Complete chloroplast genome sequence and phylogenetic analysis of Mallotus paniculatus (Lam.) Müll. Arg. (Euphorbiaceae)

Zhuowei Li\textsuperscript{a}, Nong Zhou\textsuperscript{b}, Ming Liu\textsuperscript{b} and Fuqiang Yin\textsuperscript{b}

\textsuperscript{a}College of Environmental and Chemical Engineering, Chongqing Three Gorges University, Chongqing, China; \textsuperscript{b}College of Biological and Food Engineering, Chongqing Three Gorges University, Chongqing, China

\textbf{ABSTRACT}

\textit{Mallotus paniculatus} (Lam.) Müll. Arg. 1865 (Euphorbiaceae) is a shrub or small tree with medicinal properties that is distributed across Southeast Asia. In this study, we sequenced the complete chloroplast genome of \textit{M. paniculatus} to study phylogenetic relationships within the family Euphorbiaceae Juss. The complete chloroplast genome of \textit{M. paniculatus} was 164,455 bp in length, with an overall GC content of 35.3\%. It was found to consist of a long single copy region of 89,021 bp, a small single copy region of 18,524 bp, and a pair of inverted repeats of 28,455 bp. Results indicated that the chloroplast genome contains a total of 131 genes, including 78 protein-coding genes, 37 tRNA genes, eight rRNA genes, and eight pseudogenes. The phylogenetic tree showed that \textit{M. paniculatus} is closely related to \textit{Mallotus japonicus} and \textit{Mallotus peltatus}. The total length of the chloroplast genome was 164,455 bp, and the total GC content was 35.3\%. The chloroplast genome of \textit{M. paniculatus} showed a typical quadripartite structure, including a pair of inverted repeats of 28,455 bp each, separated by a small single copy region of 18,524 bp and a long single copy region of 89,021 bp. The sequence of the chloroplast genome revealed a total of 131 genes, including 78 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Further, we identified a total of eight pseudogenes (\textit{rps16}, \textit{rbcl}, \textit{petB}, \textit{ndhF}, \textit{ndhD}, \textit{ndhG}, and a pair of reverse repeats of \textit{ycf1}). To study their phylogenetic relationships, the chloroplast genome sequences of 26 Euphorbiaceae species and two
Daphniphyllaceae species were downloaded from the GenBank database. Sequence alignment was performed using MAFFT (v. 7.427) (Katoh and Standley 2013). A maximum-likelihood phylogenetic tree was constructed using RAxML (Stamatakis 2014), with 1000 bootstrap replicates, and by applying the GTR+GAMMA model. The maximum likelihood phylogenetic tree showed that among the species analyzed, *M. japonicus* and *M. peltatus* were the closest relatives of *M. paniculatus* (Figure 1). This study provides a scientific basis to exploit these resources and a foundation for further phylogenetic analyses of *M. paniculatus*.

**Ethics statement**
*Mallotus paniculatus* (Lam.) Müll. Arg. is not a protected plant, and our research did not damage any of the *M. paniculatus* population. Therefore, no special permission was needed.

**Authors’ contributions**
Fuqiang Yin was primarily responsible for the design of the experiment and approved the final version of the paper; Zhuowei Li was primarily responsible for the writing and revision of the paper; Nong Zhou and Ming Liu participated in sequencing data assembly and annotation work; Fuqiang Yin and Zhuowei Li analyzed and interpreted the data. All authors are accountable for all aspects of the work.

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

**Funding**
This work was supported by the Chongqing Natural Science Foundation [cstc2018jcyjAX0267] and the Scientific and Technological Research Program of the Chongqing Municipal Education Commission [KJCK2020046; KJQN201901222].

**Data availability statement**
The data that support the findings of this study are openly available in GenBank of NCBI at [https://www.ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov), accession number MZ597547. The associated BioProject, SRA, and BioSample numbers are PRJNA758621, SRR15663411, and SAMN21031881, respectively.

**References**
Bolger AM, Lohse M, Usadel B. 2014. Trimmmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.
Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.
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 tool kit 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.
Lamarck JB. 1865. *Mallotus paniculatus* (Lam.) Müll.Arg. *Linnaea*. 34(2): 189.
Qiu HX. 1996. *Flora Reipublicae Popularis Sinicae* (44-2). Beijing: Science Press; p. 35–36.
Rivière C, Nguyen Thi Hong V, Tran Hong Q, Chataigné G, Nguyen Hoai N, Dejaegher B, Tistaert C, Nguyen Thi Kim T, Vander Heyden Y, Chau Van M, et al. 2010. Mallotus species from Vietnamese mountainous areas: phytochemistry and pharmacological activities. *Phytochem Rev.* 9(2):217–253.
Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47(W1):W65–W73.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
Tillich M, Lehwark P, Pellizzer 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.
Wang WJ, Jiang JH, Chen YG. 2013. Steroids from *Mallotus paniculatus*. *Chem Nat Compd.* 49(3):577–578.
Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour.* 14(5):1024–1031.
Zhu CL, Ma JX. 2014. Chemical constituents of *Mallotus paniculatus*. *Zhong Yao Cai*. 37(8):1385–1387.