

The complete chloroplast genome sequence of *Tussilago farfara* (Asteraceae)

Duan Yizhong^{a,b}, Ke Lu^b, Du Zhongyu^{b,c*} and Yehua Shen^a

^aCollege of Chemistry and Materials Science, Northwest University, Xi'an, China; ^bShaanxi Key Laboratory of Ecological Restoration in Northern Shaanxi Mining Area, Yulin University, Yulin, China; ^cBreeding Base for State Key Laboratory of Land Degradation and Ecological Restoration in Northwest China, Ningxia University, Yinchuan, China

ABSTRACT

Tussilago farfara is a member of the family Asteraceae. In this paper, we reported the complete chloroplast (cp) genome sequence of *T. farfara*. The results showed that *T. farfara* complete chloroplast genome comprises 151,325 bp, containing a large single copy (LSC) region of 83,370 bp, a small single copy (SSC) region of 18,273 bp, and a pair of inverted repeats (IRs) region of 24,841 bp. The genome has a GC content of 37.4%. The LSC, SSC, and IR regions represent 35.5%, 30.6%, and 43.0% of the *T. farfara* chloroplast genome length. We annotated 132 genes, comprising 87 protein-coding genes (PCGs), eight rRNA genes, and 36 tRNA genes. A phylogenetic analysis based on 31 cpDNA genomes suggested that the *T. farfara* is closely related to *Farfugium japonicum*.

ARTICLE HISTORY

Received 27 September 2021
Accepted 6 November 2021

KEYWORDS

Tussilago farfara; complete chloroplast genome; phylogenetic analysis

Tussilago farfara (Linnaeus 1753) belongs to the genus *Tussilago* of the family Asteraceae, it is an important medicinal material, flower buds and leaves are used as medicine to relieve cough, moisten lung and reduce phlegm (Song et al. 2021). It is also a honey sea plant, which is widely distributed in medicinal nurseries, valleys, wetlands, and forests all over China (Fedina et al. 2020). Complete chloroplast (cp) genome is the theoretical basis and experimental material for studying speciation, genetic diversity of germplasm resources, and gene engineering technology of higher plants (Lu et al. 2020). At present, it mainly focuses on physiological and biochemical aspects (Song et al. 2020). In this study, we reported the cpDNA genome sequences of *T. farfara*. We hope to provide a theoretical basis for the chloroplast genome characteristics and the phylogenetic relationship of this species.

The fresh leaves of *T. farfara* were provided from Yulin Forestry Industry Center, Shaanxi, China (109°44'20.31"E, 38°14'58.58"N) in July 2021. The specimens of *T. farfara* (Accession Number: 20210910YI02) were deposited at the Herbarium of Yulin University (<https://www.yulinu.edu.cn/>, Duan Yizhong, duanyizhong2006@163.com), Shaanxi, China. Total DNA was extracted from the fresh leaves according to a modified cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle 1987). The subsequent high-throughput sequencing was completed with the IlluminaHiSeq Ten system. The *Farfugium japonicum* (MT929248) cpDNA genome was used as the reference sequence to annotate. The cpDNA genome of *T. farfara* was annotated using the Geneious 8.0.2

(Kearse et al. 2012). The physical map of *T. farfara* was visualized by OGDRAW online tool (Lohse et al. 2013). We aligned all 31 cpDNA genome sequences used MAFFT (Kazutaka et al. 2002) and the MEGA 7.0 program (Kumar et al. 2016) to construct a phylogenetic tree according to the Neighbour-joining (NJ) method, with a bootstrap value of 1000. The annotated *T. farfara* cpDNA genome sequence has been deposited into the GenBank database (accession number: MW760850).

Our result showed that complete chloroplast genome of *T. farfara* is 151,325 bp, similar with the chloroplast structure of most other plants (Shinozaki et al. 1986), with a typical quadripartite structure containing a large single copy (LSC) region of 83,370 bp, a small single copy (SSC) region of 18,273 bp, and a pair of inverted repeats (IRs) region of 24,841 bp. The genome has a GC content of 37.4%. The LSC, SSC, and IR regions represent 35.5%, 30.6% and 43.0% of the *T. farfara*. We annotated 132 genes, including 87 protein-coding genes, 8 rRNAs, and 36 tRNAs. We constructed a phylogenetic tree based on 31 cpDNA genome sequences, with the closely related two species of *Nymphoides*, *Nymphoides coreana* (NC 041481) and *Nymphoides coronata* (NC 041484) cpDNA genome serving as an outgroup (Figure 1). The result showed that the *T. farfara* (MW760850) is closely related to *Farfugium japonicum* (MT929248).

Disclosure statement

The authors declare that they have no competing interests.

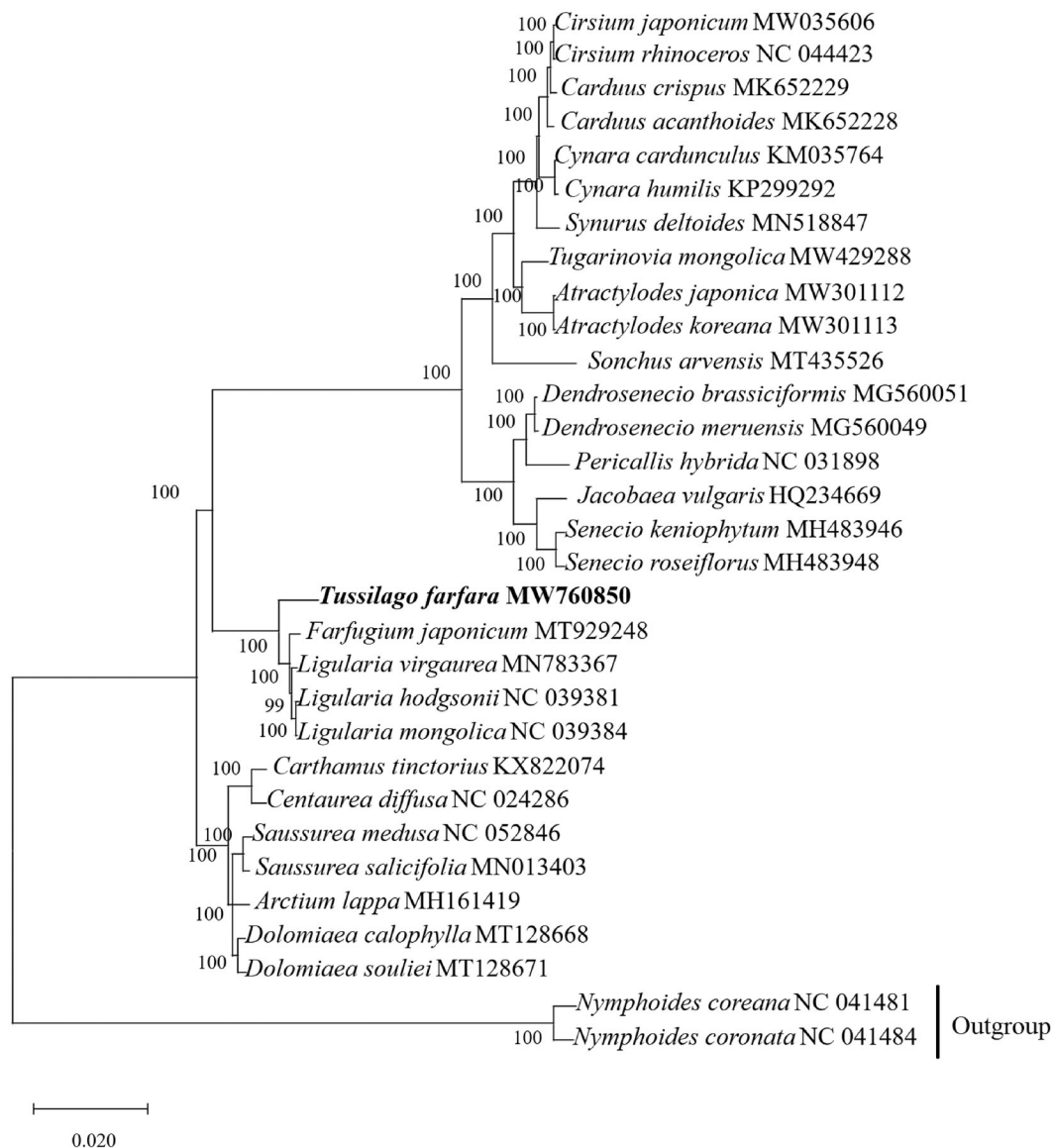


Figure 1. Phylogenetic tree constructed based on 31 species of chloroplast genome sequences. Accession numbers: *Cirsium japonicum* (MW 035606); *Cirsium rhinoceros* (NC 044423); *Carduus crispus* (MK652229); *Carduus acanthoides* (MK652228); *Cynara cardunculus* (KM035764); *Cynara humilis* (KP299292); *Synurus deltoides* (MN518847); *Tugarinovia mongolica* (MW429288); *Atractylodes japonica* (MW301112); *Atractylodes koreana* (MW301113); *Sonchus arvensis* (MT435526); *Dendrosenecio brassiciformis* (MG560051); *Dendrosenecio meruensis* (MG560049); *Pericallis hybrida* (NC 031898); *Jacobaea vulgaris* (HQ234669); *Senecio keniophytum* (MH483946); *Senecio roseiflorus* (MH483948); *Tussilago farfara* (MW760850); *Farfugium japonicum* (MT929248); *Ligularia virgaurea* (MN783367); *Ligularia hodgsonii* (NC 039381); *Ligularia mongolica* (NC 039384); *Carthamus tinctorius* (KX822074); *Centaurea diffusa* (NC 024286); *Saussurea medusa* (NC 052846); *Saussurea salicifolia* (MN013403); *Arctium lappa* (MH161419); *Dolomiaea calophylla* (MT128668); *Dolomiaea souliei* (MT128671); *Nymphoides coreana* (NC 041481); *Nymphoides coronata* (NC 041484).

Funding

This study was supported by National Natural Science Foundation of China [Nos. 21675125, 32060095].

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>] (<https://www.ncbi.nlm.nih.gov/>) under the accession no. MW760850. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA765479, SRR16017860, and SAMN21561337 respectively.

References

- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phy Bull.* 19:11–15.
- Fedina LA, Kuprin AV, Ogorodniko VEM. 2020. *Tussilago farfara* (Asteraceae) in the South of the Far East of Russia. *Russ J Biol Invasions.* 11(1):88–91.
- Kazutaka K, Kazuharu M, Kei-Ichi K, Takashi M. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12):1647–1649.

- 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.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41(Web Server issue):W575–W581.
- Lu K, Mao W, Du ZY, He YM, Fan C, Zhang K, Wang LJ, Liu GM, Duan YZ. 2020. The complete chloroplast genome sequence of *Artemisia ordosica*. *Mitochondrial DNA B.* 5(2):1663–1664.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *Embo J.* 5 (9):2043–2049.
- Song XQ, Sun J, Yu JH, Zhang JS, Bao J, Zhang H. 2020. Prenylated indole alkaloids and lignans from the flower buds of *Tussilago farfara*. *Fitoterapia.* 146:104729.
- Song XQ, Yu JH, Sun J, Liu KL, Zhang JS, Zhang H. 2021. Bioactive sesquiterpenoids from the flower buds of *Tussilago farfara*. *Bioorg Chem.* 107:104632.