

The complete chloroplast genome sequence of *Campanula zangezura* (Campanulaceae)

Kyung-Ah Kim^{a,b}, Ki-Oug Yoo^a and Kyeong-Sik Cheon^c

^aDepartment of Biological Sciences, Kangwon National University, Chuncheon, South Korea; ^bEnvironmental Research Institute, Kangwon National University, Chuncheon, South Korea; ^cDepartment of Biological Science, Sangji University, Wonju, South Korea

ABSTRACT

The complete chloroplast genome sequence of *Campanula zangezura* was determined by Illumina pair-end sequencing. The complete cp genome was 166,331 bp in length, containing a large single-copy (LSC) region of 100,034 bp and a small single-copy (SSC) region of 7781 bp, which were separated by a pair of 29,258 bp inverted repeats (IRs). A total of 113 unique genes were annotated, including 79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. Among these genes, 17 genes contained one or two introns. The ML tree based on 74 protein-coding genes showed that *C. zangezura* formed a sister to the *Campanula punctata* and *Campanula takesimana* clade.

ARTICLE HISTORY

Received 4 December 2019
Accepted 8 December 2019

KEYWORDS

Campanulaceae; *Campanula*; *Symphyandra*; chloroplast genome; phylogeny

The sect. *Symphyandra* was considered to be an independent genus due to the character of the connate anther. At present, however, it is commonly considered to belong to the genus *Campanula* due to the formed polyphyly in the Campanula species clade (Eddie et al. 2003; Haberle et al. 2008; Raab-Straube and Raus 2014). *Campanula* L. is the largest genus in Campanulaceae, consisting of 350–500 species distributed mainly in the Northern Hemisphere (Federov 1957). Among the *Campanula* species, *Campanula zangezura*, discussed in this study, is very valuable as a horticultural plant owing to its attractive flowers. Unfortunately, genome studies of this species have yet to be carried out. Moreover, Campanulaceae are known to have a highly rearranged chloroplast genome (Cosner et al. 1997, 2004; Haberle et al. 2009; Cheon et al. 2017). Therefore, the structure and sequences of the plastid genome are useful data for clarifying their phylogeny and evolutionary tendencies. However, the cp genome data remain insufficient to clarify the phylogenetic relationships of Campanulaceae.

In this study, we report the complete chloroplast genome sequence of *C. zangezura* to provide genetic information that can be used in various studies in the future. The plant materials for this study were used after seeding and cultivation in the greenhouse of Kangwon National University. A voucher specimen was deposited into the Kangwon National University Herbarium (KWNU), with voucher number KWNU89889. Total DNA was extracted using a DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA). Genomic DNA was used for sequencing with the Illumina Miseq (Illumina Inc., San Diego, CA, USA) platform. *C. zangezura* was sequenced to produce 6,736,470 raw reads and 301 bp were obtained

from them. The plastid genome assembly was performed by the *de novo* assembly protocol (Cho et al. 2015) via the Phyzen bioinformatics pipeline (<http://phyzen.com>). Annotation of the chloroplast genome was based on an online available program, DOGMA (Wyman et al. 2004), coupled with manual corrections for start and stop codons. We also compared each gene to the published complete chloroplast genome sequence of Campanulaceae for correct gene annotation. The tRNAs were confirmed using tRNAscan-SE (Schattner et al. 2005). A circular plastid genome map was drawn using the OGDRAW program (Greiner et al. 2019).

The complete chloroplast genome of *C. zangezura* (GenBank accession no. MN756013) is a circular DNA molecule 166,330 bp in length with 38.9% G + C content, composed of a large single-copy (LSC) region of 100,034 bp, a small single-copy (SSC) region of 7781 bp, and two inverted repeat (IR) regions of 29,258 bp. The whole cp genome of *C. zangezura* was 3011 bp and 3221 bp shorter than those of reported two *Campanula* species, *C. punctata* and *C. takesimana*, respectively. The plastid genome contains a total of 113 unique genes constituting 79 protein-coding genes, 30 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes. Among the 113 genes, four genes (*rpl23*, *ycf15*, *infA* and *clpP*) were identified as pseudogenes, and 17 genes (*trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *atpF*, *clpP*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, and *ycf3*) contain one or two introns.

To investigate the phylogenetic relationships between *C. zangezura* and related taxa, we conducted a phylogenetic analysis. The cpDNA sequence except for the IGS (intergenic spacer region) due to many rearrangements was used to

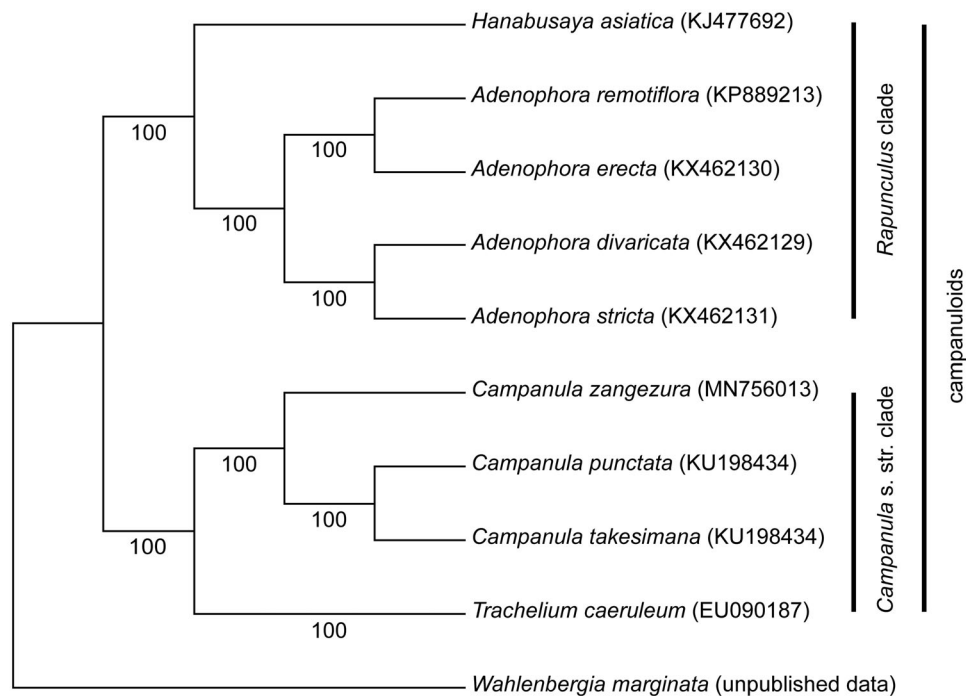


Figure 1. The ML tree based on 76 protein-coding genes from eight Campanuloid species and an outgroup.

determine the relationships among the Campanuloid species. A total of 76 protein-coding gene sequences from nine Campanuloid species and one outgroup (*Wahlenbergia marginata*) were aligned using MAFFT (Kato et al. 2002). We then conducted maximum likelihood (ML) analysis using RAxML v.7.4.2 with 1000 bootstrap replicates and the GTR + I model (Stamatakis 2006). The ML tree (Figure 1) showed that Campanuloid species formed two clades, the *Campanula* s. str. clade and the *Rapunculus* clade. In addition, *C. zangezura* formed a sister to the *C. punctata* and *C. takesimana* clade in the *Campanula* s. str. clade with a high support value (BP = 100).

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1A6A3A11029236).

References

- Cheon KS, Kim KA, Yoo KO. 2017. The complete chloroplast genome sequences of three *Adenophora* species and comparative analysis with Campanuloid species. *PLOS One*. 12(8):e0183652.
- Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ. 2015. Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). *PLOS One*. 10:e012533.
- Cosner ME, Jansen RK, Palmer JD, Downie SR. 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet*. 31(5):419–429.
- Cosner ME, Raubeson LA, Jansen RK. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol*. 4(1):27.
- Eddie WMM, Shulkina T, Gaskin J, Haberle RC, Jansen RK. 2003. Phylogeny of Campanulaceae s. str. inferred from ITS sequences of nuclear ribosomal DNA. *Ann Missouri Bot Gard*. 90(4):554–575.
- Fedorov AA. 1957. Flora of the U.S.S.R. Vol. 24. Moskva: Akademii Nauk SSSR; p. 126–450.
- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acid Res*. 47(W1):W59–W64.
- Haberle RC, Dang A, Lee T, Peñafior C, Cortes-Burns H, Oestreich A, Raubeson L, Cellinese N, Edwards EJ, Kim S-T, et al. 2009. Taxonomic and biogeographic implications of a phylogenetic analysis of the Campanulaceae based on three chloroplast genes. *Taxon*. 58(3): 715–734.
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol*. 66(4):350–361.
- Kato K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.
- Raab-Straube E, Raus T. 2014. Euro + Med-Checklist Notulae, 3. *Willdenowia*. 44(2):287–299.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res*. 33(Web Server):W686–W689.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21):2688–2690.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 20(17):3252–3255.