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.
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 open ed raw genome data from public database. The authors declare no conflict of interest and are responsible for the content.