山西大学学报（自然科学版）47（2）：464－467，2024
Journal of Shanxi University（Nat. Sci. Ed. ）

DOI：10.13451/j.sxu.ns.2023103

长尾毛蕊茶（山茶科）植物的叶绿体全基因组序列

张晓玉[1]，刘瑞瑞[2]，刘兵兵[1*]
（1.山西大学 黄土高原研究所，山西 太原 030006；
2.陇东学院 生命科学与技术学院，甘肃 庆阳 745700）

摘　要：长尾毛蕊茶(*Camellia caudata* Wall. 1832)是一种小型乔木，它主要分布于南亚。它富含抗氧化组分，是一种重要的药用植物。本文基于Illumina平台的双末端测序数据首次报道并描述了长尾毛蕊茶的叶绿体全基因组序列。完整的叶绿体基因组长度为157 043 bp，由两个26 095 bp长的反向重复序列(IRs)以及被其分开的86 599 bp长的大单拷贝区(LSC)和18 272 bp长的小单拷贝区(SSC)组成。该质体基因组包含了89个蛋白质编码基因、37个tRNA基因、8个rRNA基因和1个假基因(*ycf* 2)。质体基因组总GC含量为37.3%。基于20个相关物种全叶绿体基因组的系统发育分析表明，长尾毛蕊茶与同属的大果南山茶(*C. semiserrata* var. *magnocarpa*)和毛蕊红山茶(*C. mairei*)亲缘关系较近。

关键词：山茶属；叶绿体基因组；系统发育分析
中图分类号：Q811.4　　**文献标志码：**A　　**文章编号：**0253-2395(2024)02-0464-04

Complete Chloroplast Genome Sequence of *Camellia caudata* (Theaceae)

ZHANG Xiaoyu[1], LIU Ruirui[2], LIU Bingbing[1*]
（1. *Institute of Loess Plateau, Shanxi University, Taiyuan* 030006*, China;*
2. *College of Life Science and Technology, Longdong University, Qingyang* 745700*, China*）

Abstract: *Camellia caudata* Wall. 1832, a small tree, mainly distributes in South Asia. It is an important medicinal plant because of its high antioxidant contents. Here, for the first time, we report and characterize its complete chloroplast genome based on Illumina paired-end sequencing data. The complete chloroplast genome is 157 043 bp in length, and contains two inverted repeats (IRs) of 26 095 bp separated by a large single-copy (LSC) of 86 599 bp and a small single copy (SSC) of 18 272 bp. The plastome comprises 89 protein-coding genes, 37 tRNA genes, 8 rRNA genes and 1 processed pseudogene (*ycf* 2). The overall GC content of the plastome is 37.3%. Phylogenetic analysis based on 20 plastomes demonstrated that *C. caudata* is closely related to the congeneric *C. semiserrata* var. *magnocarpa* and *C. mairei*.

Key words: *Camellia*; chloroplast genome; phylogenetic analysis

1　Introduction

Camellia caudata Wall. 1832, is a small tree used to produce teas and medicines, belonging to the economically important genus *Camellia* (Theaceae). It is naturally distributed in the Southeast Asia, including the southern provinces of China (Guangxi, Guangdong, Hainan province, etc.),

---

收稿日期：2023-04-19；**接受日期：**2023-05-26
基金项目：国家自然科学基金（41601055）
作者简介：张晓玉（1997-），女，山西榆次人，硕士研究生，研究方向为植物生态学。E-mail：906557067@qq.com
* **通信作者：**刘兵兵（LIU Bingbing），E-mail：lbb2015@sxu.edu.cn
引文格式：张晓玉，刘瑞瑞，刘兵兵．长尾毛蕊茶（山茶科）植物的叶绿体全基因组序列[J]．山西大学学报（自然科学版），2024，47（2）：464-467. DOI：10.13451/j.sxu.ns.2023103

Burma, India and Vietnam[1]. It is well known for its tea and oil seeds[2]. In addition, *Camellia* species are also a rich source of polyphenols with antioxidant properties[3]. Despite its high economic value, little is known about its genetic composition. Chloroplast DNA sequences represent efficient DNA barcoding for species identification and genetic diversity analysis[4]. With the development of high throughput sequencing technology, the study of chloroplast genes has become possible. Here, we sequenced and characterized the whole *C. caudata* chloroplast genome using high-throughput sequencing technology, with the intention of contributing to further research on its genetics and to enhance its resource value.

2 Materials and methods

In this study, *C. caudata* leaves (Fig. 1, Shengfeng Chai, sfchai@163. com) were sampled from Guangxi Zhuang Autonomous Region of China (110°17′E, 25°02′N). The specimen has been deposited at Guangxi Institute of Botany (http://english. gxib. cn/, Sheng-feng Chai, sfchai@163. com) under the voucher number S.-F. Chai C015. The total genomic DNA was extracted from approximately 10 mg silica gel-dried leaves, following a slightly modified CTAB method[5]. The whole-genome sequencing was conducted on the Illumina NovaSeq6000 platform (Illumina, San Diego, CA) from Sangon Biotech (Shanghai, China). Around 1 Gb clean data were used for the chloroplast genome de novo assembly by the program NOVOPlasty version 4.3[6]. The chloroplast genome annotation was performed using the program Plann version 1.1[7] and corrected manually in Sequin version 16.0 (http://www. ncbi. nlm. nih. gov/). A physical map of the genome was generated by OGDRAW version 1.3.1 (http://ogdraw. mpimp-golm.mpg.de/)[8].

3 Results and discussion

The *C. caudata* plastome has a typical quadripartite structure, 157 043 bp in length (Fig. 2). It contains two inverted repeats (IRs) of 26 095 bp separated by a large single-copy (LSC) of 86 599 bp and a small single copy (SSC) of 18 272 bp. The plastome comprises 89 protein-coding genes, 37 tRNA genes, 8 rRNA genes and 1 processed pseudogene (*ycf* 2). Among the annotated genes, 20 of them have two copies. Six tRNA genes and 12 PCGs contain introns, of which 15 (*trn*A-UGC, *trn*I-GAU, *trn*G-UCC, *trn*K-UUU, *trn*L-UAA, *trn*V-UAC, *atp*F, *pet*B, *pet*D, *ndh*A, *ndh*B, *rpo*C1, *rps*16, *rpl*2 and *rpl*16) contain only one intron and 3 (*ycf*3, *rps*12 and *clp*P) contain two introns. The overall GC content of the plastome is 37.3%. The annotated

Fig. 1 The picture of *C. caudata* with fruit. It was taken by Dr. Shengfeng Chai, Guangxi Institute of Botany

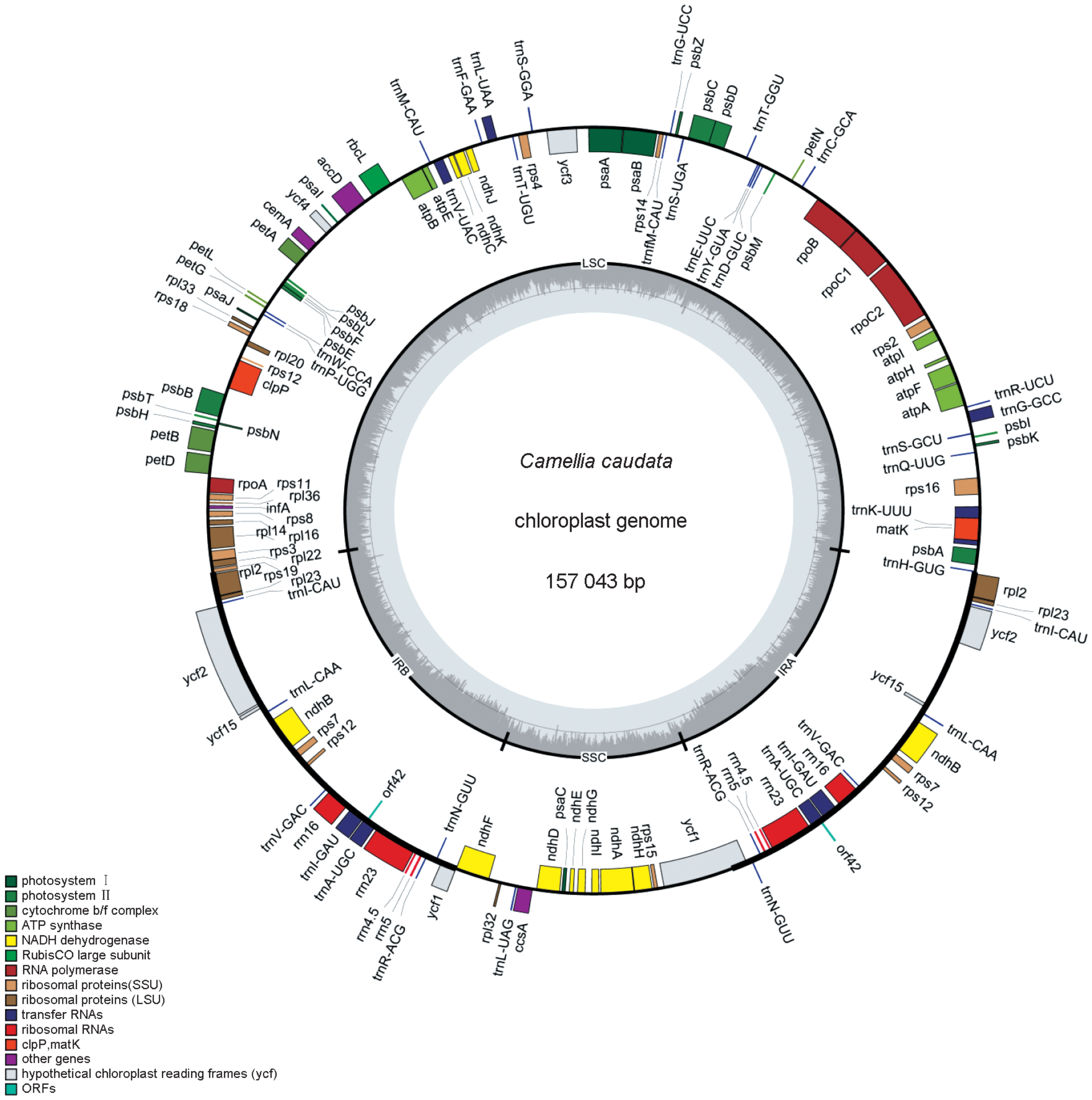

Fig. 2 Genome map of *C. caudata*. Genes within the circle are transcribed clockwise, and those outsides are transcribed counter-clockwise. Genes are colored according to functional categories. The darker grey area in the inner circle corresponds to GC content, whereas the lighter grey corresponds to AT content

chloroplast genome of *C. caudata* has been deposited in GenBank with the accession number OQ274911.

To identify the phylogenetic position of *C. caudata*, a phylogenetic tree was constructed based on whole chloroplast genomes of 20 species, using maximum likelihood (ML) methods. The sequence alignment was performed using MAFFT version 7.037b[9], and the ML tree was produced by MEGA version 7[10] using 1 000 bootstrap replicates and the GTR+G+I nucleotide substitution model. The result shows that *C. caudata* is sister to the congeneric *C. semiserrata* var. *magnocarpa* and *C. mairei* (Fig. 3). Our findings provide a foundation for phylogeographic and population genomic diversity studies of *C. caudata*. These data will also contribute to the elucidation of plastome evolution and phylogenetic relationships within the genus *Camellia*.

The genome sequence data are available in GenBank with accession No. OQ274911. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA934752, SRR23460052, and SAMN33282496, respectively.

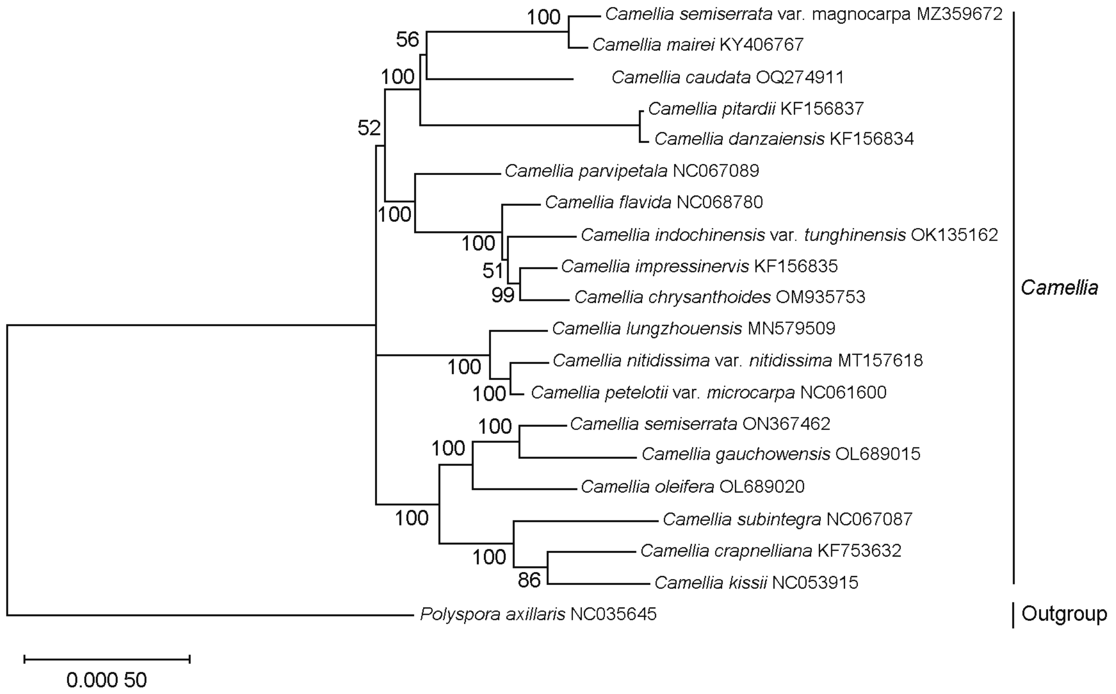

Fig. 3 The ML phylogenetic tree based on 20 complete chloroplast genomes. Numbers on the nodes are bootstrap values from 1000 replicates

References:

[1] MENG X H, LI N, ZHU H T, *et al*. Plant Resources, Chemical Constituents, and Bioactivities of Tea Plants from the Genus *Camellia* section *Thea*[J]. *J Agric Food Chem*, 2019, **67**(19): 5318-5349. DOI: 10.1021/acs.jafc.8b05037.

[2] TEIXEIRA A M, SOUSA C. A Review on the Biological Activity of *Camellia* Species[J]. *Molecules*, 2021, **26**(8): 2178. DOI: 10.3390/molecules26082178.

[3] CHITSAZAN A. Anti-cancer Properties of Green Tea Probed Viaquantum Mechanics Calculations[J]. *Orient J Chem*, 2015, **31**(1): 393-408. DOI: 10.13005/ojc/310147.

[4] DUAN N, DENG L, ZHANG Y, *et al*. Comparative and Phylogenetic Analysis Based on Chloroplast Genome of Heteroplexis (Compositae), a Protected Rare Genus[J]. *BMC Plant Biol*, 2022, **22**: 605. DOI: 10.1186/s12870-022-04000-1.

[5] DOYLE J J. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue[J]. *Phytochem Bulletin*, 1987, **19**: 11-15.

[6] DIERCKXSENS N, MARDULYN P, SMITS G. NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data[J]. *Nucleic Acids Res*, 2017, **45**(4): e18. DOI: 10.1093/nar/gkw955.

[7] HUANG D I, CRONK Q C B. Plann: a Command-line Application for Annotating Plastome Sequences[J]. *Appl Plant Sci*, 2015, **3**(8): 1500026. DOI: 10.3732/apps.1500026.

[8] GREINER S, LEHWARK P, BOCK R. OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes[J]. *Nucleic Acids Res*, 2019, **47**(W1): W59-W64. DOI: 10.1093/nar/gkz238.

[9] KATOH K, STANDLEY D M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability[J]. *Mol Biol Evol*, 2013, **30**(4): 772-780. DOI: 10.1093/molbev/mst010.

[10] KUMAR S, STECHER G, TAMURA K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets[J]. *Mol Biol Evol*, 2016, **33**(7): 1870-1874. DOI: 10.1093/molbev/msw054.