


## The complete chloroplast genome sequence of the *Cucurbita ficifolia* Bouché (Cucurbitaceae)

Zhou Cong<sup>a</sup>, Lijuan Cai<sup>b</sup>, Yu Zhang<sup>a</sup>, Wenzhen Su<sup>a</sup>, Huiying Li<sup>a</sup> and Qianglong Zhu<sup>a</sup> 

<sup>a</sup>Department of Horticulture, College of Agronomy, Jiangxi Agricultural University, Nanchang, P.R. China; <sup>b</sup>Nanchang Business College, Jiangxi Agricultural University, Nanchang, P.R. China

### ABSTRACT

*Cucurbita ficifolia* Bouché is an important germplasm resource used for rootstock and hypoglycemic food in Cucurbitaceae. The complete chloroplast genome sequence of *C. ficifolia* has been determined in this study. The total genome size is 157,533 bp in length and contains a pair of inverted repeats (IRs) of 25,639 bp, which were separated by large single copy (LSC) and small single copy (SSC) of 88,112 bp and 18,143 bp, respectively. A total of 130 genes were predicted including 86 protein-coding genes, eight rRNA genes and 36 tRNA genes. Further, Maximum-likelihood phylogenetic analysis revealed that *C. ficifolia* is a base clade of genus *Cucurbita* and closer to *Cucurbita maxima*. The chloroplast genome of *C. ficifolia* would promote the germplasm exploration, phylogenetic relationships, and molecular biology researches in *Cucurbita*.

### ARTICLE HISTORY

Received 11 August 2021  
Accepted 12 September 2021

### KEYWORDS

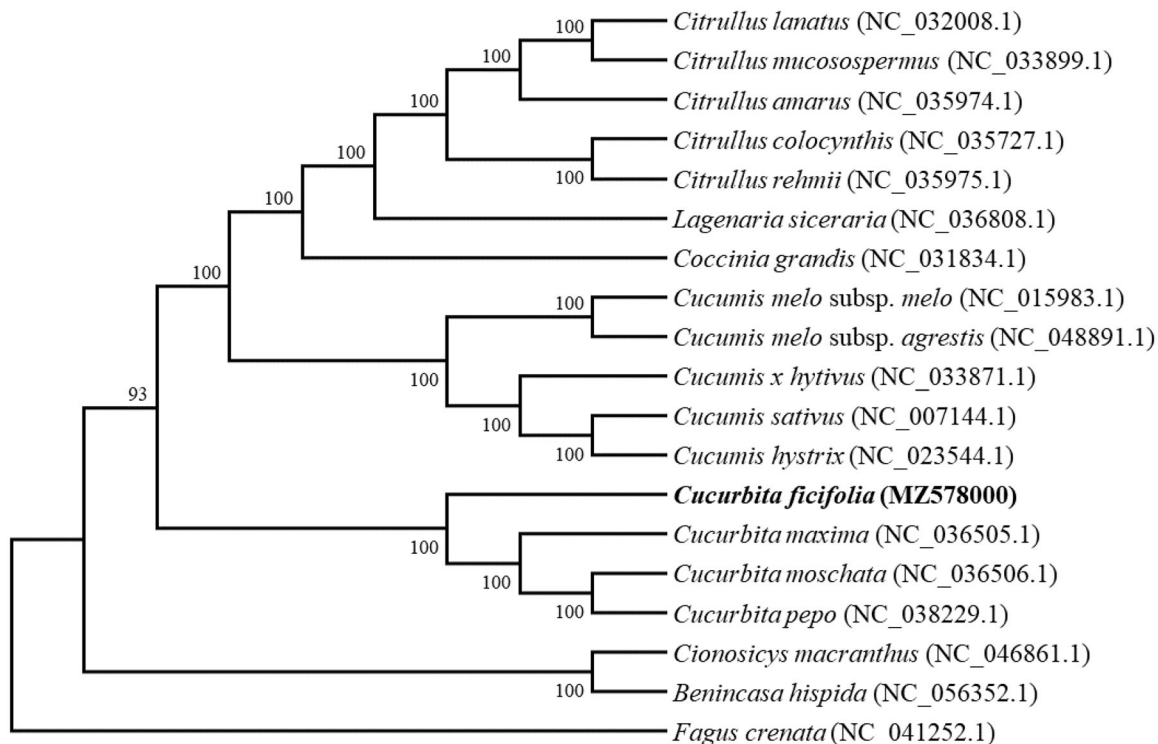
*Cucurbita ficifolia* Bouché;  
chloroplast genome; fig-  
leaf gourd

*Cucurbita ficifolia* Bouché, commonly known as ‘fig-leaf gourd’, is an important germplasm resource with strong resistance to drought, salinity, coldness and soil-borne diseases, therefore, it is worldwide used for rootstock to improve the plant tolerance to abiotic and biotic stress in Cucurbitaceae (Yuan et al. 2010; Ding et al. 2019). The fruit of *C. ficifolia* contains D-chiro-inositol with the hypoglycemic activity in man and other animals (Xia and Wang 2006), is commonly used as a vegetable insulin for remedying diabetes in Asia, Africa and South America (Moya-Hernández et al. 2020), and its seeds have a high content of unsaturated fatty acids, which enables their use as a good and healthy oil to be used in the food industry (Carrillo et al. 2018). *C. ficifolia* was origin from the high-altitude localities in Central and South America and now distributed all over the world. *C. ficifolia* belongs to the Cucurbitaceae (cucurbit) family and consists of at least five domesticated and more than ten wild species (Sun et al. 2017), it is difficult to characterize and identify diverse *Cucurbita* germplasm by traditionally morphological characters, which hinders the excavation and utilization of *C. ficifolia*. Chloroplast genome is a molecular resource for developing DNA markers for identification and domestication of *Cucurbita* (Logan et al. 2015). However, to the best of our knowledge, there are no reports that the chloroplast genome of *C. ficifolia* was taken as a molecular resource. Thus, the goal of this study is to sequence the chloroplast complete genome of *C. ficifolia* with the hope to

promoting the studies on the germplasm exploration, phylogenetic relationships, and molecular biology researches.

The *C. ficifolia* was collected at Jinggangshan, Jian, Jiangxi (114°97'E, 27°12'N) in October 2020, and the seeds were deposited and planted at the horticultural teaching station of Jiangxi Agricultural University under the voucher number: hz23. The genomic DNA was extracted from fresh and healthy leaves of *C. ficifolia* using the modified CTAB method (Porebski et al. 1997). The high-quality DNA was sent to a company and used for library construction and genome sequencing on the BGISEQ-500 (BGI, Shenzhen, China). After sequencing and base quality control, about 2 Gb of clean sequence data in fastq format was obtained. The draft chloroplast genome sequence was assembled by using the Plasmidspades.py in SPAdes 3.15.3 (Bankevich et al. 2012). Contigs representing the chloroplast genome were retrieved, ordered, and incorporated into a long scaffold by aligning to the chloroplast genome of *Cucurbita maxima* (NC 036505.1) using SAMtools and BlastN. The gaps in the scaffold were closed using GapCloser v1.12-r6, and long sequence without gaps was mapped by pair-end reads to validate its completeness. Finally, the long sequence was taken as complete genome sequence and annotated using CPGAVAS2 and GeSeq (Tillich et al. 2017; Shi et al. 2019), the problems in annotations were corrected by Sequin.

The complete chloroplast genome of *C. ficifolia* was submitted to GenBank with accession: MZ578000, the length is



**Figure 1.** Phylogenetic tree showing relationship between *Cucurbita ficifolia* and other 17 species in Cucurbitaceae, *Fagus crenata* (NC\_041252.1) was taken as the outgroup. Phylogenetic tree was constructed based on the complete chloroplast genomes using maximum-likelihood (ML) with 1000 bootstrap replicates. Numbers in each the node indicated the bootstrap support values.

157,533 bp with 36.22% GC contents, and it contains a pair of IRs (25, 639 bp) detached by the LSC (88,112 bp) and SSC (18,143 bp) regions, to exhibit a typical quadripartite structure. There is a total of 131 genes, including 86 protein-coding genes, eight rRNA genes and 36 tRNA genes; six of the protein-coding genes, six of the tRNA genes and four rRNA genes are duplicated within the IRs.

To determine the phylogenetic position of *C. ficifolia*, a phylogenetic relationship was analyzed on the complete chloroplast genome of *C. ficifolia* and other 17 species in Cucurbitaceae with Maximum-likelihood (ML) method using the program MAFFT v7.407 (Nakamura et al. 2018) and MEGA v10.0.4 (Kumar et al. 2018). The phylogenetic tree showed that *C. ficifolia* is a base clade of genus *Cucurbita* and closer to *C. maxima* (Figure 1), the conclusions further support the previous research results (Chomicki et al. 2020).

## Disclosure statement

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

## Funding

This work was funded by the National Natural Science Foundation of China [31960607], the Youth Science Foundation of Jiangxi Provincial Department of Education, China [GJJ180218], and the Construction Project of Root and Tuber Crops Industry and Technology System [JXARS-19-01].

## ORCID

Qianglong Zhu  <http://orcid.org/0000-0001-5355-8815>

## 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/MZ578000.1/> under the accession no. MZ578000.

## References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Carrillo W, Carrillo C, Carpio C, Morales D, Vilcacundo E, Alvarez M, Silva M. 2018. Characterization of fatty acids in sambo oil (*Cucurbita ficifolia* L.) from Ecuador. *Asian J Pharm Clin Res.* 11(2):403–406.
- Chomicki G, Schaefer H, Renner SS. 2020. Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytol.* 226(5):1240–1255.
- Ding YM, Xie JJ, Zhang J, Yao CX, Zhang XG, Yang ZA. 2019. Utilization and research progress on *Cucurbita ficifolia*. *China Vegetables.* 2019(2): 17–28.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 35(6):1547–1549.
- Logan K, Newsom LA, Ryan TM, Clarke AC, Smith BD, Perry GH. 2015. Gourds and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication. *P Natl Acad Sci USA.* 112(49):15107–15112.
- Moya-Hernández A, Bosquez-Molina E, Verde-Calvo JR, Blancas-Flores G, Trejo-Aguilar GM. 2020. Hypoglycemic effect and bioactive compounds associated with the ripening stages of the *Cucurbita ficifolia* Bouché fruit. *J Sci Food Agric.* 100(14):5171–5181.
- Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics.* 34(14):2490–2492.

- Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep.* 15(1):8–15.
- 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.
- Sun H, Wu S, Zhang G, Jiao C, Guo S, Ren Y, Zhang J, Zhang H, Gong G, Jia Z, et al. 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Mol Plant.* 10(10):1293–1306.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organellar genomes. *Nucleic Acids Res.* 45(W1):W6–W11.
- Xia T, Wang Q. 2006. D-chiro-inositol found in *Cucurbita ficifolia* (Cucurbitaceae) fruit extracts plays the hypoglycaemic role in streptozocin-diabetic rats. *J Pharm Pharmacol.* 58(11):1527–1532.
- Yuan H, Bie Z, He S, Hua B, Ai Z, Liu Z. 2010. Improving cucumber tolerance to major nutrients induced salinity by grafting onto *Cucurbita ficifolia*. *Environ Exp Bot.* 69(1):32–38.