

Genetics

The New Completed Genome of Purple Willow (*Salix purpurea*) and Conserved Chloroplast Genome Structure of Salicaceae

Zhiqiang Wu

Colorado State University, Department of Biology, Fort Collins, CO, USA

To compare the whole chloroplast genomes from Salicaceae, the fully assembled chloroplast genome purple willow (*Salix purpurea*) was assembled in this study based on the next generation sequencing data. The total genome size of *Salix purpurea* was 155,590 bp in length, comprising a pair of inverted repeats (IRs) of 27,459 bp, which were divided by large single copy (LSC) and small single copy (SSC) of 84,452 bp and 16,220 bp, respectively. For seven species of Salicaceae, the same sets of gene content and number are found. 110 genes were annotated, including 76 protein coding genes, 30 tRNA genes and 4 rRNA genes. Among these, 18 are duplicated in the inverted repeat regions, 14 genes contained 1 intron, and 3 genes (*rps12*, *clpP* and *ycf3*) comprised of 2 introns. *Journal of Nature and Science*, 1(3):e49, 2015.

Chloroplast genome, | Next-generation sequencing | *Salix* | *Salix purpurea*

Salix (willow) and *Populus* (poplar, aspen, and cottonwood) constitute the family Salicaceae representing a very diverse group of dioecious catkin-bearing woody plants. As an ideal model genus to study the bioenergy and biofuels, two species from this family had been finished the whole genome sequencing, including *Populus trichocarpa* (Tuskan et al., 2006) and *Salix suchowensis* (Dai et al., 2014). In genus *Salix*, majority of species are mainly distributed around the Northern Hemisphere, except a few of them native to the Southern Hemisphere (Argus 2007). Willows possess many characteristics desired for energy crops, for example the fast-growing with multiple stems to produce plentiful biomass. Genetic engineering within chloroplast genome has been used as an effective way to improve crops productivity and resistance (Devine and Daniell, 2004).

Chloroplast was acquired through endosymbiosis from ancestral free-living cyanobacteria around a billion years ago. Chloroplast (cp) genomes exhibit a typical circular double-stranded DNA arrangement, with sizes that ranged from 115 to 165 kb among land plants, but are highly conserved with regard to gene content and gene order (Ravi et al., 2008; Wu et al., 2015). Like the widely phylogenetic application of animal mitochondrial genome, cp genomes in plants, based on their conserved genome structures and comparatively high substitution rates, also keep valuable phylogenetic signals in all kinds of lineages (Wang et al., 2010, 2011; Wu and Ge, 2012). The molecular markers from cp genome can also be used to explore the biogeographical relationships among diverse plant populations (Wang et al., 2011; Chen et al., 2012). In addition, to exploit the high-efficiency of DNA barcoding marker and genetic transformation loci, the cp genomes will always provide abundant information (Group CPB, 2011; Day and Goldschmidt-Clermont, 2011).

In this study, we successfully assembled and annotated the complete cp genome of *Salix purpurea* following the method used in Wu (2014) by downloading the next generation sequencing data from NCBI database. The phylogenetic tree was built using NJ method based the whole genome alignment of seven species as described in Wu et al., (2015) and the dot-plot was employed to show the cp genome structural features among four representing Salicaceae species with out-group in Figure 1. The extremely conserved structures of chloroplast genome in all Salicaceae were demonstrated from Figure 1 with the same collinear orientation.

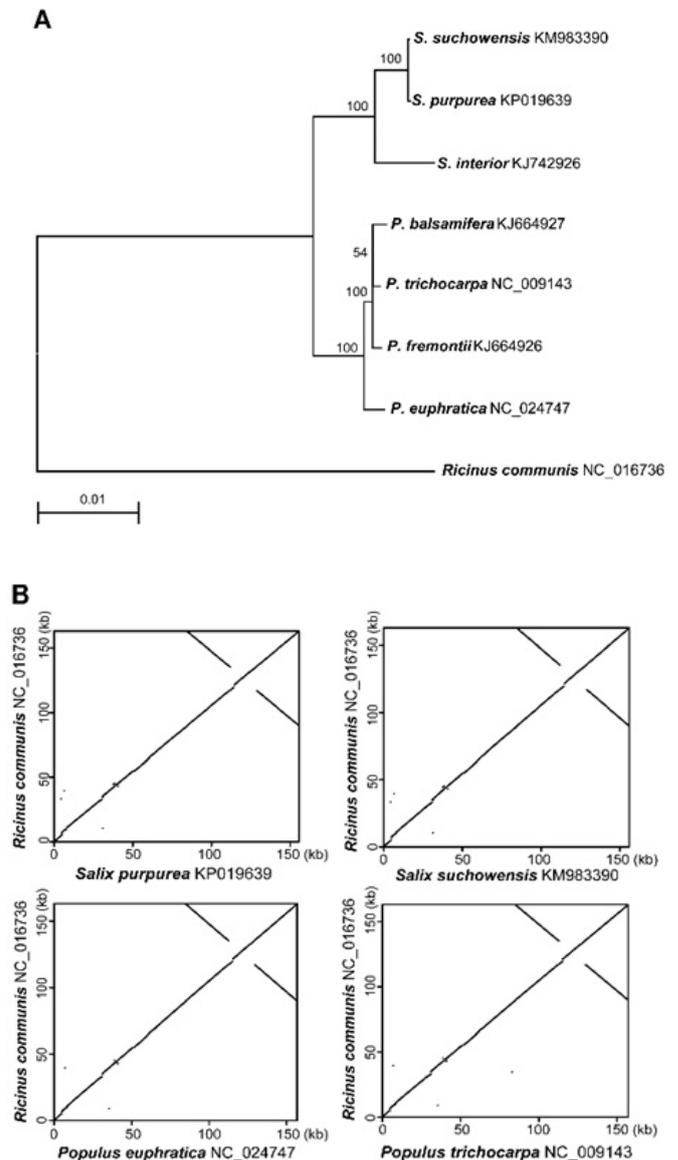


Figure 1. A. Phylogenetic trees were built using the whole genome sequence data with neighbor-joining (NJ) method of seven species from Salicaceae. Numbers above the branches are the bootstrap values of NJ with 1,000 bootstraps. Branch length is proportional to the number of substitutions, as indicated by the scale bar. B. Chloroplast genome collinear analyses using dot plot from *Populus* [*P. trichocarpa* (Tuskan et al., 2006) and *P. euphratica* (Zhang and Gao, 2014)] and *Salix* [*S. suchowensis* (Wu, 2014) and *S. purpurea* (this study)] with the out-group species *Ricinus communis* (Rivarola et al., 2011) from Euphorbiaceae

Conflict of interest: No conflicts declared.

Corresponding Author: Zhiqiang Wu, Colorado State University, Department of Biology, Fort Collins, CO, USA

Email: wu.zhiqiang.1020@gmail.com.

© 2015 by the Journal of Nature and Science (JNSCI).

The cp genome of *S. purpurea* (GenBank accession KP019639) has a total length of 155,590 bp and composed of LSC region of 84,452 bp, two IR copies 27,459 bp and SSC region of 16,220 bp. The overall GC contents of the chloroplast genome were 36.69%, and in the LSC, SSC and IR regions were 34.41%, 30.97% and 41.87%, respectively. The full annotated genes are all the same with other Salicaceae species (Tuskan et al., 2006; Wu, 2014) includes 110 unique genes, including 76 protein-coding genes, 30 tRNA genes and 4 rRNA genes. All other gene features are also the same with others, such as 17 genes contain introns with three of these genes

(*clpP*, *rps12* and *ycf3*) exhibiting two introns. In the two inverted repeat regions, 18 genes were duplicated including 7 tRNAs, 7 protein genes and 4 rRNAs. But, three genes (*infA*, *rps16* and *rpl32*) are truncated or missing from all seven Salicaceae chloroplast genomes.

Declaration of interest

We are grateful to the opened raw genome data from public database. The authors declare no conflict of interest and are responsible for the content.

1. Argus GW. (2007). *Salix* (Salicaceae) distribution maps and a synopsis of their classification in North America, north of Mexico. *Harv Pap Bot* 12: 335-368.
2. Chen D, Zhang X, Kang H, Sun X, Yin S, et al. (2012) Phylogeography of *Quercus variabilis* Based on Chloroplast DNA Sequence in East Asia: Multiple Glacial Refugia and Mainland-Migrated Island Populations. *PLoS ONE* 7(10): e47268.
3. Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, et al. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res* 24:1274-1277.
4. Day A, Goldschmidt-Clermont M. (2011). The chloroplast transformation toolbox: Selectable markers and marker removal. *Plant Biotechnol J* 9: 540-53.
5. Devine AL and Daniell H. (2004). Chloroplast genetic engineering. In S. Moller (Ed.), *Plastids* (pp. 283-320). United Kingdom: Blackwell Publisher.
6. Group CPB, Li DZ, Gao LM, Li HT, Wang H, et al. (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci USA* 108: 19641-19646.
7. Huang DI, Hefer CA, Kolosova N, Douglas CJ, Cronk QCB. (2014). Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *New Phytol* 204: 693-703.
8. Ravi V, Khurana JP, Tyagi AK, Khurana P. (2008). An update on chloroplast genomes. *Plant Syst Evol* 271:101-22.
9. Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, et al. (2011) Castor Bean Organelle Genome Sequencing and Worldwide Genetic Diversity Analysis. *PLoS ONE* 6(7): e21743.
10. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604.
11. Wang L, Qi XP, Xiang QP, Heinrichs J, Schneider H, Zhang XC. (2010). Phylogeny of the paleotropical fern genus *Lepisorus* (Polypodiaceae, Polypodiopsida) inferred from four chloroplast genome regions. *Mol Phylogenet Evol* 54(1): 211-225.
12. Wang L, Wu ZQ, Bystrakova N, Ansell SW, Xiang QP, Heinrichs J, Schneider H, et al. (2011). Phylogeography of the Sino-Himalayan fern *Lepisorus clathratus* on "the roof of the world". *PLoS one* 6 (9): e25896.
13. Wang L, Wu ZQ, Bystrakova N, Ansell SW, Xiang QP, Heinrichs J, Schneider H, et al. (2011). Phylogeography of the Sino-Himalayan fern *Lepisorus clathratus* on "The roof of the world". *PLoS ONE* 6: e25896.
14. Wu ZQ, Ge S. (2012). Phylogeny of the BEP clade in grasses revisited: Evidence from whole genome sequences of chloroplast. *Mol Phylogenet Evol* 62: 573-8.
15. Wu ZQ, Ge S. (2014). The whole chloroplast genome of wild rice (*Oryza australiensis*). *Mitochondrial DNA* [Epub ahead of print]. DOI: 10.3109/19401736.2014.928868..
16. Wu ZQ, Tembrock LR and Ge S. (2015). Are differences in genomic datasets due to true biological variants or errors in genome assembly: an example from two chloroplast genomes. *PLoS ONE* (accept).
17. Wu ZQ. (2014). The whole chloroplast genome of shrub willows (*Salix suchowensis*). *Mitochondrial DNA* [Epub ahead of print]. DOI: 10.3109/19401736.2014.982602.
18. Zhang QJ, Gao LZ. (2014). The complete chloroplast genome sequence of desert poplar (*Populus euphratica*). *Mitochondrial DNA Early Online*: 1-3.