

The complete chloroplast genome sequence of *Camellia chuongtsoensis*

Bo Yu, Ying-Bo Sun, Li-Li Huang, Ye-Chun Xu, Chao-Yi Zhao and Xiao-Fei Liu 

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, Guangdong, China

ABSTRACT

Camellia chuongtsoensis is an evergreen shrub with a single-petaled flower and golden yellow color. The complete chloroplast genome of *C. chuongtsoensis* was sequenced and analyzed in this study by Illumina sequencing. The chloroplast genome is 156,504 bp in length with a quadripartite structure containing a large single copy (LSC) region of 86,215 bp, a small single copy (SSC) region of 18,253 bp, and a pair of inverted repeat regions of 26,018 bp (IRa and IRb). The chloroplast genome of *C. chuongtsoensis* encodes 135 genes, comprising 87 protein-coding genes, 37 tRNA genes, 8 rRNA genes, and 3 pseudogenes.

ARTICLE HISTORY

Received 13 August 2020
Accepted 29 October 2020

KEYWORDS

Chloroplast genome;
Illumina sequencing;
Camellia chuongtsoensis

Camellia chuongtsoensis, a newly discovered species, set out bright yellow flowers (Li et al. 2018), which belongs to the genus *Camellia* in the family Theaceae. *C. chuongtsoensis* was planted in the Environmental Horticulture Research Institute of the Guangdong Academy of Agricultural Sciences (N23°23', E113°23', Guangzhou, China) (No: EHRGAASC001). *Camellia chuongtsoensis* had a unique and long blooming period across the summer season (Li et al. 2018), therefore, it has great horticultural application value and is a good parent source in breeding of *Camellia*.

The chloroplast genome DNA of *C. chuongtsoensis* was extracted from young leaves. Using Covaris M220 (Covaris, Woburn, MA) for breaking the DNA into 300 bp fragments, we constructed shotgun sequencing libraries according to the TruSeq™ DNA Sample Prep Kit for Illumina. Finally, whole genome sequencing was executed using the Illumina NovaSeq platform (Illumina, USA) (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). Bowtie2 v2.2.4 (Langmead and Salzberg 2012) was used to exclude reads of nuclear and mitochondrial origins. Using SPAdes 3.6.1 (Bankevich et al. 2012) to reconstruct the chloroplast genomes by *de novo* assembly, and chloroplast contigs were concatenated into larger contigs using Sequencher 5.3.2 (Gene Codes Inc., Ann Arbor, MI). A “genome walking” technique was used to remove gaps (Souza et al. 2019). Misassembled contigs were corrected by Jellyfish v2.2.3 (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 MT663341.

The complete chloroplast genome sequence of *C. chuongtsoensis* is 156,504 bp in length, including two inverted repeat regions (IRa and IRb, each 26,018 bp) separated by a large single copy (LSC) (86,215 bp) region and a small single copy (SSC) (18,253 bp) region. The AT content of the overall chloroplast genome, IR regions, LSC, and SSC are 62.65, 57.01, 64.61, and 69.40%, respectively. The AT content of the two IR regions is lower than those of the SSC and LSC, which is very common in other plants; this phenomenon is mostly attributable to tRNA and rRNA genes (Liu et al. 2019, 2020). The chloroplast genome contains 135 genes in total, including 87 protein-coding genes, 37 tRNAs, 8 rRNAs, and 3 pseudogenes.

The whole genome was used for phylogenetic tree analysis. MAFF v7.427 (Kato et al. 2005) auto mode was used to align each sequence, and the gaps in the alignment were removed using the program trimAl with ‘-nogaps’ v 1.4 (Capella-Gutierrez et al. 2009). MrBayes v3.2.7 (Fredrik et al. 2012) was used to construct the phylogenetic tree (Figure 1). This study will be useful for further analysis of genetic diversity in *Camellia*.

Author contributions

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

Prepared the resources: YS, LH, YX, CZ.

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

CONTACT Xiao-Fei Liu  liuxiaofei6996@126.com  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, No. 1, Jinying east street 1, Wushan road, Guangzhou 510640, China.

© 2021 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 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

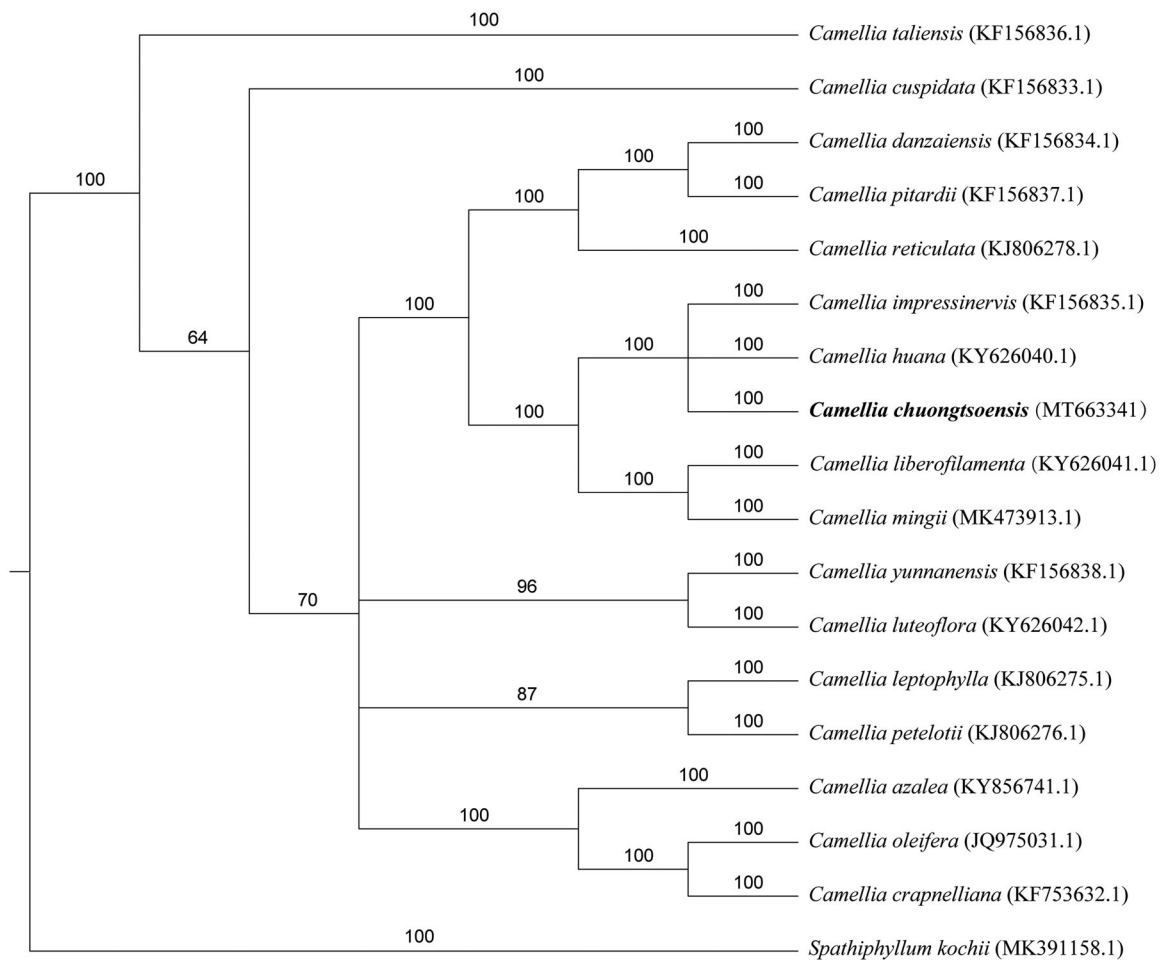


Figure 1. Phylogenetic tree reconstruction of 18 species based on sequences from whole chloroplast genomes. All the sequences were downloaded from NCBI Genbank.

Disclosure statement

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

Funding

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

ORCID

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

Data availability statement

Data were newly obtained in this study are available in the NCBI under accession number of MT663341 (<https://www.ncbi.nlm.nih.gov/nucleotide/MT663341>).

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.

Katoh 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.

Li X, Fan Z, Guo H, Ye N, Lyu T, Yang W, Wang J, Wang JT, Wu B, Li J, et al. 2018. Comparative genomics analysis reveals gene family expansion and changes of expression patterns associated with natural adaptations of flowering time and secondary metabolism in yellow *Camellia*. *Funct Integr Genomics.* 18(6):659–671.

Liu XF, Ye YJ, Liu JM, Yu B, Xu YC. 2020. Complete chloroplast genome sequence and phylogenetic analysis of *Celosia cristata* 'Xiaguang'. *Mitochondrial DNA Part 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 canifolium*. *Mitochondrial DNA B.* 4 (1): 1822–1823.

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.
- 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.
- Marçais 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.