Scientific Research Publishing

American Journal of Plant Sciences, 2019, 10, 79-100
http://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

RNA-Seq and Bulked Segregant Analysis of Genes Related to High Growth in *Ginkgo biloba* Half-Sibling Families

Limin Sun[1*], Haixia Tang[2*], Xiaoyan Men[1], Qian Zhang[3], Xia Sun[4#], Shiyan Xing[1#]

[1]Key Laboratory of Tree Germplasm Resources Research, College of Forestry, Shandong Agricultural University, Tai'an, China
[2]Shandong Institute of Pomology, Tai'an, China
[3]Taishan Academy of Forestry Sciences Tai'an City, Tai'an, China
[4]College of Horticulture Science and Engineering, Shandong Agriculture University, Tai'an, China
Email: #sunxia65@sina.com, #xingsy@sdau.edu.cn

How to cite this paper: Sun, L.M., Tang, H.X., Men, X.Y., Zhang, Q., Sun, X. and Xing, S.Y. (2019) RNA-Seq and Bulked Segregant Analysis of Genes Related to High Growth in *Ginkgo biloba* Half-Sibling Families. *American Journal of Plant Sciences*, **10**, 79-100.
https://doi.org/10.4236/ajps.2019.101008

Received: December 19, 2018
Accepted: January 11, 2019
Published: January 14, 2019

Abstract

The lifetime of *G. biloba* is very long, and its growth is relatively slow. However, little is known about growth-related genes in this species. We combined mRNA sequencing (RNA-Seq) with bulked segregant analysis (BSA) to fine map significant agronomic trait genes by developing polymorphism molecular markers at the transcriptome level. In this study, transcriptome sequencing of high growth (GD) and low growth (BD) samples of *G. biloba* half-sib families was performed. After assembling the clean reads, 601 differential expression genes were detected and 513 of them were assigned functional annotations. Single nucleotide polymorphism (SNP) analysis identified SNPs associated with 119 genes in the GD and BD groups; 58 of these genes were annotated. Two Homeobox-leucine zipper protein genes were up-regulated in the GD group compared with the BD group; therefore, these are very likely related to high growth of *G. biloba.* This study provides molecular level data that could be used for seed selection of high growth *G. biloba* half-sib families for future breeding programs.

Keywords

High Growth, *Ginkgo biloba* Half-Sibling Families, RNA-Seq and Bulked Segregant Analysis, the Transcriptome Sequencing, Differentially Expressed Genes

*These authors contributed equally to this work.

1. Introduction

Ginkgo biloba is a deciduous tree in the family Ginkgoaceae. It is the only species in China to survive the quaternary glacier and, as such, is recognized as a "living fossil" [1]. *G. biloba* has a very long lifetime, the leaf is fan-shaped, the tree is tall and straight, and its tolerance to drought and barren conditions has made it a significant ornamental, greening, edible, medicinal [2], and timber tree. In China, the cultivated area of *G. biloba* is more than 200,000 hm^2 and the number of trees has been estimated as 913,000 [3]. *G. biloba* growth is relatively slow with the average increment of timber volume reaching its maximum at about 40 years [4] [5]. The tree is generally harvested for maximum timber volume at about 60 years [6]. Until now, most studies have focused on the physiology [7] [8], phylogeny [9] [10], and sex-determining mechanism [11], and molecular biology studies about the growth mechanism of *G. biloba* are relatively few. The genome sequence of *G. biloba* is still unavailable; therefore, genomics studies are relatively difficult. mRNA sequencing (RNA-Seq) is a next-generation sequencing technology [12] [13] [14] [15] that has been used widely to authenticate and quantify normal and rare transcripts, and to provide transcript sequential structure information of specific samples [16] [17] in species without a reference genome. The recent application of RNA-Seq to *G. biloba* aseptic seedlings identified a gene that encoded chalcone isomerase (GbCHI1), one of the key enzymes in the flavonoid biosynthesis pathway, that exhibited differences in the protein sequence compared with a previously identified GbCHI [18]. Transcriptome sequencing of *G. biloba* kernels revealed 66 unigenes that were found to be responsible for terpenoid backbone biosynthesis [19]. In addition, *G. biloba* genes associated with the biosynthesis of bilobalide and paclitaxel were found by transcriptome sequencing [20]. Transcriptome sequencing of the epiphyllous ovules of *G. biloba* var. *epiphylla* identified snRNA genes associated with the adjustment and control of ovular development. However, no studies into high growth-related genes in *G. biloba* have been reported so far. The growth of *G. biloba* can be affected by a combination of the environment, inheritance, and other factors [21] [22]; therefore, we aimed to study growth-related genes in a large group of *G. biloba* plants to obtain a comprehensive overview of the genes involved.

We combined RNA-seq with bulked segregant analysis (BSA) to fine mapping genes associated with significant agronomic traits gene at the transcriptome level. BSA has been used to rapidly identify genetic markers linked to a genomic region associated with a selected phenotype [23] [24]. The fundamental principle of BSA is that extreme differences of individual phenotype or genotype can be used as the basis on which individuals are selected to obtain a DNA mixture, so that two DNA pools equivalent to near-isogenic line can be built. BSA can be used for efficient marker enrichment in a target region [25]. BSA has been used for a wide range of plant genomic applications, such as genome sequencing in barley [26] [27], Arabidopsis [28], rice [29], corn [30], and sunflower. In the

combined technology (here named BSR), RNA-Seq is used to provide BSA with genotype data in the RNA pools. Linked genes (low in false positives) can be screened out after data analysis, which greatly improves the efficiency of development and verification of linked polymorphism markers. For species with no reference genome, the RNA-Seq data are assembled to obtain unigenes that are subjected to a series of bioinformatics analysis, including genetic structure annotation, gene expression analysis, and gene function annotation.

G. biloba half-sib families from a nursery stock at the seedling stage were used in this study. High growth (GD) and low growth (BD) RNA pools were built from the group level, and BSR was used to identify candidate genes related to the high growth trait. These data will expand the existing transcriptome resources of *G. biloba*, and provide a valuable platform for further studies on developmental and metabolic mechanisms in this species. The information can also be used for functional gene studies and molecular breeding programs.

2. Materials and Methods

2.1. Genetic Materials

G. biloba seedings were obtained from the *G. biloba* germplasm resource garden of the Gaoqiao Tree Farm in Tai'an City, Shandong Province, China. The experimental field is located N 35˚54', E 116˚53', which has a continental warm temperature zone medium-latitude monsoon climate. The average annual temperature is 13.4˚C, and the maximum and minimum recorded air temperatures are 40.7˚C, and −19˚C, respectively. The annual average rainfall is 689.6 mm, average annual evaporation is 1169.8 mm, and the average number of frost-free days is 206 per year. The soil is cinnamon soil, which has a neutral to slightly alkaline reaction. The thickness of the soil layer is mostly more than 120cm, and the soil retains water and maintains fertilizer performance. The underground is weathered and broken limestone and shale. A total of 358 seeds were collected from a 200-year-old ancient Ginkgo biloba tree in Shiqiao Town, Pan County, Guizhou Province on 29 September 2013. Seeding was conducted in 2014 and 194 seedlings emerged. After planting, field management measures were uniform throughout. Seedling height was measured in December 2014 and November 2015. The 30 tallest seedlings and 30 shortest seedlings were selected from 194 ginkgo half-sib family seedlings. Three biological replicates were used for each trait, and two mixed groups, named BD and GD, respectively, were used for transcriptome analysis. The heights of the selected seedlings were recorded for 2 consecutive years, and the variable coefficient of seedling height in the families was >30%. The initial expanded second lamina at the top of the seedlings were punched and then disposed in mixing pool mode in May 2015, quick-frozen in liquid nitrogen, and stored at −80˚C until used.

2.2. Extraction of RNA from *G. biloba* Half-Sib Families Leaf Tablets

Total RNA from each sample was isolated separately using a RN38 EASY spin

plus Plant RNA kit (Aidlab Biotech, Beijing, China). Nanodrop Analyzer (Thermo Science, Wilmington, USA), Qubit 2.0 Fluorometer and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) were used to estimate the purity, concentration, and integrity of the extracted RNA.

2.3. cDNA Library Construction and Sequencing

Total mRNA was isolated by oligo (dT) selection using Dynabeads mRNA DIRECT Kit (invitrogen), and each sample was prepared 5 ug for constructing the cDNA library. The purified mRNA was fragmented at elevated temperature (90˚C), then reverse transcribed to first strand cDNA with random primer. Second strand cDNA was synthesized in the presence of DNA polymerase I and RNaseH. The cDNA was cleaned using Agencourt Ampure XP SPRI beads (Beckman Coulter). The cDNA molecules were subjected to end repair, and add an "A" base at the 3'-end. Illumina adapters were ligated to the cDNA molecules, resultant cDNA library was amplified using PCR for enrichment of adapter ligated fragments. Libraries were prepared from a 400 - 500 bp size-selected fraction following adapter ligation and 2% agarose gel separation. The cDNA library was quantified using qPCR method (>10 nm). It was then sequenced using the Illumina Hi-Seq 2500 platform.

2.4. Unigene Function Annotation

The raw reads were cleaned by removing adapter sequences, reads containing ploy-N, and low-quality sequences (Q < 30). Clean reads were aligned to the reference genome sequence using the program Tophat [31] [32].

The assembled unigene sequences were searched against the Nr, SwissProt, GO, COG, KOG, Pfam, and KEGG databases using the NCBI Basic Local Alignment Search Tool (BLAST) tools [33] to annotate the unigenes.

2.5. Unigene Structural Analysis

The CDSs of the unigenes were predicted based on their alignment to known protein sequences. The predicted CDSs were translated into amino acid sequence using the standard codon table. The unassembled clean reads in each sample were mapped to the assembled unigene sequences. SNP loci were detected using the SNP calling program in the Genome Analysis Toolkit (GATK) [34]. SNP loci were screened then we chose to measure allele segregation using Euclidean distance (ED), as a metric that does not require parental strain in-formation and is resistant to noise [35]. In order to obtain good correlation effect, The ED value was disposed in the 5 power mode, and the data were recognized as the basis for BSR relevance. Using the equation:

$$\mathrm{ED} = \sqrt{\left(A_{mut} - A_{wt}\right)^2 + \left(C_{mut} - C_{wt}\right)^2 + \left(G_{mut} - G_{wt}\right)^2 + \left(T_{mut} - T_{wt}\right)^2}$$

where each letter (A, C, G, T) corresponds to the frequency of its corresponding DNA nucleotide.

2.6. Analysis of Differential Gene Expression

Reads is compared with Unigene bank obtained by sequencing of each sample using Bowtie software [36]. The expression levels were estimated by combining with RSEM [37]. RSEM (RNA-Seq by Expectation Maximization), which implements our quantification method and provides extensions to our original model. The expression levels of the unigenes were expressed as fragments per kilobase of transcript per million mapped reads (FPKM) values to eliminate the influences of gene length and sequencing quantity difference on of the estimate gene expression. FPKM values can be used directly to compare gene expression differences between samples.

FPKM was calculated as follows:

$$\text{FPKM} = \frac{\text{cDNA Fragments}}{\text{Mapped Reads Millions} \times \text{Transcript Length kb}}$$

where "cDNA Fragments" is the number of fragments of one transcript in the sample (*i.e.*, the number of double-end reads); "Mapped Reads Millions" is the number of mapped reads (in this study it was 106); and "Transcript Length kb" is the length of the transcript.

Differential expression analysis between the GD and BD groups was conducted using DESeq [38]. Significance p-values were obtained by original hypothesis testing and adjusted using the Benjamini–Hochberg method. The FDR was used as the key index for screening the DEGs, and the screened DEGs were analyzed in a hierarchical clustering mode.

Data Availability: The raw reads of the RNA-seq are now beening processed by NCBI staff. File S1 contains SNP depth in the RNA-Seq data of *Ginkgo biloba* half-sib families. File S2 contains functional annotation of unigenes of *Ginkgo biloba* half-sib families. File S3 contains Gene Ontology annotation of unigenes of *Ginkgo biloba* half-sib families.

3. Results

3.1. Growth Analysis and Sample Collection of *G. biloba* Half-Sib Families

The average seedling height of the GD group was more than the average height of the two groups for 2 consecutive years, while the average seedling height of the BD group was lower than the average height for 2 consecutive years (**Figure 1(b)**, **Table 1**). The photosynthetic rate (Pn), which reflects the speed of carbon dioxide fixation during photosynthesis, is shown in (**Figure 1(c)**, **Table 1**). The net Pn in the GD group (8.3 μmol·m^{-2}·s^{-1}) was more than the Pn (4.23 μmol·m^{-2}·s^{-1}) in the BD group. The average chlorophyll content, which reflects photosynthetic capacity, was higher in the GD group (8.6 mg·g^{-1}·FW) than in the BD group (4.1 mg·g^{-1}·FW) (**Figure 1(d)**, **Table 1**).

3.2. Illumina HiSeq mRNA Sequencing

After quality control of the RNA-Seq reads, we obtained 30 Gb of clean reads

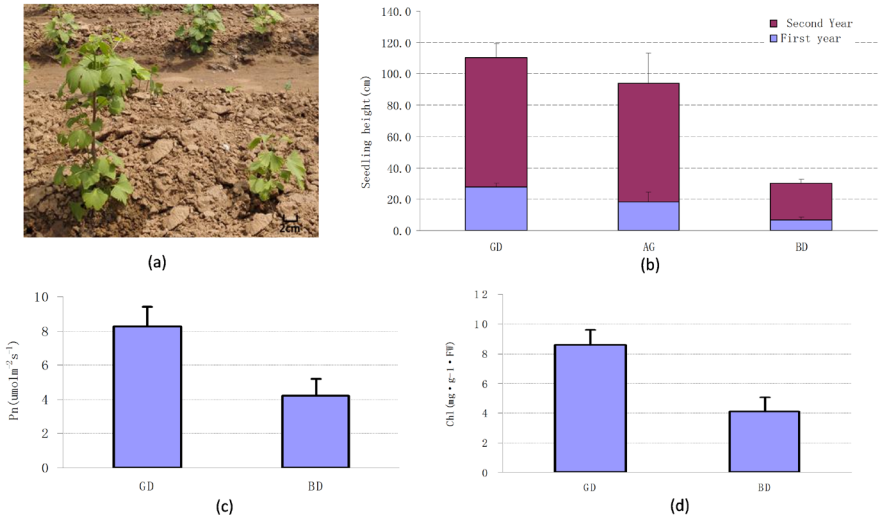

Figure 1. Growth traits of the high growth (GD) and low growth (BD) groups in the *G. biloba* half-sib families. (a) Seedlings in the GD and BD groups; (b) Average height of the seedlings in the GD and BD groups. The average seeding height of the GD group was 27.82 cm in the first year with a net increase of 82.37 cm in the second year; the average seeding height of the BD group was 6.63 cm in the first year with a net increase of 23.43 cm in the second year; AG is the average height of the two groups; (c) Net photosynthetic rate (Pn) in the GD and BD groups; (d) Average chlorophyll content in the GD and BD groups.

Table 1. Growth and physiology statistics of the *G. biloba* half-sib families.

	Height/cm	photosynthetic rate (Pn)/μmol·m^{-2}·s^{-1}	chlorophyll content/mg·g^{-1}·FW
GD	102.37	8.3	8.6
BD	30.16	4.23	4.1
Average	66.27	6.27	6.35

from the GD and BD groups; the Q30 basic group ratios were more that 90% (**Table 2**). The clean reads were assembled using Trinity software [39], and a total of 180,402 single transcripts and 142,492 unigenes are obtained. Of these, 77,069 unigenes (27.11% of the total number) were 300 - 500 bp long, and 18.15% and 15.16% were 500 - 1000 bp and 1000 - 2000 bp long, respectively. The N50 of single transcripts was 1514 bp and the N50 of the unigenes was 1081 bp, indicating that the integrity of the assembly was reasonably high (**Figure 2**).

The assembled unigenes were annotated using Clusters of Orthologous Groups (COG) [40], Eukaryotic Orthologous Groups (KOG) [40], protein family (Pfam) [42], Gene Ontology (GO) [43], Kyoto Encyclopedia of Genes and Genomes (KEGG) [44], SwissProt protein sequence (SwissProt) [45], and the NCBI non-redundant protein sequence (Nr) [46] and nucleotide sequence (Nt) [47] databases. The annotation statistics are listed in **Table 3**. The E-value for the searches against each of the databases was set as ≤1e^{-5}. A total of 41,758 (29.3%) unigenes were annotated in at least one of the databases; the remaining

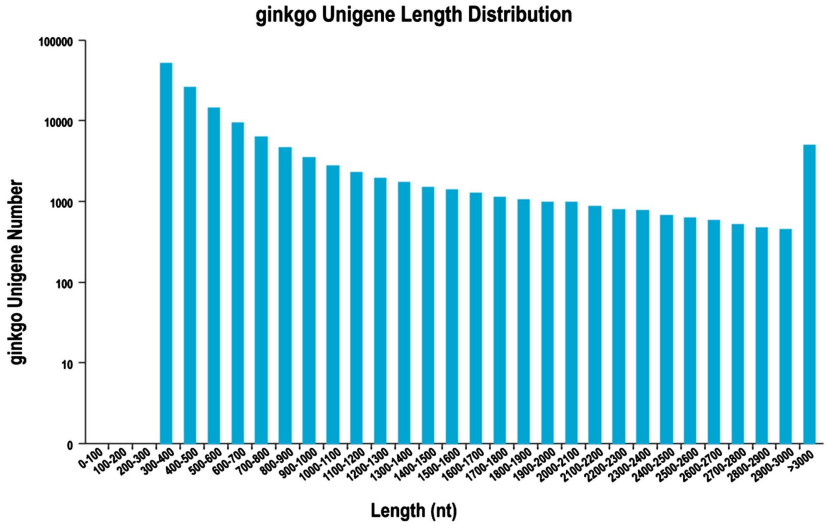

Figure 2. Length distribution of *G. biloba* half-sib families unigenes in the high growth (GD) and low growth (BD) transcriptomes.

Table 2. Statistics of the *G. biloba* half-sib families transcriptome sequencing data.

Samples	Clean reads (bp)	Clean data (bases)	GC content	Percent ≥ Q30
GD	121,909,823	30,477,455,750	44.93%	92.16%
BD	122,947,833	30,736,958,250	45.15%	91.95%

Samples: GD, high growth group, BD, low growth group. Clean reads: total number of pair-end reads in the clean data. Clean data: total number of bases in the clean data. GC content: GC content of the clean reads. Percent ≥ Q30: percentage of clean data with a quality score greater than or equal to 30 (*i.e.*, the probability that base is called incorrectly is 1 in 1000).

Table 3. Annotation statistics of the *G. biloba* half-sib families unigene.

Databases	Unigenes	≥300 bp[a]	≥1000 bp[b]
COG	11,719	4489	7230
KOG	24,483	13,310	11,173
Pfam	21,149	6775	14,374
GO	20,002	8987	11,015
KEGG	5821	1755	4066
SwissProt	22,705	9654	13,051
TrEMBL	38,598	20,023	18,575
Nr	38,991	20,416	18,575
Nt	22,101	7394	14,707

[a]≥300 bp indicates the number of annotated unigenes ≥ 300 bp long. [b]≥1000 bp indicates the number of annotated unigenes ≥ 1000 bp long.

137,734 unigenes (60.7%) were not annotated, indicating that *G. biloba* genetic information is deficient in the existing databases. The Nr database produced the highest number of annotated unigenes (38,991), while KEGG produced the least (5821).

3.3. Expression Analysis of Differentially Expressed Genes of the *G. biloba* Half-Sib Families

False discovery rate (FDR) values were adopted as a key index of differentially expressed genes (DEGs) to reduce false positives that may be caused by independent statistical hypothesis testing of expression values of a large number of genes. FDR values < 0.05 and the differential multiple fold changes (FC) ≥ 2 between two groups were used as the cutoff to identify DEGs. A scatter plot of gene expression levels in the GD and BD groups shows that most of the points fell on the diagonal (**Figure 3(a)**), indicating that the gene expression trends were similar in the two groups and the repetition correlation was high. A volcano plot of the differential gene expression between the BD and GD groups (**Figure 3(b)**) shows that the number of genes with significant −logFDR and FC values was more than the number of genes without significant −logFDR and FC values, indicating that the screening was reliable. A total of 601 DEGs were identified between the BD (control) and GD (test) groups; 400 were up-regulated and 201 were down-regulated. Hierarchical clustering analysis of the DEGs showed that genes with the same or similar expression patterns clustered together (**Figure 3(c)**).

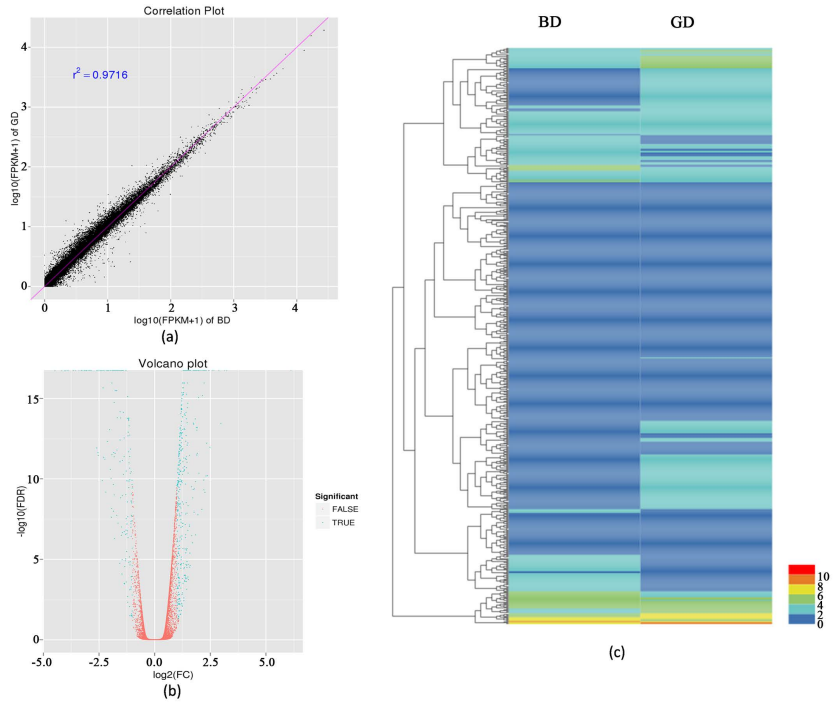

Figure 3. Analysis of gene expression in the high growth (GD) and low growth (BD) transcriptomes. (a) Scatter diagram of gene expression levels in the GD and BD groups. FPKM, fragments per kilobase of transcript per million mapped reads; (b) Volcano plot of differential gene expression between the GD and BD groups. Green indicates genes with significant −log FDR and FC values; red indicates genes without significant −log FDR and FC values. FDR, false discovery rate; FC, fold change; (c) Hierarchical clustering of DEGs with the same or similar expression patterns between the BD (control) and GD (test) groups.

3.4. Functional Annotation of Differentially Expressed Genes of *G. biloba* Half-Sib Families

The second-level GO functional annotation terms assigned to the DEGs and to all the unigenes are shown in Figure 4. Differences between the percentages of genes assigned to the different functions may be related to high growth. Under cellular component, "cell" (73, 6.3%), "cell part" (75, 6.4%), and "organelle" (55, 4.7%) were assigned to the highest number of genes; under molecular function, "catalytic activity" (124, 10.6%) and "binding" (124, 10.6%) were assigned to the highest number of genes; and under biological process, "metabolic process" (169, 14.5%), "cellular process" (126, 10.8%), and "single-organism process" (91, 7.8%) were assigned to the highest number of genes. Among the 25 COG categories (Figure 5), "Replication, recombination and repair" (39, 22.9%) and "General function prediction only" (39, 22.9%) were assigned to the highest number of DEGs, followed by "Posttranslational modification, protein turnover, chaperones" (29, 17.1%) and "Transcription" (28, 16.5%). The categories with the lowest number of DEGs were "Coenzyme transport and metabolism" (1, 0.59%), "Inorganic ion transport and metabolism" (2, 1.2%), and "Chromatin structure and dynamics" (2, 1.2%). None of the DEGs were assigned to "Nuclear structure", "Defense mechanisms", "Intracellular trafficking, secretion, and vesicular transport", or "Nucleotide transport and metabolism".

To explore the biological pathways in which the DEGs may be involved, we performed a KEGG analysis (Figure 6). Many DEGs were assigned to the Spliceosome, Protein processing in endoplasmic reticulum, RNA transport, and

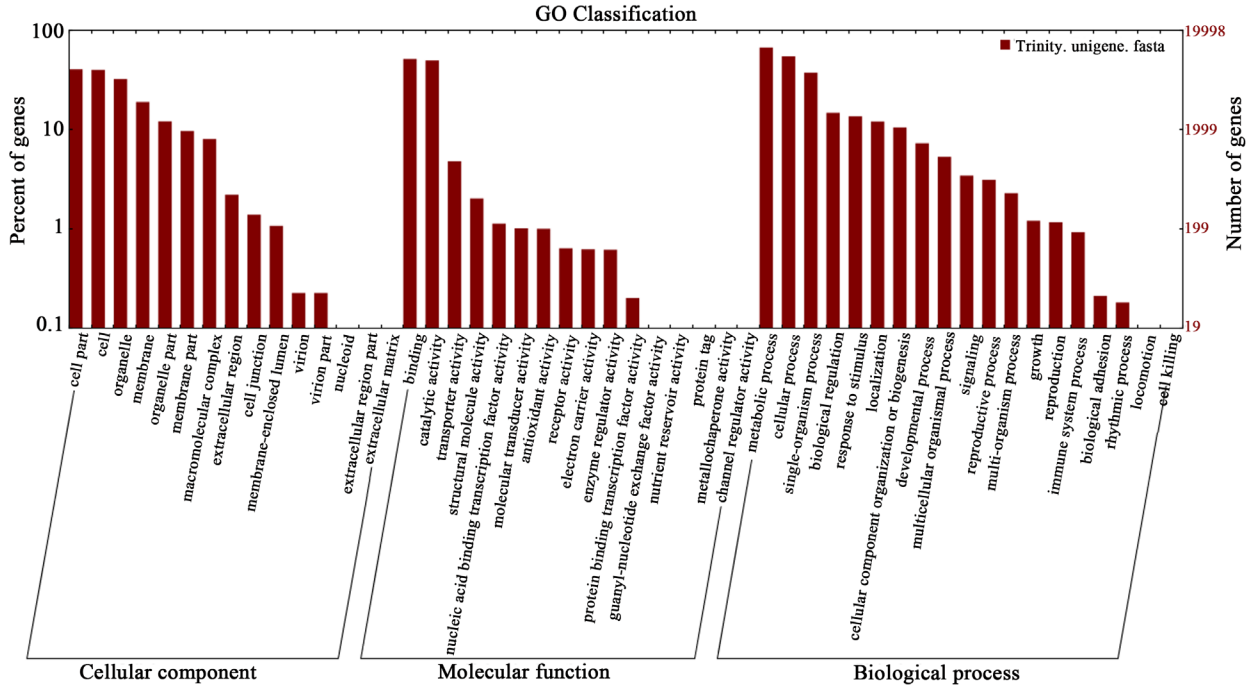

Figure 4. Gene Ontology (GO) terms assigned to differentially expressed genes and all unigenes in the *G. biloba* half-sib families transcriptomes. Second-level terms were assigned under the three GO categories: cellular component, molecular function, and biological process.

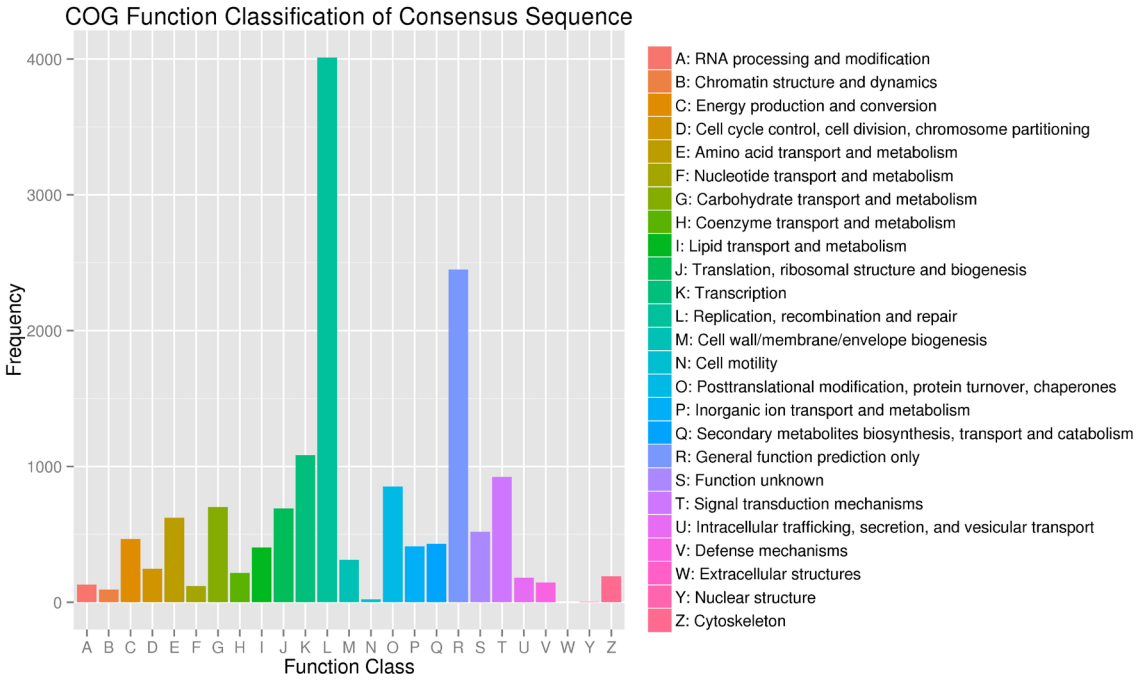

Figure 5. COG annotations assigned to differentially expressed genes in the *G. biloba* half-sib families transcriptomes.

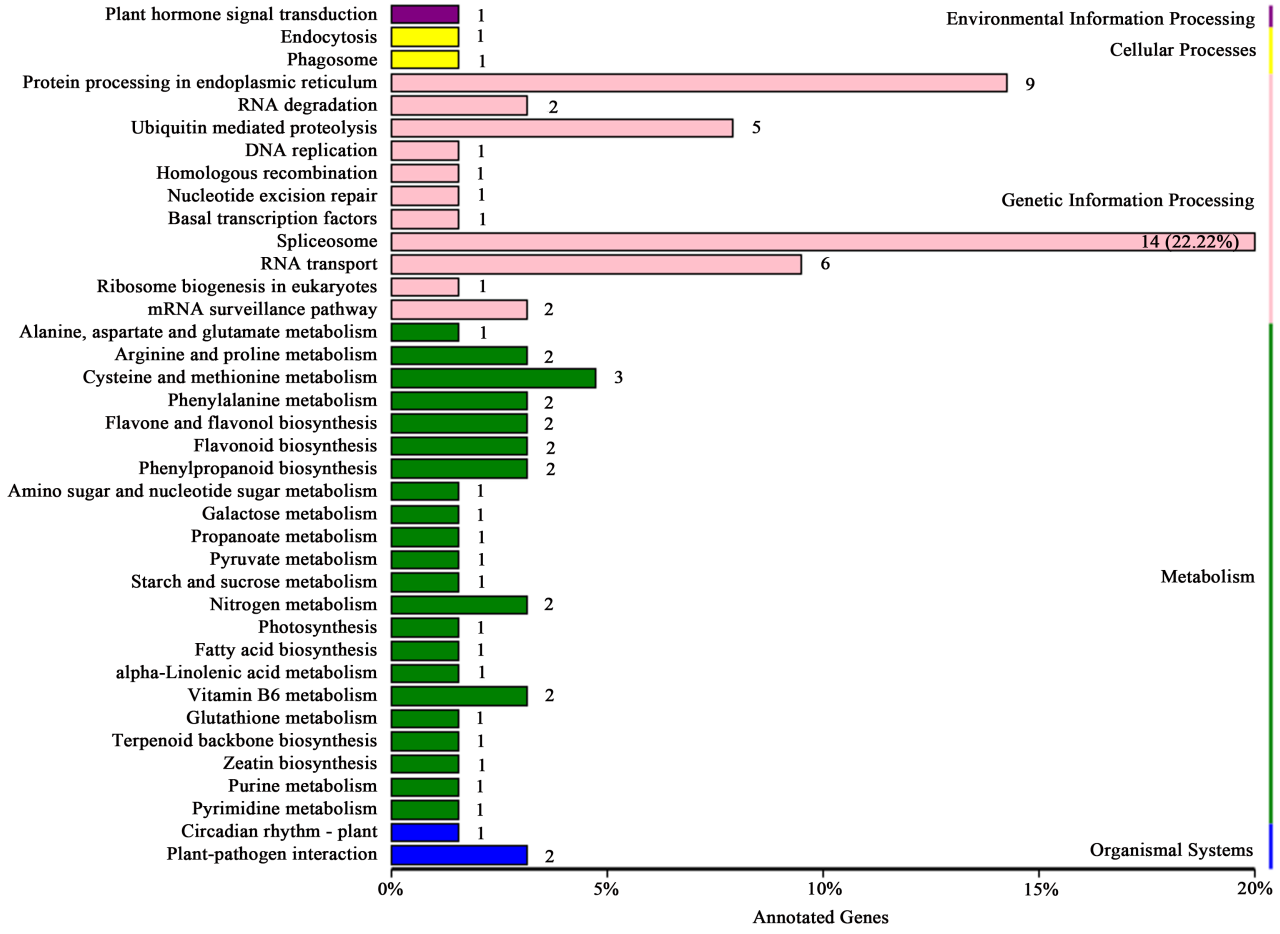

Figure 6. KEGG annotations of differentially expressed genes in the *G. biloba* half-sib families transcriptomes.

Ubiquitin-mediated proteolysis pathways. Splicing factors Prp22, Sm, SF3a, Prp6, P68, S164, Snu66, CDC5, and THOC are known to participate in mRNA splicing and genes encoding them were among the up-regulated genes in the GD group compared with BD group. The protein responsible for endoplasmic reticulum-associated degradation (ERAD) is related to *Hsp*70 and *sHSF*, which were down-regulated in GD compared with BD. Meanwhile, the genes encoding the ubiquitin-conjugating E2 enzyme (UBE20) and the ubiquitin E3 ligase (ARF-BR1) associated with the proteasome were up-regulated in GD compared with BD. Genes encoding the THOC2, Tpr, Nup62, eIF5B, and eIF4G factors, which are involved in RNA transport, were up-regulated in GD compared with BD.

3.5. Relevance of the Predicted SNPs of *G. biloba* Half-Sib Families

The unigene sequences were compared with the known sequences in three protein sequence databases (Nr, SwissProt, and KEGG) and the protein-coding sequence (CDS) information from the matched sequences was used to annotate the unigenes. The CDSs were translated into amino acid sequences according to the standard codon table. The CDSs of unigenes that did not match any of the known protein sequences were predicted using the GetORF software [48], which translates nucleotide sequences in all six reading frames. The longest amino acid sequence for each unigene was selected as the translated protein sequence for that gene. The length distributions of the CDSs and predicted protein sequences of all the unigenes are plotted in Figure 7.

The RNA-Seq reads from each group were compared with the assembled unigenes, to identify candidate single nucleotide polymorphisms (SNPs). A total of 115,517 SNP loci were acquired. After filtering SNPs with a depth of less than 3 (21,776) and undifferentiated loci (67,859) between the CD and BD pools, 25,883 SNP loci remained. SNP loci with ED^5 values (Euclidean distance) higher than the threshold value (set as 1.151) were regarded as outstanding correlative loci (Table S1). The number of genes associated with these SNPs was 119, of

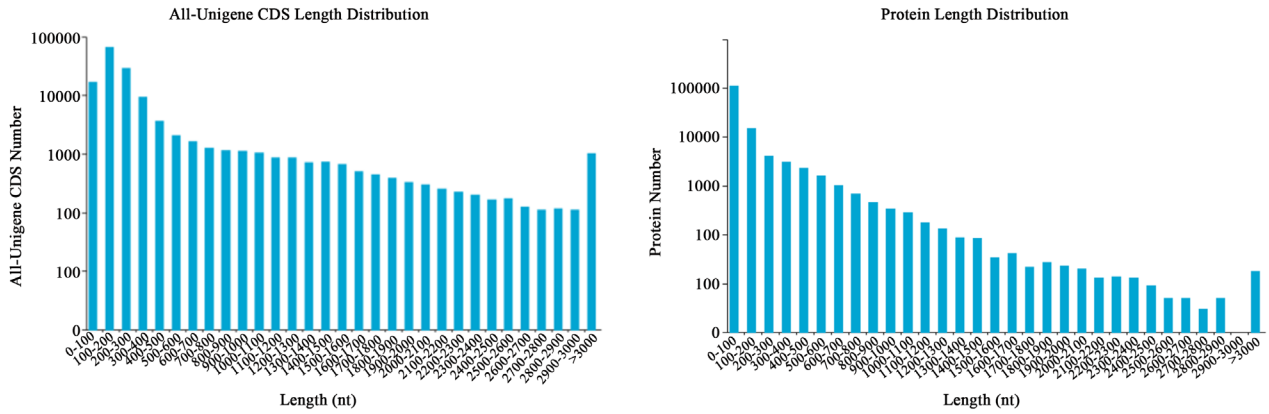

Figure 7. Length distribution of the protein-coding sequences (CDSs) and translated amino acid sequences of all the unigenes in the G. biloba half-sib families transcriptomes.

which 58 were annotated genes (Table S2). Of the 58 annotated genes, 31 had KOG annotations only under General function, Carbohydrate transport and metabolism, and Posttranslational modification, protein turnover and chaperones. Twenty-nine of the 58 genes were assigned GO terms. Under biological process, metabolic process (GO: 0008152), regulation of transcription, DNA-templated (GO: 0006355), and regulation of plant-type hypersensitive response (GO: 0010363) were highly represented; under cellular component, plasma membrane (GO: 0005886) and integral component of membrane (GO: 0016021) were highly represented; and under molecular function, binding (GO: 0005488) and metal ion binding (GO: 0046872) were highly represented (Table S3). The 58 genes were annotated with seven KEGG pathways, together with Protein processing in endoplasmic reticulum and Spliceosome, which were annotated to DEGs, Sphingolipid metabolism, Alanine, aspartate and glutamate metabolism and Carbon sequestration in photosynthetic organisms were also represented.

3.6. Expression of High Growth Trait-Related Genes

DEGs related to high growth in *G. biloba* half-sib families were predicted to participate in photosynthetic carbon sequestration. After photosynthetic carbon sequestration of CO_2, reactive enzyme activation occurs through the glycolysis process. The correlation gene (*c*28693_*g*1_*i*1) has a regulatory effect on the dehydrogenation and phosphorylation of 1,3-2-glyceric-acid phosphate to form glyceraldehyde 3-phosphate. This gene also participates in oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor, and NADP binding activities. It has been shown that improvement of plasmalemma redox activity can promote elongation growth of plants [49] [50]. The Pn and growth rate of group GD were higher than those of group BD, which may be related to the activation of genes involved in the photosynthetic carbon sequestration process of *G. biloba*.

Sphingolipids play major roles in intracellular transduction [51] and participate in many important signal transduction processes, such as adjustment of cellular growth, differentiation, senility, and programmed cell death [52]. Sphingolipid metabolism can be controlled by differential enzyme expression and is cell specific expression, and ceramidase has been implicated in different tissues [53]. Ceramidase activity has been correlated with high growth, which indicates that sphingolipids in the GD group may be related to high growth. In addition, a related enzyme involved in the activity of splicing factor Prp22 and a correlation factor associated with the snRNA component were both up-regulated in the GD group. We speculate that the spliceosome-encoding gene may effectively promote high growth in *G. biloba*.

Endoplasmic reticulum-associated protein degradation eliminates denatured proteins, paraproteins or damaged proteins, plays a major role in controlling the quality of proteins. The KEGG pathway analysis revealed that ERAD was related to the down-regulated DEGs *Hsp*70 and *sHSF*. It has been shown that degradation of ERAD substrate was coupled with the degradation pathway of ubiqui-

tin-proteasomes [54]. The DEGs *E*2 (*UBE*20) and *E*3 (*ARF-BR*1) proteasome that participate in ubiquitin-mediated proteolysis were up-regulated in the GD group. The ERAD system can preferentially degrade specific proteins and effectively protect the immune system, suggesting that it may be related to the high growth of the *G. biloba* seedlings.

3.7. Validation of DEGs Involved in Branching by Quantitative Reverse Transcription Polymerase Chain Reaction

16 DEGs that had large differences in expression between GD and BD were selected for verification by RT-qPCR (**Figure 8**). The expression profiles of these DEGs obtained by RT-qPCR corresponded to the RPKM values obtained from the transcriptome data. c106758-g1-i1, c43830-g1-i2, c88136-g1-i3, c69288-g1-i1 and c92789-g1-i1 were up-regulated in BD compared with GD, whereas c24926-g1-i1, c61897-g1-i1, c72985-g1-i1, c79250-g2-i1, c87276-g5-i1, c90996-g1-i3, c91606-g3-i1, c93631-g1-i1, c72930-g1-i1, c136359-g1-i1 and c82958-g1-i1 were up-regulated in GD compared with BD. The gene expression level trend by RT-qPCR is the same as the transcriptome sequencing data.

4. Discussion

For the BSA, the ED value of each SNP was calculated between the GD and PD RNA pools using the allele depth of the differentially occurring SNP, determine the target site, and conduct linked marker. A total of 119 genes were correlated with the identified SNPs, and 58 of them were assigned functional annotations. In Bulked Segregant Analysis and Amplified Fragment Length Polymorphism (BSA-AFLP) analysis of the resistance gene *rhm* of corn southern leaf blight, more than 222 polymorphic markers were found in a F1-generation infection resistance pool (10 plants in each pool); however, further verification found that in 80 single plants of the F2-generation, 16 of the markers were not linked with the target gene [55]. A similar result has been reported in barley [56]. It indicates that the non-linked marker can also present to polymorphic stripe of two pools. Although these issues cannot be entirely eliminated, they can be reduced by increasing the number of single plants in the mixing pools. In the present study, 30 single plants were used in each mixing pool, which made up 30.9% of the total samples and reduced the number of non-linked markers that were identified. In addition, to ensure the veracity of the gene screening, expression analysis and identification of SNP loci were performed using the RNA-Seq data to detect growth-related genes and lay the foundation for fine mapping of these genes in the *G. biloba* half-sib families.

In most woody plants, heterozygosity is strong and the genomes tend to be large and complex; therefore, studies into the genetic background of these plants have been limited. For species without a reference genome, RNA-Seq data have been used to obtain inheritance information and to build physical maps [57]. *G. biloba* is an ancient gymnosperm that is widely distributed around the world and its ability to growth and adapt to different environmental conditions suggests

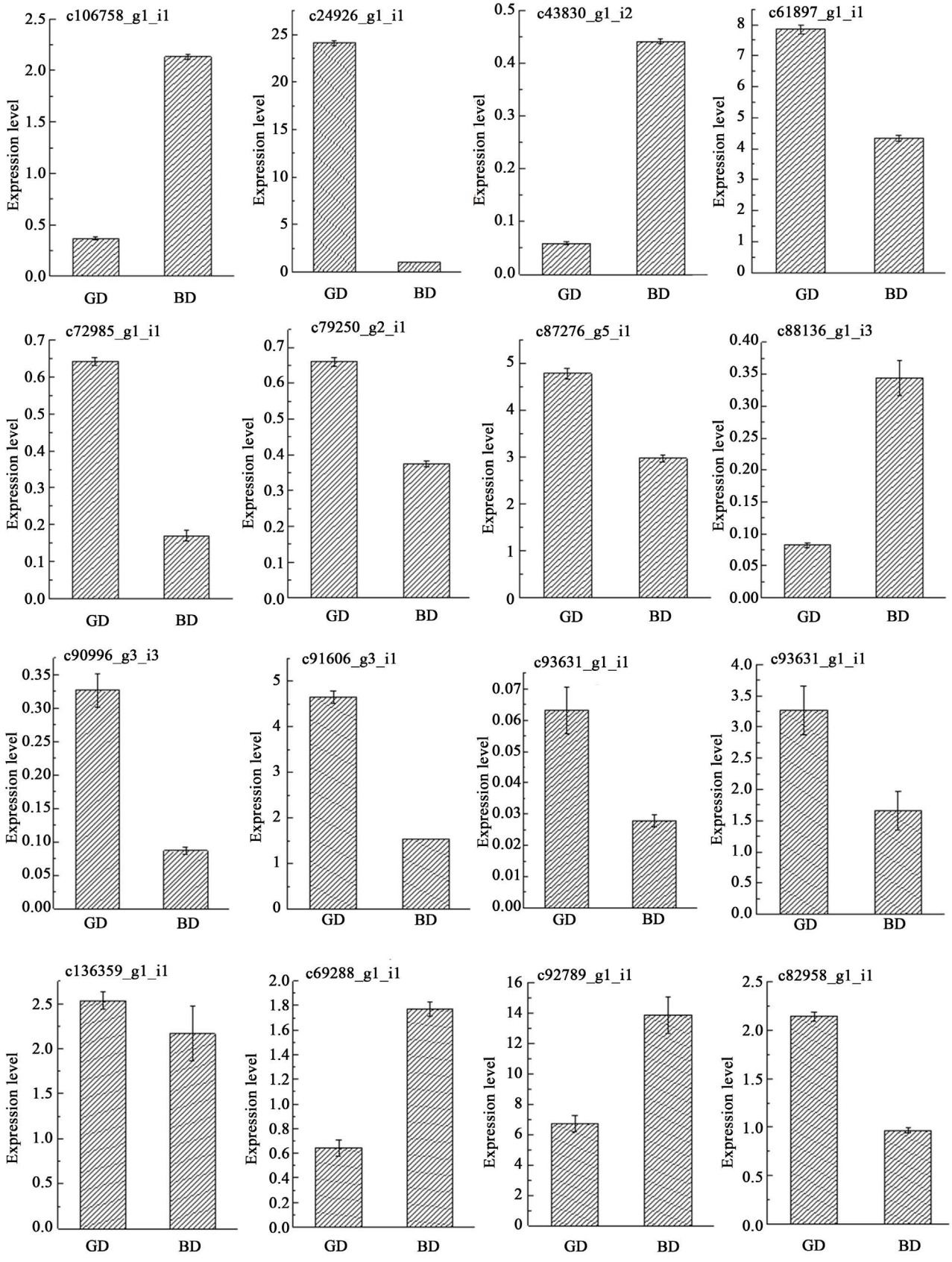

Figure 8. Expression profiles of 16 key DEGs in two groups.

that a large number of responsive genes would have evolved. Many genes and transcription factors related to growth and development of *G. biloba* are available in the related study of *G. biloba* leaves, for example, *COP9* signal corpuscle composite subunits, *AGAMOUS-like MADS-box* transcription factor [58], *glucan endo-l,* 3*-beta-glucosidase* [59], *DELLA*, *ELFB*, *homeobox-leucine zipper protein*, and *EMBRYONICFLOWER*2 [60]. Based on the expression levels of genes in different samples, 601 DEGs have been recognized and functional annotations have been assigned to 513 of them. Among them, two *Homeobox-leucine zipper protein genes* were up-regulated in the GD group compared with the BD group; therefore, these are very likely related to high growth of *G. biloba*. In addition, the DEGs and the gene associated with BSR technology were found to be associated with spliceosome activity, spliceosome metabolism, photosynthetic carbon sequestration, and endoplasmic reticulum protein processing and also to participate in growth and metabolism of *G. biloba*.

5. Conclusion

In this study, BSR technology was used to sequence transcriptomes of high growth (GD) and low growth (BD) groups in the Ginkgo biloba families in Panxian, Guizhou, and 601 DEGs were identified, of which 513 genes were functionally annotated. Among them, two Homeobox-leucine zipper protein genes were up-regulated in the GD group compared with the BD group; SNP analysis was performed on the reads of two samples with high growth and low growth, and BSR was associated with 119 genes, of which 58 were functional annotations.

Acknowledgements

Thank the National Natural Science Foundation of China: Individual Occurrence, Formation Mechanism and Systematic Significance of *Ginkgo Biloa* (31670663), National Key R & D Program: Key Technologies for Efficient Cultivation and Comprehensive Utilization of Ginkgo Biloba in Leaves and Seeds (2017 YFD0600701), National Plant Germplasm Resources Sharing Platform-National Forest Tree Germplasm Resources (including bamboo rattan flower) Platform (platform subsystem): 2013-39 funding this research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Jacobs, B.P. and Browner, W.S. (2000) Ginkgo Biloba: A Living Fossil. *The American Journal of Medicine*, **3**, 341-342. https://doi.org/10.1016/S0002-9343(00)00290-4

[2] Kato-Noguchi, H., Takeshita, S., Kimura, F., Ohno, O. and Suenaga, K. (2013) A Novel Substance with Allelopathic Activity in *Ginkgo biloba*. *Journal of Plant Phy-*

siology, **170**, 1595-1599. https://doi.org/10.1016/j.jplph.2013.07.003

[3] Xing, S.Y. (2014) Ginkgo Germplasm Resources in China. Chinese Forestry Press, Beijing.

[4] Yuan, J., Li, Q., Xiao, G.L. and Zhu, S.H. (2002) Discussion on the Utilization and Development of Ginkgo Wood. *China Forestry Science and Technology*, **16**, 6-8, 23.

[5] Cao, F.L. (2007) Chinese Ginkgo Records. Chinese Forestry Press, Beijing.

[6] Xing, S.Y., Huangpu, G.Y., Hou, J.H. and Guo, C.T. (1993) Study on the Quality of the Fine Single Plant of *Ginkgo biloba*. *Deciduous Fruits*, **4**, 15-18.

[7] Newcomer, E.H. (1954) The Karyotype and Possible Sex Chromosomes of *Ginkgo biloba*. *American Journal of Botany*, **41**, 542-545. https://doi.org/10.1002/j.1537-2197.1954.tb14375.x

[8] Echenard, V., Lefort, F., Calmin, G., Perroulaz, R. and Belhahri, L. (2008) A New and Improved Automated Technology for Early Sex Determination of *Ginkgo biloba*. *Arboriculture & Urban Forestry*, **34**, 300-307.

[9] Zhang, Q., Li, J.H., Sang, Y.L., Xing, S.Y., Wu, Q.K. and Liu, X.J. (2015) Identification and Characterization of Micro-RNAs in *Ginkgo biloba* var. epiphylla Mak. *PLoS ONE*, **10**, e0127184. https://doi.org/10.1371/journal.pone.0127184

[10] Guo, C.L., Chen, L.G., He, X.H., Dai, Z. and Yuan, H.Y. (2005) Expressions of Leafy Homologous Genes in Different Organs and Stages of *Ginkgo biloba*. *Hereditas*, **27**, 241-244.

[11] Liao, L., Liu, J., Dai, Y., Li, Q., Xie, M., Chen, Q., *et al.* (2009) Development and Application of Scar Markers for Sex Identification in the Dioecious Species *Ginkgo biloba* L. *Euphytica*, **169**, 49-55. https://doi.org/10.1007/s10681-009-9913-8

[12] Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., *et al.* (2008) Stem Cell Transcriptome Profiling via Massive-Scale mRNA Sequencing. *Nature Methods*, **5**, 613-619. https://doi.org/10.1038/nmeth.1223

[13] Fu, X., Fu, N., Guo, S., Yan, Z., Xu, Y., Hu, H., *et al.* (2009) Estimating Accuracy of RNA-Seq and Microarrays with Proteomics. *BMC Genomics*, **10**, 161-160. https://doi.org/10.1186/1471-2164-10-161

[14] Tang, F., Barbacioru, C., Wang, Y., Nordman, E. and Lee, C. (2009) MRNA-Seq Whole-Transcriptome Analysis of a Single Cell. *Nature Methods*, **6**, 377-382. https://doi.org/10.1038/nmeth.1315

[15] Wilhelm, B.T. and Landry, J.R. (2009) RNA-Seq-Quantitative Measurement of Expression through Massively Parallel RNA-Sequencing. *Methods*, **48**, 249-257. https://doi.org/10.1016/j.ymeth.2009.03.016

[16] Liu, S., Chen, H.D., Makarevitch, I., Shirmer, R., Emrich, S.J., *et al.* (2010) High Throughput Genetic Mapping of Mutants via Quantitative Single Nucleotide Polymorphism Typing. *Genetics*, **184**, 19-26. https://doi.org/10.1534/genetics.109.107557

[17] Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., *et al.* (2009) Transcriptome Sequencing to Detect Gene Fusions in Cancer. *Nature*, **458**, 97-101. https://doi.org/10.1038/nature07638

[18] Han, S.M., Wu, Z.J., Jin, Y., Yang, W.N. and Shi, H.Z. (2015) RNA-Seq Analysis for Transcriptome Assembly, Gene Identification, and SSR Mining in Ginkgo (*Ginkgo biloba* L.). *The Tree Genetics & Genomes*, **11**, 37. https://doi.org/10.1007/s11295-015-0868-8

[19] He, B., Gu, Y., Xu, M., Wang, J., Cao, F. and Xu, L. (2015) Transcriptome Analysis of *Ginkgo biloba* Kernels. *Frontiers in Plant Science*, **6**, 819.

https://doi.org/10.3389/fpls.2015.00819

[20] Nan, Z. (2013) Sequencing and Analysis of the Transcriptome of *Ginkgo biloba* L. cells. *China Biotechnology*, **33**, 112-119.

[21] Ge, Y.Q., Qiu, Y.X., Ding, B.Y. and Fu, C.X. (2003) An ISSR Analysis on Population Genetic Diversity of the Relict Plant *Ginkgo biloba*. *Chinese Biodiversity*, **11**, 276-287.

[22] Zhang, Y.Y., Ma, C.G., Lin, M.J. and Li, B.H. (2001) Study on One of Genetic Variations for *Ginkgo biloba* in China: The Variation of Breeding Fruit-Stone Characters among and within Population. *Scientia Silvae Sinicae*, **37**, 35-40.

[23] Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 9828-9832. https://doi.org/10.1073/pnas.88.21.9828

[24] Livaja, M., Wang, Y., Wieckhorst, S., Haseneyer, G., Seidel, M., Hahn, V., *et al*. (2013) BSTA: A Targeted Approach Combines Bulked Segregant Analysis with Next-Generation Sequencing and *de Novo* Transcriptome Assembly for SNP Discovery in Sunflower. *BMC Genomics*, **14**, 628. https://doi.org/10.1186/1471-2164-14-628

[25] Bauer, E., Weyen, J., Schiemann, A., Graner, A. and Ordon, F. (1997) Molecular Mapping of Novel Resistance Genes against Barley Mild Mosaic Virus (BaMMV). *Theoretical and Applied Genetics*, **95**, 1263-1269. https://doi.org/10.1007/s001220050691

[26] Steuernagel, B., Taudien, S., Gundlach, H., Seidel, M., Ariyadasa, R., Schulte, D., *et al*. (2009) *De Novo* 454 Sequencing of Barcoded BAC Pools for Comprehensive Gene Survey and Genome Analysis in the Complex Genome of Barley. *BMC Genomics*, **10**, 547. https://doi.org/10.1186/1471-2164-10-547

[27] Mackay, I.J. and Caligari, P.D.S. (2000) Efficiencies of F and Backcross Generations for Bulked Segregant Analysis Using Dominant Markers. *Crop Science*, **40**, 626-630. https://doi.org/10.2135/cropsci2000.403626x

[28] Wolyn, D.J., Borevitz, J.O., Loudet, O., Schwartz, C., Maloof, J., Ecker, J.R., *et al*. (2004) Light-Response Quantitative Trait Loci Identified with Composite Interval and Extreme Array Mapping in Arabidopsis Thaliana. *Genetics*, **167**, 907-917. https://doi.org/10.1534/genetics.103.024810

[29] Duan, Y., Li, W., Wu, W., Pan, R., Zhou, Y., Qi, J., *et al*. (2003) Genetic Analysis and Mapping of Genefzp(t) Controlling Spikelet Differentiation in Rice. *Science in china*, **46**, 328-334.

[30] Tang, H.M., Liu, S., Hillskinner, S., Wu, W., Reed, D., Yeh, C.T., *et al*. (2014) The Maize Brown *midrib*2 (*bm*2) Gene Encodes a Methylenetetrahydrofolate Reductase That Contributes to Lignin Accumulation. *Plant Journal*, **77**, 380-392. https://doi.org/10.1111/tpj.12394

[31] Yang, M., Zhu, L., Pan, C., Xu, L., Liu, Y., Ke, W. and Yang, P. (2015) Transcriptomic Analysis of the Regulation of Rhizome Formation in Temperate and Tropical Lotus (*Nelumbo nucifera*). *Scientific Reports*, **5**, Article No. 13059. https://doi.org/10.1038/srep13059

[32] Rong, L., Li, Q., Li, S., Tang, L. and Wen, J. (2016) *De Novo* Transcriptome Sequencing of *Acer palmatum* and Comprehensive Analysis of Differentially Expressed Genes under Salt Stress in Two Contrasting Genotypes. *Molecular Genetics and Genomics*, **291**, 575-586. https://doi.org/10.1007/s00438-015-1127-2

[33] Altschul, S.F., Madden, T.L., Shaffer, A., Zhang, J.H. and Zhang, Z. (1996) Gapped Blast and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, **25**, 3389-3402. https://doi.org/10.1093/nar/25.17.3389

[34] https://www.broadinstitute.org/gatk/index.php

[35] Hill, J.T., Demarest, B.L., Bisgrove, B.W., Gorsi, B., Su, Y.C. and Yost, H.J. (2013) MMAPPR, Mutation Mapping Analysis Pipeline for Pooled RNA-Seq. *Genome Research*, **23**, 687-697. https://doi.org/10.1101/gr.146936.112

[36] Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biology*, **10**, R25. https://doi.org/10.1186/gb-2009-10-3-r25

[37] Dewey, C.N. and Bo, L. (2011) RSEM, Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics*, **12**, 323. https://doi.org/10.1186/1471-2105-12-323

[38] Anders, S. and Huber, W. (2010) Differential Expression Analysis for Sequence Count Data. *Genome Biology*, **11**, R106. https://doi.org/10.1186/gb-2010-11-10-r106

[39] Grabherr, M.G., Haas, B.J., Yassour, M., *et al.* (2011) Full Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nature Biotechnology*, **29**, 644-652. https://doi.org/10.1038/nbt.1883

[40] Tatusov, R.L. (2000) The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution. *Nucleic Acids Research*, **28**, 33-36. https://doi.org/10.1093/nar/28.1.33

[41] Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., *et al.* (2004) A Comprehensive Evolutionary Classification of Proteins Encoded in Complete Eukaryotic Genomes. *Genome Biology*, **5**, R7. https://doi.org/10.1186/gb-2004-5-2-r7

[42] Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y.., Eddy, S.R., *et al.* (2014) Pfam: The Protein Families Database. *Nucleic Acids Research*, **42**, 222-230. https://doi.org/10.1093/nar/gkt1223

[43] Ashburner, M., Ball, C.A., Blake, J.A., *et al.* (2000) Gene Ontology, Tool for the Unification of Biology. *Nature Genetics*, **25**, 25-29. https://doi.org/10.1038/75556

[44] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG Resource for Deciphering the Genome. *Nucleic Acids Research*, **32**, D277-D280. https://doi.org/10.1093/nar/gkh063

[45] Apweiler, R., Bairoch, A., Wu, C.H., *et al.* (2004) UniProt, the Universal Protein Knowledgebase. *Nucleic Acids Research*, **32**, D115-D119. https://doi.org/10.1093/nar/gkh131

[46] Deng, Y.Y., Li, J.Q., Wu, S.F., Zhu, Y.P., Cai, Y.W. and He, F.C. (2006) Integrated NR Database in Protein Annotation System and Its Localization. *Computer Engineering*, **32**, 71-72.

[47] http://blast.ncbi.nlm.nih.gov/Blast.cgi

[48] http://embossgui.sourceforge.net/demo/getorf.html

[49] Qiu, Z.S. and Stern, B.R.I. (1985) Evidence for Electron Transport across the Plasma Membrane of *Zea mays* Root Cells. *Planta*, **165**, 383-391. https://doi.org/10.1007/BF00392236

[50] Cao, C.L., Li, N.Y., Lu, J.Y. and Lei, J.J. (1997) Roles of Plasma Membrane Redox System in Elongation Growth of Plants. *The Journal of Northwest Agricultural University*, No. 3.

[51] Merrill Jr., A.H. (2002) *De Novo* Sphingolipid Biosynthesis: A Necessary, but Dangerous, Pathway. *Journal of Biological Chemistry*, **277**, 25843-25846. https://doi.org/10.1074/jbc.R200009200

[52] Liu, S.-H. and Gou, P. (2009) Progress in Sphingolipids Research. *Biotechnology*, No. 2.

[53] Riebeling, C., Allegood, J.C., Wang, E., Merrill, A.H. and Futerman, A.H. (2003) Two Mammalian Longevity Assurance Gene (LAG1) Family Members, *trh*1 and *trh*4, Regulate Dihydroceramide Synthesis Using Different Fatty Acyl-CoA Donors. *Journal of Biological Chemistry*, **278**, 43452-43459. https://doi.org/10.1074/jbc.M307104200

[54] Hiller, M.M., Finger, A., Schweiger, M., *et al*. (1996) ER Degradation of a Misfolded Luminal Protein by the Cytosolic Ubiquitin-Proteasome Pathway. *Science*, **273**, 1725-1728. https://doi.org/10.1126/science.273.5282.1725

[55] Cai, H.W., Gao, Z.S., Yuyama, N. and Ogawa, N. (2003) Identification of AFLP Markers Closely Linked to the *rhm* Gene for Resistance to Southern Corn Leaf Blight in Maize by Using Bulked Segregant Analysis. *Molecular Genetics and Genomics*, **269**, 299-303. https://doi.org/10.1007/s00438-003-0837-z

[56] Molnar, S.J., James, L.E. and Kasha, K.J. (2000) Inheritance and RAPD Tagging of Multiple Genes for Resistance to Net Blotch in Barley. *Genome*, **43**, 224-23l. https://doi.org/10.1139/g99-111

[57] Li, X., Chen, G.H., Zhang, W.Y. and Zhang, X. (2010) Genome-Wide Transcriptional Analysis of Maize Endosperm in Response to *ae wx* Double Mutations. *Journal of Genetics and Genomics*, **37**, 749-762. https://doi.org/10.1016/S1673-8527(09)60092-8

[58] Shore, P. and Sharrocks, A.D. (1995) The MADS-Box Family of Transcription Factors. *European Journal of Biochemistry*, **229**, 1-13.

[59] Meirinho, S., Carvalho, M., Dominguez, A. and Choupina, A. (2010) Isolation and Characterization by Asymmetric PCR of the *ENDO*1 Gene for Glucan Endo-1,3-β-D-Glucosidase in *Phytophthora cinnamomi* Associated with the Ink Disease of *Castanea sativa* Mill. *Brazilian Archives of Biology and Technology*, **53**, 513-518. https://doi.org/10.1590/S1516-89132010000300003

[60] Lin, X., Zhang, J., Li, Y., Luo, H., Wu, Q., Sun, C., *et al*. (2011) Functional Genomics of a Living Fossil Tree, *Ginkgo*, Based on Next-Generation Sequencing Technology. *Physiologia Plantarum*, **143**, 207-218. https://doi.org/10.1111/j.1399-3054.2011.01500.x

Supplements

Table S1. SNP depth in the RN.

	#Chr	ED5	Pos	Ref	Alt	GD	Depth	AlleDp	BD	Depth	AlleDp
c20427_g1_i1	c20427_g1_i1	1.853638	378	T	C	Y	5	1.4	T	7	7.0
c28693_g1_i1	c28693_g1_i1	3.340316	268	G	C	S	10	1.9	G	6	6.0
c54675_g1_i1	c54675_g1_i1	1.299287	193	T	C	T	9	9.0	Y	6	2.4
c57187_g1_i1	c57187_g1_i1	1.222874	114	C	G	S	9	5.4	C	25	25.0
c57541_g1_i1	c57541_g1_i1	1.610099	216	C	A	M	9	2.7	C	12	12.0
c58452_g1_i1	c58452_g1_i1	2.617225	1038	G	A	G	7	7.0	R	7	1.6
c62278_g1_i1	c62278_g1_i1	1.853638	3247	G	A	G	6	6.0	R	5	1.4
c63526_g1_i1	c63526_g1_i1	2.273363	2473	A	G	A	8	8.0	R	6	1.5
c65126_g1_i1	c65126_g1_i1	1.853638	1241	G	A	R	5	1.4	G	6	6.0
c65569_g1_i1	c65569_g1_i1	1.853638	427	A	C	M	5	1.4	A	9	9.0
c67312_g2_i1	c67312_g2_i1	1.853638	851	A	G	G	5	1.4	A	9	9.0
c67468_g1_i1	c67468_g1_i1	2.617225	912	C	T	Y	7	1.6	C	7	7.0
c68921_g1_i1	c68921_g1_i1	5.656854	742	G	T	G	6	6.0	T	5	0.5
c70190_g1_i3	c70190_g1_i3	1.853638	1698	G	A	R	5	1.4	G	5	5.0
c70690_g1_i1	c70690_g1_i1	2.273363	1326	G	A	R	6	1.5	G	6	6.0
c71033_g1_i1	c71033_g1_i1	5.656854	817	C	T	Y	435	372.63	C	210	210.0
c71904_g1_i1	c71904_g1_i1	2.273363	1852	C	T	Y	6	1.5	C	5	5.0
c72003_g1_i1	c72003_g1_i1	1.853638	52	A	G	R	5	1.4	A	5	5.0
c73456_g2_i1	c73456_g2_i1	2.883309	968	T	C	Y	13	6.7	T	4	4.0
c74688_g1_i1	c74688_g1_i1	5.656854	70	A	G	A	6	4.0	G	5	0.5
c75300_g1_i1	c75300_g1_i1	5.656854	1148	T	C	Y	3	1.2	T	5	5.0
c75661_g1_i1	c75661_g1_i1	3.048366	769	A	C	C	10	0.10	M	5	3.2
c77124_g1_i1	c77124_g1_i1	1.610099	440	T	G	T	6	6.0	K	9	2.7
c77154_g1_i1	c77154_g1_i1	2.445294	1634	G	T	G	7	7.0	K	7	1.6
c78806_g1_i4	c78806_g1_i4	1.853638	293	C	T	C	7	7.0	Y	5	1.4
c80132_g2_i1	c80132_g2_i1	1.342398	327	G	A	R	8	2.6	G	5	5.0
c81702_g3_i2	c81702_g3_i2	3.139152	1422	A	C	M	9	1.8	A	9	9.0
c82414_g1_i1	c82414_g1_i1	1.191564	809	T	G	T	7	7.0	K	6	1.5
c82991_g1_i2	c82991_g1_i2	1.342398	987	C	T	T	6	0.6	Y	8	6.2
c83122_g1_i1	c83122_g1_i1	5.656854	1492	C	T	C	7	7.0	T	5	0.5
c83320_g1_i1	c83320_g1_i1	2.617225	901	C	A	C	6	6.0	M	7	1.6
c83592_g1_i1	c83592_g1_i1	2.074073	491	T	C	Y	11	2.9	T	5	5.0
c85005_g1_i2	c85005_g1_i2	1.342398	63	A	G	R	8	2.6	A	5	5.0
c86149_g2_i1	c86149_g2_i1	1.853638	481	A	G	R	5	1.4	A	6	6.0
c86611_g10_i1	c86611_g10_i1	2.273363	273	T	C	T	5	5.0	Y	6	1.5

Continued

c86973_g1_i1	c86973_g1_i1	2.273363	309	C	T	Y	6	1.5	C	10	10.0
c87053_g1_i1	c87053_g1_i1	1.342398	2642	A	G	R	18	12.6	A	3	3.0
c87211_g1_i2	c87211_g1_i2	1.309298	1621	T	G	T	24	24.0	K	37	30.7
c87395_g1_i2	c87395_g1_i2	5.656854	334	A	G	A	3	3.0	R	6	1.5
c87571_g1_i4	c87571_g1_i4	1.342398	2501	C	A	M	8	2.6	C	5	5.0
c87833_g1_i1	c87833_g1_i1	1.309009	430	C	T	Y	6	1.5	C	5	5.0
c88255_g1_i1	c88255_g1_i1	1.853638	158	A	C	M	5	1.4	A	7	7.0
c91320_g1_i1	c91320_g1_i1	1.853638	1634	T	C	Y	11	4.7	T	4	4.0
c92209_g2_i2	c92209_g2_i2	3.139152	114	C	T	Y	9	1.8	C	6	6.0
c92475_g3_i2	c92475_g3_i2	5.656854	802	G	A	A	6	0.6	G	6	6.0
c92477_g7_i1	c92477_g7_i1	2.0635	153	G	T	K	5	1.4	G	6	6.0
c92759_g3_i1	c92759_g3_i1	2.273363	443	A	G	A	7	7.0	R	6	1.5
c92783_g10_i1	c92783_g10_i1	1.853638	145	G	A	G	9	9.0	R	5	1.4
c92789_g1_i1	c92789_g1_i1	1.569725	1094	T	C	Y	126	99.27	T	227	210.17
c92867_g1_i3	c92867_g1_i3	1.610099	7856	C	G	S	9	5.4	C	4	4.0
c93001_g11_i1	c93001_g11_i1	2.617225	864	G	A	R	3	1.2	G	6	6.0
c93155_g1_i1	c93155_g1_i1	1.357662	2015	T	C	Y	6	3.3	T	3	3.0
c93654_g1_i2	c93654_g1_i2	2.074073	817	G	C	S	11	2.9	G	5	5.0
c94057_g1_i1	c94057_g1_i1	1.196559	124	T	A	T	19	19.0	W	10	3.5
c94300_g2_i1	c94300_g2_i1	1.853638	333	T	C	T	9	9.0	Y	5	1.4
c94311_g1_i6	c94311_g1_i6	1.853638	253	C	T	C	8	8.0	Y	5	1.4
c94534_g5_i2	c94534_g5_i2	1.853638	122	C	T	Y	5	1.4	C	6	6.0
c94543_g3_i2	c94543_g3_i2	2.617225	265	T	A	T	6	6.0	W	7	1.6

Table S2 shows in https://pan.wps.cn/l/s67polw.

Table S3. Gene ontology annot.

#GO_classify	All Unigene	DEG Unigene
cellular component		
extracellular region	442	1
cell	8044	10
nucleoid	15	0
membrane	3809	8
virion	45	0
cell junction	277	1
extracellular matrix	2	0
membrane-enclosed lumen	214	0
macromolecular complex	1607	2
organelle	6483	7
extracellular region part	15	0

Continued

organelle part	2426	5
virion part	45	0
membrane part	1937	3
cell part	8132	10
molecular function		
protein binding transcription factor activity	40	0
nucleic acid binding transcription factor activity	225	0
catalytic activity	9997	11
receptor activity	127	0
guanyl-nucleotide exchange factor activity	16	0
structural molecule activity	405	1
transporter activity	959	4
binding	10,354	17
electron carrier activity	124	0
antioxidant activity	200	1
channel regulator activity	1	0
metallochaperone activity	3	0
enzyme regulator activity	123	0
protein tag	5	0
nutrient reservoir activity	13	0
molecular transducer activity	202	0
biological process		
reproduction	233	0
cell killing	5	0
immune system process	185	0
metabolic process	13,486	15
cellular process	10,992	11
reproductive process	623	1
biological adhesion	42	0
signaling	688	1
multicellular organismal process	1065	1
developmental process	1459	1
growth	241	0
locomotion	6	0
single-organism process	7533	11
rhythmic process	36	0
response to stimulus	2721	3
localization	2419	5
multi-organism process	457	1
biological regulation	2964	4
cellular component organization or biogenesis	2110	2