

## SIGNIFICANT DIFFERENCES IN GENE EXPRESSION AND KEY GENETIC COMPONENTS ASSOCIATED WITH HIGH GROWTH VIGOR IN *POPULUS* SECTION *TACAMAHACA* AS REVEALED BY COMPARATIVE TRANSCRIPTOME ANALYSIS

SHIPING CHENG\*, MINGHUI CHEN, YANYAN LI, JUNQING WANG,  
XUERONG SUN AND JIANSHEG WANG

Pingdingshan University, Pingdingshan 467000, Henan Province, People's Republic of China

\*Corresponding author's e-mail: shipingcheng@163.com

### Abstract

To identify genetic components involved in high growth vigor in F<sub>1</sub> *Populus* section *Tacamahaca* hybrid plants, high and low vigor plants showing significant differences in apical dominance during a rapid growth period were selected. Apical bud transcriptomes of high and low-growth-vigor hybrids and their parents were analyzed using high-throughput RNA sequencing on an Illumina HiSeq 2000 platform. A total of 5,542 genes were differently expressed between high growth vigor hybrid and its parents, the genes were significantly enriched in pathways related to processes such as photosynthesis, pyrimidine ribonucleotide biosynthetic processes and nucleoside metabolic processes. There were 1410 differentially expressed genes between high and low growth vigor hybrid, the genes were mainly involved in photosynthesis, chlorophyll biosynthetic process, carbon fixation in photosynthetic organisms, porphyrin and chlorophyll metabolism and nitrogen metabolism. Moreover, a k-core of a gene co-expression network analysis was performed to identify the potential functions of genes related to high growth vigor. The functions of 8 selected candidate genes were associated mainly with circadian rhythm, water transport, cellulose catabolic processes, sucrose biosynthesis, pyrimidine ribonucleotide biosynthesis, purine nucleotide biosynthesis, meristem maintenance, and carbohydrate metabolism. Our results may contribute to a better understanding of the molecular basis of high growth vigor in hybrids and its regulation.

**Key words:** Apical bud; Transcriptome; RNA-Seq; High growth vigor; Heterosis; *Populus* section *Tacamahaca*.

### Introduction

Cross-breeding of *Populus* species is considered to be an effective way of creating superior clones. The production of clones that have rapid growth, high-quality, disease and insect resistance, and other beneficial characteristics by cross-breeding generates substantial economic, ecological, and social benefits (Stettler *et al.*, 1999).

Some efforts have been made to investigate the genetic basis of heterosis to allow better use of heterosis phenomenon, but the molecular mechanism of heterosis in *Populus* hybrids remains unknown (Braatne *et al.*, 1992, Pharis *et al.*, 1991). Because all genes expressed in hybrids are derived from the two parental lines, heterosis and phenotype changes exhibited in the hybrid clones should be related to changes in gene expression. Based on this idea, efforts to identify the molecular mechanisms underlying the effects of heterosis on differential gene expression have gradually intensified (Bao *et al.*, 2005, Gao *et al.*, 2013, Thiemann *et al.*, 2010, Zhai *et al.*, 2013). High-throughput RNA sequencing (RNA-Seq), a method developed recently for discovering, profiling, and quantifying RNA transcripts, has several advantages over other expression profiling technologies, including higher sensitivity and the ability to detect splicing isoforms and somatic mutations (Cheng *et al.*, 2013, Wang *et al.*, 2009, Wang *et al.*, 2010, Zhang *et al.*, 2013, Muhammad *et al.*, 2015). Transcriptome profiling technology has been used recently to identify wood formation-related genes and to establish single nucleotide polymorphisms and expressed sequences tags resources (Geraleds *et al.*, 2011, Qiu *et al.*, 2011).

Because not all of the offspring produced by cross-breeding exhibit the superior qualities that result from hybrid

heterosis. Hybrid weakness occurs often in hybrid trees with progeny showing reduced viability and poor growth phenotypes (Yang *et al.*, 2010). Breeding processes used to produce *Populus* hybrids have long been focused on the hybrid plants having positive heterosis while ignoring studies on hybrid plants that exhibit negative heterosis (Huang *et al.*, 2012). In fact, the molecular based research of heterosis for *Populus* height growth, it is better to select parents, high and low growth vigor groups as experimental materials. Currently, only by phenotypic and physiological study of the differences in the first generation of *Populus deltoides* which are from different growth vigor clones is concluded (Huang *et al.*, 2012).

To identify specific candidate genes and metabolic pathways that are active in apical buds, we employed RNA-Seq analysis to investigate the transcriptomes of apical buds of high and low growth F<sub>1</sub> hybrids and their parents during a rapid growth period. We constructed artificial hybridization groups in which all hybrid offspring were derived from the 'Zheyin3#' and 'Beijingensis' varieties acting as maternal and paternal, respectively (Wang, 2009) and selected artificial hybrid offspring with positive and negative heterosis for subsequent growth. We collected high-coverage transcriptome data, functionally analyzed differentially expressed genes among the four types of *Populus* lines, functionally annotated the differentially expressed genes, and identified core genes closely related to heterosis. This genome-wide transcriptome comparison provides a starting point for understanding the relationship between altered gene expression and hybrid phenotypic variation and the molecular mechanism underlying apical bud heterosis in *Populus*.

## Material and Methods

**Plant materials and phenotype measurement:** Heights of F<sub>1</sub> generation plants of 120 genotypes were measured after every 15 days between July and September during 2011–2013. Nine biological replications were performed for each genotype. Based on analysis of variance, nine high growth (F<sub>1</sub>-H) and nine low growth (F<sub>1</sub>-L) genotypes with significant differences in plant height were selected. Stumping of the selected materials began in autumn.

**Biological material and RNA extraction:** After stumping, annual branches were used as plant materials. One-year-old shoots 15 cm in length with no plant disease or insect pests were cut in April, 2014. Cuttings from six plants of each genotype were planted in peat soil in plastic pots (25 × 16 × 25 cm). The cuttings were watered equally on a weekly basis. The study was performed in a greenhouse of the National Engineering Laboratory for Tree Breeding. After 1 month of growth, plant height was measured every 5 days to detect the rapid growth period, which was defined as the inflection point in growth rate between the sprouting to rapid growth phase and the rapid growth to slow growth phase. Apical buds were collected during the rapid growth period from 9:00–10:00 am on June 20, 2014.

Three plants each of the Zheyin3# and Beijingensis served as three biological replicates. The F<sub>1</sub>-H and F<sub>1</sub>-L genotypes were represented by nine hybrid plants. Each set of nine plants was divided into three pools of three plants for three replicates of RNA isolation. Individuals within each group were pooled in equal amounts to provide equal representation in each RNA pool. Total RNA was separately extracted from apical buds frozen in liquid nitrogen using the Trizol reagent as per manufacturer's instructions. (Invitrogen, Carlsbad, CA, USA). RNA quality was checked using a Bioanalyzer 2100 (Aligent) and stored at -80 °C. RNA was extracted separately from three plants of each parent.

**cDNA library preparation and sequencing:** The RNA samples were subsequently used for cDNA library construction and Illumina sequencing. The cDNA libraries were prepared using a Truseq™ RNA sample prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Poly-A-containing mRNA was isolated from the total RNA, subjected to two purification rounds using poly-T oligos attached to magnetic beads, and fragmented using an RNA fragmentation kit. First-strand cDNA was generated using reverse transcriptase and random primers. Following the second-strand cDNA synthesis and adaptor ligation, 200-bp cDNA fragments were isolated using gel electrophoresis and amplified by 15 cycles of PCR. The products were then sequenced using an Illumina HiSeq™ 2000 platform and subjected to 200 cycles of paired-end (2 × 100 bp) sequencing.

**Analysis of differentially expressed genes:** Raw reads were generated using the Illumina HiSeq™ 2000 were filtered to obtain high-quality reads by removing the adapter sequences and low-quality bases at the 3' ends.

The resulting high quality reads were mapped onto the *Populus trichocarpa* reference genome (Tuskan *et al.*, 2006). Differential expression was estimated and tested using the DEGseq software (Wang *et al.*, 2010). We quantified gene expression levels using the fragments per kilobase of exon per million mapped reads (FPKM) method (Garber *et al.*, 2011). Transcripts that exhibited an FDR (false discovery rate, FDR) ≤ 0.05 and an estimated absolute log<sub>2</sub>FC (fold-change, FC) > 0.585 or < -0.585 and values of p < 0.05 were considered to indicate significantly differential expression.

**Venn diagram analysis and functional annotation of genes related to high growth vigor:** Venn diagrams provide graphical representation of interactions among sets in an easily read format (Pirooznia *et al.*, 2007). We used Venn diagram analysis to distinguish the relationships of differentially expressed genes (DEGs) between the samples and to identify genes related to plant phenotypic changes and high growth vigor. In addition, functional-enrichment analysis using the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases was performed to identify which DEGs were significantly enriched in GO terms and metabolic pathways using Bonferroni-correction ( $P < 0.05$ ). The significance of abundantly or meagerly expressed gene sets was calculated using Fisher's exact tests (as in Grosu *et al.*, 2002).

**Identification of core genes associated with high growth vigor:** Candidate genes related to high growth vigor were selected based on normalized signal intensities of expression of specific genes. The differentially expressed genes were selected to build gene co-expression networks. We calculated the Pearson's correlation and chose the significant correlation for all pairs of genes in the network (Dorogovtsev *et al.*, 2006). A k-core of a network is a subnetwork in which all nodes are connected to at least k other genes in the subnetwork. The greater the k-core value, the stronger the degree of co-expression of the differentially expressed genes (Zhang *et al.*, 2013).

## Results

Based on plant height measurements of the parents and F<sub>1</sub> hybrids over three consecutive years, we selected nine high growth genotype plants, nine low growth genotype plants, and three plants each of the Zheyin 3# and Beijingensis. Multiple comparison methods for comparing the height of the four samples. A, B, C, D represents four samples, Respectively. (i, j = A, B, C, D; i ≠ j) represents the mean of any two species, then to calculate mean absolute deviation. Which showed that the average heights of the F<sub>1</sub>-H, F<sub>1</sub>-L, Zheyin3#, Beijingensis genotypes were 279.8, 174.0, 189.0, and 218.7 cm, respectively (Fig. 1). The mean height of the high growth vigor plants measured in the greenhouse on June 20 during the rapid growth period was significantly different from the mean heights of both the low growth vigor genotypes and the parents (Table 1).

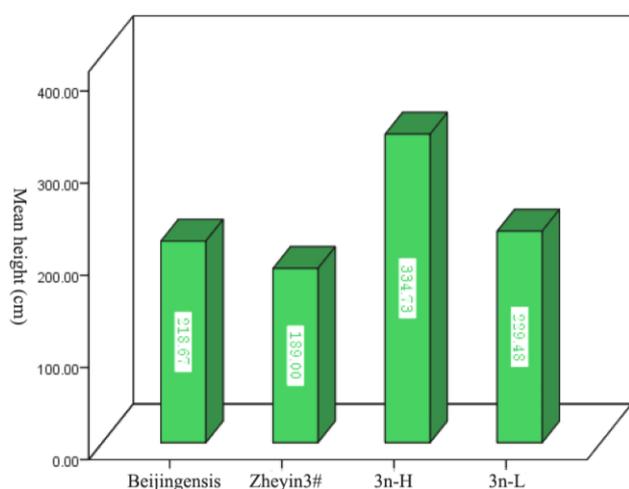


Fig. 1 The plant height in the F<sub>1</sub>-H, F<sub>1</sub>-L, 'Zheyin3#' and 'Beijingensis' genotypes.

**Sequencing, mapping, and sequence analysis:** To analyze the transcriptomes of the F<sub>1</sub>-H, F<sub>1</sub>-L, Zheyin3#, Beijingensis genotypes during the rapid growth period, poly(A)<sup>+</sup> RNA was isolated from each sample, sheered into smaller fragments, and reverse-transcribed to synthesize cDNA for analysis using an Illumina HiSeq 2000 platform. A total of 33, 433, 695; 37, 508, 442; 37, 056, 170; and 38, 696, 656 high-quality filtered reads with lengths of 100 bp were obtained for the F<sub>1</sub>-H, F<sub>1</sub>-L, Zheyin3#, Beijingensis genotypes, respectively. We used Fast-QC software to assess the quality of the sequencing data, performed *de novo* assembly of the reads, and used TopHat Software for sequence alignment with *Populus*

*trichocarpa* reference genome. The mapping rates for the F<sub>1</sub>-H, F<sub>1</sub>-L, Zheyin3#, Beijingensis genotypes were 83.6% (27, 961, 318 mapped reads), 80.1% (30, 130, 410 mapped reads), 82.0% (31, 697, 842 mapped reads), and 83.5% (30, 985, 626 mapped reads), respectively, and the unique mapping rates were 94.2%, 93.8%, 92.7%, and 92.8%, respectively (Table 2).

**Analysis of DEGs in the four genotypes:** Gene expression levels were measured as the fragments per kilobase of exon per million mapped reads (FPKM) and analyzed using DESeq software with parameters of log<sub>2</sub> FC > 0.585 or < -0.585 and p < 0.05 to identify the genes that were differentially expressed between F<sub>1</sub>-H versus Zheyin3#, F<sub>1</sub>-H versus Beijingensis, F<sub>1</sub>-H versus F<sub>1</sub>-L, and Zheyin3# versus Beijingensis.

A total of 10,445 DEGs were identified in the four comparison groups. The DEGs were classified into the following four groups: DEG<sub>PP</sub> (3493 DEGs, Zheyin3# versus Beijingensis), DEG<sub>HF</sub> (3994 DEGs, F<sub>1</sub>-H versus Beijingensis), DEG<sub>HM</sub> (1548 DEGs, F<sub>1</sub>-H versus Zheyin3#), and DEG<sub>HL</sub> (1410 DEGs, F<sub>1</sub>-H versus F<sub>1</sub>-L) (Table 3). DEGs between the hybrids and their parents in the DEG<sub>HF</sub> and DEG<sub>HM</sub> groups may be relevant to heterosis because their phenotypic differences are likely due to differences in gene expression (Song *et al.*, 2007, Yao *et al.*, 2005). Furthermore, the DEGs between high and low-growth-vigor hybrids may indicate the nature of the mechanism that confers growth advantages on the high-growth-vigor hybrids, while the DG<sub>PP</sub> group represents only the differences between the two parental lines.

**Table 1. Significance analysis of average plant height in the F<sub>1</sub>-H, F<sub>1</sub>-L, Zheyin3# (maternal) and Beijingensis (paternal) genotypes.**

Samples	Average height(cm)	$ x_i - x_F $	$ x_i - x_H $	$ x_i - x_L $
Father	218.67	29.67**	61.14**	44.71**
Mother	189.00		90.81**	15.04*
H-F <sub>1</sub>	279.81			105.85**
L-F <sub>1</sub>	173.96			

Note: Superscripts '\*' indicate a significant difference between variables at the 0.05 level H and L as high growth and low growth vigor, respectively

**Table 2. Transcriptome mapping statistics.**

Sample	Reads generated	Reads mapped	Mapping rate	Reads uniquely mapped	Unique mapping rate
F <sub>1</sub> -H	33, 433, 695	27, 961, 318	83.6%	26, 328, 101	94.2%
F <sub>1</sub> -L	37, 508, 442	30, 130, 410	80.1%	28, 255, 650	93.8%
Zheyin3#	37, 056, 170	31, 697, 842	82.0%	29, 397, 097	92.7%
Beijingensis	38, 696, 656	30, 985, 626	83.5%	28, 752, 528	92.8%

**Table 3. Number and classification of differentially expressed genes.**

Category	DG <sub>MF</sub>	DG <sub>HF</sub>	DG <sub>HM</sub>	DG <sub>HL</sub>
Number of DEGs	3493	3994	1548	1410

DG<sub>PP</sub> refers to DEGs between both parents, DG<sub>HF</sub> refers to DEGs between super hybrids and the paternal parent, DG<sub>HM</sub> refers to DEGs between super hybrids and the maternal parent, and DG<sub>HL</sub> refers to DEGs between super hybrids and inferior hybrids

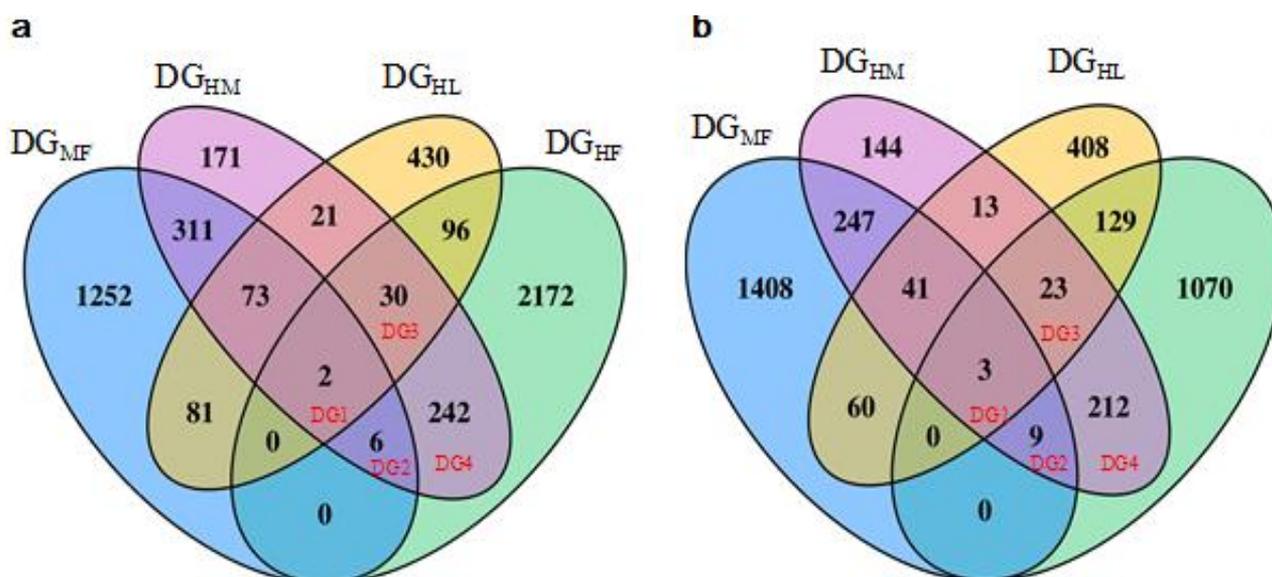


Fig. 2. Venn diagrams of differentially expressed genes. a, upregulated genes. b, downregulated genes. The DG1 and DG2 subgroups contain differentially expressed genes identified in three comparisons (DG<sub>MF</sub>, DG<sub>HM</sub>, and DG<sub>HF</sub>). The DG3 and DG4 subgroups contain DEGs unique to the DG<sub>HM</sub>, DG<sub>HF</sub>, and DG<sub>HL</sub> groups and the G<sub>HM</sub> and DG<sub>HF</sub> groups, respectively.

#### Venn diagram analysis of genes related to high growth vigor:

To identify genes related to high growth vigor in the four groups of DEGs above, a useful tool, Venn diagram should be used. Venn diagram was generated using the Venny tool from BioinfoGP to identify the transcripts uniquely expressed when different samples were compared (Fig. 2a and b). What is more, we performed Venn diagram analysis to distinguish the relationships of differentially expressed genes between the samples and to identify genes related to plant phenotypic changes and high growth vigor, which resulted in 15 DEG subgroups. Comparison of DEGs in the F<sub>1</sub>-H hybrids and both parents showed that 280 upregulated DEGs (DG1<sub>up</sub>, DG2<sub>up</sub>, DG3<sub>up</sub>, and DG4<sub>up</sub> subgroups) were in common, while 247 downregulated DEGs (DG1<sub>down</sub>, DG2<sub>down</sub>, DG3<sub>down</sub>, and DG4<sub>down</sub> subgroups) were in common. Interestingly, 8 upregulated genes (DG1<sub>up</sub> and DG2<sub>up</sub> subgroups) and 12 downregulated genes (DG1<sub>down</sub> and DG2<sub>down</sub> subgroups) were also common to both parents, suggesting that the DEGs in these two subgroups reflect differential gene expression between the two parents. The genes in the DG3<sub>up</sub>, DG4<sub>up</sub>, DG3<sub>down</sub>, and DG4<sub>down</sub> subgroups were expressed differentially between the hybrids and their parents simultaneously, while those in the DG3<sub>up</sub> and DG3<sub>down</sub> subgroups (30 upregulated genes and 23 downregulated genes, respectively) showed differential expression between the hybrids with low- and high-growth-vigor phenotypes. Therefore, the differentially expressed genes in the DG3<sub>up</sub>, DG3<sub>down</sub>, DG4<sub>up</sub>, and DG4<sub>down</sub> subgroups were most likely related to the high-growth-vigor traits of the hybrids (Gao *et al.*, 2013).

A hierarchical clustering of all 12 samples based on the RPKM values to visualize the DEGs identified by comparisons of different samples. A small cluster of differentially regulated genes in the DG3 subgroup (Fig. 3a) exhibited higher expression levels in the high-growth-vigor hybrid plants (Fig. 3b).

**Functional classification of genes in the DG3 and DG4 subgroups by GO and KEGG Pathway analysis:** To classify *Populus* genes by function, GO terms were

assigned to each group of differentially expressed genes. GO terms are a dynamically structured vocabulary that can be applied to describe functions of genes, allowing their classification into three major categories: biological process, molecular function, and cellular component and associated subcategories (Zhang *et al.*, 2012). Using Fisher's exact test, we determined the level of significance (*P* value) of each GO term to determine which GO terms were significantly enriched in the DEGs. To identify the processes associated with DEGs, we placed greater emphasis on the biological process GO terms in this study.

We analyzed the biological process categories of the DEGs in the DG3<sub>up</sub> and DG3<sub>down</sub> subgroups, which were expressed differentially in the DEG<sub>HF</sub>, DEG<sub>HM</sub>, and DEG<sub>HL</sub> groups and were likely to be related to the high-growth-vigor traits of the hybrids (Fig. 4a and b). GO terms for RNA secondary structure and DNA duplex unwinding, and cuticle hydrocarbon, proanthocyanidin, alkane, pyrimidine ribonucleotide, and wax biosynthetic processes were enriched among the upregulated DEGs, while GO terms for nucleoside metabolic, monoterpene, glutamine, and cysteine biosynthetic processes were enriched among the downregulated DEGs (Table S1).

We performed GO analysis to identify additional GO terms for DEGs in the DG4<sub>up</sub> and DG4<sub>down</sub> subgroups (Fig. 2c and d) which were expressed differentially in both the DEG<sub>HF</sub> and DEG<sub>HM</sub> groups and were possibly related to the high-growth-vigor traits of hybrids. We identified GO terms for pyrimidine ribonucleotide, nucleotide, and ribosome biosynthetic processes, RNA methylation, and protein import into the nucleus and other biosynthetic processes among the upregulated DEGs. Among the downregulated DEGs, the biological process category for hydrogen peroxide processes comprised the majority of the GO terms, followed by high light intensity and endoplasmic reticulum stress GO terms (Table S2).

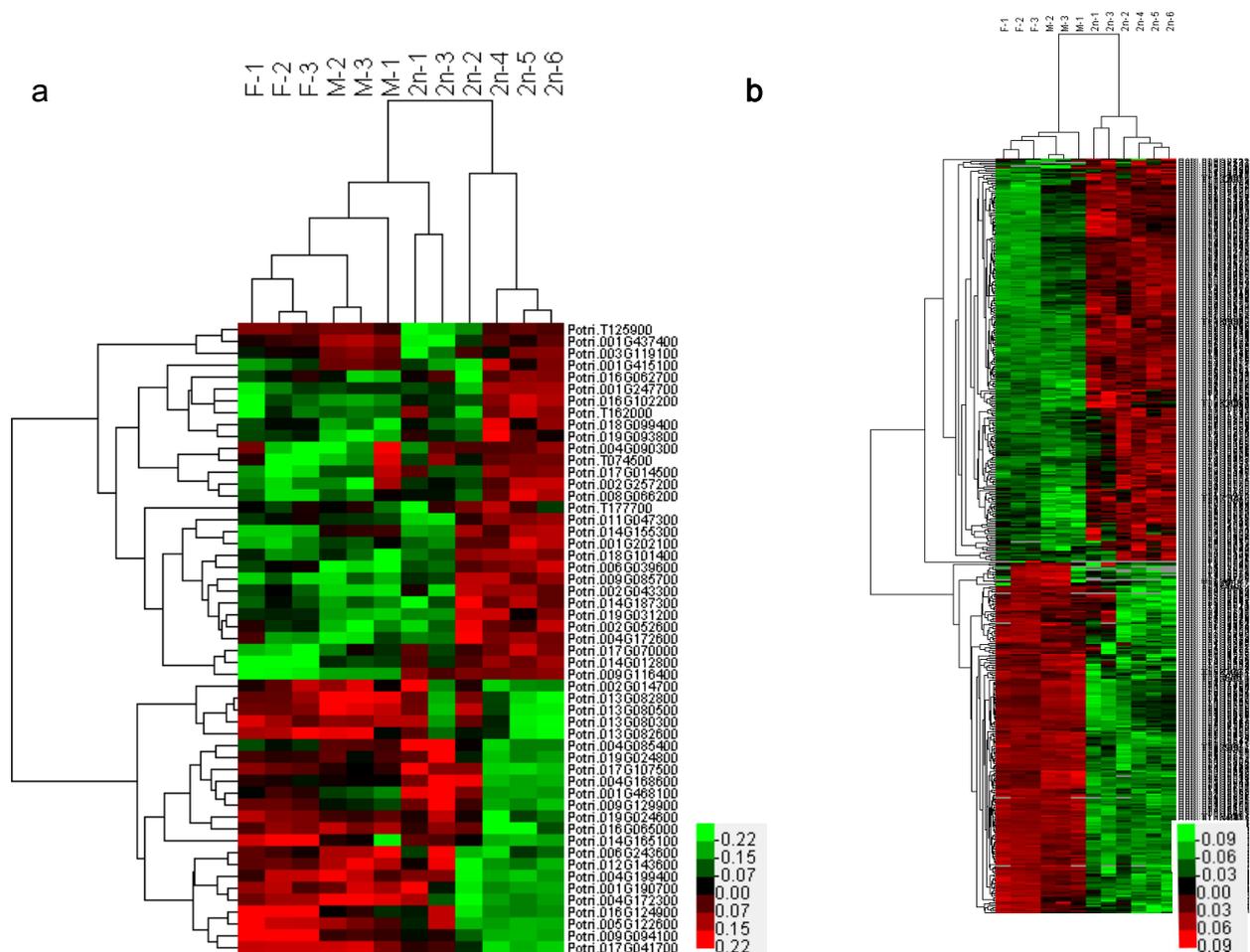


Fig. 3. The figures a and b represent hierarchical clustering of differentially expressed genes (DEGs) in the DG3 and DG4 subgroups, respectively. Three biological replicates of the paternal parent (F-1, F-2, F-3), maternal parent (M-1, M-2, M-3), low growth plants (2n-1, 2n-2, 2n-3) and high growth plants (2n-4, 2n-5, 2n-6). Upregulated genes are indicated in red. Downregulated genes are indicated in green.

To identify metabolic pathways involved in the high growth vigor of the hybrids, pathway-based analysis was performed using the KEGG pathway database. As shown in Figure 5, DEGs of the F<sub>1</sub>-H hybrids belonged mainly to the KEGG pathways for flavonoid biosynthesis, ribosome biogenesis in eukaryotes, and purine and pyrimidine metabolism. DEGs in the DG<sub>3</sub><sup>up</sup>, DG<sub>3</sub><sup>down</sup>, DG<sub>4</sub><sup>up</sup>, and DG<sub>4</sub><sup>down</sup> subgroups had three KEGG pathways in common: biosynthesis of secondary metabolites, metabolic pathways, and photosynthesis.

The most significant KEGG pathways associated with the DEGs in the DG3 subgroup were flavonoid biosynthesis, nitrogen metabolism, alanine, aspartate, and glutamate metabolism, and porphyrin and chlorophyll metabolism. In contrast, GO terms related to pyrimidine metabolism, purine metabolism, the citrate cycle (TCA cycle), plant hormone signal transduction, and starch and sucrose metabolism were highly enriched in DG<sub>HP</sub> subgroup DEGs that were expressed exclusively in the apical buds during the rapid growth period. These results suggest that there are considerable differences in apical bud physiological processes between the DEG<sub>HL</sub> and DEG<sub>HP</sub> groups. These annotations provide a valuable resource for investigating the underlying molecular genetic basis for specific phenotypic traits in *Populus* section *Tacamahaca*.

**8 genes were selected based on Co-expression network related to high growth vigor:** Gene co-expression network analysis of high-growth-vigor hybrids was performed for DEGs in the DG4 subgroup to identify core regulatory genes involved in high growth vigor which were determined by the k-core representing core degree of the regulated genes, k-core analysis is an iterative process in which the nodes are removed from the graphs in order of least connected (Wuchty & Almaas, 2005). In the DG4 subgroup, the genes with the highest k-core values had functions related primarily to heterosis (Fig. 6). These genes were Potri.002G202400, Potri.002G114800, Potri.016G109600, Potri.015G147700, Potri.015G075600, Potri.010G097400, Potri.007G071100 and Potri.003G218500. GO terms for these core genes were enriched for xylem development, cellulose catabolic processes, meristem maintenance, maintenance of shoot apical meristem identity, cell wall macromolecule metabolic processes, purine nucleotide biosynthetic processes, pyrimidine ribonucleotide biosynthetic processes, and RNA methylation (Table 4). All of the selected genes were upregulated and the GO terms had a significance level of  $p < 0.05$  (Table S1 and S2).

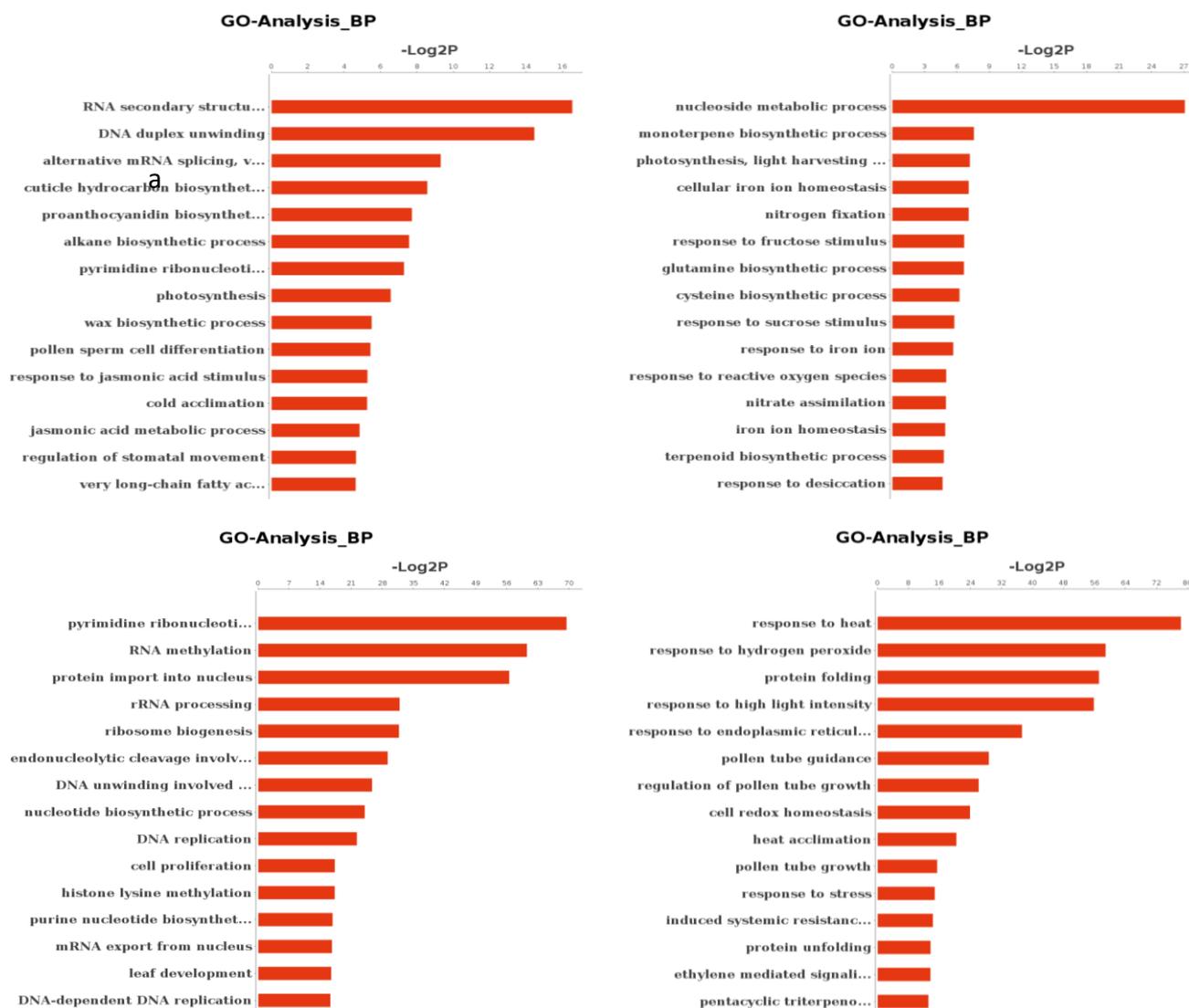


Fig. 4. Biological process classifications for GO-annotated genes associated with apical buds during the rapid growth stage. a, DG3<sub>up</sub> biological process classifications; b, DG3<sub>down</sub> biological process classifications; c, DG4<sub>up</sub> biological process classifications; d, DG4<sub>down</sub> biological process classifications.

## Discussion

Cross-breeding is an important method of genetic improvement in forestry, and the use of heterosis is key for achieving improved timber production. However, not all hybrid trees produced by cross-breeding show improved characteristics. Studies of the molecular basis of heterosis have focused on the super hybrids produced by cross-breeding, but ignored the inferior hybrids that are also produced. Currently only the phenotypic and physiological differences are studied in the first generation of *Populus deltoides* which are from different growth vigor clones (Huang *et al.*, 2012). Research on the molecular basis of heterosis in *Populus* has been limited, and the molecular and genetic mechanisms underlying the phenotypic changes in hybrids remain poorly understood.

**Comparative analysis of annotated DEGs in the DG3 and DG4 subgroups:** Comparative transcriptome analysis revealed a subset of transcripts that were differentially expressed between the hybrids and their

parents in apical buds during the rapid growth stage. Some potential regulators of heterosis in apical bud development were discovered. A large number of upregulated DEGs that were functionally related to photosynthesis and pyrimidine ribonucleotide biosynthetic processes were identified in the DG3 subgroup. Interestingly, the downregulated DEGs were associated only with photosynthesis and nucleoside metabolic processes. Another result of interest is that the DG3 and DG4 subgroups have nucleotide biosynthetic processes in common. We therefore conclude that photosynthesis and nucleic acid metabolism pathways may contribute significantly to apical bud development. Purine and pyrimidine nucleotides participate in many biochemical processes in plants, and are building blocks for nucleic acid synthesis, serve as an energy source, and are precursors for the synthesis of primary products such as sucrose, polysaccharides, and phospholipids as well as secondary products. Therefore, biosynthesis and metabolism of nucleotides are of fundamental importance in the growth and development of plants.

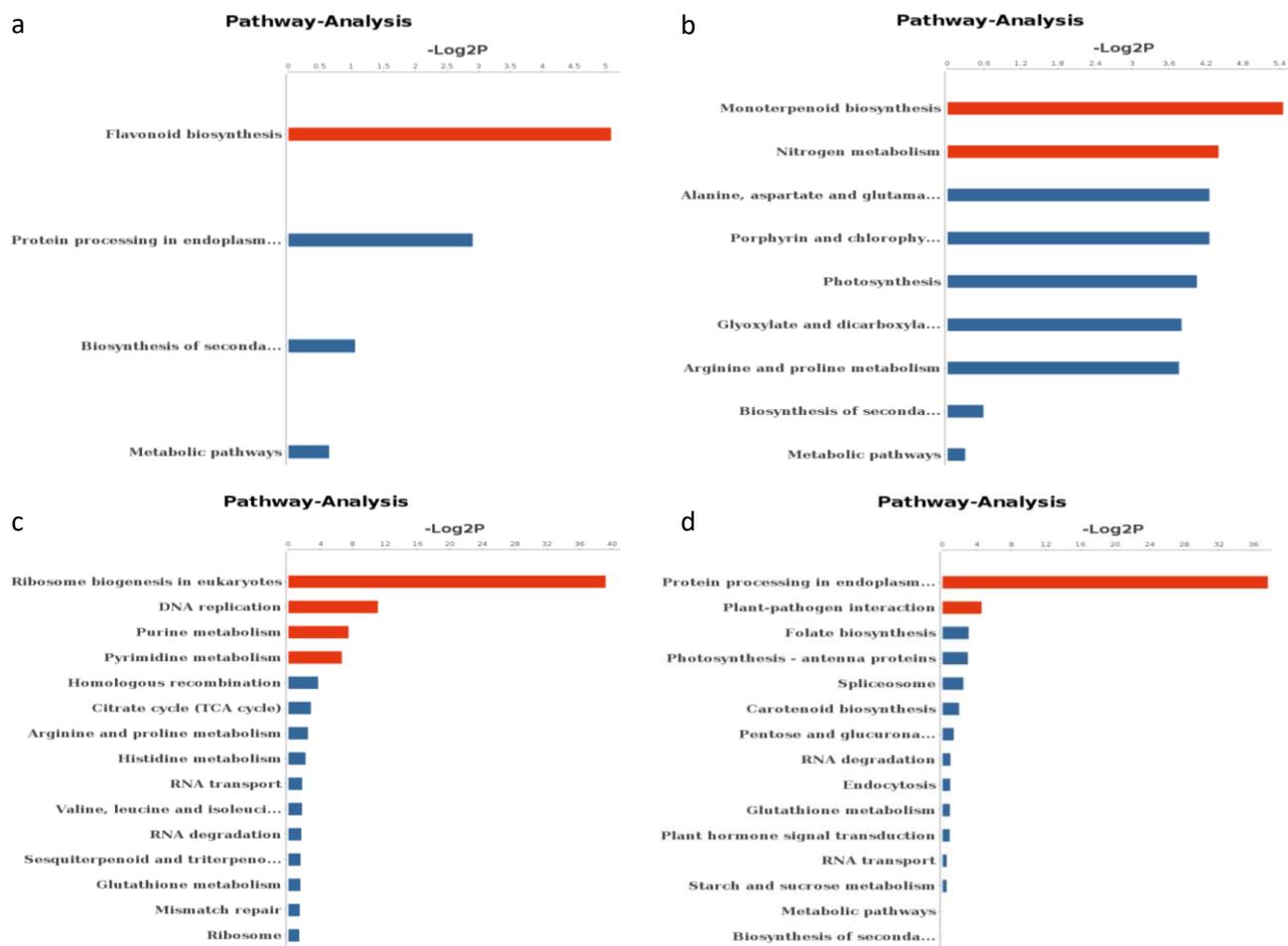


Fig. 5. KEGG Pathway assignments for apical buds during the rapid growth stage. a, DG3<sub>up</sub> pathway assignments; b, DG3<sub>down</sub> pathway assignments; c, DG4<sub>up</sub> pathway assignments; d, DG4<sub>down</sub> pathway assignments.

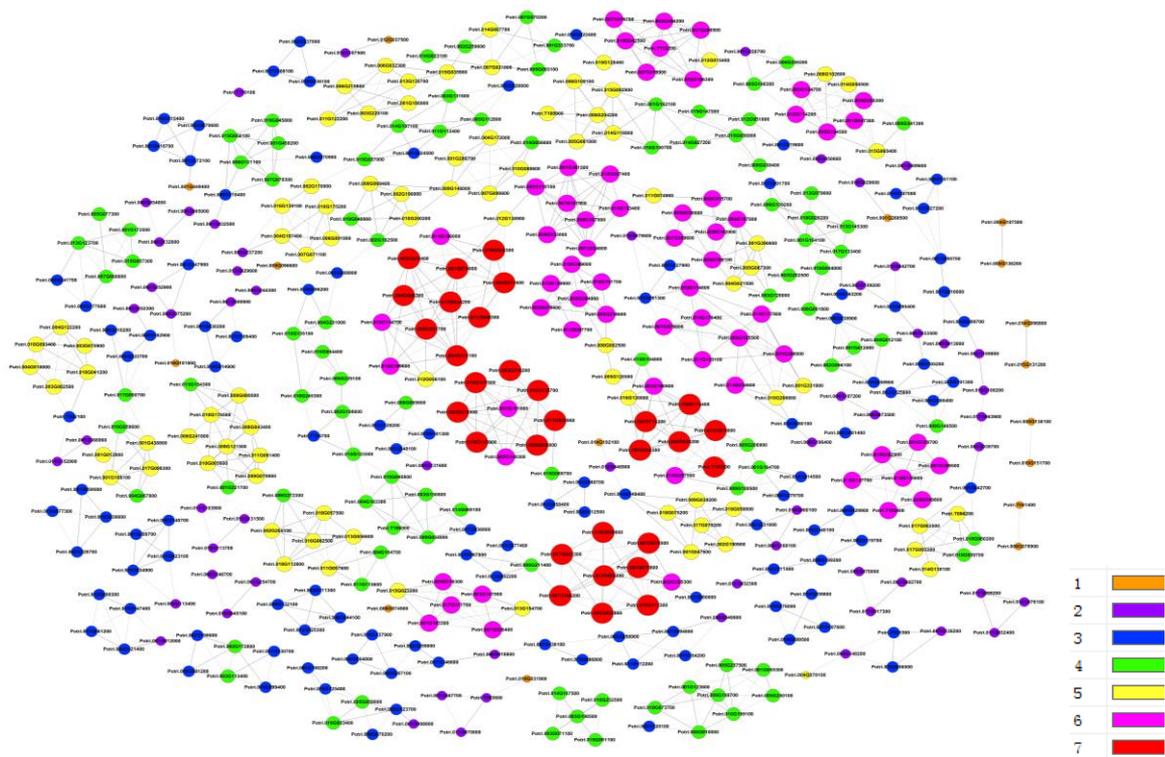


Fig. 6. Co-expression networks of genes related to high growth vigor in DG4 subgroup. For each pair of genes, we calculate the Pearson's correlation and choose the significant correlation pairs to construct the network.

**Table 4. The candidate genes and their biological functions in the k-core subnetworks.**

Gene model	Degree <sup>a</sup>	K-core <sup>b</sup>	Gene description	GO term
Potri.002G202400	10	7	GO:0010089	xylem development cellulose catabolic process cell wall macromolecule metabolic process carbohydrate metabolic process
Potri.002G114800	8	6	GO:0010089	xylem development meristem maintenance meristem initiation maintenance of shoot apical meristem identity
Potri.016G109600	7	6	GO:0005986 GO:0010228 GO:0045893 GO:0006364 GO:0006399 GO:0009658	sucrose biosynthetic process vegetative to reproductive phase transition of meristem positive regulation of transcription, DNA-dependent rRNA processing tRNA metabolic process chloroplast organization
Potri.015G147700	6	6	GO:0006164 GO:0009220 GO:0006412 GO:0006396 GO:0001510	purine nucleotide biosynthetic process pyrimidine ribonucleotide biosynthetic process translation RNA processing RNA methylation
Potri.015G075600	8	7	GO:0045893 GO:0006351 GO:0006355 GO:0009913	positive regulation of transcription, DNA-dependent transcription, DNA-dependent regulation of transcription, DNA-dependent epidermal cell differentiation
Potri.010G097400	8	6	GO:0051726 GO:0006275 GO:0010228 GO:0007049 GO:0006261 GO:0006270 GO:0048478	regulation of cell cycle regulation of DNA replication vegetative to reproductive phase transition of meristem cell cycle DNA-dependent DNA replication DNA replication initiation replication fork protection
Potri.007G071100	5	5	GO:0001510 GO:0042254 GO:0048366 GO:0010305 GO:0048367 GO:0010588 GO:0000478 GO:0006364 GO:0009220	RNA methylation ribosome biogenesis leaf development leaf vascular tissue pattern formation shoot system development cotyledon vascular tissue pattern formation endonucleolytic cleavage involved in rRNA processing rRNA processing pyrimidine ribonucleotide biosynthetic process
Potri.003G218500	8	7	GO:0005975	carbohydrate metabolic process

Note: GO annotation reference Table S1, Table S2 and Gene Ontology database. <sup>a</sup> Degrees describe the number of single gene that regulates other genes represent the size of the cycle node. The higher the degree, the more central the gene occurs within the network

**Table S1. Enriched GO terms in the biological process category in the DG<sub>HL</sub> group (DG<sub>3up</sub> and DG<sub>3down</sub> subgroups) during the rapid growth stage in apical buds (p<0.05).**

GO_term*	GO_Term_annotation	P-Value
GO:0010501_Up	RNA secondary structure unwinding	1.05557E-05
GO:0032508_Up	DNA duplex unwinding	4.47037E-05
GO:0000380_Up	alternative mRNA splicing, via spliceosome	0.001580023
GO:0006723_Up	cuticle hydrocarbon biosynthetic process	0.002631903
GO:0010023_Up	proanthocyanidin biosynthetic process	0.004732145
GO:0043447_Up	alkane biosynthetic process	0.005256474
GO:0009220_Up	pyrimidine ribonucleotide biosynthetic process	0.006393127
GO:0015979_Up	photosynthesis	0.010588813
GO:0010025_Up	wax biosynthetic process	0.021881495
GO:0048235_Up	pollen sperm cell differentiation	0.022910743
GO:0009753_Up	response to jasmonic acid stimulus	0.025697981
GO:0009631_Up	cold acclimation	0.025991612
GO:0009694_Up	jasmonic acid metabolic process	0.034665013
GO:0010119_Up	regulation of stomatal movement	0.039728781
GO:0000038_Up	very long-chain fatty acid metabolic process	0.040233608
GO:0009414_Up	response to water deprivation	0.044247156
GO:0042335_Up	cuticle development	0.047772395
GO:0009116_Down	nucleoside metabolic process	6.96864E-09
GO:0043693_Down	monoterpene biosynthetic process	0.005256474
GO:0009768_Down	photosynthesis, light harvesting in photosystem I	0.006827705
GO:0006879_Down	cellular iron ion homeostasis	0.007350864
GO:0009399_Down	nitrogen fixation	0.007350864
GO:0009750_Down	response to fructose stimulus	0.009824553
GO:0006542_Down	glutamine biosynthetic process	0.009962289
GO:0019344_Down	cysteine biosynthetic process	0.013264039
GO:0009744_Down	response to sucrose stimulus	0.01846777
GO:0010039_Down	response to iron ion	0.019819554
GO:0000302_Down	response to reactive oxygen species	0.031103558
GO:0042128_Down	nitrate assimilation	0.031613188
GO:0055072_Down	iron ion homeostasis	0.033140376
GO:0016114_Down	terpenoid biosynthetic process	0.036187106
GO:0009269_Down	response to desiccation	0.039223672

\* “Up” represents GO terms and functional annotation of upregulated genes, “Down” represents GO terms and functional annotation of downregulated genes

**The 8 candidate genes and their significant biological functions in the k-core subnetworks:** We selected 8 genes associated with high-growth-vigor hybrid plants. Potri.015G147700 gene has biological functions in RNA methylation, pyrimidine ribonucleotide and purine nucleotide biosynthetic process. Potri.002G202400 gene sequences were annotated based on sequence alignments generated using BLAST searches against NCBI databases. It encodes a *Populus trichocarpa* endo-1,4-beta-glucanase (GH9B13) gene. Endo-1,4-β-glucanases (EGases) are involved in many aspects of plant growth (Yu *et al.*, 2013). The isolation of two poplar endo-1,4-β-glucanases, PopCel1 and PopCel2, also presents indirect evidence that endo-1,4-β-glucanases can digest non-crystalline regions of cellulose allowing the enzymes to loosen walls by causing the release of xyloglucans trapped in cellulose micro-fibrils (Ohmiya *et al.*, 2000). PtrCel9A6 is another gene that encodes an endo-1,4-β-glucanase. Suppression of PtrCel9A6 expression in *Populus* caused secondary cell wall defects in xylem cells and a significant reduction in cellulose. Heterologous expression of PtrCel9A6 in *Arabidopsis* enhanced plant growth and increased fiber-cell length (Yu *et al.*, 2013). Putative membrane-anchored endo-β-1,4-D-glucanases are encoded by three genes, OsGLU1, OsGLU2, and OsGLU3, in the rice genome, all

of which are expressed ubiquitously in rice plants and promote the growth of various tissues (Zhou *et al.*, 2006). The Potri.002G114800 and Potri.016G109600 genes are involved in the maintenance of shoot apical meristem identity, regulation of cell proliferation, and the vegetative-to-reproductive phase transition of the meristem. The Potri.015G075600, Potri.010G097400, and Potri.007G071100 genes may be involved in positive regulation of transcription and RNA methylation. The Potri.003G218500 gene may regulate carbohydrate metabolic process. The identification of these 8 core genes lays the foundation for further functional analyses.

We used RNA-Seq analysis to systematically investigate the global transcriptomes of apical buds from F<sub>1</sub>-H, F<sub>1</sub>-L, and parental genotypes during the rapid growth period, thereby generating a useful resource for the *Populus* research community. We analyzed the DG<sub>HP</sub> and DG<sub>HL</sub> groups using the GO and KEGG databases to functionally screen for candidate transcripts that may contribute significantly to apical bud growth and development. The changes in the expression of the candidate transcripts may provide valuable information for future studies on the molecular mechanisms underlying apical bud heterosis.

**Table S2. Enriched GO terms in the biological process category in the DG<sub>HP</sub> group (DG<sub>4up</sub> and DG<sub>4down</sub> subgroups) during the rapid growth stage in apical buds (p<0.05).**

GO_term*	GO_Term_annotation	P-Value
GO:0009220_Up	pyrimidine ribonucleotide biosynthetic process	1.19921E-21
GO:0001510_Up	RNA methylation	5.81586E-19
GO:0006606_Up	protein import into nucleus	9.21151E-18
GO:0006364_Up	rRNA processing	2.49528E-10
GO:0042254_Up	ribosome biogenesis	2.76616E-10
GO:0000478_Up	endonucleolytic cleavage involved in rRNA processing	1.61096E-09
GO:0006268_Up	DNA unwinding involved in replication	1.8511E-08
GO:0009165_Up	nucleotide biosynthetic process	5.80297E-08
GO:0006260_Up	DNA replication	1.97592E-07
GO:0008283_Up	cell proliferation	6.19597E-06
GO:0034968_Up	histone lysine methylation	6.2729E-06
GO:0006164_Up	purine nucleotide biosynthetic process	8.76896E-06
GO:0006406_Up	mRNA export from nucleus	9.7473E-06
GO:0048366_Up	leaf development	1.09358E-05
GO:0006261_Up	DNA-dependent DNA replication	1.24171E-05
GO:0051726_Up	regulation of cell cycle	1.82239E-05
GO:0006270_Up	DNA replication initiation	2.67246E-05
GO:0006275_Up	regulation of DNA replication	3.4194E-05
GO:0006626_Up	protein targeting to mitochondrion	0.000144084
GO:0009793_Up	embryo development ending in seed dormancy	0.000163722
GO:0031120_Up	snRNA pseudouridine synthesis	0.000283013
GO:0006306_Up	DNA methylation	0.000356877
GO:0006541_Up	glutamine metabolic process	0.000495998
GO:0051567_Up	histone H3-K9 methylation	0.000539146
GO:0009909_Up	regulation of flower development	0.000593691
GO:0006396_Up	RNA processing	0.000984222
GO:0009561_Up	megagametogenesis	0.001180982
GO:0051604_Up	protein maturation	0.001638319
GO:0009555_Up	pollen development	0.001798922
GO:0006221_Up	pyrimidine nucleotide biosynthetic process	0.00192938
GO:0006265_Up	DNA topological change	0.002484393
GO:0006189_Up	'de novo' IMP biosynthetic process	0.002484393
GO:0008033_Up	tRNA processing	0.002621605
GO:0010588_Up	cotyledon vascular tissue pattern formation	0.003105508
GO:0008295_Up	spermidine biosynthetic process	0.003440475
GO:0006259_Up	DNA metabolic process	0.003846376
GO:0009451_Up	RNA modification	0.004000456
GO:0051301_Up	cell division	0.004090769
GO:0031507_Up	heterochromatin assembly	0.005353466
GO:0001522_Up	pseudouridine synthesis	0.005782671
GO:0009957_Up	epidermal cell fate specification	0.006227074
GO:0010075_Up	regulation of meristem growth	0.00865775
GO:0080057_Up	sepal vascular tissue pattern formation	0.008688913
GO:0080056_Up	petal vascular tissue pattern formation	0.008688913
GO:0048478_Up	replication fork protection	0.008688913
GO:0000469_Up	cleavage involved in rRNA processing	0.008688913
GO:0009553_Up	embryo sac development	0.008860096
GO:0009116_Up	nucleoside metabolic process	0.010312568
GO:0009664_Up	plant-type cell wall organization	0.010972461
GO:0010389_Up	regulation of G2/M transition of mitotic cell cycle	0.011875062
GO:0009832_Up	plant-type cell wall biogenesis	0.011949133
GO:0060321_Up	acceptance of pollen	0.013004817
GO:0000724_Up	double-strand break repair via homologous recombination	0.014393006
GO:0016458_Up	gene silencing	0.015756184
GO:0010817_Up	regulation of hormone levels	0.016463988
GO:0045793_Up	positive regulation of cell size	0.017301798
GO:0006527_Up	arginine catabolic process	0.017301798
GO:0001708_Up	cell fate specification	0.019571018

Table S2. (Cont'd.)

GO_term*	GO_Term_annotation	P-Value
GO:0042991_Up	transcription factor import into nucleus	0.020327583
GO:0009202_Up	deoxyribonucleoside triphosphate biosynthetic process	0.021579939
GO:0009186_Up	deoxyribonucleoside diphosphate metabolic process	0.021579939
GO:0007009_Up	plasma membrane organization	0.021579939
GO:0006177_Up	GMP biosynthetic process	0.021579939
GO:0009263_Up	deoxyribonucleotide biosynthetic process	0.021579939
GO:0061062_Up	regulation of nematode larval development	0.021579939
GO:0009446_Up	putrescine biosynthetic process	0.021579939
GO:0006281_Up	DNA repair	0.021771516
GO:0052541_Up	plant-type cell wall cellulose metabolic process	0.021876702
GO:0019654_Up	acetate fermentation	0.025839324
GO:0006529_Up	asparagine biosynthetic process	0.025839324
GO:0002943_Up	tRNA dihydrouridine synthesis	0.025839324
GO:0051131_Up	chaperone-mediated protein complex assembly	0.025839324
GO:0045995_Up	regulation of embryonic development	0.025839324
GO:0009432_Up	SOS response	0.025839324
GO:0052546_Up	cell wall pectin metabolic process	0.025953857
GO:0016444_Up	somatic cell DNA recombination	0.027663995
GO:0009932_Up	cell tip growth	0.02773987
GO:0000741_Up	karyogamy	0.029418052
GO:0009920_Up	cell plate formation involved in plant-type cell wall biogenesis	0.030080036
GO:0006557_Up	S-adenosylmethioninamine biosynthetic process	0.030080036
GO:0006168_Up	adenine salvage	0.030080036
GO:0030643_Up	cellular phosphate ion homeostasis	0.030080036
GO:0031497_Up	chromatin assembly	0.034302157
GO:0016554_Up	cytidine to uridine editing	0.034302157
GO:0006207_Up	'de novo' pyrimidine nucleobase biosynthetic process	0.034302157
GO:0006428_Up	isoleucyl-tRNA aminoacylation	0.034302157
GO:0044205_Up	'de novo' UMP biosynthetic process	0.034302157
GO:0009825_Up	multidimensional cell growth	0.036475562
GO:0006597_Up	spermine biosynthetic process	0.038505769
GO:0010388_Up	cullin deneddylation	0.039855251
GO:0000271_Up	polysaccharide biosynthetic process	0.040433363
GO:0031167_Up	rRNA methylation	0.042690953
GO:0005991_Up	trehalose metabolic process	0.042690953
GO:0006564_Up	L-serine biosynthetic process	0.042690953
GO:0000266_Up	mitochondrial fission	0.042690953
GO:0010498_Up	proteasomal protein catabolic process	0.043386952
GO:0005982_Up	starch metabolic process	0.043884176
GO:0016226_Up	iron-sulfur cluster assembly	0.043990209
GO:0045962_Up	positive regulation of development, heterochronic	0.046857792
GO:0010015_Up	root morphogenesis	0.04703583
GO:0009408_Down	response to heat	2.70281E-24
GO:0042542_Down	response to hydrogen peroxide	1.89468E-18
GO:0006457_Down	protein folding	6.19859E-18
GO:0009644_Down	response to high light intensity	1.5482E-17
GO:0034976_Down	response to endoplasmic reticulum stress	5.93181E-12
GO:0010183_Down	pollen tube guidance	2.21184E-09
GO:0080092_Down	regulation of pollen tube growth	1.36134E-08
GO:0045454_Down	cell redox homeostasis	6.40279E-08
GO:0010286_Down	heat acclimation	7.35891E-07
GO:0009860_Down	pollen tube growth	2.28794E-05
GO:0006950_Down	response to stress	3.48906E-05
GO:0009864_Down	induced systemic resistance, jasmonic acid mediated signaling pathway	4.88659E-05
GO:0043335_Down	protein unfolding	7.47698E-05
GO:0009873_Down	ethylene mediated signaling pathway	7.61678E-05
GO:0019745_Down	pentacyclic triterpenoid biosynthetic process	0.000109725

\* "Up" represents GO terms and functional annotation of upregulated genes, "Down" represents GO terms and functional annotation of downregulated genes

## Acknowledgments

The authors wish to thank Drs. Dai Chen and Bo Zhang (Novel Bioinformatics Ltd., Co, Shanghai, China) for their technical assistance in bioinformatics analysis. This work was jointly supported by Scientific Research Foundation for the introduction of talent of Pingdingshan University (No. PXY-BSQD2016009), Key Research Project of Colleges and Universities of Henan Province (16A220004 and 152300410172) and College Students Science and Technology Innovation Project (S & TIF2017147).

## References

- Bao, J.Y., S. Lee, C. Chen, X.Q. Zhang, Y. Zhang, S.Q. Liu, T. Clark, J. Wang, M.L. Cao, H.M.
- Braatne, J.H, T.M. Hinckley and R.F. Stettler. 1992. Influence of soil water on the physiological and morphological components of plant water balance in *Populus trichocarpa*, *Populus deltoides* and their F<sub>1</sub> hybrids. *Tree. Physiol.*, 11: 325-339.
- Cheng, Y.Q., J.F. Liu, X.D. Yang, R. Ma, C.M. Liu and Q. Liu. 2013. RNA-seq analysis reveals ethylene-mediated reproductive organ development and abscission in Soybean (*Glycine max* L. Merr.). *Plant. Mol. Biol. Rep.*, 31: 607-619.
- Dorogovtsev, S.N., A.V. Goltsev and J.F. Mendes. 2006. k-Core architecture and k-core percolation on complex networks. *Physica D: Nonlinear. Phenomena.*, 224: 7-19.
- Gao, Y., H. Zhang, Q. Gao, L.L. Wang, F.C. Zhang, V.S. Siva, Z. Zhou, L.S. Song and S.C. Zhang. 2013. Transcriptome analysis of artificial hybrid pufferfish *Jiyan-1* and its parental species: implications for Pufferfish heterosis. *PLoS. one.*, 8: e58453.
- Garber, M., M.G. Grabherr, M. Guttman and C. Trapnell. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods.*, 8: 469-477.
- Geraldes, A., J. Pang, N. Thiessen, T. Cezard, R. Moore, Y.J. Zhao, A. Tam, S.C. Wang, M. Friedmann, I. Birol, S.J.M. Jones, QCB. Cronk, C.J. Douglas. 2011. SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. *Mol. Ecol. Resour.*, 11: 81-92.
- Grosu, P., J.P. Townsend, D.L. Hartl and D. Cavalieri. 2002. Pathway processor: a new tool for analyzing gene expression data. *Genome. Res.*, 12: 1121-1126.
- Huang, G.W., X.H. Su and Q.J. Huang. 2012. Differences in growth and physiological characteristics in different growth vigor clones of *Populus deltoids*. *Scientia. Silvae. Sinicae.*, 3: 53-54.
- Huang, G.W., X.H. Su and Q.J. Huang. 2012. Research progress of mechanism of heterosis in plant. *World. Forestry. Research.*, 25: 13-18.
- Muhammad, Y.K.B., M. Qasim and M. Din. 2015. Profiling micorRNAs and their targets in radish (*Raphanus sativus* L.). *Pak. J. Bot.*, 47(1): 171-176.
- Ohmiya, Y., M. Samejima, M. Shiroyoshi, Y. Amano, T. Kanda, F. Sakai and T. Hayashi. 2000. Evidence that endo-1,4-betaglucanases act on cellulose in suspension-cultured poplar cells. *Plant. J.*, 24: 147-158.
- Pharis, R.P., F.C. Yeh and B.P. Dancik. 1991. Superior growth potential in trees: What is its basis, and can it be tested for at an early age. *Can. J. For. Res.*, 21: 368-374.
- Pirooznia, M., V. Nagarajan and Y.P. Deng. 2007. GeneVenn-A web application for comparing gene lists using Venn diagrams. *Bioinformatics*, 1: 420-422.
- Qiu, Q., T. Ma, Q.J. Hu, B.B. Liu, Y.X. Wu, H.H. Zhou, Q. Wang, J. Wang, J.Q. Liu and R. Sederoff. 2011. Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. *Tree. Physiol.*, 31: 452-461.
- Song, SH., H.Z. Qu, C. Chen, S.N. Hu and J. Yu. 2007. Differential gene expression in an elite hybrid rice cultivar (*Oryza sativa* L.) and its parental lines based on SAGE data. *BMC. Plant. Bio.*, 7: 49.
- Stettler, R.F., P.E. Heliman and H.D. Bradshaw. 1999. Biology of *Populus* and its implications for management and conservation. NRC. *Research. Press, Ottawa.*, pp. 57-85.
- Thiemann, A., J. Fu, T.A. Schrag, A.E. Melchinger, M. Frisch and S. Scholten. 2010. Correlation between parental transcriptome and field data for the characterization of heterosis in *Zea mays* L. *Theor. Appl. Genet.* 120: 401-413.
- Tuskan, G.A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts and A. Salamov. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313: 1596-1604.
- Wang, J. 2009. Techniques of polyploid induction in *Populus* spp. (Section Tacamahaca). Dissertation, Beijing. *Forestry. University.*
- Wang, L.K., Z.X. Feng, X. Wang and X.G. Wang. 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, 26: 136-138.
- Wang, Z., M. Gerstein and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10: 57-63.
- Wuchty, S and E. Almaas. 2005. Peeling the yeast protein network. *Proteomics.*, 5: 444-449.
- Yang, S.M. Wang and J. Yu. 2005. Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. *Plant. Physiol.*, 138: 1216-1231.
- Yang, C.C, Q.J. Huang and X.H. Su. 2010. Research progress of genetic mechanism of heterosis in wood. *World. Forestry. Res.*, 23: 25-29.
- Yao, Y.Y., Z.F. Ni, Y.H. Zhang, Y. Chen, Y.H. Ding, Z.F. Han, Z.Y. Liu and Q.X. Sun. 2005. Identification of differentially expressed genes in leaf and root between wheat hybrid and its parental inbreds using PCR-based cDNA subtraction. *Plant. Mol. Biol.*, 58: 367-384.
- Yu, L.L., J.J. Sun and L.G. Li. 2013. PtrCel9A6, an endo-1,4-β-glucanase, is required for cell wall formation during xylem differentiation in *Populus*. *Mol. Plant.*, 6: 1904-1917.
- Zhai, R.R., Y. Feng, H. Wang, X. Zhan, X. Shen, W. Wu, Y. Zhang, D.B. Chen, G.X. Dai, Z.L. Yang, L.Y. Cao and S.H. Cheng. 2013. Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC. Genomics.*, 14: 19.
- Zhang, C., Y.J. Wang, J.X. Fu, L. Dong, S.L. Gao and D.N. Du. 2013. Transcriptomic analysis of cut tree peony with glucose supply using the RNA-Seq technique. *Plant. Cell. Rep.*, 33: 111-129.
- Zhang, Y.G, J. Zhu and H.Y. Dai. 2012. Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq. *Plant. Mol. Biol. Rep.*, 30: 957-965.
- Zhang, Z., Z.Z. Sun, X. Xiao, S.X. Zhou, X.C. Wang, J. Gu, L.L. Qiu, X.H. Zhang, Q.J. Xu, B.H. Zhen, X.R. Wang and S.L. Wang. 2013. Mechanism of BDE209-induced impaired glucose homeostasis based on gene microarray analysis of adult rat liver. *Arch. Toxicol.*, 87: 1557-1567.
- Zhou, H.L., S.J. He, Y.R. Cao, T. Chen, B.X. Du, C.C. Chu, J.S. Zhang and S.Y. Chen. 2006. OsGLU1, a putative membrane-bound endo-1,4-beta-D-glucanase from rice, affects plant internode elongation. *Plant. Mol. Biol.*, 60: 137-151.