RESEARCH ARTICLE

Identification and characterization of long noncoding RNA in *Paulownia tomentosa* treated with methyl methane sulfonate

Zhe Wang¹ · Bingbing Li¹ · Yongsheng Li^{1,2} · Xiaoqiao Zhai³ · Yanpeng Dong^{1,2} · Minjie Deng^{1,2} · Zhenli Zhao^{1,2} · Yabing Cao¹ · Guoqiang Fan^{1,2}

Received: 5 June 2017 / Revised: 18 January 2018 / Accepted: 24 January 2018 / Published online: 8 February 2018
© Prof. H.S. Srivastava Foundation for Science and Society 2018

Abstract *Paulownia* is a tree native to China, with important ecological and economic value. Long noncoding RNAs (lncRNAs) are known to play important roles in eukaryotic gene regulation. However, no lncRNAs have been reported in *Paulownia* so far. We performed RNA sequencing of two *Paulownia tomentosa* lncRNA libraries constructed from the terminal buds of normal untreated seedlings and 60 mg L⁻¹ MMS-treated seedlings, and obtained a total of 2531 putative lncRNAs. The average length of the lncRNA transcripts was much less than the average length of the mRNA transcripts in the *P. tomentosa* libraries. A few of the *Paulownia* lncRNAs were conserved among ten species tested. We identified seven lncRNAs as precursors of 13 known miRNAs, 15 lncRNAs may act as target mimics of 19 miRNAs, and 351 unique noncoding sequences belonging to 133 conserved lncRNA families. In addition, we identified 220 lncRNAs responsive to methyl methane sulfonate (MMS), including seven phytohormone-related lncRNAs and one lncRNAs involved in base excision repair. This is the first time that lncRNAs have been

explored in *Paulownia*. The lncRNA data may also provide new insights into the MMS-response in *P. tomentosa*.

Keywords *Paulownia* · lncRNAs · Target gene · MMS

Introduction

The genomes of most organisms contain DNA that does not encode proteins (Wilusz et al. 2009). This DNA is called noncoding DNA. Some noncoding DNA is transcribed into functional noncoding RNA molecules (e.g. transfer RNA, ribosomal RNA (rRNA), and regulatory RNAs), the remaining noncoding DNA is either not transcribed or produces noncoding RNAs (ncRNAs), which are predicted to play essential roles in a variety of biological processes (Pauli et al. 2011). The ncRNAs mainly take three forms, housekeeping RNAs, small ncRNAs, and long ncRNAs (lncRNAs) (Knauss and Sun 2013). Some housekeeping and small ncRNAs have been characterized, and lncRNAs have been implicated to play roles in regulating gene expression and cellular functions, especially in cancer (Pan et al. 2017; Su et al. 2017).

In plants, lncRNAs have been found to play important roles in growth, development, and differentiation, and in stress responses (Ben Amor et al. 2009). A cold-induced *Arabidopsis* lncRNA was shown to play an early role in the epigenetic silencing of the FLC transcription factor, by silencing FLC transcription transiently (Swiezewski et al. 2009). In rice, a lncRNA, XLOC_057324, was demonstrated to play a role in panicle development and fertility, and a large number of lncRNAs involved in sexual reproduction has been identified (Zhang et al. 2014). Another rice lncRNA was found to regulate photoperiod-sensitive male sterility, which is an essential component of hybrid

Zhe Wang and Bingbing Li have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12298-018-0513-8>) contains supplementary material, which is available to authorized users.

✉ Guoqiang Fan
zlx64@126.com; fanguoqiangdr@163.com

¹ Institute of *Paulownia*, Henan Agricultural University, Zhengzhou, Henan, People's Republic of China

² College of Forestry, Henan Agricultural University, Zhengzhou, Henan, People's Republic of China

³ Henan Academy of Forestry, Zhengzhou, Henan, People's Republic of China

rice (Ding et al. 2012). In cotton, several lncRNAs might be involved in fiber initiation and elongation (Wang et al. 2015a). In *Fragaria vesca*, tissue-specific lncRNA expression, important to strawberry flower and fruit development, has been identified (Kang and Liu 2015). lncRNAs were suggested to influence growth and wood properties of *Populus tomentosa* (Tian et al. 2016). lncRNAs also showed pivotal functions in plants exposed to biotic and abiotic stresses. In *Brassica napus*, 931 differentially expressed lncRNAs were identified in response to *Sclerotinia sclerotiorum* infection (Joshi et al. 2016). Tomato lncRNA16397 induced SIGRX22 expression to reduce reactive oxygen species accumulation and alleviate cell membrane injury, resulting in enhanced resistance to *Phytophthora infestans* (Cui et al. 2017). In *Medicago truncatula*, lncRNAs regulate the osmotic and salt stress responses of plants through complex regulatory networks with protein-coding genes (Wang et al. 2015b). In cotton, lncRNAs may be involved in regulating plant hormone pathways in response to drought stress (Lu et al. 2016). In the perennial and woody plant, *Populus trichocarpa*, 504 lncRNAs were found to be drought responsive (Shuai et al. 2014). All these studies point to diverse functions of plant lncRNAs, which indicates their importance in functional studies of plants.

Paulownia are perennial woody trees indigenous to China (Essl 2007). Their rapid growth rate, good wood properties, and ability to adapt to poor environmental conditions have given these trees considerable economic and ecological value (Yadav et al. 2013). The genus Paulownia currently comprises seven species and a number of hybrids (The International Plant Names Index). The limited germplasm resource is a restraint in Paulownia improvement. To expand the available Paulownia germplasm resources, methyl methane sulfonate (MMS), which has been used for many years as a DNA damaging agent to induce mutagenesis (Lundin et al. 2005), has been used in Paulownia molecular breeding programs. We used MMS-treated Paulownia to screen the whole genome for lncRNAs, and bioinformatics tools to identify lncRNAs that may play essential roles in a variety of biological processes (Kornienko et al. 2013). The results may provide new potential in light of Paulownia breeding programs.

Using the complete Paulownia genome sequencing (<http://paulownia.genomics.cn>), we aimed to identify and characterize the lncRNAs in Paulownia. lncRNAs target genes were predicted to help in understanding the functions of lncRNAs, and the lncRNAs associated with the MMS-response were revealed.

Materials and methods

Plant materials

All the biological materials used in this study were obtained from Institute of Paulownia, Henan Agricultural University, China. Tissue-cultured seedlings of *Paulownia tomentosa* were cultured for 30 days before being clipped from the roots. The cultivation and MMS treatment procedures were as described by Fan et al. (2015). Details were provided in the supplemental protocol 1. The terminal buds were sheared from the normal untreated seedlings (PT) and 60 mg L⁻¹ MMS-treated seedlings (PT-60) of *P. tomentosa*, and stored at -80 °C for further use.

RNA preparation, lncRNA library construction, and high-throughput sequencing

Total RNA was extracted from the terminal buds of *P. tomentosa* using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and treated with RNase-free DNase (Promega, Madison, WI) according to the manufacturer's instructions. The quality and quantity of total RNA was determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). Integrity of the isolated RNA was determined on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using the Plant RNA Nano chip assay in accordance with the manufacturer's instructions. Total RNA was used to construct RNA-Seq libraries using a TruSeq Stranded Total RNA preparation kit with RiboZero Plant (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. The libraries were sequenced on an Illumina HiSeq 2000 platform. Reads with more than 10% N (Unable to determine base information), with adapter sequence, or of low quality were removed from the raw reads to obtain clean reads. To remove the remaining ribosomal RNA (rRNA) reads, the clean reads were mapped to the SILVA rRNA database (<http://www.arb-silva.de/>) using the soft read alignment software SOAP2. Finally, the clean reads were compared with the Paulownia genome sequence using TopHat2 (Trapnell et al. 2012). The libraries preparation and deep sequencing were performed by Beijing Genomics Institute (Shenzhen, China).

lncRNA identification

The assembled transcripts were annotated using the Cuffcompare program from the Cufflinks package (Trapnell et al. 2012). The known protein-coding transcripts were identified according to the Paulownia genome sequence annotations. The remaining unknown transcripts were used

to screen for putative lncRNAs. Transcripts smaller than 200 bp were excluded. Then, the coding potential for the remaining transcripts was determined by the Coding Potential Calculator based on quality, completeness, and sequence similarity with the open reading frames of proteins in the protein databases (Kong et al. 2007). A transcript was deemed to be noncoding if the coding potential was scored to be less than 0, which suggested the transcript had no protein-coding capacity. The MMS-response lncRNAs were identified according to the methods of previous studies (Cui et al. 2017; Tian et al. 2016).

Genomic characterization of lncRNAs

We determined the distribution of the identified lncRNA on the *Paulownia* chromosomes according to the description of Li et al. (2016). The lncRNAs were classified into several categories according to their genomic location and the method of Roberts et al. (2011). To explore lncRNA conservation, all the lncRNA sequences identified in this study were aligned against CANTATadb (<http://yeti.amu.edu.pl/CANTATA>) using BLASTN (E-value < 1e−5). To identify lncRNAs that may act as microRNA (miRNA) precursors, we aligned the lncRNA sequences to the miRNA sequences in miRBase (Release 21, <http://www.mirbase.org/>), criteria were sequence coverage more than 90%. The secondary structures of the lncRNAs and the miRNA precursors were predicted using RNAfold in the Vienna RNA package (<http://rna.tbi.univie.ac.at/>). Analysis of lncRNAs as target mimic of miRNAs were done according to the method in a previous study (Wu et al. 2013).

Target gene prediction

Many studies have indicated that one major function of lncRNAs is regulating the expression of neighboring protein-coding genes (Mercer et al. 2009; Tian et al. 2016; Wang et al. 2015b). The relative loci between lncRNAs and their neighbors can be exhibited using Integrative Genomics Viewer (Robinson et al. 2011). We used an algorithm that searches for potential cis target genes that were physically close (within 10 kb) to the lncRNAs (Jia et al. 2010). We used another algorithm to search for potential trans target genes in the *Paulownia* mRNA database. This algorithm is based on mRNA sequence complementarity and RNA duplex energy prediction, which assess the impact of lncRNA binding on complete mRNA molecules. First, we used BLAST to select target sequences complementary to the lncRNAs, setting E-value < 1e−5 and identity ≥ 95%. Then, we used the RNAplex software to calculate the complementary energy between the lncRNA and target gene sequences to select

potential trans-acting target genes (Tafer and Hofacker 2008).

Real-time quantitative PCR

The total RNA was reverse transcribed into cDNA and used to measure the expression of lncRNAs by quantitative Real-Time PCR (qRT-PCR). The qRT-PCR was performed on a DNA Engine Opticon 2 machine (MJ Research, Waltham, MA, US) using a LightCycler FastStar DNA master SYBR Green I kit (Roche Diagnostics, Mannheim, Germany). The primers were designed using Primer Express 3.0 (Applied Biosystems, Stockholm, Sweden) and the specificity of primer pairs was checked by sequencing the PCR products. All qRT-PCR amplifications were carried out in triplicate, with the standard reaction program. The specificity of the amplified fragments was checked using the generated melting curve. The generated real-time data were analyzed using the Opticon Monitor Analysis Software 3.1 tool and standardized to the levels of 18S RNA using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001). The primers used for qRT-PCR are listed in Table S1.

Gene ontology and pathway enrichment analysis

Gene Ontology (GO) terms were assigned to the predicted target genes with AgriGO (<http://bioinfo.cau.edu.cn/agriGO>) (Du et al. 2010). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.ad.jp/kegg>) was used to assign the target genes to pathways, and an enrichment analysis was conducted to analyze the potential functions of the target genes.

Results

Identification and characterization of lncRNAs in *Paulownia*

We used the total RNA extracted from the terminal buds of normal untreated seedlings (PT) and 60 mg L^{−1} MMS-treated seedlings (PT-60) of *P. tomentosa* to construct RNA-Seq libraries that were sequenced on an Illumina HiSeq 2000 platform. Details of the RNA-Seq data are shown in Table S2. A genome-wide systematic identification of lncRNAs found 2515 and 2499 lncRNAs in the PT and PT-60 libraries, respectively (Fig.S1). In total, we obtained 2531 lncRNAs, including 42 lncRNAs that had exonic overlap with a known transcript (class code ‘o’), 2071 intergenic lncRNAs (class code ‘u’), 392 exonic lncRNAs present on the opposite strand (class code ‘x’), and 26 lncRNAs with class code ‘j’ representing novel long

noncoding isoforms with at least one splice junction shared with the reference genes (Table 1). Most of the lncRNAs were intergenic, which is consistent with other studies. Detailed information of the 2531 putative lncRNAs is shown in Table S3.

We further characterized the basic genomic features of the 2531 lncRNAs. The distribution of the lncRNAs in the Paulownia chromosomes showed that they were evenly distributed in each chromosome (Fig. 1c). All the lncRNAs were searched against the genomes of *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Selaginella moellendorffii*, *Hieron*, *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Amborella trichopoda*, *Solanum tuberosum*, *Vitis vinifera*, and *Zea mays* using BLAST. The results showed that only a few of the Paulownia lncRNAs were conserved across the ten species (Table 2 and S4); the highest number matched *V. vinifera* lncRNAs, likely because Paulownia and *V. vinifera* are the most closely related in evolution. The longest conserved lncRNA sequence length was 564 nt, which matched a lncRNA in potato. Interestingly, in *Amborella*, a relatively lower plant, the number of conserved lncRNAs and the conserved lncRNA sequence length ranked at the middle level among the ten species. In addition, only nine of the Paulownia lncRNAs were conserved with *Z. mays* lncRNAs, and the longest conserved lncRNA sequence was 225 nt, which is lower than that in *Chlamydomonas* (17 conserved lncRNAs, 65nt), *Physcomitrella* (17 conserved lncRNAs, 73nt), and *G. max* (21 conserved lncRNAs, 178nt). The conservation of lncRNAs can be judged by the number of conserved lncRNAs as well as the length of conserved lncRNA sequences.

The lengths of the Paulownia lncRNA transcripts (median length of 977 nucleotides) were shorter than *Populus* lncRNA transcripts (Tian et al. 2016), while the lengths of the transcripts of Paulownia protein-coding genes (median length of 1913 nucleotides) were longer than those of *Populus* (Tian et al. 2016) (Fig. 1b). And the lncRNA transcript sequences were shorter than the transcripts of protein-coding genes, which is consistent with previous studies (Li et al. 2016).

We examined the lncRNA sequences to determine if they could be precursors of known miRNAs. We aligned

the 2531 lncRNAs to miRNA precursor sequences from miRBase, and identified seven lncRNAs as possible precursors of 13 known miRNAs (Table S5). We identified 15 lncRNAs that may act as target mimics of 19 miRNAs (Table S6). To annotate the lncRNAs from an evolutionary point of view, we classified the Paulownia lncRNAs into different ncRNA families using INFERNAL (Nawrocki et al. 2009). Based on consensus secondary structure annotation using a covariance model, we identified 350 unique ncRNA sequences belonging to 133 conserved lncRNA families (Table S7). MiRNA and small nucleolar RNA families accounted for most of the conserved lncRNA families.

Validation of transcription levels of Paulownia lncRNAs and genes

To confirm the expression of Paulownia lncRNAs and their response to MMS, we performed a qRT-PCR analysis to verify the results of the high-throughput RNA-Seq. We randomly selected twelve lncRNAs and genes for qRT-PCR validation. The expression levels of 11 of the selected lncRNAs and all of the selected mRNAs matched those of the high-throughput RNA-Seq data. The results demonstrated that in most cases, the qRT-PCR results were consistent with those of RNA-Seq (Figs. 2 and 3).

lncRNA target prediction, annotation, and enrichment analysis

The identification and analysis of target genes of lncRNAs could give insights into their role as lncRNAs have roles in regulating gene expression. We identified 6302 potential target genes for 2379 lncRNAs; among them, 2021 genes were predicted as trans-regulated genes that were targeted by 2108 lncRNAs, and 4672 genes were predicted as cis-regulated genes that were targeted by 2202 lncRNAs (Table S8). Our results show that a single target gene could be regulated by one or more lncRNAs, and one lncRNA could have more than one target gene. Among these lncRNAs, 265 lncRNAs had only one target gene each, and four lncRNAs (TCONS_00003323, TCONS_00003324,

Table 1 class code of Paulownia lncRNAs

Class code	Transcript number	Description
J	26	Novel long noncoding isoforms with at least one splice junction shared with the reference genes
O	42	lncRNAs that had exonic overlap with a known transcript
U	2071	Intergenic lncRNAs
X	392	lncRNAs that had exonic overlap with reference on the opposite strand
Total	2531	

Fig. 1 Characteristics of *P. tomentosa* lncRNAs.

a Classification of lncRNAs, class code 'u': intergenic lncRNAs, class code 'o': lncRNAs that had exonic overlap with a known transcript, class code 'x': exonic lncRNAs present on the opposite strand, class code 'j': novel long noncoding isoforms with at least one splice junction shared with the reference genes;

b lncRNAs are shorter than protein-coding transcripts;

c Distribution of lncRNAs along each chromosome. The green was representing PT and the red was representing PT-60 (color figure online)

TCONS_00011440, TCONS_00015211) had 11 target genes each. Furthermore, among these target genes, 4483 target genes were targeted by only one lncRNA each, and four target genes (PAU020236.1, TCONS_00003332, TCONS_00021835, TCONS_00021836) were targeted by 10 lncRNAs each (Table S8).

lncRNAs responsive to MMS

Differentially expressed lncRNAs and genes with P values < 0.05 and fold change > 2 or < 2 (fold change = FPKM of PT/FPKM of PT-60) were regarded as statistically significant. We identified 220 lncRNAs with MMS-responsive expression patterns; 62 were up-regulated and 158 were down-regulated (Table S9). Nine and 15 of these lncRNAs exhibited > 10 -fold increases and decreases in

Table 2 Summary of the conserved lncRNAs

	Total number	Aligned length	Identity	Coverage $\geq$ 10% number
Selaginella	3	28–38	97–100	0
Zea	9	26–225	82–100	2
Chlamydomonas	17	26–65	92–100	1
Physcomitrella	17	27–73	95–100	2
Glycine	21	29–178	80–100	6
Arabidopsis	25	25–131	82–100	6
Amborella	29	28–311	80–100	11
Oryza	38	26–215	79–100	6
Potato	79	27–564	79–100	27
Vitis	89	26–355	80–100	32

Fig. 2 Expression pattern confirmation of selected lncRNAs by qRT-PCR. **a** Changes in the relative expression levels of 12 selected lncRNAs as determined by RNA-Seq. **b** Changes in the relative expression levels of 12 selected lncRNAs as determined by qRT-PCR. 18S rRNA of *Paulownia* was used as the reference gene. Error bars represent the standard error of the mean

expression, respectively. A large proportion of the differentially expressed lncRNAs exhibited a < fourfold change, and 11 and 5 lncRNAs were expressed only in PT or PT-60, respectively. Among the MMS-responsive lncRNAs, 162 (34 up- and 128 down-regulated) belonged to class code 'u', and 54 (25 up- and 29 down-regulated), 3 (up-

regulated), and 1 (down-regulated) belonged to class codes 'x', 'j', and 'o', respectively.

We identified 681 target genes for the 220 MMS-response lncRNAs (Table S10). Among them, 173 trans-regulated target genes were predicted for 173 MMS-response lncRNAs, and 577 cis-regulated target genes were

Fig. 3 Expression pattern confirmation of selected target genes by qRT-PCR. **a** Changes in the relative expression levels of 12 selected target genes as determined by RNA-Seq. **b** Changes in the relative expression levels of 12 selected target genes as determined by qRT-PCR. PAU007003.1, transcription factor MYC2. PAU022674.1, probable inactive receptor kinase. PAU000675.1, probable leucine-rich repeat receptor-like protein kinase. PAU014164.1, phytochrome-interacting factor 3. PAU001093.1, auxin influx carrier. PAU026922.1, ethylene-responsive transcription factor 1. PAU001094.1, high mobility group protein B1. PAU003066.2, hypothetical protein. PAU004416.1, hypothetical protein. PAU005810.1, phenylalanyl-tRNA synthetase alpha chain. PAU007232.1, SQUAMOSA promoter binding protein-like 13. PAU004685.1, amino acid permease. 18S rRNA of *Paulownia* was used as the reference gene. Error bars represent the standard error of the mean

predicted for 176 MMS-response lncRNAs. Among the 681 predicted target genes, 94 were differentially expressed in PT vs. PT-60 and were targeted by 76 MMS-response lncRNAs (Table S11). Of the 94 target genes, 17 were trans-regulated target genes for 17 MMS-response lncRNAs, and 90 were cis-regulated target genes for 66 MMS-response lncRNAs. Clearly, one MMS-response lncRNA corresponded to more than one cis/trans-regulated target gene and vice versa. In total, 107 lncRNAs–target pairs were identified; 47 of the target genes were up-regulated and 60 were down-regulated. We also found different expression changes between lncRNAs and their targets. In PT vs. PT-60, the trend of expression changes was the same for 72 lncRNAs and their target genes, and opposite for 35 lncRNAs and their targets.

We performed GO analysis for the 94 differentially expressed target genes, and assigned 15 function groups under the three main GO categories, biological process, cell component, and molecular function, in which “cellular process” and “metabolic process”, “cell” and “cell part”, and “catalytic activity” were the largest groups, respectively (Fig. 4). The KEGG analysis results indicated that the target genes were enriched in 38 pathways, among which, the plant hormone signal transduction pathway was significant enriched (Table S12). Seven potential target genes for seven lncRNAs were related to this pathway; with two target genes related to signal transduction of auxin, two with jasmonic acid, and one each with gibberellin, ethylene, and brassinosteroid, respectively. These results suggested that MMS may significantly influence

Fig. 4 GO analysis for the differentially expressed target genes

hormone signal transduction in *Paulownia*. Furthermore, one target gene of a lncRNA was mapped to the base excision repair (BER) pathway, which may be involved in repairing the DNA damage caused by alkylating agents (MMS) (Lindahl and Wood 1999). This result is consistent with the idea that MMS can cause DNA damaging.

Discussion

lncRNAs have been shown to play important roles in plant development, and other biological processes. Compared with animals, including human, the study of lncRNAs in plants has lagged behind, and most studies have been carried out in *Arabidopsis*. Recently, some lncRNAs associated with growth and wood properties were identified in trees (*Populus*) (Chen et al. 2015; Shuai et al. 2014; Tian et al. 2016). Because *Paulownia* has important ecological and economic value, we investigated the expression profiles of lncRNA in *P. tomentosa* seedlings. In this study, we used a high-throughput sequencing technique to scan the whole genome

of *Paulownia*, and identified 2531 lncRNAs; among them, 26, 42, 392, and 2071 lncRNAs were identified as class code 'j', 'o', 'u', and 'x', respectively. This categorization is similar to the one reported for lncRNAs in *Brassica napus* (Joshi et al. 2016). The lncRNA sequences were shorter than the mRNA sequences, which is consistent with other studies (Hao et al. 2015; Wang et al. 2014; Zhang et al. 2014). In addition, we found that some lncRNAs might be precursors and target mimics of miRNAs.

Plant hormones have varied roles including regulation of signaling networks in response to biotic and abiotic stresses and plant developmental processes. In this study, we discovered seven lncRNAs that targeted seven plant hormone-related mRNAs responsive to MMS. The target genes were related to auxin, brassinosteroid, gibberellin, ethylene, and jasmonic acid, which suggested the lncRNAs may modulate genes involved in plant hormone signal transduction. Among the phytohormone-related lncRNAs, three were up-regulated and four were down-regulated. Among their target genes, three were up-regulated and four were down-regulated. lncRNAs are known to regulate

target genes in various ways. For example, TCONS_00001122 was up-regulated and its target gene, auxin influx carrier, was also up-regulated; while TCONS_00024558 was up-regulated and its target gene, integrator complex subunit 8, was down-regulated. We found a considerable number of mRNAs that may be modulated by multiple lncRNAs, and lncRNAs that had multiple targets. In the seven phytohormone-related lncRNAs–target pairs, one lncRNA had one target; however, TCONS_00001122 had another target that was not related to phytohormones. Plant hormones may be regulated by MMS-response lncRNAs, implying that MMS could disturb the balance of phytohormone. DNA damage caused by alkylating agents is repaired predominantly by the BER pathway. BER contributes to correction of endogenous changes such as damage base excision, uracil arising from deamination of cytosine, and i DNA demethylation (Lindahl and Wood 1999). In a previous study, we showed that MMS could influence DNA methylation levels (Cao et al. 2014). Target genes (PAU001094.1) of MMS-response lncRNAs mapped to the BER pathway, indicating MMS treatment may influence BER and result in DNA damage. In future research, we will focus on these MMS-responsive lncRNAs, especially those involved in plant hormone signal transduction and BER.

Conclusions

In this study, we reported the profiling of *Paulownia* lncRNAs on a genome-wide scale for the first time. We identified 2531 lncRNAs, and 2379 of them had 9197 potential target genes. Most lncRNAs had more than one target gene, and most target genes were regulated by multiple lncRNAs. Moreover, we identified 220 lncRNAs and 107 lncRNA–target pairs that were responsive to MMS. The lncRNA data may provide new insights into the complicated regulatory mechanisms of MMS, which could then be used to conduct extensive genetic engineering for potential improvements in *Paulownia* breeding programs. Analysis of the predicted lncRNA target genes will help to elucidate the mutagenesis effect of MMS treatments.

Acknowledgements This work was supported by the Key Science and Technology Program of Henan Province of China (No. 152107000097), the Natural Science Foundation of Henan Province of China (No. 162300410158), and the Distinguished Talents Foundation of Henan Province of China (No. 174200510001).

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

References

- Ben Amor B et al (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res* 19:57–69. <https://doi.org/10.1101/gr.080275.108>
- Cao X, Fan G, Deng M, Zhao Z, Dong Y (2014) Identification of genes related to *Paulownia* witches' broom by AFLP and MSAP. *Int J Mol Sci* 15:14669–14683. <https://doi.org/10.3390/ijms150814669>
- Chen J, Quan M, Zhang D (2015) Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq. *Planta* 241:125–143. <https://doi.org/10.1007/s00425-014-2168-1>
- Cui J, Luan Y, Jiang N, Bao H, Meng J (2017) Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lncRNA16397 conferring resistance to *Phytophthora infestans* by co-expressing glutaredoxin. *Plant J* 89:577–589. <https://doi.org/10.1111/tpj.13408>
- Ding J et al (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci USA* 109:2654–2659. <https://doi.org/10.1073/pnas.1121374109>
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:W64–W70. <https://doi.org/10.1093/nar/gkq310>
- Essl F (2007) From ornamental to detrimental? The incipient invasion of Central Europe by *Paulownia tomentosa*. *Preslia* 79:377–389
- Fan G, Cao X, Zhao Z, Deng M (2015) Transcriptome analysis of the genes related to the morphological changes of *Paulownia tomentosa* plantlets infected with phytoplasma. *Acta Physiol Plant* 37:1–12. <https://doi.org/10.1007/S11738-015-1948-Y>
- Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G (2015) Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. *PLoS ONE* 10:e0121800. <https://doi.org/10.1371/journal.pone.0121800>
- Jia H, Osak M, Bogu GK, Stanton LW, Johnson R, Lipovich L (2010) Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* 16:1478–1487. <https://doi.org/10.1261/rna.1951310>
- Joshi RK, Megha S, Basu U, Rahman MH, Kav NN (2016) Genome wide identification and functional prediction of long non-coding RNAs responsive to *Sclerotinia sclerotiorum* infection in *Brassica napus*. *PLoS ONE* 11:e0158784. <https://doi.org/10.1371/journal.pone.0158784>
- Kang C, Liu Z (2015) Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. *BMC Genom* 16:815. <https://doi.org/10.1186/s12864-015-2014-2>
- Knauss JL, Sun T (2013) Regulatory mechanisms of long noncoding RNAs in vertebrate central nervous system development and function. *Neuroscience* 235:200–214. <https://doi.org/10.1016/j.neuroscience.2013.01.022>
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 35:W345–W349. <https://doi.org/10.1093/nar/gkm391>
- Kornienko AE, Guenzl PM, Barlow DP, Pauler FM (2013) Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* 11:59. <https://doi.org/10.1186/1741-7007-11-59>
- Li H et al (2016) Genome-wide long non-coding RNA screening, identification and characterization in a model microorganism *Chlamydomonas reinhardtii*. *Sci Rep* 6:34109. <https://doi.org/10.1038/srep34109>

- Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* 286:233–271
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* (San Diego, Calif) 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
- Lu X et al (2016) Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.). *PLoS ONE* 11:e0156723. <https://doi.org/10.1371/journal.pone.0156723>
- Lundin C, North M, Erixon K, Walters K, Jenssen D, Goldman AS, Helleday T (2005) Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Res* 33:3799–3811. <https://doi.org/10.1093/nar/gki681>
- Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155–159. <https://doi.org/10.1038/nrg2521>
- Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25:1335–1337. <https://doi.org/10.1093/bioinformatics/btp157>
- Pan L et al (2017) Exosomes-mediated transfer of long noncoding RNA ZFAS1 promotes gastric cancer progression. *J Cancer Res Clin Oncol*. <https://doi.org/10.1007/s00432-017-2361-2>
- Pauli A, Rinn JL, Schier AF (2011) Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 12:136–149. <https://doi.org/10.1038/nrg2904>
- Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27:2325–2329. <https://doi.org/10.1093/bioinformatics/btr355>
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>
- Shuai P et al (2014) Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Exp Bot* 65:4975–4983. <https://doi.org/10.1093/jxb/eru256>
- Su J et al (2017) Long noncoding RNA BLACAT1 indicates a poor prognosis of colorectal cancer and affects cell proliferation by epigenetically silencing of p15 Cell. *Death Dis* 8:e2665. <https://doi.org/10.1038/cddis.2017.83>
- Swiezewski S, Liu F, Magusin A, Dean C (2009) Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* 462:799–802. <https://doi.org/10.1038/nature08618>
- Tafer H, Hofacker IL (2008) RNAplex: a fast tool for RNA–RNA interaction search. *Bioinformatics* 24:2657–2663
- Tian J et al (2016) Population genomic analysis of gibberellin-responsive long non-coding RNAs in *Populus*. *J Exp Bot* 67:2467–2482. <https://doi.org/10.1093/jxb/erw057>
- Trapnell C et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–578
- Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, Chua NH (2014) Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. *Gen Res* 24:444–453. <https://doi.org/10.1101/gr.165555.113>
- Wang M et al (2015a) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol* 207:1181–1197. <https://doi.org/10.1111/nph.13429>
- Wang TZ, Liu M, Zhao MG, Chen R, Zhang WH (2015b) Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. *BMC Plant Biol* 15:131. <https://doi.org/10.1186/s12870-015-0530-5>
- Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23:1494–1504. <https://doi.org/10.1101/gad.1800909>
- Wu HJ, Wang ZM, Wang M, Wang XJ (2013) Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol* 161:1875–1884. <https://doi.org/10.1104/pp.113.215962>
- Yadav NK, Vaidya BN, Henderson K, Lee JF, Stewart WM, Dhekney SA, Joshee N (2013) A review of *Paulownia* biotechnology: a short rotation fast growing multipurpose bioenergy tree American. *J Plant Sci* 04:2070–2082. <https://doi.org/10.4236/ajps.2013.411259>
- Zhang YC et al (2014) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Gen Biol* 15:512. <https://doi.org/10.1186/s13059-014-0512-1>