The complete chloroplast genome and phylogenetic analysis of *Cardamine circaeoides* Hook. f. et Thoms., 1861 (Brassicaceae)

Ru Wanga,b, Zhijun Denga,b,c and Yongjian Luoa,b

aHubei Key Laboratory of Biologic Resources Protection and Utilization, Hubei Minzu University, Enshi, P. R. China; bResearch Center for Germplasm Engineering of Characteristic Plant Resources in Enshi Prefecture, Hubei Minzu University, Enshi, P. R. China; cCenter for Crop Germplasm Resources, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, P. R. China

**ABSTRACT**

A species of *Cardamine circaeoides* Hook. f. et Thoms., 1861, which belongs to the Cardamine family (Brassicaceae), is an endemic species of the Wuling Mountains in Hunan and Hubei Provinces of China. Since there are many morphologically related species of *C. circaeoides*, the chloroplast (cp) genome characteristics of *C. circaeoides* were analyzed in order to explore the phylogenetic relationship between it and the closely related species. A cp genome totaling 154,838 base pairs exhibited a typical quadripartite structure with a pair of IRs (inverted repeats; 26,493 base pairs) separated by a small single-copy region of 17,938 base pairs and a large single-copy region of 83,914 base pairs. A total of 130 genes were found in the cp genome, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. There were 36.21%, 33.96%, 29.09%, and 42.35% GC content in the entire cp genome, LSC region, SSC region, and IR region, respectively. According to phylogenetic analysis, *C. circaeoides* is evolutionarily closer to *Cardamine hupingshanensis*.

*Cardamine circaeoides* belongs to the Brassicaceae family. The species is primarily found in the high mountains, valleys, and streams of the Wuling mountains of Hunan and Hubei Provinces, China (Wu et al. 2020). Besides being rich in protein, soluble sugar, amino acids, cellulose, and other nutrients, it also provides a super enrichment of selenium (Tian and Cheng 2003). In the case of selenium supplement cultivation, organic selenium concentrations can reach 3553 mg/kgDW (Both et al. 2018) and its tender stems and leaves can be eaten. In the field survey, local residents often refer to it as ‘wild rape’, ‘water rape’, or ‘rock plate vegetable’, which is a highly nutritious wild vegetable. As technology continues to develop rapidly, scientists began to pay attention to the application of modern molecular techniques in order to characterize the genetic diversity of shepherd’s purses. It has been demonstrated in several research studies that chloroplast (cp) genome sequences play an important role in plant phylogeny and genetic analysis (Wang et al. 2016; Ma et al. 2021; Yu et al. 2021). *Cardamine circaeoides*’s cp genome has been reported for the first time, and its structural characteristics have been analyzed. This study provides important genetic and phylogenetic information for understanding this species’ genetic relationships.

This study examined the fresh leaves of *C. circaeoides* collected from Enshi City, Hubei Province, China (30.301 N, 109.487 E, Altitude: 1236 m). A specimen was deposited at the Guangdong Agro-biological Gene Research Center (http://multi-omics.agrogene.ac.cn/, contact person: Yongjian Luo, and email: 851022933@qq.com) under the voucher number 20210303005. An extraction of genomic DNA from dried silica leaves was performed using Qiagen’s DNeasy Plant Mini Kit. The 1% agarose gel and Thermo Scientific Nanodrop spectrophotometer were used to confirm the quality and quantity of the samples (Figure 1). A high-throughput sequencing platform (HiSeq 4000) was used to sequence the complete genomes of *C. circaeoides*. The paired-end sequencing reads were then assembled using the software package *u HM_Chip* (Cao et al. 2015). A total of 183 genes were identified through a comparison of ribosomal RNA genes and trn-L, trn-R, trn-S, trn-F, trn-C, and trn-G genes in the plastome. The plastome of *C. circaeoides* contains 42 rRNA genes (8 copies of the 16S rRNA gene and 4 copies of the 23S rRNA gene), 42 tRNA genes (8 copies of the trn-L, trn-R, trn-S, trn-F, trn-C, and trn-G genes), and 21 genes coding for protein synthesis, including 18 protein-coding genes (8 copies of the 8S rRNA gene and 4 copies of the 23S rRNA gene) and 1 copy of the trn-G gene. The plastome of *C. circaeoides* contains 34,404 nucleotides, with a GC content of 36.21%. The LSC region contains 14,306 nucleotides, with a GC content of 36.21%, and the SSC region contains 13,898 nucleotides, with a GC content of 33.96%. The IR region contains 6,200 nucleotides, with a GC content of 29.33%.

**ARTICLE HISTORY**

Received 19 August 2022
Accepted 24 October 2022

**KEYWORDS**

*Cardamine circaeoides*; chloroplast genome; Brassicaceae; sequence alignment

**CONTACT**

Yongjian Luo 202030361@hbmzu.edu.cn School of Forestry and Horticulture, Hubei Minzu University, Enshi 445000, P. R. China; Zhijun Deng dengzhijun@hbmzu.edu.cn School of Forestry and Horticulture, Hubei Minzu University, Enshi 445000, P. R. 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.
sequence reads were trimmed to remove adaptors and low-quality sequences via Trimmomatic 0.39. Clean reads were assembled using GetOrganelle v1.5 (Jin et al. 2020). Chloroplast genome annotation was performed using CPGAVAS2 (http://47.96.249.172:16019/analyzer/home) (Shi et al. 2019) and GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al. 2017). An accession number (OL634846) has been assigned to the annotated cp genomic sequence in GenBank.

There are 154,838 base pairs in the cp genome of *C. circaeoides* (Figure 2). In the circular cp genome of *C. circaeoides*, a quadripartite structure was observed with an LSC (83,914 bp), an SSC (17,938 bp), and two IR regions (IRA and IRB, each of 26,493 bp). A total of 36.21% of *C. circaeoides‘* cp genome is composed of GCs (LSC, 33.96%; SSC, 29.09%; IR, 42.35%). GC content and gene order were similar to those of *Cardamine amaraeform* (Raman et al. 2021). A total of 130 genes were found in the cp genome, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. The introns were present in the following 12 protein-coding genes (atpF, clpP1, ndhA, ndhB, paf, petB, petD, rpl16, rpl2, rpoC1, rps16), and 6 tRNA genes (tRNA-UGC, tRNA-UCC, tRNA-GAU, tRNA-UGU, tRNA-UAA, and tRNA-UUC). In *C. circaeoides*, 13 cis-clipping genes including *rps16, atpF, rpoC1, paf, clpP1, petB, petD, rpl16, rpl2, ndhB, ndhA*, and a trans-splicing gene *rps12* were detected by CPGview-RSG (http://www.herbalgenomics.org/cpgview/). In addition, 95 SSRs were identified, including 89 single nucleotides (A/C/T, 93.68%) and 5 double nucleotides (AT/TA, 6.32%). At the same time, 23 tandem repeats were identified, with a total length of 15–42 bp. 33 dispersed repeats were identified with a threshold of 1E-4, and the length of repeat units ranged from 30 to 67.

Figure 2. Graphic representation of features identified in *C. circaeoides* plastome by using CPGview-RSG (http://www.herbalgenomics.org/cpgview). The map contains seven circles. From the center going outward, the first circle shows the distributed repeats connected with red (the forward direction) and green (the reverse direction) arcs. The next circle shows the tandem repeats marked with short bars. The third circle shows the microsatellite sequences as short bars. The fourth circle shows the size of the LSC and SSC. The fifth circle shows the IRA and IRB. The sixth circle shows the GC contents along the plastome. The seventh circle shows the genes having different colors based on their functional groups.
For the purpose of analyzing the phylogenetic placement of C. circaeoides within Cardamine, we used the cp genome sequence of C. circaeoides and 31 other related Brassicaceae species, Carica papaya, and Tarenaya hassleriana interior as an outgroup to construct the phylogenetic tree. In order to align the sequences, MAFFT v7.450 was used with the default parameters (Katoh and Standley 2013). A maximum likelihood (ML) tree was then constructed using MEGE version X with a Generalized Time-Reversible (GTR) model, with 1000 bootstrap replications. According to the phylogenetic tree (Figure 3), the species of the Cardamine genera are monophyletic. Phylogenetic analysis demonstrated that C. circaeoides is closely related to Cardamine hupingshanensis with a strong bootstrap value (98.8% for ML).

**Ethical approval**

Neither the samples used in this study nor the sampling site are listed on either the IUCN Red List of Threatened Species or the List of State-protected Plant Species. Both the field study and laboratory study were conducted in accordance with the guidelines established by Hubei Minzu University.

**Authors’ contributions**

Yongjian Luo and Zhijun Deng designed the experiment and approved the final version of the paper. Zhijun Deng and Ru Wang wrote and revised the paper; Zhijun Deng Yongjian Luo and Ru Wang analyzed and interpreted the data. All authors are responsible for all aspects of their work.

**Disclosure statement**

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

**Funding**

We acknowledge support from the National Natural Science Foundation of China [31860073, 81303169], and the Open Fund of Hubei Key Laboratory of Biologic Resources Protection and Utilization [PT012008].

**ORCID**

Yongjian Luo [http://orcid.org/0000-0002-3788-524X]

**Data availability statement**

These findings are supported by genome sequence data available in GenBank of NCBI (https://www.ncbi.nlm.nih.gov/) under accession number OL634846, associated BioProject, SRA, Bio-Sample numbers are PRJNA867123, SRR20954704, and SAMN30184410, respectively.

**References**

Both EB, Shao S, Xiang J, Jókai Z, Yin H, Liu Y, Magyar A, Derovics M. 2018. Selenolanthionine is the major water-soluble selenium compound in the selenium tolerant plant C. circaeoides. Biochim Biophys Acta Gen Subj. 1862(11):2354–2362.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21(1):241.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Ma C-X, Yan H-F, Ge X-J. 2021. The complete chloroplast genome of Lithocarpus hancei (Benth.) Rehd (Fagaceae) from Zhejiang, China. Mitochondrial DNA B Resour. 6(7):2022–2023.

Raman G, Park KT, Park S. 2021. The complete chloroplast genome of an endemic plant to Korea, Cardamine amaraeformis Nakai.: genome

Figure 3. Based on 34 chloroplast genome sequences, a maximum-likelihood phylogenetic tree was constructed for C. circaeoides. A bootstrap value is indicated by the number on each node.
structure and phylogenetic analysis. Mitochondrial DNA B Resour. 6(9):2725–2726.

Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVA52, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47(W1):W65–W73.

Tian L, Cheng L. 2003. Analysis of nutrients in Cardamine hupingxhanensis and its cultivation in foreign land. Sheng Ming ke Xue Yan Jiu. 7(2):167–172.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6–W11.

Wang L, Wuyun T, Du H, Wang D, Cao D. 2016. Complete chloroplast genome sequences of Eucommia ulmoides: genome structure and evolution. Tree Genet Gen. 12(1):12.

Wu M, Cong X, Li M, Rao S, Li Y, Guo J, Zhu S, Chen S, Xu F, Cheng S, et al. 2020. Effects of different exogenous selenium on Se accumulation, nutrition quality, elements uptake, and antioxidant response in the hyperaccumulation plant. Ecotoxicol Environ Saf. 204:111045.

Yu Y, Ouyang Z, Guo J, Zeng W, Zhao Y, Huang L. 2021. Complete chloroplast genome sequence of Erigeron breviscapus and characterization of chloroplast regulatory elements. Front Plant Sci. 12:758290.