

Chloroplast genome of *Corydalis impatiens* (Pall.) Fisch. ex DC. (Papaveraceae), a Tibetan medical herb

Digao Wan^a, Haijuan Bao^b, Qupei Danzeng^c, Xiao Guo^a and Qien Li^a

^aTibetan Medicine Research Center, Tibetan Medical College, Qinghai University, Xining, People's Republic of China; ^bCollege of Ecological Environment and Resources, Qinghai Minzu University, Xining, People's Republic of China; ^cUniversity Tibetan Medicine, Lhasa, People's Republic of China

ABSTRACT

Corydalis impatiens (Pall.) Fisch. 1821. (Papaveraceae) is a Tibetan medical herb used to reduce pain, treat skin injuries, cure hepatitis, and benefit the circulatory system. In the current study, the chloroplast genome of *C. impatiens* was sequenced. This complete genome is a circular 197,317 bp sequence consisting of a small single-copy (SSC, 3105 bp) region, a large single-copy (LSC, 89,790 bp) region, and a pair of inverted repeats (IRs, 52,211 bp). This chloroplast genome encodes a total of 127 functional genes, including 81 protein-coding, 38 transfer RNA, and eight ribosomal RNA genes. Furthermore, this chloroplast genome contains six pseudogenes, including a pair of *ndhB* a pair of *ndhD*, one *ndhC*, and one *ndhK*. The phylogenetic relationship within the genus *Corydalis* was inferred with the maximum-likelihood method, and the result showed that *C. impatiens* was most closely related to *C. conspersa*.

ARTICLE HISTORY

Received 5 April 2022
Accepted 17 July 2022

KEYWORDS

Chloroplast genome;
Corydalis impatiens (Pall.)
Fisch 1821; evolutionary
analysis; Papaveraceae;
Tibetan medical herb

Corydalis impatiens (Pall.) Fisch., 1821. (Papaveraceae) is an annual herb of the family Fumariaceae, and it grows in foothills, crevices in rocks, shady slopes under forests and thickets. *C. impatiens* also known as Pa Xia Ga, is a Tibetan medical herb (Pan et al. 2019). This traditional medicine can reduce pain, treat skin injuries, cure hepatitis, and benefit the circulatory system (Niu et al. 2013). A clear understanding of chloroplast genome information not only contributes to species identification, phylogenetic analysis and molecular breeding, but also provides a molecular basis for important cash crops, horticultural variety improvement and conservation of rare and endangered plants (Moore et al. 2010; Nie et al. 2012; Shaw et al. 2014; Zhu et al. 2021). At present, although there are some studies on the medicinal value of *C. impatiens* (Niu et al. 2013), there are few studies on its genetics. Here, the chloroplast genome of *C. impatiens* was sequenced, and a phylogenetic analysis of *C. impatiens* and its allies was carried out.

Fresh young leaves of *C. impatiens* were collected from Huangzhong County, Qinghai Province, China (36.27°N, 101.68°E). The voucher specimen (specimen accession number: LQE-2020-070) was deposited in the Specimen Room of the Tibetan Medicine Research Center of Qinghai University (<https://zyxy.qhu.edu.cn/jgsw/jxkysw/zyyyjzx/index.htm>, Qien Li, qienli@qhu.edu.cn). Total DNA was extracted from Silica gel dried young leaves with a Plant Genomic DNA Kit (DP305, TIANGEN Biotech (Beijing) Co., Ltd., Beijing, China). Qualified DNA fragmentation was carried out by Ultrasonic

Processor, the length of insert fragment was approximately 350 bp. Then terminal repair, add base A, add sequence adapter, purification, PCR amplification were implemented to complete the 350 bp library preparation. Whole-genome sequencing was conducted by Novogene Co., Ltd. (Tianjin, China) with the Illumina NovaSeq 6000 Sequencing System (Illumina, San Diego, CA). Approximately, 15.8 GB of clean data were generated. SPAdes version 3.10.1 (Bankevich et al. 2012) and SSPACE version 2.0 (Boetzer et al. 2011) were used to assemble the chloroplast genome with default settings. This process does not use a reference genome. The assembled complete chloroplast genome was annotated with CPGAVAS2 (<http://www.herbalgenomics.org/cpgavas2>) (Shi et al. 2019) and the sequence coordinates for the genes were verified by BLAST search against the *C. inopinata* (GenBank accession number: NC_052866.1) chloroplast genome.

The assembled *C. impatiens* chloroplast genome was 197,317 bp in length with a GC content of 40.68%. As seen in most chloroplast genomes, this complete genome showed a typical quadripartite structure comprising a pair of inverted repeats (IRs 52,211 bp), one large single-copy (LSC, 89,790 bp) region, and one small single-copy (SSC, 3,105 bp). The chloroplast genome of *C. impatiens* encodes a total of 127 functional genes, including 81 protein-coding, 38 transfer RNA, and eight ribosomal RNA genes, accounting for 63.78%, 29.92%, and 6.30% of all annotated functional genes, respectively. Moreover, this chloroplast genome contains six

CONTACT Qien Li  qienli@outlook.com; Xiao Guo  gx_139417@163.com  Tibetan Medical College, Qinghai University, 251 Ningda Road, Xining, Qinghai Province 810016, People's Republic of China

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

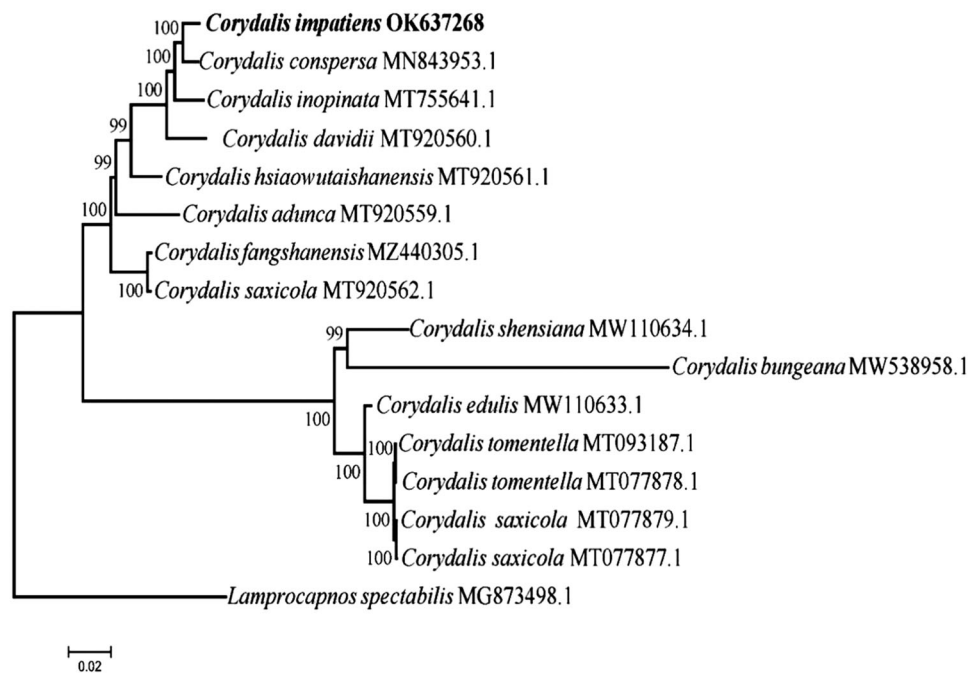


Figure 1. Maximum-likelihood (ML) tree of 16 species based on the complete chloroplast sequences. Numbers above branches are bootstrap percentages (based on 500 replicates).

pseudogenes, namely a pair of *ndhB*, a pair of *ndhD*, one *ndhC*, and one *ndhK*.

The phylogenetic relationship within the genus *Corydalis* was inferred with the maximum-likelihood (ML) method based on the Tamura–Nei model (Tamura and Nei, 1993), with *Lamprocapnos spectabilis* used as the out-group. In particular, a total of 16 complete chloroplast sequences of *C. impatiens* and its allies were aligned by MAFFT version 7.475 (Kato and Standley 2013) using default settings. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood approach, and then topology with superior log likelihood value was selected. The tree was drawn to scale, with branch lengths indicating substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 90,759 positions in the final dataset. Phylogenies were generated using the MEGA7 program based on general time-reversible (GTR)/GTR+I+G nucleotide substitution models of ML (Kumar et al. 2016), and the strengths of phylogenies were evaluated by resampling with 200 bootstrap replications. The phylogenetic analysis showed that *C. impatiens* was most closely related to *C. conspersa* (Figure 1).

The present study enriched the understanding of the maternal genetic information of *C. impatiens*, and provided additional data for reconstructing species relationships within the genus *Corydalis*.

Ethical approval

The materials used in this study are not included in the IUCN Red List of Threatened Species or the List of State-protected

Plant Species, and the sampling site is not located in any protected area. The field study and laboratory study were conducted in accordance with guidelines provided by Qinghai University.

Author contributions

Digao Wan and Qien Li were involved in the conception and design; Haijuan Bao, Qupei Danzeng, and Xiao Guo analyzed and interpreted the data; Digao Wan drafted the paper; Xiao Guo and Qien Li revised it critically for intellectual content. All authors approved the final version to be published, and agreed to be accountable for all aspects of the work.

Disclosure statement

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

Funding

This research was supported by the Natural Science Foundation of Qinghai Science & Technology Department [2019-ZJ-907] and the Collaborative Innovation Project of Tibetan Medicine in Tibet Autonomous Region [2019XTCX003].

Data availability statement

The genome sequence data obtained in this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov/> under the accession number OK637268. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA776704, SRX12869145, and SAMN22814974, respectively.

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 27(4):578–579.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- 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.
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A.* 107(10):4623–4628.
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Song W. 2012. Complete chloroplast genome sequence of a major invasive species, Crofton weed (*Ageratina adenophora*). *PLOS One.* 7(5):e36869.
- Niu XF, Xu HB, Liu X, Fan T, Qi L. 2013. Isoquinoline alkaloids from *Corydalis impatiens*. *Chem Nat Compd.* 49(1):187–189.
- Pan G, Shen J, Ma Y, He Y, Bao Y, Li R, Wang S, Wang Q, Lin P, Dang J. 2019. Preparative separation of isoquinoline alkaloids from *Corydalis impatiens* using a middle-pressure chromatogram isolated gel column coupled with two-dimensional liquid chromatography. *J Sep Sci.* 42(20):3182–3190.
- Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot.* 101(11):1987–2004.
- 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.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10(3):512–526.
- Zhu B, Qian F, Hou Y, Yang W, Cai M, Wu X. 2021. Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PLOS One.* 16(3):e0248556.