


## The complete chloroplast genome sequence of economical standard tea plant, *Camellia sinensis* L. cultivar Sangmok, in Korea

Dong-Jun Lee<sup>a</sup>, Chang-Kug Kim<sup>a</sup> , Tae-Ho Lee<sup>a</sup>, So-Jin Lee<sup>b</sup>, Doo-Gyung Moon<sup>b</sup>, Yong-Hee Kwon<sup>b</sup> and Jeong-Yong Cho<sup>c</sup>

<sup>a</sup>Genomics Division, National Institute of Agricultural Sciences(NAS), Jeonju, South Korea; <sup>b</sup>Research Institute of Climate Change and Agriculture, Jeju-do, South Korea; <sup>c</sup>Division of Food Science, Chonnam National University, Gwangju, South Korea

### ABSTRACT

The complete chloroplast genome sequence of *Camellia sinensis* L. cultivar Sangmok was determined using high-throughput sequencing technology. We sequenced Sangmok chloroplast genome and performed comparative with 21 published other *Camellia* and species from different genus for phylogenetic analysis. Chloroplast genome was 153,044 bp in length, containing a pair of 24,627 bp inverted repeat (IR) regions, which were separated by small and large single-copy regions (SSC and LSC) of 19,155 and 64,665 bp, respectively. The chloroplast genome contained 97 genes (63 protein-coding genes, 29 *tRNA* genes, and 5 *rRNA* genes). The overall GC content of the chloroplast genome was 37.2%. The phylogenetic analysis among species in number of the genus *Camellia* provided that *C. sinensis* L. cultivar Sangmok is closely related to KJ806277 *Camellia pubicosta*.

### ARTICLE HISTORY

Received 19 June 2020  
Accepted 27 June 2020

### KEYWORDS

*Camellia sinensis* L.;  
chloroplast genome;  
Sangmok; Korea

Today, tea is commercially cultivated on more than 3.80 million hectares of land on a continent-wide scale, and 5.56 million metric tons of teas worldwide were produced annually in 2014. The *Camellia sinensis* L. cultivar Sangmok, which belongs to economical standard green Tea in Korea, is an evergreen small tree or shrub. Genus *Camellia* is an economic plant (Yang et al. 2013; Huang et al. 2014) including importantly well-known camellias (e.g. *Camellia japonica*, *Camellia reticulata*, *Camellia sasanqua*, and *Camellia oleifera*) (Ming and Bartholomew 2007; Meng et al. 2018) and it provides excellent materials for becoming increasingly popular owing to special health benefits and in plant science studies (Yang et al. 2013). The chloroplast genome information has been applied in plant biology, phylogenetic inference, and species identification (Kane et al. 2012; Ruhfel et al. 2014). In this study, we assembled the complete chloroplast genomes in *C. sinensis* L. cultivar Sangmok using Pacbio Sequel II sequencing technology. We have deposited the complete chloroplast genome sequence into GenBank database (Accession Number: LC488797).

The *C. sinensis* L. cultivar Sangmok was collected from the Research Institute of Climate Change and Agriculture in Jeju island, Republic of Korea (33.28N 126.31E) and registration of PVP Right (No 4837) in Korea Seed & Variety Service (KSVS, [http://seed.go.kr/sites/seed\\_eng/index.do](http://seed.go.kr/sites/seed_eng/index.do)). We extracted genomic DNA from *C. sinensis* L. cultivar Sangmok leaves using SMRTbell Express Template Prep Kit version 2.0 (Pacific Biosciences, Menlo Park, CA) according to the manufacturer's instructions. Library construction and sequencing were

carried out using Pacbio Sequel II System at DNALINK Inc. (Seoul, South Korea). The data were produced using 152 Giga base pairs (Gbp) clean reads representing with a total of 128 Gbp of 10 kbp paired-end reads. We performed assembling of chloroplast genome by Celera Assembler 8.3rc2 and obtained 1 contigs. Chloroplast genome was confirmed by primer working method (Nadalin et al. 2012). We performed annotations from the KJ806281.1 *Camellia* species sequence.

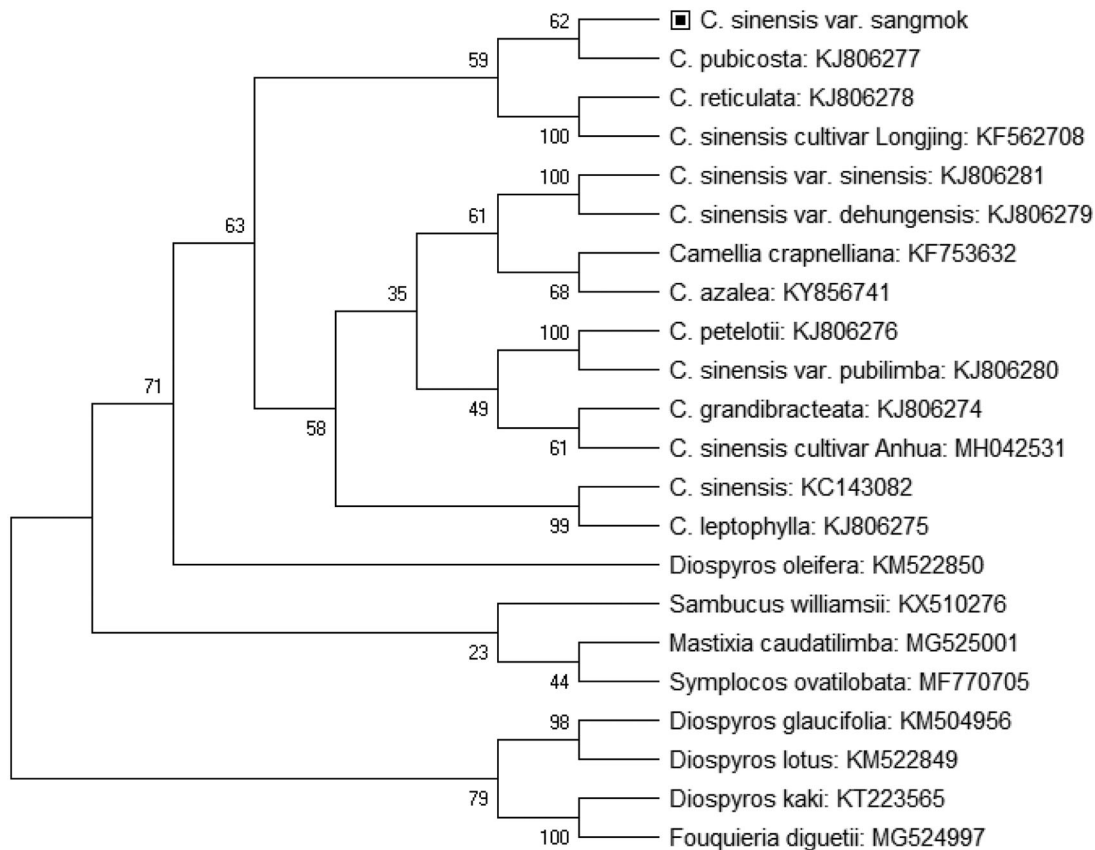
The complement chloroplast genome of *C. sinensis* L. cultivar Sangmok was 153,044 bp length of circular form including four typical regions of chloroplast such as large single copy (LSC) region of 64,665 bp and small single-copy (SSC) region of 19,155 bp, separated by a pair of inverted repeats of 24,627 bp (IRa and IRb). The chloroplast genome contained 71 protein-coding genes, 37 *tRNA* genes, and 13 *rRNA* genes. The overall GC content was 37.2% containing a pair of 24,627 bp IR regions, which were separated by SSC and LSC of 19,155 and 64,665 bp, respectively. The chloroplast phylogenetic analysis was performed using previously published chloroplast genome sequences of 22 species belonging to *Camellia* family and different genus by neighbor-joining method in MEGA version 7.0 (Kumar et al. 2016) with 1000 bootstrap value (Figure 1). According to the phylogenetic tree, *C. sinensis* L. cultivar Sangmok was closely related to KJ806277 *Camellia pubicosta*, KF562708 *Camellia sinensis* var. Longjing, and KJ806278 *Camellia reticulata*.

We expect that this sequence will help to clarify the hybrid status of *C. sinensis* L. cultivar Sangmok and provide additional genomic resources for *C. sinensis* studies.

**CONTACT** Dong-Jun Lee  [leemos1004@gmail.com](mailto:leemos1004@gmail.com)  Genomics Division, National Institute of Agricultural Sciences(NAS), Jeonju, South Korea

© 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.** Neighbor-joining tree of *C. sinensis* L. cultivar Sangmok and related species using chloroplast genome sequences. Numbers on the nodes are bootstrap values from 1000 replicates.

## Disclosure statement

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

## Funding

This work was conducted with support from the National Agricultural Genome Program [Project No. PJ013484] of the National Institute of Agricultural Science, Rural Development Administration, Republic of Korea.

## ORCID

Chang-Kug Kim  <https://orcid.org/0000-0002-8770-1736>

## Data availability statement

The data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>, accession number LC488797.

## References

- 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.
- Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JM, Cronk Q. 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot.* 99(2):320–329.
- 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.
- Nadalin F, Vezzi F, Policriti A. 2012. Gapfiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinf.* 13(S14) S8.
- Meng D, Liu S, Zhenggang X, Zhiyuan H, Wenzhen K, Liang W. 2018. The complete chloroplast genome of an economic plant, *Camellia sinensis* cultivar Anhua. *Chin Mitochondrial DNA B.* 3(2):558–559.
- Ming T, Bartholomew B. 2007. Theaceae. *Flow Chin.* 12:367–412.
- Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol Biol.* 14:23–27.
- Yang JB, Yang SX, Li HT, Yang J, Li DZ. 2013. Comparative chloroplast genomes of *Camellia* species. *PLoS One.* 8(8):e73053.