Bangladesh J. Bot. **51**(1): 51-55, 2022 (March) DOI: https://doi.org/10.3329/bjb.v51i1.58820

THE COMPLETE CHLOROPLAST GENOME ANALYSIS OF *CINNAMOMUM CASSIA* PRESL

SHAOFENG JIANG, YINGYING LIU[1], SHAN CUI, LINGJIAN GUI[1*] AND JINQING HUANG[*]

Guangxi Key Laboratory of Tumor Immunology and Microenvironmental Regulation, Guilin Medical University, Guilin 541199, China

Keywords: *Cinnamomum cassia* Presl, chloroplast genome, SSR, phylogeny, conservation genetics

Abstract

Cinnamomum cassia Presl is widely known as an important aroma and medicinal tree species growing in tropical and subtropical regions of Asia. The complete chloroplast genome of *C. cassia* which is 152,763 bp in length was determined. A total of 126 genes were detected in the chloroplast genome, including 36 tRNA genes, 8 rRNA genes, and 82 CDS. The total number of SSRs in this genome is 157. Phylogenetic analysis of this species based on 51 shared chloroplast CDS is provided. These results will provide useful resources for the study of molecular phylogeny and genetics of *C. cassia.*

Introduction

Cinnamomum cassia Presl, a medium-large tree species of Lauraceae may have originated in southern China and is now widely cultivated in the tropical and subtropical Asia (Li *et al.* 2008). It contains many aromatic compounds (Miyamura *et al.* 1983), especially cinnamic alcohol. This compound has a mild, comfortable and long-lasting aroma, and often used as food and cosmetic flavor additive (Zhang *et al.* 2019). In the modern pharmaceutical field, Cinnamic alcohol extracted from *C. cassia* is often used to prepare cinnamyl chloride, which is an excellent raw material for the preparation of cardio-cerebral vascular diseases drugs such as Cinnarizine and Flunarizine (Kirtane *et al.* 2019, Stubberud *et al.* 2019).

Therefore, *C. cassia* has played an important role in cosmetic, food and medicine engineering. At present, the chloroplast genome information of this species is still insufficient, which limits the further research on its genetic protection, especially for wild populations. In the present study the complete chloroplast genome of *C. cassia* (GenBank accession number: MN812496) was assembled and analyzed based on the next-generation sequencing method. The present study aimed to provide useful information for molecular phylogeny and genetics of this species.

Materials and Methods

Total genomic DNA was extracted from silica-dried leaves of *C. cassia* sampled from Guangxi Botanical Garden of Medicinal Plants, Nanning, China (22°51′N, 108°22′E).

The chloroplast genome of *C. cassia* was sequenced based on the Illumina Genome Platform (HiseqPE150) of Beijing Novogene (China). NOVOPlasty (Dierckxsens *et al*. 2017) was used to assemble the clean data after removing adapters to a complete sequence. After referring to the annotations of *Amborella trichopoda* (GenBank accession number AJ506156) and related species. PGA-master (Qu *et al.* 2019) was used to annotate the assembled chloroplast genome sequence, and manually modified it in Geneious 10.2 (Kearse *et al.* 2012) as required. The gene map of *C.*

*Author for correspondence: <auhhuangjinqing@foxmail.com>; <gxcnglj@foxmail.com>. [1]Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China.

cassia chloroplast genome in OGDRAW (Lohse *et al.* 2007) was Codon usage of *C. cassia* was used to analyze by CodonW (Peden 1997).

Using the software of imperfect microsatellite extractor (IMEX) (http://43.227.129.132 :8008/ IMEx / imex_ advance.html) (Nagarajaram, 2007), SSRs were found in the whole chloroplast genomes of *C. cassia*. Specific specifications are set as follows: minimum repeat number of mononucleotide: 10, the rest are dinucleotides: 5, trinucleotides: 5, tetranucleotides: 4, pentanucleotides: 3 and hexanucleotides: 3; the repeat types are imperfect, and the imperfect percentages were 10, 10, 15, 20, 5 and 5% in the above order, mismatches allowed in pattern for Mono, Di, Tri, Tetra, Penta, Hexa was 1, 1, 2, 4, 0, 3, respectively, and the level of standardization was level 1 standardization.

The previous phylogenetic study (Wu *et al.* 2017) did not involve the complete chloroplast genome of *C. cassia*, so there was still uncertainty. In order to further verify the phylogenetic relationship of *C. cassia* and related species, 16 chloroplast genomes (including 11 taxa) of Lauraceae were achieved from NCBI. From these genome sequences, 51 shared CDS were extracted and aligned by MAFFT. A maximum likelihood (ML) tree of Lauraceae was used to analyze by the software of Mega 7 with 500 bootstrap (Kumar *et al.* 2016).

Results and Discussion

The total length chloroplast genome of the *C. cassia* was 152,763 bp with the total GC content was 39.2%, which contained a small single copy region, a large single copy region and a pair of IR regions with lengths of 18,935, 93,696 and 20,066 bp, respectively. The 126 genes were determined, including 36 tRNA genes, 8 rRNA genes and 82 CDS (Fig. 1, Table 1). Gene *ycf1* and *ycf2* were found to stretch across the IRA / SSC and IRB / LSC borders, respectively.

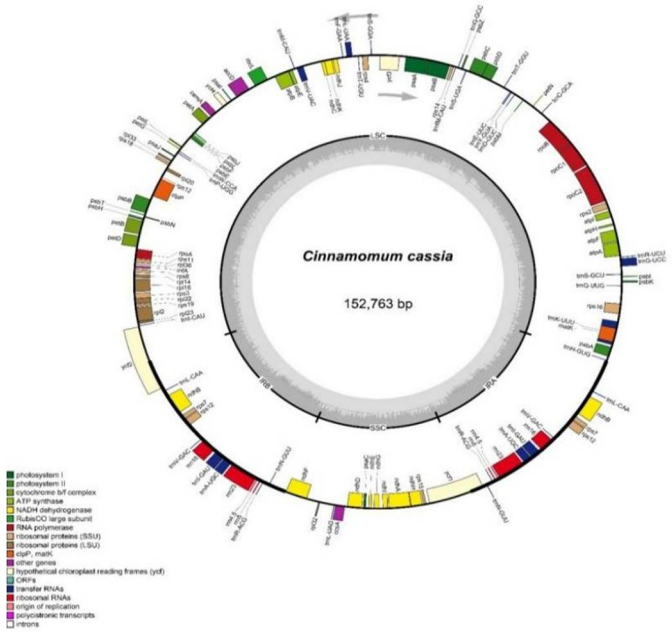

Fig. 1. Gene map of the *Cinnamomum cassia* chloroplast genome. Genes shown outside of the circle are transcribed clockwise, and genes inside are transcribed counterclockwise. Genes belonging to different functional groups are color coded. Dashed area in the inner circle indicates the GC content of the chloroplast genome.

Table 1. List of genes in the chloroplast genome of *Cinnamomum cassia*.

Category	Groups of genes	Name of genes
Protein synthesis and DNA-replication	Transfer RNAs	*trnY-GUA,trnW-CCA,,trnV-GAC*,trnT-UGU,trnT-GGU,trnS-UGA,trnS-GGA,trnS-GCU*,trnR-UCU,trnR-ACG*,trnQ-UUG,trnN-GUU*,trnM-CAU*,trnL-UAG,trnL-CAA*,trnI-GAU*,trnI-CAU*,trnH-GUG*,trnfM-CAU*,trnF-GAA,trnE-UUC,trnD-GUC,trnC-GCA,trnA-UGC**
	Ribosomal RNAs	*rrn23*,rrn16*,rrn5*,rrn4.5**
	Ribosomal protein small subunit	*rps19,rps18,rps16*,rps15,rps14*,rps12*,rps8,rps7*,rps4,rps3,rps2*
	Ribosomal protein large subunit	*rpl36,rpl33,rpl32,rpl23,rpl22,rpl20,rpl16,rpl14,rpl2**
	Subunits of RNA polymerase	*rpoC2,rpoC1*,rpoB,rpoA*
Photosynthesis	ATP synthase	*atpI,atpH,atpF,atpE,atpB*,atpA*
	Cythochrome b/f complex	*petN,petL,petG,petD,petB,petA*
	NADH-dehydrogenase	*ndhK,ndhJ,ndhI,ndhH,ndhG,ndhF,ndhE,ndhD,ndhC,ndhB*,ndhA**
	Photosystem I	*psaJ,psaI,psaC,psaB,psaA**
	Photosystem II	*psbZ,psbN,psbT,psbM,psbL,psbK*,psbJ,psbI*,psbH,psbF,psbE,psbD*,psbC*,psbB,psbA*
	Large subunit Rubisco	*rbcL*
Miscellaneous group	Translation initiation factor IF-1	*infA*
	Acetyl-CoA carboxylase	*accD*
	Cytochrome c biogenesis	*ccsA*
	Maturase	*matK*
	ATP-dependent protease	*clpP*
	Inner membrane protein	*cemA*
Pseudogene unknown function	Conserved hypothetical chloroplast ORF	*ycf1,ycf2,ycf3,ycf4*

*Duplicated gene

Microsatellites (SSRs), which have been proved to be precious resources for species genetic analysis (Nagarajaram 2007), are widely distributed in the chloroplast genome, haveing 1 to 6 bp of repetitive sequences. It was found that there were six perfect types of SSRs (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) in the chloroplast genome of *C. cassia*, and the total number of SSRs was 157. Compound SSR primers were used for *C. cassia* (Table 2).

Reconstruction of 51 shared chloroplast CDS phylogenetic tree showed that *C. cassia* was most closely related to *C. camphora* (MG021326) in this study, but different populations of *C. camphora* were divided into more than one branch. Related species such as *Sassafras tzumu* was found nested in the *Cinnamomum* branch (Fig. 2), indicating that more samples should be sequenced for the further improvement of *Cinnamomum* chloroplast phylogeny (Wu *et al.* 2017).

Table 2. Compound SSR primers used for *Cinnamomum cassia*.

No	Compound SSR	Left primer	Right primer
1	$(a)_{10}$-x_0-$(g)_{10}$	TCGATGGGGCCATTAAAATA	GCCTGTCTTTTTCGATTGCT
2	$(tc)_7$-x_{-2}-$(t)_{18}$	GAGCCAAAGAGGAGGTCTGG	ACCAACACAGCCACGTAGAA
3	$(ctt)_5$-x_{-5}-$(t)_{12}$	TCCGTCTATAATCCCCGATG	AGGGAAAGGGTACAAAATCCA
4	$(c)_{24}$-x_0-$(t)_{12}$	AAAGCCCCTTGTCTTGCTTT	TCTGGCAATAAGAAGCGTCA
5	$(t)_{11}$-x_5-$(ttat)_4$	TTTATGGGTGCCTTCCAGTT	GCGGGAATTGAGACAGTTGA
6	$(t)_{11}$-x_7-$(ttc)_5$	TGCTCGGGGTAGAAGTTTTG	CCGGGAAAGGGTAGAAGAAC
7	$(ata)_5$-x_{-8}-$(tat)_5$	TGCTCGGGGTAGAAGTTTTG	CCGGGAAAGGGTAGAAGAAC
8	$(tta)_5$-x_8-$(at)_5$	CCATGTTTGAGCTGGAGGAT	TAGGTATTGGACCGGGCATA
9	$(aca)_5$-x_{-14}-$(aca)_{11}$	TTTTTGATTGATCCCCGGTA	TCGAAACTGTTTACCCCAAGA

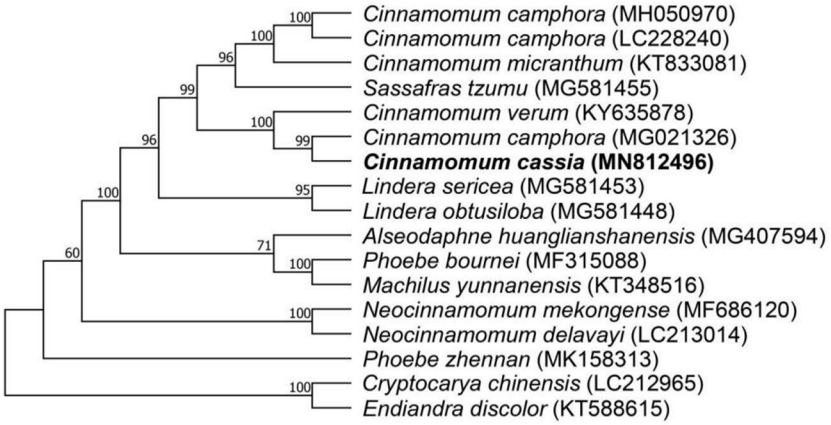

Fig. 2. Maximum likelihood tree based on 51 shared chloroplast CDS of 15 species of Lauraceae. Numbers above the branches are the bootstrap values.

The complete chloroplast genome analysis of *C. cassia* provides valuable and abundant resources and new insights into the molecular phylogeny and conservation genetics.

Acknowledgements

The research work was completed with the support of Guangxi Science and Technology Program (No.AD1850002) and Open Funds of the Guangxi Key laboratory of Tumor Immunology and Microenvironmental Regulation (No. 2019KF006, No.2020KF007).

References

Dierckxsens N, Mardulyn P, Smits G 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. **45**(4): e18.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. **28**(12): 1647-1649.

Kirtane MV, Bhandari A, Narang P, Santani R 2019. Cinnarizine: a contemporary review. Indian Journal of Otolaryngology and Head & Neck Surgery. **71**(2): 1060-1068.

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.

Li HW, Li J, Huang PH, Wei FN, Cui HB, van der Werff H 2008. Lauraceae. **7**. Science Press & Missouri Botanical Garden Press. (Flora of China).

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. Current genetics. **52**(5): 267-274.

Miyamura M, Nohara T, Tomimatsu T, Nishiokat I 1983. Seven aromatic compounds from bark of *Cinnamomum cassia*. Phytochemistry. **22**(1): 215-218.

Nagarajaram M H A 2007. IMEx: imperfect microsatellite extractor[J]. Bioinformatics **23**(10): 1181-1187.

Peden J 1997. CodonW, Trinity College, Dublin, Ireland.

Qu XJ, Moore MJ, Li DZ, Yi TS 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant methods. **15**(1):50.

Stubberud A, Flaaen NM, Mccrory DC, Pedersen SA, Linde M 2019. Flunarizine as prophylaxis for episodic migraine: a systematic review with meta-analysis. Pain. **160**(4): 762-772.

Wu C, Wang T, Wu C, Wang Y, Chaw S 2017. Plastome evolution in the sole hemiparasitic genus *Laurel* Dodder (Cassytha) and insights into the plastid phylogenomics of Lauraceae. Genome Biol. Evol. **9**(10): 2604-2614.

Zhang C, Fan L, Fan S, Wang J, Luo T, Tang Y, Chen Z, Yu L 2019. *Cinnamomum cassia* Presl: a review of its traditional uses, phytochemistry, pharmacology and toxicology. Molecules. **24**(19): 3473.

(Manuscript received on 29 June 2020; revised on 10 January 2022)