic DNA was extracted using the Exgene Plant kit (GeneAll, Seoul, Republic of Korea) following the manufacturer's protocol. Genomic DNA libraries were constructed using the QIAseq Ultralow Input Library kit (QIAGEN, Hilden, Germany). DNA shearing was performed using Q-Sonica 800 (QSonica, Newtown, CT) following the manufacturer’s protocol for obtaining ~500 bp fragments. DNA was fragmented using the following enzymatic reaction: blunt-ends were repaired, 3’-adenylated, and ligated with multiplex compatible adapters to construct an Illumina-compatible DNA library. Size selection of the 350–650 bp DNA fragments was carried out using Agencourt AMPure XP SPRI beads. PCR enrichment selectively amplified the DNA containing adapters at both ends. Library validation was performed using a Labchip GX (PerkinElmer, Waltham, Massachusetts) and then quantified using the PicoGreen dsDNA HS Assay Kit and the KAPA qPCR kit for the library. High-throughput sequencing was performed using the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA) at the Genome Analysis Center of National Instrumentation Center for Environmental Management (NICEM, Seoul, Republic of Korea). The run mode was a Rapid Pair End with 250 cycles. We obtained genomic DNA NGS sequences of approximately 10 Gb. De novo assembly was conducted using CLC Genomic Workbench version 10.0.0.1 (CLCBio, Aarhus, Denmark https://www.qiagenbioinformatics.com/), and chloroplast genome annotations were performed using Glimmer (http://ccb.jhu.edu/software/glimmer/index.shtml). The chloroplast genome sequence was deposited in GenBank under accession number MN255822.
The complete chloroplast genome of *S. cephalantha* had a typical quadripartite structure with a size of 153,016 bp. This was composed of a small single copy (SSC) region of 17,922 bp, a large single copy (LSC) region of 84,124 bp, and a pair of inverted repeat (IR) regions of 25,485 bp. The GC content of the genome was 38%. The IR regions had a higher GC content (43.2%) compared with that of the LSC (36.1%), and SSC (32.2%) regions. There were a total of 133 genes, and they included 88 protein-coding genes, 37 tRNAs, and 8 rRNAs in the genome. Twenty-one genes were duplicated in the IR regions, of which eight were protein-coding genes (*ndhB, rpl2, rpl23, rps12, rps7, ycf1, ycf2, and ycf15*), five were tRNA genes (*trnA-UGC, trnL-CAA, trnN-GUU, trnR-ACG, and trnV-GAC*), and four were rRNA genes (*rrn16S, rrn23S, rrn4.5S, and rrn5S*).

To identify the phylogenetic position of *S. cephalantha*, we generated a maximum likelihood tree of seven Scrophulariaceae species based on their complete chloroplast genomes (Figure 1). The sequences were aligned using Geneious Prime version 2019.2.3 (https://www.geneious.com, Auckland, New Zealand). Phylogenetic analyses were performed using the CIPRES Science Gateway webserver (RAxML-HPC on XSEDE version 8.2.10) with 1000 bootstrap replicates (Stamatakis 2014). The phylogenetic tree supported the tribe classification of *Scrophulariaceae* reported in previous studies (Olmstead et al. 2001; Tank et al. 2006).

To identify the phylogenetic position of *S. cephalantha*, we generated a maximum likelihood tree of seven *Scrophulariaceae* species based on their complete chloroplast genomes (Figure 1). The sequences were aligned using Geneious Prime version 2019.2.3 (https://www.geneious.com, Auckland, New Zealand). Phylogenetic analyses were performed using the CIPRES Science Gateway webserver (RAxML-HPC on XSEDE version 8.2.10) with 1000 bootstrap replicates (Stamatakis 2014). The phylogenetic tree supported the tribe classification of *Scrophulariaceae* reported in previous studies (Olmstead et al. 2001; Tank et al. 2006). The phylogenetic analysis placed *S. cephalantha* closely related to *S. buergeriana* and *S. ningpoensis* from the series *Kakudenses* (Yamazaki 1949; Jang 2016), with 100% bootstrap support.

**Disclosure statement**

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

**Funding**

This work was supported by a grant from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea [NIBR202107101].

**ORCID**

Min-Su Park https://orcid.org/0000-0002-9070-6669

**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/ under the accession no. MN255822. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA718412, SRX10472095, and SAMN18534161, respectively.

**References**

Hong DY, Yang HB, Jin CL, Holmgren NH. 1998. *Scrophulariaceae*. In: Wu ZY, Raven PH, editors. Flora of China, Vol. 18. Beijing and St. Louis, China: Science Press and Missouri Botanical Garden Press; p. 1–212.

Hong DY. 1983. The distribution of *Scrophulariaceae* in the Holarctic with special reference to the floristic relationships between eastern Asia and eastern north America. Ann Missouri Bot Garden. 70(4):701–712.

Hwang SY, Byun JH, Song HJ, Shin MK. 1998. A herbalogical study on the plants of *Scrophulariaceae* in Korea. Korean J Herbol. 13:241–270.

Jang HD, Oh BU. 2013. A taxonomic study of Korean *Scrophularia* L. (*Scrophulariaceae*) based on morphological characters. Korean J Plant Res. 26(2):271–283.

Jang HD. 2016. A systematic study of the genus *Scrophularia* (*Scrophulariaceae*) in Northeast Asia. [Ph.D. thesis]. Cheongju, South Korea: Chungbuk National University; p. 212.

National Herbal Medicine Information. 2021. The Korean Pharmacopoeia Twelfth Edition. Available from https://www.nifds.go.kr.

Olmstead RG, DePamphilis CW, Elisons WJ, Reeves PA. 2001. Disintegration of the Scrophulariaceae. Am J Bot. 88(2):348–361.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312–1313.

Stiefelhagen H. 1910. Systematische und Pflanzengeographische Studien zur Kenntnis der Gattung *Scrophularia*. Botanische Jahrbücher für Systematik, Pflanzengeographie, 44(4):406–496.

Tank DC, Beardsley PM, Kelchner SA, Olmstead RG. 2006. Review of the systematics of *Scrophulariaceae* s.l. and their current disposition. Aust Syst Bot. 19(4):289–307.

Yamazaki T. 1949. *Scrophularia* Asiae Orientalis (1). J Japn Bot. 23:79–88.