


The complete mitochondrial and plastid genomes of *Rhododendron simsii*, an important parent of widely cultivated azaleas

Jie Xu^{a*}, Hang Luo^{a*}, Shuai Nie^a, Ren-Gang Zhang^b and Jian-Feng Mao^a 

^aBeijing Advanced Innovation Center for Tree Breeding by Molecular Design, National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China; ^bBeijing Ori-Gene Science and Technology Co. Ltd., Beijing, China

ABSTRACT

The genus *Rhododendron* of the heather family (Ericaceae) is well known and widely cultivated for their highly ornamental value. The most widely cultivated *Rhododendron* species is *Rhododendron simsii* (Indoor azalea). In this study, we assembled the complete linear mitochondrial genome (GenBank accession number MW030508) and quadripartite plastid genome (GenBank accession number MW030509). The mitochondrial genome is 802,707 bp in length with containing 53 unique genes (33 protein-coding, 17 tRNA, and 3 rRNA genes), while the 152,214 bp long plastid genome is smaller and containing 105 unique genes (4 rRNA, 26 tRNA, and 75 protein-coding genes). Phylogenetic analysis showed that the same species relationship with APG system as well as the low supports of branches which is the common characteristic of resolved Ericales phylogenetics.

ARTICLE HISTORY

Received 30 September 2020
Accepted 8 March 2021

KEYWORDS

Rhododendron simsii; azalea; mitochondrial genome; plastid genome



The genus *Rhododendron* of the heather family (Ericaceae) is well known for the outstanding beauty and great diversity of corolla among its more than 1000 species and 30,000 cultivars (Stevenson 1930; Sleumer 1949; Galle 1985; Yan et al. 2015). *Rhododendron simsii* Planch. 1853. (Indoor azalea), is the most widely cultivated *Rhododendron* species (Ding 2009; Kobayashi et al. 2013; De Riek et al. 2018). As the primary ancestors of azalea cultivars, *R. simsii* was introduced into Europe in the 18th century from China and its breeding for ornamental use began in England and was well developed in Belgium (Galle 1985; De Keyser et al. 2010; De Riek et al. 2018). Then, *R. simsii* hybrids become one of the most important pot plants in Belgium, with an annual production of approximately 40 million pots (De Keyser et al. 2010). Now azalea cultivars have become one of the most popular pot plants and landscape shrubs in Europe, North American, and Asia (Galle 1985; De Riek et al. 2018). Here, we determine mitochondrial and plastid content and structure of *R. simsii* through whole-genome sequencing, as well as detected its identity and phylogenetic relationship among Ericales. This study could bring more desirable information for evolutionary and functional studies in the future.

Leaves were obtained from a 20-year-old shrub from Jingshan, Hubei Province, China. Total DNA was isolated and extracted from the leaves using the DNeasy Plant Mini Kit (QIAGEN, Inc.) and then purified using the Mobio PowerClean Pro DNA Clean-Up Kit (MO BIO Laboratories, Inc.). A specimen

was deposited at the Herbarium of Beijing Forestry University (<http://www.bjfu.edu.cn/>, Jian-Feng Mao jianfeng.mao@bjfu.edu.cn), under the voucher number mao_20190311.

For PacBio SMRT (single-molecule real-time) sequencing, libraries with 20-kb DNA inserts were prepared and sequenced on a PacBio RSII platform using P6-C4 chemistry (6 SMRT cells). A total of 6.5 million PacBio long reads were generated, yielding 51.15 Gb (roughly 100× coverage) with an average read length of 7705 bp. For Illumina sequencing, 150-bp paired-end (PE) libraries were constructed for sequencing on an Illumina HiSeq X Ten platform. Finally, ~91.49 Gb (roughly 170× coverage) of raw sequencing data were obtained.

Preceding the filtered and corrected PacBio reads, genomic reads were mapped on both organelle genomes of closely related species by minimap2 v2.11-r797 (Li 2018); *Vaccinium macrocarpon* (NC_023338.1), *Rhazya stricta* (NC_024293.1), *Hesperelaea palmeri* (NC_031323.1), *Corchorus capsularis* (NC_031359.1), and *Vitis vinifera* (NC_012119.1) for mitochondrial assembly; and *Cymbidium ensifolium* (NC_028525.1), *Diospyros kaki* (NC_030789.1), *Pouteria campechiana* (NC_033501.1), *Diospyros blancoi* (NC_033502.1), and *Vaccinium macrocarpon* (NC_019616.1) for plastid assembly. All mapped reads were extracted for the following assemblies. Firstly, we used Canu v1.7 (Koren et al. 2017) and SMARTdenovo v1.0.0 (<https://github.com/ruanjue/smartdenovo>) (Liu et al. 2020) to generate two primary assemblies.

CONTACT Jian-Feng Mao  jianfeng.mao@bjfu.edu.cn  Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Qinghua East Road NO35, Haidian Dist. Beijing, 100083, China

*These authors contributed equally to this paper.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

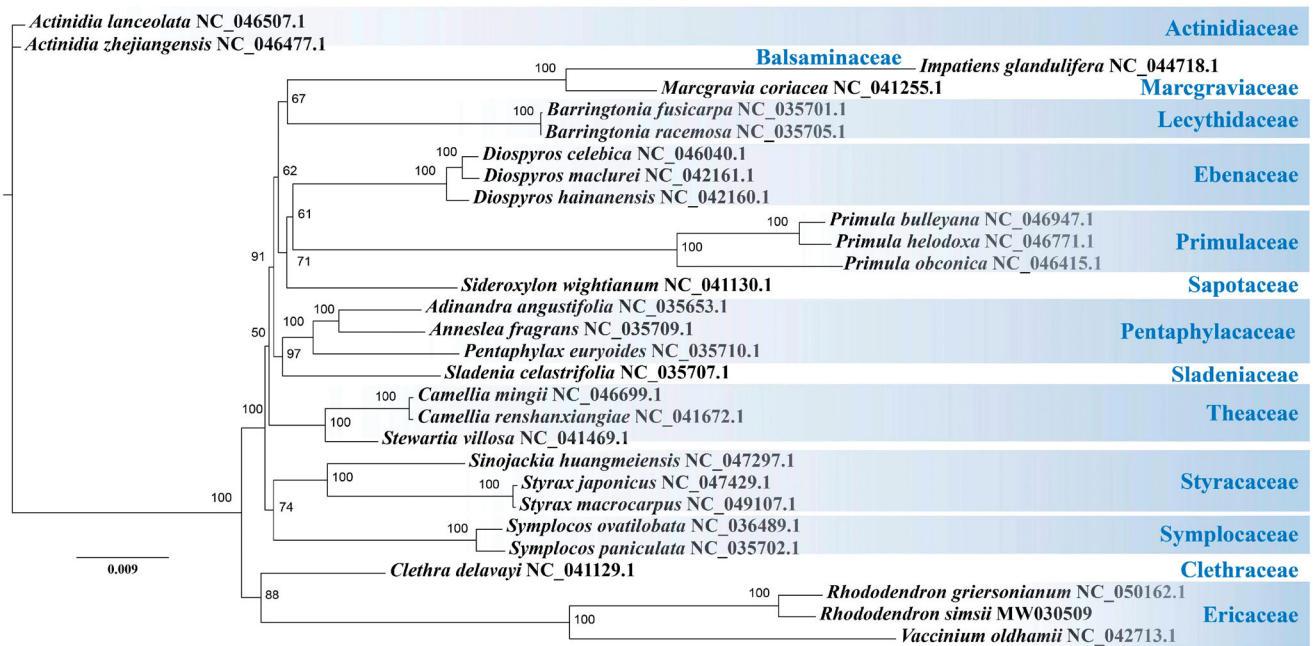


Figure 1. The maximum-likelihood tree based on single-copy genes from plastid genome of *R. simsii* and other 28 Ericales species. The numbers on the nodes indicate bootstrap values from 1000 replicates.

For the plastid genome, the assembly from SMARTdenovo was selected for high quality by checking the continuity of the assembly. Similarly, contigs from Canu were used to assemble the mitochondrial genome using SeqMan v11 (Swindell and Plasterer 1997). Finally, we annotated and illustrated the two organelle genomes using the OGAP pipeline (<https://github.com/zhangrengang/OGAP>). Within the pipelines, Exonerate v2.2.0 (Slater and Birney 2005) and AUGUSTUS v3.3.1 (Stanke et al. 2006) were employed for identity protein-coding genes, tRNAscan-SE v2.0.5 (Lowe and Eddy 1997) and BALT v36 (Kent 2002) were used to confirm tRNA and rRNA respectively.

Finally, the mitochondrial genome gave a linear scaffold of 802,707 bp, with 45.87% GC content. A total of 53 unique genes, consisting of 33 protein-coding, 17 tRNA, and 3 rRNA genes. Among these genes, eight protein-coding genes (*atp4*, *ccmB*, *ccmC*, *mttB*, *nad4L*, *nad9*, *rpl10*, *rps1*), one tRNA genes (*trnW-CCA-cp*) and one rRNA gene (*rrn26*) have a duplicated copy, *trnM-CAT-cp* gene has three copies and *trnM-CAT* gene has four copies. In addition, *nad1* and *nad2* gene was trans-spliced. Comparatively, the 152,214 bp long plastid genome is much smaller, with a much lower GC content of 35.74%. Moreover, this plastid genome has the quadripartite structure found in most land plant plastid genomes, containing 105 unique genes (including 4 rRNA, 26 tRNA and 75 protein-coding genes). All the protein-coding genes and rRNA genes are single copy, with the exception of *trnI-CAU*, *trnM-CAU* and *trnV-UAC* gene which occur in quadruplicated, triplicated and duplicated copy, respectively. Besides, *rps12* gene was trans-spliced.

Phylogenetic analysis was performed with the single-copy genes of plastid genomes from *R. simsii* and other 28 plant species in the Ericales order. MAFFT v7.471 (Katoh et al. 2002) was used to prepare sequences alignment which then was

trimmed with trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009). Finally, the phylogenetic analysis was executed by IQ-TREE v2.0.3 (Nguyen et al. 2015) with the model of GTR + F + R2 and 1000 bootstrap replicates. As expected, the resolved topology was consistent with that of the APG (Angiosperm Phylogeny Group, <http://www.mobot.org/MOBOT/research/APweb/>) (Chase et al. 2016), and the low supports were observed for some branches, in consistent with the previously reported common characteristic of the Ericales (Rose et al. 2018; Larson et al. 2019) (Figure 1).

Disclosure statement

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

Funding

This work was financially supported by grants from the Project of Construction of World Class Universities in Beijing Forestry University [2019XKJS0308] and the Fundamental Research Funds for the Central Universities in Beijing Forestry University [2018BLCB08].

ORCID

Jian-Feng Mao  <http://orcid.org/0000-0001-9735-8516>

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. MW030508 (mitochondrial genome) and MW030509 (plastid genome). The associated BioProject, SRA, and Bio-Sample numbers are PRJNA588298, SRS5624703, and SAMN13241185, respectively.

References

- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972–1973.
- Chase MW, Christenhusz M, Fay M, Byng J, Judd WS, Soltis D, Mabberley D, Sennikov A, Soltis PS, Stevens PF. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc*. 181(1):1–20.
- De Keyser E, Shu QY, Van Bockstaele E, De Riek J. 2010. Multipoint-likelihood maximization mapping on 4 segregating populations to achieve an integrated framework map for QTL analysis in pot azalea (*Rhododendron simsii* hybrids). *BMC Mol Biol*. 11(1):1.
- De Riek J, De Keyser E, Calsyn E, Eeckhaut T, Van Huylenbroeck J, Kobayashi N. 2018. Azalea. In: *Ornamental crops*. Amsterdam (The Netherlands): Springer; p. 237–271.
- Ding BY. 2009. Taxonomic study on *Rhododendron* subgen. Tsutsusi Sleumer (Ericaceae). Jiangsu (China): Abe Books.
- Galle FC. 1985. *Rhododendron or Azalea*. Azalea (Portland): Timber Press; p. 19–23.
- Katoh 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.
- Kent WJ. 2002. BLAT-the BLAST-like alignment tool. *Genome Res*. 12(4):656–664.
- Kobayashi N, Matsunaga M, Nakatsuka A, Mizuta D, Shigyo M, Akabane M. 2013. Chimeric inheritance of organelle DNA in variegated leaf seedlings from inter-subgeneric crossing of azalea. *Euphytica*. 191(1):121–128.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 27(5):722–736.
- Larson DA, Walker JF, Vargas OM, Smith SA. 2020. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *American Journal of Botany* 107(5):773–789.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34(18):3094–3100.
- Liu H, Wu S, Li A, Ruan J. 2020. SMARTdenovo: A de novo Assembler Using Long Noisy Reads. Preprints, 2020090207. doi: [10.20944/preprints202009.0207.v1](https://doi.org/10.20944/preprints202009.0207.v1).
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 25(5):955–964.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Rose JP, Kleist TJ, Löfstrand SD, Drew BT, Schönenberger J, Sytsma KJ. 2018. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol Phylogenet Evol*. 122:59–79.
- Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 6(1):31.
- Sleumer H. 1949. Ein system der Gattung *Rhododendron* L. *Bot Jahrb*. 74:511–553.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 7(1):62.
- Stevenson JB. 1930. The species of *Rhododendron*. England: Rhododendron Society.
- Swindell SR, Plasterer TN. 1997. Seqman. In: *Sequence data analysis guidebook*. Amsterdam (The Netherlands): Springer; p. 75–89.
- Yan LJ, Liu J, Möller M, Zhang L, Zhang XM, Li DZ, Gao LM. 2015. DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya–Hengduan Mountains. *Mol Ecol Resour*. 15(4):932–944.