Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2025) **23**, pp. 857–859 doi: 10.1111/pbi.14543

Brief Communication

Genome and CRISPR/Cas9 system of a widespread forest tree (*Populus alba*) in the world

Yan-Jing Liu, Peng-Fei Jiang, Xue-Min Han, Xiao-Yuan Li, Hai-Ming Wang, Yun-Jiao Wang, Xiao-Xia Wang and Qing-Yin Zeng*

State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China

Received 13 July 2023;
revised 15 November 2024;
accepted 23 November 2024.
**Correspondence* (Tel +86 10 62824020; fax +13 69 9127045;
email qingyinzeng@163.com)

Keywords: *Populus alba*, CRISPR/Cas9, genome, chromosomal feature.

Trees account for approximately 90% of the Earth's biomass and provide humans with various necessities for survival, such as clean air and water, wood, fibre and fuel (Petit and Hampe, 2006; Tuskan *et al.*, 2006). Compared with annual plants, trees have many significant features, such as perennial growth, large size, secondary growth from a vascular cambium and dormancy (Douglas, 2017). Trees should be considered as a model system in plant biology and provide possibilities to answer questions that cannot be easily solved in the annual model systems of *Arabidopsis* and rice.

Trees of the genus *Populus* are prominent forest species in temperate regions of the Northern Hemisphere. *Populus* trees, as the model systems for plant biology, have several advantages, including rapid growth, small genome, facile transgenesis and easy cloning (Bradshaw *et al.*, 2000). *Populus trichocarpa* is now widely used as a model system in the United States and Europe. *Populus trichocarpa* native to western North American cannot grow well in the fields and forests of China, which limits its breeding and application potential in China. In China, *Populus tomentosa* used to be a study system in basic research. It is a stabilized interspecific hybrid species widespread in Asia and is extensively used in breeding or forestry industries. Besides, poplar 84 K (*Populus alba* × *P. glandulosa*) and poplar 741 (*Populus alba* × (*P. davidiana* × *P. simonii*) × *P. tomentosa*) are also widely used as study systems in China. These two cultivars and *P. tomentosa* are not the most suitable strains for tree genetic study system due to their hybridization background.

The white poplar (*P. alba*) is a widespread forest tree in the world, distributed in Europe, Asia, North America, South America, Africa and Oceania (Figure 1a). *Populus alba* is not only a beautiful tree (Figure 1b) but also has extensive adaptability to different ecological environments (Stölting *et al.*, 2015). As a fast growing tree, *P. alba* is included in the forest tree breeding programme in China and the European programme of forest genetic resources. Our previous study found that with the natural expansion of *P. alba* from Europe to China, the natural populations in China experienced a bottleneck effect. Average pooled heterozygosity value of *P. alba* populations in China was much lower than that in Italy and Hungary (Liu *et al.*, 2019). As a pure and highly adaptable natural species, *P. alba* has been used as a hybrid parent in the breeding history of China. Due to the widespread distribution of *P. alba* in the Eurasian continent, using *P. alba* as a research system is conducive to the promotion of research results. Thus, *P. alba* in China is suitable as a study system for tree biology.

Populus alba v1.0 assembly was fragmented and comprised 1285 contigs with N50 of 1181 Kb, supplemented by two organellar genome fragments (Liu *et al.*, 2019). Here, we present a new assembly based on chromosome conformation capture, while combining the previous Pacific Biosciences (PacBio) single-molecule real-time (SMRT) and whole-genome shotgun sequences. A total of 124.17 gigabases (Gb) of Illumina clean data were obtained for high throughput chromosome conformation capture (Hi-C) analysis (Table S1). Based on these data, contigs of the *P. alba* were clustered into 97 scaffolds with a final N50 of 22.7 Mb (Table S2). These 97 scaffolds include 19 chromosome-length pseudomolecules, covering 412.7 Mb of the genome sequence and 78 unplaced scaffolds, representing 3.7 Mb unintegrated sequences. The order of 19 chromosome-length pseudomolecules was designated according to their collinearity with the *P. trichocarpa* v4 genome (Figure 1c). All the sequences considered as mitochondrial and plastidal sources in *P. alba* v1.0 assembly were removed before scaffolding and were assembled into complete chloroplast and mitochondria genomes (Figures S1 and S2), separately.

The expected gene contents of Benchmarking Universal Single-Copy Orthologs (BUSCO), Core Eukaryotic Gene (CEGMA) and LTR Assembly Index (LAI) were detected to qualify the completeness of the genome assembly (Table S3). The results showed that 97.8% of eudicots, 98.4% of embryophyta and 98.4% of eukaryota orthologs could be detected in *P. alba* v2.0 genome (Figure 1d). The varying degrees of increase in evaluation results of BUSCO, CEGMA and LAI demonstrated an improvement in the continuity and completeness of the *P. alba* v2.0 genome compared with v1.0. The BUSCO and CEGMA completeness of *P. alba* v2.0 genome was close to that of *P. trichocarpa* v4. The LAI score was greater than 10, which indicated that the assembly had reached the level of forest reference quality.

By optimized annotation procedures relying on expression data and gene sets of related species, 34 010 predicted protein-coding gene models were obtained in annotation version 2.0. A total of 29 847 (87.76%) of them were functional annotated with at least one database (Figure 1e). Among all gene models, 21 143 (62.16%) could found segmental duplication derived paralogues (Figure 1f). Slightly more repetitive elements were found in *P. alba* v2.0 than v1.0 (Table S4). The vast majority of TEs are LTR-RT, dominated by Gypsy-type elements, accounting for 45.3% of all repetitive elements, followed by Copia-type with 17.9%.

Studies on *A. thaliana* and *P. trichocarpa* have shown that the centromere region is gene-sparse, low SNP density, Gypsy

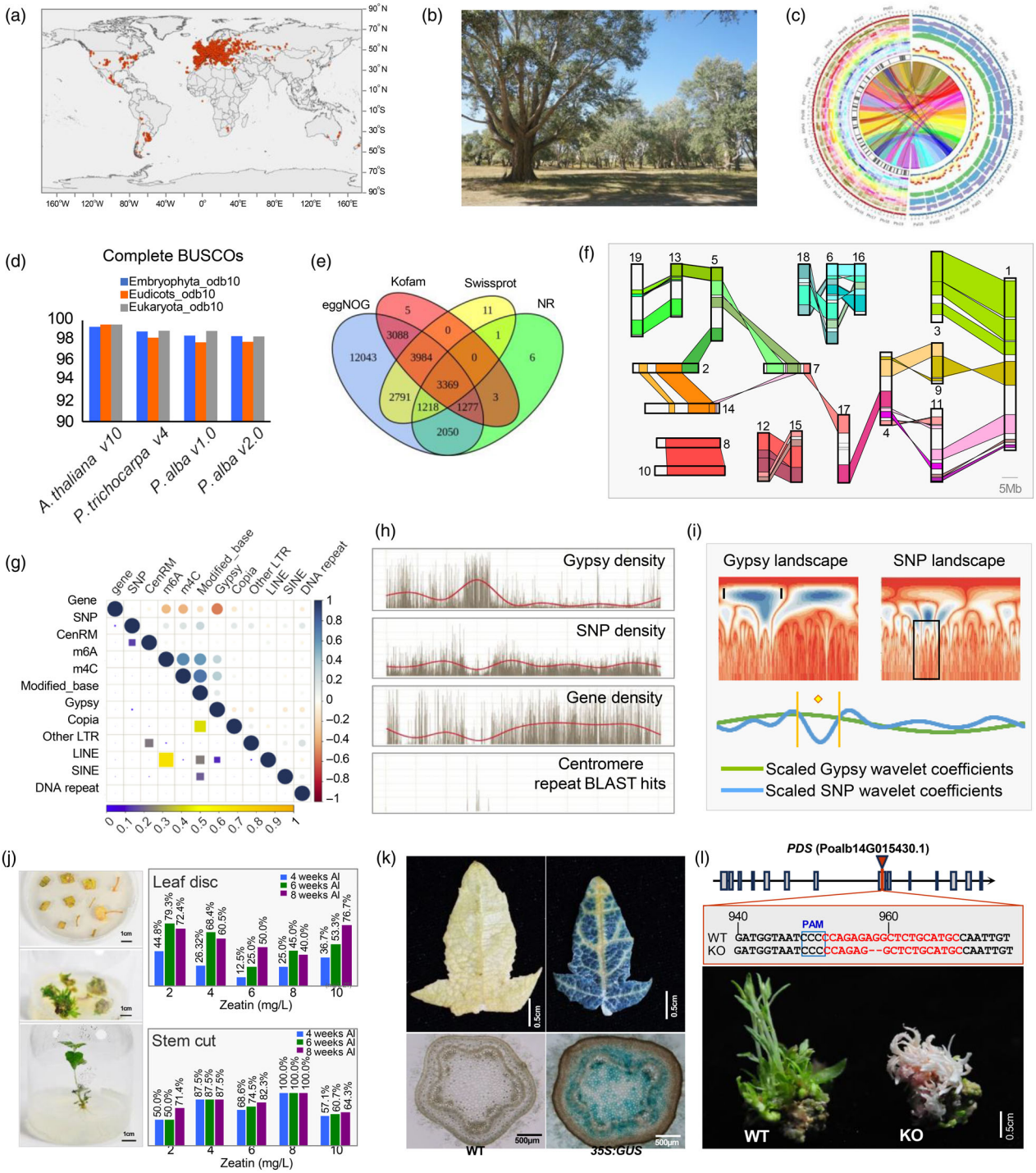

Figure 1 Features of *Populus alba* v2.0 genome and CRISPR/Cas9 system (details in Text S1).

elements enriched and high cytosine methylated (Jiang *et al*., 2003; Natali *et al*., 2015; Weighill *et al*., 2019). Centromere position was predicted using wavelet-based genomic signal analysis described by Weighill *et al*. (2019) (Figure 1g–i; Figure S3). Except for 2, 7, 8, 15, 18 and 19, relative centromeres positions (centre bin (bp)/chromosome length(bp)) of the rest 13 pseudochromosomes in *P. alba* are similar to those in *P. trichocarpa* (Table S5; Figure S4). By mapping centromeric repeat sequences from *P. trichocarpa* and various plants onto v2.0 chromosomes, the position of centromere/pericentromere on chromosome 1, 3, 6, 9, 11, 12, 18 and 19 could be further confirmed (Figure S3).

Based on the standard *Agrobacterium*-mediated transformation method, we optimized the transformation procedures for the sequenced *P. alba* plant (Supplemental Materials and Methods). The explants were originally obtained from sterilized tender stems and then subcultured every 4 weeks for sustainable use. The reporter gene, β-glucuronidase (*GUS*), and the *P. alba* phytoene desaturase (*PDS*) gene were used to test transformation and genome editing efficiency, respectively. The leaf discs and stem

segments of tissue culture seedlings were used for transformation. Zeatin (ZT) is an effective plant hormone for stimulating the division in non-meristematic tissues. To improve transformation efficiency, we assessed the calli regeneration rate of leaf discs and stem segments cultured in medium supplemented with 2 ~ 100 mg/L ZT. The results showed that a relative low concentration of ZT (2 ~ 10 mg/L) was sufficient for calli regeneration within 4 weeks (Figure 1j). The optimized ZT concentrations for leaf discs and stem segments were different. Following the protocol presented, we successfully induced the regeneration of sprout and root within 4 ~ 5 weeks and 10 ~ 12 weeks after co-culturing with *Agrobacterium*, respectively. The final transformation efficiency was 31.96%. The genome editing efficiency was 69.47% for *PDS* (Figure 1k,l). This efficient genetic transformation method will assist in the functional research of *P. alba*.

Funding

This project was funded by the National Key R&D Program of China (2022YFD2200102).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Q.-Y.Z. and Y.-J.L. designed research. Y.-J.L. analysed genome and experimental data. P.-F.J., X.-M. H., X.-Y. L., H.-M. W., Y.-J. W. and X.-X.W. conducted experiments. Q.-Y.Z. and Y.-J.L. wrote the article.

Data availability statement

Raw sequencing data of Hi-C and the assembly and annotation of *P. alba* v2.0 genome will be released in NCBI under BioProject PRJNA491245 upon publication.

References

Bradshaw, H.D., Ceulemans, R., Davis, J. and Stettler, R. (2000) Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. *J. Plant Growth Regul.* **19**, 306–313.

Douglas, C.J. (2017) *Populus* as a model tree. In *Comparative and Evolutionary Genomics of Angiosperm Trees. Plant Genetics and Genomics: Crops and Models*(Groover, A. and Cronk, Q., eds), pp. 61–84. New York, NY, USA: Springer.

Jiang, J., Birchler, J.A., Parrott, W.A. and Dawe, R.K. (2003) A molecular view of plant centromeres. *Trends Plant Sci.* **8**, 570–575.

Liu, Y.J., Wang, X.R. and Zeng, Q.Y. (2019) *De novo* assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Sci. China Life Sci.* **62**, 609–618.

Natali, L., Cossu, R.M., Mascagni, F., Giordani, T. and Cavallini, A. (2015) A survey of Gypsy and Copia LTR-retrotransposon superfamilies and lineages and their distinct dynamics in the *Populus trichocarpa* (L.) genome. *Tree Genet. Genomes* **11**, 107.

Petit, R.J. and Hampe, A. (2006) Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* **37**, 187–214.

Stölting, K.N., Paris, M., Meier, C., Heinze, B., Castiglione, S., Bartha, D. and Lexer, C. (2015) Genome-wide patters of differentiation and spatially varying selection between postglacial recolonization lineage of *Populus alba* (Salicaceae), a widespread forest tree. *New Phytol.* **207**, 723–734.

Tuskan, G.A., 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.

Weighill, D., Macaya-Sanz, D., DiFazio, S.P., Joubert, W., Shah, M., Schmutz, J., Sreedasyam, A. *et al*. (2019) Wavelet-based genomic signal processing for centromere identification and hypothesis generation. *Front. Genet.* **10**, 487.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Chloroplast genome of *P. alba*.
Figure S2 Mitochondria genome of *P. alba*.
Figure S3 The frequency of features on *P. alba* chromosome in 10 kb non-overlap windows.
Figure S4 Dot-plot alignments of the *P. alba* and *P. trichocarpa* pseudomolecules.
Figure S5 Genome wide evaluation of gene editability by CRISPR Cas9 system.
Figure S6 Calli induction efficiency under different concentrations of zeatin.
Figure S7 Shoot induction on medium CM3 and CM4 with ZT and 6-BA, respectively.
Table S1 Hi-C paired-end reads.
Table S2 Updated *Populus alba* v2.0 assembly.
Table S3 Quality of *Populus alba* v2.0 assembly.
Table S4 Classification of repetitive elements in the *Populus alba*, *Populus trichocarpa*, *Populus euphratica* and *Salix suchowensis* genome.
Table S5 Putative centromere position of *Populus alba* v2.0 assembly.
Table S6 Statistics of PDS editing results induced by CRISPR/Cas9.
Text S1 Materials and Methods.