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\(^1,2,3,*\)

\(^1\)Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P.R. China, \(^2\)National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, P.R. China, and \(^3\)Key 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. In contrast to protein-coding genes, most lncRNAs lack strong sequence conservation between species. LncRNAs are usually expressed at low levels and often exhibit tissue-specific expression. 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 lncRNA ‘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 molecular mechanisms underlying plant’s secondary growth are mainly focused on functional studies of protein-coding genes, such as the genes identified in the phenylpropanoid pathway.9 Nevertheless, few studies have systematically 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 systematically 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 ≤ 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 TrueSeq 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 v2.1.17 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 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 HMMER3.0 to eliminate transcripts with potential protein-coding ability (cutoff E-value ≤ 0.001). Next, we employed the Coding Potential Calculator (CPC) software and Coding–non-coding Index (CNCI) software 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 IncRNAs. IncRNAs were classified into four categories (‘i’, ‘o’, ‘u’, and ‘x’), consisting of IncRNAs without overlap with any genes (intergenic IncRNAs, ‘u’), antisense overlapping IncRNAs (antisense IncRNAs, ‘x’), intronic overlapping IncRNAs (intronic IncRNAs, ‘i’) and generic exonic overlapping IncRNAs (sense IncRNAs, ‘o’), according to their genomic locations relative to neighbouring genes as defined by cuffcompare programme in Cufflinks suite. The difference in IncRNA expression was calculated as the fold change (FC) = FPKM of ‘sample1’/FPKM of ‘sample2’. IncRNA transcripts were considered to be differentially expressed (DE) if they meet the criteria of FC ≥ 2 or ≤ 0.5 with P-value < 0.01.

2.4. Genomic characterization and specific expression of Populus IncRNAs

GC contents of the identified IncRNAs were calculated with EMBoss explorer’s gecce tool (http://emboss.sourceforge.net/apps/cvs/emboss/apps/gecece.html). To explore IncRNA conservation, all the IncRNA sequences identified here were aligned with BLASTN against the genome sequences of A. thaliana and rice with a cut-off E-value < 1e-10. The genomes were downloaded from Phytozome (v9.1) (http://www.phytozome.net/). The IncRNAs that had >20% of their sequence matched to other genomes were defined as conserved IncRNAs. Motifs in IncRNAs 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 sever with q-value < 0.05. The tissue specificity of IncRNA 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. The index was calculated as: tissue-specific index = \sum_{i=1}^{n} \left( \frac{1}{\exp_{max}} \right) \cdot \frac{\exp_{i}}{n}, where n is the number of tissues; \exp_{i} is the expression value of each IncRNA in tissue, i; and \exp_{max} is the maximum expression value of each IncRNA among all tissues. Only the IncRNAs showing a tissue-specific index > 0.9 were considered to be tissue-specific.

2.5. Prediction of PTGs of Populus IncRNAs

Two independent algorithms were then used to predict PTGs of the DE IncRNAs based on the regulatory effects of the IncRNA (cis- or trans-acting). For the first algorithm, genes transcribed within a 10 kb window upstream or downstream of each IncRNA were considered as potential cis target genes by using genome annotation and a genome browser, as described by Jia et al. 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 IncRNA binding on mRNA molecules, as described previously. Specifically, the algorithm first employed BLAST to select target sequences complementary to the IncRNAs with parameters of e-value < 1e-5 and identity ≥ 95% and then used RNAplex to calculate the complementary energy between two sequences for further screening and selection of potential trans target genes. 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 IncRNAs as potential precursors, targets, and target mimics of miRNAs

To explore whether some IncRNAs 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 IncRNAs and secondary structure prediction was then executed using the Vienna RNA package RNAfold web (http://rna.albany.edu). IncRNAs with classic stem-loop hairpins were regarded as putative precursors of miRNAs. Furthermore, to investigate whether some IncRNAs 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 cTMs from Populus IncRNAs were predicted with local scripts by using the algorithm developed by Wu et al. Quiet different from the authentic targets that are nearly perfectly complementary to miRNAs, cTMs 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 IncRNAs was modelled with Cytoscape 3.2.

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-ΔΔCt or 2-ΔΔACt method.

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. These trees were selected from the collection of 1,047 native individuals collected from the entire
natural distribution of *Populus 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.* The wood property traits were microfibril angle (MFA), fibre length (FL), fibre width (FW), holocellulose content (HOC), hemicellulose content (HemC), 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 (<50% of nucleotides with a quality score < Q20). Then the paired-end short reads were aligned and mapped to the *Populus* reference genome using SOAPaligner/SOAP2 v2.20 with default options. 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%. Then the SNPs among the association population were mined using the Variant Call Format tool (version 4.1). BLASTN and BLASTX were employed with a cutoff E-value < 1e-10 to obtain information in the location of the IncRNAs genes that represent DNA sequences encoding the IncRNAs and their PTGs, including promoter regions (2,000 bp upstream), or gene fragments, respectively. The genomic DNA sequences of the PTGs of the *Populus tomentosa* IncRNAs have been deposited in GenBank under the Accession Number XX155245-XX155490 (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, \( \pi \), and the number of segregating sites, \( \theta_{w} \), were used for calculation of SNP diversity using TASSEL v.5.0 (http://www.maizegenetics.net/). Also, Tajima’s \( D^{*} \) 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 IncRNAs 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. Based on 20 genomic simple sequence repeat markers described previously, the pair-wise kinship coefficients \( K \) were evaluated by SPAGeDi 1.3. and the association population structure \( \Theta \) was obtained based on significant subpopulations \( (k = 3) \) using STRUCTURE 2.3.4. 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. A \( q \)-value of 0.10 was considered as the significance threshold.

Epistasis analysis: EPISNP v2.0 (http://animalgene.umn.edu/epis npmpi/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. It incorporated tests of five epistasis effects for each pair of SNPs, including two-locus interaction, additive × additive (AA), additive × dominance (AD) or dominance × additive (DA), and dominance × dominance (DD), based on the extended Kemphorne model. The contribution rate of the pair-wise SNPs was calculated by the formula: \( c = \frac{\text{SNP1} \times \text{SNP2}}{\text{Var} \text{p}} \), where \( \text{Var} \text{p} \) is the variance of the significant \( (P < 0.001) \) SNP1 × SNP2 interactive effect (AA, AD, DA, or DD), and the \( \text{Var} \text{p} \) 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 IncRNAs in *Populus tomentosa*

The systematic genome-wide identification of *Populus tomentosa* IncRNAs 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 IncRNAs 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 IncRNAs. When compared with protein-coding genes, the *Populus* IncRNAs are more evenly distributed with no obvious location preferences across chromosomes (Fig. 1B). According to their genomic locations, these IncRNAs were portioned into 15,402 sense IncRNAs, 160 intergenic IncRNAs, 77 antisense IncRNAs, and 52 intronic IncRNAs (Fig. 1C). The IncRNAs 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 *Populus 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* IncRNAs are conserved in *Arabidopsis* and rice, respectively (Fig. 1E). Moreover, we identified that only 11 *Populus* IncRNAs could be processed to be 15 mature miRNAs (Fig. 1F; Supplementary Table S2), implying that the majority of the IncRNAs identified here undergo processing by miRNA-independent pathways.

#### 3.2. Expression profiles of *Populus* IncRNAs in secondary vascular tissues

We next explored the expression of the IncRNAs in secondary vascular tissues from the two biomass pools. First, for the maximal expression levels (\( \text{exp}_{\text{max}} \)), we compared expression levels of each transcript among all sequenced samples and found that IncRNAs...
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 (exp_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.
GO analysis of these PTGs found that they were representatively enriched in 'cellulose biosynthetic process', 'response to hormone stimulus', 'transcriptional activity', 'cell wall modifications', and '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 (expmax > 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 (>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
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 \( Pttr-PAL \), 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
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). 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 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 log$_{2}$FPKM. The genes encoding the following enzymes are included: phenylalanine ammonium lyase (PAL), 4-coumarate-CoA ligase (4CL), cinnamoyl-CoA reductase, cinnaryl alcohol dehydrogenase, hydroxycinnamoyl-CoA quinate shikimate p-hydroxycinnamoyltransferase (HCT), chalcone synthase (CHS), chalcone isomerase, flavonone 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 (SNP948) 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 SNP948 within its potential target Ptr-4CL had distinct epistatic effects on HoC and LiC in different genotypic combinations.
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 \( \text{Ptr-4CL} \) has an epistatic effect with SNP1096 from its upstream gene \( \text{Ptr-PAL} \) on V (Fig. 4C). For the downstream genes \( \text{Ptr-HCT} \) and \( \text{Ptr-CHS} \) which were assigned to two different branches of the monolignol biosynthesis pathway, we also detected an epistatic interaction with \( \text{Ptr-4CL} \) 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 \( \text{Ptr-4CL} \) 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 \( \text{Populus} \)
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...
networks, and miRNA regulation. 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. In addition, *ptr-miR397*, which targets of miR168, which could be involved in auxin signaling and secondary cell wall biosynthesis, exhibits near-perfect complementarity to its target. 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, exhibits near-perfect complementarity to its target *TCONS_000108161*. 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. In addition, *ptr-miR397*, which could be a negative regulator of laccase genes affecting LiC, was predicted to target three lncRNAs (*TCONS_00013182*, *TCONS_00108161*, and *TCONS_00078539*).

### 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. Moreover, interesting cross-regulation between miRNAs and lncRNAs has recently become apparent, 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.
Target mimicry is a recently identified mechanism of miRNA regulation initially studied in Arabidopsis, and subsequently in mammals. lncRNAs can act as miRNA decoys or sponges to sequester miRNAs, thus favouring the expression of repressed target miRNAs. Indeed, many eTMs have been functionally confirmed in Arabidopsis and rice. In our study, 28 lncRNAs were predicted to be eTMs of 14 miRNAs (Supplementary Material S10). Interestingly, consistent with the previous studies, 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, was predicted to be absorbed by five eTMs. Additionally, in some cases, like miR408, which is involved in lignin deposition and polymerization, 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. 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 can be pervasive in plants. For example, miR164, which may affect lignin metabolism by regulating transcription factors that control secondary growth and wood composition, was predicted to be absorbed by five eTMs. Additionally, in some cases, like miR408, which is involved in lignin deposition and polymerization, 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.

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

1. Ma, L., Bajic, V. B. and Zhang, Z. 2013, On the classification of long non-coding RNAs, RNA Bio., 10, 924–33.
2. Necsulea, A., Soumillon, M., Warnéfors, M., et al. 2014, The evolution of IncRNA repertoires and expression patterns in tetrapods. Nature, 305, 635–40.
3. Cabili, M.N., Trapnell, C., Goff, L., et al. 2012, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev., 25, 1915–27.
4. Ponting, C.P., Oliver, P.L. and Reik, W. 2009, Evolution and functions of long noncoding RNAs. Cell, 136, 629–41.
5. Świeżewski, S., Liu, F., Magusin, A. and Dean, C. 2009, Cold-induced silencing by long antisense transcripts of an Arabidopsis polycistron target. Nature, 462, 799–802.
6. Heo, J.B. and Sung, S. 2011, Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science, 331, 76–9.
7. Franco-Zorrilla, J.M., Valli, A., Todesco, M., et al. 2007, Target mimicry by a long intronic noncoding RNA. Plant Cell Physiol., 51, 144–63.
8. Shastry, B.S. 2009, SNPs: impact on gene function and phenotype. In: Single Nucleotide Polymorphisms, Berlin: Springer, pp. 3–22.
9. 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.
10. 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.
11. Necsulea, A., Soumillon, M., Warnefors, M., et al. 2014, The evolution of long non-coding RNAs, RNA Bio., 16, 1479–87.
12. Li, L., Eichten, S.R., Shimizu, R., et al. 2014, Genome-wide discovery and characterization of maize long non-coding RNAs, RNA Bio., 16, 1479–87.
13. Trapnell, C. 2012, DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics, 27, 1635–9.
14. Buske, F.A., Bodin, M., Bauer, D.C. and Bailey, T.L. 2010, Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics, 26, 860–6.
15. 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.
16. 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, 1596–1604.
17. Sun, L., Luo, H., Bu, D., et al. 2013, Utilizing sequence intrinsic composition to classify non-coding and long non-coding transcripts. Nucleic Acids Res., 41, e166.
18. Roberts, A., Pimentel, H., Trapnell, C. and Pachter, L. 2011, Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics, 27, 2125–9.
19. 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.

20. Uyang, S., Zhu, W., Hamilton, J., et al. 2007, The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res., 35, D883–7.
21. Bailey, T.L. 2011, DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics, 27, 1635–9.
22. Buske, F.A., Bodin, M., Bauer, D.C. and Bailey, T.L. 2010, Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics, 26, 860–6.
23. Yañez, I., Benjamín, H., Shmoish, M., et al. 2005, Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics, 21, 650–9.
24. 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, 1479–87.
25. Han, L., Zhang, K., Shi, Z., et al. 2012, IncRNA profile of glioblastoma reveals the potential role of IncRNAs in contributing to glioblastoma pathogenesis. Int. J. Oncol., 40, 2004–12.
26. Tafer, H. and Hofacker, I.L. 2008, RNAplex: a fast tool for RNA-RNA interaction search. Bioinformatics, 24, 2657–63.
27. 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.
28. Saito, R., Smoot, M.E., Ono, K., et al. 2012, A travel guide to Cytoscape plugins. Nat. Methods, 9, 1069–76.
29. 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.
30. Du, Q., Xu, B., Pan, W., et al. 2013, Allelic variation in a cellulose synthase gene (ProcXcA4) associated with growth and wood properties in Populus tomentosum. G3, 3, 2069–84.
31. 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.
32. 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.
33. Li, R., Li, Y., Kristiansen, K. and Wang, J. 2008, SOAP: short oligonucleotide alignment program. Bioinformatics, 24, 713–4.
34. Danecck, P., Auton, A., Abecasis, G., et al. 2011, The variant call format and VCFTools. Bioinformatics, 27, 2156–8.
35. Nei, M. 1987, Molecular Evolutionary Genetics. Columbia University Press.
36. Watterson, G. 1975, On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol., 7, 256–276.
37. Tajima, F. 1989, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585–95.
38. Yu, J., Pressor, G., Briggs, W.H., et al. 2004, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet., 38, 203–8.
39. 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.
40. 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.
41. Hubisz, M.J., Falush, D., Stephens, M. and Pritchard, J.K. 2009, Inference of population structure with the assistance of sample group information. Mol. Ecol. Resour., 9, 1322–32.
42. Storley, J.D. and Tibshirani, R. 2003, Statistical significance for genome-wide studies. Proc. Natl Acad. Sci. USA, 100, 9440–5.
43. Mao, Y., London, N.R., Ma, L., Dvorkin, D. and Da, Y. 2007, Detection of SNP epistasis effects of quantitative traits using an extended Kemphorne model. Physiol. Genomics, 28, 46–52.
