

Complete chloroplast genome sequences of *Corydalis edulis* and *Corydalis shensiana* (Papaveraceae)

Yan-Yan Liu^a, Sheng-Long Kan^c, Jun-Li Wang^a, Ya-Nan Cao^a and Jia-Mei Li^b 

^aCollege of Plant Protection, Henan Agriculture University, Zhengzhou, China; ^bCollege of Life Sciences, Henan Agriculture University, Zhengzhou, China ^cState Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China

ABSTRACT

Corydalis DC., the largest genus of Papaveraceae, was recognized as one of the most taxonomically challenging plant taxa. Due to the lack of genetic information used in previous studies, species discrimination and taxonomic assignment in *Corydalis* have not been fully solved. Here, the complete chloroplast genomes were reported for *Corydalis edulis* Maxim. and *Corydalis shensiana* Liden, with their genome sizes being 154,395 and 155,938 bp, respectively. Both of the chloroplast genomes comprise two inverted repeat (IR) regions, separated by a large single-copy (LSC) region and a small single-copy (SSC) region, and encode 130 genes, including 85 protein-coding genes, 8 ribosomal RNA genes, 37 transfer RNA genes. Our study will provide novel insight into the molecular phylogeny and classification of *Corydalis*.

ARTICLE HISTORY

Received 14 October 2020
Accepted 26 November 2020

KEYWORDS

Chloroplast genome;
Corydalis; Papaveraceae

Corydalis DC., the largest genus of Papaveraceae, contains about 400 species (Zhang et al. 2008). This genus is an important component of the biodiversity in the Himalaya–Hengduan Mountains and was recognized as one of the most taxonomically challenging plant taxa. Due to the lack of genetic information used in previous studies, species discrimination and taxonomic assignment in *Corydalis* have not been fully solved (Wang 2006; Ren et al. 2019). In recent years, the whole chloroplast (cp) genomes have become valuable resources for molecular phylogeny and species identification due to the maternal mode of inheritance, dense gene content, and slower evolutionary rates relative to those of nuclear and mitochondrial genomes (Wicke et al. 2011). In this study, we reported the complete cp genomes of *Corydalis edulis* Maxim. and *C. shensiana* Liden, which will provide novel insight into the molecular phylogeny and classification of *Corydalis*.

The fresh leaves of *C. edulis* and *C. shensiana* were collected from Nanyang, Henan Province, China (E111°15'41", 33°25'11") and Fengxian, Shaanxi Province, China (E106°36'27", N34°12'21"), respectively. The voucher specimens were deposited in Henan Agricultural University Herbarium (LYY1933001 and LYY19051101). Total genomic DNA was extracted from silica gel-dried leaves with the CTAB method (Rogers and Bendich 1988) and sequenced using Illumina HiSeq2500 platform at Suzhou Jinweizhi Biotechnology Institute. In total, 42.1 and 39.6 million (M) high-quality raw reads (150 bp PE read length, with Q30 > 91%) were generated for *C. edulis* and *C. shensiana*, respectively. The raw reads were filtered using CLC Genomics Workbench (<http://www.clcbio.com>) to remove low-quality

reads and those containing adaptors with the default settings. The clean reads were assembled into the draft cp genome by CLC Genomics Workbench and GENEIOUS V11.01 (<http://www.geneious.com>) with *Coreanomecon hylomeconoides* Nakai as the reference genome (GenBank accession number: NC_031446.1). The assembled cp genomes were annotated using PGA (Plastid Genome Annotator) (Qu et al. 2019). To validate the assembly, PCR amplifications and sanger sequencing were performed to confirm the four junction regions between inverted repeat (IRs) and large single-copy region (LSC)/ small single copy region (SSC) and the region with great difference with the reference. Then, the start/stop codons and intron/exon boundaries of genes were subsequently manually modified based on the reference sequences, and the online program OGDRAW (OrganellarGenomeDRAW) (Greiner et al. 2019) was used to generate the graphical genome map of the cp genomes.

The full length of *C. edulis* cp genome (GenBank accession number: MW110633) was 154,395 bp and comprised of an LSC (82,391 bp), an SSC (19,504 bp), and two IRs (26,250 bp, each). And the complete cp genome of *C. shensiana* (GenBank accession number: MW110634) was 155,935 bp in length and contained two IRs (26,344 bp, each), an LSC (82,752 bp), and an SSC (20,495 bp). The overall GC content of *C. edulis* and *C. shensiana* cp genomes were 40.24% and 40.57%, respectively. Both of the two cp genomes contained 130 genes, including 85 protein-coding genes (*ycf1* and *ycf2* are two pseudogenes, and *rps16* or *clpP* are partial sequence), 8 ribosomal RNA genes, 37 transfer RNA genes. Of those protein-coding genes, 9 (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*,

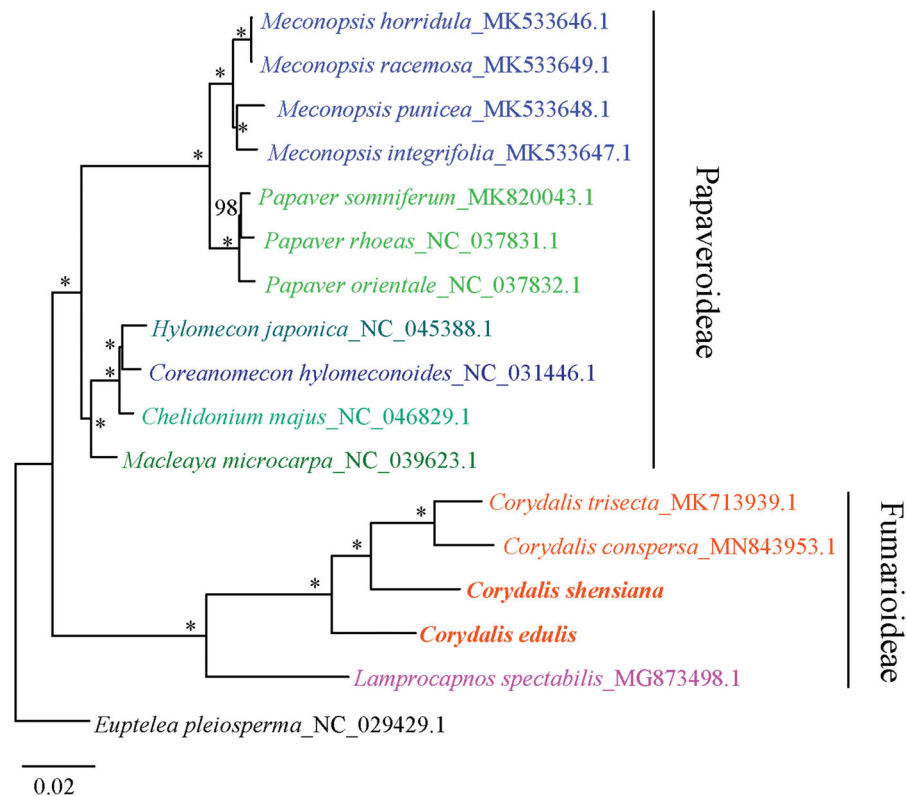


Figure 1. ML tree inferred from the coding sequences of 16 chloroplast genomes in Papaveraceae. Numbers above each node indicate bootstrap values. The asterisks show nodes supported by a bootstrap value of 100%.

rpoC1, and *rps16*) contained one intron and 3 (*clpP*, *rps12*, and *ycf3*) contained two introns. The overall structure, gene content, and arrangement of the cp genomes of *C. edulis* and *C. shensiana* were quite similar to but with higher quality than two previously reported *Corydalis* species, in which several subunits of NADH-dehydrogenase genes were absent or with partial sequence (such as *ndhC*, *ndhD*, *ndhF*, and *ndhI*) (Kanwal et al. 2019).

Sixteen cp genomes of Papaveraceae were fully aligned with MAFFT v7.3 (Katoh and Standley 2013), and the maximum-likelihood (ML) tree was constructed using all coding sequences under the GTRGAMMA model with 100 bootstrap replicates by RAxML v.8.2.1 (Stamatakis 2014). *Euptelea pleiosperma* J. D. Hooker & Thomson was chosen as an outgroup. The result showed that the four *Corydalis* species (*C. edulis*, *C. shensiana*, *C. trisecta* and *C. conspersa*) are strongly supported as monophyletic (Figure 1).

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 [32000170, 31800179] and the Foundation of He'nan Educational Committee [19B180004].

ORCID

Jia-Mei Li  <http://orcid.org/0000-0003-3434-6141>

Data availability statement

The assembled cp genomes of this study are openly available in NCBI at <https://www.ncbi.nlm.nih.gov/WebSub/?form=history&tool=genbank>, reference number (MW110633 and MW110634). The raw data that support the findings of this study are available on request from the first author LYY. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

References

- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47(W1):W59–W64.
- Kanwal N, Zhang X, Afzal N, Yang J, Li ZH, Zhao GF. 2019. Complete chloroplast genome of a Chinese endemic species *Corydalis trisecta* Franch. (Papaveraceae). *Mitochondrial DNA Part B.* 4(2):2291–2292.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4): 772–780.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Meth.* 15:50.
- Ren FM, Wang YW, Xu ZC, Li Y, Xin TY, Zhou JG, Qi YD, Wei XP, Yao H, Song JY. 2019. DNA barcoding of *Corydalis*, the most taxonomically complicated genus of Papaveraceae. *Ecol Evol.* 9(4):1934–1945.
- Rogers SO, Bendich AJ. 1988. Extraction of DNA from plant tissues. *Plant Mol Biol.* A6:1–10.
- Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9):1312–1313.
- Wang YW. 2006. Study on the phylogenetic of *Corydalis* [Ph.D. Dissertation]. Beijing (China): Institute of Botany, Chinese Academy of Sciences.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76(3–5):273–297.
- Zhang ML, Su ZY, Liden M. 2008. *Corydalis* DC. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China*. Vol. 7. Beijing (China): Science Press; St. Louis (MO): Missouri Botanical Garden Press. p. 295–428.