Ehretia dicksonii is a deciduous botany of the family Ehretiaceae. In this work, the complete chloroplast (cp) genome sequence of *E. dicksonii* was probed by next generation sequencing in an effort to provide genomic resources useful for promoting its conservation. The complete cp genome of *E. dicksonii* is 156,623 bp in length, including a large single-copy (LSC) region of 86,853 bp, and a small single-copy (SSC) region of 18,150 bp. It contains 133 genes, including 37 tRNA genes, 8 rRNA genes, and 88 protein-coding genes. The overall GC content of *E. dicksonii* chloroplast genome is 37.85%. The phylogenetic analysis suggests that *E. dicksonii* is in the clade of Ehretiaceae other than Boraginaceae. Also, *E. dicksonii* has a close relationship with *Ehretia acuminata* Brown 1810 in Ehretiaceae.

The fresh leaves of *E. dicksonii* were collected from the campus of Nanjing Forestry University, Jiangsu Province (E 118° 48' 33", N 32° 4' 45") in China. A specimen was deposited at Nanjing Forestry University (contact person: Xuehong Ma; email: xuehongma@njfu.edu.cn) under the voucher number NF2021040. According to the International Union for Conservation of Nature (IUCN) policy on endangered species research, the sample collection and the study was conducted with permission from Arboretum of Nanjing Forestry University. Then total genomic DNA was extracted with Plant DNA Kit (Genepioneer Biotechnologies, Nanjing, China). The complete cp genome sequence of *E. dicksonii* (GeneBank accession number: MZ555766) was characterized based on Illumina paired-end sequencing data to provide a valuable complete cp genomic resource. The DNA fragments were passivated, repaired, and bonded by ultrasonic wave and selected by agarose gel electrophoresis. The sample of genome sequencing library was formed by PCR amplification, which was carried out on Illumina Novaseq platform by Nanjing Genepioneer Biotechnologies Inc. (Nanjing, China), and read long for PE150 sequencing.

The original reading was filtered by fastp (version 0.20.0), and the clean data were assembled into chloroplast genome using SPAdes (Bankevich et al. 2012). Next, the reference sequence (Genebank accession number: MF179500.1) was used for quality control after assembly. Finally, the assembled genome was annotated using CpgAVAS (Liu et al. 2012).

The complete chloroplast genome sequence of *E. dicksonii* was 156,623 bp in length. The genome had a typical quadripartite structure including a pair of IR (IRa and IRb) regions of 25,810 bp that were separated by an LSC region of 86,853 bp and a SSC region of 18,150 bp. A total of 133 genes were encoded, including 8 rRNA genes (4 rRNA species), 37 tRNA genes (28 tRNA species), and 88 protein-coding genes (80 CDS species). Most of the genes occurred in a single copy; however, 8 protein-coding genes (*ndhB*, *rp12*, *rp123*, *rps12*, *rps7*, *ycf1*, *ycf15* and *ycf2*), 9 tRNA genes (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*,...
trnL-GAU, trnL-CAA, trnM-CAU, trnN-GUU, trnR-ACG and trnV-GAC), and 4 distinct rRNA genes (23S, 16S, 5S and 4.5S) are duplicated. A total of 10 protein-coding genes (atpF, ndhA, ndhB, petB, petD, rpl16, rpl2, rpcC1, rps12, rps16) contained 1 intron while the other 2 genes (clpP, ycf3) had 2 introns each. The overall GC content of the chloroplast genome is 37.85%. In addition, the GC contents of the LSC, SSC and IR regions are 35.87%, 32.2% and 43.15%, respectively.

To reveal the phylogenetic evolution of *E. dicksonii*, the phylogenetic tree (phylogram) was constructed based on 4 cp genomes from Ehretiaceae, 7 cp genomes from Boraginaceae and 1 cp genome from Carlemanniaceae as outgroups.

After sequence alignment by MAFFT (Rozewicki et al. 2019), IQTREE (Gao et al. 2018) was used to perform maximum Likelihood (ML) tree with the TVM + F + R3 model. The bootstrap method was used to test the reliability of phylogeny with 1000 replicates. The phylogenetic analysis result supported that Ehretiaceae and Boraginaceae are in 2 clades and belong to distinct taxa, *E. dicksonii* and *E. acuminata* are sister species, and they are in relative late differentiation stage in the clade of Ehretiaceae (Figure 1). As *E. dicksonii* had a close relationship with *E. acuminata* in Ehretiaceae, it can be inferred that *E. dicksonii* probably contains similar natural products with *E. acuminata* (Li et al. 2010), which can promote the study of its resource value, which needs to be further studied.

### Funding

This work was supported by the [The Biodiversity Investigation, Observation and Assessment Program of Ministry of Ecology and Environment of China #1] under Grant [number 20191226]; [The Special Fiscal Funds for Repair and Purchase in National Public Institutions #2] under Grant [number 2010002002]; [Jiangsu Forestry Science and Technology Innovation and Extension Program #3] under Grant [number LYKJ[2018]13].

### 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/] under the accession no. MZ555766. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA745580, SRR15100910, and SAMN20169761 respectively.

### References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.

Brown R. 1810. *Prodromus Florae Novae Hollandiae et Insulae van-Diemen* 1: 497.

Gao F-L, Shen J-G, Liao F-R, Cai W, Lin S-Q, Yang H-K, Chen S-L. 2018. The first complete genome sequence of *Narcissus latent* virus from *Narcissus*. *Arch Virol.* 163(5):1383–1386.

Hance HF. 1862. *Ehretia dicksonii* Hance. *Ann Sci Nat Bot sér 4*(18):224.

Li L, Shi R-B, Wulan T-N, Xu L-J, Peng Y, Xiao P-G. 2010. Chemical constituents in leaves of *Ehretia thyrsiflora*. *Chin J Chin Mater Med.* 35(03):331–332.

Liu C, Shi L-C, Zhu Y-J, Chen H-M, Zhang J-H, Guan X-J. 2012. CpgAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* 13(715):715.

Liu S-L, Wu B-C. 2013. *Boraginaceae*. In: *Li Q-X, editor. Flora of Jiangsu. Vol. 4 (Boraginaceae).* Jiangsu: Jiangsu Phoenix Science Press; St. Louis: Missouri Botanic Garden Press. p. 334.

### Disclosure statement

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

### Author contributions

Among the members of the author group, Xiaogang Xu, Lili Tong contributed to substantial conception or design of the work; Yao Cheng and Lu Tian were in charge of acquisition, analysis, and interpretation of data for the work; Xiaogang Xu, Chongli Xia and Yao Cheng contributed to manuscript preparation (Drafting the work or revising it critically for important intellectual content); Lili Tong, Xiaogang Xu and Chongli Xia contributed to final approval of the version to be published. All authors agree on the final version and to be accountable for all aspects of the work.