

The complete chloroplast genome sequence of *Populus davidiana*, and a comparative analysis with other *Populus* species

Chong Sun^a, Xia Gong^b and Xia Liu^a

^aCollege of Landscape Architecture and Life Science, Chongqing University of Arts and Sciences, Chongqing, China; ^bSichuan Academy of Botanical Engineering, Sichuan, China

ABSTRACT

Populus davidiana plays an important ecological role in boreal and temperate forests, serving as wildlife habitats and watersheds. The complete chloroplast genome sequence of *P. davidiana* was characterized from Illumina pair-end sequencing. The chloroplast genome of *P. davidiana* was 155,325 bp in length, containing a large single-copy region (LSC) of 84,679 bp, a small single-copy region (SSC) of 16,862 bp, and two inverted repeat (IR) regions of 26,892 bp. The overall GC content is 36.80%, while the corresponding values of the LSC, SSC, and IR regions are 34.5%, 30.5%, and 42.5%, respectively. The genome contains 131 complete genes, including 86 protein-coding genes (62 protein-coding gene species), 37 tRNA genes (29 tRNA species), and 8 rRNA genes (4 rRNA species). The neighbour-joining phylogenetic analysis showed that *P. davidiana* and *P. hopeiensis* clustered together as sisters to other *Populus* species.

ARTICLE HISTORY

Received 14 May 2020
Accepted 16 May 2020

KEYWORDS

Populus davidiana;
chloroplast genome;
phylogenetic analysis;
genetic information

Introduction

Populus davidiana occurs in northern and central parts of China, plus Mongolia, Korea, and the Far East of Russia (Zheng et al. 2017; Hou et al. 2018). *Populus davidiana* is widely distributed in the Northern Hemisphere and plays an important ecological role in boreal and temperate forests, serving as wildlife habitats and watersheds; they can dominate riparian forests, but are ecologically adaptable. *Populus davidiana* has wide geographic distribution, high intraspecific polymorphism, adaptability to different environments, combined with a relatively small genome size. Consequently, *P. davidiana* represents an excellent model for understanding how different evolutionary forces have sculpted the variation patterns in the genome during the process of population differentiation and ecological speciation (Neale and Antoine 2011). Moreover, we can develop conservation strategies easily when we understand the genetic information of *P. davidiana*. In the present research, we constructed the whole chloroplast genome of *P. davidiana* and understood many genome variation information about the species, which will provide beneficial help for population genetics studies of *P. davidiana*.

The fresh leaves of *P. davidiana* were collected from Lijiang city (100°23'N, 26°88'E). Fresh leaves were silica-dried and taken to the laboratory until DNA extraction.

The voucher specimen (SY002) was laid in the Herbarium of Chongqing University of Arts and Sciences and the extracted DNA was stored in the –80 °C refrigerator of the Key Laboratory of College of Landscape Architecture and Life Science. We extracted total genomic DNA from 25 mg silica-gel-dried leaf using a modified CTAB method (Doyle 1987). The whole-genome sequencing was then conducted by Biodata Biotechnologies Inc. (Hefei, China) with Illumina HiSeq platform. The Illumina HiSeq 2000 platform (Illumina, San Diego, CA) was used to perform the genome sequence. We used the software MITObim 1.8 (Hahn et al. 2013) and metaSPAdes (Nurk et al. 2017) to assemble chloroplast genomes. We used *P. tremula* (GenBank: NC_027425) as a reference genome. We annotated the chloroplast genome with the software DOGMA (Wyman et al. 2004), and then corrected the results using Geneious 8.0.2 (Campos et al., 2016) and Sequin 15.50 (<http://www.ncbi.nlm.nih.gov/Sequin/>).

The complete chloroplast genome of *P. davidiana* (National Genomics Data Center accession number GWHAMJQ01000000) was characterized from Illumina pair-end sequencing. The complete chloroplast genome sequence of *P. davidiana* was characterized from Illumina pair-end sequencing. The chloroplast genome of *P. davidiana* was 155,325 bp in length, containing a large single-copy region (LSC) of 84,679 bp, a small single-copy region (SSC) of 16,862 bp, and two inverted repeat (IR) regions of 26,892 bp.

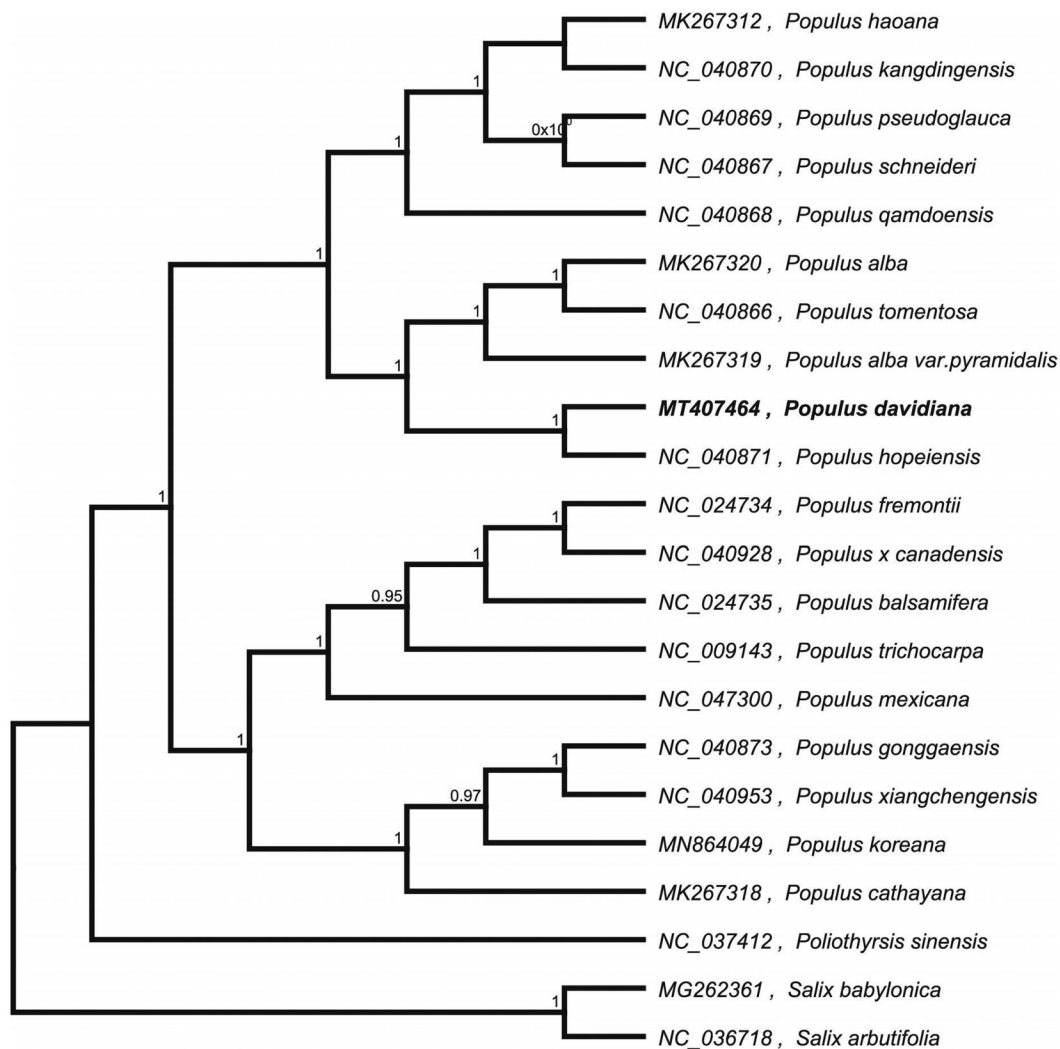


Figure 1. Neighbour-joining (NJ) analysis of *P. davidiana* and other related species based on the complete chloroplast genome sequence.

The overall GC content is 36.80%, while the corresponding values of the LSC, SSC, and IR regions are 34.5%, 30.5%, and 42.5%, respectively. The genome contains 131 complete genes, including 86 protein-coding genes (62 protein-coding gene species), 37 tRNA genes (29 tRNA species), and 8 rRNA genes (4 rRNA species).

To confirm the phylogenetic location of *P. davidiana* within the family of *Populus*, we used the complete chloroplast genomes sequence of *P. davidiana* and 21 other related species of *Populus* and *Salix babylonica* and *Salix arbutifolia* as outgroup to construct phylogenetic tree. The 22 chloroplast genome sequences were aligned with MAFFT (Kato and Standley 2013), and then the neighbour-joining tree was constructed by MEGA 7.0 (Kumar et al. 2016). The results confirmed that *P. davidiana* was clustered with *P. hopeiensis* (Figure 1).

Disclosure statement

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

Funding

Financial support for this research was provided by the Science and Technology Research Project of Chongqing Education Commission [KJQN201801336].

Data availability statement

The data that support the findings of this study are openly available National Genomics Data Center at <https://bigd.big.ac.cn/search?dbid=gwh&q=GWHAMJQ01000000>, accession number GWHAMJQ01000000.

References

- Campos FS, Kluge M, Franco AC, Giongo A, Valdez FP, Saddi TM, Brito WMED, Roehe PM. 2016. Complete genome sequence of porcine parvovirus 2 recovered from swine sera. *Genome Announc.* 4(1): e01627–01615.
- Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19(1):11–15.
- Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a

- baiting and iterative mapping approach. *Nucleic Acids Res.* 41(13): e129–e129.
- Hou Z, Wang Z, Ye Z, Du S, Liu S, Zhang J. 2018. Phylogeographic analyses of a widely distributed *Populus davidiana*: further evidence for the existence of glacial refugia of cool-temperate deciduous trees in northern East Asia. *Ecol Evol.* 8(24):13014–13026.
- 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.
- Neale DB, Antoine K. 2011. Forest tree genomics: growing resources and applications. *Nat Rev Genet.* 12(2):111–122.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27(5): 824–834.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20(17): 3252–3255.
- Zheng H, Fan L, Milne RI, Lei Z, Wang Y, Mao K. 2017. Species delimitation and lineage separation history of a species complex of aspens in China. *Front Plant Sci.* 8:375.