Transcriptome analysis of diploid and triploid *Populus tomentosa*

Wen Bian, Xiaozhen Liu, Zhiming Zhang, Hanyao Zhang

1 Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming, Yunnan, China
2 Key Laboratory of Biodiversity Conservation in Southwest China, State Forest Administration, Southwest Forestry University, Kunming, Yunnan, China

Corresponding Author: Hanyao Zhang
Email address: zhanghanyao@swfu.edu.cn

Triploid Chinese white poplar (*Populus tomentosa* Carr., Salicaceae) has stronger advantages in growth and better stress resistance and wood quality than diploid *P. tomentosa*. Using transcriptome sequencing technology to identify candidate transcriptome-based markers for growth vigor in young tree tissue is of great significance for the breeding of *P. tomentosa* varieties in the future. In this study, the cuttings of diploid and triploid *P. tomentosa* were used as plant materials, transcriptome sequencing was carried out, and their tissue culture materials were used for RT-qPCR verification of the expression of genes. The results showed that 12,240 differentially expressed genes in diploid and triploid *P. tomentosa* transcripts were annotated and enriched into 135 metabolic pathways. The top six pathways that enriched the most significantly different genes were plant-pathogen interaction, phenylpropanoid biosynthesis, MAPK signalling pathway-plant, ascorbate and aldarate metabolism, diterpenoid biosynthesis, and betalain biosynthesis pathway. Ten growth-related genes were selected from pathways of plant hormone signal transduction and carbon fixation in photosynthetic organisms for RT-qPCR verification. The expression levels of *MDH* and *CYCD3* in tissue-cultured and greenhouse planted triploid *P. tomentosa* were higher than those in tissue-cultured diploid *P. tomentosa*, which was consistent with the TMM values calculated by transcriptome.
Transcriptome analysis of diploid and triploid *Populus tomentosa*

Wen Bian¹,#, Xiaozhen Liu²,#, Zhiming Zhang², Hanyao Zhang¹,*

¹Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, Kunming Yunnan, China;
²Key Laboratory of Biodiversity Conservation in Southwest China, State Forest Administration, Southwest Forestry University, Kunming, Yunnan, China.

*equal first author.

*Corresponding author:

Professor Hanyao Zhang, PhD

*Key Laboratory for Forest Resources Conservation and Utilization in the Southwest Mountains of China, Ministry of Education, Southwest Forestry University, 300# Bailongsi, Kunming Yunnan 650224, China

Email address: zhanghanyao@hotmail.com

Phone number: (+86 871 63863575)

Fax number: (+86 871 63863022)

Abstract

Triploid Chinese white poplar (*Populus tomentosa* Carr., Salicaceae) has stronger advantages in growth and better stress resistance and wood quality than diploid *P. tomentosa*. Using transcriptome sequencing technology to identify candidate transcriptome-based markers for growth vigor in young tree tissue is of great significance for the breeding of *P. tomentosa* varieties in the future. In this study, the cuttings of diploid and triploid *P. tomentosa* were used as plant materials, transcriptome sequencing was carried out, and their tissue culture materials were used for RT-qPCR verification of the expression of genes. The results showed that 12,240 differentially expressed genes in diploid and triploid *P. tomentosa* transcripts were annotated and enriched into
135 metabolic pathways. The top six pathways that enriched the most significantly different genes were plant-pathogen interaction, phenylpropanoid biosynthesis, MAPK signalling pathway-plant, ascorbate and aldarate metabolism, diterpenoid biosynthesis, and betalain biosynthesis pathway. Ten growth-related genes were selected from pathways of plant hormone signal transduction and carbon fixation in photosynthetic organisms for RT-qPCR verification. The expression levels of MDH and CYCD3 in tissue-cultured and greenhouse planted triploid P. tomentosa were higher than those in tissue-cultured diploid P. tomentosa, which was consistent with the TMM values calculated by transcriptome.

Key words: Populus tomentosa; transcriptome analysis; GO categories; KEGG categories; RT-qPCR analysis.

1 Introduction

Populus tomentosa is an important timber and ecological tree species in northern China, and it is one of the most important fast-growing native tree species in farmland shelterbelt, timber forest, and landscape tree from North China to Northwest China (Fan et al., 2005; Li et al., 2019a). In terms of wood, it has good wood quality (with long fiber and white wood), high-fiber content, good paint, and cementation properties. In terms of growth characteristics, it has fast growth, long life (compared to other trees in northern China), drought and salinity resistance, strong smoke and pollution resistance, and a crown shape (Ci et al., 2019). Based on its excellent characteristics and the needs of ecological and economic construction, researchers have paid more and more attention to the field of forest breeding in China (Zong et al., 2019).

At present, due to the problems of high genetic heterozygosity and the long breeding cycle of forest trees, the genetic improvement process of conventional cross-breeding is seriously hindered, and the productivity of varieties used in forestry production is difficult to meet the needs of industrial development for wood, biomass, paper, fuel and biomaterials (Harfouche et al., 2012; Zong et al., 2019). An in-depth study of the biological basis of wood formation and analysis of the genetic regulation mechanism of wood formations was essential to accelerate the process of improved forest varieties (Lin et al., 2017). The study of tree transcriptomes enables us to analyze the genetic basis of the formation of tree traits (Hao et al., 2011; Zhang et al., 2019; Sun et al., 2019). Because of long cycles, breeding of new varieties with expected traits cannot be rapidly achieved through traditional crossing methods alone (Harfouche et al., 2012). The combination of conventional breeding and...
modern breeding techniques represented by genomics is the main trend in the field of forest breeding in the future (Harfouche et al., 2012). Combining conventional breeding methods with transcriptome sequencing technology to screen tree growth-related genes is not only a breakthrough in variety selection but also a new field of exploration, and to identify candidate transcriptome-based markers for growth vigor in young tree tissue is very important (Hao et al., 2011; Sun et al., 2019).

Polyploidy is important for the evolution of plants (Sattler et al., 2016; Liu and Sun, 2019). RNA interference and dosage compensation in a polyploid cell often leads to epigenetic changes as it alters the gene expression levels (Osborn et al., 2003; Soltis et al., 2004). Many studies have found significant differences between the expression levels of genes in diploid and polyploid plants (Osborn et al., 2003; Gutierrez-Gonzalez and Garvin, 2017; Li et al., 2019b). Polyploid plants often exhibit commercially beneficial qualities, e.g. increased vigour, improved product quality, enlarged organs, enhanced tolerance to both biotic and abiotic stresses and increased heterozygosity and heterosis, in contrast to their diploid relatives (Sattler et al., 2016; Liu and Sun, 2019). Polyploidy often results in downregulated fertility because the expression of fertility-related genes is lower than that of compared to their diploid relatives (Li et al., 2019b). Developing polyploidy in plants has been the focus of many plant breeders for some time (Huang et al., 1990; Bancroft et al., 2011; Rambani et al., 2014; Li et al., 2019b; Shenton et al., 2020).

There are many advantages of triploid *P. tomentosa* plants compared with diploid ones (Zhu et al., 1995; Chen et al., 2017). Studies have shown that triploid *P. tomentosa* is superior to diploid *P. tomentosa* in both volumes per plant and papermaking (Chen et al., 2017; Li and Zhang, 2000). *P. tomentosa* has a great advantage in the construction of shelterbelt and fast-growing and high-yield forest (Li et al., 2019a). However, the related studies on the differences in growth between diploid and triploid *P. tomentosa* are mostly focused on phenotypic studies, while there are few studies on the genes that control the differences in their growth characteristics. In this study, the transcriptome analysis was subjected to an analysis of the differences in gene expression between diploid and triploid *P. tomentosa*, which derived from genome doubling and hybridization with different genotypes (Zhang et al., 2008), and RT-qPCR analysis was used for verification of the results of the transcriptome analysis. It would lay a foundation to screen growth-related genes for providing high-quality, fast-growing industrial timber and speeding up the genetic improvement of forest trees.
Materials and Methods

2.1 Materials

Diploid and triploid *P. tomentosa* root sprouts were collected in Kunming World Horticultural Expo Park and on the campus of Southwest Forestry University in May 2019. These sites were adjacent to each other. The triploid poplar [(*P. tomentosa* × *P. bolleana*) × *P. tomentosa*] was produced from Beijing Forestry University (Zhang et al., 2008), whereas the diploid plants were native to the area. To obtain samples of the same physiological age, about one-year-old (variability in their ages about one month) well-grown root sprouts were selected and sampled at 9 a.m. The leaves were removed, and about one gram of stem segments under the fifth leaf down from the tip of the branch was sent to Wuhan Huada Gene Company for transcriptome sequencing (Ye et al., 2020). Among them, diploid *P. tomentosa* stem segments were used as the control group (D1, D2, D3), and triploid *P. tomentosa* stem segments were considered as the treatment group (T1, T2, T3). D1, D2, D3, T1, T2 and T3 were used to represent biological replicates.

2.2 Sequencing data filtering

RNA was extracted using the Tiangen kit (TIANGEN Biotech Beijing Co., Ltd.), which was produced by the Wuhan Huada Gene Company. The library was generated according to the operation manual of the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). The samples were sequenced using the Bgiseq-500 platform. The data were paired-end, and fastq formatted files were used at the beginning of the analysis. The data was filtered using Trimmomatic v0.36 software (Bolger et al., 2014), and Soapnuke v1.4.0 (Chen et al., 2018) was used for statistics. A trimming of the polyA tails was conducted, and a cut-off threshold of Q-score 30 was chosen and bases with Q-score less than 30 were trimmed.

2.3 Gene annotation

After obtaining clean reads, Bowtie2 v2.2.5 (http://bowtie-bio.sourceforge.net/Bowtie2/index.shtml) was used to compare clean reads to *P. trichocarpa* reference genome (Tuskan et al., 2006) (ftp://plantgenie.org/Data/PopGenIE). Then RSEM was used to calculate the expression levels of genes and transcripts (Li and Dewey, 2011). Seven major functional database annotations – KEGG Ortholog, KO; Gene Ontology, GO; NCBI
104 nonredundant protein sequences, NR; NCBI nucleotide sequences, NT; A manually annotated and reviewed
105 protein sequence database, SwissProt; Protein family, PFAM; and Clusters of Orthologous Groups of proteins,
106 KOG – were performed on the assembled UniGenes. BLAST was used for the functional annotation of NT,
107 Diamond was used for NR, KOG, SwissProt and KEGG, Hmmscan for PFAM and Blast2GO for GO, with the
e-value of $1e^{-10}$ (Feng et al., 2019).

2.4 Detection of differentially expressed genes (DEGs)

110 Using the transcriptome data of D1, D2, D3, T1, T2 and T3, DESeq2 software (Anders and Huber, 2010) was
111 employed to screen DEGs with parameters of Fold Change $\geq 2$ and adjusted P-value $\leq 0.001$. The genes with a
112 fold of more than twice in triploid *P. tomentosa* compared to diploid *P. tomentosa* and Q-value $\leq 0.001$ were
113 screened as differentially expressed genes. QUANT (https://www.quantsoftware.com/) was used to normalise
114 and reduce the variance and help detect DEGs. The transcripts were filtered to show the minimum count before
115 assessing the differential expression. TMM values were calculated by using the edgeR software (Maza, 2016).

2.5 Analysis of gene ontology (GO) function and KEGG function of DEGs

116 GO annotations were performed using nucleotide data. According to the GO annotation and KEGG annotation
117 results, as well as the functional annotation, the differential genes were functionally classified. The detection of
118 transcription factor families was based on these annotations. Meanwhile, the phyper function (Evans et al., 2000)
119 in R software was used for enrichment analysis, the p-values were calculated and then corrected by False
120 Discovery Rate (FDR) (Burger, 2018). Finally, the function with Q-value $\leq 0.05$ was deemed significant
121 enrichment. The test conducted was one tailed for enrichment.

122 GetORF was used to detect the ORF of UniGene, while HMMER 3.0 hmmsearch was used to compare the
123 ORF. The characteristics of the transcription factor family were then analysed by PlantTFDB (Sanseverino et
124 al., 2010). The gene alignments were annotated into the plant-disease resistance gene database (PRGdb) by using
125 DIAMOND (https://github.com/bbuchfink/diamond) v0.8.31 software (Buchfink et al., 2015).

127 The candidate coding region in UniGenes was identified by TransDecoder v2.0.1 software, and then the
128 PFAM protein homologous sequence was examined by BLAST alignment SwissProt database and Hmmscan
3.0 (http://hmmer.org), thus predicting the coding region CDSs. MISA (http://pgrc.ipk-gatersleben.de/misa) was used to detect simple sequence repeats (SSR) in UniGene with the parameters set as default (Sen et al., 2018).

2.6 Tissue culture of diploid and triploid *P. tomentosa*

The leaves of diploid and triploid *P. tomentosa* were inoculated on a callus induction medium (MS+1.2 mg/l 6-BA+0.6 mg/l NAA) for 15 days. After which, they were transferred to an aseptic differentiation medium (MS+1.0 mg/l 6-BA+0.4 mg/l NAA) for 28 days. Once the adventitious buds had grown by 2 to 3 cm, they were moved to a rooting medium (1/2 MS + 0.4mg/L IBA). After the adventitious roots had grown by 2 to 3 cm, the seedlings were then moved to a greenhouse.

2.7 Real-time quantitative polymerase chain reaction (RT-qPCR) validation

To verify the reliability of transcriptome sequencing results and the expression of key genes, diploid and triploid Chinese white poplar tissue culture seedlings of the same age and growth conditions, which were derived from the same trees used for transcriptome, tissue culture plants and planted in the greenhouse, at 1 month, 4 months, 7 months, 10 months and 13 months old plants, were used for RT-qPCR validation. As in transcriptome sequencing, stem segments under the fifth leaf down from the tip of a branch were selected for validation. The method of tissue culture was carried out according to the method reported by Hu et al. (2005). The stem cuttings were sampled at 9 o'clock in the morning.

Ten growth-related genes were selected from the pathway of plant hormone signal transduction (Ko04075), carbon fixation in photosynthetic organisms (Ko00710), nitrogen metabolism (Ko00910) and tryptophan metabolism (Ko00380) for RT-qPCR validation. These genes were up-regulated in triploid *P. tomentosa* compared to diploid *P. tomentosa*, and taken as candidate markers for plant breeders in the future. Using the *EF1a* as an internal reference gene, and using the gene sequences of the transcriptome, the primers were designed using Primer primer 5 (Lalitha 2000), as listed in Table 1. The RT-qPCR processes were described according to Li et al. (2019b) on a Bio-Rad CFX96TM Real-time PCR Detection System (Bio-Rad, California, USA), with a final volume of 25.0 μl, containing 2.5 units of Taq DNA polymerase, 0.4 mM deoxyribonucleotides (dNTPs), 20 μl of ddH₂O, 1 μl of cDNA, and 500 nM of each primer. Each sample was carried out in triplicate for the RT-qPCR reactions. The data were normalised with the housekeeping gene *EF1a*, and the method was employed in
3 Results and Analysis

3.1 Unigene function annotations

The length distribution of unigenes is shown in Figure 1. The unigenes were annotated with seven major functional database annotations (KEGG, GO, NR, NT, SwissProt, Pfam, and KOG). Finally there were 6,554 (NR: 86.43%), 71,647 (NT: 93.04%), 50,143 (SwissProt: 65.12%), 51,971 (KOG: 67.49%), 52,404 (KEGG: 68.05%), 50,210 (GO: 65.20%), and 50,881 (Pfam: 66.07%) unigenes with access to functional annotations. A total of 55,479 Coding DNA Sequences (CDSs) were detected. At the same time, 21,163 SSRs were detected distributed in 16,304 unigenes, and 3,443 unigene encoding transcription factors were predicted.

It could be seen that the annotations obtained in NT (NCBI nucleic acid sequence database) were the most, 93.04%; the overall annotation rate was 93.81%; and the lowest was 65.12% in the Swissprot database. A total of 86.43% of the annotations were obtained in the NR (NCBI protein database). According to the specific species distribution chart of the annotations, a total of five species were matched. The highest of the first three were P. trichocarpa, with 53.95% of the annotations, followed by P. euphratica, with 35.68% of the annotations, and finally, P. tomentosa, with 3.26% annotation.

3.2 Transcription factor (TF) prediction

The TF prediction results showed that the genes belonged to a total of 55 transcription factor families, of which, with the largest number of genes, was the MYB gene family, with a total of 438 genes involved in the expression, followed by the AP2-EREBP gene family with 290 genes, and finally the bHLH gene family with 239 genes (see Figure 2). Among these transcription factor families, both the MYB family and the WRKY family (including 208 genes) are involved in plant growth and development processes, which can provide relevant information for our subsequent screening of growth-related genes.

3.3 Cluster analysis and GO classification of DEGs

The number of up-regulated genes in diploid P. tomentosa compared with triploid P. tomentosa was 15,690 and
the down-regulated gene was 16,971. The scatter plot of DEGs showed that the difference of transcriptional profiles between triploid and diploid samples was obvious (see Figure 3). The GO function was divided into three branches: molecular function, cellular component, and biological process. Figure 4 shows the functional classification based on differential gene detection.

A total of 22,375 differentially expressed genes that had GO annotations were obtained in GO classification entries by using the classification of 32,661 common differential genes. There were 13,720 DEGs in biological processes, including 6,481 up-regulated genes and 6,879 down-regulated genes; 15,963 DEGs in cell composition, including 7,687 up-regulated genes and 8,276 down-regulated genes; and 1,092 DEGs in molecular function, including 8,553 up-regulated genes and 9,539 down-regulated genes.

Among the three branches of GO function entries, the number of DEGs of binding was the largest in molecular function, the number of DEGs of the cell and cell part was the largest in the cellular component, and the number of DEGs of the cellular process was the largest in the biological process. There are 44 DEGs that belonged to lignin production-related transcripts (see Table 2), which are important for tree breeders.

A total of 38 entries were enriched in GO function (Q ≤ 0.05) (see Table 3). Biological processes accounted for 36.8% of the total, of which catalytic activity accounted for the largest proportion of biological processes, at 42.86%; cell components accounted for 18.5% of the total, of which cell accounted for 42.85%; and molecular functions accounted for 44.7% of the total, of which catalytic activity accounted for 47.05%.

3.4 Analysis of pathway function of DEGs

Using Q-value ≤ 0.05 as the standard, 32,661 differential genes were separately analysed by pathway enrichment using the KEGG database. There were 16 significantly enriched KEGG metabolic pathways. Among them, the most frequently occupied pathways were the metabolism branch, with 13, and a total of 19,469 DEGs involved; the second was the biological system branch, with two, and a total of 1,447 DEGs involved; and finally, the environmental information process branch, with one, with 868 DEGs involved.

The results showed that there were 12,240 significantly different genes annotated in the KEGG pathways of diploid and triploid samples. The top differentially expressed genes significantly enriched six pathways were
as follows: 1,115 differential gene expression in the plant-pathogen interaction (KO: ko04626) pathway; 396
differential genes in the phenylpropane biosynthesis (KO: ko00940) pathway; 868 differential genes in the
mitogen-activated protein kinase (MAPK) plant signalling pathway (KO: ko04016); 170 differential genes
expression in the ascorbate and aldarate metabolism pathway (KO: ko00053); 64 differential genes in the
diterpenoid biosynthesis pathway (KO: ko00904); and 43 differential genes in the beets red pigment biosynthesis
pathway (KO: ko00965). All differentially expressed genes significantly enriched pathways were shown in Table
4. The results showed that there were 496 DEGs upregulated and 619 downregulated in the plant-pathogen
interaction pathway, and that there were 864 DEGs enriched in the organismal systems category of GO
annotations, 217 DEGs in environmental information processing, 25 DEGs in genetic information processing
and nine DEGs in metabolism.

The differential genes in the plant-pathogen interaction (KO: ko04626) pathway could be enriched in 46
GO function entries. These GO function entries are integral components to fungus and ribosome biogenesis in
the following components: the membrane, signal transduction, defence response, DNA-templated transcription,
calcium ion binding, ADP binding, regulation of membrane potential, cell surface receptor signalling pathway,
absicic acid-activated signalling pathway, protein phosphorylation, peptidyl-serine phosphorylation, lipid
metabolic process, intracellular signal transduction, proteolysis, protein autophosphorylation, ATP binding,
plasma membrane, mRNA transcription, protein folding, 2-alkenal reductase [NAD(P)] activity, activation of
protein kinase activity, kinase activity, regulation of mitotic cell cycle, calcium ion homeostasis, signal
transduction by protein phosphorylation, intracellular, response to stress, membrane, extracellular region, metal
ion transport, primary miRNA processing, stress-activated protein kinase signalling cascade, cellular transition
metal ion homeostasis, response to temperature stimulus, calmodulin-dependent protein kinase activity,
mitochondrion, endosome, protein kinase activity, developmental process involved in reproduction, DNA-
templated regulation of transcription, plant-type hypersensitive response, defence response to bacterium,
 incompatible interaction, lipid catabolic process, pre-miRNA processing and defence response. Among them,
the top five entries were 184 differential genes in the integral component of membrane, 174 in the signal
transduction, 145 in defence response, 128 in DNA-templated transcription and 53 in calcium ion binding. In
In addition, there were 104 differential genes in the plant-pathogen interaction pathway that had zero expression in the triploid poplar.

In the process of analysing the growth-related pathways, it was found that the growth-related genes were up-regulated and down-regulated, so it was difficult to explain the difference between diploid and triploid samples. The NCBI database was used for gene information annotation; the details are shown in Table 5.

### 3.5 RT-qPCR validation

Genes related to growth which up-regulated in the triploid *P. tomentosa* compared to diploid *P. tomentosa*, as candidate markers for plant breeders in the future, were selected for RT-qPCR validation. There was no peak in the dissolution curves of RT-qPCR products of *GH3*, *CA*, and *A-ARR* gene. And the other seven genes showed a single peak curve, indicating that their amplification products did not contain primer dimers or nonspecific amplification products, and each primer PCR reaction was specific. The results of comparison of the transcriptome analysis and the RT-qPCR analysis are shown in Figure 5A. In diploid plants, the expression levels of the *SAUR*, *FDP*, *ALDH*, *AUX1* and *ABF* in the RT-qPCR analysis were significantly higher than those in the transcriptome analysis. In triploid plants, the expression levels of the *ALDH* and *AUX1* in the RT-qPCR analysis were significantly lower than those in the transcriptome analysis, whereas the expression levels of the malate dehydrogenase (*MDH*) and *CYCD3* in the RT-qPCR analysis were significantly higher than those in the transcriptome analysis. The expression levels of *AUX1*, *CYCD3*, and *MDH* in tissue-cultured triploid poplar were higher than those of tissue-cultured diploid samples, which were consistent with the changes of TMM values calculated by transcriptome. Among them, the expression levels of the *AUX1* gene in triploid samples were significantly higher than those in diploid samples (about 6.88 times, logFC value). However, the expression of the *ABF* gene in diploid samples was higher than that of triploid samples, which was not in accordance with the TMM values. In further experiments, the expression levels of *MDH* and *CYCD3* in tissue-cultured and greenhouse planted triploid poplar were significantly higher than those of tissue-cultured and greenhouse planted diploid samples. And the expression levels of *MDH* and *CYCD3* increased with the age gradually (Figure 5, B and C).
4 Discussion

The transcriptome analysis of the new rooting stem segments of diploid and triploid *P. tomentosa* showed that most of the significantly different genes were concentrated in plant-pathogen interaction, phenylpropane biosynthesis pathway, and MAPK signalling pathway-plant. Under the condition of comprehensive screening of the GO function enrichment and KEGG function analysis of transcriptome data, it is difficult to determine the expression of specific genes when the genes associated with plant growth appear up-regulated and down-regulated. Huang et al (1990) shown there to be larger branches, leaves and fruits in the triploid variety of pear when compared to the diploid variety. And the production of tetraploid radish is 20% higher than that of ordinary diploids (Liu et al., 2003). Zhu et al. (1995) reported that allotriploidy of *P. tomentosa* had greater values than diploidy under the same growth conditions in tree height, diameter at breast height, and single plant volume at the age of eight years.

It is interesting that although the transcriptome of the diploidy and triploidy of *P. tomentosa* was analysed here, according to the annotations of the specific species distribution chart, only 3.26% was annotated by *P. tomentosa*. This may partly be due to the fact that the triploid materials used in this study were obtained by the hybridised *P. bolleana*, and some of the mRNAs did not belong in *P. tomentosa*. Of course, it may also partly be due to the fact that the plant stems were taken from annual plants in this study, which is different to previous studies on *P. tomentosa*. Previous studies on the transcriptome analysis of *P. tomentosa* were mostly based on aseptic seedlings, root one-month-old stem and leaf sample materials (An et al., 2011; Wang et al., 2018), and the expression of mRNA was organ-specific (Ohtsuki et al., 2005).

The phytohormone signal transduction pathway controls plant cell division, cell elongation, cell enlargement, and stem elongation, which is closely related to plant growth and development (Guo et al., 2020). Compared with previous data of transcriptome analysis on phytohormone signal transduction, we selected *AUX1*, *GH3*, *A-ARR*, *CYCD3*, *ABF* and five other genes for the RT-qPCR validation. Huge differences were found between the expression levels of the selected genes in the RT-qPCR analysis and those of the transcriptome analysis, and we inferred that it might be due to the different growth stages of the experiment’s materials.

Photosynthesis is an important metabolic process in plants, and its strength has an important effect on plant
growth, development, and stress resistance. Li and Zhang (2006) found that the lower leaves of the fast-growing triploid *P. tomentosa* clones could maintain a higher photosynthetic rate when measuring the leaf net photosynthetic rate of diploid and triploid *P. tomentosa*. Photosynthesis includes a series of complex reactions in which carbon fixation is a central link in the regulation of photosynthesis (Feng et al., 2006). We compared the data of carbon fixation pathways in previous photosynthetic organism kiwifruit (Li et al., 2019b) and selected the up-regulated malate dehydrogenase (*MDH*) for the RT-qPCR validation, and the expression levels of *MDH* gene were up-regulated in triploid *P. tomentosa* plants, compared to the diploid ones.

With the possible exception of the stomatal dimension, the response to polyploidy can be very variable and complex. It has been proved in oilseed rape (Bancroft et al., 2011), sugarcane (Manners and Casu, 2011), cotton (Rambani et al., 2014), wheat (Leach et al., 2014), kiwifruit (Li et al., 2019b), rice (Shenton et al., 2020) and numerous studies in Arabidopsis species, some of which specifically consider triploidy (Fort et al., 2016; 2017; Pacey et al., 2019). Related to the phenomenon, the gene expression levels were not upregulated in any polyploid plant, and over-, under- or mixed-expression of genes were found in the polyploid plant (Osborn et al., 2003; Gutierrez-Gonzalez and Garvin, 2017; Li et al., 2019b). Some genes in polyploid plants were upregulated while others were downregulated compared to diploid plants, which was also found in this study. It illustrated that the allopolyplody – in particular in *P. tomentosa* – also alter the gene expression profile and levels as well as those in autopolyploid plants, compared to their diploid relatives. In this study, the expression levels that related to the growth genes such as *MDH* and *CYCD3* in triploid *P. tomentosa* were higher than those of diploid *P. tomentosa*.

*MDH* is mainly involved in the metabolism of plant photosynthesis (Sawada et al., 2002). The main function of the protein encoded by the *MDH* gene is to control the carbon dioxide levels during photosynthesis (Sawada et al., 2002). Higher expression levels of the *MDH* gene was proved to be related to higher rates of photosynthesis (Kandoi et al., 2018), which in turn contributes to higher timber yield. CYCLIN D3 (*CYCD3*) is a cell-cycle gene, and overexpression of CYCLIN D3;1 (*CYCD3;1*) in transgenic plants can increase mitotic cycles and reduce endocycles (Menges et al., 2006). *CYCD3* was found to regulate cambial cell proliferation and secondary growth, and the protein encoded by the *CYCD3* gene is required for normal vascular development in Arabidopsis (Collins et al., 2015). Previous studies have indicated a relationship between faster growth and the increased expression of the *CYCD3* gene. Hence, the upregulation of *MDH* and *CYCD3* in the triploids is biologically...
There were many DEGs between the diploid and triploid poplar plants that were enriched for a plant-pathogen interaction pathway, stress resistance and several growth-related transcripts too. Genes in the plant-pathogen interaction pathway are known to be diverse and large in plants and are known to be involved in reproduction isolation across Arabidopsis and Poplars too (Liao et al., 2014; Qian et al., 2018). In this study, one differential gene of the developmental process involved in reproduction was also found in the plant-pathogen interaction pathway. Further, the enrichment of differential expression transcripts in this category could provide useful information to tree breeders who intend to generate heterotic F1s. The pathway enrichment approach also shows several other interesting candidates that would be worth elaborating on in the future relating to studies demonstrating higher stress resistance in the triploid poplar.

5 Conclusion

A total of 32,661 DEGs were identified in triploid and diploid Chinese white poplar, of which 15,690 were up-regulated and 16,971 were down-regulated in triploidy compared to diploidy. Through the comprehensive analysis of GO functional enrichment analysis and the pathway functional annotation of transcriptome data of diploid and triploid *P. tomentosa*, no significantly enriched entries and pathways related to growth were found. Compared to diploidy, the growth-related genes were found to be up-regulated and down-regulated in the natural diploid and triploid *P. tomentosa* trees. Although the expression levels of genes were unstable in the different environments and different growth stages, the expression levels of *MDH* and *CYCD3* in triploid *P. tomentosa* were higher than those of diploid *P. tomentosa* in young tree tissue, which was consistent with the values calculated using the transcriptome data.

Availability of data and material

All data generated or analyzed during this study are included in this published article. RNA-Seq data were presented at the Genome Sequence Archive of the Beijing Institute of Genomics (BIG) Data Center (accession number PRJCA002269, https://bigd.big.ac.cn/search?dbId=biosample&q=PRJCA002269).
Acknowledgments

The manuscript was proofread by Proofed Inc.(UK).

References

An XM, Wang DM, Wang ZL, Li B, Bo WH, Cao GL, Zhang ZY. Isolation of a LEAFY homolog from *Populus tomentosa*: expression of PtLFY in *P. tomentosa* floral buds and PtLFY-IR-mediated gene silencing in tobacco (*Nicotiana tabacum*). Plant Cell Reports, 2011, 30(1):89-100. doi: 10.1007/s00299-010-0947-0.

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology, 2010, 11: R106.

Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X, Liu S, Trick M. Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nature Biotechnology, 2011, 29(8):762-6. doi: 10.1038/nbt.1926.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014, 30(15):2114-20. doi: 10.1093/bioinformatics/btu170.

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature Methods, 2015, 12: 59-60.

Burger T. Gentle introduction to the statistical foundations of false discovery rate in quantitative proteomics. Journal of Proteome Research, 2018, 17(1):12-22. doi: 10.1021/acs.jproteome.7b00170.

Chen TY, Wang B, Wu YY, Wen JL, Liu CF, Yuan TQ, Sun RC. Structural variations of lignin macromolecule from different growth years of triploid of *Populus tomentosa* Carr. International Journal of Biological Macromolecules, 2017, 101:747-757. doi: 10.1016/j.ijbiomac.2017.03.146.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and
preprocessing of high-throughput sequencing data. Gigascience, 2018, 7(1):1-6. doi: 10.1093/gigascience/gix120.

Ci D, Tian M, Song Y, Du Q, Quan M, Xuan A, Yu J, Yuan Z, Zhang D. Indole-3-acetic acid has long-term effects on long non-coding RNA gene methylation and growth in *Populus tomentosa*. Molecular Genetics and Genomics, 2019, 294(6):1511-1525. doi: 10.1007/s00438-019-01593-5.

Collins C, Maruthi NM, Jahn CE. *CYCD3* D-type cyclins regulate cambial cell proliferation and secondary growth in Arabidopsis. Journal of Experimental Botany, 2015, 66(15):4595-606. doi: 10.1093/jxb/erv218.

Evans M, Hastings N, Peacock B. *Statistical Distributions*, New York: Wiley-Interscience, Third edition, 2000.

Fan B, Li X, Zhang J, Chen W, Dong H. Decomposition of triploid *populus tomentosa* fine root and *Lolium multiflorum* grass root in a composite ecosystem and their nutrient dynamics. Ying Yong Sheng Tai Xue Bao, 2005, 16(11):2030-4. in Chinese.

Feng X, Jia Y, Zhu R, Chen K, Chen Y. Characterization and analysis of the transcriptome in *Gymnocypris selincuoensis* on the Qinghai-Tibetan Plateau using single-molecule long-read sequencing and RNA-seq, *DNA Research*, 2019, 26(4): 353–363. https://doi.org/10.1093/dnares/dsz014.

Feng Y, Chen X, Shi D, Wang Q. Expression of rice cytoplasmic *FBA* gene in Anabaena 7120 and its regulation on photosynthesis. Plant Research, 2006 (06): 691-698. in Chinese.

Fort A, Ryder P, McKeown PC, Wijnen C, Aarts MG, Sulpice R, Spillane C. Disaggregating polyploidy, parental genome dosage and hybridity contributions to heterosis in *Arabidopsis thaliana*. New Phytologist, 2016, 209(2):590-9. doi: 10.1111/nph.13650. Epub 2015 Sep 23.

Fort A, Tuteja R, Braud M, McKeown PC, Spillane C. Parental-genome dosage effects on the transcriptome of F1 hybrid triploid embryos of *Arabidopsis thaliana*. Plant Journal, 2017, 92(6):1044-1058. doi: 10.1111/tpj.13740. Epub 2017 Nov 28.
Guo J, Lu C, Zhao F, Gao S, Wang B. Improved reproductive growth of euhalophyte *Suaeda salsa* under salinity is correlated with altered phytohormone biosynthesis and signal transduction. Functional Plant Biology, 2020. doi: 10.1071/FP19215.

Gutierrez-Gonzalez JJ, Garvin DF. De Novo Transcriptome Assembly in Polyploid Species. Methods in Molecular Biology, 2017, 1536: 209-221. doi: 10.1007/978-1-4939-6682-0_15.

Hao S, Zhao T, Xia X, Yin W. Genome-wide comparison of two poplar genotypes with different growth rates. Plant Molecular Biology, 2011;76(6):575-591. doi:10.1007/s11103-011-9790-0.

Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, Scarascia Mugnozza G. Accelerating the domestication of forest trees in a changing world. Trends in Plant Science, 2012, 17(2): 64-72. doi: 10.1016/j.tplants.2011.11.005.

Hu L, Lu H, Liu Q, Chen X, Jiang X. Overexpression of *mtlD* gene in transgenic *Populus tomentosa* improves salt tolerance through accumulation of mannitol. Tree Physiology, 2005, 25(10):1273-81.

Huang L, Li S, Cong P. Comparison of polyploid and diploid characters of pear. Chinese Fruit trees, 1990 (03): 30-31+20. in Chinese

Jaikishan I, Rajendrakumar P, Hariprasanna K, Balakrishna D, Bhat BV, Tonapi VA. Identification of differentially expressed transcripts at critical developmental stages in sorghum [*Sorghum bicolor* (L.) Moench] in relation to grain yield heterosis. *J Biotech*. 2019;9(6):239. doi:10.1007/s13205-019-1777-6.

Kandoi D, Mohanty S, Tripathy BC. Overexpression of plastidic maize NADP-malate dehydrogenase (ZmNADP-MDH) in *Arabidopsis thaliana* confers tolerance to salt stress. Protoplasma, 2018, 255(2):547-563. doi: 10.1007/s00709-017-1168-y.

Lalitha S. Primer premier 5. Biotech Software & Internet Report, 2000, 1: 270–272.

Leach LJ, Belfield EJ, Jiang C, Brown C, Mithani A, Harberd NP. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics. 2014, 15:276. doi: 10.1186/1471-2164-15-276.
Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 2011, 12:323. doi: 10.1186/1471-2105-12-323.

Li J, Gao K, Lei B, Zhou J, Guo T, An X. Altered sucrose metabolism and plant growth in transgenic Populus tomentosa with altered sucrose synthase PtSS3. Transgenic Research, 2019a, doi:10.1007/s11248-019-00184-9.

Li J, Zhang Z. Study on photosynthetic characteristics of triploid Populus tomentosa clones. Journal of Beijing Forestry University, 2000 (06): 12-15. in Chinese

Li S, Liu X, Liu H, Zhang X, Ye Q, Zhang H. Induction, identification and genetics analysis of tetraploid Actinidia chinensis. Royal Society Open Science, 2019b, 6: 191052.

Liao W, Ji L, Wang J, Chen Z, Ye M, Ma H, An X. Identification of glutathione S-transferase genes responding to pathogen infestation in Populus tomentosa. Functional and Integrative Genomics. 2014,14(3):517-529. doi:10.1007/s10142-014-0379-y.

Lin YJ, Chen H, Li Q, Li W, Wang JP, Shi R, Tunlaya-Anukit S, Shuai P, Wang Z, Ma H, Li H, Sun YH, Sederoff RR, Chiang VL. Reciprocal cross-regulation of VND and SND multigene TF families for wood formation in Populus trichocarpa. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(45):E9722-E9729. doi: 10.1073/pnas.1714422114.

Liu B, Sun G. Transcriptome and miRNAs analyses enhance our understanding of the evolutionary advantages of polyploidy. Critical Reviews in Biotechnology, 2019, 39(2):173-180. doi: 10.1080/07388551.2018.1524824.

Liu W, Wang M, Yan Z. Advances in polyploid breeding of vegetable crops. Changjiang vegetables, 2003 (01): 29-33. in Chinese

Liu XZ, Liu XP, Zhang ZM, Sang M, Sun XD, He CZ, Xin PY & Zhang HY. Functional analysis of the FZF1 genes of Saccharomyces uvarum. Frontiers in Microbiology, 2018, 9: 96.
Manners, JM, Casu, RE (2011) Transcriptome analysis and functional genomics of sugarcane. Trop Plant Biology, 4: 9-21.

Maza E. In Papyro comparison of TMM (edgeR), RLE (DESeq2), and MRN normalization methods for a simple two-conditions-without-replicates RNA-Seq experimental design. Frontiers in Genetics, 2016, 7: 164. doi: 10.3389/fgene.2016.00164.

Menges M, Samland AK, Planchais S, Murray JA. The D-type cyclin CYCD3;1 is limiting for the G1-to-S-phase transition in Arabidopsis. Plant Cell. 2006, 18(4):893-906. doi:10.1105/tpc.105.039636.

Ohtsuki T, Otsuki M, Murakami Y, Maekawa T, Yamamoto T, Akasaka K, Takeuchi S, Takahashi S. Organ-specific and age-dependent expression of insulin-like growth factor-I (IGF-I) mRNA variants: IGF-IA and IB mRNAs in the mouse. Zoological Science, 2005, 22(9):1011-21. doi: 10.2108/zsj.22.1011.

Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA. Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics, 2003, 19:141–147. doi:10.1016/S0168-9525(03)00015-5.

Pacey EK, Maherali H, Husband BC. The influence of experimentally induced polyploidy on the relationships between endopolyploidy and plant function in Arabidopsis thaliana. Ecology Evolution, 2019, 10(1):198-216. doi: 10.1002/ece3.5886.

Qian C, Yan X, Yin H, Fan X, Yin X, Sun P, Li Z, Nevo E, Ma XF. Transcriptomes divergence of Ricotia lunaria between the two micro-climatic divergent slopes at "Evolution Canyon" I, Israel. Frontiers in Genetics, 2018, 9: 506. doi:10.3389/fgene.2018.00506.

Rambani A, Page JT, Udall JA. Polyploidy and the petal transcriptome of Gossypium. BMC Plant Biology, 2014, 14:3. doi: 10.1186/1471-2229-14-3.

Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano R M. PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Research, 2010, 38(Database issue):D814-D821. doi:10.1093/nar/gkp978.
Sattler MC, Carvalho CR, Clarindo WR. The polyploidy and its key role in plant breeding. Planta. 2016, 243(2):281-96. doi: 10.1007/s00425-015-2450-x.

Sawada S, Sakamoto T, Sato M, Kasai M, Usuda H. Photosynthesis with single-rooted Amaranthus leaves. II. Regulation of ribulose-1,5-bisphosphate carboxylase, phosphoenolpyruvate carboxylase, NAD-malic enzyme and NAD-malate dehydrogenase and coordination between PCR and C4 photosynthetic metabolism in response to changes in the source-sink balance. Plant and Cell Physiology, 2002, 43(11):1293-301. doi: 10.1093/pcp/pcf153.

Sen S, Dehury B, Sahu J, Rathi S, Yadav RNS. Mining and comparative survey of EST-SSR markers among members of Euphorbiaceae family. Molecular Biology Reports, 2018, 45(4):453-468. doi: 10.1007/s11033-018-4181-0.

Shenton M, Kobayashi M, Terashima S, Ohyanagi H, Copetti D, Hernández-Hernández T, Zhang J, Ohmido N, Fujita M, Toyoda A, Ikawa H, Fujiyama A, Furuumi H, Miyabayashi T, Kubo T, Kudrna D, Wing R, Yano K, Nonomura KI, Sato Y, Kurata N. Evolution and diversity of the wild rice Oryza officinalis complex, across continents genome types, and ploidy levels. Genome Biology Evolution, 2020 Mar 3. pii: evaa037. doi: 10.1093/gbe/evaa037.

Soltis DE, Soltis PS, Tate JA. Advances in the study of polyploidy since Plant speciation. New Phytologist, 2004, 161:173–191. doi:10.1046/j.1469-8137.2003.00948.x.

Sun P, Jia H, Zhang Y, Li J, Lu M, Hu J. Deciphering genetic architecture of adventitious root and related shoot traits in Populus using QTL mapping and RNA-Seq data. International Journal of Molecular Sciences, 2019, 20(24):6114. doi: 10.3390/ijms20246114.

Tuskan GA, 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 RR, Bhalerao RP, Blauze D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J,
Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science, 2006, 313:1596–1604.

Wang L, Wu R, Bo W. Transcriptome profiling of PeCRY1 transgenic Populus tomentosa. Genes and Genomics. 2018, 40(4):349-359. doi: 10.1007/s13258-017-0631-7.

Ye Q, Liu X, Bian W, Zhang Z, Zhang H. Over-expression of transcription factor ARK1 gene leads to down-regulation of lignin synthesis related genes in hybrid poplar ‘717’. Scientific Reports, 2020, 10: 8549. https://doi.org/10.1038/s41598-020-65328-y.

Zhang Q, Zhang ZY, Lin SZ, Zheng HQ, Lin YZ, An XM, Li Y, Li HX. Characterization of resistance gene analogs with a nucleotide binding site isolated from a triploid white poplar. Plant Biology (Stuttg). 2008, 10(3):310-22. doi: 10.1111/j.1438-8677.2008.00029.x.

Zhang Y, Xiao Z, Zhan C, Liu M, Xia W, Wang N. Comprehensive analysis of dynamic gene expression and investigation of the roles of hydrogen peroxide during adventitious rooting in poplar. BMC Plant Biology, 2019, 19(1):99. doi: 10.1186/s12870-019-1700-7.

Zhu Z. Polyploid induction, growth and photosynthetic characteristics of Populus davidiana and Populus simonii × P. nigra. Master thesis, Northeast Forestry University, Harbin, China, 2010. in Chinese

Zong D, Gan P, Zhou A, Li J, Xie Z, Duan A, He C. Comparative analysis of the complete chloroplast genomes of seven Populus species: Insights into alternative female parents of Populus tomentosa. PLoS One, 2019, 14(6):e0218455. doi: 10.1371/journal.pone.0218455.
Figure 1 Length distribution of unigenes.

Figure 2 Classification of the family of transcription factors.

Figure 3 Scatter plot of log diploid and triploid expression data. Red box: representing upregulated transcripts after FDR; gray box: representing non-DEGs; blue box: representing downregulated transcripts.

Figure 4 GO secondary node annotation statistics of differential expression genes. Note: The abscissa was the number of genes and the left side of the ordinate was the GO classification.

Figure 5 Comparison of the transcriptome analysis and the RT-qPCR analysis.
A: TDP, transcriptome data of diploid plants; RDP, RT-qPCR data of diploid plants; RTP, RT-qPCR data of triploid plants; TTP, transcriptome data of triploid plants. B,C: DP, Diploid plants; TP, triploid plants. These plants were tissue culture plants and planted in greenhouse. 1, 1 month old; 2, 4 months old; 3, 7 months old; 4, 10 months old; 5, 13 months old. The whiskers are standard deviation bars. **, the difference is very significant (P-value<0.01). RT-qPCR was performed on 3 diploid and 3 triploid plants, which were driven from the same tree used for transcriptome analysis, normalized with housekeeping gene EF1α, repeated 3 times. The $2^{(-\Delta\Delta Ct)}$ method was utilized to process the data (Liu et al., 2018). Because the difference of the values in Figure 5A was too big, the y-axis of Figure 5A was changed to a log scale.
Figure 1

Figure 1 Length distribution of unigenes.
Figure 2

Figure 2 Classification of the family of transcription factors.
Figure 3

Figure 3 Scatter plot of log diploid and triploid expression data.

Red box: representing upregulated transcripts after FDR; gray box: representing non-DEGs; blue box: representing downregulated transcripts.
Figure 4

Figure 4 GO secondary node annotation statistics of differential expression genes.

Note: The abscissa was the number of genes and the left side of the ordinate was the GO classification.
Figure 5

Figure 5 Comparison of the transcriptome analysis and the RT-qPCR analysis.

A: TDP, transcriptome data of diploid plants; RDP, RT-qPCR data of diploid plants; RTP, RT-qPCR data of triploid plants; TTP, transcriptome data of triploid plants. B,C: DP, Diploid plants; TP, triploid plants. These plants were tissue culture plants and planted in greenhouse. 1, 1 month old; 2, 4 months old; 3, 7 months old; 4, 10 months old; 5, 13 months old. **, the difference is very significant (P-value<0.01). RT-qPCR was performed on 3 diploid and 3 triploid plants, which were driven from the same tree used for transcriptome analysis, normalized with housekeeping gene EF1α, repeated 3 times. The $2^{(-\Delta\Delta C_t)}$ method was utilized to process the data (Liu et al., 2018). Because the difference of the values in Figure 5A was too big, the y-axis of Figure 5A was changed to a log scale.
Manuscript to be reviewed
Table 1 (on next page)

Table 1 The primers used in the RT-qPCR analysis.
Table 1: The primers used in the RT-qPCR analysis.

| Gene Symbol | Upstream and Downstream primer sequence (5→3) | Product length |
|-------------|-----------------------------------------------|----------------|
| EF1a        | F: GGCAAGGAGAAGGTACACAT R: CAATCACACGCTTGCTCAATA | 204            |
|             | F: TGGATCTGTCAATTCATTATGTTGCT | 143            |
| AUX1        | R: AAATACCGTATGATAAAGGAGGTAT | 211            |
| GH3         | F: GGACACCGAAAGGAAGGAAGGT | 211            |
|             | R: CCCTGAAAACATCTAATCAAAGCTAC | 100            |
| A-ARR       | F: AAATCAGGGGGAGCTCTCTTT | 100            |
|             | R: TTCTACACTCTCTCTCTGGACTGT | 143            |
| CYCD3       | F: TGCTCTGCTCTCCTCTTTGGTGAG | 216            |
|             | R: CCCTGAAAATCTCAGCAGCAATC | 153            |
| ABF         | F: TGGAAGTGCGCAAGTGGAAG | 134            |
|             | R: TCAAGACACTGGCAAAGGCA | 134            |
| MDH         | F: CCGGCTTCAATCCAATAGACTC | 153            |
|             | R: GAAGGAAAGGGTGTACGCAAT | 153            |
| CA          | F: AGAGATTATAATGAGCGCAAGCAG | 178            |
|             | R: TGGCCCCCTTATCTTTTCCTT | 178            |
| FDP         | F: ACTCCAAACACAAAACGAG | 136            |
|             | R: AGCCCACCTTGGTGATATGGAGC | 136            |
| SAUR        | F: TGGCAAGCAGTTTCCGCTTC | 105            |
|             | R: ACTGGAACCACAAAATCGCTTC | 105            |
| 1       | F: TGCTGGTGAGCTTGAGGATT | 168 |
|---------|-------------------------|-----|
|         | **ALDH**                |     |
|         | R: ATCAAAGAAATGGAGAATAGGCAGA |     |
Table 2 (on next page)

Table 2 DEGs belonged to lignin production related transcripts.

Note: a, large categories only includes biological process, cellular component and molecular function; b, subcategories under each large category.
Table 2. DEGs belonged to lignin production related transcripts.

| Gene ID          | GO Term                              | Level 1a          | Level 2b          | log2(Triploid/Diploid) |
|------------------|--------------------------------------|-------------------|-------------------|------------------------|
| CL3256.Contig2_  | GO:0009809/lignin biosynthetic process | Biological_process | Membrane          | 1.817457211            |
| All              | CL3256.Contig4_ GO:0009809/lignin biosynthetic process | Cellular_component | Membrane          | -1.85198633            |
| All              | CL3256.Contig5_ GO:0009809/lignin biosynthetic process | Molecular_function | Membrane          | -1.008154399           |
| All              | CL34.Contig3_All GO:0046274/lignin catabolic process | Molecular_function | Extracellular region | 1.395293042         |
| All              | CL34.Contig7_All GO:0046274/lignin catabolic process | Biological_process | Extracellular region | 1.050358353         |
| All              | CL34.Contig8_All GO:0046274/lignin catabolic process | Cellular_component | Extracellular region | 4.052332527         |
| All              | CL3608.Contig10_All GO:0046274/lignin catabolic process | Molecular_function | Extracellular region | -3.273514187         |
| All              | CL3608.Contig1_All GO:0046274/lignin catabolic process | Cellular_component | Extracellular region | 4.625729696         |
| All              | CL3608.Contig3_All GO:0046274/lignin catabolic process | Biological_process | Extracellular region | 1.866743079         |
| All              | CL3608.Contig6_All GO:0046274/lignin catabolic process | Biological_process | Extracellular region | 1.65429845          |
| All              | CL3608.Contig8_All GO:0046274/lignin catabolic process | Biological_process | Extracellular region | 13.4946612          |
| All              | CL3608.Contig9_All GO:0046274/lignin catabolic process | Cellular_component | Extracellular region | -3.794382392         |
CL4447.Contig1_ GO:0009809//lignin biosynthetic process Biological_process Cellular process 2.0065603
All

CL4447.Contig3_ GO:0009809//lignin biosynthetic process Biological_process Cellular process 5.546749329
All

CL4447.Contig4_ GO:0009809//lignin biosynthetic process Biological_process Cellular process 2.188480553
All

CL4447.Contig5_ GO:0009809//lignin biosynthetic process Biological_process Cellular process 1.719058134
All

CL4519.Contig1_ GO:0046274//lignin catabolic process Cellular_component Extracellular region -4.213801176
All

CL4519.Contig2_ GO:0046274//lignin catabolic process Biological_process Extracellular region 2.420173712
All

CL4519.Contig3_ GO:0046274//lignin catabolic process Cellular_component Extracellular region 2.474187548
All

CL4519.Contig4_ GO:0046274//lignin catabolic process Cellular_component Extracellular region 1.824853683
All

CL4519.Contig5_ GO:0046274//lignin catabolic process Molecular_function Extracellular region 1.92369436
All

CL4519.Contig6_ GO:0046274//lignin catabolic process Biological_process Extracellular region 1.658798156
All

CL5519.Contig1_ GO:0046274//lignin catabolic process Molecular_function Extracellular region 2.035174788
All

CL5519.Contig2_ GO:0046274//lignin catabolic process Molecular_function Extracellular region 1.871722631
All

CL7312.Contig1_ GO:0046274//lignin catabolic process Biological_process Extracellular region 1.766925603
All

CL7312.Contig4_ GO:0046274//lignin catabolic Cellular_component Extracellular region 2.018171
| Accession | Description | GO Term | Category | Location | Fold Change |
|-----------|-------------|---------|----------|----------|-------------|
| CL7312.Contig5 | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 2.303556773 |
| CL8534.Contig3 | biosynthetic process | GO:0009809/lignin | Biosynthetic | Membrane | 1.847328294 |
| CL914.Contig4_Al | metabolic process | GO:0009808/lignin | Metabolic | Membrane | -2.280202421 |
| CL914.Contig5_Al | metabolic process | GO:0009808/lignin | Metabolic | Membrane | -5.156601442 |
| CL9567.Contig1 | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | -2.96038825 |
| CL9787.Contig1 | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 1.036570867 |
| Unigene14552_All | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 3.092109481 |
| Unigene15649_All | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 9.382022244 |
| Unigene15651_All | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 2.619508994 |
| Unigene17889_All | biosynthetic process | GO:0009809/lignin | Biosynthetic | Cellular process | -5.163595454 |
| Unigene23438_All | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 2.998409872 |
| Unigene2393_All | catabolic process | GO:0046274/lignin | Catabolic | Extracellular region | 2.709284131 |
| Gene   | GO:0046274//lignin process   | Process      | Component       | Region                  |
|--------|------------------------------|--------------|-----------------|-------------------------|
| Unigene24809_All | GO:0046274//lignin process | catabolic    | Cellular_component | Extracellular region 4.060836332 |
| Unigene314_All   | GO:0046274//lignin process | catabolic    | Cellular_component | Extracellular region 2.01007043 |
| Unigene428_All   | GO:0046274//lignin process | catabolic    | Cellular_component | Extracellular region 1.572211841 |
| Unigene5164_All  | GO:0009808//lignin process | metabolic    | Biological_process | Membraneactivity -1.626207392 |
| Unigene8384_All  | GO:0046274//lignin process | catabolic    | Biological_process | Extracellular region 1.219895563 |

Note: a, large categories only includes biological process, cellular component and molecular function; b, subcategories under each large category.
Table 3 (on next page)

Table 3 Entries were enriched in GO function.

Note: a, large categories only includes biological process, cellular component and molecular function; b, subcategories under each large category.
1. Table 3: Entries were enriched in GO function.

| GO Term ID   | GO Term                                | Level 1a         | Level 2b                        | Rich Ratio | Q-value   |
|--------------|----------------------------------------|------------------|---------------------------------|------------|-----------|
| GO:0043531   | ADP binding                            | Molecular function| Binding                         | 0.649      | 5.07E-40  |
| GO:0007165   | Signal transduction                    | Biological process| Biological regulation          | 0.607      | 1.54E-32  |
| GO:0006952   | Defense response                       | Biological process| Response to stimulus           | 0.613      | 2.75E-24  |
| GO:0046914   | Transition metalion binding            | Molecular function| Binding                         | 0.675      | 1.49E-09  |
| GO:0046916   | Cellular transition metalion homeostasis| Biological process| Biological regulation         | 0.677      | 1.49E-09  |
| GO:0030001   | Metalion transport                     | Biological process| Localization                   | 0.644      | 2.10E-08  |
| GO:0005886   | Plasma membrane                        | Cellular component| Cell                           | 0.494      | 1.64E-06  |
| GO:0016021   | Integral component of membrane        | Cellular component| Membrane part                  | 0.461      | 3.73E-06  |
| GO:0007205   | Protein kinase C-activating G-protein coupled receptor signaling pathway | Biological process | Biological regulation          | 0.8        | 9.92E-05  |
| GO:0031347   | Regulation of defense response         | Biological process| Biological regulation          | 0.618      | 9.92E-05  |
| GO:0004143   | Diacylglycerol kinase activity         | Molecular function| Catalytic activity              | 0.8        | 9.99E-05  |
| GO:0006351   | Transcription, DNA-templated           | DNA-Biological process| Cellular process              | 0.491      | 0.001     |
| GO:0001228   | Transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific | Molecular function | Transcription regulator          | 0.592      | 0.002     |
| GO:0046658 | Anchored component of plasma membrane | Cellular component | Cell | 0.565 | 0.002 |
|GO:0009611 | Response to wounding | Biological process | Response to stimulus | 0.613 | 0.002 |
|GO:0048046 | Apoplast | Cellular component | Extracellular region | 0.609 | 0.002 |
|GO:2000022 | Regulation of jasmonic acid mediated signaling pathway | Biological process | Biological regulation | 0.625 | 0.002 |
|GO:1903507 | Negative regulation of nucleic acid-templated transcription | Biological process | Biological regulation | 0.629 | 0.003 |
|GO:000977 | RNA polymerase II regulatory region sequence-specific DNA binding | Molecular function | Binding | 0.570 | 0.003 |
|GO:0003700 | DNA binding transcription factor activity | Molecular function | Transcription regulator activity | 0.497 | 0.003 |
|GO:0003839 | Gamma-glutamylcyclotransferase activity | Molecular function | Catalytic activity | 0.941 | 0.004 |
|GO:0006979 | Response to oxidative stress | Biological process | Response to stimulus | 0.596 | 0.006 |
|GO:0005576 | Extracellular region | Cellular component | Extracellular region | 0.541 | 0.006 |
|GO:0016298 | Lipase activity | Molecular function | Catalytic activity | 0.698 | 0.006 |
|GO:0016020 | Membrane | Cellular component | Membrane | 0.501 | 0.007 |
|GO:0006629 | Lipid metabolic process | Biological process | Metabolic process | 0.550 | 0.010 |
|GO:0016747 | Transferase activity, transferring acyl groups other than amino-acyl groups | Molecular function | Catalytic activity | 0.578 | 0.012 |
| GO:001085 RNA polymerase II transcription factor binding | Molecular function | Binding | 0.635 0.018 |
| GO:0031408 Oxylipin biosynthetic process | Biological process | Cellular process | 0.627 0.018 |
| GO:0004190 Aspartic-type endopeptidase activity | Molecular function | Catalytic activity | 0.559 0.021 |
| GO:0020037 Heme binding | Molecular function | Binding | 0.521 0.0211 |
| GO:0003714 Transcription corepressor activity | Molecular function | Transcription regulator activity | 0.589 0.023 |
| GO:0004612 Phosphoenolpyruvate carboxykinase (ATP) activity | Molecular function | Catalytic activity | 0.786 0.027 |
| GO:0003951 NAD+ kinase activity | Molecular function | Catalytic activity | 0.638 0.0418 |
| GO:0004601 Peroxidase activity | Molecular function | Antioxidant activity | 0.607 0.0418 |
| GO:0006814 Sodium ion transport | Biological process | Localization | 0.778 0.0418 |
| GO:0005618 Cell wall | Cellular component | Cell | 0.536 0.042 |
| GO:0009041 Uridylate kinase activity | Molecular function | Catalytic activity | 0.842 0.042 |

Note: a, large categories only includes biological process, cellular component and molecular function; b, subcategories under each large category.
Table 4  (on next page)

Table 4 Differentially expressed genes significantly enriched KEGG pathways.

Note: a, large categories only includes cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems; b, subcategories under each large category.
### Table 4: Differentially expressed genes significantly enriched KEGG pathways.

| Pathway ID | Pathway Name | Level 1a | Level 2b | Rich Ratio | Q value       |
|------------|--------------|----------|----------|------------|---------------|
| ko04626    | Plant-pathogen interaction | Organismal Systems | Environmental adaptation | 0.572 | 1.62E-20 |
| ko00940    | Phenylpropanoid biosynthesis | Metabolism | Biosynthesis of other secondary metabolites | 0.561 | 1.29E-05 |
| ko04016    | MAPK signaling pathway - plant | Environmental Processing | Signal transduction | 0.522 | 5.41E-05 |
| ko00053    | Ascorbate and aldarate metabolism | Metabolism | Carbohydrate metabolism | 0.574 | 0.003 |
| ko00904    | Diterpenoid biosynthesis | Metabolism | Metabolism of terpenoids and polyketides | 0.653 | 0.003 |
| ko00965    | Betalain biosynthesis | Metabolism | Biosynthesis of other secondary metabolites | 0.694 | 0.005 |
| ko00591    | Linoleic acid metabolism | Metabolism | Lipid metabolism | 0.604 | 0.007 |
| ko04712    | Circadian rhythm-plant | Organismal Systems | Environmental adaptation | 0.532 | 0.007 |
| ko00944    | Flavone and flavonol biosynthesis | Metabolism | Biosynthesis of other secondary metabolites | 0.743 | 0.011 |
| ko00052    | Galactose metabolism | Metabolism | Carbohydrate metabolism | 0.529 | 0.027 |
| ko00561    | Glycerolipid metabolism | Metabolism | Lipid metabolism | 0.533 | 0.033 |
| KEGG ID | Description                                      | Category               | Full Path Description              | Score | P-value |
|---------|--------------------------------------------------|------------------------|-----------------------------------|-------|---------|
| ko00941 | Flavonoid biosynthesis                           | Metabolism             | Biosynthesis of other secondary metabolites | 0.557 | 0.033   |
| ko0073  | Cutin, suberine and wax biosynthesis             | Metabolism             | Lipid metabolism                  | 0.582 | 0.037   |
| ko0903  | Limonene and pinene degradation                  | Metabolism             | Metabolism of terpenoids and polyketides | 0.649 | 0.037   |
| ko0900  | Terpenoid backbone biosynthesis                  | Metabolism             | Metabolism of terpenoids and polyketides | 0.545 | 0.045   |
| ko0945  | Stilbenoid, diarylheptanoid and gingerol biosynthesis | Metabolism             | Biosynthesis of other secondary metabolites | 0.604 | 0.045   |

Note: a, large categories only includes cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems; b, subcategories under each large category.
Table 5 (on next page)

Table 5 Information of growth-related differential genes.
### Table 5 Information of growth-related differential genes.

| KEGG pathway | GIDs               | Gene annotation                                           | Gene function                                                                 |
|--------------|--------------------|-----------------------------------------------------------|-------------------------------------------------------------------------------|
| AUX1         | XM_01104 9129.1    | *Populus euphratica* auxin transporter-like protein 2 (LOC105141777), transcript variant X1, mRNA | Cell enlargement, plant growth                                                |
| GH3          | XM_01103 2075.1    | *P. euphratica* probable indole-3-acetic acid-amido synthetase GH3.1 (LOC105129841), mRNA | Indirect regulation of cell expansion and plant growth                        |
| A-ARR        | XM_01100 2137.1    | *P. euphratica* two-component response regulator *ARR8* (LOC105108010), mRNA | Cell division and bud formation                                               |
| CYCD3        | XM_01101 0474.1    | *P. euphratica* cyclin-D3-3-like (LOC105114061), mRNA | Indirect regulation of cell division and cell elongation                      |
| KO04075      |                    | *P. euphratica* ABSCISIC ACID-INSSENSITIVE 5-like protein 2 (LOC105137304), transcript variant X2, mRNA | Participation in stress resistance processes such as low temperature, high salt and oxidation stress, and involvement in plant growth and development |
| KO00710      | MDH                | *P. euphratica* oligopeptide transporter 4 (LOC105109875), mRNA | Participation in the TCA cycle, photosynthesis, C4 cycle, and other metabolic pathways to promote plant growth and development |