

Full Paper

Identification and allelic dissection uncover roles of lncRNAs in secondary growth of *Populus tomentosa*

Daling Zhou^{1,2,3}, Qingzhang Du^{1,2,3}, Jinhui Chen^{1,2,3}, Qingshi Wang^{1,2,3}, and Deqiang Zhang^{1,2,3,*}

¹Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P.R. China, ²National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P.R. China, and ³Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P. R. China

*To whom correspondence should be addressed. Tel. +86 10 62336007. Fax. +86 10 62336164. Email: deqiangzhang@bjfu.edu.cn

Edited by Dr. Satoshi Tabata

Received 24 November 2016; Editorial decision 28 March 2017; Accepted 4 April 2017

Abstract

Long non-coding RNAs (lncRNAs) function in various biological processes. However, their roles in secondary growth of plants remain poorly understood. Here, 15,691 lncRNAs were identified from vascular cambium, developing xylem, and mature xylem of *Populus tomentosa* with high and low biomass using RNA-seq, including 1,994 lncRNAs that were differentially expressed (DE) among the six libraries. 3,569 *cis*-regulated and 3,297 *trans*-regulated protein-coding genes were predicted as potential target genes (PTGs) of the DE lncRNAs to participate in biological regulation. Then, 476 and 28 lncRNAs were identified as putative targets and endogenous target mimics (eTMs) of *Populus* known microRNAs (miRNAs), respectively. Genome re-sequencing of 435 individuals from a natural population of *P. tomentosa* found 34,015 single nucleotide polymorphisms (SNPs) within 178 lncRNA loci and 522 PTGs. Single-SNP associations analysis detected 2,993 associations with 10 growth and wood-property traits under additive and dominance model. Epistasis analysis identified 17,656 epistatic SNP pairs, providing evidence for potential regulatory interactions between lncRNAs and their PTGs. Furthermore, a reconstructed epistatic network, representing interactions of 8 lncRNAs and 15 PTGs, might enrich regulation roles of genes in the phenylpropanoid pathway. These findings may enhance our understanding of non-coding genes in plants.

Key words: association studies, epistatic interaction, long non-coding RNA, target mimics, *Populus tomentosa*

1. Introduction

Long non-coding RNAs (lncRNAs) are non-coding transcripts longer than 200 nucleotides (nts). According to their genomic location, lncRNAs can be classified into intergenic, intronic, sense, and antisense

types.¹ In contrast to protein-coding genes, most lncRNAs lack strong sequence conservation between species.² LncRNAs are usually expressed at low levels and often exhibit tissue-specific expression.³ Emerging evidence shows that lncRNAs play vital roles in numerous biological

processes by regulating the expression of genes in close proximity (acting in *cis*) or at a distance (acting in *trans*) in the genome.⁴

With the advent of high-throughput sequencing, thousands of lncRNAs have been identified in several plant species. However, the detailed functional analysis of plant lncRNAs is still in its infancy. For example, vernalization in *Arabidopsis thaliana* is influenced by lncNAT 'COOLAIR' and intronic lncRNA 'COLDAIR'.^{5,6} lncRNAs also function as endogenous target mimics (eTMs) of microRNAs (miRNAs), providing a new mechanism for regulation of miRNA activity.⁷

Trees, as model systems for research on plant's secondary growth, promise to be one of the most abundant renewable sources for the production of wood biomass.⁸ Therefore, a better understanding of the regulation of secondary growth and wood biosynthesis is essential for improving wood quality and biomass. Previous studies of molecular mechanisms underlying plant's secondary growth are mainly focused on functional studies of protein-coding genes, such as the 95 genes identified in the phenylpropanoid pathway.⁹ Nevertheless, few studies have systemically investigated the roles of lncRNAs in secondary growth of plants.

Single nucleotide polymorphisms (SNPs) represent the most frequent genetic variants among individuals and can affect gene expression, function, and phenotypes.¹⁰ Recent studies have linked SNPs in lncRNAs to the abnormal expression and dysregulation of the lncRNAs.¹¹ For example, a SNP in the lncRNA long-day-specific male-fertility-associated RNA affected the expression of the lncRNA, resulting in photoperiod-sensitive male sterility in rice.¹² Genome-wide association studies in human have assigned 93% of the disease- or trait-predisposing SNPs to non-coding regions that could affect the expression of non-coding RNAs.^{13,14} Therefore, identification of phenotype-related SNPs within lncRNA loci and association analysis that link the genetic variations (SNPs) to phenotypic variations can provide new clues for their potential functions of lncRNAs.

Here, we systemically identified and characterized lncRNAs from the main vascular tissues (cambium, developing xylem, and mature xylem) of secondary growth in *Populus tomentosa* with high and low biomass. Prediction of functional motifs, potential target genes (PTGs), and miRNAs regulation for the lncRNAs illustrated their proposed functions in secondary growth. We found SNPs within the lncRNA loci and their PTGs in a natural population of *P. tomentosa*. Genetic association was then used to associate these SNPs with phenotype variations to further explore functions of lncRNAs. Overall, the findings provide insights into the roles of lncRNAs in secondary growth of plants.

2. Materials and methods

2.1. Plant materials

This study was carried out in a hybrid population of 722 F₁ individuals established by controlled crossing between two poplar parents, clone 'Pt-3' (*P. tomentosa*) as female and 'Pt-TB14' (*P. tomentosa*) as male. The progeny of this population were grown in 2009 in Guan Xian County, Shandong Province, China (36°23'N, 115°47'E) using a randomized complete block design with three replications. All individuals were scored for three growth traits, including diameter at breast height (DBH), tree height (H), and stem volume (V). Based on the measurement (Supplementary Fig. S1), three genotypes with highest biomass and three with lowest biomass (measured as wood volume) were selected to produce two pooled samples (high- and

low-biomass pools) with three biological replicates. A *t*-test (SAS v8.2) showed a significant difference between the high and low biomass groups ($P \leq 0.01$). For each pooled sample, the stem bark of each tree was gently removed in strips (100 mm wide and 350–400 mm long) and then the soft cambial cells were gently scraped off. After that, the developing xylem cells (newly formed xylem about 3 mm away from the exposed surface) below the cambium were immediately scraped away and the mature xylem cells were obtained by scraping deeper into the wood, using the technique described by Dalessandro and Northcote.¹⁵ These tissues, including vascular cambium, developing xylem, and mature xylem from the high-biomass pools (HC, HD, and HM) and low-biomass pools (LC, LD, and LM), respectively, were immediately collected and frozen in liquid nitrogen and then stored at -80°C for RNA extraction.

2.2. RNA isolation, library construction and RNA-seq

Total RNA from HC, HD, HM, LC, LD, and LM was extracted using the Qiagen RNeasy kit (Qiagen China, Shanghai, China) according to the manufacturer's instructions. Additional on-column DNase digestions were performed during the RNA purification using RNase-Free DNase Set (Qiagen). The six total RNA samples were monitored by NanoDrop ND-1000 and Agilent Bioanalyzer 2100 before RNA-seq and then used to construct strand-specific RNA-seq libraries separately according to the TruSeq RNA Sample Preparation Guide. After being quantified with the Qubit 2.0 Fluorometer and Agilent 2100 bioanalyzer, the six strand-specific libraries were sequenced on the Illumina HiSeq2500 platform with the paired-end programme. Library construction and Illumina sequencing were carried out by Shanghai Biotechnology Corporation (Shanghai, China). The sequencing data have been deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>) under the accession number SRP073689.

2.3. Prediction of *Populus* lncRNAs

Pre-processing of raw sequencing data were performed using FASTX-Toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) with default parameters by removing adapter sequences, low-quality reads ($Q20 < 20$), sequences shorter than 20 nts. To rule out rRNAs, high-quality reads that passed the quality filters were aligned to *Populus* rRNA sequences using the Short Oligonucleotide Analysis Package (SOAP2; <http://soap.genomics.org.cn/soapaligner.html>). Supplementary Table S1 shows the summary of the sequencing data. Then clean reads were aligned to the *Populus trichocarpa* genome (version 3.0)¹⁶ using Tophat,¹⁷ with 3 base mismatches allowed. The reads with no more than three mismatches were used to construct transcripts of each sample separately using Cufflinks v2.1.1¹⁷ based on the *P. trichocarpa* genome reference. To reduce transcriptional noise, only those assembled transcript isoforms that were detected in two or more experiments were retained for further analyses, as performed by Li *et al.*¹⁸ Then, the Cuffcompare programme was used to merge the RefSeq, Ensembl, and UCSC *Populus* known genes into one set of gene annotations for comparison with the assembled transcripts. Expression levels (FPKM, fragments per kilobase of transcript per million fragments) of these assembled transcripts were calculated by Cufflinks v2.1.1.¹⁷ The prediction of lncRNAs from RNA-seq data was performed according to Sun *et al.*¹⁹ and the pipeline was shown in Supplementary Fig. S2. We only retained transcripts longer than 200 nts with an open reading frame shorter than 100 amino acids and optimum expression threshold of $\text{FPKM} > 1.0$ in at least one

sample. After that, we used double filters to evaluate the protein-coding potential of the remaining transcripts. Transcripts were first aligned to the Pfam database using HMMER3.0²⁰ to eliminate transcripts with potential protein-coding ability (cutoff *E*-value ≤ 0.001). Next, we employed the Coding Potential Calculator (CPC) software²¹ and Coding–non-coding Index (CNCI) software²² to evaluate the protein-coding potential of the remaining transcripts with default parameters. When using CPC, we used the protein-coding transcripts of *P. trichocarpa* as a reference. Only transcripts that did not pass the protein-coding-score test (CPC score < 0 , CNCI score < 0) were classified as putative lncRNAs. LncRNAs were classified into four categories ('i', 'o', 'u', and 'x'), consisting of lncRNAs without overlap with any genes (intergenic lncRNAs, 'u'), antisense overlapping lncRNAs (antisense lncRNAs, 'x'), intronic overlapping lncRNAs (intronic lncRNAs, 'i') and generic exonic overlapping lncRNAs (sense lncRNAs, 'o'), according to their genomic locations relative to neighbouring genes as defined by cuffcompare programme in Cufflinks suite.²³ The difference in lncRNA expression was calculated as the fold change (FC) = FPKM of 'sample1'/FPKM of 'sample2'. LncRNA transcripts were considered to be differentially expressed (DE) if they meet the criteria of $FC \geq 2$ or ≤ 0.5 with *P*-value < 0.01 .

2.4. Genomic characterization and specific expression of *Populus* lncRNAs

GC contents of the identified lncRNAs were calculated with EMBOSS explorer's geecee tool (<http://emboss.sourceforge.net/apps/cvs/emboss/apps/geecee.html>). To explore lncRNA conservation, all the lncRNA sequences identified here were aligned with BLASTN against the genome sequences of *A. thaliana*²⁴ and rice²⁵ with a cutoff *E*-value $< 1e-10$. The genomes were downloaded from Phytozome (v9.1) (<http://www.phytozome.net/>). The lncRNAs that had $>20\%$ of their sequence matched to other genomes were defined as conserved lncRNAs. Motifs in lncRNAs were identified using DREME online software specially designed to find relatively short motifs with *E*-value < 0.05 and the possible roles of each identified motif were predicted by using the GOMo (Gene Ontology for Motifs) annotation server with *q*-value < 0.05 .^{26,27} The tissue specificity of lncRNA expression was evaluated according to the tissue-specific index, which ranges from 0 for housekeeping genes to 1 for tissue-restricted genes, as described by Yanai *et al.*²⁸ The index was

calculated as: tissue-specific index = $\frac{\sum_{i=1}^n (1 - \frac{exp_i}{exp_{max}})}{n - 1}$, where *n* is the number of tissues; *exp_i* is the expression value of each lncRNA in tissue, *i*; and *exp_{max}* is the maximum expression value of each lncRNA among all tissues. Only the lncRNAs showing a tissue-specific index > 0.9 were considered to be tissue-specific.

2.5. Prediction of PTGs of *Populus* lncRNAs

Two independent algorithms were then used to predict PTGs of the DE lncRNAs based on the regulatory effects of the lncRNA (*cis*- or *trans*-acting). For the first algorithm, genes transcribed within a 10 kb window upstream or downstream of each lncRNA were considered as potential *cis* target genes by using genome annotation and a genome browser, as described by Jia *et al.*²⁹ The second algorithm for the prediction of potential *trans* target genes is based on sequence complementarity and RNA duplex energy prediction to access the impact of lncRNA binding on mRNA molecules, as described previously in.³⁰ Specifically, the algorithm first employed BLAST to select target sequences complementary to the lncRNAs with parameters of

e-value $< 1e-5$ and identity $\geq 95\%$ and then used RNAplex to calculate the complementary energy between two sequences for further screening and selection of potential *trans* target genes.³¹ The predicted target genes were then annotated by PopGenie (<http://www.popgenie.org/>) and gene ontology (GO) terms were analysed by agriGO (<http://bioinfo.cau.edu.cn/agriGO/index.php>) to find significant GO categories [false discovery rate (FDR) ≤ 0.05]. In addition, we used the KEGG database (<http://www.genome.ad.jp/kegg/>) and a hypergeometric statistic test to analyses their potential roles in the pathways.

2.6. Prediction of lncRNAs as potential precursors, targets, and target mimics of miRNAs

To explore whether some lncRNAs may act as precursors of miRNAs, the 401 *Populus* known miRNAs in miRBase (Release 20.0, <http://www.mirbase.org/>) were aligned to the sequences of the novel lncRNAs and secondary structure prediction was then executed using the Vienna RNA package RNAfold web (<http://rna.tbi.univie.ac.at/>). LncRNAs with classic stem-loop hairpins were regarded as putative precursors of miRNAs. Furthermore, to investigate whether some lncRNAs may be targeted by miRNAs, the target genes of miRNAs were predicted by psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) with expectation ≤ 3 based on the near-perfect complementarity between miRNAs and target genes and target-site accessibility. The miRNA eTMs from *Populus* lncRNAs were predicted with local scripts by using the algorithm developed by Wu *et al.*³² Quiet different from the authentic targets that are nearly perfectly complementary to miRNAs, eTMs can bind to a miRNA with a 3-nt bulge between the 5' end 9th to 12th positions of the miRNA, thus serving as decoys for the miRNA to interfere with the binding of the miRNA to its authentic targets. Based on these prediction results, a potential interaction network of miRNAs, miRNAs, and lncRNAs was modelled with Cytoscape 3.2.³³

2.7. Real-time quantitative PCR

Real-time quantitative PCR (qRT-PCR) was performed on a 7500 Fast RT-PCR System using the SYBR Premix Ex Taq as described in the manufacturer's instructions. The cDNA template for the reactions was reverse-transcribed using total RNA extracted from cambium, developing xylem, and mature xylem of the high- and low-biomass samples, as described earlier. Primers were designed using Primer Express 3.0 software (Supplementary Material S1). The efficiency of the primers was calculated by performing RT-PCR on several dilutions of first-strand cDNAs. Efficiencies of the different primer sets were similar. The specificity of each primer set was checked by sequencing PCR products. All qRT-PCR amplifications were carried out in triplicate, with the standard reaction programme and the specificity of the amplified fragments was checked using the generated melting curve. The generated real-time data were analysed using the Opticon Monitor Analysis Software 3.1 tool and standardized to the levels of poplar ACTINII-like (Accession number: EF145577) using the $2^{-\Delta C_t}$ or $2^{-\Delta\Delta C_t}$ method.³⁴

2.8. Association population

The association population used for association studies consisted of 435 unrelated individuals of *P. tomentosa* ($2n = 38$), representing almost the entire geographic distribution of *P. tomentosa* as described in our previous study.³⁵ These trees were selected from the collection of 1,047 native individuals collected from the entire

natural distribution of *P. tomentosa* (30–40°N, 105–125°E) in 1982 by using root cuttings, and were grown using a randomized complete block design with three replications in Guan Xian County, Shandong Province, China (36°23'N, 115°47'E).

The 435 individuals of the population were then scored for 10 quantitative traits, with at least three clones measured per genotype. The three growth traits (DBH, H, and V) and seven wood property traits have been measured and analysed by Du *et al.*³⁶ The wood property traits were microfibril angle (MFA), fibre length (FL), fibre width (FW), holocellulose content (HoC), hemicellulose content (HemC), α -cellulose content (α C), and lignin content (LiC). The phenotypic data used for association analysis are shown in Supplementary Material S2.

2.9. Re-sequencing of the association population and SNP calling

The set of 435 individuals from the association population were re-sequenced using the Illumina GA2 instrument with an average of 15-fold coverage, and the libraries were constructed based on genomic DNA, following the manufacturer's recommendations (Illumina). The sequence quality of paired-end short reads of 100 base pair (bp) was controlled by removing low-quality reads ($\leq 50\%$ of nucleotides with a quality score $< Q20$).³⁷ Then the paired-end short reads were aligned and mapped to the *Populus* reference genome using SOAPaligner/SOAP2 v2.20 with default options.³⁸ To get high-quality SNPs, only the reads that could be matched to a unique genomic location were used for SNP calling; reads that mapped to at least two genomic positions were excluded. After SNP calling, we used our previous SNP data from 10 candidate genes in 120 individuals identified by PCR-Sanger sequencing to validate our SNP calling results and found that the accuracy of SNP calling reached 97.5%.³⁷ Then the SNPs among the association population were mined using the Variant Call Format tool (version 4.1).³⁹ BLASTN and BLASTX were employed with a cutoff *E*-value $< 1e-10$ to obtain information in the location of the lncRNAs genes that represent DNA sequences encoding the lncRNAs and their PTGs, including promoter regions (2,000 bp upstream), or gene fragments, respectively. The genomic DNA sequences of the PTGs of the *P. tomentosa* lncRNAs have been deposited in GenBank under the Accession Number KX155245-KX155490 (Supplementary Material S3). We then used UltraEdit 3.2 (<http://www.ultraedit.com/>) to capture the genotypes of the SNPs from the SNP-calling pipeline. Low-quality SNP markers with missing data ($>10\%$) were excluded from further analysis.

SNP diversity and linkage disequilibrium (LD) analysis: On the basis of the SNP genotyping data, the average number of pair-wise differences per site between sequences, π ,⁴⁰ and the number of segregating sites, θ_w ,⁴¹ were used for calculation of SNP diversity using TASSEL v5.0 (<http://www.maizegenetics.net/>). Also, Tajima's *D*⁴² for neutrality tests was estimated using the genotypic data. LD tests were performed in the association population using common SNPs with missing data ($<25\%$) and minor allele frequencies (MAFs) > 0.05 in the natural population. The squared correlation of allele frequencies value (r^2) was calculated in TASSEL v5.0. To assess the extent of LD, the decay of LD with physical distance (bp) between the common SNPs was estimated by nonlinear regression for the genes and lncRNAs within the same chromosome.

2.10. Association analysis

Single SNP-based associations: we used the mixed linear model (MLM) in TASSEL v5.0 to identify the SNP-trait associations in the

association population. This model takes into account the effects of population structure and relatedness among individuals for marker-trait associations.⁴³ Based on 20 genomic simple sequence repeat markers described previously in,⁴⁴ the pair-wise kinship coefficients (*K*) were evaluated by SPAGeDi 1.3,⁴⁵ and the association population structure (*Q*) was obtained based on significant subpopulations ($k = 3$) using STRUCTURE 2.3.4.⁴⁶ Specifically, the genotypic effects of associated SNPs could be effectively decomposed into additive and dominant effects under the model. Then, corrections for the testing of *P*-values for all the associations were performed using the FDR through QVALUE.⁴⁷ A *q*-value of 0.10 was considered as the significance threshold.

Epistasis analysis: EPISNP v2.0 (<http://animalgene.umn.edu/episnpmpi/download.html>) provides an ideal genetic modelling method for multiple SNPs to investigate epistatic interactions and was developed for testing epistatic SNP effects on quantitative traits.⁴⁸ It incorporated tests of five epistasis effects for each pair of SNPs, including two-locus interaction, additive \times additive (AA), additive \times dominance (AD) or dominance \times additive (DA), and dominance \times dominance (DD), based on the extended Kempthorne model.⁴⁹ The contribution rate of the pair-wise SNPs was calculated by the formula: $c = \frac{SS_{SNP1 \times SNP2}}{Var_p}$, where $SS_{SNP1 \times SNP2}$ is the variance of the significant ($P < 0.001$) SNP1 \times SNP2 interactive effect (AA, AD, DA, or DD), and the Var_p is the phenotype variance. An *F*-test was used to test the significance of the two-locus interaction effect. The same FDR method was also used to correct for multiple tests.

3. Results

3.1. Genome-wide identification and characterization of lncRNAs in *P. tomentosa*

The systematic genome-wide identification of *P. tomentosa* lncRNAs was performed by high-throughput RNA-seq of six libraries of secondary vascular tissues (HC, HD, HM, LC, LD, and LM). In total, we identified 15,691 lncRNAs with the number ranging from 6,384 (HC) to 9,732 (LM) across the six sample pools (Fig. 1A). We then characterize the basic genomic features of these lncRNAs. When compared with protein-coding genes, the *Populus* lncRNAs are more evenly distributed with no obvious location preferences across chromosomes (Fig. 1B). According to their genomic locations, these lncRNAs were portioned into 15,402 sense lncRNAs, 160 intergenic lncRNAs, 77 antisense lncRNAs, and 52 intronic lncRNAs (Fig. 1C). The lncRNAs ranged from 201 to 7,862 nt in length with an median of 1,010 nt (Fig. 1D), which were shorter than protein-coding genes of *P. tomentosa* (median length of 1,888 nt). The mean GC content was 41.3%, a little lower than that of protein-coding sequences (42.6%) (Supplementary Material S4). Conservation analysis showed that only 42.4 and 20.2% of *P. tomentosa* lncRNAs are conserved in *Arabidopsis* and rice, respectively (Fig. 1E). Moreover, we identified that only 11 *Populus* lncRNAs could be processed to be 15 mature miRNAs (Fig. 1F; Supplementary Table S2), implying that the majority of the lncRNAs identified here undergo processing by miRNA-independent pathways.

3.2. Expression profiles of *Populus* lncRNAs in secondary vascular tissues

We next explored the expression of the lncRNAs in secondary vascular tissues from the two biomass pools. First, for the maximal expression levels (exp_{max}), we compared expression levels of each transcript among all sequenced samples and found that lncRNAs

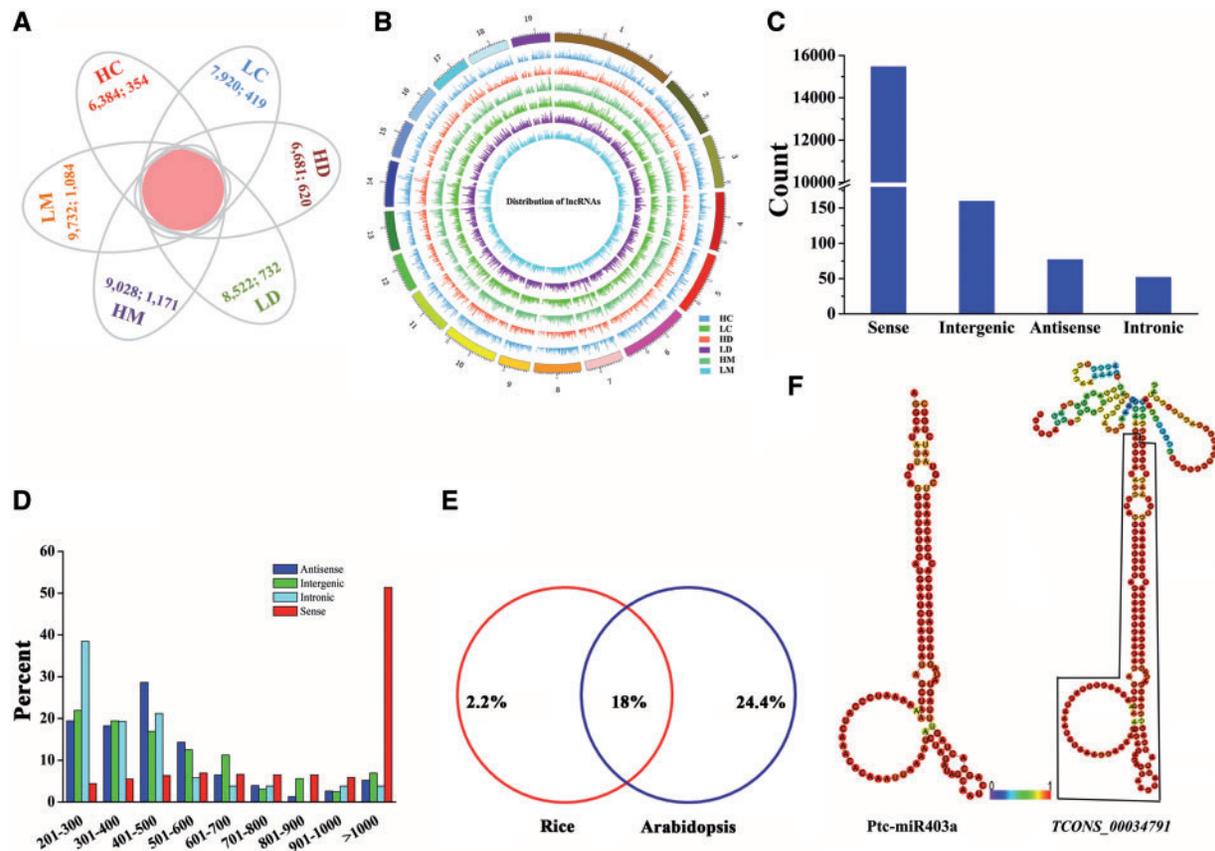


Figure 1. Characteristics of *P. tomentosa* lncRNAs. (A) Venn diagram showing the number of totally and uniquely identified lncRNAs in each sample. (B) Distribution of lncRNAs along each chromosome for each sample generated using Circos. (C) Distribution of four types of lncRNAs, including sense, intergenic, antisense, and intronic lncRNAs. (D) Length distribution of lncRNAs. (E) Sequence conservation of *Populus* lncRNAs in *Arabidopsis* and rice. (F) Predicted secondary structure of the lncRNA *TCONS_00034791* (partial) and the precursor of *ptc-miR403a*. HC, vascular cambium from high-biomass pools; LC, vascular cambium from low-biomass pools; HD, developing xylem from high-biomass pools; LD, developing xylem from low-biomass pools; HM, mature xylem from high-biomass pools; LM, mature xylem from low-biomass pools.

and mRNAs exhibit different density peaks and that the density peaks of mRNAs lag behind those of lncRNAs (Fig. 2A). Indeed, expression categorization showed that the majority of lncRNAs displayed low expression levels, with only 2% highly expressed ($\text{exp}_{\text{max}} > 50$ FPKM) (Fig. 2B). The largest number of *Populus* lncRNAs showed expression peaks in LM (25.6%) followed by HM (24.1%) (Fig. 2C). Then, using the tissue-specific expression index (see ‘Materials and methods’ section), we found 2,746 *Populus* lncRNAs showed tissue-specific expression (Fig. 2D), and more than half of these (1,512 of 2,746) were preferentially expressed in stem xylem (907 in HM and 605 in LM; Fig. 2E). Our qRT-PCR revealed nearly perfect concordance with the RNA-seq results for the expression patterns and verified the expression specificity of the tested lncRNAs (Fig. 2F).

3.3. Functional sequence motifs and PTGs of the DE lncRNAs

Using pair-wise comparisons among the six samples, we identified 1,994 DE lncRNAs (Supplementary Material S5), representing differential expression within or between high- and low-biomass samples (Fig. 3A). To explore whether these lncRNAs have conserved elements, 55 sequence motifs were identified in the DE lncRNAs with biological functions (Supplementary Material S6). Intriguingly,

three of the motifs specifically participate in ‘vascular tissue pattern formation’, ‘plant cell wall’, ‘xylem development’, as well as response to plant hormone stimulus (Fig. 3B). Since lncRNAs play important roles in regulating gene expression, identification and analysis of their target genes may help us to explore their potential functions. Computational prediction identified a set of 5,352 PTGs corresponding to 8,931 lncRNA–target pairs, including 3,569 *cis*-regulated PTGs for 1,812 lncRNAs and 3,297 *trans*-regulated PTGs for 1,824 lncRNAs (Supplementary Material S7).

We then analysed the relationship between expression of the lncRNAs and the PTGs for each of the nine comparisons. Among the 8,931 lncRNA–target pairs, 28% pairs showed the same trend in expression and 6% showed the opposite trend in at least one comparison (Supplementary Materials S5 and S7). Also, for a specific gene pair, lncRNA and the corresponding PTG showed different trends in different comparisons. The non-coherent expression relationship between lncRNAs and their PTGs (*cis* and *trans* targets) were also confirmed by our qRT-PCR results (Fig. 3C). For example, in the comparison of HM vs LM, three lncRNAs (*TCONS_00108374*, *TCONS_00011246*, and *TCONS_00185640*) showed the same expression trend as their potential *cis* and *trans* targets. In addition, *TCONS_00011246* and its *trans* target *LAC2* showed the opposite expression trend in the comparison of HC vs LC, while *TCONS_00108374* showed the same trend as its target *Ptr-IAA27*.

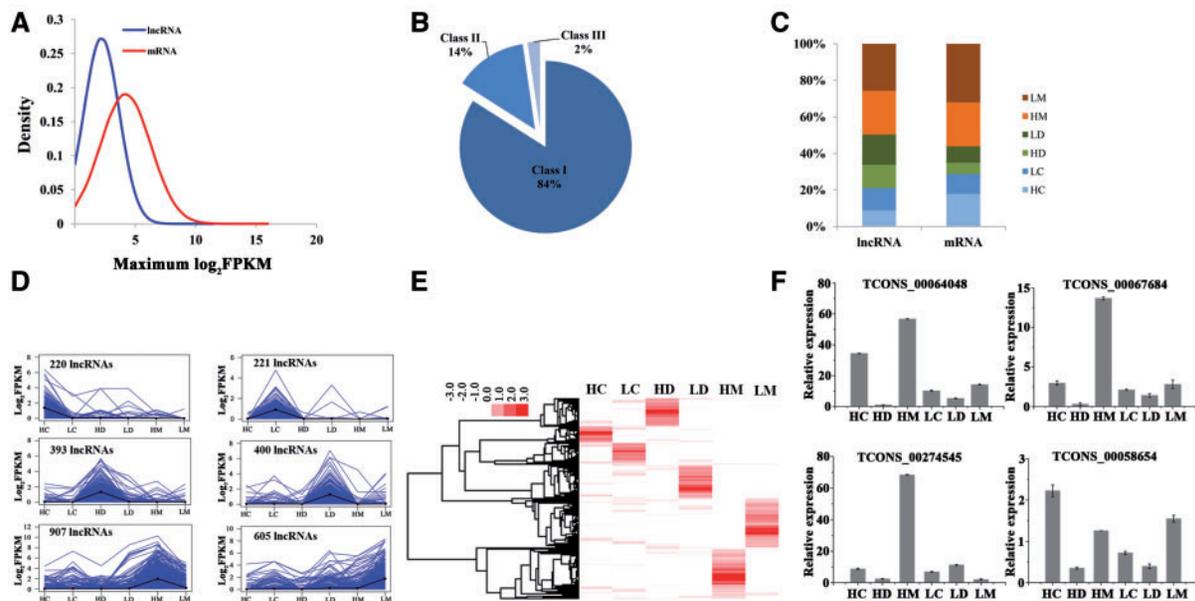


Figure 2. Expression of *Populus* lncRNAs in different vascular tissues of high- and low-biomass pools. (A) Density plot of maximum expression levels of lncRNAs and mRNAs across the six samples (high and low biomass \times vascular cambium, developing xylem, and mature xylem). The expression levels are normalized to \log_2 FPKM. (B) The proportion of lncRNAs showing different expression levels across all samples. The expression levels (FPKM) of lncRNAs can be divided into three classes: Class I ($\text{exp}_{\text{max}} \leq 10$ FPKM), Class II ($\text{exp}_{\text{max}} > 10$ FPKM and ≤ 50 FPKM), Class III ($\text{exp}_{\text{max}} > 50$ FPKM). exp_{max} , the maximum expression levels of lncRNAs across all samples. (C) Tissue distribution of expression peaks of lncRNAs and mRNAs. (D) The tissue-specific lncRNAs. The total numbers of lncRNAs in each group is given. (E) Heat maps of lncRNAs with tissue-specific expression (\log_2 FPKM). Colour intensity represents the fractional density across the row of log-normalized FPKM. (F) Expression validation by qRT-PCR. Error bars represent standard deviations ($n = 3$). HC, vascular cambium from high-biomass pools; LC, vascular cambium from low-biomass pools; HD, developing xylem from high-biomass pools; LD, developing xylem from low-biomass pools; HM, mature xylem from high-biomass pools; LM, mature xylem from low-biomass pools.

GO analysis of these PTGs found that they were representatively enriched in ‘cellulose biosynthetic process’, ‘response to hormone stimulus’, ‘transcriptional activity’, ‘cell wall modifications’, as well as ‘programmed cell death’ (Supplementary Fig. S3). Underlying these GO terms are many developmental regulators of secondary growth, including many transcription factors and genes encoding cell division and expansion proteins, cell wall-associated proteins, and stress-responsive proteins (Supplementary Material S8). Furthermore, we conducted KEGG pathway analysis and revealed that these PTGs were assigned into many pathways concerning secondary metabolite biosynthesis, such as ‘starch and sucrose metabolism’, ‘flavonoid biosynthesis’, and ‘phenylpropanoid biosynthesis’. Among these, 31 PTGs of 26 DE lncRNAs were detected in the phenylpropanoid pathway (Fig. 4A; Supplementary Table S3).

3.4. lncRNAs may interact with miRNAs as potential targets or target mimics of *Populus* miRNAs

Studies have shown that lncRNAs may act as targets or eTMs of miRNAs. In this study, we identified 476 lncRNAs as potential targets of 213 miRNAs from 90 families (Supplementary Material S9) and 28 lncRNAs as eTMs of 14 miRNAs (Supplementary Material S10). Interestingly, in the case of seven miRNAs (ptc-miR169r, ptc-miR169y, ptc-miR171g-5p, ptc-miR408-5p, ptc-miR473a-5p, ptc-miR530a, and ptc-miR6459b), each of the miRNAs not only targeted lncRNAs but also was targeted by eTMs of lncRNAs. For example, ptc-miR169y was predicted to target *TCONS_00237001* (Fig. 5A) and also have an eTM, *TCONS_00013311* (Fig. 5B). Further experimental validation by qRT-PCR revealed a negative

relationship between the expression of miRNAs and their targeting lncRNAs (Fig. 5A and B). Based on these results, we reconstructed an interaction network among miRNAs, lncRNAs, and their PTGs (mRNAs) (Supplementary Fig. S4). However, all the results are predicted preliminarily based on bioinformatic analyses and need to be further validated.

3.5. Genotyping and LD tests for SNPs within lncRNA genes and PTGs in the *P. tomentosa* association population

We selected the most abundant ($\text{exp}_{\text{max}} > 50$ FPKM) lncRNAs genes (see above) and their PTGs, which may play important roles in developmental processes. Among these, 22 PTGs of 16 lncRNAs were enriched in the phenylpropanoid pathway (Fig. 4A). Based on the genome re-sequencing data of 435 *P. tomentosa* individuals from a natural population, we obtained 87,347 SNPs from high-quality genomic sequences of 178 lncRNA genes and 522 PTGs, after SNPs with missing data ($>10\%$) were excluded. The SNPs were not evenly distributed among genes, ranging from 1 to 556, with an average density of 32 SNPs per gene (Supplementary Material S11). SNPs in lncRNA genes had lower diversity ($\pi_T = 0.01804$ and $\theta_w = 0.0454$) than SNPs in PTGs ($\pi_T = 0.03422$ and $\theta_w = 0.06733$). After eliminating SNPs with missing data $> 25\%$ or MAFs < 0.05 in the natural population, we obtained 34,015 common SNPs, including 6,590 SNPs in 178 lncRNA genes and 32,428 SNPs in 522 PTGs (Supplementary Material S12). To perform association analysis for these markers, we first conducted LD tests. The nonlinear regression showed a rapid decay of LD at the chromosome level within an

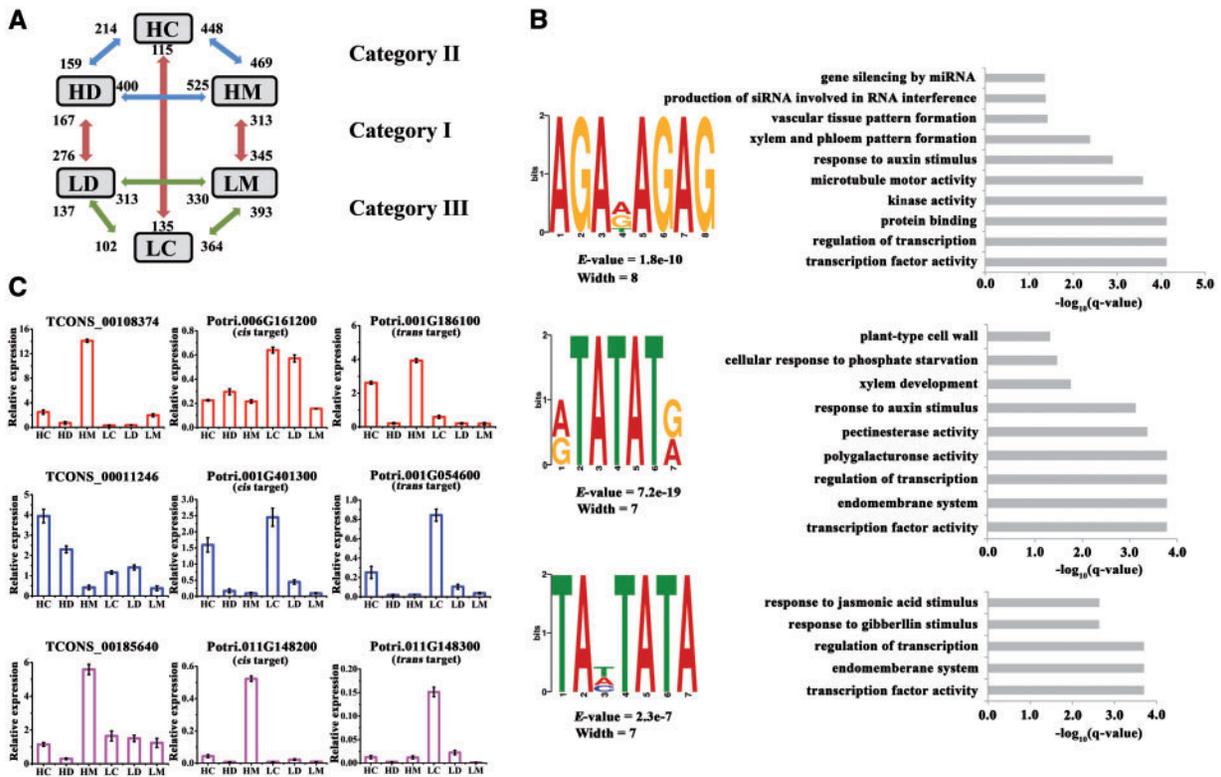


Figure 3. The DE lncRNAs within or between high-biomass and low-biomass pools in the *P. tomentosa* population. (A) Numbers of the DE ($FC \geq 2$ or ≤ 0.5 , $P < 0.01$) lncRNAs among the six samples. Category I (red double-headed arrow) represents the comparison between high- and low-biomass samples. Category II (blue double-headed arrow) and Category III (green double-headed arrow) represent the comparisons among tissue types within the high- and low-biomass samples, respectively. HC, vascular cambium from high-biomass pools; LC, vascular cambium from low-biomass pools; HD, developing xylem from high-biomass pools; LD, developing xylem from low-biomass pools; HM, mature xylem from high-biomass pools; LM, mature xylem from low-biomass pools. (B) Sequence motifs of the DE lncRNAs (left panel) and their corresponding GO terms (right panel). The y-axis of the sequence logo represents information contents in bits. *E*-values for the motifs were calculated by comparison with shuffled sequences. (C) The expression levels of lncRNAs and their potential *cis-trans* targets measured by qRT-PCR. Error bars indicate standard deviations ($n = 3$).

average distance of 2,443 bp, where r^2 values declined to 0.1 (Supplementary Fig. S5).

3.6. Single-SNP association testing for additive and dominant effects

We used association analysis to examine the linkages between phenotypes and SNPs in the lncRNA genes and PTGs. We first conducted SNP-trait association tests between the 34,015 common SNPs and 10 tree growth and wood property traits by using a MLM in TASSEL v5.0. This detected 2,993 significant ($P < 0.001$) associations, representing 1,163 unique SNPs in 83 lncRNA genes and 328 PTGs. The total numbers of significant associations varied across the 10 traits, and the phenotypic variance explained by individual SNPs ranged from 1.2 to 29.3% (Table 1). Among the 1,163 trait-related markers, 329 SNPs from 19 lncRNA genes and 98 PTGs were simultaneously associated with two to three traits (Supplementary Material S13). Interestingly, similar association patterns were identified for SNPs within the same gene. For example, we found that 11 SNPs from *Ptr-PAL*, which functions in phenylpropanoid biosynthesis, were associated with four traits (DBH, HoC, MFA, and V). Two of the SNPs (SNP1108 and SNP1109) simultaneously associated with three common traits (DBH, MFA, and V). Most (~77%)

trait-associated lncRNAs shared at least one common trait with their corresponding PTGs (Fig. 6).

For additive effects, we detected 455 significant associations representing 410 unique SNPs from 249 genes (38 lncRNAs and 211 PTGs) across all 10 traits (Table 2; Supplementary Material S14). We found that 45 of the 410 SNPs were associated with two to three traits, indicating a remarkable pleiotropic effect of these markers. Correspondingly, for each trait, 1–102 genes exhibited additive effects on tree growth and wood properties. For dominant effects, we detected 266 significant associations representing 249 unique SNPs from 187 genes (33 lncRNA genes and 154 PTGs) (Supplementary Material S14). Of these associations, 151 showed a positive dominance effect and 115 had a negative value. For these 249 SNPs, each was associated with one to three traits with positive or negative dominant effects. Correspondingly, around 21% (40 of 187) of the genes associated with more than one trait. Generally, for a specific trait, most of the SNPs from the same gene exhibited concordant dominance. For example, one SNP (SNP_57423) within *TCONS_00141756* and five SNPs from the PTG *Potri.008G056000* were commonly linked to LiC with positive dominance. We also found a small portion of SNPs showing discordant dominance. For instance, two unique SNPs (SNP_4287 and SNP_4431) from the same gene (*Potri.001G030000*) had opposite dominant effects on MFA, suggesting the complex allelic combinations underlying stem

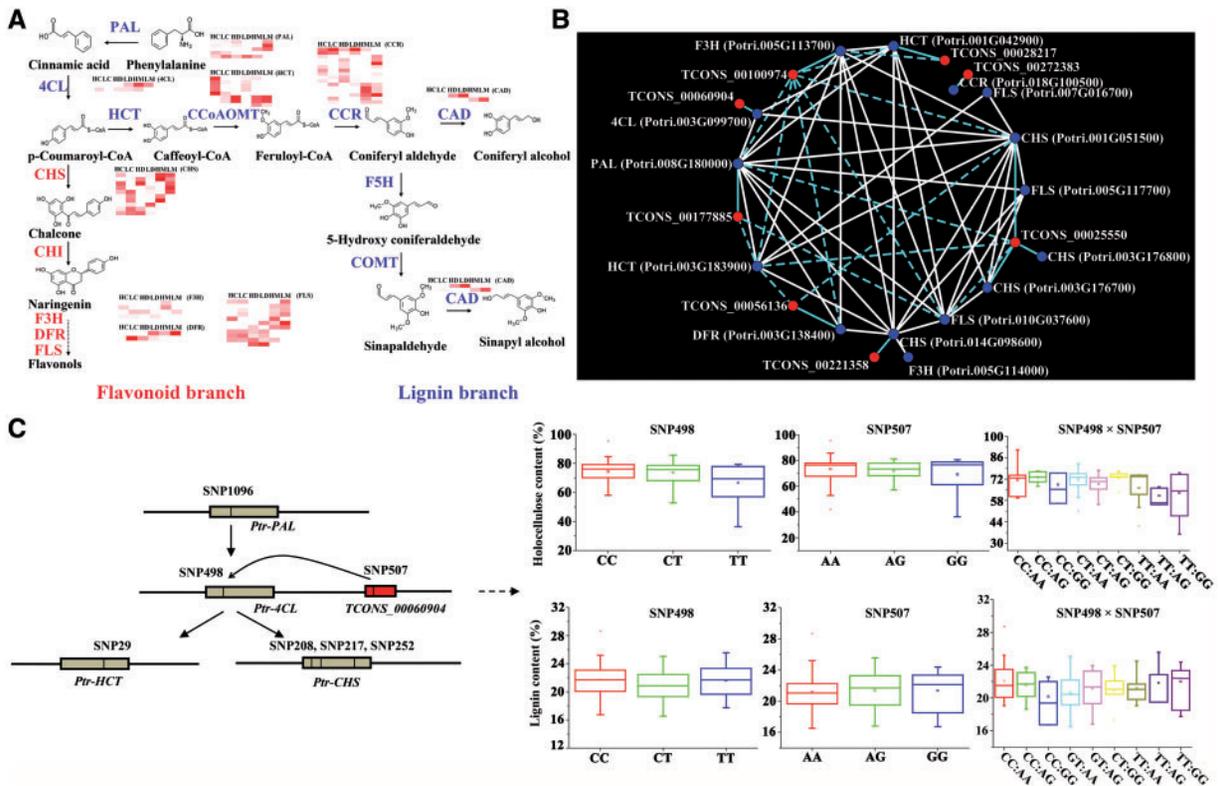


Figure 4. The lncRNAs enriched in the phenylpropanoid pathway exhibited epistatic interactions with their PTGs. (A) Simplified phenylpropanoid pathway. Pathway information was obtained from <http://www.genome.ad.jp/kegg/>. Pathway genes from the lignin branch (right) and flavonoid branch (left) are shown in blue and red, respectively. Relative abundance of these transcripts and their corresponding lncRNAs from the six samples are shown by heatmap based on RNA-seq data. The expression values are normalized to log₂FPKM. The genes encoding the following enzymes are included: phenylalanine ammonium lyase (PAL), 4-coumarate-CoA ligase (4CL), cinnamoyl-CoA reductase, cinnamyl alcohol dehydrogenase, hydroxycinnamoyl-CoA:quinic acid shikimate p-hydroxycinnamoyltransferase (HCT), chalcone synthase (CHS), chalcone isomerase, flavanone 3-hydroxylase, dihydroflavonol 4-reductase, and flavonol synthase. Supplementary Table S3 gives detailed information on these genes. HC, vascular cambium from high-biomass pools; LC, vascular cambium from low-biomass pools; HD, developing xylem from high-biomass pools; LD, developing xylem from low-biomass pools; HM, mature xylem from high-biomass pools; LM, mature xylem from low-biomass pools. (B) The epistatic network of lncRNAs and their PTGs enriched in the phenylpropanoid pathway. This network contains 24 lncRNA–mRNA pairs (light blue lines) and 36 mRNA–mRNA pairs (white lines). Also, pairs of lncRNAs and their corresponding targets are connected by light blue, dashed lines. Eight lncRNAs and 15 PTGs are represented as red and blue circles, respectively. (C) Example of the epistatic network. A SNP (SNP498) within *Ptr-4CL* had epistatic interactions with other SNPs from genes (grey), including an upstream gene (*Ptr-PAL*) and two downstream genes (*Ptr-CHS* and *Ptr-HCT*), and the lncRNA, TCONS_00060904 (red). TCONS_00060904 is 5.23 kb downstream of its potential *cis* target *Ptr-4CL* (left panel). The right panel shows SNP507 in TCONS_00060904 and SNP498 within its potential target *Ptr-4CL* had distinct epistatic effects on HoC and LiC in different genotypic combinations.

growth and wood biosynthesis. Overall, we detected 600 associations with significant ($P < 0.001$) additive or/and dominant effects for the 10 traits, representing 533 unique SNPs from 52 lncRNA genes and 254 PTGs (Table 2). Of these associations, around 20% (121 of 600) have a combination of additive and dominant effects for a certain trait.

3.7. Epistasis analysis of epistatic effects

We next examined epistatic effects of the SNPs to further probe the complexity of this regulatory network. We used EPISNP v2.0 to analyze the pair-wise interactive effects for 34,015 SNPs from lncRNA genes and PTGs across the 10 traits. This detected 17,656 significant ($P < 0.001$) epistatic pairs for all 10 traits, representing 4,334 SNPs from 111 lncRNA genes and 448 PTGs (Table 3; Supplementary Material S15). These 4,334 interactive SNPs, representing a coverage level of 8.6% of the tested markers, did not display detectable additive/dominant effects. These 17,656 pair-wise epistatic interactions could be further interpreted

into 3,918 AA, 8,866 AD or DA, and 5,132 DD effects for the traits (Table 3). The contribution rate of the significant epistatic SNP pairs ranged from 0.01% to 13.63% (Table 3), which was lower than that of single-locus associations, with only 76 epistatic pairs having a contribution rate >1% (Supplementary Material S15).

To improve our understanding of the genetic architecture that affects complex phenotypic variation, we drew the interconnected gene–gene networks for each trait, based on the distribution of the epistatic pairs of SNPs between chromosomes (13,412 pairs) and within chromosomes (3,983 pairs) (Supplementary Fig. S6). In total, 8,898 gene–gene pairs among 111 lncRNA genes and 448 PTGs were detected for all 10 traits (Table 3). These epistatic gene pairs could be categorized into 1,963 lncRNA–lncRNA, 183 lncRNA–mRNA, and 6,653 mRNA–mRNA pairs (Supplementary Fig. S6). Notably, among the 183 lncRNA–mRNA pairs, we found 166 (~91%) lncRNA–target interactions involving 79 lncRNA genes and 146 corresponding PTGs for all 10 traits (Supplementary Fig. S6). Thus this work may provide evidence for possible interactions between the lncRNAs and PTGs.

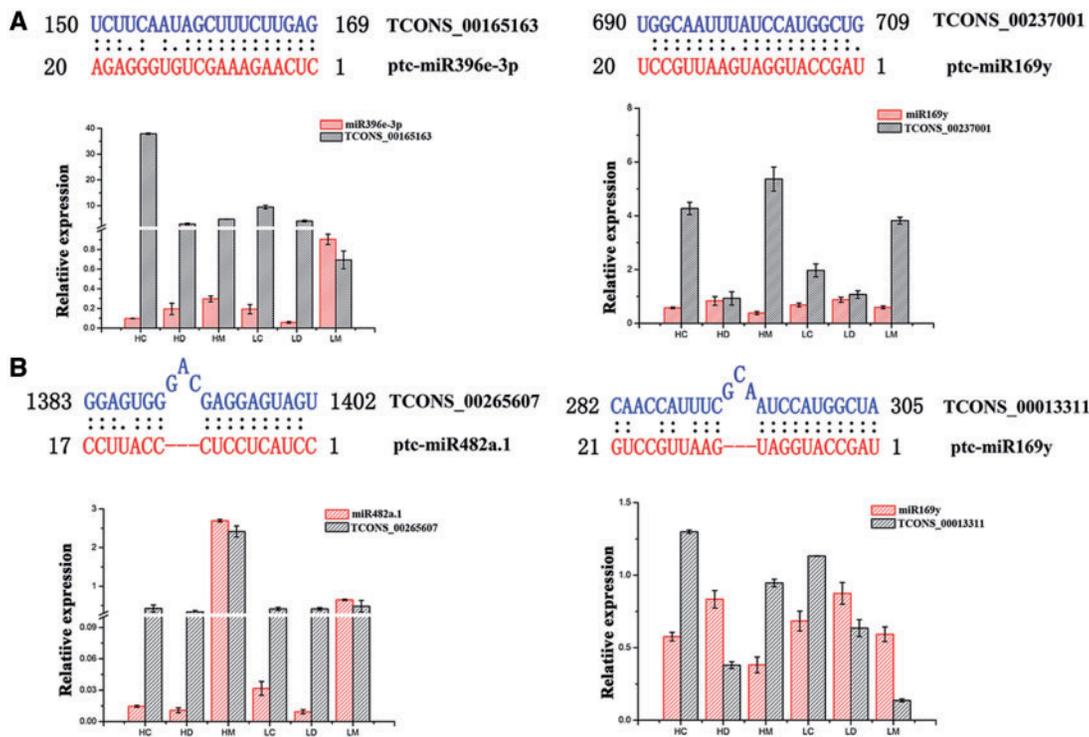


Figure 5. Expression analysis of *Populus* lncRNAs and miRNAs by qRT-PCR. (A) Expression of lncRNAs as potential targets of miRNAs. (B) Expression of lncRNAs as target mimics of miRNAs. Error bars indicate standard deviations ($n = 3$). HC, vascular cambium from high-biomass pools; LC, vascular cambium from low-biomass pools; HD, developing xylem from high-biomass pools; LD, developing xylem from low-biomass pools; HM, mature xylem from high-biomass pools; LM, mature xylem from low-biomass pools.

Table 1. Summary of significant ($P < 0.001$) associations with tree growth and wood properties in the association population of *P. tomentosa* detected by TASSEL v5.0

Trait	Numbers of associations			R^2 (%) ^a
	SNPs	LncRNAs	Targets	
α C (%)	35	6	31	14.15–26.86
DBH (cm)	666	48	189	14.15–29.06
FW (μ m)	61	13	53	14.05–29.02
FL (mm)	10	0	11	14.89–23.87
HemC (%)	29	1	26	14.03–24.90
HoC (%)	78	6	49	1.23–29.25
LiC (%)	10	1	8	14.04–22.01
MFA (degree)	128	23	97	17.13–27.26
V (m^3)	473	33	155	14.23–28.98
H (m)	6	3	4	14.01–23.84
<Total>	1,163	83	328	1.23–29.25

^aPercentage of the phenotypic variance explained.

We then focused on the epistatic interactions among the 16 lncRNAs and 22 PTGs in the phenylpropanoid pathway (Fig. 4A). In total, we detected 60 gene–gene pairs among 8 lncRNA genes and 15 PTGs across all 10 traits, including 36 mRNA–mRNA and 24 lncRNA–mRNA pairs (Fig. 4B). For example, the SNP498 within *Ptr-4CL* has an epistatic effect with SNP1096 from its upstream gene *Ptr-PAL* on V (Fig. 4C). For the downstream genes *Ptr-HCT* and *Ptr-CHS* which were assigned to two different branches of the monolignol biosynthesis pathway, we also detected an epistatic interaction with *Ptr-4CL* for the MFA (Fig. 4C). These 60 pairs included 10 interacting

pairs between lncRNA genes and the corresponding PTGs. For instance, different genotype combinations between SNP507 within *TCONS_00060904* and SNP498 from the PTGs *Ptr-4CL* showed considerable non-additive effects on LiC and HoC, although the two genes displayed no significant additive/dominant effect (Fig. 4C).

4. Discussion

To date, systematic researches for lncRNAs have been conducted in only a few plants. In this study, we identified 15,691 *Populus*

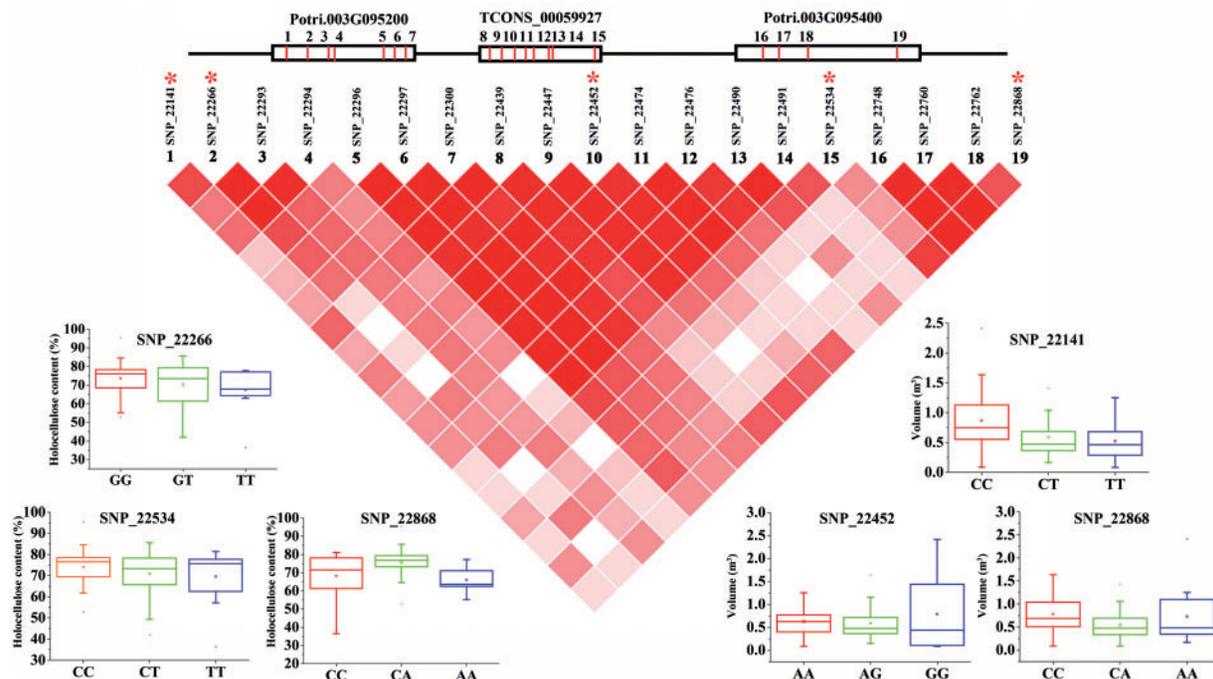


Figure 6. Dissection of the complex genotype effects (additive/dominant) for the lncRNA (TCONS_00059927) and its potential *cis*-acting target genes. Five significant SNPs (SNP_22141, SNP_22266, SNP_22452, SNP_22534, and SNP_22868) within the lncRNA and its PTGs, Potri.003G095200 (upstream) and Potri.003G095400 (downstream) have additive/dominant effects on HoC and stem volume. The two target genes are 1.56 and 6.15 kb from the lncRNA, respectively.

Table 2. The additive and dominant effects of significant ($P < 0.001$) SNPs associated with tree growth and wood properties in the association population of *P. tomentosa* detected by TASSEL v5.0

Trait		Additive effect			Dominant effect		
		Numbers of SNPs		Range of effect	Numbers of SNPs		Range of effect
		LncRNAs	Targets		LncRNAs	Targets	
Tree growth	DBH (cm)	9	62	2.38–8.47	5	19	–10.91–11.72
	H (m)	1	0	1.98	1	5	–3.42 to –2.03
	V (m ³)	25	133	0.16–0.95	12	67	–1.06 to 1.55
Wood properties	α C (%)	1	25	5.01–14.36	2	23	–23.65 to –19.64
	FL (mm)	1	10	0.06–0.16	1	7	–0.34 to 0.18
	FW (μ m)	4	15	1.40–4.12	2	10	1.60–5.30
	HemC (%)	2	26	6.81–20.71	2	14	–17.24 to 15.52
	HoC (%)	8	49	5.62–21.00	6	27	–32.75 to 20.77
	LiC (%)	1	6	1.67–5.52	2	3	–4.23 to 3.85
	MFA (degree)	24	105	3.55–12.77	14	80	–17.88 to 10.67
<Total>		69	390	0.06–21.00	44	239	–32.75 to 20.77

lncRNAs with a focus on secondary growth and wood biosynthesis (Fig. 1). Although there are some limitations for the definition of lncRNAs that merely rely on transcript size and coding potential, a relatively robust and reliable list of *Populus* lncRNAs is provided here for the relatively strict bioinformatics criteria used for the definition as lncRNAs. These lncRNAs shared many features of other species;^{18,50,51} they were shorter in length and less conserved in sequence compared with protein-coding transcripts, and expressed in a tissue-specific manner (Figs 1 and 2). These findings suggested that these *Populus* lncRNAs may exhibit a biological function, rather

than merely being transcriptional noise. Thus, this work may provide new insight into the study of secondary growth of plants.

4.1. Prediction of potential targets and functional motifs of *Populus* lncRNAs revealed that they may be involved in wood biosynthesis

Functional characterization of lncRNAs remains at an early stage. The commonly used methods for lncRNA functional prediction are based on protein binding,⁵² epigenetic modification,⁵³ co-expression

Table 3. The epistatic effect of significant ($P < 0.001$) SNPs associated with tree growth and wood properties in the association population of *P. tomentosa* detected by EPISNP v2.0

Trait	No. of epistatic SNP–SNP pairs ^a			No. of interacting SNPs	No. of epistatic gene–gene pairs	No. of interactive genes		Range of contribution rate (c , %)
	AA	AD (DA)	DD			lncRNA	target	
DBH (cm)	404	473	194	897	790	89	249	0.03–0.83
H (m)	107	240	104	548	418	71	176	0.03–1.00
V (m ³)	1,230	2,217	1,380	1,670	3,221	97	359	0.03–12.95
α C (%)	448	751	455	1,329	1,340	89	302	0.01–0.91
FL (mm)	208	505	324	860	806	72	257	0.01–13.63
FW (μ m)	242	1,025	657	1,309	1,374	88	307	0.01–0.67
HemC (%)	359	998	487	1,375	1,461	91	299	0.02–3.78
HoC (%)	601	1,088	656	1,441	1,644	96	308	0.01–0.78
LiC (%)	142	332	167	694	563	77	208	0.01–1.06
MFA (degree)	495	1,784	996	1,599	2,249	95	329	0.03–2.28
<Total>	3,918	8,866	5,132	4,334	8,898	111	448	0.01–13.63

No., number.

^aTotal numbers of significant ($P < 0.001$) SNP–SNP pairs with different types of epistatic effects. The SNP–SNP epistatic effects can be interpreted into four types: AA, additive \times additive effect; AD, additive \times dominant effect; DA, dominant \times additive effect; and DD, dominant \times dominant effect.

networks,⁵⁴ and miRNA regulation.⁵⁵ Here, we employed a target prediction programme for *Populus* lncRNAs based on adjacent gene functions (*cis*-regulation) and sequence complementarity (*trans*-regulation). A total of 1,994 (12.6%) lncRNAs were DE between/within high-biomass and low-biomass pools (Fig. 3A), comparable to protein-coding genes (8,573) under the same differential expression threshold. This comparison supports the idea that the lncRNAs have a biological purpose and are subject to complex transcriptional regulation during wood biosynthesis. Interestingly, lncRNAs in mature xylem exhibited more (2- to 3-fold) differential expression compared with vascular cambium and developing xylem from the high- and low-biomass pools, although approximately evenly split between genes that were differentially up- and down-regulated in each pair-wise comparison (Fig. 3A). This may suggest that the regulation of lncRNAs is subject to transcriptional rearrangement during wood biosynthesis and more lncRNAs are involved in xylem development.

The importance of transcriptional regulation is illustrated by the coordinated expression of specific classes of genes in related biological processes occurring at specific stages of wood formation, which initiates in the vascular cambium and involves cell division, expansion, secondary wall formation and lignification in developing xylem, and finally, programmed cell death to form mature xylem.⁵⁶ Supporting this, using a target prediction programme for the DE lncRNAs, we found many important PTGs encoding cell division and expansion proteins, cell wall-associated proteins, and stress-responsive proteins, as well as many transcription factors, including 31 PTGs in the phenylpropanoid pathway. GO analysis showed that the PTGs of lncRNAs highly expressed in mature xylem were enriched in many biological processes related to programmed cell death; however, the PTGs of lncRNAs highly expressed in vascular cambium and developing xylem were assigned to many GO terms, such as ‘carbohydrate metabolic process’, ‘polysaccharide biosynthetic process’, ‘cell morphogenesis’, and ‘response to hormone stimulus’ (Supplementary Fig. S3). Thus, our observation of the activation of lncRNAs in mature xylem and relatively little difference between vascular cambium and developing xylem is expected. Additionally, we found that the PTGs of lncRNAs DE between these two pools were significantly enriched in many represented biological

processes, such as ‘carbohydrate metabolic process’, ‘oxidation reduction’, ‘response to hormone stimulus’, and ‘regulation of protein stability’ (Supplementary Fig. S3), indicating that the coordinated regulation of lncRNAs and their potential targets in these biological processes may be responsible for the differentiation between these two biomass pools.

Noteworthy, all the target genes of lncRNAs predicted in our study are potential targets and need further experimental confirmation in future studies. Although the functions of lncRNAs remain largely unknown, previous studies have reported that lncRNAs may participate in various biological processes without the need to be mutually exclusive. Given the complex functions of lncRNAs and the limited strategies that merely rely on *cis/trans*-regulation for the prediction of lncRNA targets in this study, more target genes likely remain to be explored and other approaches for prediction of target genes for lncRNAs need to be applied.

4.2. *Populus* lncRNAs may play roles in miRNA-mediated regulatory networks in secondary wall formation

MiRNAs function to destabilize mRNAs and repress the translation of protein-coding genes and play essential roles in plant development and physiology.⁵⁷ Moreover, interesting cross-regulation between miRNAs and lncRNAs has recently become apparent,⁵⁸ implying that lncRNAs may interact with miRNAs apart from protein-coding genes. Thus, the intriguing mechanism of lncRNA–miRNA crosstalk may reveal a new layer of functions of lncRNAs. Indeed, in the present study, we identified 476 lncRNAs as potential targets of 213 miRNAs from 90 families (Supplementary Material S9). Among these miRNA families targeting lncRNAs, the stress-responsive miRNA, ptc-miR473a, which plays an important role in biosynthesis of cell wall metabolites and plant development,⁵⁹ exhibits near-perfect complementarity to its target *TCONS_00108161*. Also, we found that *TCONS_00078539* was predicted to be the potential target of miR168, which could be involved in auxin signaling and secondary cell wall biosynthesis.^{60,61} In addition, ptr-miR397, which could be a negative regulator of laccase genes affecting LiC,⁶² was predicted to target three lncRNAs (*TCONS_00013182*,

TCONS_00015036, and *TCONS_00028534*). These results indicated that *Populus* lncRNAs may play roles in secondary cell wall biosynthesis and wood formation as potential targets of many regulatory miRNAs.

Target mimicry is a recently identified mechanism of miRNA regulation initially studied in *Arabidopsis*,⁷ and subsequently in mammals;⁶³ lncRNAs can act as miRNA decoys or sponges to sequester miRNAs, thus favouring the expression of repressed target mRNAs. Indeed, many eTMs have been functionally confirmed in *Arabidopsis* and rice.³² In our study, 28 lncRNAs were predicted to be eTMs of 14 miRNAs (Supplementary Material S10). Interestingly, consistent with the previous studies,^{51,64} we identified novel eTMs for four miRNAs (miR160, miR164, miR169, and miR482a.1). This may indicate that eTM-mediated regulation for these four miRNAs may be pervasive in plants. For example, miR164, which may affect lignin metabolism by regulating transcription factors that control secondary growth and wood composition,⁶⁵ was predicted to be absorbed by five eTMs. Additionally, in some cases, like miR408, which is involved in lignin deposition and polymerization,⁵⁹ might target seven lncRNAs and be absorbed by three eTMs. Therefore, based on these analyses, we infer that these lncRNAs potentially function via interacting with miRNAs in complicated networks involving interactions among lncRNAs, mRNAs, and miRNAs (Supplementary Fig. S4). However, all the findings are mainly based on our bioinformatics prediction, and the specific regulatory mechanism requires further investigation and validation, such as the cellular localization of miRNAs and lncRNAs predicted as miRNA targets and eTMs using *in situ* hybridization. We believe that the emergence of important roles for lncRNAs in secondary cell wall biosynthesis and tree growth and our investigation of lncRNA–miRNA interaction network may provide an important resource for further studies.

4.3. Allelic variations of *Populus* lncRNAs and their PTGs indicated that they were associated with growth and wood properties

Association mapping provides a valuable tool allowing us to identify the natural allelic variation responsible for a particular phenotype.⁶⁶ In this study, we performed association studies for lncRNAs and PTGs, integrating additive, dominant, and epistatic effects of allelic variations underlying wood biosynthesis and stem biomass. Thus, this study initiated the dissection of potential genetic regulatory mechanisms involving lncRNAs and their PTGs. In total, 1,163 unique SNPs from lncRNAs and their PTGs were significantly associated with growth and wood properties based on single-SNP associations (Table 1). Also, ~28% (329 of 1,163) SNPs were associated with at least two traits and each trait could be associated with 6 SNPs (H) to 666 SNPs (DBH) (Table 1), consistent with a previous study that genetic association studies often found that a locus associated with multiple, sometimes seemingly distinct traits. This observation also highlights the possibility that traits share common genetic pathways and underscores the relevance of pleiotropy.⁶⁷ Intriguingly, ~77% of the lncRNAs and their corresponding PTGs associated with at least one trait in common. For example, we found that SNPs within the lncRNA *TCONS_00059927* and its potential *cis* PTGs simultaneously associated with HoC and V under additive/dominant effects (Fig. 6), indicating the potential genetic interaction between the lncRNA and its corresponding target genes in the pathways underlying this complex genetic network.

In quantitative genetics, epistasis, defined as the interaction between genotypes at two or more loci, is an important component of the genetic basis of complex traits and can also involve

modification of the additive or dominant effects of the interacting loci. Therefore, phenotypic variations cannot be predicted simply by summing the effects of individual loci.⁶⁸ At the gene level, epistasis is a phenomenon whereby the effects of a given gene on a biological trait are masked or enhanced by one or more other genes.⁶⁹ Previous studies have documented that epistasis analysis has considerable potential to reveal interactions between genes and pathways.⁷⁰ In this study, we detected 17,656 significant SNP–SNP epistatic pairs, covering 8.6% of all investigated SNPs, associated with tree growth and wood properties (Table 3). Interestingly, we did not observe significant additive and/or dominant effects within these epistatic pairs. The result revealed pervasive epistatic effects among allelic mutations without significant main effects in the natural population of *Populus*, supporting the idea that epistasis plays a key role in genetic architecture and also providing essential information for a deeper understanding of the mechanisms of gene interactions.⁶⁸

Examining a specific pathway, we found an epistatic interaction network among 8 lncRNAs and 15 PTGs involved in the phenylpropanoid pathway (Fig. 4B). Among the 15 target genes, we detected 36 epistatic pairs across all 10 traits (Fig. 4B). Consistent with the steps in the pathway, we found *Ptr-4CL* had epistatic interactions with its upstream gene *Ptr-PAL*, which encodes the first enzyme of the pathway and also the two downstream genes *Ptr-HCT* and *Ptr-CHS*, which encode enzymes assigned to the two different branches in the pathway (Fig. 4C). In addition, we identified epistatic interactions between genes where the biological evidence for how they interacted had not yet been found (Fig. 4C). Intriguingly, our results are consistent with studies in vertebrates and sponges, where lncRNAs have been shown to be highly interconnected with multiple protein-coding genes based on expression profiles.^{2,71} In the epistatic network, 24 interacting pairs between 8 lncRNAs and 12 genes showed that the lncRNAs can have epistatic interactions with their corresponding target genes and also interact with other non-target genes predicted by our target prediction programme, suggesting that the lncRNAs may interact with their target genes by different mechanisms in the pathway. However, the precise mechanisms underlying such gene–gene interactions are largely unknown. Our findings therefore may provide a reference for follow-up studies to further explore this hypothesis. Collectively, our epistasis analysis of lncRNAs and their potential target genes may enrich the interaction networks among genes in the pathway, since these genes may be part of a common regulatory network.

Accession number

SRA accession: SRP073689.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online.

Funding

This work was supported by the State ‘13.5’ Key Research Programme of China (No. 2016YFD0600102), and the Project of the National Natural Science Foundation of China (No. 31670333).

References

- Ma, L., Bajic, V. B. and Zhang, Z. 2013, On the classification of long non-coding RNAs, *RNA Bio.*, **10**, 924–33.
- Necsulea, A., Soumillon, M., Warnefors, M., et al. 2014, The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, **505**, 635–40.
- Cabili, M.N., Trapnell, C., Goff, L., et al. 2011, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–27.
- Ponting, C.P., Oliver, P.L. and Reik, W. 2009, Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–41.
- Swiezewski, S., Liu, F., Magusin, A. and Dean, C. 2009, Cold-induced silencing by long antisense transcripts of an Arabidopsis polycomb target. *Nature*, **462**, 799–802.
- Heo, J.B. and Sung, S. 2011, Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**, 76–9.
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., et al. 2007, Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, **39**, 1033–7.
- Demura, T. and Ye, Z.-H. 2010, Regulation of plant biomass production. *Curr. Opin. Plant Biol.*, **13**, 298–303.
- Shi, R., Sun, Y.-H., Li, Q., Heber, S., Sederoff, R. and Chiang, V.L. 2010, Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.*, **51**, 144–63.
- Shastry, B.S. 2009, SNPs: impact on gene function and phenotype. In: *Single Nucleotide Polymorphisms*, Berlin: Springer, pp. 3–22.
- Tang, J.-Y., Lee, J.-C., Chang, Y.-T., et al. 2013, Long noncoding RNAs-related diseases, cancers, and drugs. *Scientific World J.*, **2013**: 943539.
- Ding, J., Lu, Q., Ouyang, Y., et al. 2012, A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 2654–9.
- Hindorf, L.A., Sethupathy, P., Junkins, H.A., et al. 2009, Potential etiological and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 9362–7.
- Kumar, V., Westra, H.-J., Karjalainen, J., et al. 2013, Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.*, **9**, e1003201.
- Dalessandro, G. and Northcote, D. 1977, Changes in enzymic activities of nucleoside diphosphate sugar interconversions during differentiation of cambium to xylem in sycamore and poplar. *Bio. J.*, **162**, 267–79.
- Tuskan, G.A., Difazio, S., Jansson, S., et al. 2006, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596–1604.
- Trapnell, C., Roberts, A., Goff, L., et al. 2012, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Prot.*, **7**, 562–78.
- Li, L., Eichten, S.R., Shimizu, R., et al. 2014, Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.*, **15**, R40.
- Sun, L., Zhang, Z., Bailey, T.L., et al. 2012, Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study. *BMC Bioinformatics*, **13**, 1.
- Finn, R.D., Clements, J. and Eddy, S.R. 2011, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–37.
- Kong, L., Zhang, Y., Ye, Z.-Q., et al. 2007, CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.*, **35**, W345–9.
- Sun, L., Luo, H., Bu, D., et al. 2013, Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.*, **41**, e166.
- Roberts, A., Pimentel, H., Trapnell, C. and Pachter, L. 2011, Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*, **27**, 2325–9.
- Lamesch, P., Berardini, T.Z., Li, D., et al. 2012, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–10.
- Ouyang, S., Zhu, W., Hamilton, J., et al. 2007, The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.*, **35**, D883–7.
- Bailey, T.L. 2011, DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, **27**, 1653–9.
- Buske, F.A., Bodén, M., Bauer, D.C. and Bailey, T.L. 2010, Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics*, **26**, 860–6.
- Yanai, I., Benjamin, H., Shmoish, M., et al. 2005, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–9.
- Jia, H., Osak, M., Bogu, G.K., Stanton, L.W., Johnson, R. and Lipovich, L. 2010, Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA*, **16**, 1478–87.
- Han, L., Zhang, K., Shi, Z., et al. 2012, LncRNA profile of glioblastoma reveals the potential role of lncRNAs in contributing to glioblastoma pathogenesis. *Int. J. Oncol.*, **40**, 2004–12.
- Tafer, H. and Hofacker, I.L. 2008, RNAplex: a fast tool for RNA–RNA interaction search. *Bioinformatics*, **24**, 2657–63.
- Wu, H.-J., Wang, Z.-M., Wang, M. and Wang, X.-J. 2013, Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol.*, **161**, 1875–84.
- Saito, R., Smoot, M.E., Ono, K., et al. 2012, A travel guide to Cytoscape plugins. *Nat. Methods*, **9**, 1069–76.
- Livak, K.J. and Schmittgen, T.D. 2001, Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *Methods*, **25**, 402–8.
- Du, Q., Xu, B., Pan, W., et al. 2013, Allelic variation in a cellulose synthase gene (ProCesA4) associated with growth and wood properties in *Populus tomentosa*. *G3*, **3**, 2069–84.
- Du, Q., Xu, B., Gong, C., et al. 2014, Variation in growth, leaf, and wood property traits of Chinese white poplar (*Populus tomentosa*), a major industrial tree species in Northern China. *Can. J. For. Res.*, **44**, 326–339.
- Du, Q., Gong, C., Wang, Q., et al. 2016, Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol.*, **209**, 1067–1082.
- Li, R., Li, Y., Kristiansen, K. and Wang, J. 2008, SOAP: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713–4.
- Danecek, P., Auton, A., Abecasis, G., et al. 2011, The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–8.
- Nei, M. 1987, *Molecular Evolutionary Genetics*. Columbia University Press.
- Watterson, G. 1975, On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, **7**, 256–276.
- Tajima, F. 1989, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–95.
- Yu, J., Pressoir, G., Briggs, W.H., et al. 2006, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, **38**, 203–8.
- Du, Q., Wang, B., Wei, Z., Zhang, D. and Li, B. 2012, Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered.*, **103**, 853–62.
- Hardy, O.J. and Vekemans, X. 2002, SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes*, **2**, 618–20.
- Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. 2009, Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.*, **9**, 1322–32.
- Storey, J.D. and Tibshirani, R. 2003, Statistical significance for genome-wide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–5.
- Ma, L., Runesha, H.B., Dvorkin, D., Garbe, J.R. and Da, Y. 2008, Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. *BMC Bioinformatics*, **9**, 315.
- Mao, Y., London, N.R., Ma, L., Dvorkin, D. and Da, Y. 2007, Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. *Physiol. Genomics*, **28**, 46–52.

50. Liu, J., Jung, C., Xu, J., et al. 2012, Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell*, **24**, 4333–4345.
51. Zhang, Y.-C., Liao, J.-Y., Li, Z.-Y., et al. 2014, Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.*, **15**, 512.
52. Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H. and Qu, L.-H. 2013, ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.*, **41**, D177–87.
53. Sati, S., Ghosh, S., Jain, V., Scaria, V. and Sengupta, S. 2012, Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.*, **40**, 10018–31.
54. Guttman, M., Amit, I., Garber, M., et al. 2009, Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–7.
55. Keniry, A., Oxley, D., Monnier, P., et al. 2012, The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nature Cell Biol.*, **14**, 659–65.
56. Hertzberg, M., Aspeborg, H., Schrader, J., et al. 2001, A transcriptional roadmap to wood formation. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 14732–37.
57. Mallory, A.C. and Vaucheret, H. 2006, Functions of microRNAs and related small RNAs in plants. *Nat. Genet.*, **38**, S31–6.
58. Yoon, J.-H., Abdelmohsen, K. and Gorospe, M. 2014, Functional interactions among microRNAs and long noncoding RNAs. In: *Seminars in Cell and Developmental Biology*. Academic Press, **34**, 9–14.
59. Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L. and Chiang, V.L. 2005, Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from Arabidopsis. *Plant Cell*, **17**, 2186–203.
60. Liu, Q. and Chen, Y.-Q. 2009, Insights into the mechanism of plant development: interactions of miRNAs pathway with phytohormone response. *Biochem. Biophys. Res. Co.*, **384**, 1–5.
61. Ong, S.S. and Wickneswari, R. 2012, Characterization of microRNAs expressed during secondary wall biosynthesis in *Acacia mangium*. *PLoS One*, **7**, e49662.
62. Lu, S., Li, Q., Wei, H., et al. 2013, Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA*, **110**, 10848–53.
63. Tay, Y., Kats, L., Salmena, L., et al. 2011, Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*, **147**, 344–57.
64. Shuai, P., Liang, D., Tang, S., et al. 2014, Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.*, **65**, 4975–83.
65. Ong, S.S. and Wickneswari, R. 2011, Expression profile of small RNAs in *Acacia mangium* secondary xylem tissue with contrasting lignin content-potential regulatory sequences in monolignol biosynthetic pathway. *BMC Genomics*, **12**, S13.
66. Thumma, B.R., Nolan, M.F., Evans, R. and Moran, G.F. 2005, Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics*, **171**, 1257–65.
67. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. and Smoller, J.W. 2013, Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.*, **14**, 483–95.
68. Mackay, T.F. 2014, Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.*, **15**, 22–33.
69. Moore, J.H. 2005, A global view of epistasis. *Nat. Genet.*, **37**, 13–4.
70. Phillips, P.C. 2008, Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.*, **9**, 855–67.
71. Gaiti, F., Fernandez-Valverde, S.L., Nakanishi, N., et al. 2015, Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. *Mol. Biol. Evol.*, **32**, 2367–82.