




GENOME NOTE

REVISED Characterization of the Complete Chloroplast Genome of *Gaultheria nummularioides* D.Don 1825 (Ericaceae) [version 2; peer review: 2 approved]

Xiao-Juan Cheng ¹, Yan-Ling Xu¹, Catherine M. Bush², Yan-Jun Lin¹, Xin-Yu Du³, Lu Lu¹

¹School of Pharmaceutical Sciences and Yunnan Key Laboratory of Pharmacology for Natural Products, Kunming Medical University, Kunming, China

²Department of Biology, University of North Carolina at Greensboro, Greensboro, NC, USA

³Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

V2 First published: 21 Nov 2022, 11:1358
<https://doi.org/10.12688/f1000research.127937.1>
 Latest published: 04 Sep 2023, 11:1358
<https://doi.org/10.12688/f1000research.127937.2>

Abstract











Gaultheria nummularioides D.Don 1825 (Ericaceae) is a traditional Chinese medicinal plant used to treat rheumatoid arthritis. The complete chloroplast genome of *G. nummularioides* has been sequenced and assembled. The genome is 176,207 bp in total with one large single copy (LSC: 107,726 bp), one small single copy (SSC: 3,389 bp), and two inverted repeat regions (IRa and IRb; each 32,546 bp). The chloroplast genome encoded a total of 110 unique genes; the GC content of these genes is 36.6%. The results based on phylogenetic analysis of the complete chloroplast genome suggests that *G. nummularioides* diverged later than *G. praticola*, the sister relationship between *G. nummularioides* and the clade comprising *G. fragrantissima* Wall. 1820 and *G. hookeri* C.B. Clarke 1882 was strongly supported. This study provides additional information on the genetic diversity of *G. nummularioides*, its closely related taxa, and further exploration of chloroplast genomes in the Ericaceae family.


Keywords

Ericaceae; *Gaultheria nummularioides*; plastid genome; phylogeny

Open Peer Review

Approval Status  

	1	2
version 2		
(revision)		
04 Sep 2023		
version 1		
21 Nov 2022		

1. **Xing Huang**, Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China
2. **Hoang Dang Khoa Do** , Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the **Genomics and Genetics** gateway.

Corresponding author: Lu Lu (lulukmu@163.com)

Author roles: **Cheng XJ:** Writing – Original Draft Preparation; **Xu YL:** Data Curation; **Bush CM:** Writing – Review & Editing; **Lin YJ:** Supervision, Writing – Review & Editing; **Du XY:** Data Curation; **Lu L:** Project Administration, Supervision

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the National Natural Science Foundation of China under Grant number 31960080 and 42175139; the Scientific Research Fund Project of Yunnan Education Department under Grant number 2022Y197; the Program Innovative Research Team in Science and Technology in Yunnan Province under Grant number 202005AE160004.

Copyright: © 2023 Cheng XJ *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cheng XJ, Xu YL, Bush CM *et al.* **Characterization of the Complete Chloroplast Genome of *Gaultheria nummularioides* D.Don 1825 (Ericaceae) [version 2; peer review: 2 approved]** F1000Research 2023, 11:1358 <https://doi.org/10.12688/f1000research.127937.2>

First published: 21 Nov 2022, 11:1358 <https://doi.org/10.12688/f1000research.127937.1>

REVISED Amendments from Version 1

We explain in more detail the processing of the raw data, the addition of the chloroplast genome map, and making the sentences flow better.

Any further responses from the reviewers can be found at the end of the article

Introduction

As a traditional medicinal plant, *Gaultheria nummularioides* D. Don 1825 is often used to treat rheumatoid arthritis (Luo et al. 2018). This species is a small shrub predominantly distributed in alpine meadows from 1300 to 4600 m in China (mainly in the Himalaya-Hengduan Mountains region [HHM] in Yunnan, Sichuan, and Tibet), as well as in other regions in Southeast Asia (Fritsch et al. 2008). It is morphologically distinct from most other *Gaultheria* species from the HHM region based on characteristics such as growth habit and leaf and corolla indumentum (Middleton 1991). Middleton (1991) therefore placed *G. nummularioides* into a separate section, i.e., sect. *Monoanthe mona* Middleton from ser. *Leucothoides* (Airy-Shaw) Middleton of sect. *Brossaea* (L.) Middleton. Although *G. nummularioides* did fall within the Leucothoides clade within Gaultherieae, it is not a monophyletic species, as its evolution may have involved gene introgression or gene capture based on four cpDNA genic regions (*matK*, *rpl16*, *trnL-trnF*, and *trnS-trnG*) indicated one sample of *G. nummularioides* from Bhutan diverged earlier but other five samples from China later than those of

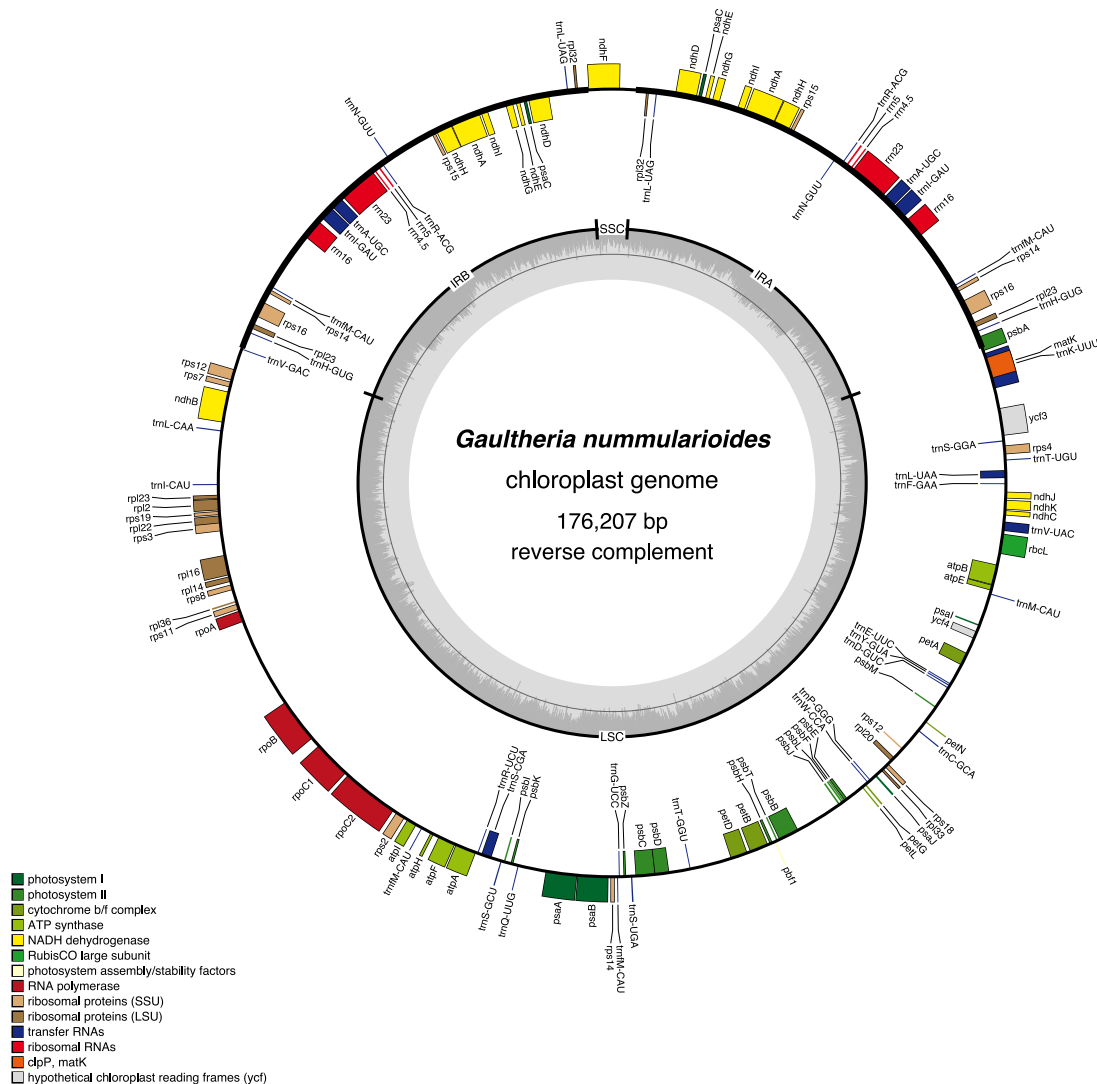


Figure 1. Chloroplast genome map of *Gaultheria nummularioides*.

G. praticola C.Y. Wu 1981 (Lu et al. 2010). Chloroplast DNA is typically maternally inherited and is characterized by a relatively small genome and slow mutation rate, so the complete chloroplast genome is of great value for understanding the phylogenetic relationships and maternal inheritance within the species and closely related taxa (Palmer et al. 1988; Jung et al. 2014; Xu et al. 2021). However, the sequence and characteristics of *G. nummularioides* had not been sequenced. This study presents the complete chloroplast genome of *G. nummularioides* and its resulting phylogenetic relationship.

Methods

The sample (Collected on August 3, 2007) was collected from Duoxiongla in Motuo County, Tibet, China (29°30'37"N, 94°51'25"E). The voucher specimen was deposited at the herbarium of the Kunming Institute of Botany (collection number: LL-07304; contact person: Lu Lu, lulukmu@163.com; <https://www.cvh.ac.cn/spms/detail.php?id=ea952b43>) under the voucher number 1248321. The plants were collected and studied in accordance with the regulations of the author's institution and national or international regulations; no specific permits were required.

After collecting fresh leaves in the field, they were dried using silica gel with discoloration. Subsequently, genomic DNA was extracted using the CTAB method (Doyle and Doyle 1987). Specifically, 1000 µL of 4×CTAB solution, preheated at 65°C and containing a small amount of mercaptoethanol, was added to ground leaves. Sequencing was carried out on the Illumina HiSeq 2000 platform, generating 2-3 Gb of paired-end reads with a length of 150 bp for each sample. The raw data obtained from sequencing can contain low-quality data or data with connectors. To ensure the quality of subsequent information analysis, the raw data was filtered to eliminate these low-quality data or paired reads with connectors. The specific characteristics of the data to be filtered and removed are as follows: (1) Single-end sequencing reads with an N content exceeding 10% of the read length and (2) Single-end sequencing reads with a number of low-quality bases (Q ≤ 5) that exceeded 50% of the read length. The resulting filtered data is referred to as clean data and is used for subsequent analyses.

The clean reads were *de novo* assembled (a assemble method for constructing genomes without a reference sequence) using GetOrganelle v1.7.5.0 (Jin et al. 2020). Bandage v0.8.1 was used for visualization of the assembly results (Wick et al. 2015). Annotation of the genome was conducted and then manually modified by Geneious v8.0.2

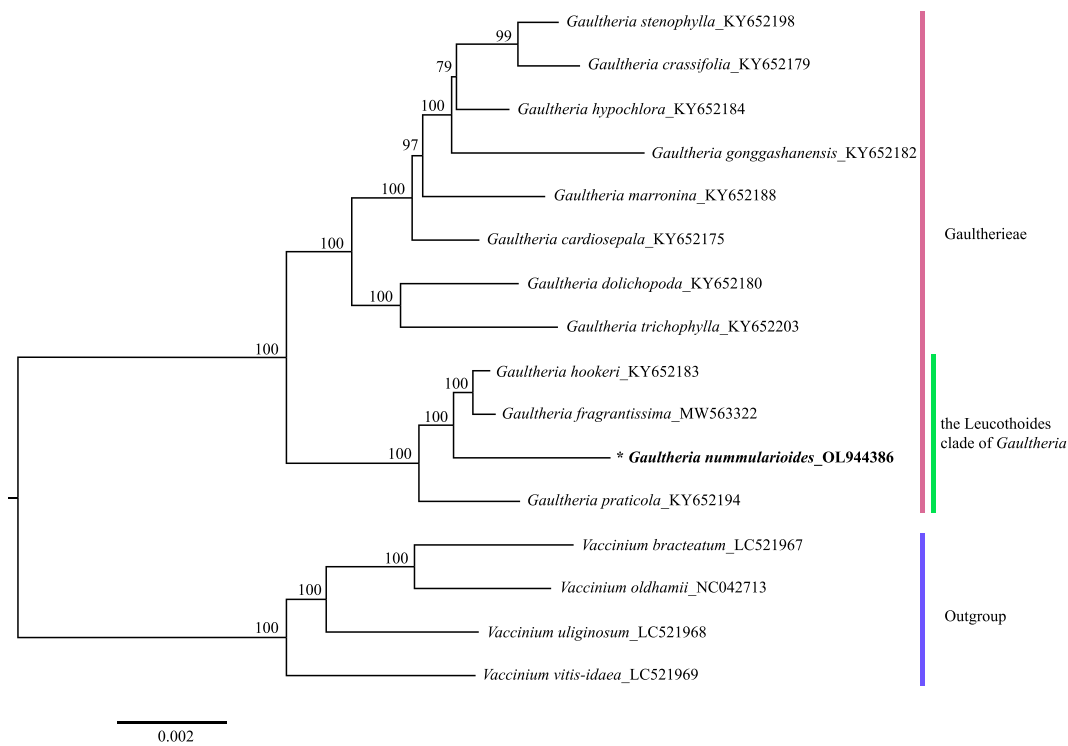


Figure 2. Maximum likelihood tree based on 16 complete chloroplast genomes of Ericaceae, including four outgroup species. GenBank accession numbers are listed beside the Latin name. The bootstrap support values based on 1000 replicates are shown next to the nodes. *Gaultheria nummularioides* is marked in bold and with an asterisk.

(Biomatters Ltd., Auckland, New Zealand). **OrganellarGenomeDRAW** v1.3.1 was used to draw the chloroplast genome map (Figure 1, Greiner et al. 2019). The complete chloroplast genome of *G. nummularioides*, eleven complete chloroplast genomes of related species from the genus *Gaultheria*, and four samples from the Vaccinieae tribe of Ericaceae as outgroup were aligned using **HomBlocks** v1.0 (Bi et al. 2018). A maximum likelihood phylogenetic tree was reconstructed by **RAxML** v8.2.X with GTRGAMMA substitution model and 1000 rapid bootstrap replicates (Stamatakis 2014).

Results

The complete chloroplast genome of *G. nummularioides* (GenBank accession no. OL944386) is a typical quadripartite structure with a total length of 176,207 bp. It is composed of a large single-copy (LSC: 107,726 bp) region, a small single-copy (SSC: 3,389 bp) region, and a pair of inverted repeats (IRs: 32,546 bp). The GC content in the chloroplast genome is 36.6%. This chloroplast genome encoded a total of 110 unique genes, of which 25 were duplicated once in the IR regions. Three genes were characterized by their multiple duplicates: *rpl23* and *rps14* have three duplicates each and *trnFM* has four duplicates in both LSC and IR regions. Out of the 110 genes, there were 76 protein-coding, 30 tRNA, and 4 rRNA genes. Maximum likelihood (ML) analysis results show that *G. nummularioides* diverged later than *G. praticola*. The results strongly support the sister-group relationship between *G. nummularioides* and the clade comprising *G. fragrantissima* Wall. 1820 and *G. hookeri* C.B. Clarke 1882 (Figure 2). The four taxa comprise the Leucothoides clade.

Data availability

Underlying data

GenBank: *Gaultheria nummularioides* chloroplast, complete genome, Accession number OL944386: <https://identifiers.org/ncbi/insdc:OL944386>.

BioProject: *Gaultheria nummularioides* Raw sequence reads, Accession number PRJNA744357: <https://identifiers.org/NCBI/bioproject:PRJNA744357>.

SRA: The complete chloroplast genome: *Gaultheria nummularioides*, Accession number SRX11377412: <https://www.ncbi.nlm.nih.gov/sra/?term=SRX11377412>.

Bio-Sample: Plant sample from *Gaultheria nummularioides*, Accession number SAMN20088414: <https://identifiers.org/biosample:SAMN20088414>.

References

- Bi GQ, Mao YX, Xing QK, et al.: **HomBlocks: a multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching.** *Genomics*. 2018; **110**(1): 18–22. [PubMed Abstract](#) | [Publisher Full Text](#)
- Doyle JJ, Doyle JL: **A rapid DNA isolation procedure for small quantities of fresh leaf tissue.** *Phytochem. Bull.* 1987; **19**: 11–15.
- Fritsch PW, Zhou LH, Lu L, et al.: **The Flowering Plant Genus *Gaultheria* (Ericaceae) in the Gaoligong Shan, along the Border Region of China and Myanmar.** *Proc. Natl. Acad. Sci.* 2008; **59**(6): 147–214.
- Greiner S, Lehwark P, Bock R: **OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organelle genomes.** *Nucleic Acids Res.* 2019; **47**: W59–W64. [PubMed Abstract](#) | [Publisher Full Text](#)
- Jin JJ, Yu WB, Yang JB, et al.: **GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes.** *Genome Biol.* 2020; **21**(1): 241. [PubMed Abstract](#) | [Publisher Full Text](#)
- Jung J, Kim KH, Yang K, et al.: **Practical application of DNA markers for high-throughput authentication of *Panax ginseng* and *Panax quinquefolius* from commercial ginseng products.** *J. Ginseng Res.* 2014; **38**(2): 123–129. [PubMed Abstract](#) | [Publisher Full Text](#)
- Lu L, Fritsch PW, Cruz BC, et al.: **Reticulate evolution, cryptic species, and character convergence in the core East Asian clade of *Gaultheria* (Ericaceae).** *Mol. Phylogenet. Evol.* 2010; **57**(1): 364–379. [Publisher Full Text](#)
- Luo BS, Gu RH, Kennelly EJ, et al.: ***Gaultheria* ethnobotany and bioactivity: Blueberry relatives with anti-inflammatory, antioxidant, and anticancer constituents.** *Curr. Med. Chem.* 2018; **25**(38): 5168–5176. [PubMed Abstract](#) | [Publisher Full Text](#)
- Middleton DJ: **Infrageneric classification of the genus *Gaultheria* L. (Ericaceae).** *Bot. J. Linn. Soc.* 1991; **106**(3): 229–258. [Publisher Full Text](#)
- Palmer JD, Jansen RK, Michaels HJ, et al.: **Chloroplast DNA Variation and Plant Phylogeny.** *Ann. Mo. Bot. Gard.* 1988; **75**(4): 1180–1206. [Publisher Full Text](#)
- Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics.* 2014; **30**(9): 1312–1313. [PubMed Abstract](#) | [Publisher Full Text](#)
- Wick RR, Schultz MB, Zobel J, et al.: **Bandage: interactive visualization of de novo genome assemblies.** *Bioinformatics.* 2015; **31**(20): 3350–3352. [Publisher Full Text](#)
- Xu YL, Du XY, Li YR, et al.: **The complete chloroplast genome of *Gaultheria fragrantissima* Wall. (Ericaceae) from Yunnan, China, an aromatic medicinal plant in the wintergreens.** *Mitochondrial DNA B: Resour.* 2021; **6**(6): 1761–1762.

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 14 September 2023

<https://doi.org/10.5256/f1000research.155153.r203746>

© 2023 Huang X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Xing Huang

Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

The authors have addressed most of the comments and revised the manuscript accordingly. I don't have new questions.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics and molecular biology of plants

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 05 September 2023

<https://doi.org/10.5256/f1000research.155153.r203745>

© 2023 Do H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Hoang Dang Khoa Do

Nguyen Tat Thanh University, Ho Chi Minh City, Ho Chi Minh, Vietnam

Dear Authors,

Thank you very much for your revised article which was modified based on the reviewer's comments. I do not have any questions for the version 2.

Best regards.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Phylogeny

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 29 March 2023

<https://doi.org/10.5256/f1000research.140483.r167786>

© 2023 Do H. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Hoang Dang Khoa Do

¹ Nguyen Tat Thanh University, Ho Chi Minh City, Ho Chi Minh, Vietnam

² Nguyen Tat Thanh University, Ho Chi Minh City, Ho Chi Minh, Vietnam

Thank you very much for studying on chloroplast genome of *Gaultheria nummularioides* which provides new information about genomic data of *Gaultheria* as well as the Ericaceae family. The content of the manuscript is suitable for a Genome Note of F1000Research.

I have some comments as the following:

1. Please ask an English editing service for checking the writing of the manuscript. Some sentences are very long.
2. Please provide more details of the sequencing process (i.e., single-end or paired-reads, size of reads 150 bp, 250 bp or 300 bp).
3. Please provide details on the step of filtering raw reads to get clean reads for *de novo* assembly.
4. Please use italic font for species names in the Data availability part.

Thank you.

Are the rationale for sequencing the genome and the species significance clearly described?

Partly

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomics, Phylogeny

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 25 Aug 2023

Xiaojuan Cheng

Question 1: Please ask an English editing service for checking the writing of the manuscript. Some sentences are very long.

Response: One of our co-author, Catherine M. Bush is a native speaker of English from USA and published many papers in English. English has been polished by her.

Question 2: Please provide more details of the sequencing process (i.e., single-end or paired-reads, size of reads 150 bp, 250 bp or 300 bp).

Response: Sequencing was carried out on the Illumina HiSeq 2000 platform, generating 2-3 Gb of paired-end reads with a length of 150 bp for each sample.

Question 3: Please provide details on the step of filtering raw reads to get clean reads for de novo assembly.

Response: The raw data obtained from sequencing may contain low-quality data or data with connectors. To ensure the quality of subsequent information analysis, it is necessary to eliminate these low-quality data or paired reads with connectors, referred to as filtering the raw data. The specific characteristics of the data to be filtered and removed are as follows: (1) Single-end sequencing reads with an N content exceeding 10% of the read length. (2) Single-end sequencing reads with a number of low-quality bases ($Q \leq 5$) exceeding 50% of the read length. The resulting filtered data is referred to as clean data and is used for subsequent analyses.

Question 4: Please use italic font for species names in the Data availability part.

Response: We correct that in the latest version.

Competing Interests: No competing interests were disclosed.

Reviewer Report 23 January 2023

<https://doi.org/10.5256/f1000research.140483.r160447>

© 2023 Huang X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Xing Huang

¹ Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

² Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

The manuscript has assembled and characterized the chloroplast genome of *Gaultheria nummularioides*, which is meaningful to Ericaceae family. However, several major concerns should be addressed:

1. In methods part, how was the dried leaves processed for DNA isolation by CTAB method? Which should be described.
2. The Illumina Solexa platform was used according to the methods, which is different from the SRA information. The authors should thoroughly check it.
3. How were the clean reads generated? Raw reads were directly generated by the sequencing machine.
4. In methods part, 'OrganellarGenomeDRAW v1.3.1 (Greiner *et al.* 2019) was used to draw the chloroplast genome map'. And where is the map presented in the manuscript?
5. The sequence of *Gaultheria griffithiana* (NC_057623.1) should be added in phylogenetic analysis.
6. The details of sequencing data should be described in results part.
7. The tRNA and rRNA couldn't be considered as unique gene.

Are the rationale for sequencing the genome and the species significance clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of the sequencing and extraction, software used, and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a usable and accessible format, and the assembly and annotation available in an appropriate subject-specific repository?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetics and molecular biology of plants

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 25 Aug 2023

Xiaojuan Cheng

Question 1: In methods part, how was the dried leaves processed for DNA isolation by CTAB method? Which should be described.

Response: After collecting fresh leaves in the field, they were dried using silica gel with discoloration. Subsequently, genomic DNA was extracted using the CTAB method (Doyle and Doyle, 1987). Specifically, 1000 μ L of 4 \times CTAB solution, preheated at 65°C and containing a small amount of mercaptoethanol, was added to leaf grinding.

Question 2: The Illumina Solexa platform was used according to the methods, which is different from the SRA information. The authors should thoroughly check it.

Response: Upon further review, the sequencing methodology utilizes the Illumina HiSeq 2000 platform and has been rectified.

Question 3: How were the clean reads generated? Raw reads were directly generated by the sequencing machine.

Response: Raw data obtained from sequencing contains some low-quality data or data with connectors. In order to ensure the quality of subsequent information analysis, it is necessary to remove these low-quality data or paired reads with connectors, i.e., to filter the raw data. The specific characteristics of the data to be filtered and removed are as follows: (1) the N content in single- end sequencing reads exceeds 10% of the length of the reads; (2) the number of low-quality ($Q \leq 5$) bases in single-end sequencing reads exceeds 50% of the length of the reads. The filtered data are called clean data and are used for subsequent analyses.

Question 4: In methods part, 'OrganellarGenomeDRAW v1.3.1 (Greiner et al. 2019) was used to draw the chloroplast genome map'. And where is the map presented in the manuscript?

Response: Thank you for the reminder! We add the map in the latest version.

Question 5: The sequence of *Gaultheria griffithiana* (NC_057623.1) should be added in phylogenetic analysis.

Response: This manuscript focuses on reporting the chloroplast genome structure of a

single species within the genus *Gaultheria*. We have another manuscript for displaying the plastid phylogenomic relationships of the whole clade (ca. 60 species) within *Gaultheria* at species level (unpublished data). Add one more species cannot help on resolving the phylogenetic position of this species in such a large clade.

Question 6: The details of sequencing data should be described in results part.

Response: we reference to the similar description of the other articles and revise it accordingly.

Question 7: The tRNA and rRNA couldn't be considered as unique gene.

Response: We know each of tRNA gene or rRNA gene has two copies in the IR region. The "unique gene" we used in our study means that tRNA and rRNA have not appeared in any other place but in the IR region. We described this in the way following the literature such as Henriquez et al. (2020) and (Gao et al., 2023). We revised this accordingly.

(a) The genomes contain 114 unique genes, including four rRNA genes, 80 protein-coding genes, and 30 tRNA genes (Henriquez et al., 2020).

(b) The chloroplast genomes of *R. rubiginosa*, *R. hybrida*, and *R. acicularis* were conserved and contained 115 unique genes, of which 80 were protein-coding genes, 31 were transfer RNA (tRNA) genes, and 4 were ribosomal RNA (rRNA) genes (Gao et al., 2023).

Reference:

Henriquez, C. L., Abdullah, Ahmed, I., Carlsen, M. M., et al. 2020. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). *Planta*, 251, 1–16.

Gao, C., Li, T., Zhao, X., et al. 2023. Comparative analysis of the chloroplast genomes of *Rosa* species and RNA editing analysis. *BMC Plant Biology*, 23(1), 318.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research