

The complete chloroplast genome assembly of *Castanopsis orthacantha* Franch. 1899 (Fagaceae) in Southwestern China

Si Yin and Yong Gao

College of Biological Resource and Food Engineering, Qujing Normal University, Qujing, Yunnan, China

ABSTRACT

Castanopsis orthacantha Franch. 1899 is one representative tree species distributed in the evergreen broad-leaved forests of southwestern China. This species is an important source of timber for manufacturing furniture, floors, and paper. It also plays a significant role in maintaining ecological balance and stability. To help with the genetic diversity assessment of *C. orthacantha*, we sequenced and assembled the first complete chloroplast genome. The length of the chloroplast genome was 160,588 bp, with a typical quadripartite structure (GenBank accession no. OR900101). The large single-copy (LSC), small single-copy (SSC), and two inverted repeats (IRs) were 90,237 bp, 18,953 bp, and 25,699 bp in length, respectively. We annotated one hundred and thirty genes across the chloroplast genome, including 86 protein-coding genes (79 are unique), 37 tRNA genes (29 are unique), and eight rRNA genes (four are unique). Seventeen genes had one intron, and four were detected with two introns. The maximum likelihood phylogeny suggested that *C. orthacantha*, *C. lamontii*, *C. sclerophylla*, and *C. hainanensis* formed a clade with a high bootstrap value. This newly sequenced chloroplast genome assembly will aid in the population genetic and phylogenetic studies of *Castanopsis* species in the future.

ARTICLE HISTORY

Received 21 December 2023
Accepted 9 April 2024

KEYWORDS

Castanopsis orthacantha;
Fagaceae; chloroplast
genome; phylogenetic
analysis

Introduction

The *Castanopsis* genus, which belongs to the Fagaceae family, is a group of evergreen trees that comprises about 120 species that are widely distributed in tropical and subtropical Asia (Huang et al. 1999). The *Castanopsis* species have been considered of great ecological and economic value in China (Chen et al. 2018). Most species are significant components of natural forests, and the timbers are valuable materials for making furniture, floors, and paper (Hinsinger and Strijk 2017). Approximately 58 *Castanopsis* species are found in China, and the natural distribution region of this genus in China is tropical and subtropical areas in southern and southwestern China (Liu and Zhou 2006). Among these species, *C. orthacantha* Franch. 1899 mainly distributed across Yunnan, Guizhou, and Sichuan provinces in southwestern China, with an elevation ranging from 1500 to 3200 meters. The adult trees of *C. orthacantha* are 10–15 m tall, and the branches, leaf blades, and rachis of inflorescences are glabrous (Huang et al. 1999). The forest of *C. orthacantha* is one representative vegetation type in central Yunnan evergreen broad-leaved forests and plays a significant role in maintaining ecological balance and stability. However, no genetic diversity assessment of *C. orthacantha* has been conducted to date, which hinders the protection and utilization of this species. In this

study, we sequenced and assembled the complete chloroplast genome sequence of *C. orthacantha*, which could help the phylogeographic and population genetic study of this species.

Materials and methods

One individual of *C. orthacantha* was sampled in Maxiong Mountain, Qujing, Yunnan province, China during the field investigation in 2023 (25°54′8.6″ N, 103°56′53.4″ E). The specimen (yinsi_YJK2-20230520, Yin Si, May 2023) was deposited into the herbarium of Qujing Normal University (Yong Gao, 562698574@qq.com, gaoyong@mail.qjnu.edu.cn). Genomic DNA was isolated from fresh leaves using a commercial DNA isolation kit (Tiangen, Beijing, China). Novogene Bioinformatics Technology (Beijing, China) conducted the construction of a DNA sequencing library and next-generation sequencing. Briefly, genomic DNA was randomly fragmented by Covaris instrument (Covaris, Woburn, USA) to a size of 350 bp. DNA fragments were selected and ligated with sequencing adapters, followed by PCR amplification. Finally, the DNA library was pair-end sequenced with a read length of 150 bp on the Illumina Novaseq 6000 platform (Illumina, San Diego, CA, USA).

CONTACT Yong Gao  562698574@qq.com, gaoyong@mail.qjnu.edu.cn  College of Biological Resource and Food Engineering, Qujing Normal University, Qujing, Yunnan, 655011, China.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2024.2342940>.

© 2024 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-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

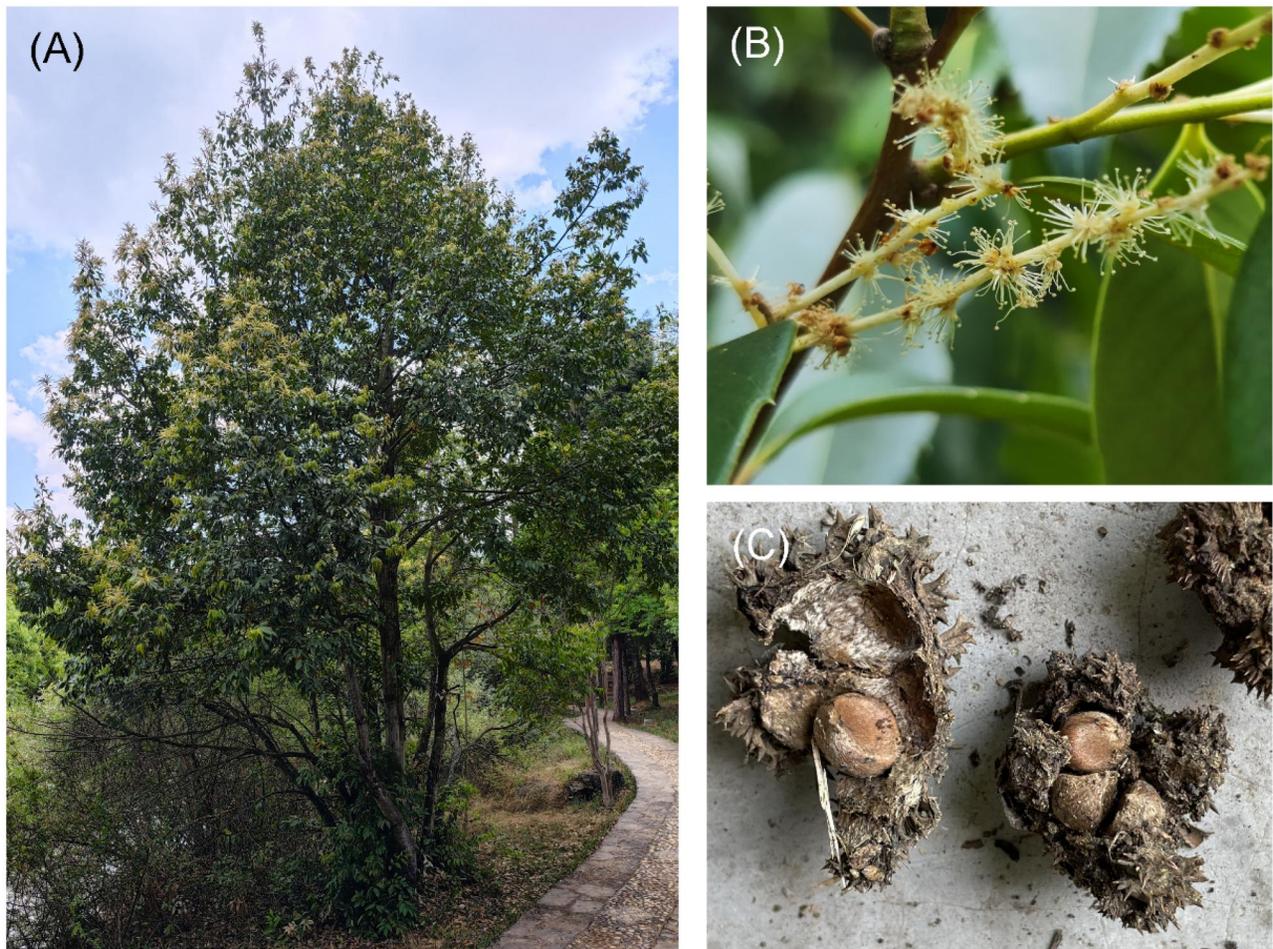


Figure 1. Morphological characteristics of the plant (a), flower (B), and fruit (C) of *Castanopsis orthacantha*. The photos were taken by the author Yong Gao. The adult trees of *C. orthacantha* are 10–15 m tall. The branches, leaf blades, and rachis of inflorescences are glabrous. There are usually three female flowers and nuts per cupule.

Before assembling the genome, we applied quality control to the raw sequencing data of *C. orthacantha*. Sequencing reads with unidentified nucleotides or low qualities were removed. The chloroplast genome assembling was conducted using GetOrganelle V1.7.8.1 (Jin et al. 2020). The k-mers were set as 75, 95, 115, and 127, and the maximum extension rounds (R) were set to 40. The coverage of reads across the genome was calculated using the software BWA v0.7.17 (Li and Durbin 2009). The genomic features of the *C. orthacantha* chloroplast genome were annotated with PGA and CPGVAS 2 (<http://47.96.249.172:16019/analyzer/home>) (Shi et al. 2019, Qu et al. 2019). The annotation results were manually adjusted when necessary. Using the web-based platform CPGView (<http://www.1kmpg.cn/cpgview/>), we drew the circular map and the detailed structure of transcripts in the chloroplast genome (Liu et al. 2023). To assess the phylogenetic status of *C. orthacantha*, chloroplast genome sequences of 14 *Castanopsis* species were downloaded from the NCBI GenBank database. Additional chloroplast sequences of two species in the *Castanea* genus (*C. crenata* and *C. mollissima*) were downloaded as outgroups. The sequences of all 17 species were aligned using MAFFT v7.475 (Katoh and Standley 2013), and the best nucleotide substitution model was computed using ModelFinder (Kalyaanamoorthy et al. 2017). The maximum likelihood (ML) phylogeny was

constructed using IQ-TREE v1.6.12 with 1,000 bootstraps (Nguyen et al. 2015). In addition, Bayesian inference (BI) of the chloroplast genomes was performed using MrBayes 3.2.6 (Ronquist et al. 2012). Five million Markov chain Monte Carlo generations were analyzed with a sampling frequency of 500 and 25% generations as burn-in (Figures 1 and 2).

Results

A total of 11.01 Gb of raw data was produced using next-generation sequencing, and 10.94 Gb of clean data was kept after filtering. The assembled chloroplast genome of *C. orthacantha* was 160,588 bp in length, with a GC content of 37%. The coverage depth of sequencing reads across the genome ranged from 476 to 34,167 (Figure S1). The genome was deposited into NCBI GenBank with the accession number OR900101. We found a typical quadripartite structure in the genome, and the length of LSC, SSC, and two IRs were 90,237 bp, 18,953 bp, and 25,699 bp, respectively. One hundred and thirty genes were annotated in the chloroplast genome, including 86 protein-coding genes (79 are unique), 37 tRNA genes (29 are unique), and eight rRNA genes (four are unique). Seventeen genes (*rps16*, *atpF*, *rpoC1*, *petB*, *rpl2*, *ndhB*, *ndhA*, *ndhB*, *rpl2*, *trnK-UUU*, *trnG-GCC*, *trnL-UAA*, *trnV-UAC*, *trnI*-

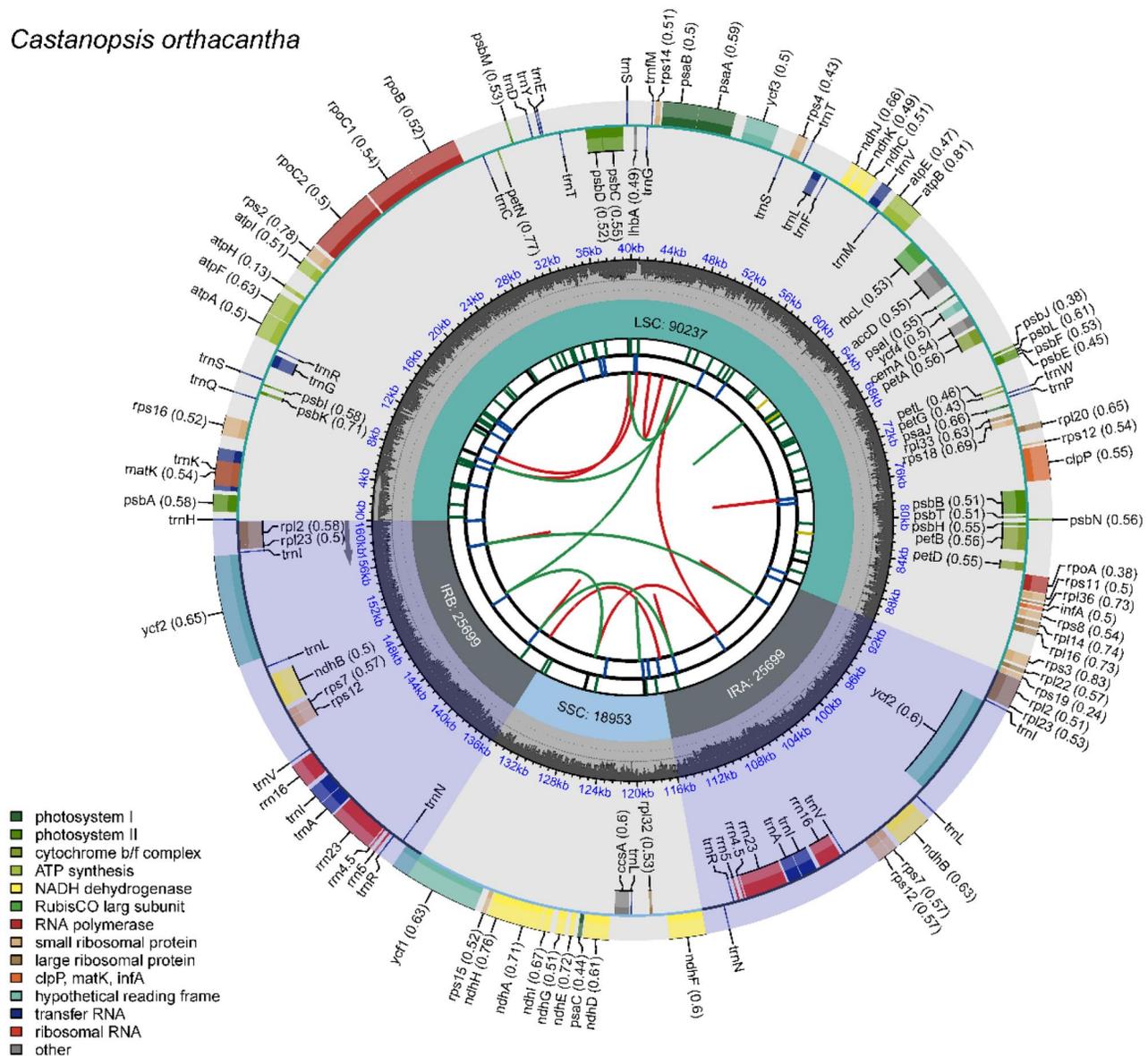
Castanopsis orthacantha

Figure 2. The circular map of the chloroplast genome of *Castanopsis orthacantha*. Genes belonging to different functional groups are plotted in the outer circle. The quadripartite structure, which consists of the LSC, the SSC, and two IR regions, is shown. The dark gray in the inner circle indicates the GC content of the chloroplast genome.

GAU, *trnA-UGC*, *trnA-UGC* and *trnI-GAU*) had one intron, and four genes (*clpP*, *ycf3*, and two copies of *rps12*) were found with two introns (Figures S2, S3). The TIM + F + R2 model was determined to be the best nucleotide substitution model based on the Bayesian information criterion. ML and BI phylogenetic analyses all suggested that *C. orthacantha*, *C. lamontii*, *C. sclerophylla*, and *C. hainanensis* formed a clade with a high support value (Figure 3).

Discussion and conclusion

In this study, we sequenced and assembled the complete chloroplast genome of *C. orthacantha* for the first time. This chloroplast genome comprised a typical quadripartite structure, like most species in the Fagaceae family (Peng et al. 2021, Zhang et al. 2021, Wang et al. 2023). A total of 130 genes were found in the chloroplast genome of *C.*

orthacantha, which was fewer than the closely related species, *C. lamontii* (132) and *C. sclerophylla* (131) (Ye et al. 2019a). It should be attributed to the missing one copy of the hypothetical gene (*ycf1*) in the chloroplast genome of *C. orthacantha*. The phylogenetic analyses based on chloroplast genomes indicated that *C. orthacantha* and three other species formed a clade. However, some branches of the ML phylogenetic tree showed low support values. Using only chloroplast genomes might not be enough to fully resolve the evolutionary relationships within the *Castanopsis* genus. Previous studies also suggested that genetic introgression and hybridization happened between some closely related *Castanopsis* species (Chen et al. 2023). The synonymous substitution rate of nuclear genomes in plants has been reported to evolve two times faster than that of chloroplast genomes (Wolfe et al. 1987, Wang et al. 2024). It should give a better understanding of the phylogenetic topologies of the *Castanopsis* genus by combining the chloroplast and

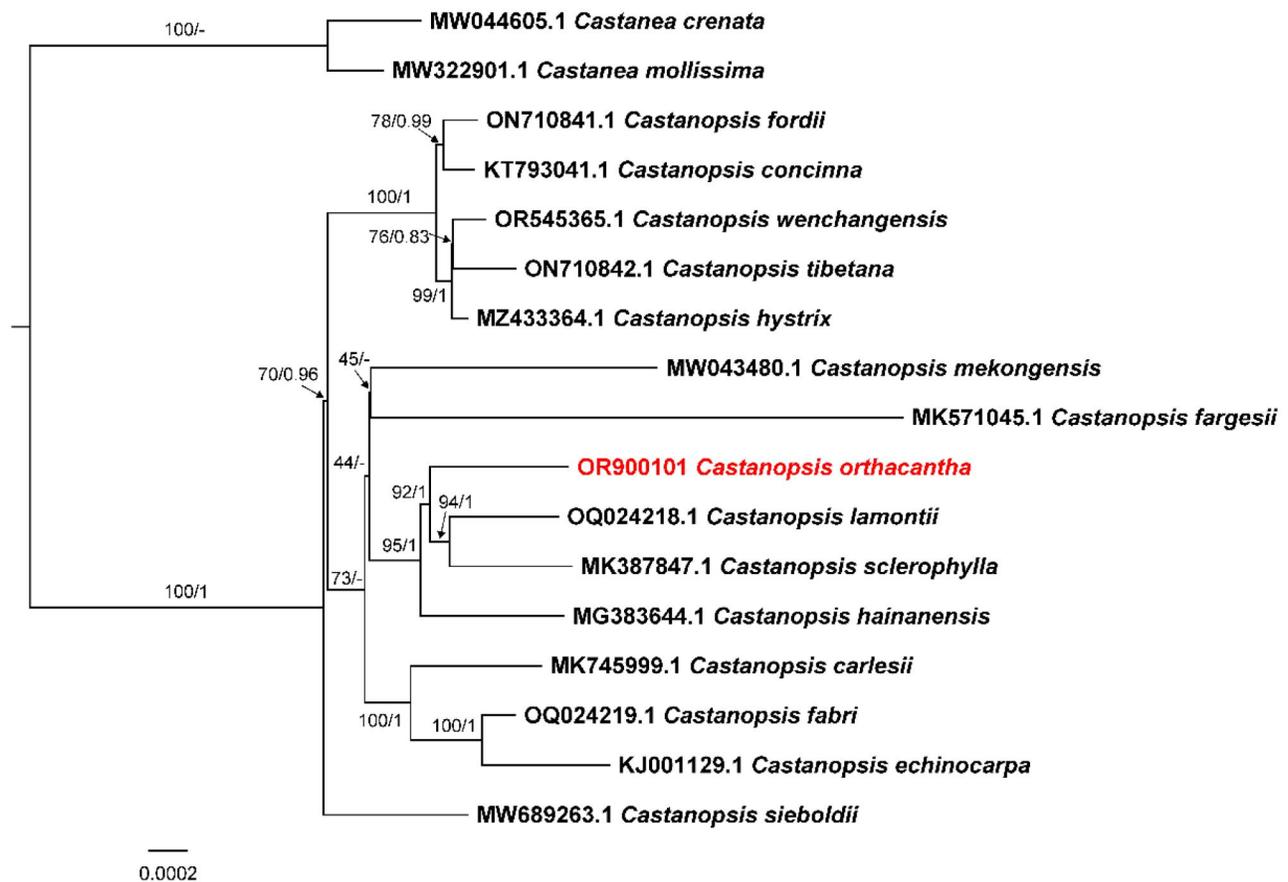


Figure 3. The phylogenetic tree of *Castanopsis orthacantha* and 16 related taxa from the family fagaceae. *Castanopsis crenata* and *Castanopsis mollissima* are used as outgroups. The scale bar represents the number of substitutions at each locus. Bootstrap value and Bayesian posterior probability are shown at each branch. The following sequences are used: MW044605 (unpublished); MW322901 (Zhang et al. 2021); ON710841 (Wang et al. 2023); KT793041 (Hinsinger and Strijk 2017); OR545365 (Chen et al. 2023); ON710842 (unpublished); MZ433364 (unpublished); MW043480 (Peng et al. 2021); MK571045 (Ye et al. 2019b); OR900101 (this study); OQ024218 (unpublished); MK387847 (Ye et al. 2019a); MG383644 (Chen et al. 2018); MK745999 (Sun et al. 2019); OQ024219 (unpublished); KJ001129 (unpublished); MW689263 (unpublished).

genomic data in the future. In conclusion, the chloroplast genome of *C. orthacantha* produced in this study could help the population genetic and phylogenetic studies of *Castanopsis* species.

Authors' contributions

Yong Gao conceived the project, and Si Yin assembled the genome and analyzed the data. Yong Gao and Si Yin drafted the paper. The authors approved the final version of this manuscript.

Disclosure statement

The authors declare no potential conflict of interest.

Ethical approval

This study includes no human, animal, or endangered plant samples. No permissions are needed during the collection of samples.

Funding

This work was supported by the Scientific Research Found Project of the Yunnan Provincial Education Commission (2024J0937), the Special Basic Cooperative Research Programs of Yunnan Provincial Undergraduate Universities (202101BA070001-011), and the Key Laboratory of Yunnan

Provincial Department of Education of the Deep-Time Evolution on Biodiversity from the Origin of the Pearl River.

Data availability statement

The chloroplast genome assembly of *C. orthacantha* was deposited into the NCBI GenBank database with accession number OR900101. The associated BioProject, Bio-Sample, and SRA numbers are PRJNA1048665, SAMN38656432, and SRR27061277, respectively.

References

- Chen X, Feng Y, Chen S, Yang K, Wen X, Sun Y. 2023. Species delimitation and genetic relationship of *Castanopsis hainanensis* and *Castanopsis wenchangensis* (Fagaceae). *Plants*. 12(20):3544. doi:10.3390/plants12203544.
- Chen X, Yang J, Yang Y, Zhang X, Du X, Zhao G. 2018. Characterization of the complete plastid genome of *Castanopsis hainanensis* Merril. *Conserv Genet Resour*. 10(4):825–828. doi:10.1007/s12686-017-0940-9.
- Hinsinger DD, Strijk JS. 2017. Complete chloroplast genome sequence of *Castanopsis concinna* (Fagaceae), a threatened species from Hong Kong and South-Eastern China. *Mitochondrial DNA Part A*. 28(1):65–66.
- Huang CJ, Chang YT, Bartholomew B. 1999. *Fagaceae*. *Flora China*. Beijing: science Press.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21(1):241. doi:10.1186/s13059-020-02154-5.

- Kalyaanamoorthy S, Minh BQ, Wong T, Haeseler AV, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589. doi:10.1038/nmeth.4285.
- Katoh K, Standley D. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780. doi:10.1093/molbev/mst010.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14):1754–1760. doi:10.1093/bioinformatics/btp324.
- Liu MQ, Zhou ZK. 2006. Modern and Geological distribution of *Castanopsis* (Fagaceae). *Acta Bot Yunnanica*. 28(3):223–235.
- Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour*. 23(3):694–704. doi:10.1111/1755-0998.13729.
- 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. doi:10.1093/molbev/msu300.
- Peng J, Yang L, Duan A, Wang D, Li S. 2021. The complete chloroplast genome of *Castanopsis mekongensis* A. Camus (Fagaceae). *Mitochondrial DNA B Resour*. 6(1):168–170. doi:10.1080/23802359.2020.1860707.
- 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):50. doi:10.1186/s13007-019-0435-7.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542. doi:10.1093/sysbio/sys029.
- 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. doi:10.1093/nar/gkz345.
- Sun R-X, Ye X-M, Wang Z-L, Lin X-F. 2019. The complete chloroplast genome of *Castanopsis carlesii* (Hemsl.) Hay. *Mitochondrial DNA B Resour*. 4(2):2591–2592. doi:10.1080/23802359.2019.1641437.
- Wang J, Kan S, Liao X, Zhou J, Tembrock LR, Daniell H, Jin S, Wu Z. 2024. Plant organellar genomes: much done, much more to do. *Trends Plant Sci*. 13(23):S1360–S1385. doi:10.1016/j.tplants.2023.12.014.
- Wang YF, Huang YL, He RJ, Yang BY, Liu ZB. 2023. The complete chloroplast genome sequence of *Castanopsis fordii* Hance (Fagaceae). *Mitochondrial DNA B Resour*. 8(2):236–239. doi:10.1080/23802359.2023.2167477.
- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84(24):9054–9058. doi:10.1073/pnas.84.24.9054.
- Ye X, Hu D, Guo Y, Sun R. 2019a. Complete chloroplast genome of *Castanopsis sclerophylla* (Lindl.) Schott: genome structure and comparative and phylogenetic analysis. *PLoS One*. 14(7):e0212325. doi:10.1371/journal.pone.0212325.
- Ye XM, Guo YP, Lei XG, Sun RX. 2019b. The complete chloroplast genome of *Castanopsis Fargesii* Franch. (Fagaceae). *Mitochondrial DNA Part B*. 4(1):1656–1657. doi:10.1080/23802359.2019.1605850.
- Zhang S, Zhu C, Bai X, Li M, Chen Y, Zhao Y. 2021. The complete chloroplast genome of *Castanea mollissima* 'Chuizhili. *Mitochondrial DNA B Resour*. 6(3):1160–1161. doi:10.1080/23802359.2021.1903360.