The complete chloroplast genome of *Miliusa glochidioides* (Annonaceae) and phylogenetic analysis

Yangying Gan\(^a\), Qi Zhang\(^b\) and Jingyao Ping\(^c\)

\(^a\)Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Economics and Information, Guangdong Academy of Agricultural Sciences, Guangzhou, China; \(^b\)Guangdong Provincial Key Laboratory for Crop Germplasm Resources Preservation and Utilization, Agro-biological Gene Research Center, Guangdong Academy of Agricultural Sciences, Guangzhou, China; \(^c\)College of Life Sciences, South China Agricultural University, Guangzhou, China

**ABSTRACT**

The chloroplast (cp) genome of *Miliusa glochidioides* has been fully sequenced. The cp genome of this species has a typical quadripartite structure comprised of four parts: a large single copy (LSC; 88,782 bp) region, a small single copy (SSC; 18,949 bp) region, and two inverted repeat (IR; 26,029 bp each) regions. The full length of the cp genome is 159,789 bp; its GC content is 36.7\%, and it encodes a total of 129 genes including 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Among the protein-coding genes, nine (*rps16*, *rpl2*, *rpl16*, *atpF*, *rpoC1*, *petB*, *petD*, *ndhA*, and *ndhB*) contain one intron, and three (*rps12*, *clpP*, and *ycf3*) have two introns. A maximum-likelihood (ML) phylogenetic tree shows that *M. glochidioides* is a sister to Chieniodendron hainanense.

*Miliusa glochidioides* Handel-Mazzetti 1933 is an evergreen shrub with a height of 1.5 meters that belong to family Annonaceae, and is distributed on forested slopes below 900 meters in Guangxi, China, where it is endemic. *Miliusa* plants are often used as ropes or kindling due to their strong and slender stems (Li et al. 2011). In the Annonaceae Checklist (http://herbarium.botanik.univie.ac.at/annonaceae/listTax.php), this species is treated as a synonym of *Orophae polyocarpa* (Rainer and Chatrou 2006). However, serious discrepancies exist in the descriptions of these two taxa within China (Li et al. 2011). Obtaining the complete chloroplast (cp) genome may provide a molecular basis for investigating phylogenetic relationships and population variation.

Fresh leaves were collected from the South China Botanical Garden of the Chinese Academy of Sciences (E113°36', N23°18'). The specimens were stored in the Herbarium of South China Agricultural University (Jingyao Ping, email: pingjnyao@foxmail.com) under specimen code PJY-GXYDH2110 (herbaria acronyms follow Thiers, 2021, continuously updated). Total DNA was extracted using the E.Z.N.A.\(^\circledR\) Plant DNA Kit (OMEGA, Shanghai). After the genomic DNA extracted from a sample was tested for quality, the DNA was broken through physical methods (ultrasonication), and then purified to construct a sequencing library. The steps of this process are as follows: DNA end repair, addition of A to the 3' end, ligation of sequencing adapters, recovery of the target fragment through agarose gel electrophoresis, polymerase chain reaction amplification of the target fragment, and finally, construction of a sequence library. The Illumina NovaSeq6000 platform was used for paired-end sequencing with a 150-bp read length. The sequences were spliced using the stitching software NOVOPlasty V4.2 (https://github.com/ndierckx/NovoPlasty) (Dierckxsens et al. 2017), and GetOrganelle V1.7.0 (https://github.com/Kinggerm/GetOrganelle) (Jin et al. 2020) was used for identification of the optimal assembly results. The assembled cp genome was annotated using PGA (Qu et al. 2019) and GeSeq (Tillich et al. 2017) software. The program MAFFT v7.311 (Katoh and Standley 2013) was used for multiple sequence alignment of the complete cp genome of *M. glochidioides* with those of other eight plants downloaded from GenBank. RAxML v8.0 software (Stamatakis 2014) was used to construct an maximum-likelihood (ML) tree with 1000 bootstrap replicates using the GTRGAMMAI substitution model.

The cp genome of *M. glochidioides* (GenBank accession number: OM047203) has a typical quadripartite structure with a total length of 159,789 bp and GC content of 39.2\%. The lengths of LSC, SSC and IR are 88,782, 18,949 and 26,029 bp, respectively. This cp genome encodes a total of 129 genes, including 84 protein-coding genes, 37 tRNA, and eight rRNA genes. Among the protein-coding genes, nine (*rps16*, *rpl2*, *rpl16*, *atpF*, *rpoC1*, *petB*, *petD*, *ndhA*, and *ndhB*) contain one intron, and three (*rps12*, *clpP*, and *ycf3*) contain two introns. The ML tree strongly supports *M. glochidioides* and Chieniodendron hainanense as sister groups (Figure 1). *M. glochidioides* is the first species of its genus whose cp genome has been sequenced, providing reliable information for phylogenetic and cp genomics studies of Annonaceae.
Ethics statement

Research on plants (cultivated or wild), including the collection of plant material has been carried out conformed to the requirement of our institution and International ethics, which did not cause damage to the local environment. Field studies have complied with local legislation and appropriate permissions/license were granted while taking samples from a preserved/protected land. There are no ethical issues and other conflicts of interest in this study.

Author contributions

YG and JP sampled the material, analyzed some of the data, made recommendations and wrote the manuscript. QZ designed the study and modified the manuscript. All authors read and contributed to the final version of the manuscript.

Disclosure statement

The authors declare that they do not have any conflict of interest.

Figure 1. Maximum-likelihood tree based on the sequences of nine complete chloroplast genomes. Magnolia alba and Liriodendron chinense were selected as outgroup. Numbers in the nodes were bootstrap values from 1000 replicates.

Funding

This work was supported by the National Natural Science Foundation of China (31800184).

Data availability statement

The genome sequence data that support the findings of this study are available in GenBank of NCBI at [https://www.ncbi.nlm.nih.gov] under the accession no. MT742546, MT742547, MK035708, MH924590, MH992130, KU563738, NC_037005, NC_030504, and OM047203. These data were derived from the following resources available in the public domain: [https://www.ncbi.nlm.nih.gov/nuccore/]. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA793237, SRR17406795, SAMN24519632, respectively.

References

Dierckxsens N, Mardulyn P, Smits G. 2017. Novoplasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21(1):241.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Li BT, Gilbert MG, Schatz G. 2011. Annonaceae. Flora of China. Vol. 19. Beijing: Science Press (In Chinses).

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 15(1):1–12.

Rainer H, Chatrou LW. 2006. AnnonBase: world species list of Annonaceae. Wilson Bulletin. 101(4):606–620.

Stamatakis A. 2014. RaxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312–1313.

Tillich M, Lehward P, Pellizzier T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6–W11.