



Characterization of the complete chloroplast genome of *Camellia brevistyla*, an oil-rich and evergreen shrub

Yupeng Wang^{a,b,c} , Jiyuan Li^{b,c}, Zhengqi Fan^{b,c}, Dongyang Wu^a , Hengfu Yin^{b,c} and Xinlei Li^{b,c}

^aCollege of Information Science and Technology, Nanjing Forestry University, Nanjing, China; ^bState Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical, Forestry, Chinese Academy of Forestry, Hangzhou, China; ^cKey Laboratory of Forest Genetics and Breeding, Research Institute of Subtropical, Forestry, Chinese Academy of Forestry, Hangzhou, China

ABSTRACT

Camellia brevistyla is an economic species for its seeds with high oil content and ornamental value, which is cultivated widely across southern China. In this study, the complete chloroplast (cp) genome sequence of *C. brevistyla* was assembled and annotated in order to future genetic research. The whole cp genome of *C. brevistyla* is 159,281 bp in size, composed of a small single copy (SSC) region of 15,662 bp and a large single copy (LSC) region of 86,251 bp separated by a pair of inverted repeats (IRs, IRA: 130598: 159281, IRB: 86252: 114935). The overall GC content of *C. brevistyla* cp genome is 37.19%, with the base content A (31.03%), T (31.78%), C (18.94%), and G (18.25%). Phylogenetic analysis of 20 species based on 74 protein-coding genes shows that *C. brevistyla* is evolutionarily closest to *Camellia danzaiensis*.

ARTICLE HISTORY

Received 26 November 2019
Accepted 7 December 2019

KEYWORDS

Camellia brevistyla;
chloroplast genome;
phylogenetic analysis

Camellia brevistyla, one of more than 250 species of the genus *Camellia*, grows mainly in thickets or pine forests on hillsides at an elevation of about 600 m and is found commonly in southeast China. *Camellia brevistyla* is listed as an economic tree species for oil production and horticulture (Su et al. 2017). Meanwhile, terpene synthase genes could be isolated from *C. brevistyla*, which was identified as hedycaryol synthase; it was shown to specifically expressed in flowers and has certain biological application value (Hattan et al. 2016). However, there are few studies on the genomic information and genetic diversity of *C. brevistyla*. Here, we report the first complete cp genome of *C. brevistyla* (Genbank accession: MN640791) and discuss the phylogenetic relationship among its populations, which will provide important reference for its future biological research.

The samples of fresh leaves of *C. brevistyla* were collected from Jinhua International Camellia Species Park (Zhejiang, China; Coordinates: 29°7'10.1208"N, 119°35'52.1088"E). Voucher specimen (CD_01) was deposited in State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry. *C. brevistyla* was sequenced using Illumina HiSeq 2500 sequencing systems (Illumina, USA) at Genesky Biotechnologies (Shanghai, China). The 24,696,066 raw reads were under quality control by Trimmomatic (Bolger et al. 2014). The strategy

for assembly and annotation the cp genome was adapted from Wang et al. (2018). Finally, the cp genome map of *C. brevistyla* was generated by OGDRAW version 1.3.1 (Greiner et al. 2019).

The complete cp genome sequence of *C. brevistyla* is 159,281 bp with the typical quadripartite structure (Wang et al. 2016), including a SSC region of 15,662 bp, a LSC region of 86,251 bp and a pair of IRs region of 28,684 bp. The total GC content of sequence was 37.19%, and the GC content of IR regions was higher than that of SSC region (30.70%) and SSC region (35.33%). The cp genome of *C. brevistyla* has 131 functional genes, which are composed of 83 protein-coding genes and 48 RNA genes (44 tRNA genes and 4 rRNA genes). The longest genes of protein-coding genes, tRNAs and rRNAs is *ycf2_2* with 6867 bp, *rrn23* with 2809 bp and tRNA-Ile3 with 239 bp, respectively.

A Minimum-Evolution tree based on 74 protein-coding genes that were screened by the Perl script (Wang et al. 2019) was reconstructed by MEGA7 (Kumar et al. 2016) to study the phylogenetic relationship of 20 *Camellia* species. The evolutionary distances were computed using the JTT matrix-based method (Jones et al. 1992) and there were a total of 22,648 positions in the final dataset. As illustrated in Figure 1, the cp genome of *C. brevistyla* is closest to *Camellia danzaiensis* (NC_022460.1) and also clustered closely to *Camellia pitardii* (NC_022462.1).

CONTACT Xinlei Li  lixinlei2020@163.com  State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical, Forestry, Chinese Academy of Forestry, Hangzhou, Zhejiang, 311400, China; Dongyang Wu  eassun2000@sina.com  College of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China

© 2020 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.

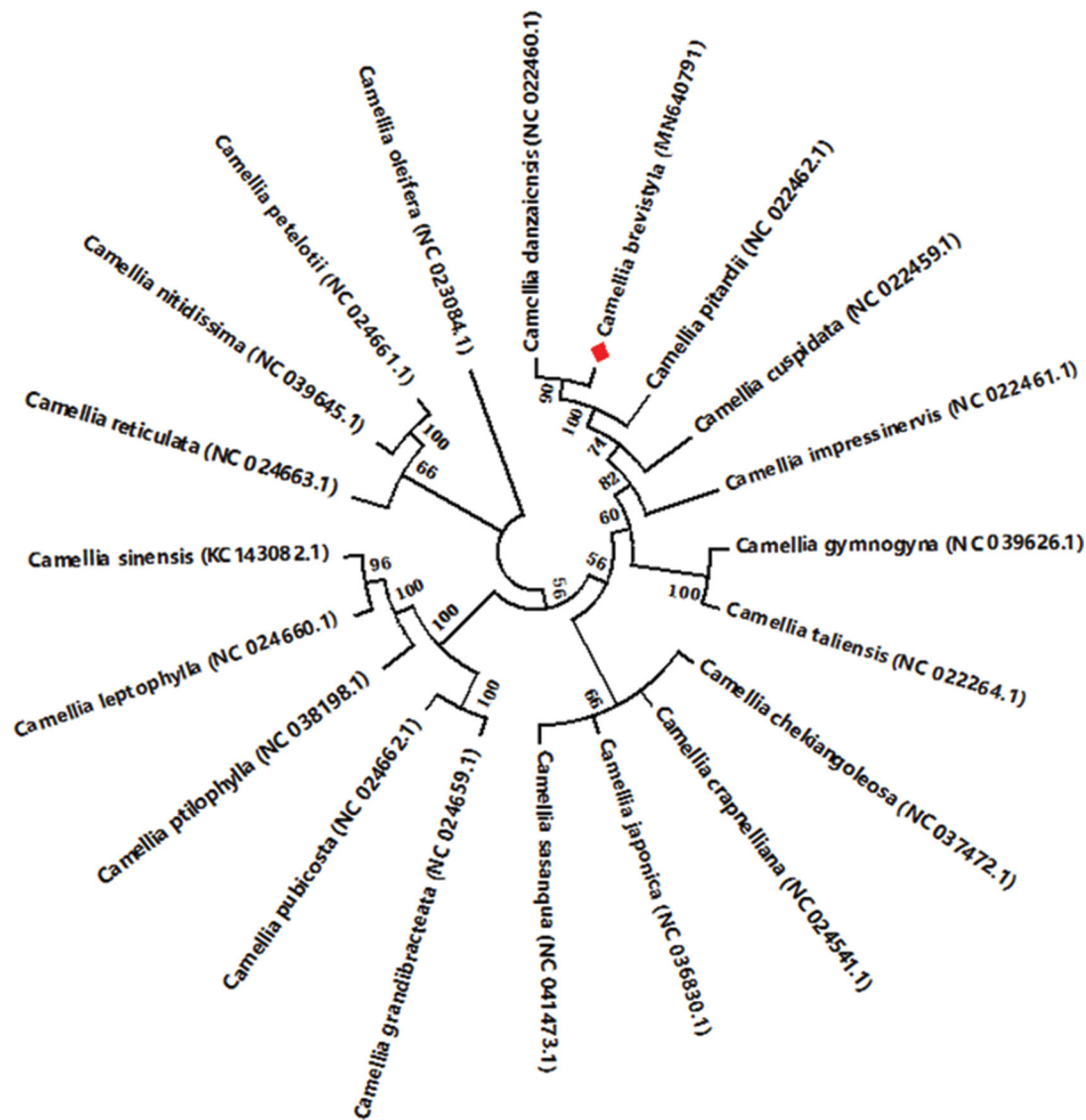


Figure 1. The Minimum-Evolution (ME) tree of 20 *Camellia* cp genomes based on 74 protein-coding genes were conducted in MEGA7. The bootstrap values from 1000 replicates are listed for each node.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported by Nonprofit Research Projects [CAFYBB2017MB007] of Chinese Academy of Forestry.

ORCID

Yupeng Wang  <http://orcid.org/0000-0003-4170-3952>

Dongyang Wu  <http://orcid.org/0000-0002-2570-6494>

References

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120.
- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. 47(W1):W59–W64.
- Hattan J-I, Shindo K, Ito T, Shibuya Y, Watanabe A, Tagaki C, Ohno F, Sasaki T, Ishii J, Kondo A, et al. 2016. Identification of a novel hedycaryol synthase gene isolated from *Camellia brevistyla* flowers and floral scent of *Camellia* cultivars. *Planta*. 243(4):959–972.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Appl Biosci*. 8: 275–282.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33(7):1870.
- Su MH, Hsu TH, Wang CN, Lin KH, Chiang MC, Kang RD, Nguyen KM, Nguyen H. 2017. Genetic diversity of a novel oil crop, *Camellia brevistyla*, revealed by ISSR DNA markers. *Hortic Sci Technol*. 35(5):588–598.
- Wang X, Bi C, Wang C, Ye Q, Yin T, Ye N. 2019. Genome-wide identification and characterization of WUSCHEL-related homeobox (WOX) genes in *Salix suchowensis*. *J for Res*. 30(5):1811–1822.
- Wang X, Bi C, Xu Y, Wei S, Dai X, Yin T, Ye N. 2016. The whole genome assembly and comparative genomic research of *Thellungiella parvula* (Extremophile Crucifer) mitochondrion. *Int J Genomics*. 2016:1–13.
- Wang X, Cheng F, Rohlsen D, Bi C, Wang C, Xu Y, Wei S, Ye Q, Yin T, Ye N, et al. 2018. Organellar genome assembly methods and comparative analysis of horticultural plants. *Hortic Res*. 5(1):3.