

The complete chloroplast genome of *Camellia confuse* Craib 1914, an economically valuable oil crop

Si Wu^{a,b} , Mei Ying Yang^a, Meng Long Fan^a, Ying Zhang^a, Xin Lei Li^a, Heng Fu Yin^a and Ji Yuan Li^a

^aResearch Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, China; ^bNanjing Forestry University, Nanjing, P R of China

ABSTRACT

Camellia confuse Craib 1914 is an industrially valuable oil crop from southern China for which little genetic information is available. Here, we found that its complete chloroplast genome is a circular sequence (156,905 bp) with a large single-copy region (LSC) of 67,724 bp, a small single copy region (SSC) of 18,400 bp, and two inverted repeats (IRs). In total, 130 genes were identified, including 86 protein-coding genes, 36 transfer RNAs, and 8 rRNA genes. Phylogenetic analysis showed that *C. confusa* is close to *C. meiocarpa*. These results provide valuable information for accelerating research on the evolution of camellias.

ARTICLE HISTORY

Received 26 September 2021
Accepted 4 June 2022

KEYWORDS

Camellia confusa;
chloroplast genome;
phylogeny

Camellia confuse Craib 1914, best known for its beautifully shaped flowers and industrially valuable oil, is a woody plant that is mainly cultivated in the Yunnan and Guizhou provinces of China, as well as in Laos, Thailand, and Vietnam (Liu et al. 2018). It belongs to the ancient genus *Camellia*, which contains about 200 species (Fan et al. 2021). Nonetheless, little genetic information for *C. confusa* is available. The chloroplast is an indispensable organelle for photosynthetic organisms and an ideal model for taxonomic classification because of its small size and conserved structure. Here, we describe the complete chloroplast genome of *C. confusa* and compare it with those of its relatives.

Leaves of *C. confusa* were sampled from the Research Institute of Subtropical Forestry, Chinese Academy of Forestry (RISF), Hangzhou, China (119°95'E, 30°07'N). All collections were approved by the head of RISF. All samples were collected by SiWu in May 10, 2021. The specimens were stored in the laboratory at the RISF, and the voucher number was YL914711 (XinLei Li, lixinlei2020@163.com). Total DNA was extracted using the MiniBest Plant Genomic DNA Extraction Kit (Takara, Dalian, China), and sequenced on the Illumina HiSeq 4000 platform (Illumina, San Diego, California, USA) at Genesky Biotechnologies (Shanghai, China). We obtained a total of 25,129,048 reads, and 23,938,766 clean reads remained after quality control with Trimmomatic (Bolger et al. 2014). SPAdes version 3.10.1 (Bankevich et al. 2012) was used to assemble the chloroplast genome, and Velvet Optimizer version 2.2.5 was used to maximize the splicing results. We used CpGAVAS2 (Shi et al. 2019) to annotate the

final genome. BLASTp at NCBI was used to confirm the annotation of the protein-coding sequences. Information on the chloroplast genome was uploaded to NCBI (MW034673).

The chloroplast genome of *C. confusa* was 156,905 bp in length with a GC content of 37.31% and showed the typical quadripartite structure, with a large single-copy (LSC) region of 67,724 bp, a small single copy (SSC) region of 18,400 bp, and two inverted repeat (IR) regions of 24,662 bp. Overall, it contained 86 protein-coding genes, 36 transfer RNAs, and 8 ribosomal RNAs. Among these genes, 15 (atpF, ndhA, ndhB, petB, petD, rpl16, rpl2, rpoC1, rps12, rps16, trnA-UGC, trnG-GCC, trnI-GAU, trnK-UUU, trnL-UAA, and trnV-UAC) contained one intron, and two genes (clpP and ycf3) contained two introns.

To investigate the evolutionary relationships of *C. confusa* with other members of the genus, we performed phylogenetic analysis of thirteen *Camellia* species. MEGA v7.0.14 was used to construct the phylogenetic tree by the maximum likelihood method (Kumar et al. 2016). The result revealed that *C. confusa* was closely related to *C. meiocarpa* (Figure 1).

Author contributions

SiWu drafted the manuscript; Meiying Yang and Menglong Fan performed the data analysis; Ying Zhang and Xinlei Li designed this study and revised the manuscript critically for intellectual content; Hengfu Yin and Jiyuan Li carried out literature search, data acquisition and manuscript editing; all authors contributed to the final approval of the version to be

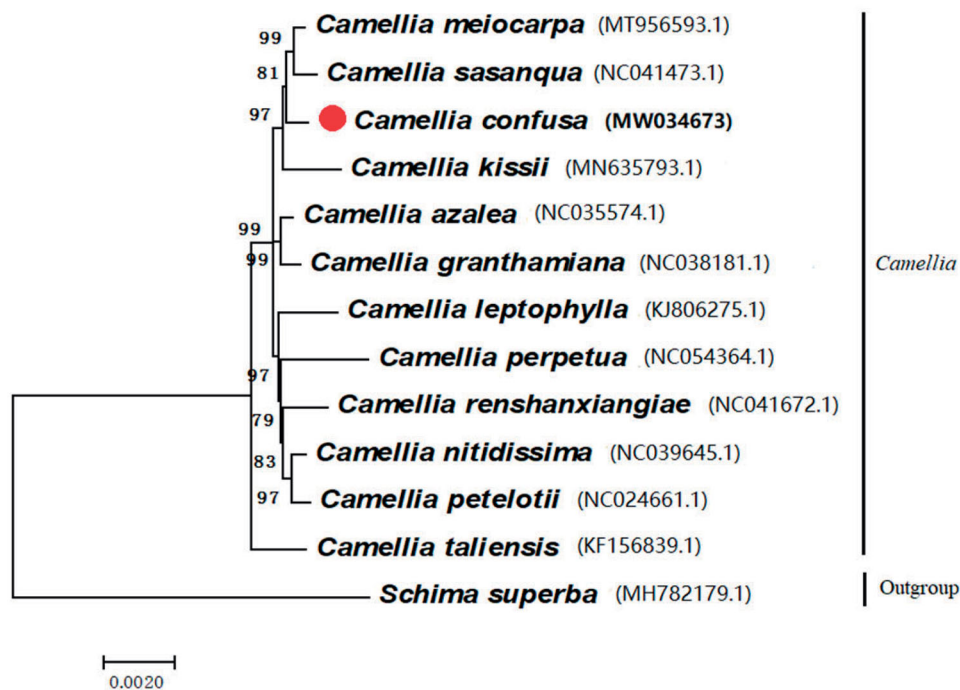


Figure 1. Maximum likelihood phylogenetic tree based on protein-coding genes of *C. confusa* and other *Camellia* species, with *Schima superba* as an outgroup.

published and all agree to be accountable for all aspects of the work.

under the accession No.MW034673. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA745398, SRR15097284, and SAMN20165932 respectively.

Disclosure statement

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

Funding

This work was supported by Zhejiang Basic and Public Welfare Research Program [LGN20C160006] and National Key R&D Program of China [2020YFD1000500].

ORCID

Si Wu  <http://orcid.org/0000-0001-7529-1373>

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at [<https://www.ncbi.nlm.nih.gov>]

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120.
- Fan ML, Yang K, Zhou R, Liu QH, Guo X, Sun YK. 2021. Temporal transcriptome profiling reveals candidate genes involved in cold acclimation of *Camellia japonica* (Naidong). *Plant Physiol Biochem.* 167:795–805.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.
- Liu C, Longsheng C, Wei T, Shaofeng P, Meiqun L, Nan D, Yongzhong C. 2018. Predicting potential distribution and evaluating suitable soil condition of oil tea *Camellia* in China. *Forests.* 9(8):487.
- 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.