

MITOGENOME ANNOUNCEMENT



Characterization of the first complete chloroplast genome sequence of *Celtis sinensis* (Cannabaceae) and its phylogenetic implications

Guozheng Wang^{a*}, Yunyan Zhang^{a*}, Jiehan Zhou^a, Mengyuan Zhang^a, Yimin Hu^b, Yongjing Tian^a, Jingbo Zhou^a and Zhongsheng Wang^a

^aCollege of Life Sciences, Nanjing University, Nanjing, China; ^bDepartment of Forest Resources, Anhui Academy of Forestry, Hefei, China

ABSTRACT

Celtis sinensis Pers. is a popular garden landscape tree in riparian areas and a valuable resistant tree in many extreme environments. Here, we determined the first complete chloroplast genome of *C. sinensis* using high-throughput sequencing technology. Our results showed the chloroplast genome of *C. sinensis* was 159,092 bp long and displayed a typical quadripartite structure consisting of a pair of inverted repeats with a length of 26,895 bp and separating by two single-copy regions (LSC, 86,085 bp and SSC, 19,217 bp). Besides, the chloroplast genome of *C. sinensis* totally contained 131 genes, including 87 protein coding genes, 36 transfer RNAs and eight ribosomal RNAs. Additionally, a maximum likelihood phylogenetic analysis based on the 19 chloroplast genomes demonstrated the monophyly of Cannabaceae and *C. sinensis* formed a sister clade to *Celtis biondii*.

ARTICLE HISTORY

Received 28 September 2019
Accepted 9 November 2019

KEYWORDS

Celtis sinensis; resistant plant; plastid genome; phylogenomics

Celtis sinensis Pers., a deciduous tree native to Japan, Korea and China (Panetta 2001), belongs to the genus *Celtis* within the eudicot family Cannabaceae *sensu* APG IV (The Angiosperm Phylogeny Group 2016). It is widely cultivated as a popular garden landscape tree in riparian areas because of its characteristics of elegant dense shade, hard trunk and fast growth speed (Siebert et al. 2018). Besides, it performs strong resistance to various adverse conditions, such as saline-alkali soil, industrial pollution areas and coastal inland transition zones (Jim and Liu 2001; Bhuju and Ohsawa 1999). Previous studies about *C. sinensis* mainly focused on its community structure, reproductive regeneration mechanism and ecological adaptability (Lee et al. 2004; Zhang et al. 2016; Siebert et al. 2018). However, researches regarding its genetic backgrounds are extremely scarce. Fortunately, with the advent of next-generation sequencing technologies, abundant genomic resources of our interested species have been produced in a rapid and cost-effective way in recent years (Hao et al. 2011; Malé et al. 2014). To comprehensively understand the genetic background of *C. sinensis* and facilitate its utilization, in this study, we characterized the first complete chloroplast genome sequence of *C. sinensis* based on the Illumina pair-end sequencing data and registered it into the GenBank with the accession number MN481989.

The silica-gel dried leaves of *C. sinensis* were sampled from Mo Mountain (China; 34°30.8589'N, 117°18.2520'E) and the voucher specimen (accession number: ZYY190729) was deposited at the Herbarium of Zhejiang University (HZU). Total genomic DNA was isolated with a modified CTAB

method (Doyle and Doyle 1987). Subsequently, paired-end sequencing library was constructed and sequenced on the Illumina HiSeq X-Ten platform at Beijing Genomics Institute (Shenzhen, China). In results, approximately 3.5 Gb of raw data were generated and we de novo assembled the complete chloroplast genome using GetOrganelle pipeline (<https://github.com/Kinggerm/GetOrganelle>). The genome was then automatically annotated using CpGAVAS (Liu et al. 2012), with manual adjusting and confirming the annotated genes in Geneious v11.0.4 (<http://www.geneious.com/>; Kears et al. 2012).

The complete chloroplast genome of *C. sinensis* was 159,092 bp in length, displaying a typical circular quadripartite structure with a pair of inverted repeat regions (IR, 26,895 bp) separated by a large single-copy region (LSC, 86,085 bp) and a small single-copy region (SSC, 19,217 bp). It encoded 112 unique genes, including 78 protein-coding genes, 30 tRNA genes and four rRNA genes. Nine protein-coding genes (*rps16*, *atpF*, *roc1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *ndhA*) and six tRNA gene (*trnK*-UUU, *trnG*-UCC, *trnL*-UAA, *trnV*-UAC, *trnL*-GAU, *trnA*-UGC) contained a single intron, while three protein-coding genes (*ycf3*, *clpP*, *rps12*) possessed two introns. The gene *rps12* was trans-spliced; with the 5'-end exon located in the LSC region and two copies of 3'-end exon and intron in the IR regions. The GC content of the total chloroplast genome was 36.3%, with the corresponding values of 34.1, 29.8 and 42.3% for the LSC, SSC and IR regions, respectively.

CONTACT Zhongsheng Wang ✉ wangzs@nju.edu.cn College of Life Sciences, Nanjing University, Nanjing, 210023, China

*These authors have contributed equally to this work.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

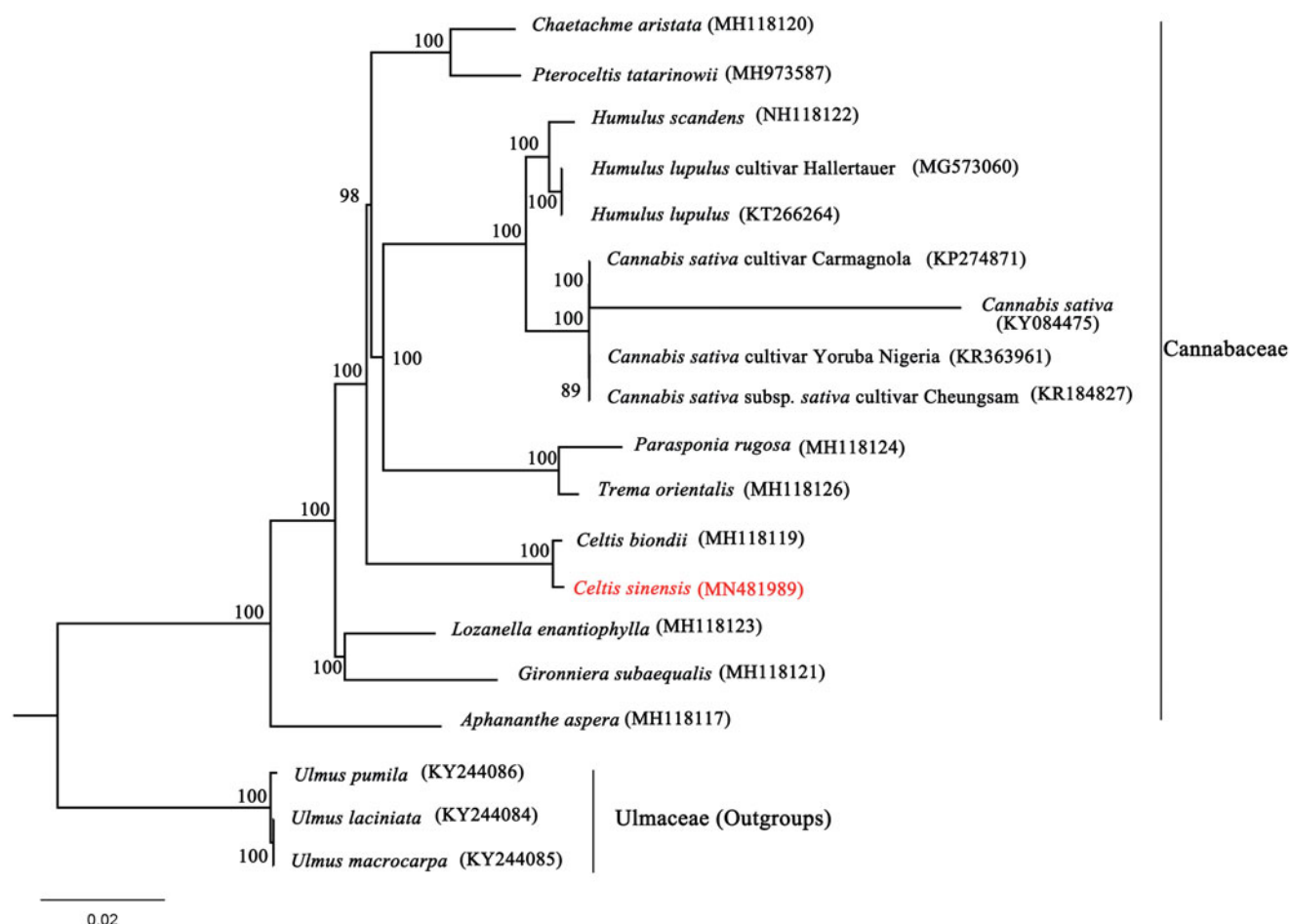


Figure 1. The phylogenetic tree based on 19 complete chloroplast genome sequences. Relative branch lengths are indicated. Numbers near the nodes represent ML bootstrap values.

Furthermore, to ascertain phylogenetic position of *C. sinensis* in the family Cannabaceae, we selected 16 published plastomes of Cannabaceae and the three taxa of Ulmaceae as outgroups, employing the GTR + R + I model and 1000 bootstrap replicates under the maximum-likelihood (ML) inference in RAxML-HPC v.8.2.10 on the CIPRES cluster (Miller et al. 2010). Our reconstructed phylogenetic tree strongly supported the monophyly of Cannabaceae and *C. sinensis* was sister to *Celtis biondii* with full bootstrap support (Figure 1). These novel genomic resources and plastid phylogenomics here will largely enrich the genetic resources and enhance the molecular breeding of *C. sinensis*.

Disclosure statement

The authors are really grateful to the opened raw genome data from public database. The authors report no conflicts of interest and are responsible for the content and writing of the paper.

Funding

This research was supported by the National Natural Science Foundation of China [Grant No. 30970512].

References

- Bhuj D, Ohsawa M. 1999. Species dynamics and colonization patterns in an abandoned forest in an urban landscape. *Ecol Res.* 14(2): 139–153.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Hao Y, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Sahalie J. 2011. Next-generation sequencing to generate interactome datasets. *Nat Methods.* 8(6):478–480.
- Jim CY, Liu HT. 2001. Species diversity of three major urban forest types in Guangzhou city, China. *Forest Ecol Manag.* 146(1–3): 99–114.
- 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.
- Lee CS, Lee KS, Hwangbo JK, You YH, Kim JH. 2004. Selection of tolerant plants and their arrangement to restore a forest ecosystem damaged by air pollution. *Water Air Soil Poll.* 156(1): 251–273.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* 13:715.
- Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, Lhuillier E, Scotti-Saintagne C, Tinaut A, Chave J. 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Mol Ecol Resour.* 14(5):966–975.

- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Paper presented at Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp 1–8.
- Panetta FD. 2001. Seedling emergence and seed longevity of the tree weeds *Celtis sinensis* and *Cinnamomum camphora*. *Weed Res.* 41(1):83–95.
- Siebert SJ, Struwig M, Knoetze L, Komape DM. 2018. *Celtis sinensis* Pers. (Ulmaceae) naturalised in northern South Africa and keys to distinguish between *Celtis* species commonly cultivated in urban environments. *Bothalia*. 48(1):45–56.
- The Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc.* 181:1–20.
- Zhang XB, Yao R, Zhang GF. 2016. Interspecific relationships of dominant species from *Celtis sinensis* communities in Laoshan National Forest Park, China. *J Nanjing Norm Univ.* 39(4):56–64.