

The complete chloroplast genome sequence of the Japanese *Camellia* (*Camellia japonica* L.)

Seon-Hee Kim^a, Chung Hyun Cho^a, Misuk Yang^b and Seung-Chul Kim^a

^aDepartment of Biological Sciences, Sungkyunkwan University, Suwon, Gyeonggi-do, Republic of Korea; ^bAmorePacific Corporation, R&D Center, Yongin-si, Gyeonggi-do, Republic of Korea

ABSTRACT

The complete chloroplast genome sequence of valuable ornamental tree, *Camellia japonica* L. (Theaceae), was determined. The genome size was 156,971 bp in length, containing a pair of 25,798 bp inverted repeat (IR) regions, which were separated by small and large single copy regions (SSC and LSC) of 18,394 and 86,673 bp, respectively. The cp genome contained 134 genes, including 91 coding genes, six rRNA genes, and 37 tRNA genes. The overall GC content of the chloroplast genome was 37.3%. Phylogenetic analysis revealed the position of *C. japonica* being sister to the clade containing *C. crapnelliana* and *C. oleifera* (subgenus *Camellia*).

ARTICLE HISTORY

Received 15 August 2017
Accepted 24 August 2017

KEYWORDS

Camellia japonica;
chloroplast genome;
phylogenetic relationships

The genus *Camellia* L., containing more than 200 species, is the largest and economically most important genus in the family Theaceae (Vijayan et al. 2009). It is mainly distributed in East and Southeast Asia and is widely cultivated around the world (Vijayan et al. 2009). Species of the *Camellia* provide many valuable products, including nonalcoholic tea, edible oils, and ornamental plants (Eden 1958; Kondo 1975; Ming 2000; Gao et al. 2005; Zhang et al. 2007; Vijayan et al. 2009). Especially, *C. japonica* is the most important and attractive ornamental plant. In addition, it is one of the important members on forest environment of coastal Japan and the southern and western of Korea Peninsula (Numata 1974; Chung and Kang 1996).

We sampled *C. japonica* in Wimi-ri, Jeju Island, Korea (voucher specimen: SKK1508014001) and obtained the complete chloroplast genome sequences. Sequencing was done using the Illumina HiSeq 2000 (Illumina, San Diego, CA) at Macrogen Inc. (Seoul, South Korea) and assembled by SOAP *de novo* 2 (Luo et al. 2012) and CLC Genomics Workbench v.5.5.1 (CLC Bio, Aarhus, Denmark) with Geneious v.8.1.6 (Biomatters Ltd., Auckland, New Zealand). Gene annotation was done using Blast X, Geneious v.8.1.6 (Biomatters Ltd., Auckland, New Zealand), and then manually corrected for start and stop codons and for intron/exon boundaries. Large and small subunits of ribosomal RNA (rRNA) were identified using RNAmmer 1.2 Server (Lagesen et al. 2007), whereas the transfer RNAs (tRNA) were predicted using ARAGORN v1.2.36 (Laslett and Canback 2004). The complete chloroplast genome of *C. japonica* (GenBank ID: KU951523) was then aligned

with 14 representative congeneric species of *Camellia* using MAFFT v.7 (Katoh and Standley 2013). Maximum likelihood (ML) analysis was conducted based on the concatenated 128 chloroplast coding genes using IQ-TREE v.1.4.2 (Nguyen et al. 2015).

The complete chloroplast genome of *C. japonica* has a total length of 156,971 bp, which is composed of large single copy (LSC) region of 86,673 bp, small single copy (SSC) region of 18,394 bp and two inverted repeat (IRa and IRb) regions of 25,952 bp each. The GC contents of the entire chloroplast genome were 37.3%, and those of LSC, SSC, and IR regions were 35.3%, 34.6%, and 43%, respectively. The chloroplast genome included 134 genes, including 91 protein-coding genes, six rRNA genes, and 37 tRNA genes. A gene loss or genome structural change was not detected. The *ycf1* gene in the junction region of IRb and SSC was the only pseudogene found, which was created by incomplete duplication of the normal copy of *ycf1* in the IRa and SSC junction region.

The phylogenetic relationship of *C. japonica* was inferred by its comparison with other 14 chloroplast genomes in *Camellia*, representing two subgenera and six sections (Ming 2000). The ML tree showed that *C. japonica* (sect. *Camellia*) was sister to the clade containing *C. crapnelliana* (sect. *Heterogenea*) and *C. oleifera* (sect. *Oleifera*) of subgenus *Camellia* (Figure 1). Most sections and two subgenera were not monophyletic, reflecting highly specious and reticulating lineage of genus *Camellia* (Vijayan et al. 2009).

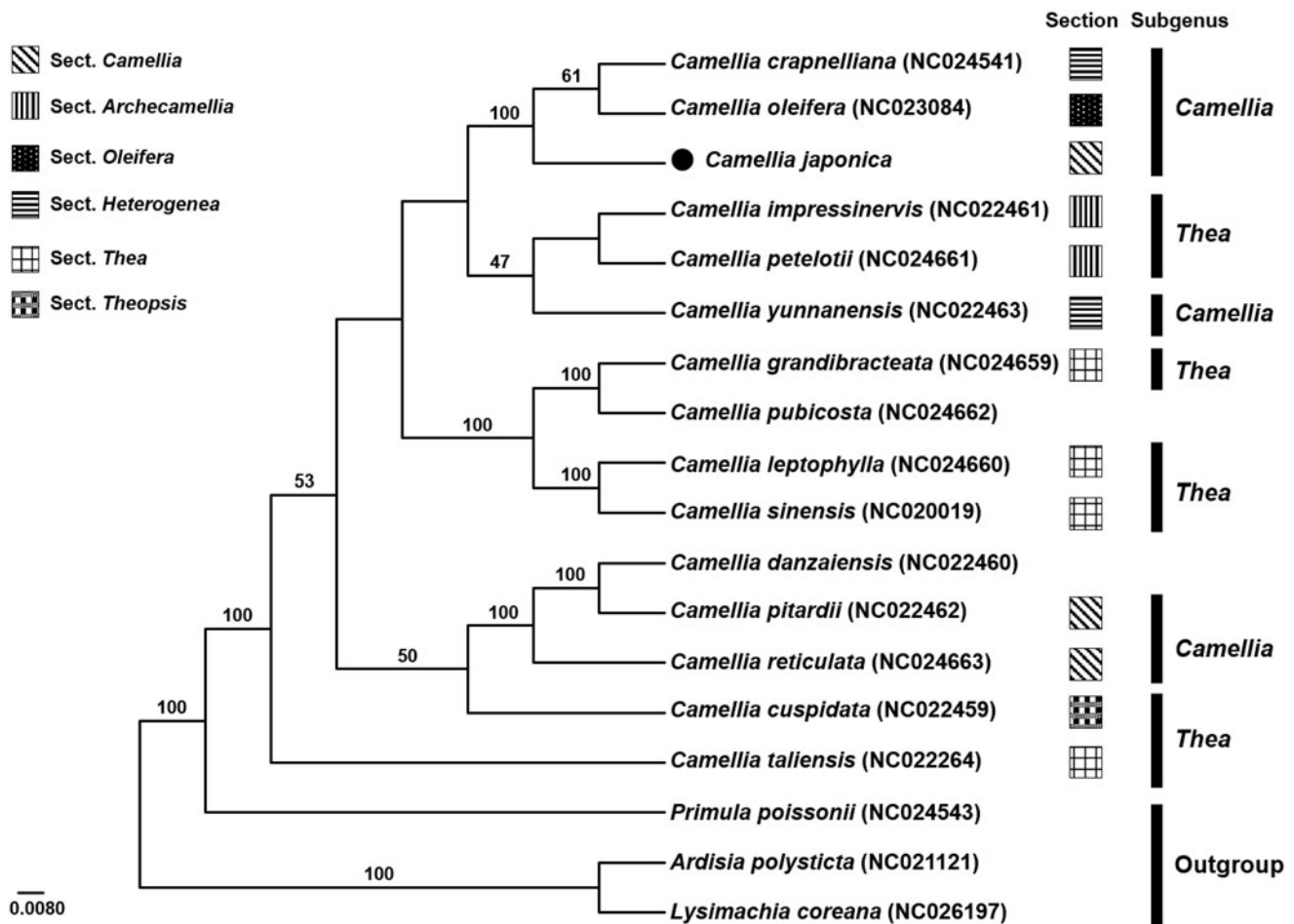


Figure 1. Maximum likelihood tree of the genus *Camellia* based on representative complete chloroplast genome sequences, including *C. japonica* (GenBank ID: KU951523) sequenced in this study. The bootstrap support values are shown above the branches. Three representative taxa of Ericales (*Primula poissonii*: NC024543, *Ardisia polysticta*: NC021121, *Lysimachia coreana*: NC026197) were used as outgroups.

Disclosure statement

The authors declare that there is no conflict of interest regarding the publication of this article. The authors alone are responsible for the content and writing of the paper.

Funding

This study was supported by a grant from the Amore-Pacific Corporation in South Korea, entitled 'Multidisciplinary approach for discovering emerging values of heritage plant, *Camellia japonica*, in Korea'.

References

- Chung MG, Kang SS. 1996. Genetic variation within and among populations of *Camellia japonica* (Theaceae) in Korea. *Can J Res.* 26:537–542.
- Eden T. 1958. The development of tea culture. In: Eden T, editor. *Tea*. London: Longman; p. 1–4.
- Gao JY, Parks CR, Du YQ. 2005. Collected species of the genus *Camellia*; an illustrated outline. Huangzhou, Zhejiang: Zhejiang Science and Technology Press; 302 pp.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kondo K. 1975. Cytological studies in cultivated species of *Camellia* [PhD thesis]. Chapel Hill: University of North Carolina; 260 pp.
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
- Laslett D, Canback B. 2004. ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes. *Nucleic Acids Res.* 32:11–16.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience.* 1:18.
- Ming T. 2000. Monograph of the genus 'Camellia'. Kunming: Yunnan Science and Technology Press; 352 pp.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.
- Numata M. 1974. The flora and vegetation of Japan. Tokyo: Kodansha; 294 pp.
- Vijayan K, Zhang W-J, Tsou C-H. 2009. Molecular taxonomy of *Camellia* (Theaceae) inferred from nrITS sequences. *Am J Bot.* 96:1348–1360.
- Zhang D, Yu J, Chen Y, Zhang R. 2007. Ornamental tea oil from *Camellia* cultivars and their hypocotyl graft propagation. *SNA Res Conf.* 52:257–260.