Taylor & Francis
Taylor & Francis Group

PLASTOME ANNOUNCEMENT

🔓 OPEN ACCESS ![Check for updates]

The complete chloroplast genome of *Coffea liberica* (Gentianales: Rubiaceae)

Xuehui Bai[a]*, Hongyu Zheng[b]*, Xing Huang[c] (ID), Jinhong Li[a], Tieying Guo[a], Qin Luo[a], Zhirun Zhang[a], Weihuai Wu[c] and Kexian Yi[c,d,e] (ID)

[a]Dehong Tropical Agriculture Research Institute of Yunnan, Ruili, PR China; [b]China State Farms Economic Development Center, Beijing, PR China; [c]Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, PR China; [d]Key Laboratory of Integrated Pest Management on Tropical Crops, Ministry of Agriculture and Rural Affairs, Haikou, PR China; [e]Hainan Key Laboratory for Monitoring and Control of Tropical Agricultural Pests, Haikou, PR China

ABSTRACT

Coffee is one of the most popular beverages around the world. As one of the best-known coffee species, Liberian coffee (*Coffea liberica* Bull ex Hiern 1876) has a high resistance to leaf rust, a devasting disease caused by *Hemileia vastatrix*. However, there are few reports on the systematic position and phylogenetic relationship of *C. liberica* at the chloroplast (cp) genome level. Thus, we successfully assembled its cp genome. The full length is 154,799 bp with a GC content of 37.48%. We have further annotated the cp genome and predicted 85 protein-coding genes together with 8 rRNAs and 37 tRNAs. Furthermore, a large single copy region (LSC), a small single copy region (SSC), an inverted repeat region a (IRa) and an inverted repeat region b (IRb) are identified with lengths of 84,868 bp, 18,121 bp, 25,905 bp and 25,905 bp, respectively. The phylogenetic tree indicates that *C. liberica* is closely related to *C. canephora*, which is consistent with a previous result obtained from genotyping-by-sequencing.

Background

Coffee is one of the most popular beverages around the world. The three best-known coffee species for coffee production are Arabica (*Coffea arabica* L.), Robusta (*C. canephora* L. Linden) and Liberian coffees (*C. liberica* Bull ex Hiern 1876) (Patay et al. 2016). To date, *C. arabica* has the largest cultivation areas for coffee production, but it is threatened by leaf rust, a devasting disease caused by *Hemileia vastatrix* (Talhinhas et al. 2017). In contrast, high leaf rust resistance has been identified in *C. canephora* and *C. liberica*, which has been successfully used for breeding resistant varieties in *C. arabica* (Prakash et al. 2004). However, there are few reports on the systematic position and phylogenetic relationship of *C. liberica* at the chloroplast (cp) genome level. The cp genome could provide reliable evidence of the evolution and origin of plant species, such as Solanaceae (Mehmood, Shahzadi, et al. 2020; Mehmood, Ubaid, Bao, et al. 2020; Mehmood, Ubaid, Shahzadi, et al. 2020). Thus, we successfully sequenced and assembled the cp genome of *C. liberica*, which will benefit related studies in the future.

Methods and results

Young leaves of *C. liberica* were cut from a five-year-old tree in the coffee germplasm garden of the Dehong Tropical Agriculture Research Institute of Yunnan in Ruili, China (24.0256°N, 97.8596°E) and used for DNA extraction. The specimen has been preserved in the Herbarium of the Dehong Tropical Agriculture Research Institute of Yunnan (http://www.dtari.org.cn/, Xuehui Bai, 13529520059@163.com) under the voucher number DTARI-cl202101. The fresh leaves were rapidly soaked in liquid nitrogen and broken into powder for total DNA extraction by using the CTAB method (Doyle and Doyle 1987). The DNA sample was used for library construction and Illumina sequencing after being delivered to Biozeron Biotech (Shanghai, China). The Illumina NovaSeq platform was selected for paired-end short reads sequencing after the DNA sequences were broken into 300–500 bp fragments. After Illumina sequencing, we deposited a total of 3.81 Gb raw data in the SRA database with the accession number PRJNA771824. A total of 3.78 Gb clean data was filtrated in order to assemble the scaffolds of the cp genome by using NOVOPlasty v4.2 (Dierckxsens et al. 2017). The gaps between scaffolds were filled with GapCloser v1.12 to obtain

CONTACT Xing Huang ✉ hxalong@gmail.com 🖅 Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, PR China; Jinhong Li ✉ lijinhong1969@hotmail.com 🖅 Dehong Tropical Agriculture Research Institute of Yunnan, Ruili, PR China; Weihuai Wu ✉ weihuaiwu2002@163.com; Kexian Yi ✉ yikexian@126.com 🖅 Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, PR China
*These authors contributed equally to this work.
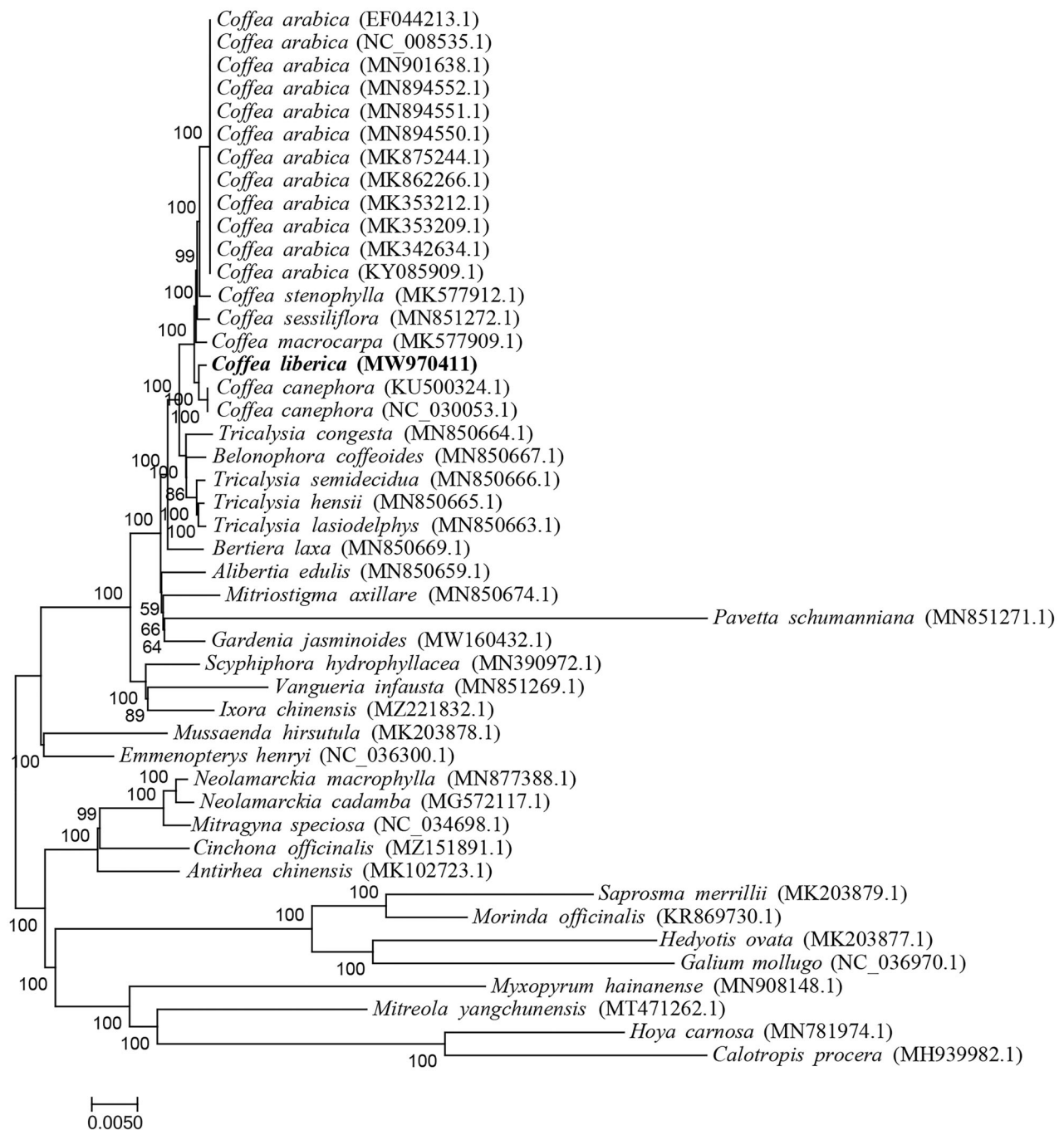
Figure 1. Phylogenetic tree of 46 chloroplast genomes.

the full cp genome (Luo et al. 2012). The cp genome of *C. liberica* contained 154,799 bp with a GC content of 37.48%, which was deposited in GenBank under the accession number MW970411. We selected the GeSeq and CPGAVAS2 software to annotate the cp genome and predicted 85 protein-coding genes together with 8 rRNAs and 37 tRNAs (Tillich et al. 2017; Shi et al. 2019). We selected Geneious v11.0.3 to screen the regional boundaries (Kearse et al. 2012). As a result, a large single copy region (LSC), a small single copy region (SSC), an inverted repeat region a (IRa) and an inverted repeat region b (IRb) were identified with lengths of 84,868 bp, 18,121 bp, 25,905 bp and 25,905 bp, respectively.

We selected 46 cp genome sequences to conduct the phylogenetic analysis. There are 42 species in Rubiaceae and four other species as outgroup, comprising *Myxopyrum hainanense*, *Mitreola yangchunensis*, *Hoya carnosa* and *Calotropis procera* (Amenu et al. 2022). All these cp genomes were aligned using MAFFT v7.0 (Katoh and Standley 2013). The phylogenetic tree was constructed by the Maximum Likelihood method with bootstrap values of 1000 replicates in MEGA 7.0.26 (Kumar et al. 2016). The result has indicated that *C. liberica* is closely related to *C. canephora* (Figure 1), which is consistent with a previous result obtained from genotyping-by-sequencing (Bawin et al. 2021). This study will

benefit future studies related to chloroplast in the *Coffea* genus.

Ethics approval and consent to participate

The study involved only a cultivated crop without any threatened/endangered species. It was exempted from ethical approval and didn't need any permissions to carry it out.

Authors' contribution

Xuehui Bai, Hongyu Zheng and Xing Huang conceived and designed the experiments. Xuehui Bai and Hongyu Zheng analyzed the data and drafted the manuscript. Jinhong Li, Tieying Guo, Qin Luo and Zhirun Zhang contributed to the species identification and sample preparation. Xing Huang, Weihuai Wu and Kexian Yi revised the manuscript. All authors read and approved the final manuscript.

Disclosure statement

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

ORCID

Xing Huang 🔟 http://orcid.org/0000-0003-3950-0080
Kexian Yi 🔟 http://orcid.org/0000-0003-0496-3157

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/nuccore/) under the accession number MW970411. The accession numbers of BioProject, SRA and Bio-Sample are PRJNA771824, SRX12645655 and SAMN22346234, respectively.

References

Amenu SG, Wei N, Wu L, Oyebanji O, Hu G, Zhou Y, Wang Q. 2022. Phylogenomic and comparative analyses of Coffeeae alliance (Rubiaceae): deep insights into phylogenetic relationships and plastome evolution. BMC Plant Biol. 22(1):88.

Bawin Y, Ruttink T, Staelens A, Haegeman A, Stoffelen P, Mwanga JCI, Roldán-Ruiz I, Honnay O, Janssens SB. 2021. Phylogenomic analysis clarifies the evolutionary origin of *Coffea arabica*. J Syst Evol. 59(5): 953–963.

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

Doyle JJ, Doyle JL. 1987. A Rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem Bull. 19:11–15.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

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

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.

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. GigaSci. 1(1):1.

Mehmood F, Shahzadi I, Ahmed I, Waheed MT, Mirza B. 2020. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genomics. 112(2):1522–1530.

Mehmood F, Ubaid Z, Bao Y, Poczai P, Mirza B. 2020. Comparative plastomics of ashwagandha (Withania, Solanaceae) and identification of mutational hotspots for barcoding medicinal plants. Plants. 9(6):752.

Mehmood F, Ubaid Z, Shahzadi I, Ahmed I, Waheed MT, Poczai P, Mirza B. 2020. Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). PeerJ. 8:e9552.

Patay ÉB, Bencsik T, Papp N. 2016. Phytochemical overview and medicinal importance of *Coffea* species from the past until now. Asian Pac J Trop Med. 9(12):1127–1135.

Prakash NS, Marques DV, Varzea VM, Silva MC, Combes MC, Lashermes P. 2004. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. Theor Appl Genet. 109(6): 1311–1317.

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.

Talhinhas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, Tavares S, Pereira AP, Azinheira HG, Guerra-Guimarães L, et al. 2017. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. Mol Plant Pathol. 18(8):1039–1051.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6–W11.