

Characterization and phylogenetic analysis of the complete chloroplast genome of *Cercis canadensis* 'Forest Pansy'

Lijuan Feng^a, Jihan Tao^a, Xuemei Yang^a, Qiqing Jiao^a, Chuanzeng Wang^b, Yun Cheng^c and Yanlei Yin^a

^aShandong Institute of Pomology, Taian, Shandong, China; ^bShandong Academy of Agricultural Sciences, Jinan, Shandong, China; ^cLiaocheng Natural Resources and Planning Bureau, Liaocheng, Shandong, China

ABSTRACT

Cercis canadensis 'Forest Pansy' is a tree species with high ornamental value, which complete chloroplast (cp) genome was sequenced, assembled, and annotated. The genome size is 158,960 bp with a total GC content of 36.17%. The cp genome is made up of a large single-copy region (88,114 bp), a small single-copy region (19,590 bp), and two inverted repeat regions (25,628 bp each). It contains 128 genes, including 84 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. Eighteen genes were duplicated in IRs. The maximum-likelihood (ML) phylogenetic analysis indicated that the Leguminosae species are grouped together, and *C. canadensis* 'Forest Pansy' is closely related to *C. canadensis*. The result would provide valuable information for genetic studies on *Cercis* genus.

ARTICLE HISTORY

Received 11 November 2019
Accepted 23 November 2019

KEYWORDS

Cercis canadensis; Forest Pansy; chloroplast genome; phylogenetic relationship

'Forest Pansy' is a variant of the *Cercis canadensis*, which belong to *Cercis* of Leguminosae family (Davis et al. 2002). It is deciduous shrubs or small landscape tree, which exhibits considerable morphological diversity, including variation in plant architecture, plant size, and flower and leaf colors (Roberts et al. 2015). It is native to Canada, and is mainly distributed in the USA, China, and Africa (Zhang et al. 2009). It has higher ornamental value with beautiful pink flowers and purple leaves. It is widely cultivated in botanical gardens, parks, and on the roadside in China (Li et al. 2017). In this present study, we reported and characterized the complete chloroplast (cp) genome of Forest Pansy based on Illumina pair-end sequencing and compared it with other genus cp genome sequences. The result would supply valuable information for the evolution process and conservation genetics of *C. canadensis*.

The fresh leaves of *C. canadensis* 'Forest Pansy' (Voucher specimen Accession No. SDJNDZJ0069) was collected from the Taidong field of Shandong Institute of Pomology (36.20°N, 117.12°E), Shandong Province, China. Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands). cpDNA sequencing was performed with an Illumina HiSeq 2500 platform by Nanjing Genepioneer Biotechnologies (Nanjing, China). The raw paired-end reads of cpDNA were filtered using fastp program (Chen et al. 2018), and *de novo* assembly performed using GetOrganelle (Jin et al. 2018). The cp genome was annotated using the program DOGMA (Wyman et al. 2004), with the cp genome of *Cercis glabra* (GenBank Accession No. KY806281) (Wang et al. 2017) serving as the reference. A circular genome map of the genome was generated with OGDRAW (<http://ogdraw>.

mpimp-golm.mpg.de/) (Lohse et al. 2013). The complete cp genome was deposited in the GenBank (Accession: MN562096).

The complete cp genome of *C. canadensis* 'Forest Pansy' was 158,960 bp in length with a total GC content of 36.17%. It is shorter than the reference species *Cercis glabra* (159,181 bp) and exhibited a typical quadripartite structure with known *Cercis* cp genomes. The cp genome is made up of a large single-copy region (LSC) of 88,114 bp, a small single-copy region (SSC) of 19,590 bp, and two inverted repeat regions (IRs) of 25,628 bp each. The complete cp genome encoded 128 unique genes, which contained 84 protein-coding genes, 36 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes. The tRNA genes are distributed throughout the genome with 21 in the LSC, one in the SSC, and 14 in the IR regions, while rRNAs are only situated in the IR regions. There were 18 duplicated genes in IRs, including seven protein-coding genes (*rpl2*, *rpl23*, *ndhB*, *rps7*, *rps12*, *ycf2*, and *ycf15*), seven tRNA genes (*trnICAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnRACG*, and *trnN-GUU*) and four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*). Among the protein-coding genes, two genes (*clpP*, *ycf3*) contained two introns, and other eight genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpoC1*, and *rps12*) had one intron each.

To ascertain the phylogenetic position of *C. canadensis* 'Forest Pansy', 25 complete cp genomes within Leguminosae family were selected, *Fraxinus excelsior* (GenBank Accession No. NC_037446.1) and *Platanus occidentalis* (GenBank Accession No. DQ923116.1) as outgroup. The cp genomes of these species were aligned Using MAFFT v7.3 (Kazutaka and Standley 2013). The maximum-likelihood (ML) phylogenetic

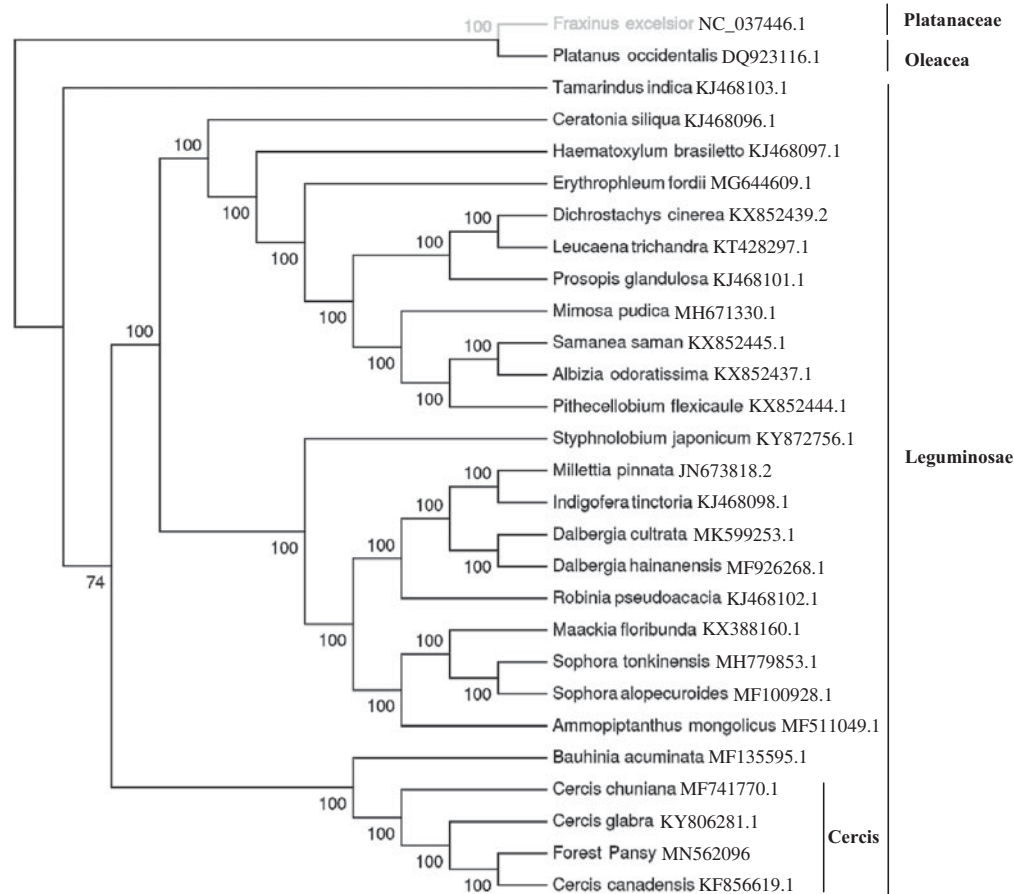


Figure 1. The best maximum-likelihood (ML) phylogenetic tree based on the 28 complete chloroplast genome sequences. The number on each node indicates bootstrap support value.

tree was constructed by the IQ-TREE with the best-fit model identified using ModelFinder (Kalyaanamoorthy et al. 2017). The result showed that the Leguminosae species are grouped together, and *C. canadensis* 'Forest Pansy' is closely related to *C. canadensis* (Figure 1). It will be valuable for the genetic study on *Cercis* genus.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Shandong Key Research and Development Projects [2018GNC111009, 2018JHZ003], Shandong Agricultural Improvement Seed Projects [2017LZN023], Agricultural Science and Technology Innovation Engineering Discipline Team of Shandong Academy of Agricultural Sciences [CXGC2018E22].

References

- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34(17):i884–i890.
- Davis CC, Fritsch PW, Donoghue L. 2002. Phylogeny and biogeography of *Cercis* (Fabaceae): evidence from nuclear ribosomal ITS and chloroplast *ndhF* sequence data. *Syst Bot*. 2:289–302.
- Jin JJ, Yu WB, Jun B, Yang JB, Song Y, Yi TS, Li DZ. 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *bioRxiv*. 4: 256479.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Meth*. 14(6):587–589.
- Kazutaka K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
- Li YM, Wang LM, Wang HJ, Meng YE, Zhang HC, Dong XY, Fu ZZ. 2017. Study on two-step rooting culture of *Cercis canadensis* 'Forest Pansy' seeding in vitro. *J Henan Agr Sci*. 46:114–117.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. Organellar Genome-DRAW: a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl Acids Res*. 41(W1):W575–W581.
- Roberts DJ, Werner DJ, Wadi PA, Trigiano RN. 2015. Inheritance and allelism of morphological traits in eastern redbud (*Cercis canadensis* L.). *Hortic Res*. 2(1):15049.
- Wang YH, Wang H, Yi TS, Wang YH. 2017. The complete chloroplast genomes of *Adenolobus garipensis* and *Cercis glabra* (Cercidoideae, Fabaceae). *Conserv Genet Resour*. 9(4):635–638.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 20(17):3252–3255.
- Zhang Y, Zhuo LH, Wang CZ. 2009. Effects of shade treatments on net photosynthetic characteristics and chlorophyll fluorescence parameters of *Cercis canadensis* 'Forest Pansy'. *J Agr Univ Hebei*. 32: 52–56.