

## Complete plastome sequence of *Xylosma longifolia* Clos. (Salicaceae)

Peng Chen<sup>a</sup>, Xiao-Feng Zhang<sup>a</sup>, Jacob B. Landis<sup>b</sup>, Zhi-Xin Zhu<sup>a</sup> and Hua-Feng Wang<sup>a</sup>

<sup>a</sup>Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou, China;

<sup>b</sup>School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY, USA

### ABSTRACT

*Xylosma longifolia* is a tree species within Salicaceae and is distributed in Guizhou, Yunnan, Fujian, Guangxi, Guangdong, and Hainan provinces of China as well as in Vietnam, Laos, and India. There are no studies utilizing the complete plastome of *Xylosma longifolia* in the current literature. Therefore, this report provides a reference for the plastid gene sequence of *Xylosma longifolia*, and it contributes to the phylogenetic placement and species identification. In this report, we described the complete plastome sequence of *Xylosma longifolia*. The complete plastome length of *Xylosma longifolia* is 156,938 bp and has the typical quadripartite structure and gene content of angiosperms, including two inverted repeat (IR) regions of 27,514 bp, a large single-copy (LSC) region of 85,221 bp and a small single-copy (SSC) region of 16,689 bp. The plastome contains 130 genes, including 86 protein coding genes, 36 tRNA genes, eight rRNA genes (5S rRNA, 4.5S rRNA, 16S rRNA, and 23S rRNA). The GC content of the plastome is 36.8%. The complete plastome sequence will be a valuable resource for studies involving the phylogenetic inference of Salicaceae.

### ARTICLE HISTORY

Received 22 December 2020  
Accepted 24 February 2021

### KEYWORDS

*Xylosma longifolia*;  
Salicaceae; plastome;  
genome structure;  
phylogenetic

*Xylosma longifolia* is a tree species within Salicaceae, occurring in Guizhou, Yunnan, Fujian, Guangxi, Guangdong, and Hainan provinces of China as well as in Vietnam, Laos, and India. *Xylosma longifolia* is a small evergreen tree or shrub, growing to approximately four to seven meters in height with grayish brown bark. Its leaves and barks could be used as medicine. Compared to the nuclear genome, chloroplast genomes (plastomes) have several advantages including a haploid nature, maternal inheritance, conserved typical structure in vascular plants, and minimal gene duplications (Wang et al. 2020). Complete plastome sequences have become a powerful tool for resolving plant phylogenies. Therefore, the study of plastomes has significant benefits for species identification and systematic position. However, there are no reports of the complete plastome sequence of *X. longifolia*.

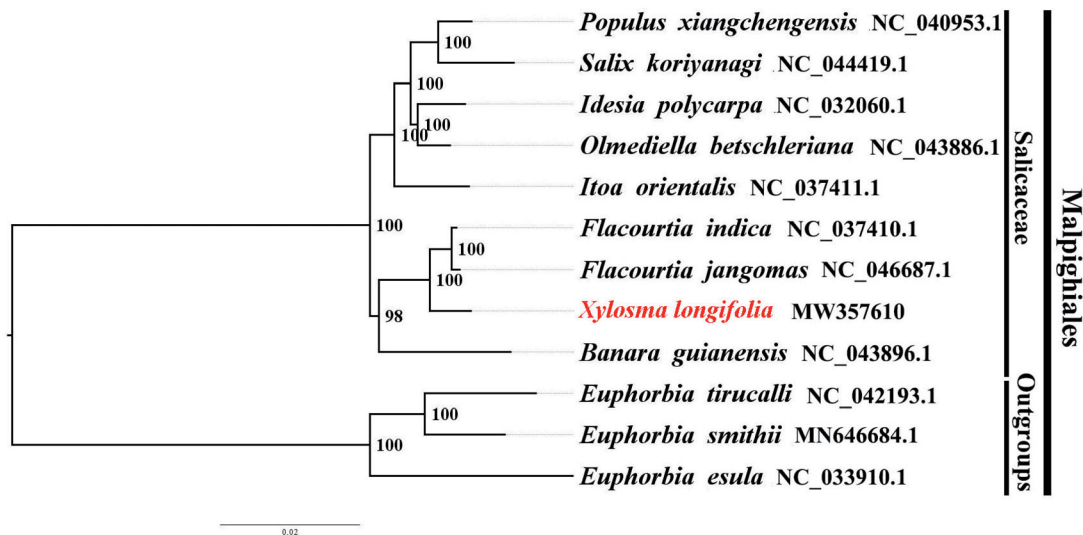
In this report, we describe the complete plastome sequence of *X. longifolia* (GenBank accession number: MW357610) for promoting the protection of germplasm and providing useful genomic resources. The plant sample of *X. longifolia* was collected from Forest Park, Wanding Town, Ruili county, Yunnan Province, China (100.22°E, 26.87°N). The voucher specimen (voucher code, RL0607) and associated DNA were deposited in the Herbarium of the Institute of Herbarium of China National GeneBank (code of herbarium: HCNGB).

We assembled approximately six Gb of clean data using the plastome of *Populus alba* (GenBank accession number: AP008956) (Rivarola et al. 2011) as a reference with MITO bim v1.8 (Hahn et al. 2013). We aligned all assembled

contigs to the reference with BLAST as implemented in Geneious R11.0.5 (Biomatters Ltd., Auckland, New Zealand) followed by mapping clean reads to the assembly to verify sequencing depth and contig overlap using Geneious R11.0.5. The assembled plastome of *Xylosma longifolia* was annotated with DOGMA (Dual Organellar Genome Annotator). Intron/exon boundaries were further determined with alignments in MAFFT v7 against the reference plastomes of *Populus alba* (AP008956). The annotated plastomes sequences were deposited in GenBank with accession number (MW357610). We used OGDRAW to draw genome maps with subsequent manual editing. Sequences of *Xylosma longifolia* and related species were aligned using PROGRESSIVE MAUVE v2.4.0 to compare the structure and gene content among the plastomes.

The results showed that the length of the plastome is 156,938 bp, which has the typical quadripartite structure of angiosperms, including two inverted repeats (IR), one large single copy (LSC) region and one small single copy (SSC) region. The plastome contains 130 genes, including 79 unique protein coding genes (seven of which are repeated in the IR), 29 unique tRNA genes (seven of which are repeated in the IR), and eight rRNA genes (5S rRNA, 4.5S rRNA, 16S rRNA and 23S rRNA are repeated in the IR). The GC content of the plastome overall is 36.8%, with the LSC, SSC, and IRs having a GC content of 34.5%, 42.1%, and 42.1%, respectively.

We used RaxML with 1000 bootstraps and the GTRGAMMAI substitution model, infer the Maximum likelihood (ML) phylogenetic tree of twelve published complete



**Figure 1.** The ML phylogeny inferred from 12 complete plastome sequences using RAxML. *Xylosma longifolia* (GenBank accession number, MW357610, this study), *Populus xiangchengensis* NC040953.1; *Salix koriyanagi*.NC044419.1; *Idesia polycarpa* NC032060.1; *Olmediella betschleriana* NC043886.1; *Itoa orientalis* NC037411.1; *Flacourtia indica* NC037410.1; *Flacourtia jangomas* NC046687.1; *Banara guianensis* NC\_043896.1; Outgroups: *Euphorbia tirucalli* NC042193.1; *Euphorbia smithii* MN646684.1; *Euphorbia esula* NC033910.1.

plastomes of Malpighiales, using *Euphorbia tirucalli* NC\_042193.1, *Euphorbia smithii* MN646684.1, and *Euphorbia esula* NC\_033910.1 as outgroups. By reconstructing the phylogenetic relationships of *X. longifolia* and published plastomes of related taxa in the order, we found that *X. longifolia* is most closely related to *Flacourtia indica* and *F. jangomas* within Salicaceae (Figure 1). Most nodes in the inferred ML trees were highly supported, providing an initial beneficial understanding to the phylogenetic relationships of Salicaceae. After comparing the results obtained here with a previously published phylogeny of *X. longifolia* (e.g. Rivarola et al. 2011), we find that the phylogenetic resolution is improved over previous inferred trees relying on few plastid or nuclear markers (e.g. Hamzeh and Dayanandan 2004; Liu et al. 2016). Our results demonstrate the power of plastomes in phylogenomics to improve estimates of phylogeny among genera and subfamilies, and it provides new insights into plastome evolution within Salicaceae. The complete plastome sequence of *Xylosma longifolia* will provide valuable genetic information for the protection of this species along with the phylogenetic study for Salicaceae.

### Disclosure statement

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

### Funding

This study was funded by the startup fund from Hainan University [kyqd1633 and kydq(zr)1840].

### Data availability statement

The genome sequence data supporting the results of this study are publicly available on GenBank of NCBI (<https://www.ncbi.nlm.nih.gov/>) with registration number MW357610. The associated SRA, BioProject and BioSample numbers are SRS3260867, PRJNA438407 and SAMN08770789, respectively.

### References

- 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.
- Hamzeh M, Dayanandan S. 2004. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast TRNT-TRNF region and nuclear rDNA. *Am J Bot.* 91(9):1398–1408.
- Liu X, Wang ZS, Wang DS, Zhang JG. 2016. Phylogeny of *Populus*-*Salix* (Salicaceae) and their relative genera using molecular datasets. *Biochem Systemat Ecol.* 68:210–215.
- Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, Melake-Berhan A, Huot Creasy H, Puiu D, Rosovitz MJ, et al. 2011. Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLOS One.* 6(7):e21743.
- Wang H-X, Liu H, Moore MJ, Landrein S, Liu B, Zhu Z-X, Wang H-F. 2020. Plastid phylogenomic insights into the evolution of the Caprifoliaceae s.l. (Dipsacales). *Mol Phylogenet Evol.* 142:106641
- Wu J, Nyman T, Wang D-C, Argus GW, Yang Y-P, Chen J-H. 2015. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evol Biol.* 15(1):31
- Zhu ZX, Mu WX, Wang JH, Zhang JR, Zhao KK, Friedman CR, Wang HF. 2018. Complete plastome sequence of *Dracaena cambodiana* (Asparagaceae): a species considered “Vulnerable” in Southeast Asia. *Mitochondrial DNA Part B.* 3(2):620–621.