RESEARCH PAPER

Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa

Peng Shuai*, Dan Liang*, Sha Tang, Zhoujia Zhang, Chu-Yu Ye, Yanyan Su, Xinli Xia† and Weilun Yin†

College of Biological Sciences and Technology, National Engineering Laboratory of Tree Breeding, Beijing Forestry University, mailbox 69, No. 35 Qinghua East Road, Haidian District, Beijing 100083, P.R. China

* These authors contributed equally to this work.
† To whom correspondence should be addressed. E-mail: xiaxl@bjfu.edu.cn; yinwl@bjfu.edu.cn

Received 3 March 2014; Revised 9 May 2014; Accepted 16 May 2014

Abstract

Protein-coding genes are considered to be a dominant component of the eukaryotic transcriptome; however, many studies have shown that intergenic, non-coding transcripts also play an important role. Long intergenic non-coding RNAs (lincRNAs) were found to play a vital role in human and Arabidopsis. However, lincRNAs and their regulatory roles remain poorly characterized in woody plants, especially Populus trichocarpa (P. trichocarpa). A large set of Populus RNA-Seq data were examined with high sequencing depth under control and drought conditions and a total of 2542 lincRNA candidates were identified. In total, 51 lincRNAs and 20 lincRNAs were identified as putative targets and target mimics of known Populus miRNAs, respectively. A total of 504 lincRNAs were found to be drought responsive, eight of which were confirmed by RT-qPCR. These findings provide a comprehensive view of Populus lincRNAs, which will enable in-depth functional analysis.

Key words: Drought, lincRNA, miRNA, non-coding RNA, Populus trichocarpa, target mimicry.

Introduction

According to recent studies, more than 90% of eukaryotic genomes are transcribed, yet only 1–2% has protein-coding capacity (Kim and Sung, 2012; Hangauer et al., 2013). To characterize the complete Populus genome, all transcripts must be examined. While much is known about Populus protein-coding genes, recent studies have suggested that eukaryotic genomes also encode a large number of functional transcripts of non-coding RNAs (ncRNAs), including housekeeping and regulatory RNAs (Chen and Carmichael, 2010; Shuai et al., 2013; Tang et al., 2013). One regulatory ncRNA, long ncRNA (lincRNA), has been reported to be a vital component of eukaryotic gene regulation (Ng and Ng, 2010; Guttmann et al., 2011; Nagano and Fraser, 2011; Kim and Sung, 2012; Kornienko et al., 2013).

According to the general location in the genome, there are two types of long ncRNAs; namely, long intron ncRNAs and long intergenic ncRNAs (lincRNAs) (Liu et al., 2012). Bumgarner et al. (2009) identified two lincRNAs in yeast that contribute to the control of variegated gene expression. A growing number of lincRNAs are known to be key regulators in higher eukaryotic organisms based on large-scale sequencing. Ulitsky et al. used chromatin marks, poly(A)-site mapping and RNA-Seq data to identify more than 550 distinct lincRNAs in zebrafish (Ulitsky et al., 2011). Large-scale identification of human lincRNAs has been accomplished by analysing various cell types (Khalil et al., 2009; Jia et al., 2010; Cabili et al., 2011; Hangauer et al., 2013; Kumar et al., 2013). A total of 1119 candidate lincRNAs loci have been identified in the fruit fly, some of which may be important in the nervous system (Young et al., 2012). Li et al. (2012) identified 281 novel lincRNAs in chicken skeletal muscle using next-generation sequencing.
In plants, genome-wide identification of lincRNAs has only been conducted in maize and Arabidopsis (Boerner and McGinnis, 2012; Liu et al., 2012). In maize, a computational pipeline using the programming language Python was developed and applied to full-length cDNA sequences to identify, classify, and localize potential lincRNAs. A total of 439 maize lincRNAs were identified (Boerner and McGinnis, 2012). Using a reproducibility-based bioinformatics strategy to analyse 200 Arabidopsis thaliana transcriptome data sets, Liu et al. (2012) identified 13 230 intergenic transcripts, of which 6480 could be classified as lincRNAs. In Populus, studies of regulatory RNAs have focused on miRNAs; however, none have addressed lincRNAs (Li et al., 2011, 2013; Shuai et al., 2013).

In this study, the high-throughput RNA-Seq method, which allows for the sensitive detection of transcripts with low expression and does not depend on current gene annotations, was applied. It is therefore ideal for detecting novel transcripts, especially lincRNAs. Recent studies have shown that lincRNAs are developmentally regulated and responsive to external stimuli (Ben Amor et al., 2009; Xin et al., 2011; Wu et al., 2012). Environmental stressors due to climate change, especially drought stress, could make forests increasingly vulnerable to disease and die-offs (Allen et al., 2010). Drought is known to be an important abiotic stress (Hamanishi and Campbell, 2011). In this study, transcriptome sequencing was conducted in a control library (CL) and drought library (DL) and 2542 lincRNAs were identified. To investigate these novel lincRNAs further, differential expression of lincRNAs between the two treatments was analysed. A total of 504 drought-responsive lincRNAs were identified, some of which were confirmed by RT-qPCR.

In addition, the relationship between lincRNAs and miRNAs was investigated. Some lincRNAs were identified as putative targets and target mimics of miRNAs. Furthermore, the lincRNAs identified here were compared with those identified by previous studies. Very few Populus lincRNAs have been identified. Overall, our findings revealed that lincRNAs play key regulatory roles in Populus. In this study, a basic annotation set of 2542 lincRNAs is provided, which will increase our understanding of the intergenic functional non-coding genes in plants.

Materials and methods

Plant materials and RNA-Seq

Seedlings (female P. trichocarpa ‘Nisqually 1’) from tissue culture (5 cm tall) were planted in individual pots (15 l) containing loam soil and placed in a greenhouse. After 3 months of growth, they were ~45 cm tall. For the drought-stress treatment, different relative soil moisture contents (RSMCs) were used for control and drought conditions. Seedlings from soil with good irrigation (RSMC 70–75%) were used as a control and a low soil water-content level (RSMC 15–20%) was chosen for the drought treatment (Li et al., 2011; Shuai et al., 2013). Leaf water potential (WP) was measured using a PsyPro WP data logger (Wescor). Photosynthetic rate, stomatal conductance, and transpiration rate were measured using the Li-6400 Photosynthesis System (Li-Cor). For another three abiotic stresses, the seedlings grown in vitro were untreated or treated with cold (4 °C for 24 h), heat (37 °C for 24 h), or water (cover the plants with water for 14 h). For material harvest, mature leaves from the same position of three individual plants were collected and frozen immediately in liquid nitrogen. For RNA extraction, total RNA was extracted using the standard CTAB method for plants (Chang et al., 1993). Beads with oligo(dT) were used to isolate poly(A) mRNA after total RNA was collected. Fragmentation buffer was added to interrupt mRNA into short fragments. First-strand and second-strand cDNA was synthesized. After that, the short fragments were connected with sequencing adaptors. For PCR amplification, 200–700 bp fragments were selected as templates, with respect to the results of agarose gel electrophoresis. Two libraries of CL and DL were submitted to the Beijing Genomics Institute (BGI) for Illumina sequencing (HiSeq™ 2000).

Assembling RNA transcripts and identifying novel transcriptional units

After filtering out low-quality reads and trimming the adaptor sequences, a total of ~260 million pair-end clean reads were obtained. Clean reads were mapped to the P. trichocarpa (version 2.2) genome and gene sequences, respectively, using SOAP2 (Li et al., 2009). Mismatches of no more than five bases were allowed in the alignment. The lincRNA gene structure was optimized according to the read distribution, information of paired-ends, and the genome annotation. The distribution of reads in the genome was obtained by aligning the continuous and overlapping reads to form a Transcription Active Region (TAR). According to paired-end data, the different TARs were connected to construct a potential gene model. The disjunctive parts in the lincRNA gene model were considered lincRNA exons. The other components of the model, which were not detected in sequencing reads, were considered lincRNA introns. Gene models that did not map to the known transcript and were located in the intergenic region (200 bp from upstream or downstream genes) were selected for further analysis. To distinguish transcriptional cDNA from genomic DNA contamination, the relatively high expression-level transcripts were subjected to further analysis. Read numbers per base pair of less than 2 were not considered.

After novel intergenic transcriptional units were obtained, they were adapted to four filter processes to identify lincRNA candidates. First, the length of TUs had to be longer than 200 bp to exclude small intergenic transcripts. Second, the longest ORF of the TU had to be smaller than 100 AA (the
longest ORF predicted by OrfPredictor (http://proteomics.ysu.edu/tools/OrfPredictor.html) (Min et al., 2005). Both strands of the TUs were used for prediction. Third, to ensure that our results were not influenced by genomic DNA contamination of the cDNA library, the lincRNA candidates must have appeared in both CL and DL. Fourth, these sequences were analysed using BLAST against the miRBase and RepPop to remove miRNA precursors and repetitive elements (Zhou and Xu, 2009; Kozomara and Griffiths-Jones, 2011).

Validating the non-coding capacity of lincRNAs using CPC and codon usage

The coding potential of all putative lincRNAs were submitted to the CPC (http://cpc.cbi.pku.edu.cn/) (Kong et al., 2007). The exonic bases for each transcript in a model were analysed separately and in both orientations (forward and reverse strand). A similar analysis was also performed for 45 033 coding transcripts. A total of 2.3% of genes annotated as protein coding by Phytozome were designated as non-coding by CPC (score less than 0.5) (see Supplementary Table S4 at JXB online). If all transcripts within an intergenic model were considered non-coding, it was defined as a lincRNA locus (Young et al., 2012).

Calculation of codon usage was conducted using the longest ORF from OrfPredictor by the online tool (http://www.biinformatics.org/sms2/codon_usage.html). The Codon Usage database (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=3694) for the P. trichocarpa CDS was used as a reference (Nakamura et al., 2000).

miRNA target and target mimicry prediction

The target was predicted by submitting P. trichocarpa miRNAs and the lincRNAs to the psRNATarget (http://plantgrn.noble.org/psRNATarget/) (Dai and Zhao, 2011), with a total of no more than three mismatches and G/U pairs within the lincRNA and miRNA pairing regions. The target mimics were predicted using psRNATarget combined with local scripts and the rules established by Wu et al. (2013).

Differential expression analysis of lincRNAs between the two treatments

The lincRNA sequence reads of the two libraries were normalized to FPKM (fragments per kilobase of transcript per million mapped reads) values in each sample (Trapnell et al., 2010). Calculation of the P-value for comparison of lincRNA expression between the two libraries was based on an established method (Audic and Claverie, 1997; Man et al., 2000). The following P-value formula was used:

\[
p(x|y) = \left(\frac{N_2}{N_1}\right)^y \frac{(x+y)!}{x!y!} \frac{1+\sum_{k=0}^{y} \binom{x+y}{k} \left(\frac{N_2}{N_1}\right)^k}{\sum_{k=0}^{\infty} \binom{x+y}{k} \left(\frac{N_2}{N_1}\right)^k}
\]

where \(N_1\) is the total number of reads in the sequencing library of the control, \(N_2\) is the total number of reads in the sequencing library of the drought treatment, \(x\) is the number of reads for an lincRNA in the control library, and \(y\) is the number of reads for a lincRNA in the drought treatment library. Specifically, the log2 ratio formula was: \(\log_2 \text{ratio} = \log_2 \left(\frac{\text{FPKM in DL}}{\text{FPKM in CL}}\right)\). According to these calculations, lincRNAs with a log2 ratio larger than 1 and \(P\)-value less than 0.001 were considered to be drought responsive.

Quantitative real-time polymerase chain reaction (RT-qPCR) analysis

In this study, there were 504 drought-responsive lincRNAs. Among them, 60 changed by 4-fold between the control and drought treatments. The 60 lincRNAs were considered the most up- and down-regulated candidates. To validate the lincRNA high-throughput sequencing results, RT-qPCR was performed for eight randomly selected drought-responsive lincRNAs from these 60 lincRNAs. RNA was extracted from leaves using the CTAB method (Chang et al., 1993). All primers used in this study are listed in Supplementary Table S10 at JXB online. RT-qPCR was performed using an ABI StepOnePlus instrument. RT-qPCR results were subjected to the following calculations: Sample cycle threshold (Ct) values were determined and standardized relative to the three endogenous control genes (ACTIN, 18S, and HIS), and the \(2^{-\Delta\Delta C_T}\) method was used to calculate the relative changes in gene expression based on the RT-qPCR data (Livak and Schmittgen, 2001).

Results

Physiological characterization of P. trichocarpa in response to drought stress

P. trichocarpa plants were exposed to soil water deficiency at two relative soil moisture contents (RSMC) levels. These levels were set at 70–75% as control and 15–20% for the drought-stress group. The leaf water potential (WP) was detected as a measure of the water status in plants and of the ability of plants to absorb water. A significant decrease (\(P < 0.001\)) in leaf WP from \(-1.40\) MPa in controls to \(-3.23\) MPa under drought conditions was observed (Fig. 1). At the same time, it was found that the net photosynthetic rates, transpiration rate, and stomatal conductance of leaves under drought conditions significantly decreased compared with the controls (Fig. 1). These physiological differences indicate that plants under drought conditions may show significant changes in gene expression (including lincRNAs). These changes should be identified using RNA-Seq and compared between control and drought conditions.

Deep sequencing and prediction of novel intergenic transcripts

To identify and investigate the lincRNAs in Populus during drought stress, Illumina sequencing of transcripts from CL and
DL was performed. A total of ~269 million clean reads (90 bp) was obtained (see Supplementary Table S1 at *JXB* online). The average read depth of this sequencing was 57.3-fold that of the whole *Populus* genome. This large data set allowed for the detection of both rare and tissue-specific transcription. Reads of known protein-coding genes accounted for 85% of the total reads. However, the other 15% of reads could not be mapped to known genes and contained regions of high expression within intergenic regions. Transcripts from intergenic regions were then identified. TUs (novel intergenic transcript units) that were 200 bp from upstream or downstream genes were explored further. Totals of 11 292 and 11 275 TUs were obtained by high-throughput sequencing from CL and DL, respectively (see Supplementary Table S2 at *JXB* online). These TUs were unknown intergenic transcripts that did not match any known protein-coding transcripts of *P. trichocarpa*.

**Identification of lincRNA candidates**

To identify novel and drought-responsive lincRNAs, these TUs were analysed further using computational and experimental methods. A pipeline for the characterization of *Populus* lincRNAs was constructed (Fig. 2). Totals of 11 292 and 11 275 novel intergenic transcriptional units were analysed in this pipeline. Four filter processes were applied to distinguish lincRNAs from transcript units. To characterize long ncRNAs, the minimum transcript length was 200 bp. A total of 220 (CL) and 6 532 (DL) TUs were found to be longer than 200 bp in the two libraries. OrfPredictor was used to identify protein-coding regions in TUs (Min et al., 2005), and to calculate the longest possible ORF of each strand. The putative protein-coding RNAs were then filtered using a maximum possible ORF length of 100 amino acids (AA). After these two steps, TUs found in both libraries were selected as putative lincRNAs. After these two steps, 3372 putative lincRNA loci found in both libraries were chosen for the identification and exclusion of repetitive elements and microRNA precursors (see Supplementary Table S3 at *JXB* online). To exclude the repetitive elements, 3372 putative lincRNAs were compared using the RepPop database (Zhou and Xu, 2009). A total of 583 putative lincRNAs showed high similarity with repetitive elements. To identify miRNA precursors from putative lincRNAs, all 352 *Populus* miRNA precursors from miRBase (version 20) were compared with the 3372 putative lincRNAs (Kozomara and Griffiths-Jones, 2011). A total of 28 lincRNAs were identified as miRNA precursors (see Supplementary Table S3 at *JXB* online), which corresponded to ~1% of the total lincRNAs. These selection processes identified 2761 putative lincRNA loci for further investigation (see Supplementary Table S3 at *JXB* online).

**Independent non-coding evidence of the most predicted lincRNAs**

As stated above, thousands of putative lincRNAs were distinguished based on a cut-off transcript ORF length. However, it is important to validate the coding status of these lincRNAs loci using an independent method. Therefore, the putative lincRNAs were examined using the Coding Potential Calculator (CPC) filter (Kong et al., 2007). CPC facilitates assessment of the protein-coding potential of large-scale transcripts using sequence features and support vector machines. The protein-coding capacities of TUs on both strands were assessed by CPC. The 45 033 protein-coding transcripts of *P. trichocarpa* (version 2.2) were used as a reference, of which only 2.3% were characterized as non-coding (see Supplementary Table S4 at *JXB* online; Table 1) (Tuskan et al., 2006). The quality of the lincRNA discovery pipeline was demonstrated based on the non-coding potential of 45 033 transcripts.
Approximately 92.1% of the 2761 putative lincRNAs were non-coding (CPC score <0.5) (see Supplementary Table S3 at JXB online; Table 1). The non-coding rate was higher than those of 6220 and 6532 raw TUs, whose non-coding rates were 90.2% and 90.9%, respectively. A total of 219 putative lincRNA loci classified as coding were discarded. It was also found that the average non-coding potential of lincRNA with more exons was higher than those with fewer exons (Table 1).

Based on this non-coding evidence, the codon usage and GC content of the predicted ORFs in lincRNA candidates was analysed further. This information on coding sequences of *Populus trichocarpa* is available in the Codon Usage Database. It was found that the codon usage in the longest predicted ORF in lincRNAs was not correlated with that in the database (see Supplementary Table S5 at JXB online). Glutamic acid and aspartic acid content in the CDS was significantly higher than that in lincRNAs. The coding sequences of *Populus* contained 42.6% GC nucleotides, while the mean GC content of predicted ORFs in lincRNAs was 38.8%. The disparity between the protein-coding RNAs and lincRNAs is indicative of different evolutionary pressures in ORFs. These results further support the non-coding potential of these lincRNAs.

**Length and scaffold distribution of lincRNAs**

Based on the above results, 2542 putative lincRNAs were selected for further analysis (see Supplementary Table S3 at JXB online). The length distribution of these lincRNA loci ranged from 200 bp to 4241 bp, yet more than 80% ranged from 200 bp to 600 bp (Fig. 3A). The most abundant length was 300–400 bp. Furthermore, the distribution of lincRNAs in *Populus* scaffolds was examined (as the 19 chromosomes of version 2.2 were not well assembled). In the eighth scaffold, there were 8.3 lincRNAs within every 1 Mbp, which was the highest lincRNA packing density. The seventh scaffold (containing 4.3 lincRNAs per 1 Mbp of nucleotides) had the lowest packing density (Fig. 3B).

**LincRNAs as putative targets of miRNAs**

The relationship between miRNA and lincRNA is an important issue. Comprehensive miRNA regulation patterns of 90 lincRNAs, some of which are important in breast cancer, were examined in humans (Juan et al., 2013). These miRNAs may play roles in promoting the degeneration of lincRNAs. A total of 30 miRNAs were predicted to target to the sense strand of lincRNAs, while 21 were also found to target the antisense strand (see Supplementary Table S6 at JXB online; Fig. 4A, B).

**Table 1. Independent assessment of the non-coding capacity of our lincRNA candidate**

| Classifications                                      | Number | Non-coding | %    |
|------------------------------------------------------|--------|------------|------|
| Protein-coding genes in *P. trichocarpa*             | 45 033 | 1059       | 2.3  |
| TU longer than 200bp in CL                           | 6220   | 5611       | 90.2 |
| TU longer than 200bp in DL                           | 6532   | 5988       | 90.9 |
| lincRNA candidates from pipeline                     | Total  | 2761       | 2542 | 92.1 |
|                                                      | 1 Exon | 1500       | 1363 | 90.9 |
|                                                      | 2 Exons | 765       | 714  | 93.3 |
|                                                      | >2 Exons | 496       | 465  | 93.8 |
Functional prediction of lincRNAs as putative target mimics of miRNAs

A mature miRNA leads to the degradation of a specific target; however, this can be inhibited through target mimicry (Hansen et al., 2013). Target mimicry is an important function of lincRNAs in plants. This miRNA–lincRNA relationship has recently been observed in plants (Franco-Zorrilla et al., 2007; Todesco et al., 2010; Meng et al., 2012). Arabidopsis IPS1 is the first target mimic identified that is complementary to, and regulates the expression of, miR-399. In this study, the target mimics of these lincRNAs were predicted according to the rules of Wu et al. (2013). In total, eight and 12 target mimics were identified on the sense and antisense strand, respectively (see Supplementary Table S7 at JXB online; Fig. 4C, D).

Three miRNAs (ptc-miR482a.1, ptc-miR476a, and ptc-miR156c) with different functions (target and target mimicry, respectively) were identified. LincRNA262 and lincRNA2623 is the target and target mimic of ptc-miR156c, respectively. LincRNA1851 has two target mimics (lincRNA20 and lincRNA1795). LincRNA1310 is a target of ptc-miR476a. For ptc-miR482a.1, the relationship was more complex than the other two, and a possible regulatory mechanism is shown in Fig. 5. Ptc-miR482a.1 regulates four lincRNAs (lincRNA1078, lincRNA1203, lincRNA2213, and lincRNA2252) and 27 disease resistance transcripts through degradation. However, three lincRNAs (lincRNA1128, lincRNA1828, and lincRNA2623) regulate ptc-miR482a.1 as target mimics and inhibit its function.

Identification of drought responsive lincRNAs

To identify drought-responsive lincRNAs from P. trichocarpa, the number of normalized lincRNA reads of CL and DL were calculated using FPKM (fragments per kilobase of transcript per million mapped reads) and compared. Based on the sequencing results, lincRNAs with a differential expression of greater than 2-fold and *P*-values less than 0.001 were considered differentially expressed. A total of 504 lincRNAs were identified (see Supplementary Table S9 at JXB online), and eight were subjected to experimental validation by quantitative real-time polymerase chain reaction (RT-qPCR). As shown in Fig. 6, the expression patterns indicated by the sequencing and RT-qPCR results of drought-responsive lincRNAs were consistent, although the relative expression levels of all lincRNAs by RNA-Seq were greater than those by RT-qPCR. Therefore, six lincRNAs were identified as up-regulated after drought treatment, while two lincRNAs were down-regulated.

Discussion

High-throughput RNA-Seq has been used to identify both protein-coding and non-protein-coding transcripts, regardless of their expression levels. The data provide a comprehensive view of the transcriptome of P. trichocarpa, which can be used to identify functional elements and understand the molecular mechanisms underlying drought resistance.
of whether they are known or unknown (Hangauer et al., 2013). In this study, 2542 lincRNA loci were identified by analysing 269 million sequences. The number of lincRNAs identified by sequencing depends largely on the sequencing depth and species. In the initial screen of more than 20 000 TUs, lincRNAs were selected based on strict criteria. Although this may have excluded many lincRNAs, these 2542 lincRNA loci constitute a reliable set of *Populus* lincRNAs. This is the first study of lincRNA in model woody plants, and will provide a basic overview of lincRNAs in *Populus*.

Due to the inherent limitations of using 90-bp pair-end reads in transcriptome sequencing, the complete structure of a transcript was difficult to obtain. In addition, because our RNA-Seq data lack strand information, it was not possible to determine from which strand the lincRNAs were produced. These limitations are ubiquitous in recent lincRNAs studies using the RNA-Seq method (Li et al., 2012; Kumar et al., 2013). Thus, determination of the full structure and strand information requires improved sequencing technology. Although there are limitations to using lincRNAs, the data can also be valuable. Therefore, lincRNAs from either strand were analysed and their differential expression was evaluated using a fragment of the transcript.

The sequences of 6480 lincRNAs in *Arabidopsis* and 1119 lincRNAs in the fruit fly were compared with 2542 lincRNAs from other species (Liu et al., 2012; Young et al., 2012). Only six lincRNAs identified in this study had homologues in *Arabidopsis*, yet there were no significant matches with the fruit fly (see Supplementary Table S8 at *JXB* online). These results suggest that the majority of lincRNAs identified from our study were not conserved with currently known lincRNAs among plant and animal species. This was also reported in previous studies of other species such as wheat, mouse, and fruit fly (Guttman et al., 2010; Xin et al., 2011; Young et al., 2012).

The FANTOM consortium uses a cut-off of 300 nt (100 AA) to identify putative protein-coding mRNAs (Okazaki et al., 2002). In this study, this ORF threshold was also used to distinguish protein-coding from non-protein-coding transcripts. The two important criteria (ORF and length thresholds) of the pipeline were compared with the number of lincRNAs identified in recent studies (Table 2) (Boerner and McGinnis, 2012; Li et al., 2012; Liu et al., 2012; Young et al., 2012; Hangauer et al., 2013). Four out of six previous studies used an ORF threshold of 100 AA to identify lincRNAs. In chicken, the aggressive criteria of our pipeline resulted in the identification of only a few lincRNAs (Table 2). Based on these results, the criteria (100 AA and 200 nt) used in this work were appropriate. The CPC filter applied here was also used to assess lincRNAs in previous studies (Jia et al., 2010; Boerner and McGinnis, 2012; Young et al., 2012). Based on the low non-coding potential of protein-coding transcripts using CPC in humans and in our study, CPC is a reliable evaluation tool (Jia et al., 2010).

As predicted by Juan et al. (2013), lincRNAs are similar in nature to mRNA in that miRNAs can bind lincRNAs and trigger degradation. Previously, lincRNAs were identified as targets of miRNAs in human studies. The repression of lincRNAs could be a novel component of miRNA regulation. In this study, the miRNA–lincRNA interaction was also observed.

Target mimicry is a newly identified miRNA regulation mechanism first studied in *Arabidopsis*. Franco-Zorrilla et al. (2007) reported that over-expression of the non-coding gene IPS1 inhibited miR399 and increased expression of the target of miR399. Computational methods have been applied to identify target mimics. However, previous studies focused mainly on target mimics derived from annotated genes; however, no target mimics have yet been identified in *Populus*. In this study, predictions were performed using 2542 lincRNAs and potential target mimics for about 20 miRNAs were identified. Two target mimics (lincRNA432 and lincRNA1174) for

### Table 2. Comparison of ORF and length thresholds

| Species    | ORF threshold | Length threshold | lincRNAs found | Reference       |
|------------|---------------|------------------|----------------|----------------|
| Maize      | 120 AA        | 200 nt           | 439            | Axtell (2013)   |
| Arabidopsis| 100 AA        | 200 nt           | 6480           | Liu et al. (2012) |
| Populus    | 100 AA        | 200 nt           | 2542           | –              |
| Fruit fly  | 100 AA        | 200 nt           | 1119           | Young et al. (2012) |
| Human      | 100 AA        | 200 nt           | 58 465         | Hangauer et al. (2013) |
| Chicken    | 60 AA         | 300 nt           | 281            | Li et al. (2012) |

![Fig. 6. Differential expression analysis of eight lincRNAs under drought stress. RT-qPCR was performed for eight randomly selected drought-responsive lincRNAs from the 60 most up- and down-regulated candidate lincRNAs. Three internal controls (ACTIN, 18S, and HIS) were used for data normalization. The normalized lincRNA levels of the control were set arbitrarily to 1.](Image 319x319)
miR160 were also identified in *Populus*, which were conserved with the 13 endogenous target mimics identified by Wu et al. (2013) in *Arabidopsis* and rice. Ptc-miR482a.1 has been investigated previously and regulates specific disease-resistance proteins in *P. trichocarpa*. This miRNA is known to be responsive to four abiotic stresses (cold, heat, salt, and dehydration) in *Populus* (Lu et al., 2008). Our data suggest involvement of ptc-miR482a.1 in the regulatory network in combination with seven lincRNAs. Three lincRNAs could potentially regulate ptc-miR482a.1 through target mimicry. These lincRNAs may be associated with abiotic stress tolerance in *Populus*.

LincRNAs are known to respond to biotic and abiotic stresses in plants, such as *Arabidopsis* and wheat (Xin et al., 2011; Liu et al., 2012). Drought-responsive lincRNAs identified in this study were selected using aggressive criteria and were confirmed experimentally. However, some of these drought-responsive lincRNAs were also responsive to other abiotic stresses (water, and cold stresses) (Fig. 7). Drought-induced lincRNA2962 and lincRNA1039 are also down-regulated and up-regulated by cold stress, respectively. LincRNA3241 is down-regulated by water and cold stress.

Interestingly two drought-responsive lincRNAs (lincRNA20 and lincRNA2752) that were target mimics of miRNAs, were identified. Drought-responsive lincRNA20 adsorbed ptc-miR476, which is a specific family in *Populus* according to miRBase (Kozomara and Griffiths-Jones, 2011). However, lincRNA20 is also specific to *Populus*, and may thus represent a *Populus*-specific regulatory mechanism. In addition, drought-responsive lincRNA2752 is a target mimic of ptc-miR169, and could reduce the expression of ptc-miR169. MiR169 is known to regulate the *NF-YA* transcription factor in plants, which is important in drought stress regulation (Ni et al., 2013). This network may be involved in the lincRNA2752-regulation of drought tolerance through miR169 and *NF-YA*. However, the specific regulatory mechanism requires further investigation, and knock out and over-expression of the lincRNA genes in *P. trichocarpa* should be performed to increase our understanding of the regulatory mechanisms.

**Supplementary data**

Supplementary data can be found at JXB online.

Supplementary Table S1. Summary of *P. trichocarpa* RNA-Seq data.

Supplementary Table S2. 11 292 TUs in the control library and 11 275 TUs in the drought library.

Supplementary Table S3. The 3 372 lincRNA candidates from the pipeline.

Supplementary Table S4. CPC scores of 45 033 *P. trichocarpa* transcripts.

Supplementary Table S5. Comparison of the codon frequency of CDS and lincRNAs in *P. trichocarpa*.

Supplementary Table S6. *Populus* microRNAs target lincRNAs (sense strand and antisense strand).

Supplementary Table S7. *Populus* lincRNAs as target mimics to microRNAs (sense strand and antisense strand).

Supplementary Table S8. Similarities of lincRNAs in *Populus* and *Arabidopsis*.

Supplementary Table S9. Differential expression of lincRNAs.

Supplementary Table S10. Primers designed for RT-qPCR.

**Acknowledgements**

We thank Ms Tingyan Zhong for her helpful suggestions. This research was supported by the National Natural Science Foundation of China (31270656), The Program for Changjiang Scholars and Innovative Research Teams in University (IRT13047), the Joint Programs of Scientific Research and Graduate Training from BMEC (Stress Resistance Mechanism of Poplar), and the ‘111’ Project of Beijing Forestry University (B13007).

**Conflict of interest**

The authors have no conflict of interest to declare.

**References**

Allen CD, Macalady AK, Chenchouni H, et al. 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259, 660–684.

Audic S, Claverie JM. 1997. The significance of digital gene expression profiles. *Genome Research* 7, 986–995.

Axtell MJ. 2013. Classification and comparison of small RNAs from plants. *Annual Review of Plant Biology* 64, 137–159.

Ben Amor B, Wirth S, Merchan F. et al. 2009. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Research* 19, 57–69.

Boerner S, McGinnis KM. 2012. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS ONE* 7, e43047.

Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, Fink GR. 2009. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences, USA* 106, 18321–18326.

Cabiil MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic non-coding RNAs. *Genome Research* 21, 1275–1284.
noncoding RNAs reveals global properties and specific subclasses. Genes and Development 25, 1915–1927.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter 11, 113–116.

Chen L-L, Carmichael GG. 2010. Decoding the function of long non-coding RNAs. Current Opinion in Cell Biology 22, 357–364.

Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Research 39, W155–W159.

Franco-Zorrilla JM, Valli A, Todesco M. et al. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics 39, 1033–1037.

Guttman M, Donaghey J, Carey BW. et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300.

Guttman M, Garber M, Levin JZ. et al. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28, 503–510.

Hamanishi ET, Campbell MM. 2011. Genome-wide responses to drought in forest trees. Forestry 84, 273–283.

Hangauer MJ, Vaughn IW, McManus MT. 2013. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genetics 9, e1003569.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. Nature 495, 384–388.

Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L. 2010. Genomewide computational identification and manual annotation of human long non-coding RNA genes. RNA 16, 1478–1487.

Juan L, Wang G, Radovich M, Schneider BP, Clare SE, Wang Y, Liu Y. 2013. Potential roles of microRNAs in regulating long intergenic noncoding RNAs. BMC Medical Genomics 6, S7.

Khali AM, Guttman M, Huarte M. et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences, USA 106, 11167–11172.

Kim ED, Sung S. 2012. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends in Plant Science 17, 16–21.

Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research 35, W345–W349.

Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. BMC Biology 11, 59.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39, D152–D157.

Kumar V, Westra HJ, Karjalainen J. et al. 2013. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genetics 9, e1003201.

Li B, Duan H, Li J, Deng XW, Yin W, Xia X. 2013. Global identification of miRNAs and targets in Populus euphratica under salt stress. Plant Molecular Biology 81, 525–539.

Li B, Qin Y, Duan H, Yin W, Xia X. 2011. Genome-wide characterization of new and drought stress responsive microRNAs in Populus euphratica. Journal of Experimental Botany 62, 3765–3779.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25, 1966–1967.

Li T, Wang S, Wu R, Zhou X, Zhu D, Zhang Y. 2012. Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. Genomics 99, 292–298.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Nam-Hai C. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant Cell 24, 4333–4345.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2^(-Delta Delta CT) method. Methods 25, 402–408.

Lu S, Sun YH, Chiang VL. 2008. Stress-responsive microRNAs in Populus. The Plant Journal 55, 131–151.

Man MZ, Wang X, Wang Y. 2000. POWER_SAGE: comparing statistical tests for SAGE experiments. Bioinformatics 16, 953–959.

Meng Y, Shao C, Wang H, Jin Y. 2012. Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants. BMC Genomics 13,197.

Min XJ, Butler G, Storms R, Tsang A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Research 33, W677–W680.

Nagano T, Fraser P. 2011. No-nonsense functions for long noncoding RNAs. Cell 145, 178–181.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Research 28, 292.

Ng J-H, Ng H-H. 2010. LincRNAs join the pluripotency alliance. Nature Genetics 42, 1035–1036.

Ni Z, Hu Z, Jiang Q, Zhang H. 2013. GmNFYA3, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. Plant Molecular Biology 82, 113–129.

Okazaki Y, Funuro M, Kasukawa T. et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420, 563–573.

Shuai P, Liang D, Zhang Z, Yin W, Xia X. 2013. Identification of drought-responsive and novel Populus trichocarpa microRNAs by high-throughput sequencing and their targets using degradome analysis. BMC Genomics 14, 233.

Tang S, Liang H, Yan D, Zhao Y, Han X, Carlson JE, Xia X, Yin W. 2013. Populus euphratica: the transcriptomic response to drought stress. Plant Molecular Biology 83, 539–557.

Todesco M, Rubio-Somoza I, Paz-Ares J, Weigel D. 2010. A Collection of Target Mimics for Comprehensive Analysis of MicroRNA Function in Arabidopsis thaliana. PLoS Genetics 6, e1001031.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28, 511–515.

Tuskan GA, Difazio S, Jansson S. et al. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313, 1596–1604.

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147, 1537–1550.

Wu H-J, Wang Z-M, Wang M, Wang X-J. 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. Plant Physiology 161, 1875–1884.

Wu J, Okada T, Fukushima T, Tsudzuki T, sugiura M, Yukawa Y. 2012. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis. RNA Biology 9, 302–313.

Xin M, Wang Y, Yao Y. et al. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. BMC Plant Biology 11, 61.

Yan J, Gu Y, Jia X, Kang W, Pan S, Tang X, Chen X, Tang G. 2012. Effective small RNA destruction by the expression of a short tandem target mimic in Arabidopsis. The Plant Cell 24, 415–427.

Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu J-L, Ponting CP. 2012. Identification and properties of 1,119 candidate lincRNA loci in the Drosophila melanogaster genome. Genome Biology and Evolution 4, 427–442.

Zhou F, Xu Y. 2009. RepPop: a database for repetitive elements in Populus trichocarpa. BMC Genomics 10, 14.