



Complete chloroplast genome sequence and phylogenetic analysis of *Camellia fraterna*

Bo Yu, Ying-bo Sun , Xiao-fei Liu , Li-li Huang, Ye-chun Xu and Chao-yi Zhao

Environmental Horticulture Institute, Guangdong Academy of Agricultural Sciences, Guangdong Key Lab of Ornamental Plant Germplasm Innovation and Utilization, Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture, Guangzhou, China

ABSTRACT

Camellia fraterna belongs to the genus *Camellia* in the family Theaceae. We sequenced and analyzed the complete chloroplast genome of *C. fraterna* by Illumina sequencing in this study. The full length of the complete chloroplast genome is 156,902 bp, containing a pair of inverted repeat regions of 26,030 bp (IRa and IRb) separated by a large single-copy (LSC) region of 86,583 bp and a small single-copy (SSC) region of 18,259 bp. The *C. fraterna* chloroplast genome encodes 135 genes, comprising 87 protein-coding genes, 37 tRNA genes, eight rRNA genes, and three pseudogenes. This study will be useful for further study on genetic diversity and molecular breeding.

ARTICLE HISTORY

Received 13 August 2020
Accepted 14 October 2020

KEYWORDS

Chloroplast genome;
Camellia fraterna;
phylogenetic analysis

Camellia fraterna was planted in the Environmental Horticulture Research Institute of the Guangdong Academy of Agricultural Sciences (N23°23', E113°23', Guangzhou, China) (no. EHRIGAASC002); evergreen shrubs or small trees. The flower is small, white, and single-petaled with a strong aroma. It can be used for garden or potted ornamental viewing. It is also a precious breeding parent material for cultivating aromatic camellia varieties.

The chloroplast genome DNA of *C. fraterna* was extracted from young leaves. Covaris M220 (Covaris, Woburn, MA) was used for breaking the DNA into 300 bp fragments, and we constructed shotgun sequencing libraries according to the TruSeq™ DNA Sample Prep Kit for Illumina. Whole genome sequencing was executed using the Illumina NovaSeq platform (Illumina, San Diego, CA) (Genepioneer Biotechnologies Co. Ltd., Nanjing, China). Pair-end Illumina raw reads were cleaned from adaptors and barcodes and then quality filtered using Trimmomatic (Bolger et al. 2014). Then, reads were mapped to the chloroplast genome of the reference species (GenBank accession number: NC_024663), using Bowtie2 v2.2.4 (Langmead and Salzberg 2012) to exclude reads of nuclear and mitochondrial origins. *De novo* assembly to reconstruct the chloroplast genomes using SPAdes 3.6.1 (Bankevich et al. 2012), and chloroplast contigs were concatenated into larger contigs using Sequencher 5.3.2 (Gene Codes Inc., Ann Arbor, MI). A 'genome walking' technique, using the Unix 'grep' function, was used to find reads that could fill any gaps between contigs that did not assemble in the initial set of analyses (Souza et al. 2019). Misassembled

contigs were corrected by Jellyfish v.2.2.3 (Marcais and Kingsford 2011). Annotation of the chloroplast genomes was generated by CpGAVAS (Liu et al. 2012) and a circular representation was drawn using the online tool OGDRAW (Lohse et al. 2007). The complete chloroplast genome sequence has been submitted to GenBank with the accession number of MT663342.

The length of chloroplast genome sequence of *C. fraterna* is 156,902 bp, including two inverted repeat regions (IRa and IRb, each 26,030 bp) separated by an large single-copy (LSC) (86,583 bp) region and an small single-copy (SSC) (18,259 bp) region. The GC content of the overall chloroplast genome, IR regions, LSC, and SSC is 37.34, 43.02, 35.33, and 30.58%, respectively. The GC content of the two IR regions is higher than those of the SSC and LSC, which is similar with *Celosia cristata* (Liu et al. 2020), *Spathiphyllum 'Parrish'* (Liu et al. 2019), and *Spathiphyllum cannifolium* (Liu et al. 2019). The chloroplast genome contains 135 genes in total, including 87 protein-coding genes, 37 tRNAs, eight rRNAs, and three pseudogenes.

The whole genome was used for phylogenetic tree analysis. First, we use MAFF v7.427 (Kazutaka et al. 2005) – auto mode to align each sequence. The gaps in the alignment were removed using the program trimAl with '-nogaps' v 1.4 (Capella-Gutierrez et al. 2009). Finally, MrBayes v3.2.7 (Fredrik et al. 2012) was used to construct the phylogenetic tree (Figure 1). We found that *C. fraterna* is closely related to *C. reticulata*.

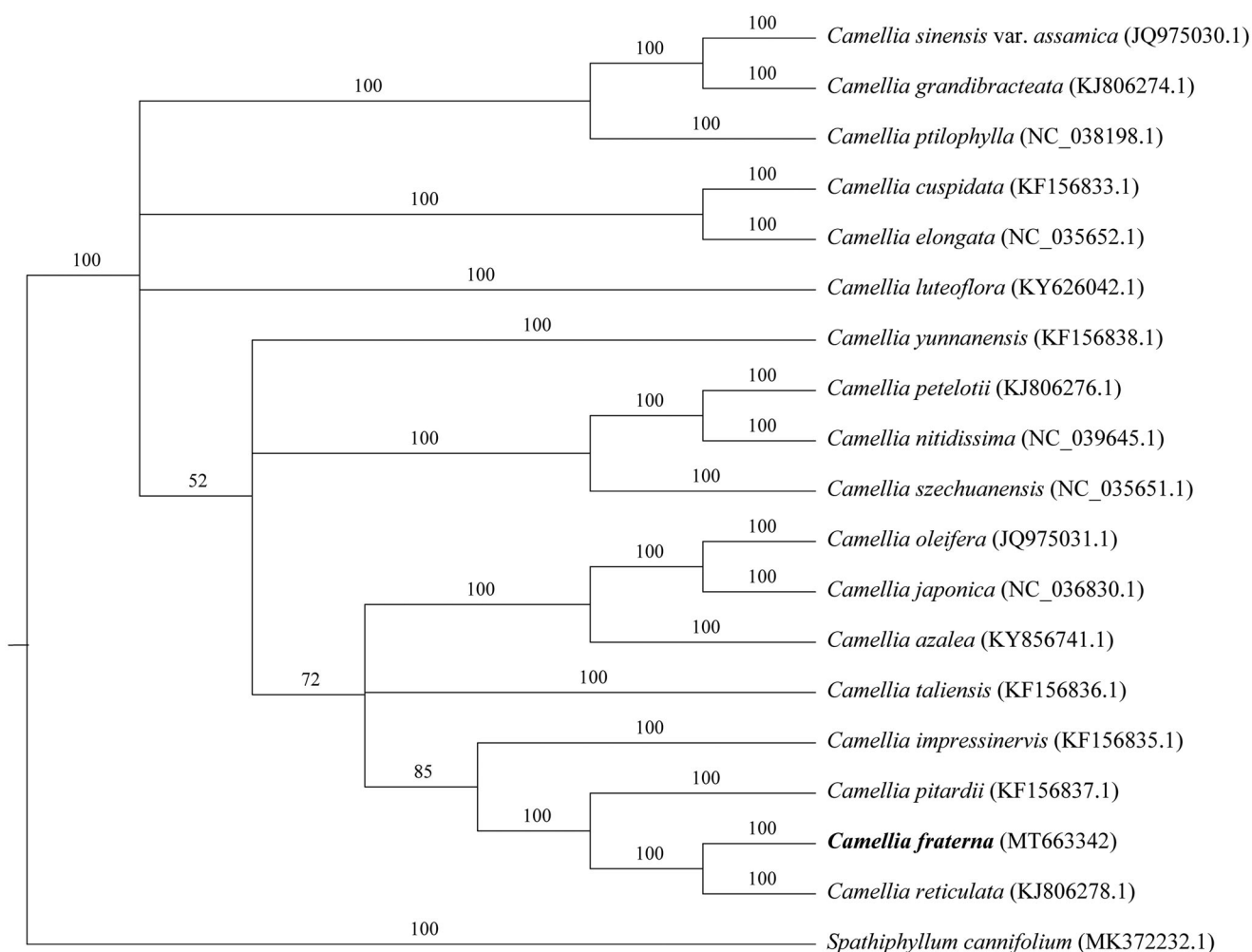


Figure 1. Phylogenetic tree reconstruction of 19 species based on sequences from whole chloroplast genomes. All the sequences were downloaded from NCBI GenBank.

Disclosure statement

The authors declare no conflict of interest.

Funding

This study is funded by the Key-Area Research and Development Program of Guangdong Province (2020B020220005, 2018B020202002), Guangzhou Science and Technology Project (201807010016, 201604020031), Guangdong provincial Science and Technology Department (2018A050506054, 2016LM3169, 2014B070706016).

ORCID

Ying-bo Sun  <http://orcid.org/0000-0002-3577-559X>

Xiao-fei Liu  <http://orcid.org/0000-0003-2955-1538>

Data availability statement

The data that newly obtained at this study are available in the NCBI under accession number of MT663342 (<https://www.ncbi.nlm.nih.gov/nuccore/MT663342>).

Author Contributions

BY: Performed the experiments investigation, project administration, writing the original draft and data curation. LH YX CZ: Prepared the

resources. YS XL: Supervised the project and made revisions to the manuscript.

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.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973.
- Fredrik R, Maxim T, Paul VDM, Daniel LA, Aaron D, Sebastian H, Bret L, Liang L, Marc AS, John PH. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539–542.
- Kazutaka K, Kei-Ichi K, Hiroyuki T, Takashi M. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33:511–518.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and Genbank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* 13:715.

- Liu XF, Ye YJ, Liu JM, Yu B, Xu YC. 2020. Complete chloroplast genome sequence and phylogenetic analysis of *Celosia cristata* 'Xianguang'. *Mitochondrial DNA B*. 5(2):1338–1339.
- Liu XF, Zhu GF, Li DM, Wang XJ. 2019. Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum* 'Parrish'. *PLoS One*. 14(10):e0224038.
- Liu XF, Zhu GF, Li DM, Wang XJ. 2019. The complete chloroplast genome sequence of *Spathiphyllum cannifolium*. *Mitochondrial DNA B*. 4(1): 1822–1823.
- Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 52(5–6):267–274.
- Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 27(6):764–770.
- Souza UJBd, Nunes R, Targueta CP, Diniz-Filho JAF, Telles MPC. 2019. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae – Caesalpinioideae): comparative analysis with related Mimosoid species. *Sci Rep*. 9(1):14206.