


Complete chloroplast genome sequence of *Populus euphratica* from PacBio Sequel platform

Zhen-Bo Jiang^{a,b,c}, Jun-Tuan Zhai^{a,b,c}, Pei-Pei Jiao^{a,b,c}, Tian-Ge Yang^d, Zhi-Hua Wu^{a,d}  and Zhi-Jun Li^{a,b,c}

^aXinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Tarim University, Alar, PR China; ^bDesert Poplar Research Center, Tarim University, Alar, PR China; ^cCollege of Life Science, Tarim University, Alar, PR China; ^dHubei Provincial Key Laboratory for Protection and Application of Special Plant Germplasm in Wuling Area of China & Key Laboratory of State Ethnic Affairs Commission for Biological Technology, College of Life Sciences, South-Central University for Nationalities, Wuhan, PR China;

ABSTRACT

Populus euphratica Oliv., one of tall arbors growing in desert areas, has great stress resistance. The complete chloroplast genome was reported in this study using the PacBio Sequel Platform. The chloroplast genome with a total size of 157,881 bp consisted of two inverted repeats (IRs) (27,666 bp) separated by a large single-copy region (85,906 bp) and a small single-copy region (16,643 bp). Further annotation revealed the chloroplast genome contains 111 genes, including 77 protein-coding genes, 30 *tRNA* genes, and four *rRNA* genes. The information of the chloroplast genome will be useful for study on the evolution of *P. euphratica* in the future.

ARTICLE HISTORY

Received 2 August 2020
Accepted 16 December 2020

KEYWORDS




Populus euphratica;
chloroplast genome; PacBio
Sequel; evolution

Populus euphratica Oliv., is the natural arbor species that can survive in the serious desert environments and exhibits remarkable resistance to environmental stresses (Lv et al. 2014). Due to its greater ability to cope with environmental stresses, *P. euphratica* is widely considered as an ideal model system when studying the molecular mechanisms of abiotic stress responses in woody species (Sun et al. 2009; Ding et al. 2010). In this study, to obtain the new insight into the evolution of *P. euphratica*, we sequenced, assembled, and annotated the accurate chloroplast genome with PacBio Sequel platform.

The materials of *P. euphratica* in this study were collected from *P. euphratica* forest in the headwater region of the Tarim River on the northwestern margin of the Tarim basin in Xinjiang province of China (81°17'56.52"E, 40°32'36.90"N, 980 m above sea level). The voucher specimens were deposited at the Herbarium of Tarim University (TD-00301). The leaves total genomic DNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) method and sequenced using the PacBio platform. The raw sequencing data (SRR12959747) generated 35,960 reads with the N50 of 10,213 bp. The whole Chloroplast genomes were assembled from whole genome sequencing data using Canu (Koren et al. 2017) and got 15 contigs with the N50 of 21,246 bp. To discard nuclear DNA sequences and obtain the complete chloroplast genome sequence, we aligned the contigs of a preliminary assembly to the whole chloroplast data from

NCBI. Then the draft genome was polished with Arrow (SMRT link-6.0.0, Pacific Biosciences, Menlo Park, CA). Due to the special structure of the chloroplast genome, we mapped the scaffolds to the reference to find the IR region and manually adjusted. Then annotated using CPGAVAS2 (Shi et al. 2019) and PGA (Qu et al. 2019). The complete chloroplast genome was 157,881 bp (MT818237) and composed of two inverted repeats (IRs) of 27,666 bp each, which divide a large single copy (LSC) region of 85,906 bp and a small single copy (SSC) region of 16,643 bp, the average GC content was 36.53%. The chloroplast genomes encoded 111 genes, including 77 protein-coding genes, 30 *tRNA* genes, and four *rRNA* genes.

According to the previously published chloroplast genome of *P. euphratica* from NCBI with Illumina platform (NC_024747), we aligned the *P. euphratica* chloroplast of Illumina and PacBio platforms using BLASTN. PacBio RS data can produce high-quality sequence assemblies covering a greater proportion of the genome than can be achieved by Illumina sequencing alone. We found that the cp genome got from PacBio platform was slightly longer. After designing the primers (5'-AATGTAGGATTAGCGGTTCT-3' and 5'-GCTGTATTCATGCCGTTTCG-3', 5'-TAACCTGCTCTGCTGGACT-3', and 5'-CTTGACTTGCTGCTTGCTT-3') for different places between the genome with two platforms, we verified the real existence of the insertion assembled by PacBio through Sanger. The result showed that the PacBio has the advantage

CONTACT Zhi-Hua Wu  zhwu@scuec.edu.cn  Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Tarim University, Alar, China; Hubei Provincial Key Laboratory for Protection and Application of Special Plant Germplasm in Wuling Area of China & Key Laboratory of State Ethnic Affairs Commission for Biological Technology, College of Life Sciences, South-Central University for Nationalities, Wuhan, China; Zhi-Jun Li  lizhijun0202@126.com Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Tarim University, Alar, China; Desert Poplar Research Center of Tarim University, Tarim University, Alar, China; College of Life Science, Tarim University, Alar, China

© 2021 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.

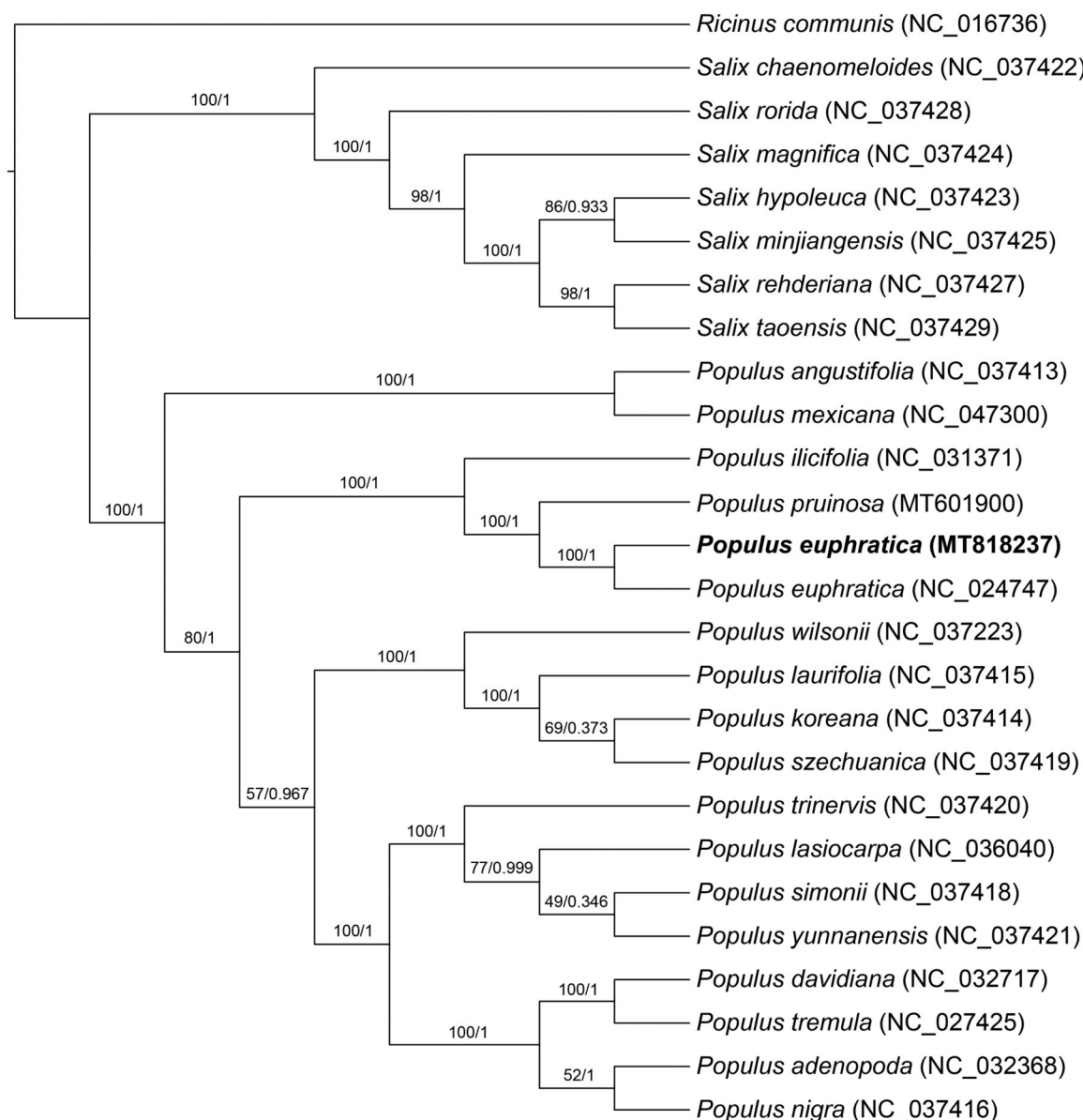


Figure 1. Phylogenetic tree reconstructed by maximum-likelihood (ML) and Bayesian inference (BI) analysis based on the whole chloroplast protein-coding genes of these 26 species.

of getting more complete chloroplast genome, which is also reported in other plants (Wu et al. 2014).

In our study, to explore the phylogenetic relationship of *P. euphratica* within Salicaceae, additional 25 species from Salicaceae were studied. With the species of *Ricinus communis* L. as the outgroup, the phylogenetic trees were built from the whole protein-coding gene matrix by maximum-likelihood (ML) and Bayesian inference (BI) (Figure 1). The ML tree was generated using IQ-TREE (Nguyen et al. 2015) based on the best model of TVM + F + R3 and 1000 bootstrap replicates, and BI analysis was performed in MrBayes version 3.2.7 (Ronquist et al. 2012). This result showed that the *P. euphratica* was closer to the species of *P. pruinosa*.

Disclosure statement

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

Funding

This work was financially supported by the National Natural Sciences Foundation of China [U1803231, 30660018] and the Innovative Team Building Plan for Key Areas of Xinjiang Production and Construction Corps [2018CB003], Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin [BRZD2003], and Hubei Provincial Natural Science Foundation of China [2019CFB214].

ORCID

Zhi-Hua Wu  <http://orcid.org/0000-0002-2895-7256>

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>] under the accession no. MT818237. The associated 'BioProject', 'SRA', and

'Bio-Sample' numbers are PRJNA673650, SRR12959747, and SAMN16619580, respectively.

References

- Ding M, Hou P, Shen X, Wang M, Deng S, Sun J, Xiao F, Wang R, Zhou X, Lu C, et al. 2010. Salt-induced expression of genes related to Na(+)/K(+) and ROS homeostasis in leaves of salt-resistant and salt-sensitive poplar species. *Plant Mol Biol.* 73(3):251–269.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736.
- Lv F, Zhang H, Xia X, Yin W. 2014. Expression profiling and functional characterization of a CBL-interacting protein kinase gene from *Populus euphratica*. *Plant Cell Rep.* 33(5):807–818.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Qu X-J, Moore MJ, Li D-Z, Yi T-S. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods.* 15(1):50.
- Ronquist F, Teslenko M, van der Mark P, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology.* 61(3):539–42.
- 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.
- Sun J, Chen S, Dai S, Wang R, Li N, Shen X, Zhou X, Lu C, Zheng X, Hu Z, et al. 2009. NaCl-induced alternations of cellular and tissue ion fluxes in roots of salt-resistant and salt-sensitive poplar species. *Plant Physiol.* 149(2):1141–1153.
- Wu Z, Gui S, Quan Z, Pan L, Wang S, Ke W, Liang D, Ding Y. 2014. A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MiSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. *BMC Plant Biol.* 14(1):289.