

The complete chloroplast genome of *Crateva unilocularis* (Capparaceae)

Xin-hua Zheng^{a,b}, Hua-chao Duan^{a,b}, Shi-min Li^{a,b} and Qiong Dong^{a,b}

^aKey Laboratory of National Forestry and Grassland Administration on Biodiversity Conservation in Southwest China, Kunming, China;

^bForestry College, Southwest Forestry University, Kunming, China

ABSTRACT

Crateva unilocularis is naturally distributed in Southern China, which is an elite natural tree with high edible and medicinal value. In this study, whole chloroplast (cp) genome of *Crateva unilocularis* was assembled and characterized on the basis of Illumina pair-end sequencing data. The complete cp genome was 156,417 bp in length, containing a large single-copy region (LSC) of 85,607 bp and a small single-copy region (SSC) of 18,164 bp, which were separated by a pair of 26,323 bp inverted repeat regions (IRs). The genome contained 128 genes, including 85 protein-coding genes, 35 tRNA genes, and 8 rRNA genes. The overall GC content is 36.32%, while the corresponding values of the LSC, SSC, and IR regions were 33.98, 29.45, and 42.48%, respectively. The maximum-likelihood phylogenetic analysis showed a strong sister relationship with *Crateva tapia*. These findings provide a foundation for further investigation of cp genome evolution in *Crateva unilocularis* and other higher plants.

ARTICLE HISTORY

Received 21 July 2020

Accepted 27 November 2020

KEYWORDS

Crateva unilocularis;
chloroplast genome;
phylogenetic analysis

Crateva unilocularis is a deciduous tree species belong to the family Capparaceae which mainly distributed in southwest China (Sha 2006), it is a unique edible and valuable medicinal plant for Chinese medicine (Sha et al. 2008; Dong et al. 2011). It is especially valued for nutrition and medicine because the young leaves are rich in vitamin C, amino acids and proteins (Cheng et al. 2000). It is well known as an edible vegetable by the local ethnic minorities (Li et al. 2018). Apart from nutritional values, it is a timber species and has ecological protection threats (Nai 2005). Digging its genetic information can help to study the species diversity and evolutionary relationships.

Chloroplast genome are widely being used in the DNA barcoding (Dong et al. 2014), phylogenetic relationships, biology and species conservation (Xue et al. 2012). To date, chloroplast genomes have been reported in many species, such as *Origanum vulgare*, *Capsella rubella* and *Schrenkiella parvula* ((Lukas and Novak 2013; Wu 2015; He et al. 2016) and many others. In this paper, we report the complete chloroplast genome sequence of *Crateva unilocularis* based on the Illumina pair-end sequencing data. This will provide benefits for further studies on biological research in the Capparaceae order.

The fresh leaf samples of *Crateva unilocularis* were obtained from Kunming, Yunnan, China (geospatial coordinates: E102°45'55", N25°03'56", altitude: 1954 m), a voucher specimen (KUN 1265460) is deposited at the Kunming institute of botany, Chinese Academy of Sciences. Extraction of total genomic DNA was done using magnetic beads plant genomic DNA preparation kit. The DNA samples were stored

at the Key Laboratory of National Forestry and Grassland Administration on Biodiversity Conservation in Southwest China, Southwest Forestry University, Kunming, China.

After genomic DNA extraction, a library with the insertion size of 150 bp was constructed and high-throughput DNA sequencing was performed on an Illumina HiSeq X platform. After obtaining approximately 2.93 Gb high-quality clean reads, the raw data were used to assemble a complete Cp genome made by software from GetOrganelle with reference to *Champereia manillana* (Jin et al. 2018). The Geneious R8 (Biomatters Ltd, Auckland, New Zealand) was used to assemble and execute complete cp enzyme labeling annotations. Finally, the Cp DNA sequence of *Crateva unilocularis* was submitted to GenBank (accession number: MT679554).

The complete length of *Crateva unilocularis* cp genome was 156,417 bp (Figure 1), comprising of a large single-copy region (LSC with 85,607 bp) and a small single-copy region (SSC with 18,164 bp), which were separated by a pair of inverted repeats (IRs with 26,323 bp). The overall GC content of genome was 36.32%, the GC content of the LSC (33.98%), and SSC (29.45%) regions were relatively lower than that of the IR regions (42.48%). A total of 128 functional genes were contained in the cp genome, including 85 protein-coding genes (PCG), 8 rRNA genes, and 35 tRNA genes.

To further investigate the phylogenetic position of *Crateva unilocularis*, The maximum-Likelihood (ML) tree was constructed based on complete cp genome sequences of 24 other Capparaceae species using MEGA-X with 1000 bootstrap replicates (Kumar et al. 2016). The result of the phylogenetic analysis showed that *Crateva unilocularis* is closely

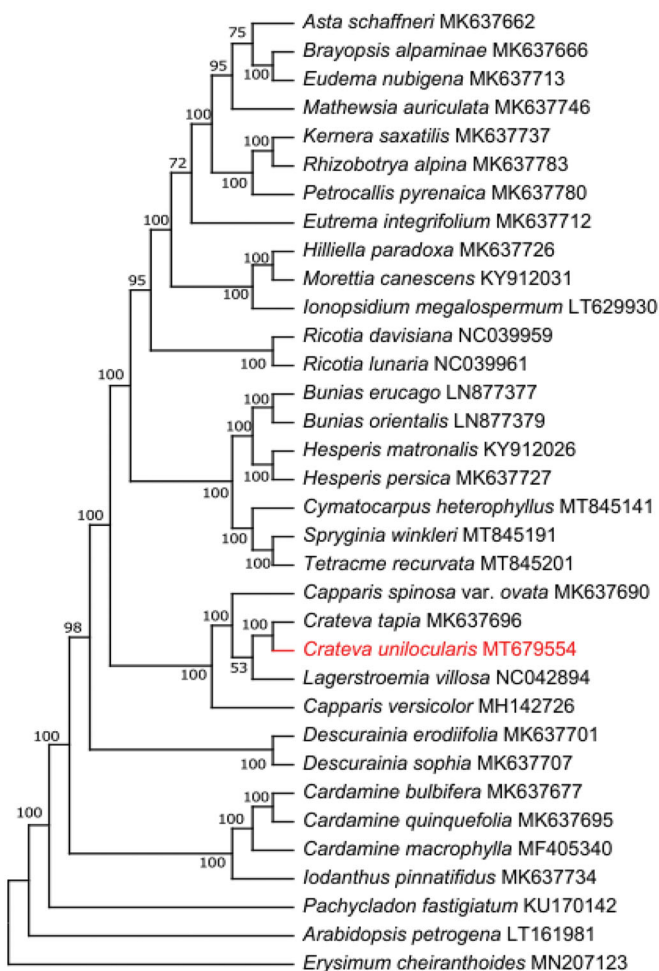


Figure 1. The best maximum likelihood (ML) phylogram inferred from 34 chloroplast genomes in Capparidaceae.

related to the species of *Crateva tapia*. In conclusion, complete chloroplast genome of *Crateva unilocularis* is decoded for the first time, which will also provide valuable information for the study of the genetic diversity and further biological analysis of *Crateva unilocularis* and Capparidaceae.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The study was supported by the National Natural Science Foundation of China [Grant No. 31260191], and the Scientific Research Fund of Yunnan Provincial Department of Education [Grant No. 2019Y0125].

Data availability statement

The data that support the findings of this study are openly available in [National Center for Biotechnology Information], [<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA671642>], accession number [MT679554].

References

- Cheng WH, Sima YK, Wang KL, Xu JC. 2000. The wild species and their use value of Capparidaceae in Southeastern Yunnan. *J West China Forest Sci.* 92(3):23–26.
- Dong Q, He Z, Xu YP. 2011. Determination of vitamin C in *Crateva unilocularis* from different regions. *Chin J Tropical Agri.* 31(09):48–50.
- Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genet.* 15:138.
- He Q, Hao G, Wang X, Bi H, Li Y, Guo X, Ma T. 2016. The complete chloroplast genome of *Schrenkiella parvula* (Brassicaceae). *Mitochondrial DNA Part A.* 27(5):3527–3528.
- Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018. Get Organelle: a simple and fast pipeline for de novo assemble of a complete circular chloroplast genome using genome skimming data. *BioRxiv.*4(14):1–11.
- 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 HY, Zhong XT, Liu Y, Chen GM. 2018. High efficient cultivation technology of woody wild herb *Crateva unilocularis*. *China Tropical Agri.* 84(05):89–90.
- Lukas B, Novak J. 2013. The complete chloroplast genome of *Origanum vulgare* L. (Lamiaceae). *Gene.* 528(2):163–169.
- Nai ZW. 2005. Seedling and cultivation techniques of wild *Crateva unilocularis*. *Appl Technol Rural Areas.* 30(4):18–19.
- Sha A. 2006. The exploitation and utilization of wild vegetable *Crateva unilocularis* in Yunan Province. *Chin Wild Plant Resour.* 25(3):35–36.
- Sha S, Wang YH, Gao C, Shen SK, Ma MX. 2008. Analysis of nutritional components in wild vegetable *Crateva unilocularis*. *J West China Forest Sci.* 37(1):100–102.
- Wu Z. 2015. The complete chloroplast genome of *Capsella rubella*. *Mitochondrial DNA.* 27(5):1.
- Xue JH, Dong WP, Cheng T, Zhou SL. 2012. Nelumbonaceae: systematic position and species diversification revealed by the complete chloroplast genome. *J Syst Evol.* 50(6):477–487.