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

Myounghai KWAK* and Rainer W. BUSSMANN

1National Institute of Biological Resources, Incheon 22689, Korea
2Institute of Botany and Bakuriani Alpine Botanical Garden, Ilia State University, Tbilisi 0105, Georgia
3Department of Botany, State Museum of Natural History, 76135 Karlsruhe, Germany

**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*

**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.
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.

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

<table>
<thead>
<tr>
<th>Gene categories</th>
<th>Gene groups</th>
<th>Gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-replication</td>
<td>Large subunit ribosomal proteins</td>
<td>rpl2, rpl14, rpl16*, rpl20, rpl22, rpl23, rpl32 (×2), rpl33, rpl36</td>
</tr>
<tr>
<td></td>
<td>DNA-dependent RNA polymerase</td>
<td>rpoA, rpoB, rpoC1, rpoC2</td>
</tr>
<tr>
<td></td>
<td>Small subunit ribosomal proteins</td>
<td>rps2, rps3, rps4, rps7, rps8, rps11, rps12***, rps14, rps15 (×2), rps16* (×2), rps18, rps19</td>
</tr>
<tr>
<td>Ribosomal RNAs</td>
<td></td>
<td>rrrn4.5S (×2), rrrn5S (×2), rrrn16S (×2), rrrn23S (×2)</td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>Subunits of ATP synthase</td>
<td>atpA, atpB, atpE, atpF*, atpH, atpI</td>
</tr>
<tr>
<td></td>
<td>Subunits of NADH-dehydrogenase</td>
<td>ndhA (×2)<em>, ndhB</em>, ndhC, ndhD (×2), ndhE (×2), ndhF, ndhG (×2), ndhH (×2), ndhI (×2), ndhJ (×2)</td>
</tr>
<tr>
<td></td>
<td>Subunits of cytochrome b/f complex</td>
<td>petA (×2), petB*, petD*, petG, petL, petN</td>
</tr>
<tr>
<td></td>
<td>Subunits of photosystem I</td>
<td>psaA, psaB, psaC (×2), psaI (×2), psaJ</td>
</tr>
<tr>
<td></td>
<td>Subunit of rubisco</td>
<td>rbcl</td>
</tr>
<tr>
<td>Other genes</td>
<td>Subunit of acetyl-CoA-carboxylase</td>
<td>accD</td>
</tr>
<tr>
<td></td>
<td>C-type cytochrome synthesis gene</td>
<td>ccsA (×2)</td>
</tr>
<tr>
<td></td>
<td>Envelop membrane protein</td>
<td>cemA (×2)</td>
</tr>
<tr>
<td></td>
<td>Translational initiation factor</td>
<td>infA</td>
</tr>
<tr>
<td></td>
<td>Maturase</td>
<td>matK</td>
</tr>
<tr>
<td>Unknown function</td>
<td>Conserved open reading frames</td>
<td>ycf1, ycf3, ycf4 (×2)</td>
</tr>
</tbody>
</table>

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

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. ndhA, 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 Therorhodion, 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 phylobiogeographic 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. Woohan 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


