

## Characterization of the complete chloroplast genome sequence of *Camellia oleifera* in Hainan, China

Wan Zhang<sup>a</sup>, Yunlin Zhao<sup>a</sup>, Guiyan Yang<sup>b</sup>, Yongcheng Tang<sup>a</sup> and Zhenggang Xu<sup>a</sup>

<sup>a</sup>Hunan Research Center of Engineering Technology for Utilization of Environmental and Resources Plant, Central South University of Forestry and Technology, Changsha, China; <sup>b</sup>College of Forestry, Northwest A&F University, Yangling, China

### ABSTRACT

*Camellia oleifera*, an evergreen small tree or shrub with high medicinal and ecological values, is mainly distributed in subtropical montane regions of China. In this study, complete chloroplast genome was reported for *Camellia oleifera* in Hainan, China. The results showed that the whole genome was 156,996 bp in length, including a pair of inverted repeats (IRs) of 26,025 bp, a large single copy (LSC) region of 86,649 bp and a small single copy (SSC) region of 18,297 bp. The genome contained a total of 115 different genes, including 81 protein-coding genes, 30 tRNA genes, and four rRNA genes. Among these genes, eight genes contain a single intron and two genes contain two introns. The total GC content of *Camellia oleifera* was 37.29%. The maximum likelihood phylogenetic analysis based on 21 chloroplast genomes showed that *Camellia oleifera* was similar to *Camellia danzaiensis*.

### ARTICLE HISTORY

Received 15 October 2017  
Accepted 17 November 2017

### KEYWORDS

*Camellia oleifera*; chloroplast genome; next generation sequencing; phylogenetic analysis



*Camellia oleifera*, an evergreen small tree or shrub of the genus *Camellia* (Theaceae), is mainly distributed in subtropical area of high mountains and hilly terrains in southern China. It is one of the four woody edible oil plants in the world, and is also a unique natural Chinese woody edible oil species (Zhuang 2008; Zhou et al. 2013). Tea oil extracted from *Camellia oleifera* seed has good health care and medicinal value (He et al. 2007). Especially, the unique climate and environmental conditions in Hainan Island gave birth to rich and distinctive oil tea resources (Zheng et al. 2016). However, the seed offspring of *Camellia oleifera* has a complex hereditary, and is easy to mutate, which still exist challenges in the interspecies relationships and phylogenetic (Yang et al. 2013; Huang et al. 2014). Therefore, it is essential to study the chloroplast genome sequence to provide a valuable genomic resource for the future genetic and phylogenetic studies about *Camellia oleifera*. Here, we presented the complete chloroplast genome of *Camellia oleifera* (GenBank accession number: MF541730) based on the Illumina paired-end sequencing data.

The seeds of *Camellia oleifera* were collected from Chengmai, Hainan Province (110.00°E, 19.75°N), and fresh leaves were obtained by seed germination. The specimen was kept in the laboratory at  $-80^{\circ}\text{C}$  under the accession number 20170519YC. Plant Chloroplast Purification Kit (BTN120308) and Column Plant DNA Extraction Kit were combined for extracting Chloroplast DNA. The library was prepared with a NEBNext<sup>®</sup> Ultra TM DNA Library Prep Kit for Illumina (NEB, Ipswich, MA). The whole-genome sequencing

was performed with the 500 bp paired-end sequencing method by the Illumina HiSeq4000 Platform. The raw data were filtered with Trimmomatic v0.32 (Bolger et al. 2014) to get Clean data for subsequent analysis. Protein-coding genes and non-coding RNAs genes were annotated with DOGMA (Wyman et al. 2004) and cpGAVAS (Liu et al. 2012). A circular map of the chloroplast genome was generated with the OGDRAW (Lohse et al. 2007) software.

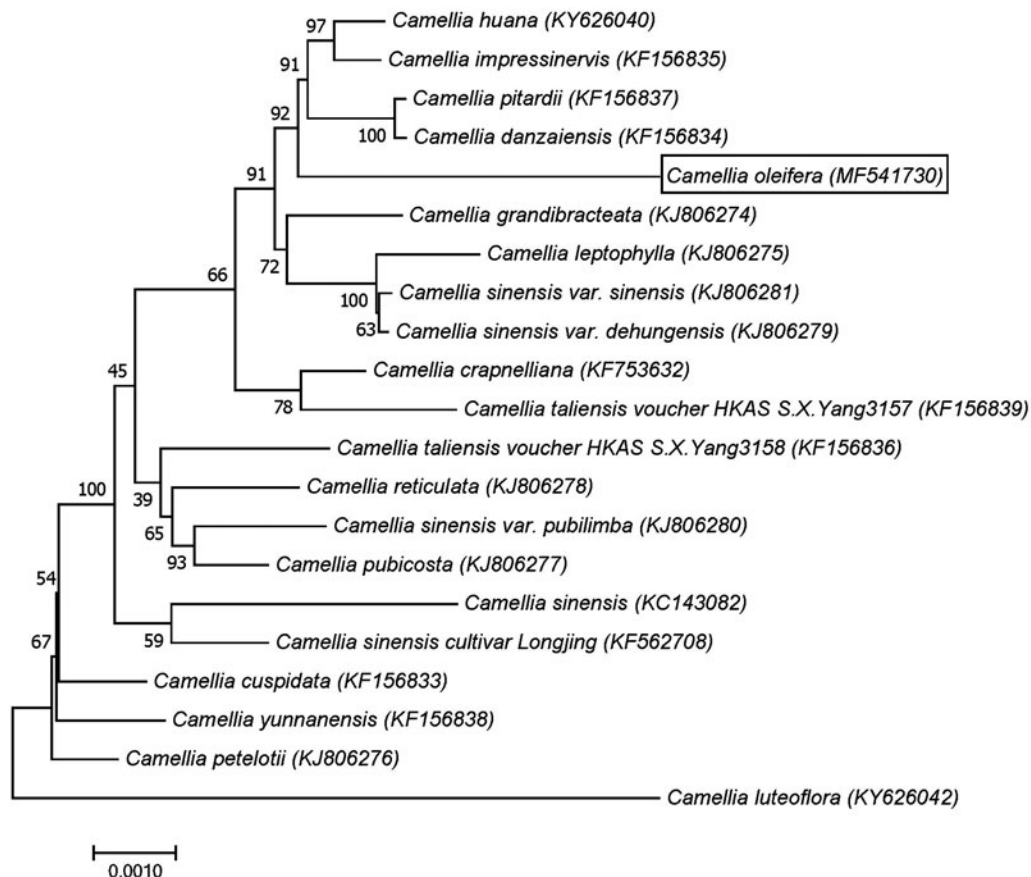
The chloroplast genome of *Camellia oleifera* is a circular molecule of 156,996 bp in length, with a small single copy (SSC) region of 18,297 bp, a large single copy (LSC) region of 86,649 bp, and a pair of inverted repeats (IRs) regions of 26,025 bp. The total GC content of chloroplast genome is 37.29%, and the corresponding values of LSC, SSC, and IRs are 35.29%, 30.55%, and 42.98%, respectively. It contains 115 different genes, including 81 PCG species, 30 tRNA genes, and four rRNA genes. Among these genes, eight genes (*atpF*, *rpoC1*, *rpl2* × 2, *ndhA*, *ndhB* × 2, *ycf1*) contain a single intron and two genes (*ycf3* and *clpP*) contain two introns.

The evolutionary analyses between *Camellia* species were conducted in MEGA7 (with 1000 bootstrap replicates) (Kumar et al. 2016). The analysis involved 21 amino acid sequences (Figure 1). The phylogenetic analysis showed that *Camellia oleifera* was similar to *Camellia danzaiensis*. The chloroplast genome reported here provided useful genomic resources not only for the exchange of information between nuclear genomes, but also for the population genetics of *Camellia oleifera* and its phylogenetic and evolutionary studies.

**CONTACT** Zhenggang Xu  [rssq198677@163.com](mailto:rssq198677@163.com)  Hunan Research Center of Engineering Technology for Utilization of Environmental and Resources Plant, Central South University of Forestry and Technology, Changsha, China

© 2017 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.** Maximum likelihood phylogenetic tree based on 21 complete chloroplast genomes from *Camellia* family.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The study was financially supported by the Key Technology R&D Program of Hunan Province (2016TP2007, 2016NK2148, 2016TP1014) and National Key Technology Research and Development Program of China (2014BAC09B03-02).

## References

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- He GX, Li SL, Xu LC, Gao H. 2007. Growth characteristic and function value of *Camellia oleifera*. *Jiangxi Forest Sci Technol*. 4:39–42.
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol*. 14:151.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33:1870–1874.
- Liu C, Shi LC, Zhu YJ, Chen HM, Zhang JH, Lin XH, Guan XJ. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom*. 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*. 52:267–274.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 20:3252–3255.
- Yang JB, Yang SX, Li HT, Yang J, Li DZ. 2013. Comparative chloroplast genomes of *camellia* species. *PLoS ONE*. 8:e73053.
- Zheng DJ, Pan XZ, Zhang DM, Xie LS, Zeng JH, Zhang ZL, Ye H. 2016. Survey and analysis on tea-oil *Camellia* resource in Hainan. *J Northw Forest Univ*. 31:130–135.
- Zhou CF, Yao XH, Lin P, Wang KL, Chang J, Mo RH. 2013. Constituent changes associated with seeds development of *Camellia oleifera* Abel. *Chin J Oil Crop Sci*. 35:680–685.
- Zhuang RL. 2008. *Camellia oleifera* of China. Beijing: China Forestry Press.