Transcriptome analysis of immature xylem in the Chinese fir at different developmental phases

Yunxing Zhang[1,2,3,*], Xiaojiao Han[1,2,*], Jian Sang[1,2], Xuelian He[2], Mingying Liu[1,2], Guirong Qiao[1,2], Renying Zhuo[1,2], Guiping He[2] and Jianjun Hu[1,4]

[1] State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China
[2] Key Laboratory of Tree Breeding of Zhejiang Province, The Research Institute of Subtropical of Forestry, Chinese Academy of Forestry, Hangzhou, Zhejiang, China
[3] Institute of Architectural and Artistic Design, Henan Polytechnic University, Jiaozuo, Henan, China
[4] Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China
[*] These authors contributed equally to this work.

ABSTRACT

Background. Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.] is one of the most important native tree species for timber production in southern China. An understanding of overall fast growing stage, stem growth stage and senescence stage cambium transcriptome variation is lacking. We used transcriptome sequencing to identify the repertoire of genes expressed during development of xylem tissue in Chinese fir, aiming to delineate the molecular mechanisms of wood formation.

Results. We carried out transcriptome sequencing at three different cultivation ages (7Y, 15Y and 21Y) generating 68.71 million reads (13.88 Gbp). A total of 140,486 unigenes with a mean size of 568.64 base pairs (bp) were obtained via *de novo* assembly. Of these, 27,427 unigenes (19.52%) were further annotated by comparison to public protein databases. A total of 5,331 (3.79%) unigenes were mapped into 118 pathways by searching against the Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG). Differentially expressed genes (DEG) analysis identified 3, 16 and 5,899 DEGs from the comparison of 7Y vs. 15Y, 7Y vs. 21Y and 15Y vs. 21Y, respectively, in the immature xylem tissues, including 2,638 significantly up-regulated and 3,280 significantly down-regulated genes. Besides, five NAC transcription factors, 190 MYB transcription factors, and 34 WRKY transcription factors were identified respectively from Chinese fir transcriptome.

Conclusion. Our results revealed the active transcriptional pathways and identified the DEGs at different cultivation phases of Chinese fir wood formation. This transcriptome dataset will aid in understanding and carrying out future studies on the molecular basis of Chinese fir wood formation and contribute to future artificial production and applications.

Subjects Evolutionary Studies, Genomics, Molecular Biology, Plant Science, Soil Science
Keywords Transcriptome, Chinese fir, RNA-Seq, Wood formation, Xylem

Submitted 12 February 2016
Accepted 10 May 2016
Published 7 June 2016

Corresponding authors
Renying Zhuo, zhuory@gmail.com
Guiping He, guipinghe@126.com

Academic editor
Yong Wang

Additional Information and Declarations can be found on page 16

DOI 10.7717/peerj.2097

Copyright
2016 Zhang et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article Zhang et al. (2016), Transcriptome analysis of immature xylem in the Chinese fir at different developmental phases. PeerJ 4:e2097; DOI 10.7717/peerj.2097

INTRODUCTION

Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.], a fast growing evergreen coniferous tree ($2n = 2x = 22$), is one of the most important native tree species for timber production in southern China and is also distributed in Vietnam. It is the third most commonly planted tree species in plantations worldwide (*Del Lungo, Ball & Carle*, *2006*). Due to its high value in terms of adaptability, growth rate, timber quality, versatility and commercial value, the planting area of Chinese fir in China is around 9.215 million ha, accounting 28.54% of all forested land (*Lei*, *2005*; *Shi, Zhen & Zheng*, *2010*; *Huang et al.*, *2012*) and for 20–30% of the total commercial timber production in China (*Li & Ritchie*, *1999*; *Orwa et al.*, *2013*). Chinese fir growth and development can be divided into three phases including fast growing stage, stem growth stage and senescence stage (*Duan et al.*, *2004*).

Wood formation involves various division and differentiation activities of cambium cells, including vascular cambium activation, secondary xylem differentiation, cell expansion, secondary wall deposition, programmed cell death, and heartwood formation (*Ye & Zhong*, *2015*). Significant progress has been made in the past decade in uncovering the molecular players involved in the developmental phases of wood formation in tree species. *Populus trichocarpa* is the first sequencing tree (*Tyler*, *2006*), creates opportunities for investigation of secondary growth, and secondary xylem (wood) development in woody plants (*Brunner, Busov & Strauss*, *2004*; *Cronk*, *2005*; *Jansson & Douglas*, *2007*). Transcriptome analyses revealed that the suite of genes, highly expressed in wood-forming cells, includes receptor kinase, transcription factors, and secondary wall biosynthesis genes.

In transcriptional network, secondary wall NAC, MYB and WRKY transcription factors act as the top-level and second-level master switches, respectively (*Zhong & Ye*, *2014*). These findings represent an important step toward elucidating the molecular mechanisms controlling wood formation. To date, genome sequences have been released for four tree species (*Ye & Zhong*, *2015*), including the angiosperms *P. trichocarpa* and *Eucalyptus grandis* (*Myburg et al.*, *2014*), and the gymnosperms *Picea abies* (*Nystedt et al.*, *2013*) and *P. glauca* (*Birol et al.*, *2013*). Furthermore, transcriptome sequences have been obtained, for *Acacia auriculiformis* (*Wong, Cannon & Wickneswari*, *2011*), *E. camaldulensis* (*Thumma, Sharma & Southerton*, *2012*), *Fraxinus spp.* (*Bai et al.*, *2011*), *C. lanceolat*a (*Huang et al.*, *2012*), *Larix leptolepis* (*Zhang et al.*, *2012*), *Populus simonii* × *Populus nigra via* (*Chen et al.*, *2012*), *P. spp.* (*Raherison et al.*, *2012*), *Pinus monticola* (*Liu, Sturrock & Benton*, *2013*), *P. glauca* (*Raherison et al.*, *2015*). The availability of these genome and transcriptome sequences together with an improvement of the methodologies used for generation of transgenic trees will enable researchers to directly employ tree species as models for studying wood formation.

Studies of Chinese fir have mainly focused on anatomical, biochemical, cytological, physiology, and ecological aspects, with only few reports relating to molecular mechanism of wood formation. Due to the limited genomic sources of Chinese fir, it is important to explore transcriptome for further molecular improvement. RNA sequencing (RNA-seq) provide a revolutionary tool with numerous applications for high-throughput functional genomics research (*Mardis*, *2008*; *Schuster*, *2008*; *Morozova, Hirst & Marra*, *2009*;

Nagalakshmi, Waern & Snyder, *2010*). It has accelerated the investigation of the complexity of gene transcription patterns, functional analyses and gene regulation networks in plants (*Wang et al.*, *2010b*).

Transcriptome analyses (*Huang et al.*, *2012*; *Qiu et al.*, *2013*; *Wang et al.*, *2013a*) have identified a number of candidate genes and transcription factors correlated with changes in wood formation in Chinese fir. In the previous studies, however, the samples were collected from the same year or during the same cultivation phase.

In the present work, we used RNA sequencing (RNA-seq) technology to characterize the transcriptome at different stages of Chinese fir growth and development. The RNA samples from three different growth and development phases were sequenced with the high-throughput Illumina deep sequencing technique. Based on the bioinformatics analysis of assembled transcriptome data, we characterized immature xylem transcriptional pathways during the different cultivation phases of Chinese fir. Furthermore, we identified the DEGs subject to regulation during xylem development. The transcriptome sequencing of Chinese fir immature xylem may help to discover new genes and pathways. The data will promote future genetic and genomics studies on the molecular mechanisms of wood formation, and contribute to future applications, including artificial wood production.

MATERIALS AND METHODS

Plant materials

The samples of Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.] were collected from three different sites in Kaihua Country Forest Farm (29°08′33.56N, 118°23′56.59E), Zhejiang Province. No specific permits were required from the Forest Farm to select samples. The Forest Farm is not privately-owned and the field studies did not involve protected species. The immature xylem tissues were collected from three trees at every three different cultivation phases (7 years, 15 years and 21 years of cultivation (7Y, 15Y and 21Y)). The each phase, samples (outer glutinous 1–1.2 mm layer comprising early developing xylem tissue) were harvested from approximately breast height (1.0–1.20 m) on the main stem after removal of the bark using razor blades as described by Huang and Eshchar (*Mizrachi et al.*, *2010*; *Huang et al.*, *2012*). All of the tissue samples were immediately frozen in liquid nitrogen and stored at −80 °C for future use.

RNA extraction, library construction and RNA-seq

Experimental procedures including sample preparation and sequencing were performed following the standard protocols (Illumina, Inc.). Total RNA was extracted separately from each sample using the R6827-01 Plant RNA Kit (Guduo, Shanghai, China). Three biological replicates were performed at three different cultivation phases. The concentration of RNA was analyzed using a spectrophotometer (UV-Vis Spectrophotometer, Quawell Q5000; Quawell, San Jose, CA, USA), and the integrity of RNA was evaluated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Equal quantities of high-quality RNA from each sample were combined into a single large pool for cDNA synthesis.

The mRNA-seq library was constructed using Illumina's TruSeq RNA Sample Preparation Kit (Illumina Inc, San Diego, CA, USA). The mRNA isolation, fragment

interruption, cDNA synthesis, adapter ligation, PCR amplification and RNASeq were performed at Beijing BioMarker Technologies (Beijing, China). The poly-A mRNA was enriched using oligo (dT) magnetic beads, and the mRNA was broken into fragments by fragmentation buffer. The cleaved RNA fragments were transcribed into first-strand cDNA using random hexamer primers, followed by second strand cDNA synthesis using DNA polymerase I and RNase H. The short fragments were purified with the QiaQuick PCR Purification Kit (Qiagen) and eluted in EB buffer for end-repaired by addition of poly(A) to 3′. Then the suitable fragments were separated by an agarose gel electrophoresis and selected for PCR amplification as sequencing templates. The constructed mRNA-seq library was sequenced on the Illumina HiSeq^{TM} 2500 sequencing platform.

Sequence data analysis and assembly

To obtain high-quality clean data for *de novo* assembly, the raw reads were filtered by removing the adapter sequences, low quality sequences (reads with ambiguous bases 'N'), and reads in which more than 20% of bases had a *Q*-value <30. Reads were assembled using the reference transcriptome sequence of Chinese fir using the Bowtie and RSEM packages (*Grabherr et al.*, *2011*). The clean reads were assembled into contigs using Trinity (http://trinityrnaseq.sourceforge.net/) (*Grabherr et al.*, *2011*). After Trinity *de novo* assembly and correction, the contigs without any gaps were linked into transcripts according to the paired-end information of the sequences. Related contigs were clustered into transcripts based on nucleotide sequence identity. The longest transcripts were regarded as unigenes redundancies were removed. Finally, the unigenes were combined to produce the final assembly used for annotation. The unigenes expression abundance was represented in reads per kilobase of exon model per million mapped reads (RPKM). The RPKM measure of read density reflects the molar concentration of a transcript for RNA length and for the total read number in the measurement.

Functional annotation

To determine the functional annotation of the unigenes, the assembled sequences were compared against the NCBI Nr database (*Deng et al.*, *2006*), SwissProt (*Apweiler et al.*, *2004*), GO (*Ashburner et al.*, *2000*), COG (*Tatusov et al.*, *2000*), and KEGG (*Kanehisa et al.*, *2004*) with an *E*-value $\leq 10^{-5}$. Gene names were assigned based on the best BLAST hit (*Altschul et al.*, *1997*). Open reading frames (ORFs) were predicted using the "GetORF" program (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). The longest ORF extracted from each unigene was defined as coding sequence (CDS), and the CDSs were translated into amino sequences using the standard codon table. The Blast2GO program was applied to obtain GO annotation of unigenes with an *E*-value $\leq 10^{-5}$ including molecular function, biological process, and cellular component categories. The unigenes sequences were aligned to COG database to classify and predict possible functions. Annotations of Chinese fir unigenes were used to predict biochemical pathways using the pathways tools. The KEGG database was used to analyze gene products related to metabolism and gene function in cellular processes.

Detection of candidate SSR markers

The assembled sequences longer than 1 kb were used for the detection of SSR markers. Potential SSR markers were detected among the 17,902 unigenes using MISA software (http://pgrc.ipk-gatersleben.de/misa/). The parameters were set for the identification of perfect dinucleotide motifs with a minimum of six repeats, and tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of five repeats (*Zeng et al.*, *2010*; *Wei et al.*, *2011*).

Identification of differentially expressed genes

DESeq was performed to detect the genes which were differentially expressed, based on a threshold false discovery rate (FDR) <0.01 and an absolute value log2ratio ≥2. If the FDR ($Q = V/R$) is required to remain below a cutoff (e.g., 0.01), then the FDR can be calculated according to the Benjamini and Hochberg algorithm as: FDR $= E(Q) = E\{V/(V+S)\} = E(V/R)$ (*Benjamini & Hochberg*, *1995*). All of the DEGs were used for the Nr, Swissport, GO, KEGG and COG Functional annotation analyses.

Sequence retrieval of transcription factors related to NAC, MYB and WRKY

Hidden Markov Model (HMM) was employed factors. The profiles of the NAC, MYB and WRKY DNA-binding domain PF01849, PF00249 and PF03106 used for the HMM search (HMMER 3.1, http://hmmer.janelia.org/) were downloaded from the Pfam database (http://pfam.sanger.ac.uk/), respectively. There were 5 NAC transcription factors, 190 MYB transcription factors and 34 WRKY transcription factors were obtained with an E-value threshold of 0.1. Ultimately, the expression levels of these transcription factors were identified according to the results of DEGs.

Sequence alignments and phylogenetic constructions of transcription factors

Alignment of the amino acid sequences of the NAC, MYB, WRKY transcription factor domains were aligned with Clustal X using the default parameters. For the phylogenetic analysis, the neighbor-joining trees was constructed by MEGA6.0. Bootstrap values obtained after 1,000 replications are indicated on the branches.

RESULTS

RNA-Seq and de novo transcriptome assembly

To obtain a global overview of the Chinese fir transcriptome at different developmental phases, nine RNA samples from immature xylem at three different cultivation stages (fast growing stage, stem growth stage and senescence stage) were sequenced with Illumina HiSeq^{TM} 2500. After stringent quality assessment and data filtering, a total of 68.71 million reads and 13.88 gigabase pairs (Gbp) were generated (Table 1). The reads with base quality greater than 30 ($Q \geq 30$) and no ambiguous "N" were defined as high-quality reads.

Reads were mapped against the reference transcriptome sequences of Chinese fir. After the removal of adaptor sequences and exclusion of contaminated or short reads, 33,966,473 high-quality reads were assembled into 6,590,556 contigs (https://figshare.com/s/fdf6af8b8ae6aa02bd52) using SOAPdenovo (*Li et al.*, *2010*). Using the

Table 1 Summary of Illumina transcriptome sequencing for Chinese fir.

Sample	Total reads	Total bases	GC (%)	Q30%
Chinese fir	68 719 634	13 879 230 822	43.24	96.42

Table 2 Length distribution of assembled contigs, transcripts and unigenes.

Nucleotide length(bp)	Contigs	Transcripts	Unigenes
0–300	6,508,442	87,745	69,385
300–500	42,221	51,541	35,452
500–1,000	21,099	36,109	17,747
1,000–2,000	11,681	31,744	10,593
2,000+	7,113	24,999	7,309
Total number	6,590,556	232,138	140,486
Total length	404,940,125	190,238,773	79,886,333
N50 length	64	1,635	882
Mean length	61.44	819.51	568.64

Trinity *de novo* assembly program, next-generation short-read sequences were assembled into 232,138 transcripts with mean length of 819.51 base pairs (bp) and N50 length of 1,635 bp. The transcripts were subjected to cluster and assembly analyses. Finally 140,486 unigenes with a mean size of 568.64 bp were obtained, these included 17,902 unigenes (12.74%) with length greater than 1 kb. An overview of the contigs, transcripts and unigenes is shown in Table 2.

The length distributions of contigs, transcripts and unigenes were shown in Table 2. As expected for a randomly fragmented transcriptome, there was a positive relationship between the length of a given unigene and the number of reads assembled into it (Fig. 1). Open Reading Frame (ORF) prediction analysis performed with GetORF (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html) identified 71,044 unigenes (50.57%) as having ORFs starting with an 'ATG' codon. The raw reads of Chinese fir produced in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (accession number: SRS959453).

Mapped read depth (reads per kilo base per million reads (RPKM)) was used as a metric for the expression of each unigenes. The expression of the unigenes varied similarly with sequencing depth (Table S1). The expression of unigenes ranged from 0 to 7,443.78 RPKM with an average of 9.32 RPKM. Unigenes with low RPKM values were removed, because they may not have been reliable due to low abundance or statistical errors. Of 71,102 unigenes remaining, 60,631 (85.27%) had a very low expression level of less than 10 RPKM. Unigenes with high RPKM values included those related to metabolism, cell wall biogenesis and remodeling, signal transduction and stress, such as laccase, poly-ubiquitin, ARF-L1 protein, thaumatin-like protein.

Functional annotation and classification

The reads of Chinese fir in different cultivation phases (7Y, 15Y and 21Y) were assembled, and several complementary approaches were utilized to annotate the assembled sequences

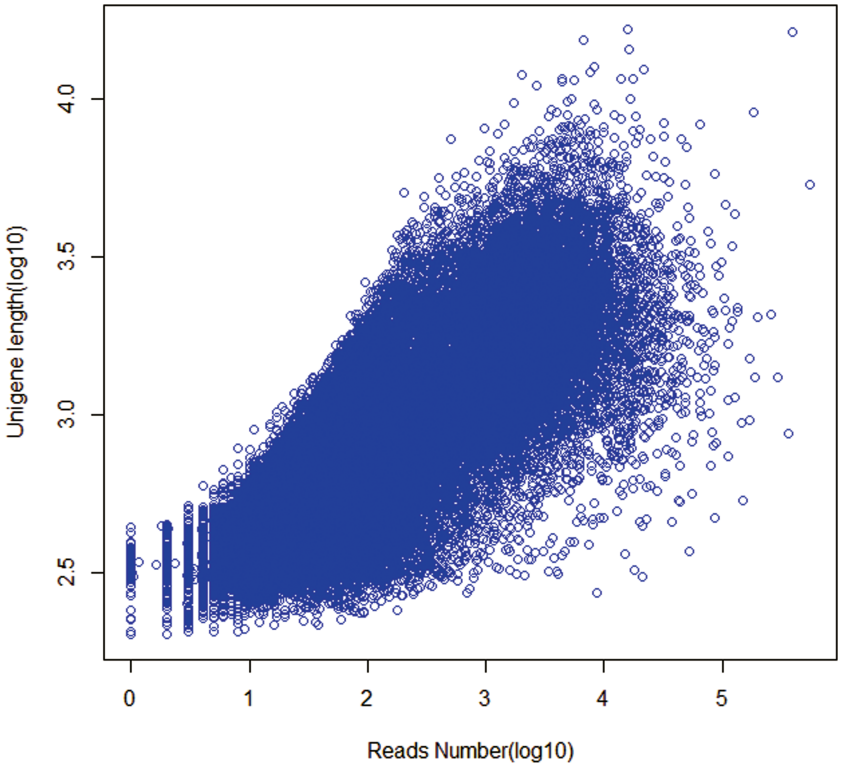

Figure 1 **Dependence of unigene lengths on the number of reads assembled into that unigenes.**

Table 3 **Functional annotation of Chinese fir unigenes.**

Annotated databases	Unigenes	≥300 bp	≥1,000 bp	Percentage of unigenes
Nr-annotation	26,305	12,136	14,169	18.72%
COG-annotation	9,942	3,797	6,145	7.08%
GO-annotation	15,085	6,206	8,879	10.74%
KEGG-annotation	5,331	1,854	3,477	3.79%
Swissprot-annotation	17,407	7,276	10,131	12.39%
Total	27,427	12,899	14,528	19.52%

(Table 3). Only 19.52% of the unigenes (27,427) were able to be annotated based on aligning with sequences deposited in diverse protein databases, including the National Center for Biotechnology Information (NCBI) nonredundant protein (Nr) database, Cluster of Orthologous Groups of proteins (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and UniProt/Swiss-Prot. According to the BLASTX results, 26,305 (18.72%) unigenes had homologous proteins in the Nr protein database. We found that for 29% of the unigenes the most similar proteins sequence was from *P. sitchensis*, whereas 13% were most similar to sequences from *Vitis vinifera*, and 5% to *Theobroma cacao* (Fig. 2).

Gene Ontology (GO) analysis was used for functional classification of the assembled transcripts and gene products in terms of their likely associated biological processes, cellular components, and molecular functions. There were 26,305 unigenes annotated in the Nr

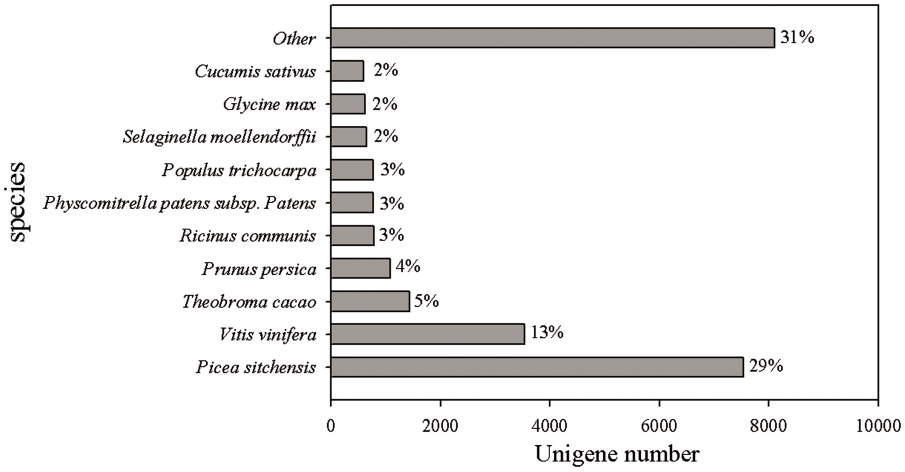

Figure 2 Species distribution of the top BLAST hits in Nr dababase. Top BLASTX results for 26,305 unigenes were calculated. Species with proportions of more than 1% are shown.

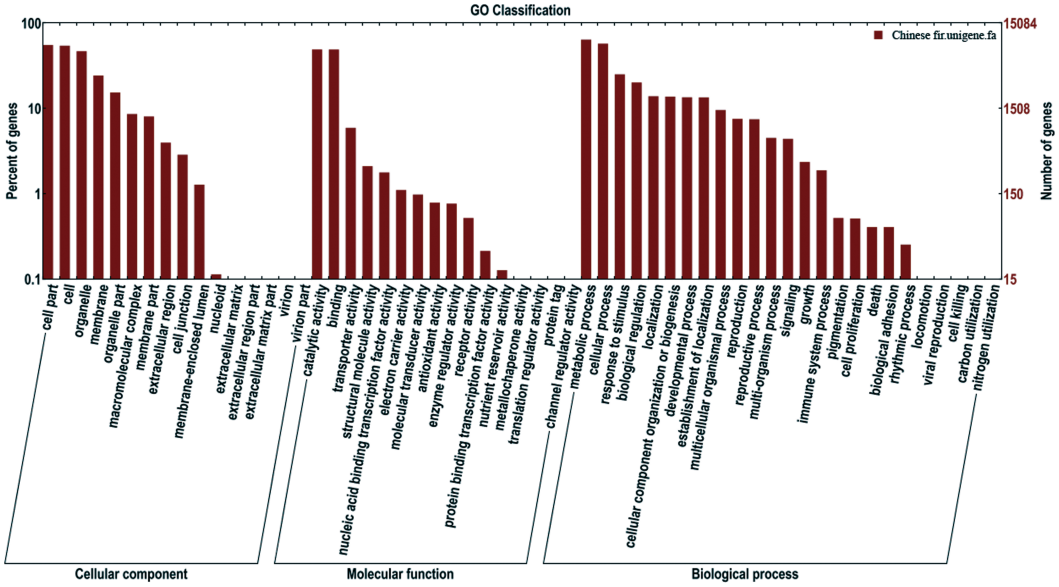

Figure 3 Functional annotation of assembled sequences based on gene ontology (GO) categorization. GO analysis was summarized into three categories: cellular component, molecular function and biological process.

database, among which 15,085 unigenes were assigned one or more GO terms, with 37.2% in cellular components, 18.9% in molecular functions, and 43.9% in biological processes (Fig. 3). To better review GO cellular components, the GO terms were further clustered to their parent terms. For biological processes, genes involved in metabolic processes, cellular processes, and response to stimulus were highly represented. For molecular functions, catalytic activity, binding, and transporter activity were the most highly represented. For cellular components, the three top classifications were cell part, cell and organelle.

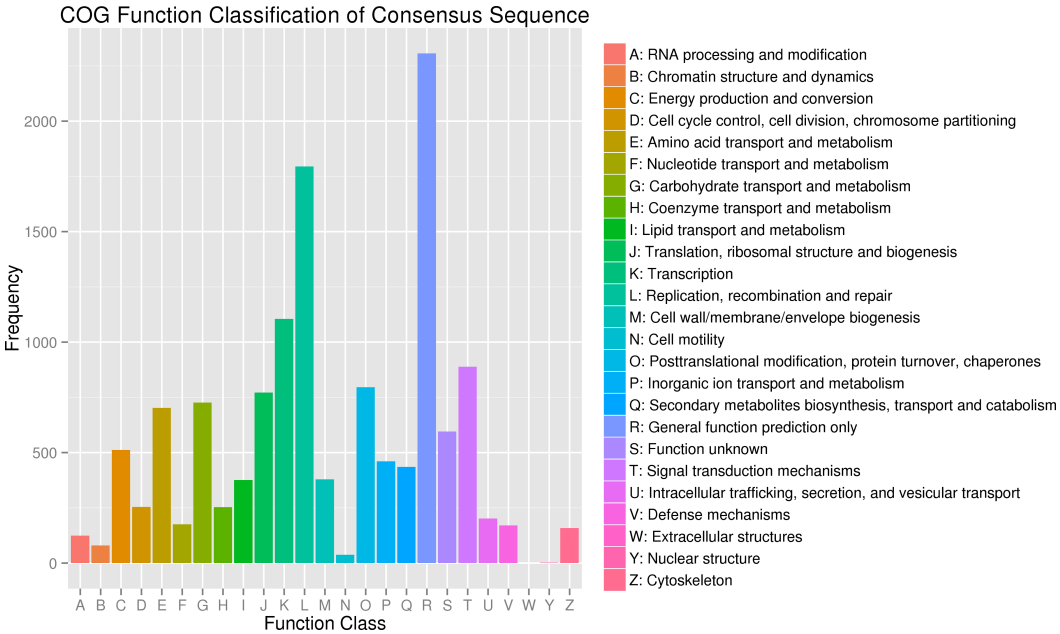

Figure 4 **Clusters of orthologous group (COG) classification.**

In addition, all unigenes were aligned to the COG database for further functional prediction and classification. Overall, 9,942 of the 140,486 sequences were assigned to 24 COG categories, including RNA processing and modification, chromatin structure and dynamics, energy production and conversion, cell cycle control, cell division, and chromosome partitioning (Fig. 4). The category of general function prediction only represented the largest group (2,306; 17.32%), followed by replication, recombination and repair (1,795; 13.48%), transcription (1,105; 8.30%). Only a few unigenes were assigned to chromatin structure and dynamics, cell motility and nuclear structure (80, 38 and 2 unigenes, respectively). Furthermore, 379 unigenes were assigned to cell wall/membrane/envelope biogenesis and 158 unigenes were assigned to cytoskeleton. No unigene was assigned to extracellular structures.

KEGG is a public database for networks of molecular interactions in cells and their variants specific to particular organisms. To further examine the usefulness of the Chinese fir unigenes generated in the present study, the unigenes were compared with the KEGG database using BLASTX and the corresponding pathways were established. Only 5,331 (3.79%) unigenes were assigned 118 pathways (Table S2). The pathways with highest unigene representation were Ribosome (ko03010, 213 unigenes, 3.69%), followed by RNA transport (ko03013, 195 unigenes, 3.38%) and Spliceosome (ko03040, 173 unigenes, 3.00%).

SSR marker discovery

SSRs can be used as powerful molecular markers for genetics, evolution and breeding studies. To explore SSR profiles in the unigenes of Chinese fir, the 17,902 unigene sequences were searched for SSRs. In total, 2,784 sequences containing 3,267 SSRs were obtained,

Table 4 Frequency of candidate SSRs in Chinese fir.

Motif	Repeat number											Total	%
	5	6	7	8	9	10	11	12	13	14	>14		
Di	–	210	73	41	39	15	21	1	0	1	0	401	31.53
Tri	518	196	94	17	1	1	0	0	0	0	0	827	65.02
Tetra	–	23	3	0	1	0	0	0	0	0	0	27	2.12
Penta	–	5	1	0	0	0	0	0	0	0	0	6	0.47
Hexa	–	3	4	3	1	0	0	0	0	0	0	11	0.86
Total	518	437	175	61	42	16	21	1	0	1	0	1,272	100
%	40.72	34.36	13.76	4.80	3.30	1.26	1.65	0.08	0	0.08	0	100	

Table 5 Number of up- and down-regulated DEGs in xylem of Chinese fir at different ages.

Comparison	Number of DEGs	Up	Down
7Y vs. 15Y	3	1	2
7Y vs. 21Y	16	14	2
15Y vs. 21Y	5,899	2,623	3,276

with 394 unigene sequences containing more than one SSR. Tri-nucleotide repeat motifs (65.02%) were the most abundant, followed by Di-nucleotide repeats (31.53%) (Table 4). The most abundant repeat type was AAG/CTT (232, 18.24%), followed by AG/CT (166, 13.05%), and AT/AT (161, 12.66%).

Identification of differentially expressed genes

A total of 140,486 unigenes were detected from the clean reads of all three samples as described above. To detect DEGs between the samples harvested from tress at different stages, DESeq was used with the criteria FDR $\leq$ 0.01 and log2Ratio $\geq$ 2. Whereas only a few DEGs were identified for the comparisons of the samples from 7-year-old trees, by far the most, DEGs were identified from the 15Y vs. 21Y comparisons (Table 5). The representative genes of up- and down-regulated DEGs in xylem of Chinese fir at different phases were shown in Fig. 5 and Table S3. For the 7Y vs. 21Y comparison, most DEGs were up-regulated. Whereas the 15Y vs. 21Y comparison produced a roughly similar number of up-regulated DEGs and down-regulated DEGs.

The up- and down-regulated DEGs were further analyzed based on GO component, GO function and GO process (Tables 6–8). The main GO component categories were cell part, cell, and organelle. In GO function ontology, the major classifications for the DEGs were catalytic activity, binding, and transporter activity. Most of the DEGs were classified into GO process categories of metabolic process, cellular process, and single-organism process. These results indicate that most of the DEGs were related to metabolism, cell wall biogenesis and remodeling, signal transduction and stress.

Transcription factors of interest

Recent studies have demonstrated that many transcription factors, such as NAC, MYB, and WRKY gene families, regulate the formation of secondary wall (*Rogers & Campbell*,

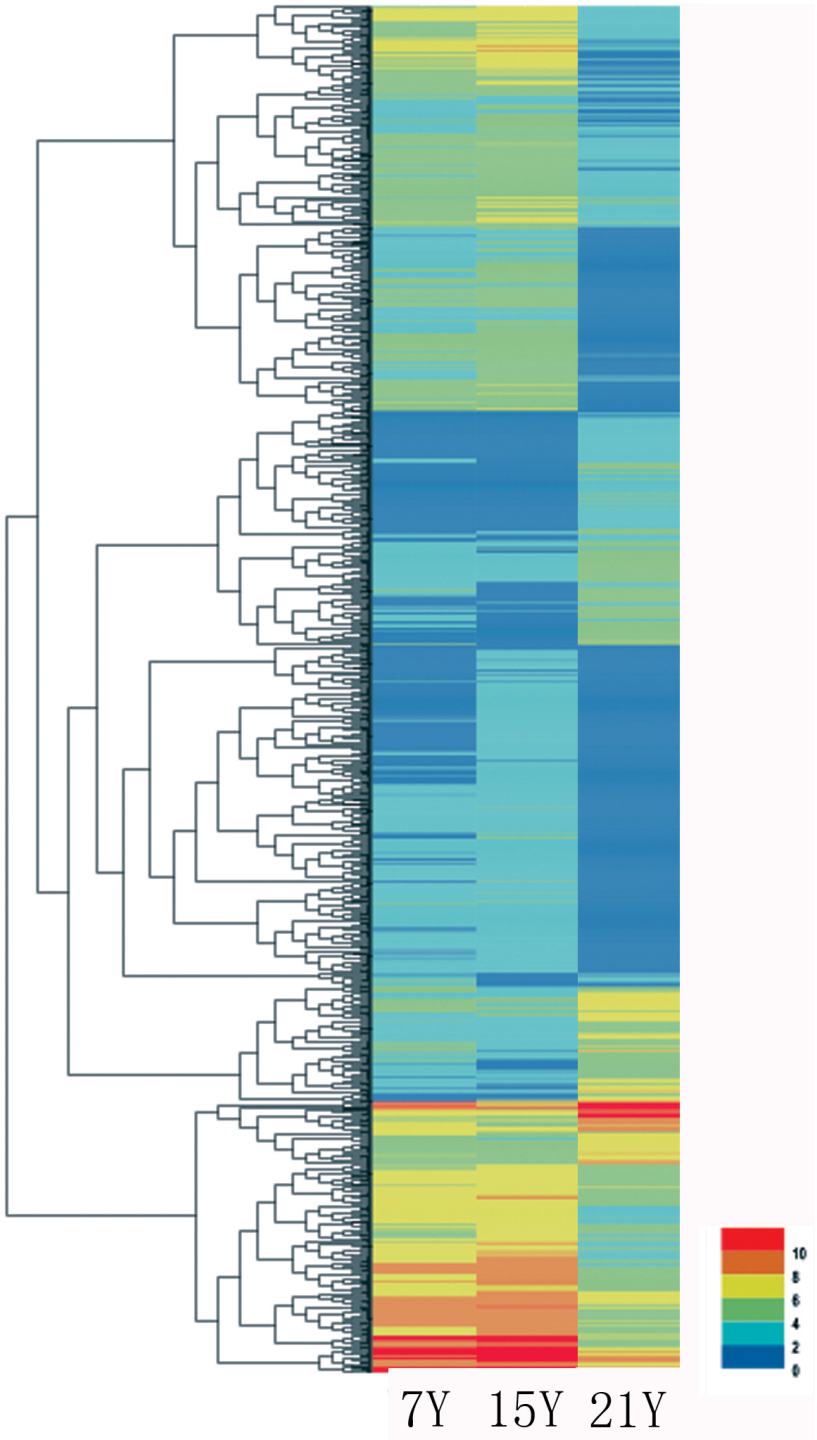

Figure 5 Heatmap of the relative expression levels of differentially expressed genes.

Table 6 Up- and down-regulated Chinese fir DEGs by GO component ontology.

GO component ontology	7Y vs. 15Y			7Y vs. 21Y			15Y vs. 21Y		
	Up	Down	*P*-value	Up	Down	*P*-value	Up	Down	*P*-value
Cell part	0	1	8.72E–01	4	0	5.78E–01	772	747	1.00E+00
Cell	0	1	8.55E–01	3	0	8.97E–01	749	734	1.00E+00
Organelle	0	1	7.36E–01	3	0	7.15E–01	599	591	9.98E–01
Membrane	0	1	3.82E–01	3	0	1.59E–01	345	403	4.27E–09
Organelle part	0	0	1.00E+00	0	0	1.00E+00	234	206	2.02E–01
Macromolecular complex	0	0	1.00E+00	0	0	1.00E+00	117	82	9.91E–01
Membrane part	0	0	1.00E+00	0	0	1.00E+00	107	149	3.44E–04
Cell junction	0	1	4.51E–02	0	0	1.00E+00	60	45	1.70E–04
Extracellular region	0	1	6.24E–02	0	0	1.00E+00	58	98	4.34E–08
Membrane-enclosed lumen	0	0	1.00E+00	0	0	1.00E+00	5	5	1.00E+00
Nucleoid	0	0	1.00E+00	0	0	1.00E+00	3	1	3.49E–01
Extracellular matrix	0	0	1.00E+00	0	0	1.00E+00	1	3	2.22E–01
Extracellular matrix part	0	0	1.00E+00	0	0	1.00E+00	0	1	6.20E–01
Extracellular region part	0	0	1.00E+00	0	0	1.00E+00	0	8	9.85E–04

Table 7 Up- and down-regulated Chinese fir DEGs by GO function ontology.

GO function ontology	7Y vs. 15Y		7Y vs. 21Y		15Y vs. 21Y	
	Up	Down	Up	Down	Up	Down
Catalytic activity	0	1	0	1	600	657
Binding	0	1	1	0	511	561
Transporter activity	0	0	1	0	76	86
Structural molecule activity	0	0	0	0	56	20
Nucleic acid binding transcription factor activity	0	0	0	0	32	29
Electron carrier activity	0	0	0	0	21	15
Molecular transducer activity	0	0	0	0	14	15
Enzyme regulator activity	0	0	0	0	11	8
Antioxidant activity	0	0	0	0	10	9
Receptor activity	0	0	0	0	4	4
Guanyl-nucleotide exchange factor activity	0	0	0	0	2	2
Protein binding transcription factor activity	0	0	0	0	1	0
Nutrient reservoir activity	0	0	0	0	1	1
Metallochaperone activity	0	0	0	0	0	1

2004; *Kubo et al.*, *2005*; *Zhong & Ye*, *2009*; *Wang et al.*, *2010a*). Five NAC transcription factors, 190 MYB transcription factors, and 34 WRKY transcription factors were identified respectively from Chinese fir transcriptome. Phylogenetic analyses of the three clusters of transcription factors were shown in the Figs. S1–S3. The transcription factors of NAC, MYB, WRKY were clustered into two, three, three main classes separately. The expression profiles of differentially expressed transcription factors were analyzed in different developmental phases (Fig. 6). The transcript levels of 32 MYB and 3 WRKY transcription factors increased from 7Y to 21Y, while 28 MYB and 4 WRKY exhibited a reduced expression

Table 8 Up- and down-regulated Chinese fir DEGs by GO process ontology.

GO function ontology	7Y vs. 15Y		7Y vs. 21Y		15Y vs. 21Y	
	Up	Down	Up	Down	Up	Down
Metabolic process	0	1	2	1	782	765
Cellular process	0	1	2	0	661	718
Single-organism process	0	1	1	0	601	672
Response to stimulus	0	1	0	0	357	354
Biological regulation	0	1	0	0	281	278
Localization	0	1	1	0	194	222
Cellular component organization or biogenesis	0	1	0	0	177	188
Developmental process	0	1	0	0	172	211
Multicellular organismal process	0	1	0	0	132	131
Reproductive process	0	1	0	0	89	102
Multi-organism process	0	1	0	0	59	75
Signaling	0	0	0	0	53	75
Growth	0	1	0	0	36	62
Reproduction	0	0	0	0	28	27
Immune system process	0	0	0	0	23	28
Biological adhesion	0	0	0	0	6	15
Rhythmic process	0	0	0	0	3	5
Biological phase	0	0	0	0	2	1
Locomotion	0	0	0	0	1	2

profiles. We chose all of the NAC, WRKY transcription factors and 60 MYB transcription factors (full sequence score > 100) matching to the DEGs results, and discovered five WRKY transcription factors including three up-regulated DEGs (c100548.graph_c0, c49459.graph_c0, c97419.graph_c0) and two down-regulated DEGs (c48904.graph_c0, c70863.graph_c0). Furthermore, 17 MYB transcription factors were dug out including five up-regulated DEGs (c89668.graph_c0, c84067.graph_c0, c90780.graph_c0, etc.) and 12 down-regulated DEGs (c104123.graph_c0, c106541.graph_c0, c98343.graph_c0, etc.) in 15Y vs. 21Y (Table S4).

DISCUSSION

RNA-seq has emerged to be a valuable tool to discover molecular markers and identify novel genes. In recent years, the growing number of species for which significant genetic resources are available is sparking a new era of plant genetic study (*Morozova & Marra*, *2008*; *Coppe et al.*, *2010*). In this study, RNA-seq technology was applied to the Chinese fir transcriptome using Illumina HiSeq^{TM} 2500 platform, and the transcriptome at three different cultivation phases was systematically investigated.

A total of 68.71 million reads and 13.88 gigabase pairs (Gbp) were generated, and 140,486 unigenes with a mean size of 568.64 bp were obtained. About 19.52% of the unigenes were successfully annotated, and the information regarding putative nucleotide variations offers intriguing leads for the analysis of the transcriptome and biological functions in Chinese fir. Among the unigenes 29%, 13%, and 5% appeared to be most closely related to genes

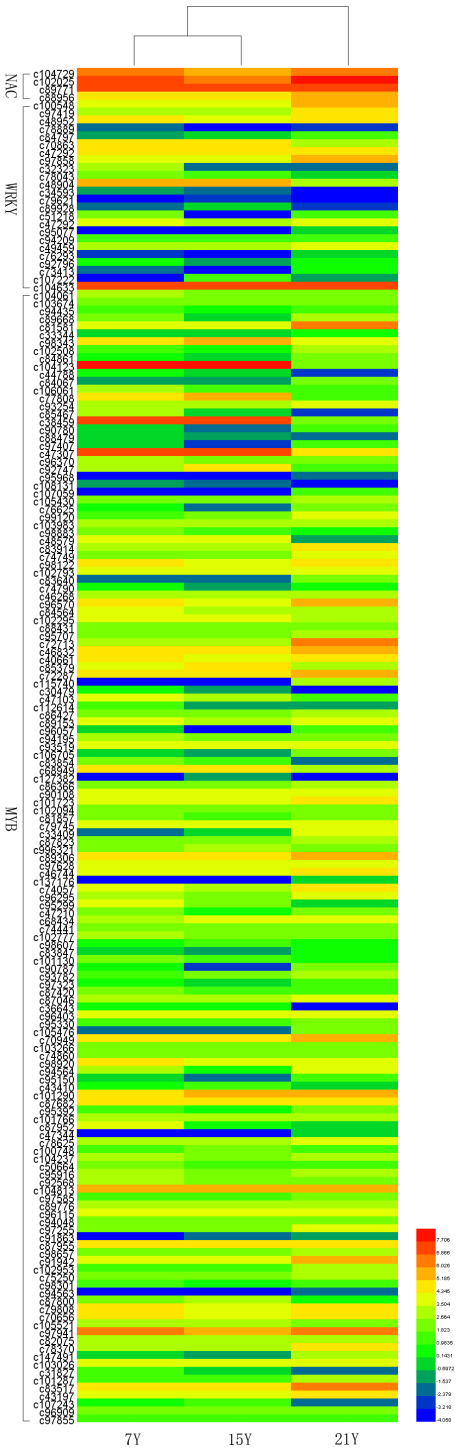

Figure 6 **The differential expression patterns of representative TFs in different developmental phases.**

from *P. sitchensis, Vitis vinifera*, and *Theobroma cacao* respectively. Chinese fir and *P. sitchensis* gymnosperms assigned to the Coniferopsida Coniferae. Thus, there is a close relationship between *P. sitchensis* and Chinese fir based on both systematic botany and molecular analysis. This research moves us toward identifying candidate genes for wood formation and clarifying the functions of the relevant pathways in Chinese fir. Compared with those from other conifer trees, our results using samples from different cultivation phases identified a much larger number of unigenes. In addition, the mean length of unigenes (568 bp) that we obtained is much longer than those in previous studies using the same technology, which reported 449 bp (*Huang et al.*, *2012*), 505 bp (*Wang et al.*, *2013a*), and 497 bp (*Qiu et al.*, *2013*). To the best of our knowledge, this study represents the first attempt at de novo sequencing and assembly of the Chinese fir trancriptome using RNA-seq, to focus on different cultivation phases including fast growing stage, stem growth stage and senescence stage respectively. The results obtained in this research demonstrated that our final assembly quality was satisfactory and it therefore provides sequence resources and facilitates further gene cloning and functional analyses.

Several genes encoding the biosynthesis of wood components (cellulose, xylan, glucomannan, and lignin), such as Cel/TDIF/CLE/PXY-WOX4/MYB (*Plomion, Leprovost & Stokes*, *2001*; *Ito et al.*, *2006*; *Hirakawa et al.*, *2008*; *Etchells & Turner*, *2010*; *Hirakawa, Kondo & Fukuda*, *2010*; *Ji et al.*, *2010*; *Suer et al.*, *2011*; *Wang et al.*, *2011*), have been identified in angiosperms (*Jansson & Douglas*, *2007*). Unfortunately, only few of the related genes have been identified and functionally characterized in gymnosperms. In this study, the lack of a reference genome for Chinese fir, hampered our efforts to determine gene and their functions. BLAST hits Perhaps these unigenes might play an important roles in Chinese fir and quite different from the other species.

Differences in gene expression profiles can yield insight into mechanisms underlying physiological changes, and DEGs were found among different developmental stages, tissues, treatments, and species (*Zeng et al.*, *2010*; *Logacheva et al.*, *2011*; *Wang et al.*, *2013b*; *Wu et al.*, *2013*; *Wu et al.*, *2014*). The 7Y, 15Y and 21Y samples that we collected represented the different cultivation development phases (fast growing stage, stem growth stage and senescence stage), and there were vary a few DEGs in 7Y compared to either 15Y or 21Y. However, many up-regulated (2,623) and down-regulated (3,276) DEGs were detected in 15Y compared to 21Y. Furthermore, most DEG unigenes in 15Y vs. 21Y were annotated to specific pathways using the KEGG database, including metabolic pathways, and biosynthesis of secondary metabolites among others. These results indicate considerable changes of gene expression in immature xylem during the transition from stem growth stage and senescence stage. A similar phenomenon was reported for the transition from the active stage to the dormant stage in vascular cambium, in which expression of the core cell cycle genes in vascular cambium correlated well with the cessation of cambial cell division (*Brown et al.*, *2005*; *Druart et al.*, *2007*; *Li et al.*, *2009*; *Galindo González et al.*, *2012*; *Qiu et al.*, *2013*). We can conclude that the key cycle protein transcripts expressed preferentially at stem growth stage and senescence stage of Chinese fir may be essential for wood formation, and these DEGs are involved in a broad range of physiological functions between stem growth stage and senescence stage.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by Major State Basic Research Development Program of China (No. 2012CB114500), National "Twelfth Five-year" Plan for Science & Technology Support Development Program of China (No. 2012BAD01B0201) and the Bamboo/Tree Breeding Project of Zhejiang Province "Twelfth Five-Year" Plan (No. 2012C12908-11). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Major State Basic Research Development Program: 2012CB114500.
Bamboo/Tree Breeding Project: 2012BAD01B0201.
Science & Technology Support Development Program: 2012C12908-11.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yunxing Zhang and Xiaojiao Han conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Jian Sang performed the experiments, contributed reagents/materials/analysis tools.
- Xuelian He, Mingying Liu and Guirong Qiao analyzed the data.
- Renying Zhuo conceived and designed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, reviewed drafts of the paper.
- Guiping He and Jianjun Hu conceived and designed the experiments.

Data Availability

The following information was supplied regarding data availability:

The National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (accession number: SRS959453).

Assembly contigs can be found in Figshare: https://figshare.com/s/fdf6af8b8ae6aa02bd52.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.2097#supplemental-information.

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25(17)**:3389–3402 DOI 10.1093/nar/25.17.3389.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M. 2004. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **32(suppl 1)**:D115–D119 DOI 10.1093/nar/gkh131.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25(1)**:25–29 DOI 10.1038/75556.

Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, Mittapalli O. 2011. Transcriptomic signatures of ash (*Fraxinus spp.*) phloem. *PLoS ONE* **6(1)**:e16368 DOI 10.1371/journal.pone.0016368.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57(1)**:289–300.

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Saint Yuen MM, Keeling CI, Brand D, Vandervalk BP. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* **29(12)**:1492–1497.

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR. 2005. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *The Plant Cell Online* **17(8)**:2281–2295 DOI 10.1105/tpc.105.031542.

Brunner AM, Busov VB, Strauss SH. 2004. Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends in Plant Science* **9(1)**:49–56.

Chen S, Jiang J, Li H, Liu G. 2012. The salt-responsive transcriptome of *Populus simonii* × *Populus nigra* via DGE. *Gene* **504(2)**:203–212 DOI 10.1016/j.gene.2012.05.023.

Coppe A, Pujolar JM, Maes GE, Larsen PF, Hansen MM, Bernatchez L, Zane L, Bortoluzzi S. 2010. Sequencing, de novo annotation and analysis of the first Anguilla anguilla transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics* **11(1)**:635 DOI 10.1186/1471-2164-11-635.

Cronk Q. 2005. Plant eco-devo: the potential of poplar as a model organism. *New Phytologist* **166(1)**:39–48 DOI 10.1111/j.1469-8137.2005.01369.x.

Del Lungo A, Ball J, Carle J. 2006. Global planted forests thematic study. Results and analysis. In: *Planted forests and trees working papers*. F AO.

Deng Y, Li J, Wu S, Zhu Y, Chen Y, He F. 2006. Integrated nr database in protein annotation system and its localization. *Computer Engineering* **32(5)**:71–74.

Druart N, Johansson A, Baba K, Schrader J, Sjödin A, Bhalerao RR, Resman L, Trygg J, Moritz T, Bhalerao RP. 2007. Environmental and hormonal regulation of the activity–dormancy cycle in the cambial meristem involves stage-specific modulation of transcriptional and metabolic networks. *The Plant Journal* **50(4)**:557–573 DOI 10.1111/j.1365-313X.2007.03077.x.

Duan A, Zhang J, He C, Tong S. 2004. Study on the change laws of biomass of Chinese Fir plantations. *Forest Research* **18(2)**:125–132.

Etchells JP, Turner SR. 2010. The PXY-CLE41 receptor ligand pair defines a multi-functional pathway that controls the rate and orientation of vascular cell division. *Development* **137(5)**:767–774 DOI 10.1242/dev.044941.

Galindo González LM, El Kayal W, JU CJT, Allen CC, King-Jones S, Cooke JE. 2012. Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant, Cell and Environment* **35(4)**:682–701 DOI 10.1111/j.1365-3040.2011.02444.x.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29(7)**:644–652 DOI 10.1038/nbt.1883.

Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF peptide signaling regulates vascular stem cell proliferation via the WOX4 homeobox gene in Arabidopsis. *The Plant Cell Online* **22(8)**:2618–2629 DOI 10.1105/tpc.110.076083.

Hirakawa Y, Shinohara H, Kondo Y, Inoue A, Nakanomyo I, Ogawa M, Sawa S, Ohashi-Ito K, Matsubayashi Y, Fukuda H. 2008. Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proceedings of the National Academy of Sciences of the United States of America* **105(39)**:15208–15213 DOI 10.1073/pnas.0808444105.

Huang H-H, Xu L-L, Tong Z-K, Lin E-P, Liu Q-P, Cheng L-J, Zhu M-Y. 2012. De novo characterization of the Chinese fir (*Cunninghamia lanceolata*) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. *BMC Genomics* **13(1)**:648 DOI 10.1186/1471-2164-13-648.

Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H. 2006. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science* **313(5788)**:842–845 DOI 10.1126/science.1128436.

Jansson S, Douglas CJ. 2007. Populus: a model system for plant biology. *Annual Review of Plant Biology* **58**:435–458 DOI 10.1146/annurev.arplant.58.032806.103956.

Ji J, Strable J, Shimizu R, Koenig D, Sinha N, Scanlon MJ. 2010. WOX4 promotes procambial development. *Plant Physiology* **152(3)**:1346–1356 DOI 10.1104/pp.109.149641.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32(suppl 1)**:D277–D280 DOI 10.1093/nar/gkh063.

Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T. 2005. Transcription switches for protoxylem and metaxylem vessel formation. *Genes & Development* **19(16)**:1855–1860 DOI 10.1101/gad.1331305.

Lei J. 2005. *Forest resources of China*. Beijing: Chinese Forestry, 172–173.

Li W-F, Ding Q, Chen J-J, Cui K-M, He X-Q. 2009. Induction of PtoCDKB and PtoCYCB transcription by temperature during cambium reactivation in Populus tomentosa Carr. *Journal of Experimental Botany* **60(9)**:2621–2630 DOI 10.1093/jxb/erp108.

Li M, Ritchie GA. 1999. Eight hundred years of clonal forestry in China: I. Traditional afforestation with Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.). *New Forests* **18(2)**:131–142.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20(2)**:265–272 DOI 10.1101/gr.097261.109.

Liu J-J, Sturrock RN, Benton R. 2013. Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to Cronartium ribicola. *BMC Genomics* **14(1)**:884 DOI 10.1186/1471-2164-14-884.

Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ, Penin AA. 2011. De novo sequencing and characterization of floral transcriptome in two species of buckwheat (Fagopyrum). *BMC Genomics* **12(1)**:30 DOI 10.1186/1471-2164-12-30.

Mardis ER. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* **9**:387–402 DOI 10.1146/annurev.genom.9.081307.164359.

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. 2010. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics* **11(1)**:681 DOI 10.1186/1471-2164-11-681.

Morozova O, Hirst M, Marra MA. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics* **10**:135–151 DOI 10.1146/annurev-genom-082908-145957.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92(5)**:255–264 DOI 10.1016/j.ygeno.2008.07.001.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D. 2014. The genome of *Eucalyptus grandis*. *Nature* **510(7505)**:356–362.

Nagalakshmi U, Waern K, Snyder M. 2010. RNA-Seq: a method for comprehensive transcriptome analysis. *Current Protocols in Molecular Biology* **chapter 4**:Unit 4.11.1–Unit 4.11.13 DOI 10.1002/0471142727.mb0411s89.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497(7451)**:579–584 DOI 10.1038/nature12211.

Orwa C, Mutua A, Kindt R, Jamnadass R, Anthony S. 2013. *Agroforestry database: a tree reference and selection guide version 4 0*. Kenya: World Agroforestry Centre.

Plomion C, Leprovost G, Stokes A. 2001. Wood formation in trees. *Plant Physiology* **127(4)**:1513–1523 DOI 10.1104/pp.010816.

Qiu Z, Wan L, Chen T, Wan Y, He X, Lu S, Wang Y, Lin J. 2013. The regulation of cambial activity in Chinese fir (*Cunninghamia lanceolata*) involves extensive transcriptome remodeling. *New Phytologist* **199(3)**:708–719 DOI 10.1111/nph.12301.

Raherison ESM, Giguère I, Caron S, Lamara M, MacKay JJ. 2015. Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures. *New Phytologist* **207(1)**:172–187 DOI 10.1111/nph.13343.

Raherison ES, Rigault P, Caron S, Poulin P-L, Boyle B, Verta J-P, Giguère I, Bomal C, Bohlmann J, MacKay J. 2012. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. *BMC Genomics* **13(1)**:434 DOI 10.1186/1471-2164-13-434.

Rogers LA, Campbell MM. 2004. The genetic control of lignin deposition during plant growth and development. *New Phytologist* **164(1)**:17–30 DOI 10.1111/j.1469-8137.2004.01143.x.

Schuster SC. 2008. Next-generation sequencing transforms today's biology. *Nature Methods* **5(1)**:16–18.

Shi JS, Zhen Y, Zheng RH. 2010. Proteome profiling of early seed development in *Cunninghamia lanceolata* (Lamb.) Hook. *Journal of Experimental Botany* **61(9)**:2367–2381.

Suer S, Agusti J, Sanchez P, Schwarz M, Greb T. 2011. WOX4 imparts auxin responsiveness to cambium cells in Arabidopsis. *The Plant Cell Online* **23(9)**:3247–3259 DOI 10.1105/tpc.111.087874.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28(1)**:33–36 DOI 10.1093/nar/28.1.33.

Thumma BR, Sharma N, Southerton SG. 2012. Transcriptome sequencing of *Eucalyptus camaldulensis* seedlings subjected to water stress reveals functional single nucleotide polymorphisms and genes under selection. *BMC Genomics* **13(1)**:364 DOI 10.1186/1471-2164-13-364.

Tyler B. 2006. The genome of black cottonwood, *Populus trichocarpa*. *Science* **313**:1596–1604 DOI 10.1126/science.1128691.

Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA. 2010a. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proceedings of the National Academy of Sciences of the United States of America* **107(51)**:22338–22343 DOI 10.1073/pnas.1016436107.

Wang Z, Chen J, Liu W, Luo Z, Wang P, Zhang Y, Zheng R, Shi J. 2013a. Transcriptome characteristics and six alternative expressed genes positively correlated with the phase transition of annual cambial activities in Chinese Fir (*Cunninghamia lanceolata* (Lamb.) Hook). *PLoS ONE* **8(8)**:e71562.

Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y. 2010b. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (Ipomoea batatas). *BMC Genomics* **11(1)**:726.

Wang H, Soler M, Yu H, Camargo ELO, San Clemente H, Savelli B, Ladouce N, Paiva J, Grima-Pettenati J. 2011. Master regulators of wood formation in Eucalyptus. In: *BMC proceedings*. BioMed Central Ltd.

Wang Y, Xu L, Chen Y, Shen H, Gong Y, Limera C, Liu L. 2013b. Transcriptome profiling of radish (Raphanus sativus L.) root and identification of genes involved in response to lead (Pb) stress with next generation sequencing. *PLoS ONE* **8(6)**:e66539 DOI 10.1371/journal.pone.0066539.

Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X. 2011. Characterization of the sesame (Sesamum indicum L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* **12(1)**:451 DOI 10.1186/1471-2164-12-451.

Wong MM, Cannon CH, Wickneswari R. 2011. Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and Acacia mangium via de novo transcriptome sequencing. *BMC Genomics* **12(1)**:342 DOI 10.1186/1471-2164-12-342.

Wu D, Austin RS, Zhou S, Brown D. 2013. The root transcriptome for North American ginseng assembled and profiled across seasonal development. *BMC Genomics* **14(1)**:564 DOI 10.1186/1471-2164-14-564.

Wu Z-J, Li X-H, Liu Z-W, Xu Z-S, Zhuang J. 2014. De novo assembly and transcriptome characterization: novel insights into catechins biosynthesis in Camellia sinensis. *BMC Plant Biology* **14(1)**:277 DOI 10.1186/s12870-014-0277-4.

Ye Z-H, Zhong R. 2015. Molecular control of wood formation in trees. *Journal of Experimental Botany* **66(14)** Epub ahead of print March 5 2015 DOI 10.1093/jxb/erv081.

Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe BA, Wang Y. 2010. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, Epimedium sagittatum (Sieb. Et Zucc.) Maxim. *BMC Genomics* **11(1)**:94 DOI 10.1186/1471-2164-11-94.

Zhang Y, Zhang S, Han S, Li X, Qi L. 2012. Transcriptome profiling and in silico analysis of somatic embryos in Japanese larch (*Larix leptolepis*). *Plant Cell Reports* **31(9)**:1637–1657 DOI 10.1007/s00299-012-1277-1.

Zhong R, Ye Z-H. 2009. Transcriptional regulation of lignin biosynthesis. *Plant Signaling & Behavior* **4(11)**:1028–1034 DOI 10.4161/psb.4.11.9875.

Zhong R, Ye Z-H. 2014. Complexity of the transcriptional network controlling secondary wall biosynthesis. *Plant Science* **229**:193–207 DOI 10.1016/j.plantsci.2014.09.009.