

## The complete chloroplast genome sequence of *Lindera fragrans* (Lauraceae): genome structure and phylogenetic analysis

Xue Bai, Juan Peng and Biao Xiong

College of Tea Science, Guizhou University, Guiyang, China

### ABSTRACT

*Lindera fragrans*, naturally distributed in southern China, is economic and ecological tree. Its leaves are one of the primary raw materials for processing *Hawk tea*. Here, we assembled the complete chloroplast genome of *L. fragrans* using Illumina pair-end sequencing. The chloroplast genome size was 152,739 base pairs (bp) with 39% GC content, containing a pair of inverted repeats (IRA/B) of 20,067 bp, separated by a large- and a small single-copy region (LSC/SSC) of 93,711 bp and 18,894 bp, respectively. The genome encoded 126 genes, including 82 protein-coding, 36 transfer RNA (tRNA), and eight ribosomal RNA (rRNA) genes. Phylogenetic analysis based on chloroplast genome sequences of 22 species suggested that *L. fragrans* was closely related to species of *Lindera* genus than other genera in Lauraceae. These results can be a valuable genome resource for further genetics and molecular biology studies.

### ARTICLE HISTORY

Received 29 March 2022  
Accepted 12 June 2022

### KEYWORDS

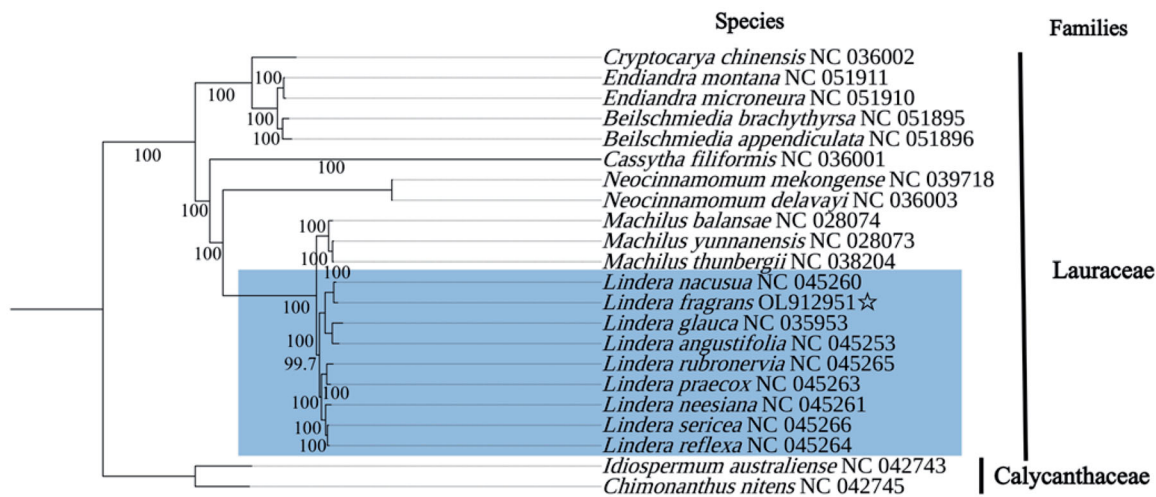
*Lindera fragrans*; complete chloroplast genome; phylogeny

*Lindera fragrans* Oliver (Hooker's Icon. Pl. 18: t. 1788. 1888), a small indecuous tree with lanceolate leaf, is widely distributed in southern China. *L. fragrans* contains compounds such as terpenoids, flavonoids, and antioxidants. Its leaves can prepare *Hawk tea*, which is a traditional folk beverage or a Chinese medicine due to its high flavonoid content (Tan et al. 2016; Jia et al. 2017; Feng et al. 2019). For angiosperm, the gene sequence and structure of the chloroplast genome can play a role in revealing the phylogeny of species and interpreting the origin and domestication of cultivated crops. Although the chloroplast genome of *L. fragrans* was already present in GenBank (unverified: MN453265.1 and MW800996.1), it was not assembled into circular and had no genetic annotation information. Therefore, this study provided the whole chloroplast genome of *L. fragrans* for the first time and meaningful information for the phylogeny of Lauraceae and laid a foundation for subsequent research (NCBI accession number OL912951).

A young *L. fragrans* tree was collected from Baoshan City, Yunnan, China (99°17'E, 25°11'N) and transplanted to the Laboratory of Guizhou University for Cultivation. No endangered or protected individual was involved in the study, and no specific permissions were required for the sample. The voucher samples used in this study were deposited in the Herbarium of Forestry College at Guizhou University (Xingyong Cui, [cuixy0520@163.com](mailto:cuixy0520@163.com)) under the accession number BS202111LF03. The total DNA was extracted from 1 g of the fresh leaves using a modified CTAB method (Doyle and Doyle 1987), and then the quality and quantity of genomic DNA were determined by ultraviolet-visible

spectrophotometer and agarose gel electrophoresis. A sequencing library (PE150) was constructed using the qualified DNA, and then sequenced using an Illumina Hiseq 4000 platform. All raw reads were filtered through NGS QC tool-kit\_v2.3.3 (Patel and Jain 2012). The obtained data were assembled by GetOrganelle (Jin et al. 2020), and then formed a circular chloroplast genome using the Bandage program (Wick et al. 2015). The predictive genes of *L. fragrans* were annotated using CPGAVAS2 (Shi et al. 2019) and compared with the chloroplast sequence of *L. communis* as a reference. The tRNA genes were verified using tRNAscan-SE (Schattner et al. 2005). Phylogenetic tree construction and reliability assessment of internal branches, based on the maximum-likelihood (ML) method, were conducted using the ML method MAFFT v7.271 (Katoh and Standley 2013), and the bootstrap was performed using IQ-TREE v1.6.12 (Minh et al. 2020). GTR + F + I + G4 chosen according to BIC was selected as the best-fit model according to the built-in ModelFinder.

The chloroplast genome of *L. fragrans* was 152,739 base pairs (bp) with 39% GC content, containing a pair of inverted repeats (IRA/B) of 20,067 bp, separated by a large- and a small single-copy region (LSC/SSC) of 93,711 bp and 18,894 bp, respectively. A total of 126 genes are included in the quadripartite structure, including 82 protein-coding genes, 36 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. The phylogenetic analysis based on the complete chloroplast genomes of 20 species in Lauraceae and two from Calycanthaceae suggested that *L. fragrans* was more closely related to species of the *Lindera* genus than other genera in Lauraceae (Figure 1). These results can



**Figure 1.** Maximum-likelihood phylogenetic tree of *L. fragrans* based on the complete chloroplast genomes of 22 previously reported species (all the sequences were downloaded from NCBI GenBank; numbers on the nodes are bootstrap values from 1000 replicates).

provide meaningful data for exploring and utilizing the tea-like species resources.

### Author contributions

BX conceived the project and designed the study; XB and JP performed the sampling and experiments; XB performed the data analysis and wrote the manuscript; BX edited the manuscript; and all authors read and approved the final manuscript.

### Disclosure statement

The authors declare there are no competing interests.

### Data availability statement

The data that support the findings of this study are openly available in GenBank at <https://www.ncbi.nlm.nih.gov/>. The complete chloroplast genome has been deposited in GenBank with accession number OL912951. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA797553, SRS11758197, and SAMN25010772, respectively.

### Funding

This work was supported by the College and Education committee project of Guizhou University [703/702534183301], Guizhou Science and Technology Plan Project [Qiankehe Basics-ZK [2021] General 151], and Guizhou Provincial Postgraduate Research Fund [YJSKYJJ [2021] 005].

### References

- Doyle J, Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19(1):11–15.
- Feng J, Yang J, Chang Y, Qiao L, Dang H, Luo K, Guo H, An Y, Ma C, Shao H, et al. 2019. Caffeine-free hawk tea lowers cholesterol by reducing free cholesterol uptake and the production of very-low-density lipoprotein. *Commun Biol.* 2:173.
- Jia X, Li P, Wan J, He C. 2017. A review on phytochemical and pharmacological properties of *Litsea coreana*. *Pharm Biol.* 55(1):1368–1374.
- Jin JJ, Yu WB, Yang JB, Song Y, de Pamphilis C, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21(1):241.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534.
- Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLOS One.* 7(2):e30619.
- Schattner P, Brooks Angela N, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33(Web Server issue):W686–W689.
- Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47(W1):W65–W73.
- Tan LH, Zhang D, Wang G, Yu B, Zhao SP, Wang JW, Yao L, Cao WG. 2016. Comparative analyses of flavonoids compositions and antioxidant activities of hawk tea from six botanical origins. *Ind Crops Prod.* 80:123–130.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 31(20):3350–3352.