

The complete chloroplast genome of *Euphorbia ebracteolata* Hayata (Euphorbiaceae)

Mengli Wang, Xiufu Wan, Jiuwen Liang, Tan Li and Sheng Wang

State Key Laboratory Breeding Base of Dao-di Herbs, National Resources Center of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, PR China

ABSTRACT

Euphorbia ebracteolata is a perennial medicinal plant and widely used in China for thousands of years. The complete chloroplast genome reported here is 163,090 bp in length, including two inverted repeats (IRs) of 26,699 bp, which are separated by a large single-copy (LSC) and a small single-copy (SSC) of 91,943 and 17,749 bp, respectively. The whole chloroplast genome of *E. ebracteolata* contains 112 genes, including 78 protein-coding genes, 30 transfer RNA, and 4 ribosome RNA. Phylogenetic analysis result strongly indicated that *E. ebracteolata* is closely related to *E. helioscopia*.

ARTICLE HISTORY

Received 22 September 2020
Accepted 8 November 2020

KEYWORDS

Euphorbia ebracteolata;
chloroplast genome;
phylogenetic analysis

The genus *Euphorbia* is the largest in the spurge family, some of which have been used as medicinal plants for a long time. *Euphorbia ebracteolata* Hayata (Euphorbiaceae), a perennial herb distributed in the central and eastern of China. The root of this plant is poisonous, named 'LangDu' which is used as a traditional herbal medicine to treat edema, asthma, tuberculosis, and tumor diseases for thousands of years (Fu et al. 2006). Because 'LangDu' is widely used and extremely important, it is necessary to understand the background of the evolutionary biology of *E. ebracteolata*. Chloroplast genome is used as a source of valuable data for phylogenetic analysis, genetic diversity evaluation, and plant molecular identification (Dong et al. 2018; Sun et al. 2020). In this study, the completed chloroplast genome sequence of *E. ebracteolata* is determined which provides valuable genetic information for the phylogenetic studies.

The fresh samples of *E. ebracteolata* were collected from Huaining county Anhui province, China (30°44'1"N, 116°49'47"E). Voucher specimen was stored at the herbarium of Institute of Chinese Materia Medica (CMMI), China Academy of Chinese Medical Sciences with the specimen voucher number is 340822LY0035. Total genomic DNA from fresh leaves of a single individual using the method of Li et al. (2013). And the sequencing library was constructed and quantified following the methods introduced by Dong et al. (2017).

The whole-genome sequencing was conducted with 150 bp paired-end reads on the Illumina HiSeq X Ten platform. Contigs were assembled from the high-quality

paired-end reads by using the SPAdes version 3.6.1 program (Kmer = 95) (Bankevich et al. 2012) after low-quality reads were filtered. The chloroplast genome contigs selected by the Blast program (Altschul et al. 1990), taken *Malva parviflora* (GenBank: MK860036) as the reference. The selected contigs were assembled using Sequencher version 4.10 (Gene Codes Corporation, Ann Arbor, MI <http://www.genecodes.com>). Gene annotation was performed with Plann (Daisie et al. 2015), taken the cp genome of *E. esula* (GenBank: KY000001) as the reference and manually corrected for codons and gene boundaries using BLAST searches. The annotated *E. ebracteolata* genomic sequence has been deposited into GenBank with the accession number MT830860.

The complete chloroplast genome reported here is 1,63,090 bp in length, including two inverted repeats (IRs) of 26,699 bp, which are separated by a large single-copy (LSC) and a small single-copy (SSC) of 91,943 and 18,089 bp, respectively. The overall GC-content of the chloroplast genome was 35.5%. The chloroplast DNA of *E. ebracteolata* comprised 112 distinct genes, including 78 protein-coding genes, 4 transfer RNA, and 30 ribosome RNA, do not contain *ycf15*. In these genes, 17 harbored a single intron, whereas two (*ycf3* and *clpP*) contained double introns. The *rps12* gene is a trans-spliced gene with 5' end located in the LSC region and the 3' end located in the IR region. The gene *trnK-UUU* has the largest intron, which contains the *matK* gene.

To confirm the phylogenetic location of *E. ebracteolata* within the genus *Euphorbia*, total 11 complete cp genomes

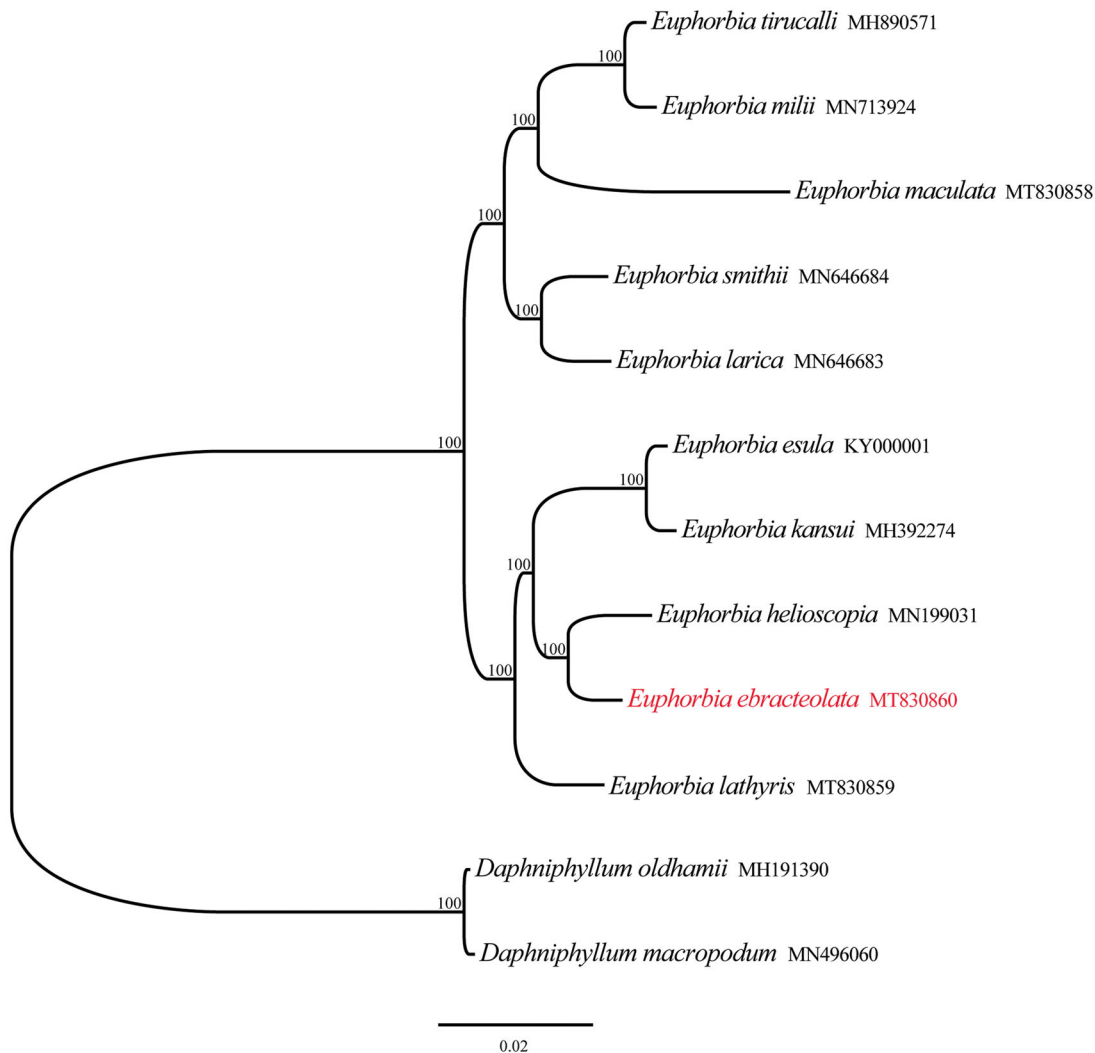


Figure 1. Phylogenetic tree reconstruction of 12 taxa using maximum likelihood (ML) methods in the chloroplast genome sequences. ML bootstrap support value presented at each node.

were downloaded from Genbank and the genus *Daphniphyllum* was taken as an outgroup. All chloroplast genome sequences were aligned using MAFFT (Kato et al. 2019). Phylogenetic analysis was conducted based on IQ-tree using PhyloSuite under the TVM + F + I + G4 model with 1000 bootstrap replicates (Nguyen et al. 2015; Zhang et al. 2020). The phylogenetic tree showed that all species of *Euphorbia* form a monophyletic group with 100% support, and *E. ebracteolata* was closely related to *E. helioscopia* (Figure 1). The chloroplast genome of *E. ebracteolata* provided a lot of genetic information for species conservation and taxonomy of genus *Euphorbia*.

Disclosure statement

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

Funding

This work was supported by National Natural Science Foundation of China [No.81703648, No.81891014], the Fundamental Research Funds for the Central public welfare research institutes [ZZXT201901, ZZ13-YQ-084]

and the National Key Research and Development Program of China [2017YFC1701405].

Data availability statement

The data that support the findings of this study are openly available in GenBank of NCBI <https://www.ncbi.nlm.nih.gov/>, reference number MT830860, raw data BioProject ID: PRJNA662166, Submission ID: SUB8183746.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- 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.
- Daisie, Huang, Quentin, and Cronk. 2015. 'Plann: A command-line application for annotating plastome sequences'. *Applications in Plant Sciences.*
- Dong W, Xu C, Li W, Xie X, Lu Y, Liu Y, Jin X, Suo Z. 2017. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *Front Plant Sci.* 8:1148.

- Dong W, Xu C, Wu P, Cheng T, Yu J, Zhou S, Hong DY. 2018. Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales. *Mol Phylogenet Evol.* 126:321–330.
- Fu GM, Qin HL, Yu SS, Yu BY. 2006. Yuexiandajisu D, a novel 18-norrosane-type dimeric diterpenoid from *Euphorbia ebracteolata* Hayata. *J. Asian Nat Prod Res.* 8(1–2):29–34.
- Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 20(4):1160–1166.
- Li JL, Wang S, Jing Y, Wang L, Zhou SL. 2013. A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany.* 48:72–78.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective Stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Sun J, Wang Y, Liu Y, Xu C, Yuan Q, Guo L, Huang L. 2020. Evolutionary and phylogenetic aspects of the chloroplast genome of *Chaenomeles* species. *Sci Rep.* 10:11466.
- Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 20(1):348–355.