

Characterization of the complete *Salix viminalis* var. *gmelinii* Turcz 1854 chloroplast genome from the northeast of China

Jie Wan^{a,b}, Huijie Tang^{a,b}, Zicheng Yu^{a,b}, Jing Wang^{a,b} , Xu Yao^{a,b} and Xiaoping Li^{a,b,c}

^aCollaborative Innovation Center of Southern Modern Forestry, Nanjing Forestry University, Nanjing, China; ^bCollege of Forestry, Nanjing Forestry University, Nanjing, China; ^cKey Laboratory of Forest Tree Genetics and Breeding and High-Efficiency Cultivating in Jiangsu Province, Nanjing Forestry University, Nanjing, China

ABSTRACT

The *Salix viminalis* var. *gmelinii* Turcz 1854 is a variant of the *Salix* genus, from the Salicaceae family, and possesses an extremely high economic value. In this study the complete chloroplast genome of the woody plant *S. viminalis* var. *gmelinii* was characterized for the first time using a high-throughput approach in conjunction with *de novo* assembly technology. The *S. viminalis* var. *gmelinii* chloroplast genome is 155,405 base pairs (bp) in length and contains 36.71% GC content. It incorporates a large single-copy region (LSC, 84,287bp) alongside one small-copy region (SSC, 16,198bp), and two inverted repeat regions (IRA and IRB, 27,460bp). Moreover, this chloroplast genome encodes 128 genes, which comprises 83 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. Furthermore, the phylogenetic analysis revealed that *S. viminalis* var. *gmelinii* is closely related to *S. cupularis* and *S. gordejevii*.

ARTICLE HISTORY

Received 31 January 2022
Accepted 17 September 2022

KEYWORDS

Chloroplast genome;
phylogenetic tree;
Salicaceae; *Salix viminalis*
var. *gmelinii*

Salix viminalis var. *gmelinii* is an exuberant shrub species that belongs to the genus *Salix*, which is the largest genus of Salicaceae worldwide and contains approximately 500 species (Argus 1997). Mainly distributed in the provinces of Heilongjiang, Jilin, Liaoning, and the autonomous region of Inner Mongolia in the northeast of China, the *S. viminalis* var. *gmelinii* grows on the banks of rivers and streams as a re-vegetation tree. As a species, it provides many economical purposes, whereby its branches are used to produce weaving baskets, the leaves act to feed silkworms, while tannins are extracted from the bark. However, the characteristics of the *Salix* genus vary significantly between species and populations, while the similar natural hybrid phenomenon blurs the morphological boundaries between the species (Chen et al. 2010). Moreover, the genomic and related molecular genetic resources are limited, which makes it difficult to analyze. This study herein provides the complete chloroplast genome sequence of the *S. viminalis* var. *gmelinii*. The publication of which will both enhance the ability to undertake the phylogenetic study of the Salicaceae family and provide essential genetic information for the further development of *S. viminalis* var. *gmelinii*.

Fresh and healthy leaves of *S. viminalis* var. *gmelinii* were collected from Mao'er Mountain (Heilongjiang, China; coordinates: 45°23'N, 127°32'E), and the voucher specimen was deposited in the Biotechnology Building, Nanjing Forestry University, China (Li Xiaoping1, xpli@njfu.edu.cn,

DBNHL2017005). The genomic DNA of *S. viminalis* var. *gmelinii* was extracted using the CTAB method (Doyle and Doyle 1987). Whole genomic sequencing was performed by an Illumina NovaSeq 6000 series sequencer (PE150) (Illumina, San Diego, CA) following the completion of the sequencing library. A total of 2.61 G raw data were generated and then trimmed using Fastp (Chen et al. 2018). The chloroplast genome of *S. viminalis* var. *gmelinii* was assembled *de novo* using NOVOPlasty (version 4.1) (<https://github.com/ndierckx/NOVOPlasty>) (Dierckxsens et al. 2017). Subsequently, the assembled chloroplast genome sequence was annotated using an online GeSeq program (Tillich et al. 2017) alongside CPGAVAS 2 (Shi et al. 2019; Zhou et al. 2021) *S. gordejevii* (NC_058001.1) was used as the reference genome and manual correction was performed. After manual correction, the result annotation of the complete chloroplast DNA was submitted to NCBI online by BankIt (GenBank accession number: OK505606.1).

S. viminalis var. *gmelinii* represents a typical quadripartite structure, it is 155,405 bp in length and contains a small single-copy region (SSC, 16,198 bp), a large single-copy region (LSC, 84,287 bp) and a pair of inverted repeat regions (IR, 27,460 bp). The total GC content of the chloroplast genome is 36.71%. A total of 128 genes were annotated from the chloroplast genome sequence, including 83 protein-coding genes, 37 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes. Moreover, a total of 18 genes,



Figure 1. Maximum likelihood phylogenetic tree based on 22 selected plants chloroplast genome sequences.

including seven protein-coding genes, seven tRNAs, and four rRNAs were duplicated in the IR regions.

To reveal the phylogenetic position of *S. viminalis* var. *gmelinii* against the close relatives of Salicaceae, the phylogenetic analysis was performed based on 22 complete chloroplast genomes of Salicaceae (Figure 1). The complete chloroplast genomes were aligned using MAFFT v7.307 (<https://mafft.cbrc.jp/alignment/software/>) (Kato and Standley 2013). A phylogenetic tree was generated through a maximum likelihood (ML) analysis using MEGA 6.0 with 1000 bootstrap replicates and setting the *Arabidopsis thaliana* to the outgroup (Tamura et al. 2013). Comparison of the *Salix viminalis* var. *gmelinii* plastome to previously published data shows a high level of gene synteny with some publicly available *Salix* sequences (Zhou et al. 2021). And, there are three *Salix viminalis* plastomes deposited in the GenBank. They were placed in the phylogenetic tree for analysis and found that they did not form a monophyletic clade. The potential reasons for this phenomenon may be that they did not come from the same ancestor, or due to different sampling locations and times. The finished phylogenetic tree indicated that the chloroplast genome of *S. viminalis* var. *gmelinii* exhibited a close relationship with the *S. cupularis* and the *S. gordejewii* (Figure 1). The complete chloroplast genome sequence together with the gene annotations was deposited in the GenBank under the accession number OK505606. The publication of the complete chloroplast genome of *S. viminalis* var. *gmelinii* will potentially provide the necessary genetic resources and background data that can enhance the evolution and application study of the Salicaceae.

Author contributions

Xiaoping Li contributed to the conception of the study; Jie Wan, Huijie Tang, Jing Wang, Zicheng Yu, and Xu Yao performed the experiment; Jie Wan contributed significantly to analysis and manuscript preparation; Jie Wan performed the data analyses and wrote the manuscript; Xiaoping Li helped perform the analysis with constructive discussions and revised it critically for intellectual content. All authors approve the version to be published and agree to be accountable for all aspects of the work.

Disclosure statement

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

Funding

This study was supported by the Key projects of Jiangsu Province for college students' innovation and Entrepreneurship [202210298003Z], Natural Science Foundation of Jiangsu Province [BK20161525] and National Natural Science Foundation of China [30970245].

Regulation statement

This study strictly follows the wild plant protection regulations of China, which were firstly promulgated by Decree No. 204 of the state council of China on Sept. 30, 1996 and amended on Oct. 7, 2017 under Decision of The State Council on Revising Some Administrative Regulations.

ORCID

Jing Wang  <http://orcid.org/0000-0002-8897-4226>

Data availability statement

The complete chloroplast genome sequence was submitted to Genbank of NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide/OK505606>) under the reference number OK505606. The raw data has been deposited in SRA under accession number SRR16582939 (<https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA774245>); The BioSample number and the BioProject number are SAMN22560793 PRJNA774245 respectively.

References

- Argus GW. 1997. Infrageneric classification of *Salix* (Salicaceae) in the new world. *Syst Bot Monographs*. 52:1–121.
- Chen J-H, Sun H, Wen J, Yang Y-P. 2010. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. *Taxon*. 59(1):29–37.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34(17):i884–i890.
- Dierckxens N, Mardulyn P, Smits G. 2017. NOVOplasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Shi L, Chen H, Jiang M, et al. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyser. *Nucleic Acids Res*. 47:65–73.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30(12):2725–2729.
- 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.
- Zhou J, Jiao Z, Guo J, Wang BS, Zheng J. 2021. Complete chloroplast genome sequencing of five *Salix* species and its application in the phylogeny and taxonomy of the genus. *Mitochondrial DNA B Resour*. 6(8):2348–2352. doi:10.1080/23802359.2021.1950055.