

The complete chloroplast genome sequence of *Lithocarpus longinux* (Fagaceae)

Chun-Ya Wu^{a,b}, Lin Lin^a, Kai-Ping Yao^a, Rui-Jie Yang^a and Min Deng^{a,b}

^aSchool of Ecology and Environmental Sciences, Yunnan University, Kunming, Yunnan, China; ^bYunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming, Yunnan, China

ABSTRACT

Lithocarpus longinux (Hu) Chun ex Y.C.Hsu & H.Wei Jen is a Critically Endangered tree distributed in Ma-li-po county in the Southeastern Yunnan Province China. Less than ten individuals have been found since the species was established 70 years ago. In this study, we assembled and annotated the complete chloroplast genome of *L. longinux*. The complete chloroplast genome of the species is 161,420 bp in length and has a GC content of 36.8%, including one large single-copy region (LSC, 90,409 bp), one small single-copy region (SSC, 19,255 bp), and two copies of inverted repeat regions (IRs, 25,878 bp). A total of 112 unique genes were detected, including 81 protein-coding genes, 29 tRNA genes, and 4 rRNA genes. Phylogenetic analysis of 31 representative chloroplast genomes of the Fagales suggests *Lithocarpus* is monophyletic with strong bootstrap support and that *L. longinux* is closely related to *L. balansae*.

ARTICLE HISTORY

Received 23 September 2021
Accepted 18 June 2022

KEYWORDS

Lithocarpus longinux;
Fagaceae; chloroplast
genome; Illumina
sequencing

Lithocarpus Blume is one of the dominant woody lineages widespread in the East Asian monsoon evergreen broad-leaved forests with ca. 400 species (Govaerts and Frodin 1998, Huang et al. 1998, Huang et al. 1999). Many species of the genus are canopy trees in the forests playing significant roles in regional ecosystem functioning and biodiversity conservation (Cannon et al. 2018).

Lithocarpus longinux is a Critically Endangered tree with a restricted distribution in Ma-li-po county of the Southeastern Yunnan Province, China. Its habitat is located in mid-altitude montane cloud limestone forests. The species was first described in 1951 (Hu 1951), called *Pasania longinux*. This species name *Lithocarpus longinux* was first published in 1976 (Chun et al. 1976). However, according to Flora of China (both Chinese and English versions), the species used to be treated as a synonym of *L. areca* (Huang et al. 1998, Huang et al. 1999). Fortunately, some researchers found that the entire leaf margin and prominent concentric rings on the cupule of *L. longinux* differ from *L. areca*. Therefore, the name *L. longinux* was reinstated by Zhou Wei et al. (Zhou et al. 2013). Moreover, based on our previous Endangered status assessments on Asian Fagaceae, less than 10 individuals of *L. longinux* were found. The species urgently requires conservation.

Like other *Lithocarpus* species, *L. longinux* is a typical biennial fruiting species. Its flowering time is from March to April, and the fruits fully ripen from September to October of the following year (Zhou et al. 2013). Despite the genus' significant ecological and economic importance, the genome information on *Lithocarpus* is limited, particularly in the

chloroplast genome. Only three whole chloroplast genomes of the genus have ever been uploaded to GenBank. Meanwhile, the phylogenetic framework on *Lithocarpus* is still not resolved due to limited sampling and low throughput of SNPs generated by Sanger-based sequencing markers, which cannot provide a robust phylogenetic framework. As a result, the phylogenetic trees on *Lithocarpus* obtained by previous works only received mediate-low credibility supports at the main topologies (Yang et al. 2018). Therefore, the use of genome data is essential for untangling the genus' evolutionary history. Compared with nuclear genomes, chloroplast genomes have multiple advantages, e.g. relatively conservative structure, uniparental inheritance, and easy sequencing (Birky et al. 1983). Therefore, the complete chloroplast genome has been widely used in the phylogenetic studies of various plant taxa (Daniell et al. 2016).

In this study, the complete chloroplast genome of *L. longinux* was sequenced, assembled, and annotated. The phylogenetic tree of Fagaceae was constructed to test the discrimination abilities of the chloroplast genome at different taxonomic ranks. Our study can provide helpful genomic information for conserving this Endangered species and to further explore the chloroplast genome evolution pattern of the genus *Lithocarpus*.

Fresh leaves of *L. longinux* were collected from Ma-li-po county, Yunnan Province, China (23.1673 N, 104.8540 E, alt. 2300 m). The voucher specimen was deposited at herbarium of Kunming Institute of Botany (Tao Deng, dengtao@mail.kib.ac.cn) under the voucher number DM22970. Total genomic DNA was extracted from silica-dried leaves based on

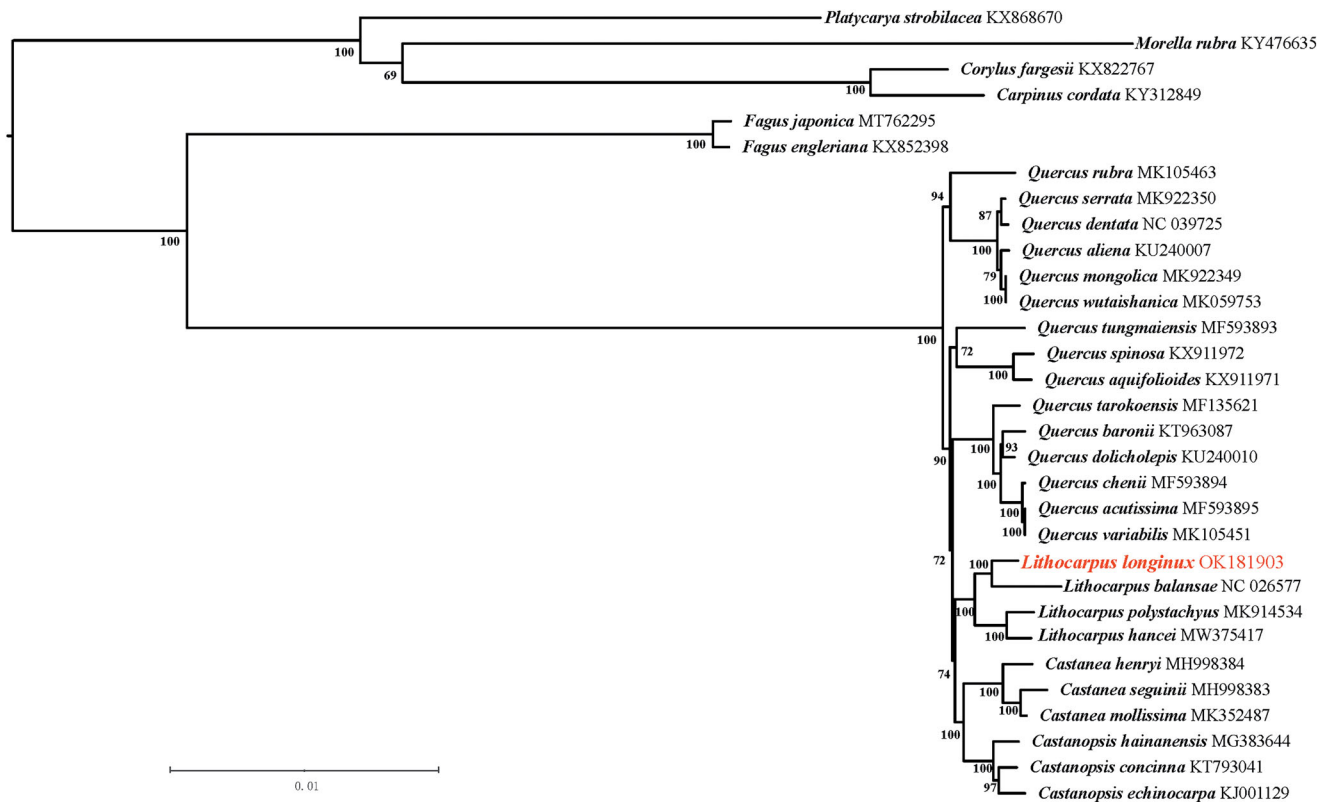


Figure 1. The phylogenetic tree of *Lithocarpus longinux* and other 30 species from Fagales (incl. 26 species from family Fagaceae, two species from family Betulaceae, one species from family Myricaceae and family Juglandaceae respectively) were reconstructed based on maximum-likelihood analysis with 1000 bootstraps under the GTR + GAMMA model. The bootstrap value for each clade is provided above.

modified CTAB method (Doyle and Doyle 1987). Whole-genome sequencing was conducted with the paired-end (PE) reads of 2×150 bp on the Illumina NovaSeq platform. Raw reads were filtered and trimmed to remove the low quality reads by Fastp (Chen et al. 2018). A total of 60,366,324 clean reads were used for *de-novo* assembly with GetOrganelle v1.7.2 (Jin et al. 2020). Gene annotation was performed by the pipeline PGA (Qu et al. 2019) and Geneious V10.2.3 (Kearse et al. 2012). Circular mapping of the *L. longinux* chloroplast genome was performed using the OGDRAW online tool (Lohse et al. 2013). The annotated genome sequence was deposited in GenBank under Accession Number OK181903.

The complete chloroplast genome of *L. longinux* is 161,420 bp in length. It has a typical quadripartite structure including one large single-copy (LSC) region (90,409 bp), one small single-copy (SSC) region (19,255 bp), and two copies of inverted repeat (IRs) regions (25,878 bp). The overall GC content of this species is approximately 36.8% (LSC, 34.6%; SSC, 31.0%; IRs, 42.7%). A total of 112 unique genes, including four rRNA genes, 29 tRNA genes, and 81 protein-coding genes, were annotated.

To infer the systematics of *L. longinux*, 31 published complete chloroplast genomes of Fagales were aligned by MAFFT 7.487 (Rozewicki et al. 2019). Finally, a maximum likelihood (ML) tree was constructed using RAxML v.7.2.6 with 1000 bootstraps under the GTRGAMMA model (Stamatakis 2014). The monophyletic status of *Lithocarpus* with two main clades within the genus was robustly supported (100% bootstrap). *L. longinux* and *L. balansae* are sister species (Figure 1).

This result suggest that complete chloroplast genome data has good potential to resolve species phylogenetic placement in *Lithocarpus*.

Author contributions

Min Deng and Chun-Ya Wu were involved in the conception and design, Min Deng and Lin Lin collected the plant materials, Chun-Ya Wu, Kai-ping Yao and Rui-Jie Yang performed experiments, Chun-Ya Wu analyzed the data, Min Deng and Chun-Ya Wu wrote the manuscript.

Ethical approval

Research and collection of plant material was conducted according to the guidelines provided by Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations. Permission was granted by the National Natural Science Foundation of China.

Disclosure statement

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

Funding

This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under [Grant number 31972858] and Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations under [Grant number PSESP2021F01].

Data availability statement

The complete chloroplast genome sequence of *Lithocarpus longinux* is deposited in the GenBank database (<https://www.ncbi.nlm.nih.gov/>) under the accession number OK181903. The associated BioProject, SRA, and BioSample numbers are PRJNA769608, SRR16297765, and SAMN22153924, respectively.

References

- Birky CW, Maruyama T, Fuerst P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*. 103(3):513–527.
- Cannon CH, Brendel O, Deng M, Hipp AL, Kremer A, Kua C-S, Plomion C, Romero Severson J, Sork VL. 2018. Gaining a global perspective on Fagaceae genomic diversification and adaptation. *New Phytol*. 218(3): 894–897.
- Chen S-F, Zhou Y-Q, Chen Y-R, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34(17):i884–i890.
- Chun , Hsu Y-C, Jen X-W. 1976. *Lithocarpus longinux* (Hu) Chun ex Y.C.Hsu & H.W.Jen, *Acta Phytotax. Sin.* 14(2):77.
- Daniell H, Lin C-S, Yu M, Chang W-J. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol*. 17(1):134.
- Doyle J, Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.
- Govaerts R, Frodin DG. 1998. World checklist and bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae and Ticodendraceae). London: Kew Publishing; p. 150–151.
- Hu H-H. 1951. Additional notes on the Fagaceae of Yunnan I. *Acta Phytotaxon Sin.* 1:103–118.
- Huang C-C, Chang Y-T, Bartholomew B. 1999. Fagaceae. In: Wu, C-Y, Raven PH, editors. *Flora of China*. Volume 4. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press; p. 380–400.
- Huang C-C, Chang Y-T, Hsu Y-C, Jen X-W, Fu L-K, Chen C-J, Tang Y-C, Kuang K. 1998. Fagaceae. In: Chen W-Y, Huang C-C, editors. Volume 22. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Pres; p. 263–332.
- 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.
- Kearse M, Moir R, Wilson A, Stones Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 28(12): 1647–1649.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 41(Web Server issue):W575–W581.
- Qu X-J, Moore MJ, Li D-Z, Yi T-S. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15(1):1–12.
- Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res*. 47(W1):W5–W10.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
- Yang C-K, Chiang Y-C, Huang B-H, Ju L-P, Liao P-C. 2018. Nuclear and chloroplast DNA phylogeography suggests an Early Miocene southward expansion of *Lithocarpus* (Fagaceae) on the Asian continent and islands. *Bot Stud*. 59(1):27.
- Zhou W, Huang Y-S, Xia N-H. 2013. Reinstatement of the name *Lithocarpus longinux* (Fagaceae). *Nord J Bot*. 31(2):186–189.