



# The complete chloroplast genome sequence of *Rhododendron caucasicum* (Ericaceae)

Myounghai KWAK<sup>1\*</sup> and Rainer W. BUSSMANN<sup>2,3</sup>

<sup>1</sup>National Institute of Biological Resources, Incheon 22689, Korea

<sup>2</sup>Institute of Botany and Bakuriani Alpine Botanical Garden, Iliia State University, Tbilisi 0105, Georgia

<sup>3</sup>Department of Botany, State Museum of Natural History, 76135 Karlsruhe, Germany

## Corresponding author

Myounghai KWAK  
E-mail: mhkwak1@korea.kr

## OPEN ACCESS

© 2023 The Author(s)  
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** *Rhododendron caucasicum* Pall. is a shrub distributed in the mountainous areas of the Caucasus from northeastern Türkiye towards the Caspian Sea. This study reports the first complete chloroplast genome sequence of *R. caucasicum*. The plastome is 199,487 base pairs (bp) long and exhibits a typical quadripartite structure comprising a large single-copy region of 107,645 bp, a small single-copy region of 2,598 bp, and a pair of identical inverted repeat regions of 44,622 bp each. It contains 143 genes, comprising 93 protein-coding genes, 42 tRNA genes, and eight rRNA genes. The large chloroplast genome size is likely due to the expansion of inverted repeats. A phylogenetic analysis of chloroplast genomes with other *Rhododendron* species supports previously recognized infrageneric relationship.

**Keywords:** complete chloroplast genome, inverted repeats, phylogenetic analysis, *Rhododendron*

**RECEIVED** 18 August 2023; **REVISED** 21 September 2023; **ACCEPTED** 21 September 2023

## INTRODUCTION

*Rhododendron* is the largest woody plant genus in the Northern Hemisphere, comprising over 1,000 species (Frodin, 2004). A recent study indicated that the genus *Rhododendron* first originated in northeast Asia in the Paleocene and then dispersed to North America in the late Eocene and Oligocene (Shrestha et al., 2018). However, the contemporary species diversity of *Rhododendron* is mainly due to extensive speciation in the tropical and subtropical regions of southern China, south Asia, and the Malay Archipelago during the 30–10 MYA period (Milne et al., 2010; Shrestha et al., 2018). A recent molecular study divided the genus *Rhododendron* into five subgenera and 11 sections (Xia et al., 2022).

The chloroplast genome has been extensively used to clarify phylogenetic relationships from the species level to deeper levels (Gitzendanner et al., 2018; Li et al., 2019; Fan et al., 2021). Chloroplast genomes are the one of best molecular markers in plant phylogenetic studies due to their abundance and lack of recombination with appropriate mutation rates. Moreover, despite some exceptions, the maternal inheritance of chloroplasts contributes to its role as a key player in identifying ancient hybrid phenomena with

comparison of the phylogenetic relationships of nuclear genes (Kawabe et al., 2018; Liu et al., 2022). Due to the high single-cell copy number and small genome size (120–160 kb) of plant chloroplasts, fast and cost-effective genome skimming is sufficient to obtain fully annotated whole genome sequences of the chloroplast.

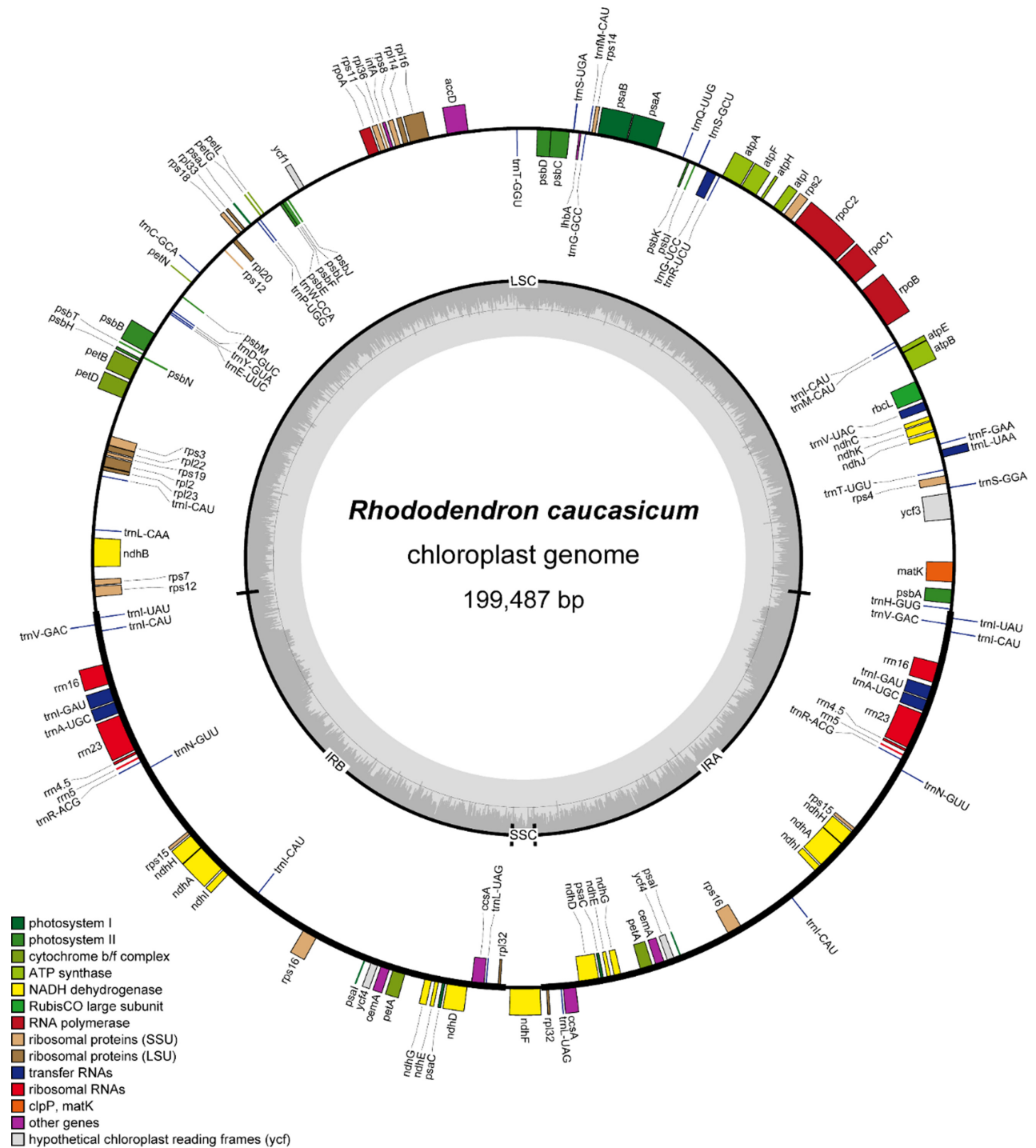
*Rhododendron caucasicum* Pall. is a shrub distributed in the mountainous areas of the Caucasus from northeastern Türkiye towards the Caspian Sea. This species is phylogenetically closely related to *R. aureum* Georgi and *R. brachycarpum* D. Don ex G. Don, found in Northeast Asia (Milne, 2004). The disjunct distribution of *R. caucasicum* from *R. aureum* and *R. brachycarpum* and their phylogenetic closeness show that *R. caucasicum* is a rare case of a tertiary relict species in southwest Eurasia. Here, we report the complete chloroplast genome sequence of *R. caucasicum*. The chloroplast genome of *R. caucasicum* will aid further investigation into the biogeography of this species group.

## MATERIALS AND METHODS

*Rhododendron caucasicum* was sampled at approximately 2,500 m in the timberline area of Tsratskharo Pass, close to

Bakuriani, Samtskhe-Yavakheti, Georgia, by R. W. Bussmann in August 2022. The voucher specimen (RBU-19784) was deposited at the Herbarium of the National Institute of Biological Resources (KB) and the National Herbarium of Georgia (TBI). Genomic DNA was extracted from the dried leaves taken from

the specimens using the cetyltrimethylammonium bromide method (Doyle and Doyle, 1987) and verified by 1% agarose gel electrophoresis. The DNA library was constructed using a TruSeq DNA Nano Kit for a 350-bp insert size according to the manufacturer’s instructions (Illumina Inc., San Diego,



**Fig. 1.** Circular map of the *Rhododendron caucasicum* complete chloroplast genome. The genes outside the circle are transcribed clockwise while those inside are transcribed counterclockwise. The dark gray plot in the inner circle corresponds to the GC content. Large single-copy, small single-copy, and inverted repeat are indicated by LSC, SSC, and IR (IRA and IRB), respectively.

CA, USA). Whole-genome sequencing was performed using the Illumina NovaSeq6000 platform (DNA Link Inc., Seoul, Korea). We retrieved 7.3 Gb of raw reads (150 bp paired-end reads), which were quality-trimmed using the Trimmomatic tool (Bolger et al., 2014). *De novo* assembly was performed with Velvet v1.2.19 (Zerino and Birney, 2008), and the obtained contigs were used to construct a draft genome with the *R. delavayi* Franch. chloroplast genome (GenBank accession no. MN711645) as a reference. The genome sequence was confirmed by aligning the raw reads against the assembled genome using BWA v0.7.17 and SAMtools v1.9 (Li, 2013). The gaps were closed using GapCloser v1.12 (Zhao et al., 2011). Annotation of the chloroplast genome was conducted using Geneious Prime v2020.2.4 (Biomatters Ltd., Auckland, New Zealand) based on the previously reported Ericaceae chloroplast genomes in the National Center for Biotechnology Information (NCBI) database. tRNA prediction was performed using the

tRNAscan-SE2.0 (Chan and Lowe, 2019), and a circular map was drawn using OGDRAW v1.31 (Greiner et al., 2019).

The complete chloroplast genome sequences of 15 *Rhododendron* species were downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) to investigate the phylogenetic relationship of *R. caucasicum* with other *Rhododendrons*. Among the previously reported complete chloroplast genomes from Ericaceae species, *Gaultheria longibracteolata* R.C. Fang and *Vaccinium myrtillus* L. were used as the outgroups. Phylogenetic analysis was performed using 74 coding sequences of *Rhododendron* species. Alignments were performed using Clustal Omega v1.2.2 as implemented in Geneious Prime software, and the alignments were concatenated. Subsequent phylogenetic analyses hereafter were performed in PhyloSuite v1.2.3 (Zhang et al., 2020; Xiang et al., 2023). The optimal partitioning strategies and evolutionary models for the coding sequences under the Bayesian information criterion were determined using

**Table 1.** List of genes annotated in the chloroplast genome of *Rhododendron caucasicum*.

Gene categories	Gene groups	Gene names
Self-replication	Large subunit ribosomal proteins	<i>rpl2, rpl14, rpl16*</i> , <i>rpl20, rpl22, rpl23, rpl32</i> (×2), <i>rpl33, rpl36</i>
	DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Small subunit ribosomal proteins	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12**†</i> , <i>rps14, rps15</i> (×2), <i>rps16*</i> (×2), <i>rps18, rps19</i>
	Ribosomal RNAs	<i>rrn4.5S</i> (×2), <i>rrn5S</i> (×2), <i>rrn16S</i> (×2), <i>rrn23S</i> (×2)
	Transfer RNAs	<i>trnA-UGC</i> (×2)*, <i>trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-M-CAU, trnG-GCC, trnG-UCC*</i> , <i>trnH-GUG, trnI-CAU</i> (×5), <i>trnI-GAU</i> (×2)*, <i>trnI-UAU*</i> (×2), <i>trnL-CAA, trnL-UAA*</i> , <i>trnL-UAG</i> (×2), <i>trnM-CAU</i> (×2), <i>trnN-GUU</i> (×2), <i>trnP-UGG, trnQ-UUG, trnR-ACG</i> (×2), <i>trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC</i> (×2), <i>trnV-UAC*</i> , <i>trnW-CCA, trnY-GUA</i>
Photosynthesis	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF*</i> , <i>atpH, atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhA*</i> (×2), <i>ndhB*</i> , <i>ndhC, ndhD</i> (×2), <i>ndhE</i> (×2), <i>ndhF, ndhG</i> (×2), <i>ndhH</i> (×2), <i>ndhI</i> (×2), <i>ndhJ</i> (×2), <i>ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA</i> (×2), <i>petB*</i> , <i>petD*</i> , <i>petG, petL, petN</i>
	Subunits of photosystem I	<i>psaA, psaB, psaC</i> (×2), <i>psaI</i> (×2), <i>psaJ</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ</i>
	Subunit of rubisco	<i>rbcL</i>
Other genes	Subunit of acetyl-CoA-carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>cesA</i> (×2)
	Envelop membrane protein	<i>cema</i> (×2)
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
Unknown function	Conserved open reading frames	<i>yef1, yef3, yef4</i> (×2)

Asterisks indicate genes containing one intron and double asterisks indicate genes containing two introns.

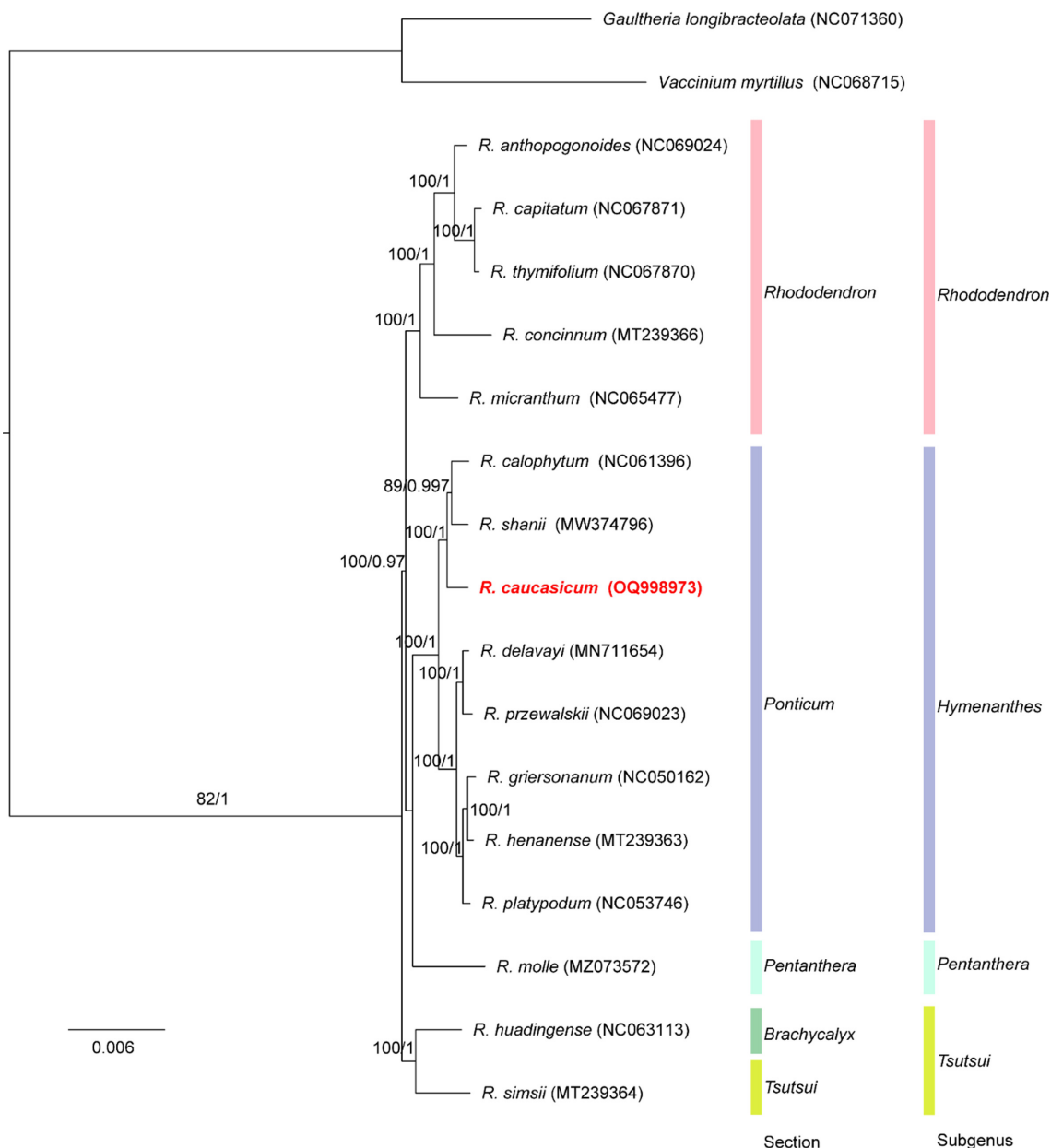
†, trans-spliced genes; (×2), genes have two copies; (×3), genes have three copies.

ModelFinder (Kalyaanamoorthy et al., 2017). The best-fit partition models are shown in Table 2. A maximum likelihood (ML) was reconstructed using IQ-tree (Nguyen et al., 2015) with 10,000 ultrafast bootstrap replicates (Minh et al., 2013). A Bayesian inference tree was built using MrBayes v3.2.7a (Ronquist et al., 2012). Markov Chain Monte Carlo runs were performed for 10 million generations, and trees were sampled every 1,000 generations. The first 25% of the trees were discarded as burn-in to ensure the chains were stationary. The

remaining trees were used to generate a strict consensus tree and calculate each node's posterior probabilities.

## RESULTS AND DISCUSSION

The chloroplast genome of *R. caucasicum* (GenBank accession no. OQ998973) consists of 199,487 bp and has four subregions: a large single-copy region (LSC) of 107,645 bp and a small single-copy region (SSC) of 2,598 bp that



**Fig. 2.** Phylogenetic tree of *Rhododendron caucasicum* and related taxa based on 74 protein-coding gene sequences of the chloroplast sequence. The phylogenetic tree was drawn based on the maximum likelihood phylogenetic tree. The number above the branches corresponds to the bootstrap support values from the maximum likelihood and posterior probability values for the Bayesian inference analyses. *Gaultheria longibracteolata* and *Vaccinium myrtillus* were used as outgroups. The numbers in parenthesis are National Center for Biotechnology Information (NCBI) GenBank accession numbers.

**Table 2.** Optimal partitioning strategies and evolutionary models selected by ModelFinder using the Bayesian information criterion.

ML	BI	Genes
TVM + F + R3	GTR + F + I + G4	<i>rpoC1, rps2, rpl2, ycf3, rps7, rps8, rps11, rps14, rps15, rps16, rps19, rpl22, rpl23, rpl33, rpl36, atpF, psaJ, rbcL, rpoB</i>
TIM3 + F + G4	GTR + F + G4	<i>ycf1</i>
TPM2u + F + R2	HKY + F + G4	<i>rps3, rps12, rps18</i>
TVM + F + G4	GTR + F + G4	<i>rps4, rpl14, rpl20, matK, rpoA</i>
TVM + F + I	GTR + F + I	<i>ycf4, lhbA, atpH, ndhC, ndhJ, petB, petL, petN, psaA, psaB, psbA, psbB, psbC, psbD, psbE, psbH, psbJ, psbK, psbN</i>
F81 + F + R2	F81 + I + G4	<i>rpl16</i>
TPM3u + F + I	HKY + F + I	<i>rpl32, atpA, atpB, atpE, atpI, cemA, ndhB, ndhI, ndhK, petA, psbF, psbM, ccsA, ndhA, ndhD, ndhE, ndhG, ndhH, petG, psaC, psal, psbI, psbL, psbT</i>
TPM3uT	GTR + F + G4	<i>ndhF</i>
TPM2u + F + R2	HKY + F + I + G4	<i>petD</i>

ML, maximum likelihood; BI, Bayesian information; TVM, transversion model, AG = CT and unequal base frequency; TIM3, transition model, AC = CG, AT = GT and unequal base frequency; TPM2u, AC = AT, AG = CT, CG = GT and unequal base frequency; TPM3u, AC = CG, AG = CT, AT = GT and unequal base frequency; F81, equal rates but unequal base frequency; GTR, general time reversible model with unequal rates and unequal base frequency; HKY, unequal transition/transversion rates and unequal base frequency; F, empirical base frequency; G4, discrete gamma model with four rate categories; I, allowing for a proportion of invariable sites; R2, freerate model parameters with two of categories; R3, freerate model parameters with tree of categories.

are separated by the inverted repeat regions (IR) of 44,622 bp (Fig. 1). The chloroplast genome's GC content is 35.9% and is 35.3, 30.0, and 36.7% in the LSC, SSC, and each of the IRs, respectively. The chloroplast contains 143 genes (93 protein-coding genes [PCGs], eight ribosomal RNAs [rRNAs], and 42 transfer RNAs [tRNA]); 24 genes (13 PCGs, four rRNAs, and nine tRNAs) are duplicated in the IR regions (Table 1). *clpP*, *ycf2*, and *ycf68* were not identified in the *R. caucasicum* cp genome, and we concluded those genes were missing since they were also missing in the previously reported *Rhododendron* cp genomes (Liu et al., 2020; Ma et al., 2021; Wang et al., 2021).

The *R. caucasicum* chloroplast genome size (199,487 bp) falls within the known size categories of *Rhododendron* genomes, ranging from 197,877 bp (*R. mole*; MZ073672) to 230,777 bp (*R. kawakamii*, NC058233), which is relatively large among the angiosperm chloroplast genomes (Daniell et al., 2016; Olejniczak et al., 2016). The *R. caucasicum* chloroplast genome has expanded IRs and contracted SSC like other previously reported *Rhododendron* cp genomes. *nhaA*, *ndhD*, *ndhE*, *ndhG*, *ndhH*, *ndhI*, *rps15*, *psaC*, *ccsA*, and *rpl32*, which are generally found in the SSC, were moved to the IR, while only *ndhF* was detected in the SSC region of *R. caucasicum*. Thus, the increased chloroplast genome size might be due to the expansion of the IRs.

The ML- and Bayesian inference-based phylogenies had the same topology with high support for each branch (Fig. 2). The sub-generic relationships shown in this study are

consistent with previous molecular phylogenetic studies (Shrestha et al., 2018; Xia et al., 2022). Except for the subgenus *Therorhodon*, which is not included in this study, two species in the subgenus *Tsutsui* diverged first from the rest. Then, the subgenus *Rhododendron* diverged from the subgenera *Hymenanthes* and *Pentanthera*.

Given that *R. caucasicum* is a tertiary relic species and the closest sister to *R. aureum* and *R. brachycarpum* (Milne, 2004), we expect that further extensive phylogeographic studies will clarify their speciation histories and provide clues to their disjunct distribution. Accordingly, the chloroplast sequence we describe of *R. caucasicum* will provide useful information for future studies to understand their phylogenetic and evolutionary relationships.

## ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Biological Resources, funded by the Ministry of Environment of the Republic of Korea (Grant No. NIBR202207101). This project was carried out in collaboration under the Memorandum of Understanding signed by National Institute of Biological Resources and Ilia State University. The authors are grateful to Prof. Ohseok Kwon at Kyungpook National University for his work on this cooperative project and to Dr. Jongsun Park and Dr. Woochan Kwon at Infoboss for their assistance on assembly and annotation.



## CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

## LITERATURE CITED

- Bolger, A. M., L. Lohse and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Chan, P. P. and T. M. Lowe. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *In Gene Prediction. Methods in Molecular Biology*, Vol. 1962. Kollmar, M. (ed.), Humana, New York. Pp. 1–14.
- Daniell, H., C.-S. Lin, M. Yu and W.-J. Chang. 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology* 17: 134.
- Doyle, J. J. and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Fan, Y., Y. Jin, M. Ding, Y. Tang, J. Cheng, K. Zhang and M. Zhou. 2021. The complete chloroplast genome sequences of eight *Fagopyrum* species: Insights into genome evolution and phylogenetic relationships. *Frontiers in Plant Science* 12: 799904.
- Frodin, D. G. 2004. History and concepts of big plant genera. *Taxon* 53: 753–776.
- Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* 105: 291–301.
- Greiner, S., P. Lehwark and R. Bock. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47: W59–W64.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.
- Kawabe, A., H. Nukii and H. Y. Furihata. 2018. Exploring the history of chloroplast capture in *Arabidopsis* using whole chloroplast genome sequencing. *International Journal of Molecular Sciences* 19: 602.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv: <https://doi.org/10.48550/arXiv.1303.3997>.
- Li, H.-T., T.-S. Yi, L.-M. Gao, P.-F. Ma, T. Zhang, J.-B. Yang, M. A. Gitzendanner, P. W. Fritsch, J. Cai, Y. Luo, H. Wang, M. van der Bank, S.-D. Zhang, Q.-F. Wang, J. Wang, Z.-R. Zhang, C.-N. Fu, J. Yang, P. M. Hollingsworth, M. W. Chase, D. E. Soltis, P. S. Soltis and D.-Z. Li. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* 5: 461–470.
- Liu, B.-B., C. Ren, M. Kwak, R. G. J. Hodel, C. Xu, J. He, W.-B. Zhou, C.-H. Huang, H. Ma, G.-Z. Qian, D.-Y. Hong and J. Wen. 2022. Phylogenomic conflict analyses in the apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern Hemisphere. *Journal of Integrative Plant Biology* 64: 1020–1043.
- Ma, L.-H., H.-X. Zhu, C.-Y. Wang, M.-Y. Li and H.-Y. Wang. 2021. The complete chloroplast genome of *Rhododendron platypodum* (Ericaceae): An endemic and endangered species from China. *Mitochondrial DNA Part B: Resources* 6: 196–197.
- Milne, R. I. 2004. Phylogeny and biogeography of *Rhododendron* subsection *Pontica*, a group with a tertiary relic distribution. *Molecular Phylogenetics and Evolution* 33: 389–401.
- Milne, R. I., C. Davies, R. Prickett, L. H. Inns and D. F. Chamberlain. 2010. Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: Between-lineage hybridization during adaptive radiation? *Plant Systematics and Evolution* 285: 233–244.
- Minh, B. Q., M. A. T. Nguyen, A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30: 1188–1195.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.
- Olejniczak, S. A., E. Łojewska, T. Kowalczyk and T. Sakowicz. 2016. Chloroplasts: State of research and practical applications of plastome sequencing. *Planta* 244: 517–527.
- Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.
- Shrestha, N., Z. Wang, X. Su, X. Xu, L. Lyu, Y. Liu, D. Dimitrov, J. D. Kennedy, Q. Wang, Z. Tang and X. Feng. 2018. Global patterns of *Rhododendron* diversity: The role of evolutionary time and diversification rates. *Global Ecology and Biogeography* 27: 913–924.
- Wang, Z.-F., L.-W. Chang and H.-L. Cao. 2021. The complete chloroplast genome of *Rhododendron kawakamii* (Ericaceae). *Mitochondrial DNA Part B: Resources* 6: 2538–2540.
- Xia, X.-M., M.-Q. Yang, C.-L. Li, S.-X. Huang, W.-T. Jin, T.-T. Shen, F. Wang, X.-H. Li, W. Yoichi, L.-H. Zhang, Y.-R. Zheng, X.-Q. Wang. 2022. Spatiotemporal evolution of the

- global species diversity of *Rhododendron*. *Molecular Biology and Evolution* 39(1): msab314.
- Xiang, C.-Y., F. Gao, I. Jakovlić, H.-P. Lei, Y. Hu, H. Zhang, H. Zou, G.-T. Wang and D. Zhang. 2023. Using PhyloSuite for molecular phylogeny and tree-based analyses. *iMeta* 2: e87.
- Zerbino, D. R. and E. Birney. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.
- Zhang, D., F. Gao, I. Jakovlić, H. Zou, J. Zhang, W. X. Li and G.T. Wang. 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20: 348–355.
- Zhao, Q.-Y., Y. Wang, Y.-M. Kong, D. Luo, X. Li and P. Hao. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: A comparative study. *BMC Bioinformatics* 12(Suppl 14): S2.