Chromosome-level genome assembly of a triploid poplar *Populus alba 'Berolinensis'*

Song Chen[1], Yue Yu[1], Xinyu Wang[1], Sui Wang[1], Tianjiao Zhang[1], Yan Zhou[2], Ruihan He[2], Nan Meng[1], Yiran Wang[1], Wenxuan Liu[1], Zhijie Liu[1], Jinwen Liu[2], Qiwen Guo[2], Haijiao Huang[1], Ronald R Sederoff[1], Guohua Wang[1], Guanzheng Qu[1], and Su Chen[1]

[1]Northeast Forestry University
[2]Affiliation not available

March 10, 2023

Abstract

Many studies have provided significant insights into polyploid breeding in recent years, but limited research has been carried out on trees. The genomic information needed to understand the growth and response to abiotic stress in polyploidy trees is largely unknown, but has become critical due to the threats to our forests imposed by climate change. Populus alba '*Berolinensis*', also known "Yinzhong poplar", is a triploid poplar from the northeast of China. This hybrid triploid poplar is widely used as a landscape ornamental and in urban forestry for its adaptation to adverse environments and fast growth than its parental diploid. It is an artificially synthesized male allotriploid hybrid, with three haploid genomes of *P. alba* 'Berolinensis' originated from different poplar species, so it is attractive for studying polyploidy genomic mechanisms in heterosis. In this study, we focused on the allelic genomic interactions in *P. alba* '*Berolinensis*', and generated a high-quality chromosome-level genome assembly consisting of 19 allelic chromosomes. Its three haploid chromosome sets are polymorphic with an average of 25.73 nucleotide polymorphism sites per kilobase. We found that some stress related genes such as *RD22* and *LEA7* exhibited sequence differences between different haploid genomes. The genome assembly has been deposited into our polyploid genome online analysis website TreeGenomes (https://www.treegenomes.com). These polyploid genomic related resources will provide a critical foundation for the molecular breeding of *P. alba* '*Berolinensis*' and help us uncover the allopolyploidization effects of heterosis and abiotic stress resistance and traits of polyploidy species deeper in the future.

Song Chen[1], Yue Yu[1], Xinyu Wang[1], Sui Wang[1,2], Tianjiao Zhang[3], Yan Zhou[1], Ruihan He[1], Nan Meng[1], Yiran Wang[1], Wenxuan Liu[1], Zhijie Liu[1], Jinwen Liu[4], Qiwen Guo[5], Haijiao Huang[1], Ronald R Sederoff[1,6], Guohua Wang[1,3,*], Guanzheng Qu[1,*] and Su Chen[1,*]

1. State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin 150040, China.

2. Key Laboratory of Soybean Biology in Chinese Ministry of Education, Northeast Agricultural University, Harbin 150030, China.

3. College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China.

4. Zhejiang Provincial Key Laboratory for Water Environment and Marine Biological Resources Protection, Wenzhou University, Wenzhou 325035, China.

5. Department Forest and Soil Sciences, Institute of Forest Ecology, University of Natural Resources and Life Sciences, Vienna 1190, Austria.

6. Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh NC 27695, USA.

Introduction

Polyploidization is a successful mechanism for plants to accommodate abiotic stress, particularly adaptation to arid and cold environments (Levin, 1983; Lourkisti et al., 2020). Polyploid plants always have a greater advantage than diploid plants when faced with extreme environments (Ehrendorfer, 1980; Lowry & Lester, 2006). Previous studies demonstrated that naturally occurring tetraploid Arabidopsis (Arabidopsis thaliana) plants exhibit increased salinity tolerance compared to diploids (Chao et al., 2013). Both tetraploid rice (Oryza sativa) and the hybrid citrange (Citrus sinensis × Poncirus trifoliata) show increased tolerance to salinity and drought because of polyploidization, which affects the expression of genes involved in stress and the phytohormone response pathways (Ruiz et al., 2016; P. M. Yang et al., 2014). Similarly, the tetraploid rootstock-grafted watermelon (Citrullus lanatus) is more tolerant to salinity stress than its diploid parental plants (Zhu et al., 2018). In addition, Rice et al. (2019) and Brochmann et al. (2004) found that polyploidy plants frequency will significantly increase from the equator to the poles with latitude.

As for the possible cause, Wu et al. (2019) attributed that the expansion of specific gene families and the resulting dosage effects significantly facilitated adaptation to changed conditions such as low temperature and salinity stress. Substantial evidence also indicates that a large number of genes encoding transcription factors, signal transducers, and enzymes are retained during polyploidization, due to the immediate selective advantage or to maintain stoichiometry (Birchler & Veitia, 2012; Song et al., 2020; Van de Peer et al., 2021). Furthermore, epigenetic remodeling was also altered in polyploid plants (Jackson, 2017; Wendel et al., 2018). In hybrid allopolyploid A. thaliana, epigenetic modifications can increase the ability to combat stress and to recover more rapidly than their parental species (Ni et al., 2009; Miller et al., 2015).

Poplars are the most widely distributed and domesticated forest tree species. Polyploid (Populus spp.) varieties are always created through interspecific hybridization, followed by asexual reproduction, which allows replication and maintenance of meiotically unstable variants (Henry et al., 2015). Populus alba 'Berolinensis', also named "Yinzhong poplar", is a typical triploid species in China. It is a fast-growing and highly adaptable poplar triploid hybrid that is resistant to drought, cold, and saline-alkaline stresses (S. Chen & Polle, 2009; Y. Li et al., 2008; Tang et al., 2010). This hybrid triploid species is widely used in the forestation of high-latitude urban areas due to its adaptation to adverse environments (Chai et al., 2002). As an artificially synthesized male allotriploid elite hybrid (J. Wang et al., 2017), the three haploid genomes of P. alba 'Berolinensis' originated from different poplar species, so it is a desired material for studying the collaboration mechanisms between allelic genes from different haploid genomes.

Recent research has provided significant insights into polyploid breeding (Niazian & Nalousi, 2020; Sattler et al., 2016; Song et al., 2012), but it has been mainly limited to mechanisms regulating the polyploid formation, cell division, and ploidy dynamics. Investigations into the cooperation of multiple allelic genes in polyploidy and heterosis lag behind those of diploid organisms (Renny-Byfield & Wendel, 2014). Although several studies have recently mapped molecular mechanisms of polyploidy in growth and response to abiotic stress (Du et al., 2020), the genomic basis for these processes is largely unknown. As an artificially synthesized allotriploid plant, P. alba 'Berolinensis' has three sets of different allelic haploid chromosomes and a significant growth advantage over other poplar species. Therefore, we used P. alba 'Berolinensis' as a research platform to study the collaboration mechanisms between allelic genes from different haploid genomes, investigate the polyploid plants' growth advantage, and explore the potential of extending the results to other species using genetic engineering.

Materials and methods

Plant materials, physiological measurements and flow cytometric analysis

Genome sequencing and assembly were performed on P. alba 'Berolinensis' YZY-HRB-1 plantlets, an allotriploid hybrid poplar from Northeast Forestry University (Harbin, China). The plantlets used in this study were all generated through tissue culture, then pre-cultured in a greenhouse with 16 h day length, 26 °C / 22 °C Day/Night temperature, and relative humidity was kept 65 - 95%. In addition, we collected seven genotypes of the parental species of P. alba 'Berolinensis' for the phenotypic observation experiment. The YBY-URC-1 (Urumqi, China) and YBY-URC-2 (Urumqi, China) represent the P. alba species in this study. The XJY-URC-1 (Urumqi, China) as the representative of P. bolleana, and OSY-HEL-1 (Central Finland Region, Finland), ZSY-MDG-1 (Mudanjiang, China), ZSY-YCH-1 (Yichun, China), and ZSY-OHE-1 (Greater Khingan Mountains, China) as the representatives of P. tremula. Using leaves as explants, all the plants of these genotypes were dedifferentiated into calli and then induced into tissue culture plantlets. When the plantlets of these genotypes grown to 10 cm in height, they were transplanted into pots. This day was defined as the first day. After 60 days of growth, thirty individuals of each genotype were used to assess the variation in phenotypic traits. Plant heights and stem diameters were measured using a measuring tape and a caliper, respectively. Leaf area was measured using LI-3000C (LI-COR Biosciences, Lincoln NE, USA).

The polyploid level of the P. alba 'Berolinensis' was confirmed using a ploidy analyzer (Partec PA, Münster, Germany). 0.5 cm2 of fresh leaves from P. alba 'Berolinensis' were chopped using a sharp razor blade in 0.5 mL of Sysmex CyStainTM UV Precise P Nuclei Extraction Buffer for 60 s. Then 2 mL of Sysmex CyStainTM UV Precise P DAPI staining solution was added held for 60 s. After filtration through a 30 μm filter, the samples were measured by a PA flow cytometer (Partec PA, Münster, Germany) equipped with a 365 nm UV mercury arc lamp and a filter combination for DAPI staining. A minimum of 10,000 nuclei was measured per sample and histograms were generated using the companion software to determine the peak position. Experimental conditions for a flow cytometry profile of its parental plants were the same as the P. alba 'Berolinensis'.

Sequencing and *de novo* assembly of genomic DNA

Genomic DNA was extracted from fresh leaves using the MGIEasy DNA Library Prep Kit (BGI, Wuhan, CN). PCR-free paired-end genome sequencing libraries were constructed according to a BGI protocol and sequenced (PE150) using the BGISEQ-500 platform (BGI, Wuhan, CN). The PacBio library with a 30-kb targeted size was generated following the manufacturer's protocol and sequenced using the PacBio Sequal II platform (PacBio, Menlo Park, USA). Hi-C sequencing was performed parallel to the genomic sequencing. The fixed DNA was sheared with the MboI restriction enzyme, and the sequencing platform was selected BGISEQ-500 (BGI, Wuhan, CN). Before assembly, clean data filtered by fastp v0.23.2 (Chen et al., 2018) were used to investigate the genomic features of P. alba 'Berolinensis' using the k-mer frequency distribution. The 27-mer frequency distribution analysis was performed using k-mer counting (KMC) (Kokot et al., 2017) and the result was interpreted by GenomeScope v2.0 (Ranallo-Benavidez et al., 2020).

The PacBio SMRT (single molecule real-time) sequencing subreads were de novo assembled using Canu v2.0 (Koren et al., 2017), and the genome size was estimated to be 430 Mb. Considering the large number (˜x395) of long-reads that were generated using the PacBio Sequal II platform, we used Canu with parameters of "corOutCoverage=1000", "minReadLength=2000", and "minOverlapLength=2000" as criteria for reads correction to avoid false positives in the assembly. P. alba 'Berolinensis' is a highly repetitive genome with a heterozygosity ratio above 3%. According to Canu, the following parameters were used to avoid collapsing the polyploid genome: "OvlMerThreshold=300", "corMhapSensitivity=normal", and "batOptions=-dg 3 -db 3 -dr 1 -ca 500 -cp 50". After assembly, the trimmed PacBio long-reads and cleaned NGS (next-generation sequencing) short reads were mapped to the draft assembly using minimap2 v2.17-r941 (H. Li, 2018) and bowtie2 v2.3.5.1 (Langmead & Salzberg, 2012). Furthermore, the assembly contigs were then polished four

rounds for short reads using Pilon v1.23 (Walker et al., 2014).

Hi-C library construction, chromosome assembly and evaluation

The raw contigs were first corrected using Hi-C reads. The Hi-C reads were aligned to contigs using Juicer v1.6 (Durand, Shamim, et al., 2016). Mis-joined contigs were corrected by detecting abrupt long-range contact patterns using the 3D-DNA v201008 pipeline (Dudchenko et al., 2017). The "release3ddna.pl" script provided by ALLHiC v0.9.13 (Zhang et al., 2019, p.) was used to transform the corrected contigs of 3D-DNA to the ALLHiC input format. Then the Hi-C reads were mapped again to the corrected contigs with HiC-Pro v3.0.0 using default parameters, which internally used bowtie2 v2.3.5.1. The Hi-C corrected contigs were further mapped onto the 57 pseudo-chromosomes in P. alba 'Berolinensis' using the ALLHiC v0.9.13 pipeline (Zhang et al., 2019). The chromosome number and orientation were renamed according to the chromosome-scale assembly of P. trichocarpa. The accuracy of Hi-C based chromosome construction was evaluated by a chromatin contact matrix. Juicebox v1.11.08 (Durand, Robinson, et al., 2016) was used for fine-tuning the assembled scaffolds graphically. Finally, three haplotypes were fully resolved at the chromosomal level.

The completeness of the assembled contigs and chromosomes was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.1.2 (Waterhouse et al., 2018). The "embryophyta_odb10 (Creation date: 2020-09-10)" lineage-specific profile that contains 1,614 BUSCO groups was used as the database for assessment.

Gene prediction and functional annotation

Repetitive sequences of the P. alba 'Berolinensis' genome were annotated using both homology-based search and de novo methods. The public TE (transposable elements) libraries RepBase v20181026 (Bao et al., 2015) and Dfam v3.0 (Storer et al., 2021) were downloaded as the homology-based search libraries. And the de novo TE libraries were built by using RepeatModeler v2.0.1 (Flynn et al., 2020) and EDTA (Extensive de novo TE Annotator) v1.9.6 (Ou et al., 2019). Finally, these four TE libraries were integrated together and used as the input libraries of RepeatMasker v4.1.1 (Tarailo-Graovac & Chen, 2009) to search the repeats in P. alba 'Berolinensis' assembly. After completing the whole genome searches, all the repetitive regions were soft-masked for further protein-coding gene annotation.

The gene structure was annotated by the MAKER v3.1.2 (Cantarel et al., 2007) pipeline, predicting protein-coding genes based on transcriptomic, homologous, and de novo methods. For transcriptome-based predictions, RNA-seq data of six samples from leaves were assembled by Trinity v2.8.5 (Haas et al., 2013). For homology-based predictions, the protein sequences of nine Populus species (NCBI Resource Coordinators et al., 2018) whose genomes have been sequenced were downloaded and de-redundancy. For de novo prediction, Augustus v3.3.3 (Stanke et al., 2006), GeneMark v4.71 (Besemer & Borodovsky, 2005), and SNAP (Semi-HMM-based Nucleic Acid Parser) v20131129 (Korf, 2004) analyses were performed on the repeat-masked genome, with parameters trained from the RNA-seq data. Finally, the transcript, homology, and de novo gene sets were merged to form a comprehensive non-redundant gene set using EVM (EvidenceModeler) v1.1.1 (Haas et al., 2008). The candidate alleles were then predicted by MCScanX v20131111 (Y. Wang et al., 2012) according to Zhang et al. (2021), and the genomic alignments were generated by SyRI (Synteny and Rearrangement Identifier) v1.6 (Goel et al., 2019) and NUCmer (NUCleotide MUMmer) v4.0.0 (Marcais et al., 2018). The tRNA genes were detected using tRNAscan-SE v2.0 (Chan et al., 2021), identification mode selected as eukaryotic.

Functional annotation was based on the comparison with the UniProtKB database (downloaded on June 1, 2020) (Boeckmann, 2003) using an E-value of 1e-5. GO terms were assigned to the annotated genes using AHRD (Automated Assignment of Human Readable Descriptions) v3.3.3 (Hallab, 2015). Further hierarchical, functional, and phylogenetic annotation on orthology assignment was annotated using the eggnog-mapper v2.0.1 (Huerta-Cepas et al., 2017, 2019). Transcription factors (TF) were predicted using

the PlantRegMap v2020 (Tian et al., 2019). In addition, the 2000 bp region upstream of all the coding genes was defined as the promoter region, in which the cis-elements were identified with PLACE database (Higo et al., 1999).

Molecular evolution and phylogenetic analysis

The protein sequences from nine Populus species and four related species were collected for phylogenetic analysis, which includes Populus alba v1.0 (Liu et al., 2019), P. bolleana v1.0 (J. Ma et al., 2019), P. deltoides v2.1 (DOE-JGI, https://phytozome.jgi.doe.gov), P. euphratica v1.0 (T. Ma et al., 2013), P. pruinose v1.0 (W. Yang et al., 2017), P. simonii v1.0 (H. Wu et al., 2020), P. tremula v2.2 (Lin et al., 2018), P. tremuloides v1.1 (Lin et al., 2018), P. trichocarpa v4.0 (Tuskan et al., 2006), Salix brachista v1.0 (J. Chen et al., 2019), Salix purpurea v5.1 (DOE-JGI, https://phytozome.jgi.doe.gov), Arabidopsis thaliana vTAIR10 (Berardini et al., 2015), and Oryza sativa v7.0 (Ouyang et al., 2007). All the orthologous genes were identified by OrthoFinder v2.5.4 (Emms & Kelly, 2019) using BLAST and the protein alignments of each single-copy family were then created using MAFFT v7.480 (Katoh & Standley, 2013). Based on this information, we reconstructed a maximum likelihood phylogenetic tree by using RAxML-NG v1.0.3 (Kozlov et al., 2019), O. sativa was selected as the outgroup.

Germline short variant discovery (SNPs + Indels)

The gatk4-data-processing v2.1.1 workflow (Broad Institute, 2017) provided by GATK official practices was used to implement the data pre-processing before variant discovery. We used it to finish the data clean-up operations, the correction of the technical biases, and the alignment between DNA-seq reads and P. alba 'Berolinensis' reference genome. Alignment was done with bwa-mem2 v2.2.1 (Vasimuddin et al., 2019), the primary alignment strategy was selected "MostDistant", the validation stringency was selected "Silent", the attributes to retain was selected "X0", the optical duplicate pixel distance was selected 2500, and all the BAM files were compressed at level 5. GATK4 HaplotypeCaller v4.2.0 (McKenna et al., 2010) was used to identify substitution and insertion/deletion variant calls. The GQBs (gvcf-gq-bands) were selected 10, 20, 30, 40, 50, 60, 70, 80, and 90. The annotation groups were selected "StandardAnnotation", "StandardHCAnnotation", and "AS_StandardAnnotation". These variants were then annotated and analyzed using SnpEff v5.0e (Cingolani et al., 2012), the "upDownStreamLen" was set 2000 bp to match the promoter length.

RNA-seq analysis of salinity stressed transcriptomes

To quantify the transcriptome expression data of Populus alba 'Berolinensis' under salinity stress, these plants were treated with a salt solution with 2% NaCl continued for 24 hours. Total RNAs were extracted from the salinity stressed samples using the MGIEasy RNA Library Prep Set (BGI, Wuhan, China), cDNA libraries were constructed according to the MGISEQ user manual. The samples were then sent to Annoroad Gene Technology (Annoroad, Beijing, China) and performed paired-end read sequencing using the MGISEQ-2000 platform (BGI, Wuhan, China). After filtering out low-quality reads, the clean data was submitted to the NCBI Sequence Read Archive (SRA).

Gene expression analysis was performed using the nf-core/rnaseq v3.2 (Ewels et al., 2020) pipeline in nextflow v21.04.1 (Di Tommaso et al., 2017) for RNA-seq data analysis. STAR (Spliced Transcripts Alignments to a Reference) v2.7.6a (Dobin et al., 2013) was used to align and map the sequencing reads to the reference genome. Gene transcript levels were determined using RSEM (RNA-Seq by Expectation-Maximization) v1.3.1 (B. Li & Dewey, 2011) and combined with the trimmed mean of M-value (TMM) normalization method. The differentially expressed genes (DEGs) were analyzed using edgeR v3.34.0 (Robinson et al., 2010), the significantly differentially expressed genes were filtered with the threshold of FDR < 0.05 and log2FC > 1.

Results

Triploid *Populus alba'Berolinensis* ' has a growth advantage

The polyploid level of the P. alba 'Berolinensis' was confirmed using a ploidy analyzer (Partec PA, Munster, Germany). Using its parental species P. alba, P. bolleana, and P. tremula as controls, P. alba 'Berolinensis' was confirmed to be triploid (Figure 1). We further compared the growth of P. alba 'Berolinensis' to its three diploid parental species and found that P. alba 'Berolinensis' had a growth advantage (Figure 2). The height of two-month-old P. alba 'Berolinensis' plantlets was 47.06 +- 2.86 cm. While the two genotypes of P. alba were 41.17 +- 2.79 and 32.64 +- 4.51 cm in heights, respectively. P. bolleana was 32.20 +- 3.75 cm in height, and the four genotypes of P. tremula were 35.64 +- 2.98, 39.72 +-2.87, 36.22 +- 4.27, and 37.21 +- 3.60 cm in heights, respectively. The stem diameter of the P. alba 'Berolinensis' was 0.39 +- 0.03 cm. The two genotypes of P. alba were 0.33 +- 0.04 and 0.22 +- 0.03 cm. P. bolleana was 0.27 +- 0.03 cm, and the four genotypes of P. tremula were 0.26 +- 0.03, 0.23 +- 0.03, 0.23 +- 0.04, and 0.21 +- 0.01 cm, respectively. The leaf area of P. alba 'Berolinensis' was 85.89 +- 9.27 cm2. The two genotypes P. alba were 53.34 +- 9.59 and 30.62 +- 6.95 cm2, respectively. P. bolleana was 48.20 +- 9.75 cm2, and the four genotypes of P. tremula were 58.84 +- 12.36, 40.09 +- 8.07, 44.16 +- 9.02, and 37.24 +- 9.36 cm2, respectively. The leaf length of the P. alba 'Berolinensis' was 12.13 +- 0.99 cm. The two genotypes of P. alba were 10.81 +- 1.29 and 8.39 +- 0.75 cm. P. bolleana was 10.68 +- 1.98 cm, and the four genotypes of P. tremula were 11.45 +- 1.03, 9.59 +- 1.03, 9.51 +- 1.12, and 8.71 +- 1.42 cm, respectively. The results indicated that the triploid P. alba 'Berolinensis' had a growth advantage in these respects over its parental plants.

De novo assembly of the *Populus alba 'Berolinensis* ' genome

We performed k-mer distribution analysis to investigate the genomic characteristics of P. alba 'Berolinensis'. The haploid genome size was estimated as 468.96 Mb based on the corrected PacBio CLR (continuous long reads). The estimated genome size is slightly larger than P. trichocarpa (Tuskan et al., 2006), P. alba (Liu et al., 2019), and P. simonii (H. Wu et al., 2020). Three distinctive peaks corresponding to haploid, diploid, and triploid were observed from the k-mer distribution curve (Figure S1). The heterozygosity of the genome was estimated as 3.15% to 3.76%, indicating 96.24% to 96.85% of sequence identity among the three haplotypes. Similar results were also obtained from the k-mer distribution from NGS short reads. Additional information on genome features was estimated by k-mer distribution (Table 1).

PacBio CLR reads and NGS short reads were used to assemble and polish the genome. 169.74 Gb long-reads, corresponding to 395x coverage of the haploid genome were generated from the PacBio Sequel II sequencing platform. The PacBio data contains 12.53 million reads with an average length of 13.54 kb and an N50 length of 22.82 kb. All the long reads then were used for correction. Some of them were trimmed due to being chimeric or having an erroneous sequence. Finally, a total of 103.78 Gb trimmed reads with an average length of 17.58 Kb were obtained, representing 241x coverage of the haploid genome. The NGS data were generated from the BGISEQ-500 platform and contained 377.40 million 150 bp paired-end reads (113.22 Gb), corresponding to 263x coverage of the haploid genome. After filtering low-quality and duplicated reads, 111.97 Gb of clean short reads were generated for genome polishing. Further details are listed in Table S1 and Table S2.

The assembly contigs were assembled using the trimmed PacBio CLR reads and polished using the clean NGS short reads. The final assembly that includes three haploid genomes is 1.56 Gb in length, with an N50 length of 1.63 Mb (Table 2). The genomic sequencing reads were mapped to the assembly to assess the assemble quality. 94.27% of the long reads and 99.87% of the NGS short reads were mapped to the assembly. Of the mapped short reads, 99.82% were properly pair mapped. ALLHiC (Zhang et al., 2019) was then used as scaffolding for the three haploid genomes by integrating 258.91 million read pairs of Hi-C data. The final assembly contains 1,469.18 Mb in 57 super-scaffolds, representing the 57 chromosomes in the three haplotypes. These chromosomes were further classified into three haploid genomes (A, B, and C) according

to the phylogenetic information available (Figure 3, S2, S3). Results showed that the three haploid genomes were close to P. alba, P. bolleana, and P. tremula.

BUSCO assessments indicated that the assembly covered at least 99.44% (1,605/1,614) of the embryophyte's single-copy ortholog datasets. The three haploid genomes contain 98.51% (1,590/1,614), 98.95% (1,597/1,614), and 99.26% (1,602/1,614) of the BUSCO genes. The Hi-C linkage map also showed that the chromosome groups were clearly defined (Figure 4, S4, S5). These results above confirm the well-organized chromosome-level assembly of P. alba 'Berolinensis' in this study.

Annotation of the *Populus alba 'Berolinensis* ' genome

The coding genes were annotated in the genome using the 515.87 million salinity stressed clean RNA-seq read pairs, 14 homologous species, and three de novo methods. 128,281 protein-coding genes were annotated in the genome. The three haploid genomes of P. alba 'Berolinensis' contained 43,255, 43,076, and 41,950 protein-coding genes. BUSCO analysis with the recently released plant dataset from OrthoDB was then applied to the annotated genes to assess completeness. 98.95% (1,597/1,614) of the core eukaryotic proteins were successfully annotated. The three haploid proteomes contain 97.83% (1,579/1,614), 98.02% (1,582/1,614), and 98.57% (1,591/1,614) BUSCO proteins. More than 80.35% of the protein-coding genes were functionally annotated by eggNOG v5.0 (Huerta-Cepas et al., 2019). In addition to the coding genes, annotation of repetitive sequences indicated that the P. alba 'Berolinensis' genome contained approximately 45.23% (664.48 Mb) repetitive sequences (Table 3).

Comparative analysis of the *Populus alba 'Berolinensis* ' genome

We inferred the phylogenic relationships of P. alba 'Berolinensis' and fourteen related species. The homology analysis for protein-coding genes among these species was carried out using OrthoFinder. 501 single-copy genes were identified, we then constructed the species tree using them via the maximum likelihood method. The three haploid genomes of P. alba 'Berolinensis' are closely related to P. alba, P. bolleana, and P. tremula (Figure 5d). We compared the protein-coding genes in P. alba 'Berolinensis' with those in P. trichocarpa, whose genome has been well annotated. 34,699 protein-coding genes account for 81.74% (A: 81.55%, B: 81.89%, C: 81.79%) in the P. alba 'Berolinensis' genome and have high similarities to 33,692 P. trichocarpa protein sequences. 66.44% (A: 66.14%, B: 66.50%, C: 66.68%) of the genes have $> 90\%$ sequence similarity with their homologs.

The three haploid genomes of P. alba 'Berolinensis' showed a high degree of gene collinearity. Synteny analysis using MCScanX revealed 1,365, 1,272, and 1,265 collinear fragments were between three groups of the chromosomes. The karyotypes of the three haploid genomes are likely to be conserved. We then mapped the clean NGS short reads to the three haploid genomes. Comparisons identified 31,552,252 SNPs and 6,522,107 InDels in the whole genome, with an average of 64.51 SNPs and 13.34 InDels per kilobase (Table 4).

The genomic comparison between P. alba 'Berolinensis' and its parental species was then performed by SyRI (Goel et al., 2019). We identified 321.28 Mb syntenic regions between P. alba genome and the haploid genome A, 319.63 Mb syntenic regions between P. bolleana genome and the haploid genome B, and 288.74 Mb syntenic regions between P. tremula genome and the haploid genome C, respectively (Figure S6). Further variant studies by SyRI identified 1,370,885 SNPs and 190,560 InDels located in exon regions, with an average of 6.63 SNPs and 0.92 InDels per kilobase, respectively.

In addition, we collected 100 samples of public WGS (Whole Genome Sequencing) raw data of the parental species from the Sequence Read Archive (SRA) database (NCBI Resource Coordinators et al., 2018) for the population genetic variation detection, including 88 individual samples and 12 mixed populations samples. The SNPs between the P. alba populations and the haploid genome A ranged from 2,879,457 (0.57%) to 7,002,530 (1.40%), which is slightly lower than the results reported by Liu et al. (2019). The SNPs between

the P. bolleana populations and the haploid genome B ranged from 1,951,890 (0.39%) to 4,707,654 (0.93%), which is similar to the heterozygosity level of P. bolleana reported by Ma et al. (2019). P. tremula has a large range of varieties, with a broad geographic distribution. The SNPs between the aspen populations and haploid genome C ranged from 5,354,053 (1.16%) to 12,861,231 (2.78%), which is similar to the results reported by Ingvarsson et al. (2020). Further statistics revealed that only 5.66% of these variants were located in exons and 23.59% were in promoters on average. Details are listed in Table S3.

Allelic variations among the three monoploid chromosomes

Genomic variation is an essential source of diversity for selection and breeding. Retaining resistance related variation from its parents may give P. alba 'Berolinensis' more resilience to harsh environments where its parental species cannot survive. To identify allelic variation among the three haploid genomes, we first manually screened for the collinear blocks from MCScanX and obtained 23,913 candidate allelic groups of genes present in the three haploid genomes. The subsequent polymorphism analysis showed an average of 25.73 nucleotide polymorphism sites per kilobase. Multiple sequence alignment revealed that many allelic genes contain structural variation. Some representative allele variants between the three haploid genomes are compared in Figure 6a.

One representative example is RESPONSIVE TO DEHYDRATION 22 (RD22), which is a molecular link between abscisic acid (ABA) signaling and abiotic stress in plants (Yamaguchi-Shinozaki et al., 1992; Yamaguchi-Shinozaki & Shinozaki, 1993). Its expression is a reliable ABA early response marker. Multiple sequence alignments indicated extensive variations between the RD22 alleles in P. alba 'Berolinensis' genome. For example, Poabs.B18G000441.v1.0 contained 21 SNPs and a 63-base deletion (235 to 297) compared to its alleles Poabs.A18G000432.v1.0 and Poabs.C18G000432.v1.0 (Figure 6a). The deletion does not result in a frameshift or a premature stop codon but results in a 21-amino-acid deletion (69 to 89) in the protein. The comparison of RD22 orthologs in their parental genome supports our conclusion. The RD22 protein orthologs in P. alba (Palba.scaffold217.34) and P. tremula (Potra2n18c32926.1) have the complete sequences, while a 21-amino-acid deletion in P. bolleana (PAYT022271.1) is consistent with Poabs.B18G000441.v1.0 (Figure 6b).

Another example is the CHALCONE ISOMERASE (CHI) gene, which encodes a chitinase involved in the ethylene/jasmonic acid-mediated signaling pathway during systemic acquired resistance (Jez et al., 2000; Muir et al., 2001). Compared with the CHI protein Poabs.B04G001821.v1.0, the other two allelic proteins have a serine insertion at base 63 to 64, destroying the original "GGGGGGGG" short tandem repeat sequence (Figure 6b). Furthermore, the Poabs.C04G001808.v1.0 also has two amino acid mutations at 272 bp and 276 bp, which fall in lysozyme-like domains. These variants were also verified by the variation identification in the populations of P. alba 'Berolinensis' parental species. The results indicated that the genome of the triploid poplar includes more genetic diversity.

In order to further identify the association between these allelic variations and those in the parental species, we performed a synteny analysis using MCScanX on genomes of P. alba, P. bolleana, and P. tremula. The comparisons among the three genomes showed a high degree of gene collinearity. The synteny analysis identified 31,417, 27,702, and 27,718 allelic genes in P. alba, P. bolleana, and P. tremula, respectively. Moreover, 1,311, 1,024, and 1,127 collinear blocks and a total of 113,790 allele pairs were identified between the three groups using whole-genome alignment (Figure S7). The results are consistent with the comparisons among the three haploid genomes of P. alba 'Berolinensis'. 93.45% of these allele pairs identified within these parental species were also observed in the triploid poplar P. alba 'Berolinensis'.

Effects of salinity stress on the three monoploid chromosomes

To further investigate the effects of allelic variation on gene expression, we grew plants under salinity stress and constructed and sequenced 18 cDNA libraries on the Illumina platform. The libraries contained six-time points and three replicates for each time point. We identified differentially expressed genes (DEGs) in the

three haploid genomes with thresholds FDR < 0.05 and log2FC > 1. A total of 4,768 allelic loci contained at least one DEG, and 8,573 DEGs were identified. The haploid genomes A, B, and C contained 2,930, 2,835, and 2,808 DEGs, respectively. About one-third of these allelic loci contained two DEGs. Of the 4,768 allelic loci, 1,838, 1,933, and 1,960 did not respond to salinity stress in the haploid genomes A, B, and C, respectively.

Promoters are typically upstream sequences of the coding strand of the transcribed gene. Cis-elements that play crucial roles in gene expression are usually located within promoter regions (Hernandez-Garcia & Finer, 2014). We extracted 2 kb of upstream sequences of the alleles and performed further analysis. The subsequent polymorphism analysis and multiple sequence alignments showed that each upstream sequence contained an average of 175.84 nucleotide polymorphism sites. The discrepancies of the upstream sequences among the three haploid genomes were about 8.79%, which is much higher than the genome heterozygosity. The upstream sequences are more variable, which may contribute to the allele specific expression under salinity stress. We performed a cis-element analysis on the upstream sequences and found that the three haploid genomes contained many different cis-elements.

For instance, LATE EMBRYOGENESIS ABUNDANT (LEA) is a plant protein abundant in seeds and vegetative tissues under stress conditions (Hundertmark & Hincha, 2008; Popova et al., 2011, 2015). Three homologs of LEA7 were identified in the P. alba 'Berolinensis' genome. AP2/ERF and DOF proteins are members of a major family of plant transcription factors. Recent studies (Gutterson & Reuber, 2004; Ruta et al., 2020; Xie et al., 2019; Yanagisawa, 2002) are disclosing their multiple roles in gene expression when associated with plant-specific phenomena including light, phytohormone and defense responses, seed development, and germination. Cis-elements analysis using PLACE showed that 21 AP2_ERF transcription factor binding sites were found within the upstream sequence of LEA7 in haploid genome C, whereas 16 and 18 were found in its two alleles in haploid genomes A and B. Similar patterns of variation were also found in DOF transcription factor binding sites. 35, 35, and 41 DOF binding sites were identified in each of three haploid genomes alleles, respectively. The expression of LEA7 in haploid genome C was more affected by salinity stress than its alleles (Figure 6c). Although the expressions of three genes were all increased under salinity stress, the increase rate in Poabs.C01G001835.v1.0 transcript abundance was still significantly greater than the other two alleles. The trimmed mean of M-value (TMM) of Poabs.C01G001835.v1.0 was up to 36.99 following 12 hours of salinity stress. However, the Poabs.A01G001939.v1.0 was only 4.57 and the Poabs.B01G001878.v1.0 was only 12.67 at the same time. The cis-elements quantitative advantage trend of LEA7 promoter was also identified in the extensive genetic diversity lineage of P. tremula.

Salinity stress will bring more abundant changes in P. alba 'Berolinensis' transcriptome compared to the diploid parents. As an allopolyploid, P. alba 'Berolinensis' exhibits novel patterns of gene expression, which result from epistatic interactions as well as combinations of individual gene expression of one parent and the expression level dominance of the other parental gene.

Development of *Populus alba 'Berolinensis '* genome database

TreeGenomes (*https://www.treegenomes.com*) is a manually curated database of *P. alba 'Berolinensis '* genome sequence and bioinformatic tools for analyzing polyploid omics data sets. All data are open and freely available for all users, including *P. alba 'Berolinensis '* genome information, an allele browser, and a gene expression viewer (Figure S6). The goal of providing TreeGenomes sequence databases and resources is to obtain an easily comprehensive understanding of *P. alba 'Berolinensis '* gene functions, identify possible engineering targets and facilitate breeding research. To achieve this, we used an open-source database PostgreSQL to model and store the genome information, and the modified open-source query tool Redash was used for the search service. This architecture could provide a natural way to query genetic entities, annotations, and their relationships. Meanwhile, database technology is an evolving field, so we will periodically reassess our technology choices to ensure that they scale to our service requirements.

Discussion

Polyploidization is a potential source of advantage for plants to accommodate abiotic stress, it was considered to contribute to the genetic and phenotypic novelty (Te Beest et al., 2012; Van de Peer et al., 2021). But polyploidy also has its limits, illustrated by its evolutionary stability and the contrast between animals and plants (Mable, 2004). The potentially disruptive effects caused by genomic instability, minority cytotype exclusion, and mitotic and meiotic abnormalities (Comai, 2005; Levin, 1975; Madlung et al., 2004) all could quickly remove the new polyploids from the population, especially in animals. Moreover, the ploidy asymmetry between female and male gametes and chromosomally unbalanced gametes hamper seed development and reduce polyploid fertility (Hojsgaard, 2018). In a natural world dominated by sexual reproduction, its heavy reliance on asexual reproduction dictates that it cannot compete with the reproductive efficiency of diploid species. Therefore, stably inherited polyploid organisms are still rare in natural conditions.

P. alba 'Berolinensis', also named "Yinzhong poplar", is a fast-growing and highly adaptable artificial allotriploid poplar hybrid, in which the three haploid chromosomes originated from three different ancestral poplars. In contrast to its parents, this species possesses distinct features, such as tolerance for cold, salinity, drought tolerance, and taller plants, all traits that were targets of artificial breeding. Artificially breeding has led to the substantial differences between three haploid chromosomes in the P. alba 'Berolinensis' genome, allowing us to separate the three haplotypes and explore further the reasons for the strong ecological adaptability of the triploid plants. The triploid hybrid species P. alba 'Berolinensis', with its new genome sequence, will be an excellent model to study polyploidy in plants.

Compared to its parents, P. alba 'Berolinensis' has more genetic diversity and stronger resistance to adversity. Genome sequencing of P. alba 'Berolinensis' suggested a large number of sequence differences between its three different haploid chromosome sets. Polymorphism analysis showed an average of up to 25.73 nucleotide polymorphism sites per kilobase. Subsequent salinity stress experiments demonstrated that some of the abiotic stress related allelic proteins differ in sequences and spatial structures because of their templates from the different haploid chromosomes. The differences in the three haploid genomes may affect the relative abundance of those proteins, resulting in improved plant resistance to adverse environments or greater resilience. The differences in parts of the ABA signaling pathway gene expression (e.g., RD22 and LEA7) between different haploid genomes may contribute to the novel features of P. alba 'Berolinensis'.

The accumulation of variants of parental alleles contributes to the high level of genetic diversity in P. alba 'Berolinensis' genome, offering possibilities for a more comprehensive exploration of phenotypic space. Alternatively, the nonlethal noncoding mutations due to the long-term vegetative propagation will also result in allelic expression differences and enrich the resources of the plants to maintain stoichiometry and homeostasis. These differences contribute to Populus species survival advantage in diverse environmental conditions and provide materials for cultivating a better poplar species. The salinity stress transcriptome data showed P. alba 'Berolinensis' gene expression to be more diversified, and it has made up for the adverse effects of insufficient monoallelic expression in the parental plants through hybridization. The different sources of P. alba 'Berolinensis' alleles have facilitated its adaptation to environmental changes and have provided variation for expression of a broad range of niche traits.

In conclusion, we present a high quality haplotype-resolved genome assemblies of the triploid P. alba 'Berolinensis'. These gene structure and expression information will provide valuable reference targets for further studying its abiotic stress essential genes. This study will also support the studies of allelic balance, population variation, and breeding of polyploid plants. Future studies would address the adaptation and evolution of genetic and epigenetic changes in this polyploid species, and leading to in-depth research on the coordination of its haploid genomes.

Acknowledgements

This work was supported by Heilongjiang Province Key Research and Development Program of China (GA21B010), Innovation Project of State Key Laboratory of Tree Genetics and Breeding (2015A01), Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team), and Zhejiang Provincial Natural Science Foundation of China (LY20C160010).

References

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* ,*6* (1), 11. https://doi.org/10.1186/s13100-015-0041-9

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome: Tair: Making and Mining the "Gold Standard" Plant Genome.*Genesis* , *53* (8), 474–485. https://doi.org/10.1002/dvg.22877

Besemer, J., & Borodovsky, M. (2005). GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* , *33* (Web Server), W451–W454. https://doi.org/10.1093/nar/gki487

Birchler, J. A., & Veitia, R. A. (2012). Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines.*Proceedings of the National Academy of Sciences* , *109* (37), 14746–14753. https://doi.org/10.1073/pnas.1207726109

Boeckmann, B. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* , *31* (1), 365–370. https://doi.org/10.1093/nar/gkg095

Broad Institute. (2017a). Gatk4-data-processing. *San Francisco (CA): GitHub; [Accessed 2020 June 1]* . https://github.com/gatk-workflows/gatk4-data-processing

Broad Institute. (2017b). Gatk4-germline-snps-indels. *San Francisco (CA): GitHub; [Accessed 2020 June 1]* . https://github.com/gatk-workflows/gatk4-germline-snps-indels

Brochmann, C., Brysting, A. K., Alsos, I. G., Borgen, L., Grundt, H. H., Scheen, A.-C., & Elven, R. (2004). Polyploidy in arctic plants: POLYPLOIDY IN ARCTIC PLANTS. *Biological Journal of the Linnean Society* , *82* (4), 521–536. https://doi.org/10.1111/j.1095-8312.2004.00337.x

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., & Yandell, M. (2007). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* , *18* (1), 188–196. https://doi.org/10.1101/gr.6743907

Chai, Y., Zhu, N., & Han, H. (2002). Dust removal effect of urban tree species in Harbin. *The Journal of Applied Ecology* , *13* (9), 1121–1126.

Chao, D.-Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., & Salt, D. E. (2013). Polyploids Exhibit Higher Potassium Uptake and Salinity Tolerance in Arabidopsis. *Science* , *341* (6146), 658–659. https://doi.org/10.1126/science.1240561

Chen, J., Huang, Y., Brachi, B., Yun, Q., Zhang, W., Lu, W., Li, H., Li, W., Sun, X., Wang, G., He, J., Zhou, Z., Chen, K., Ji, Y., Shi, M., Sun, W., Yang, Y., Zhang, R., Abbott, R. J., & Sun, H. (2019). Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nature Communications* ,*10* (1), 5230. https://doi.org/10.1038/s41467-019-13128-y

Chen, S., & Polle, A. (2009). Salinity tolerance of *Populus* : Salinity tolerance of *Populus* . *Plant Biology* ,*12* (2), 317–333. https://doi.org/10.1111/j.1438-8677.2009.00301.x

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118} ; iso-2; iso-3. *Fly* , *6* (2), 80–92. https://doi.org/10.4161/fly.19695

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology* , *35* (4), 316–319. https://doi.org/10.1038/nbt.3820

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* , *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, L., Wang, J., & Wang, G. (2020). BYASE: A Python library for estimating gene and isoform level allele-specific expression. Bioinformatics, 36(19), 4955–4956. https://doi.org/10.1093/bioinformatics/btaa636

Du, K., Liao, T., Ren, Y., Geng, X., & Kang, X. (2020). Molecular Mechanism of Vegetative Growth Advantage in Allotriploid *Populus* .*International Journal of Molecular Sciences* , *21* (2), 441. https://doi.org/10.3390/ijms21020441

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). *De novo* assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* ,*356* (6333), 92–95. https://doi.org/10.1126/science.aal3327

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* , *3* (1), 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* , *20* (1), 238. https://doi.org/10.1186/s13059-019-1832-y

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* , *38* (3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* , *117* (17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). *De novo*transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* ,*8* (8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* ,*9* (1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Hallab, A. (2015). *Protein Function Prediction Using Phylogenomics, Domain Architecture Analysis, Data Integration, and Lexical Scoring* .

Henry, I. M., Zinkgraf, M. S., Groover, A. T., & Comai, L. (2015). A System for Dosage-Based Functional Genomics in Poplar. *The Plant Cell* , *27* (9), 2370–2383. https://doi.org/10.1105/tpc.15.00349

Huang, S., He, X., Wang, G., & Bao, E. (2021). AlignGraph2: Similar genome-assisted reassembly pipeline for PacBio long reads. Briefings in Bioinformatics, 22(5), bbab022. https://doi.org/10.1093/bib/bbab022

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.*Molecular Biology and Evolution* , *34* (8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* , *47* (D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Hundertmark, M., & Hincha, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana* . *BMC Genomics* , *9* (1), 118. https://doi.org/10.1186/1471-2164-9-118

Jackson, S. A. (2017). Epigenomics: Dissecting hybridization and polyploidization. *Genome Biology* , *18* (1), 117. https://doi.org/10.1186/s13059-017-1254-7

Jez, J. M., Bowman, M. E., Dixon, R. A., & Noel, J. P. (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Structural Biology* , *7* (9), 786–791. https://doi.org/10.1038/79025

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.*Molecular Biology and Evolution* , *30* (4), 772–780. https://doi.org/10.1093/molbev/mst010

Kokot, M., Długosz, M., & Deorowicz, S. (2017). KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* , *33* (17), 2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. *Genome Research* , *27* (5), 722–736. https://doi.org/10.1101/gr.215087.116

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* , *5* (1), 1–9.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* , *35* (21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* , *9* (4), 357–359. https://doi.org/10.1038/nmeth.1923

Levin, D. A. (1983). Polyploidy and Novelty in Flowering Plants.*The American Naturalist* , *122* (1), 1–25. https://doi.org/10.1086/284115

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* , *12* (1), 323. https://doi.org/10.1186/1471-2105-12-323

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences.*Bioinformatics* , *34* (18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, Y., Su, X., Zhang, B., Huang, Q., Zhang, X., & Huang, R. (2008). Expression of jasmonic ethylene responsive factor gene in transgenic poplar tree leads to increased salt tolerance. *Tree Physiology* ,*29* (2), 273–279. https://doi.org/10.1093/treephys/tpn025

Lin, Y.-C., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., Zuccolo, A., Nystedt, B., Hvidsten, T. R., de la Torre, A., Cossu, R. M., Hoeppner, M. P., Lantz, H., Scofield, D. G., Zamani, N., Johansson, A., Mannapperuma, C., Robinson, K. M., Mähler, N., Leitch, I. J., … Street, N. R. (2018). Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American

and European aspen. *Proceedings of the National Academy of Sciences* , *115* (46), E10970–E10978. https://doi.org/10.1073/pnas.1801437115

Liu, Y.-J., Wang, X.-R., & Zeng, Q.-Y. (2019). *De novo* assembly of white poplar genome and genetic diversity of white poplar population in Irtysh River basin in China. *Science China Life Sciences* ,*62* (5), 609–618. https://doi.org/10.1007/s11427-018-9455-2

Lourkisti, R., Froelicher, Y., Herbette, S., Morillon, R., Tomi, F., Gibernau, M., Giannettini, J., Berti, L., & Santini, J. (2020). Triploid Citrus Genotypes Have a Better Tolerance to Natural Chilling Conditions of Photosynthetic Capacities and Specific Leaf Volatile Organic Compounds. *Frontiers in Plant Science* , *11* , 330. https://doi.org/10.3389/fpls.2020.00330

Ma, J., Wan, D., Duan, B., Bai, X., Bai, Q., Chen, N., & Ma, T. (2019). Genome sequence and genetic transformation of a widely distributed and cultivated poplar. *Plant Biotechnology Journal* , *17* (2), 451–460. https://doi.org/10.1111/pbi.12989

Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y., Liu, B., Qiu, Q., Wang, Z., Zhang, J., Wang, K., Jiang, D., Gou, C., Yu, L., Zhan, D., Zhou, R., Luo, W., Ma, H., Yang, Y., ... Liu, J. (2013). Genomic insights into salt adaptation in a desert poplar. *Nature Communications* , *4* (1), 2797. https://doi.org/10.1038/ncomms3797

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.*Genome Research* , *20* (9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Miller, M., Song, Q., Shi, X., Juenger, T. E., & Chen, Z. J. (2015). Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. *Nature Communications* , *6* (1), 7453. https://doi.org/10.1038/ncomms8453

Muir, S. R., Collins, G. J., Robinson, S., Hughes, S., Bovy, A., Ric De Vos, C. H., van Tunen, A. J., & Verhoeyen, M. E. (2001). Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nature Biotechnology* ,*19* (5), 470–474. https://doi.org/10.1038/88150

NCBI Resource Coordinators, Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L. Y., ... Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* , *46* (D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Ni, D. A., Sozzani, R., Blanchet, S., Domenichini, S., Reuzeau, C., Cella, R., Bergounioux, C., & Raynaud, C. (2009). The Arabidopsis MCM2 gene is essential to embryo development and its over-expression alters root meristem function. *New Phytologist* , *184* (2), 311–322. https://doi.org/10.1111/j.1469-8137.2009.02961.x

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* , *20* (1), 275. https://doi.org/10.1186/s13059-019-1905-y

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., & Buell, C. R. (2007). The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research* ,*35* (Database), D883–D887. https://doi.org/10.1093/nar/gkl976

Popova, A. V., Hundertmark, M., Seckler, R., & Hincha, D. K. (2011). Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the

presence of membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* ,*1808* (7), 1879–1887. https://doi.org/10.1016/j.bbamem.2011.03.009

Popova, A. V., Rausch, S., Hundertmark, M., Gibon, Y., & Hincha, D. K. (2015). The intrinsically disordered protein LEA7 from *Arabidopsis thaliana* protects the isolated enzyme lactate dehydrogenase and enzymes in a soluble leaf proteome during freezing and drying. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* , *1854* (10), 1517–1525. https://doi.org/10.1016/j.bbapap.2015.05.002

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* , *11* (1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Renny-Byfield, S., & Wendel, J. F. (2014). Doubling down on genomes: Polyploidy and crop plants. *American Journal of Botany* ,*101* (10), 1711–1725. https://doi.org/10.3732/ajb.1400119

Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., & Mayrose, I. (2019). The global biogeography of polyploid plants. *Nature Ecology & Evolution* , *3* (2), 265–273. https://doi.org/10.1038/s41559-018-0787-9

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26* (1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ruiz, M., Quiñones, A., Martínez-Cuenca, M. R., Aleza, P., Morillon, R., Navarro, L., Primo-Millo, E., & Martínez-Alcántara, B. (2016). Tetraploidy enhances the ability to exclude chloride from leaves in carrizo citrange seedlings. *Journal of Plant Physiology* ,*205* , 1–10. https://doi.org/10.1016/j.jplph.2016.08.002

Song, M. J., Potter, B. I., Doyle, J. J., & Coate, J. E. (2020). Gene Balance Predicts Transcriptional Responses Immediately Following Ploidy Change in *Arabidopsis thaliana* . *The Plant Cell* ,*32* (5), 1434–1448. https://doi.org/10.1105/tpc.19.00832

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research* , *34* (Web Server), W435–W439. https://doi.org/10.1093/nar/gkl200

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* ,*12* (1), 2. https://doi.org/10.1186/s13100-020-00230-y

Tang, R.-J., Liu, H., Bao, Y., Lv, Q.-D., Yang, L., & Zhang, H.-X. (2010). The woody plant poplar has a functionally conserved salt overly sensitive pathway in response to salinity stress. *Plant Molecular Biology* , *74* (4–5), 367–380. https://doi.org/10.1007/s11103-010-9680-x

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics* , *25* (1). https://doi.org/10.1002/0471250953.bi0410s25

The French–Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* ,*449* (7161), 463–467. https://doi.org/10.1038/nature06148

Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., & Gao, G. (2019). PlantRegMap: Charting functional regulatory maps in plants.*Nucleic Acids Research* , gkz1020. https://doi.org/10.1093/nar/gkz1020

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., ... Rokhsar, D. (2006). The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray).*Science* , *313* (5793), 1596–1604. https://doi.org/10.1126/science.1128691

Van de Peer, Y., Ashman, T.-L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy: An evolutionary and ecological force in stressful times. *The Plant Cell* , *33* (1), 11–26. https://doi.org/10.1093/plcell/koaa015

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* , *9* (11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, J., Chen, S., Dong, L., & Wang, G. (2021). CHTKC: A robust and efficient k-mer counting algorithm based on a lock-free chaining hash table. Briefings in Bioinformatics, 22(3), bbaa063. https://doi.org/10.1093/bib/bbaa063

Wang, J., Huo, B., Liu, W., Li, D., & Liao, L. (2017). Abnormal meiosis in an intersectional allotriploid of *Populus* L. and segregation of ploidy levels in 2x x 3x progeny. *PLOS ONE* , *12* (7), e0181767. https://doi.org/10.1371/journal.pone.0181767

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. -h., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* ,*40* (7), e49–e49. https://doi.org/10.1093/nar/gkr1293

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* , *35* (3), 543–548. https://doi.org/10.1093/molbev/msx319

Wendel, J. F., Lisch, D., Hu, G., & Mason, A. S. (2018). The long and short of doubling down: Polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* , *49* , 1–7. https://doi.org/10.1016/j.gde.2018.01.004

Wu, H., Yao, D., Chen, Y., Yang, W., Zhao, W., Gao, H., & Tong, C. (2020). De Novo Genome Assembly of *Populus simonii* Further Supports That *Populus simonii* and *Populus trichocarpa*Belong to Different Sections. *G3 Genes|Genomes|Genetics* , *10* (2), 455–466. https://doi.org/10.1534/g3.119.400913

Wu, S., Cheng, J., Xu, X., Zhang, Y., Zhao, Y., Li, H., & Qiang, S. (2019). Polyploidy in invasive *Solidago canadensis* increased plant nitrogen uptake, and abundance and activity of microbes and nematodes in soil. *Soil Biology and Biochemistry* , *138* , 107594. https://doi.org/10.1016/j.soilbio.2019.107594

Yamaguchi-Shinozaki, K., Koizumi, M., Urao, S., & Shinozaki, K. (1992). Molecular Cloning and Characterization of 9 cDNAs for Genes That Are Responsive to Desiccation in *Arabidopsis thaliana* : SequenceAnalysis of One cDNA Clone That Encodes a Putative Transmembrane Channel Protein. *Plant and Cell Physiology* , *33* (3), 217–224. https://doi.org/10.1093/oxfordjournals.pcp.a078243

Yamaguchi-Shinozaki, K., & Shinozaki, K. (1993). The plant hormone abscisic acid mediates the drought-induced expression but not the seed-specific expression of *rd22* , a gene responsive to dehydration stress in Arabidopsis thaliana. *Molecular and General Genetics MGG* , *238–238* (1–2), 17–25. https://doi.org/10.1007/BF00279525

Yang, P. M., Huang, Q. C., Qin, G. Y., Zhao, S. P., & Zhou, J. G. (2014). Different drought-stress responses in photosynthesis and reactive oxygen metabolism between autotetraploid and diploid rice. *Photosynthetica* , *52* (2), 193–202. https://doi.org/10.1007/s11099-014-0020-2

Yang, W., Wang, K., Zhang, J., Ma, J., Liu, J., & Ma, T. (2017). The draft genome sequence of a desert tree *Populus pruinosa* .*GigaScience* , *6* (9). https://doi.org/10.1093/gigascience/gix075

Zhang, X., Chen, S., Shi, L., Gong, D., Zhang, S., Zhao, Q., Zhan, D., Vasseur, L., Wang, Y., Yu, J., Liao, Z., Xu, X., Qi, R., Wang, W., Ma, Y., Wang, P., Ye, N., Ma, D., Shi, Y., ... You, M. (2021). Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis* . *Nature Genetics* , *53* (8), 1250–1259. https://doi.org/10.1038/s41588-021-00895-y

Zhang, X., Zhang, S., Zhao, Q., Ming, R., & Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* , *5* (8), 833–845. https://doi.org/10.1038/s41477-019-0487-8

Zhu, H., Zhao, S., Lu, X., He, N., Gao, L., Dou, J., Bie, Z., & Liu, W. (2018). Genome duplication improves the resistance of watermelon root to salt stress. *Plant Physiology and Biochemistry* , *133* , 11–21. https://doi.org/10.1016/j.plaphy.2018.10.019

Zou, Q., Hu, Q., Guo, M., & Wang, G. (2015). HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy. Bioinformatics, 31(15), 2475–2481. https://doi.org/10.1093/bioinformatics/btv177

Data Accessibility Statement

Data supporting the findings of this work are available within the paper and its supplementary files. The Populus alba 'Berolinensis' genome assembly and annotations have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.16902892), CNCB Genome Warehouse (BioProject: PRJCA009923), and NCBI GenBank (BioProject: PRJNA845083, PRJNA845082, and PRJNA845081). Raw sequencing data are available in the NCBI Sequence Read Archive (SRA) database (BioProject: PRJNA635615 and PRJNA721517).

Author Contributions

S. C. (Su Chen), G. Q., and G. W. conceived and supervised the study. Y. Y., X. W., S. W., T. Z., Y. Z., R. H., N. M., Y. W., W. L., Z. L., and Q. G. collected the samples and data. J. L. and H. H. provided experimental guidance. S. C. (Song Chen) performed the analyses and wrote the manuscript. R. S. and S. C. (Su Chen) revised the manuscript. All the authors have read and approved the final manuscript.

Tables and Figures

Table 1 Summary of the k -mer analysis for estimating the genome size of *P. alba* '*Berolinensis* '

	Pacbio Sequel II	BGISEQ-500
Heterozygosity	3.76%	3.15%
Genome Haploid Length	468,959,062 bp	431,981,358 bp
Genome Repeat Length	369,406,124 bp	228,163,764 bp
Genome Unique Length	99,552,938 bp	203,817,593 bp
Model Fit	91.58%	90.20%
Read Error Rate	0.35%	0.15%

Table 2 Statistics of genome and subgenome assembly of *P. alba* '*Berolinensis* '

	PacBio assembly	Hi-C assembly	Subgenome A	Subgenome B	Subgenome C
N90	38,288	19,144,057	19,843,081	19,913,313	17,228,702
L90	5,674	49	17	17	17
N75	89,961	20,922,378	21,476,760	20,922,378	19,144,057
L75	1,279	38	13	13	13

	PacBio assembly	Hi-C assembly	Subgenome A	Subgenome B	Subgenome C
N50	1,631,398	26,226,177	26,936,469	26,226,177	24,238,767
L50	170	22	8	8	8
Maximum length	24,784,932	62,333,478	60,589,463	62,333,478	55,171,902
Minimum length	2,147	16,129,400	18,697,356	19,475,028	16,129,400
Average length	133,244	25,775,032	26,418,080	26,593,606	24,313,412
Total length	1,556,566,804	1,469,176,872	501,943,521	505,278,517	461,954,834
Total N length	0	1,956,300	746,300	769,200	440,800
Total sequences	11,682	57	19	19	19
GC (%)	38.26	36.11	36.21	36.35	35.74

Table 3 Classification of repetitive elements in the *P. alba* '*Berolinensis* ' genome

	Number of elements	Length occupied	Percentage of sequence
Retroelements	617,858	283,318,362	19.28%
LINEs	2,971	1,897,312	0.13%
LTR elements	614,887	281,421,050	19.16%
Ty1/Copia	122,932	57,683,998	3.93%
Gypsy/DIRS1	243,329	148,617,876	10.12%
DNA transposons	725,548	281,369,782	19.15%
hobo-Activator	1,142	629,205	0.04%
Tourist/Harbinger	108	16,646	0.00%
Unclassified	404,753	99,796,853	6.79%
Total interspersed repeats	2,733,528	664,484,997	45.23%
Genome size	-	1,469,176,872	-

+ The names of the main classes of repetitive elements are shown in bold.

Table 4 Classification of variant effects by region in the *P. alba* '*Berolinensis* ' genome

	Subgenome A	Subgenome B	Subgenome C
Exon	1,802,308	1,788,037	1,825,586
Intron	6,877,435	6,784,399	6,873,962
Intergenic	9,351,751	9,318,644	8,999,860
Upstream	7,648,308	7,637,644	7,270,580
Downstream	7,080,570	7,047,241	6,823,592
UTR 3' prime	669,893	658,962	663,487
UTR 5' prime	431,654	433,611	433,104

Figure 1

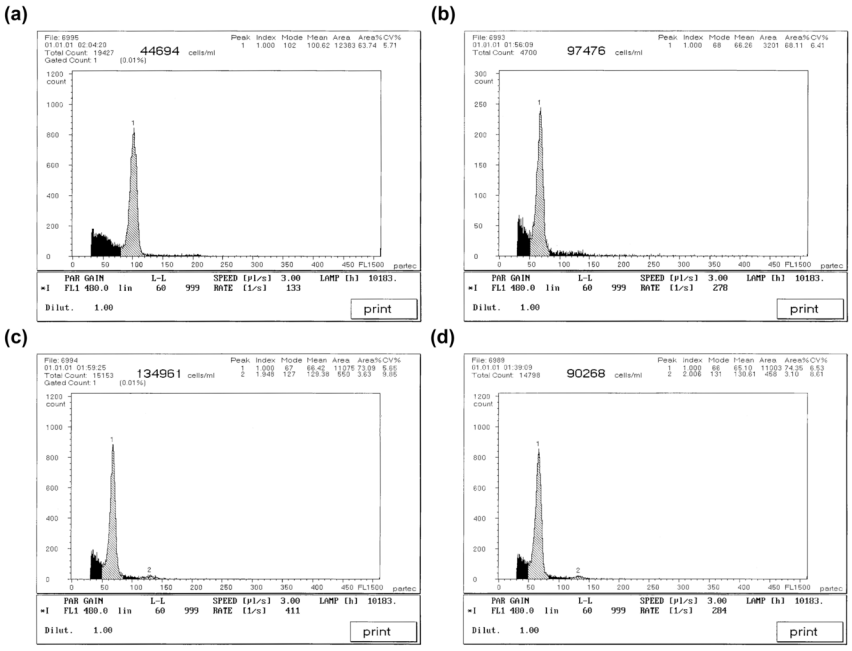

Fig. 1 Flow cytometry analysis of the *P. alba* '*Berolinensis* ' and its parents. **(a)** Histogram of the triploid *P. alba* '*Berolinensis* ' (3n = 57). **(b)** Histogram of the diploid *P. alba* (2n = 38). **(c)** Histogram of the diploid *P. bolleana* (2n = 38). **(d)** Histogram of the diploid *P. tremula* (2n = 38).

Figure 2

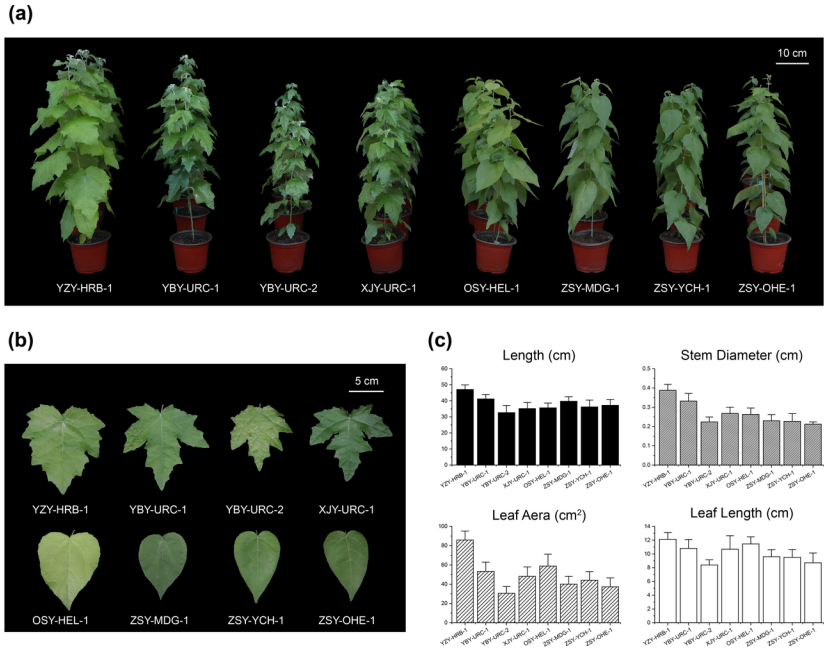

Fig.2 Plant phenotype assays of 60 days *P. alba*'*Berolinensis* ' and its parentals, the mean values ± SD are

all calculated from thirty replicates. **(a)** Photographs of seedlings. **(b)** Photographs of leaves. **(c)** Lengths (cm), stem diameters (cm), leaf areas (cm^2), and leaf lengths (cm) of different genotypes.

Figure 3

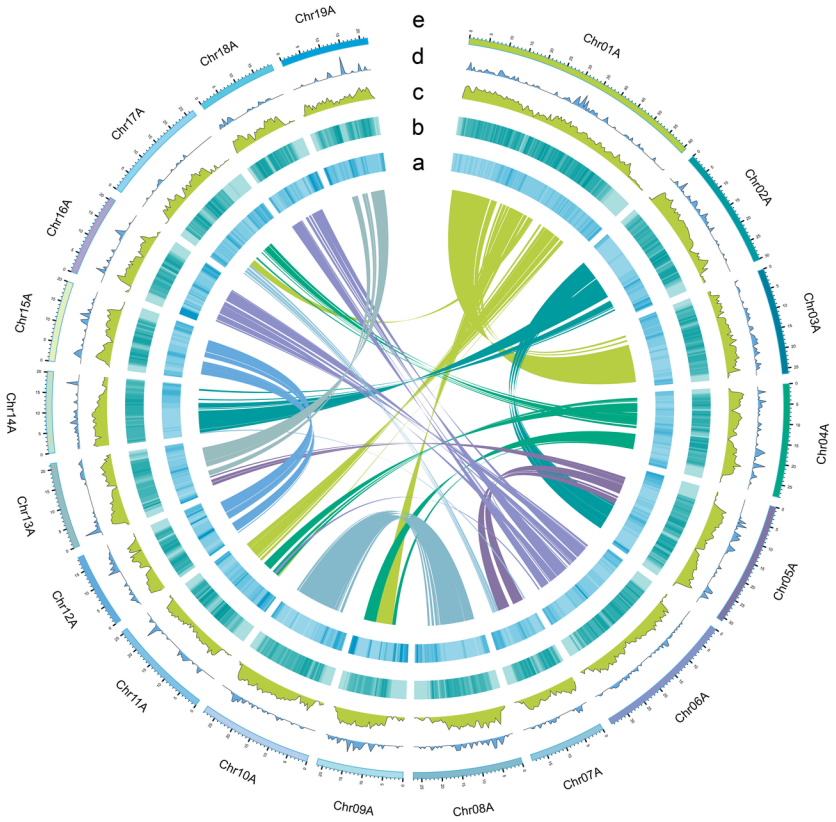

Fig.3 Chromosome features and variant distributions. The center lines represent syntenic genes pairs, the bandwidth is proportional to syntenic block size. Notes in circos plots (from outermost to innermost): **(a)** Transposable element abundance. **(b)** Germline short variant (SNPs + indels).**(c)** Gene density. **(d)** The density of tRNA. **(e)** The 19 chromosomes of Subgenome A.

Figure 4

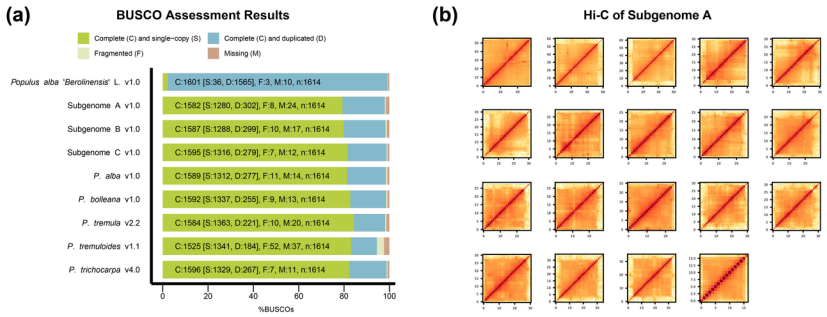

Fig.4 Genome assembly quality and completeness estimation. **(a)** BUSCO analysis of the genomes from

Populus alba'Berolinensis ', *P. alba* , *P. bolleana* , *P. tremula* , *P. tremuloides* , and *P. trichocarpa* . (C) complete; (S) single-copy; (D) duplicated; (F) fragmented; (M) missing. **(b)** Hi-C interaction matrix maps of the Subgenome A.

Figure 5

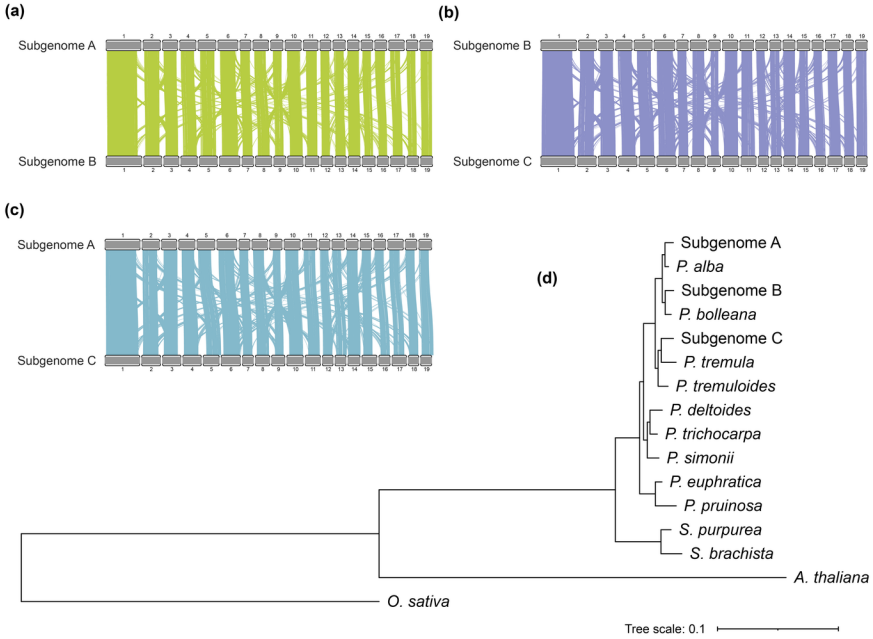

Fig.5 The evolutionary relationships of the *Populus*species. **(a)** Genomic alignments between Subgenome A and Subgenome B. **(b)** Genomic alignments between Subgenome B and Subgenome C. **(c)** Genomic alignments between Subgenome A and Subgenome C. **(d)** Species phylogeny was inferred with RAxML-NG using genes with one-to-one relationships across *Salicaceae* genomes and with 1000 bootstraps. The phylogenetic tree is based on 14 species: *Populus alba 'Berolinensis* ', *P. alba* , *P. bolleana* , *P. tremula* , *P. tremuloides* , *P. trichocarpa* , *P. deltoides* , *P.simonii* , *P. euphratica* , *P. pruinosa* , *S. purpurea* , *S. brachista* , *A. thaliana* , and *O. sativa* as outgroup species.

Figure 6

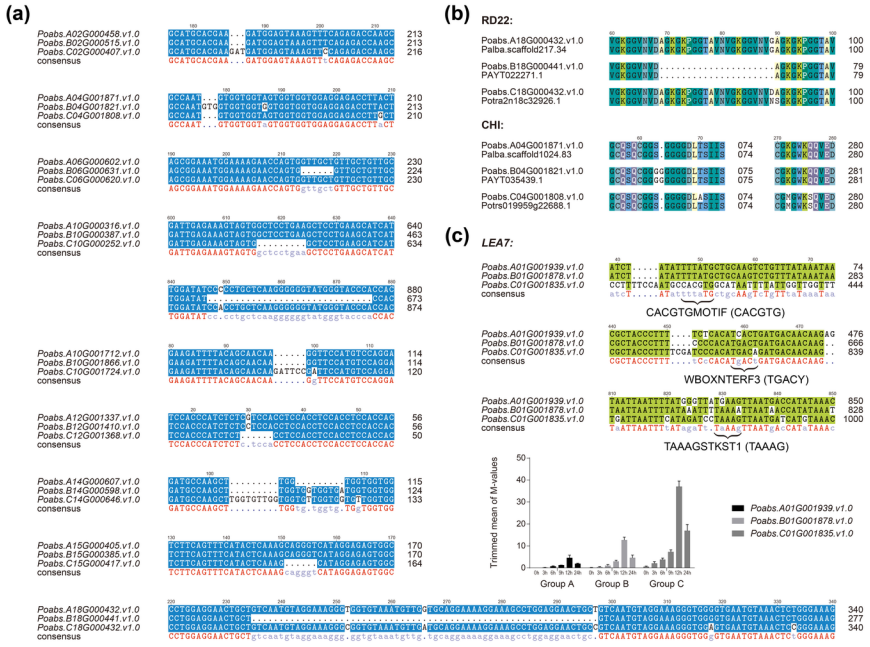

Fig.6 Allelic variations analysis in *P. alba'Berolinensis* ' and *Populus* species. **(a)** Partial representative allelic variations in *P. alba'Berolinensis* '. **(b)** Protein-coding variation caused by allelic variation of *RD22* and *CHI* in *P. alba'Berolinensis* ' and its parentals. **(c)** Promoter allele variations and gene expression changes of *LEA7* in *P. alba'Berolinensis* '. The five samples from left to right in each group are sorted by time-course of salinity stress, which is 0h, 3h, 6h, 9h, 12h, and 24h, respectively.