Article

Changes in DNA Methylation in Response to 6-Benzylaminopurine Affect Allele-Specific Gene Expression in *Populus tomentosa*

Anran Xuan [1,2], Yuepeng Song [1,2], Chenhao Bu [1,2], Panfei Chen [1,2], Yousry A. El-Kassaby [3] and Deqiang Zhang [1,2,*]

[1] National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, No. 35, Qinghua East Road, Beijing 100083, P. R. China; xuananran@bjfu.edu.cn (A.X.); yuepengsong@bjfu.edu.cn (Y.S.); BuChenhao@bjfu.edu.cn (C.B.); PanfeiChen@bjfu.edu.cn (P.C.)

[2] Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, College of Biological Sciences and Technology, Beijing Forestry University, No. 35, Qinghua East Road, Beijing 100083, P. R. China

[3] Department of Forest and Conservation Sciences, Faculty of Forestry, Forest Sciences Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; y.el-kassaby@ubc.ca(El-Kassaby YA)

* Correspondence: DeqiangZhang@bjfu.edu.cn(D.Z.)

Abstract: The cytokinins play important roles in plant growth and development by regulating gene expression at genome wide level. DNA methylation is responsive to the external environment, but whether DNA methylation changes in response to cytokinin treatment to regulate gene expression is still unclear. Here, we used bisulfite sequencing and RNA sequencing to examine genome-wide DNA methylation and gene expression patterns in poplar (*Populus tomentosa*) after treatment with the synthetic cytokinin 6-benzylaminopurine (6-BA). We identified 566 significantly differentially methylated regions (DMRs) in response to 6-BA treatment. Transcriptome analysis showed that 501 protein-coding genes, 262 long non-coding RNAs, and 15,793 24-nt small interfering RNAs were differentially expressed under 6-BA treatment. Among these, 79% were differentially expressed between alleles in *P. tomentosa*. Combined DNA methylation and gene expression analysis demonstrated that DNA methylation plays an important role in regulating allele-specific gene expression. To further investigate the relationship between these 6-BA-responsive genes and phenotypic variation, we performed SNP analysis of 507 6-BA-responsive DMRs via re-sequencing using a natural population of *P. tomentosa* and identified 206 SNPs that were significantly associated with growth and wood properties. Association analysis indicated that 53% of loci with allele-specific expression had primarily dominant effects on poplar traits. Our comprehensive analyses of *P. tomentosa* DNA methylation and the regulation of allele-specific gene expression suggest that DNA methylation is an important regulator of imbalanced expression between allelic loci.

Keywords: allele specific expression; 6-BA; DNA methylation; long noncoding RNA; siRNA; poplar

1. Introduction

Cytokinins are an important class of phytohormones whose discovery was based on their ability to promote cell division in tobacco tissue[1]. These phytohormones, which are synthesized at the root tip[2,3], play various roles in regulating plant growth, development, and differentiation, including meristem function, leaf senescence, source/sink relationships, and vascular development[4-6]. The first synthetic cytokinin, 6-benzylaminopurine (6-BA), was produced in 1952. Exogenous 6-BA treatment delays senescence and improves the quality of chlorophyll-containing vegetables by inhibiting chlorophyll degradation due to its effects on chlorophyll-related gene

expression[7-9]. However, to date, the responses of woody plants to 6-BA have not been systematically studied at the molecular level, including epigenetic and transcriptional analyses.

DNA methylation is an important epigenetic modification that plays crucial roles in regulating genomic functions. For example, the methylation of microRNA genes regulates gene expression during bisexual flower development in andromonoecious poplar[10,11]. In addition, epigenetic modifications (primarily DNA methylation) affect plant growth and development by controlling flowering time[12]. The prominent role of DNA methylation in regulating gene expression raises the question of how whole-genome methylation patterns change in response to exogenous 6-BA application. Understanding this relationship will provide valuable information for the functional analysis of DNA methylation. The analysis of whole-genome methylation patterns in response to 6-BA should uncover the precise, specific expression patterns of various transcription regulatory elements, including genes, long non-coding RNAs (lncRNAs), and small interfering RNAs (siRNAs). LncRNAs are transcripts longer than 200 nucleotides that appear to have no coding potential[13]. LncRNAs play important roles in many biological processes, such as dosage compensation, epigenetic regulation, and cell differentiation[14]. For example, gibberellin (GA)-responsive lncRNAs are associated with wood properties in *P. tomentosa*[15].

The 24-nucleotide siRNAs represent crucial links in the epigenetic regulatory network in plants. Changes in 24-nt siRNA levels affect DNA methylation at loci regulated by *RNASE THREE-LIKE PROTEIN 2* (*RTL2*) and are inversely correlated with the steady-state levels of mRNA, thus implicating *RTL2* in the regulation of protein-coding gene expression in *Arabidopsis thaliana*[16]. Similarly, changes in 24-nt siRNA levels in *Arabidopsis* hybrids likely play an epigenetic role in hybrid vigor[17]. Therefore, exploring the response patterns of *P. tomentosa* to 6-BA treatment and the interactions between various genetic regulatory elements and their effects on DNA methylation is of vital importance. The relationship between DNA methylation and other epigenetic modifications remains to be fully investigated.

Exogenous phytohormone treatment affects the diploid plant poplar at the transcriptional level not only by altering gene expression but also by influencing the expression of alleles, as allelic differences also play a critical role in gene regulation. Allelic differential gene expression or allele-specific expression (ASE) is commonly observed in the human genome[18] and often reflects the presence of putative allele-specific cis-acting factors of genetic or epigenetic origin. For example, in the human liver, DNA methylation influences the ASE of *CYP1A2*[19]. The variation in ASE levels in heterozygous loci in an individual is likely due to cis-acting polymorphisms, leading to different levels of mRNA expression[20]. The effect of cis-acting elements on gene expression in plants could be explored by identifying ASE under 6-BA treatment.

Here, we systematically identified 6-BA-responsive differentially methylated regions (DMRs) in *P. tomentosa* leaf tissue at the genome-wide scale. The DMRs were annotated to various genomic transcriptional elements, including lncRNAs, mRNAs, and 24 nt-siRNAs. We then identified SNPs in the DMRs in a natural population of *P. tomentosa*, tested the influence of allelic variation in DMRs on growth and wood properties through association analysis, and dissected their additive and dominant effects. One lncRNA (*TCONS_00053467*) contained multiple loci with prominent dominant effects. Its target gene (*Potri.002G258000*) is homologous to *MALE DISCOVERER1-INTERACTING RECEPTOR LIKE KINASE 2* (*MIK2*) in *A. thaliana*, which encodes an important regulator of responses to cell wall damage triggered by inhibition of cellulose biosynthesis. *Potri.002G258000* is likely involved in the response of poplar to exogenous 6-BA. We also performed ASE analysis of *P. tomentosa* at the whole-genome level in response to 6-BA treatment, finding that the effects of 6-BA treatment on the imbalanced expression of allelic loci is universal. The results of this study increase our understanding of the effects of 6-BA-responsive DNA methylation on allele-specific gene expression in *Populus* and lay the foundation for further research on the regulatory effects of 6-BA on plant growth.

2. Results

2.1. 6-BA treatment affects the physiological and photosynthetic characteristics of poplar

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 February 2020 doi:10.20944/preprints202002.0038.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 2117; doi:10.3390/ijms21062117

To explore whether 6-BA affects the growth and development of *P. tomentosa*, we investigated its short-term effects on physiological characteristics in leaves harvested at 0, 3, 6, 12, and 24 h of treatment. The physiological characteristics included total protein content, POD activity, SPS activity and MDA content (Figure 1A–D). At 6 h of treatment, the 6-BA treated group showed significantly higher total protein content and POD activity than the control (increases of 29.94 and 37.62%, respectively). At 3 and 6 h of treatment, the 6-BA treated group showed significantly lower SPS activity than the control (3.04- and 1.39-fold downregulation, respectively). At 12 h of treatment, the 6-BA treated group showed significantly lower MDA than the control. These results indicate that antioxidant enzyme activity significantly increased at 6 h after 6-BA treatment, and cellular damage occurred after 12 h of treatment. Moreover, they indicate that 6 h of treatment is a key time point to explore the response of poplar to 6-BA treatment; thus, the 6 h time point was selected for subsequent analyses.

To further investigate the long-term effects of 6-BA on *P. tomentosa*, we sprayed plants with 6-BA once a week for a month and measured the photosynthetic characteristics under short-term (one day) and long-term (one month) treatment. These photosynthetic characteristics included net photosynthetic rate (Pn), stomatal conductance (Gs), transpiration rate (Tr), and intercellular CO_2 concentration (Ci). Long-term treatment with 6-BA significantly reduced several of these values, with Pn, Gs, and Tr decreasing by 25.98%, 50.37%, and 37.58%, respectively, compared to the control, whereas Ci remained stable (Figure 1E–H). And water use efficiency (WUE=Pn/Tr) increased by 12.22% (Figure S1).

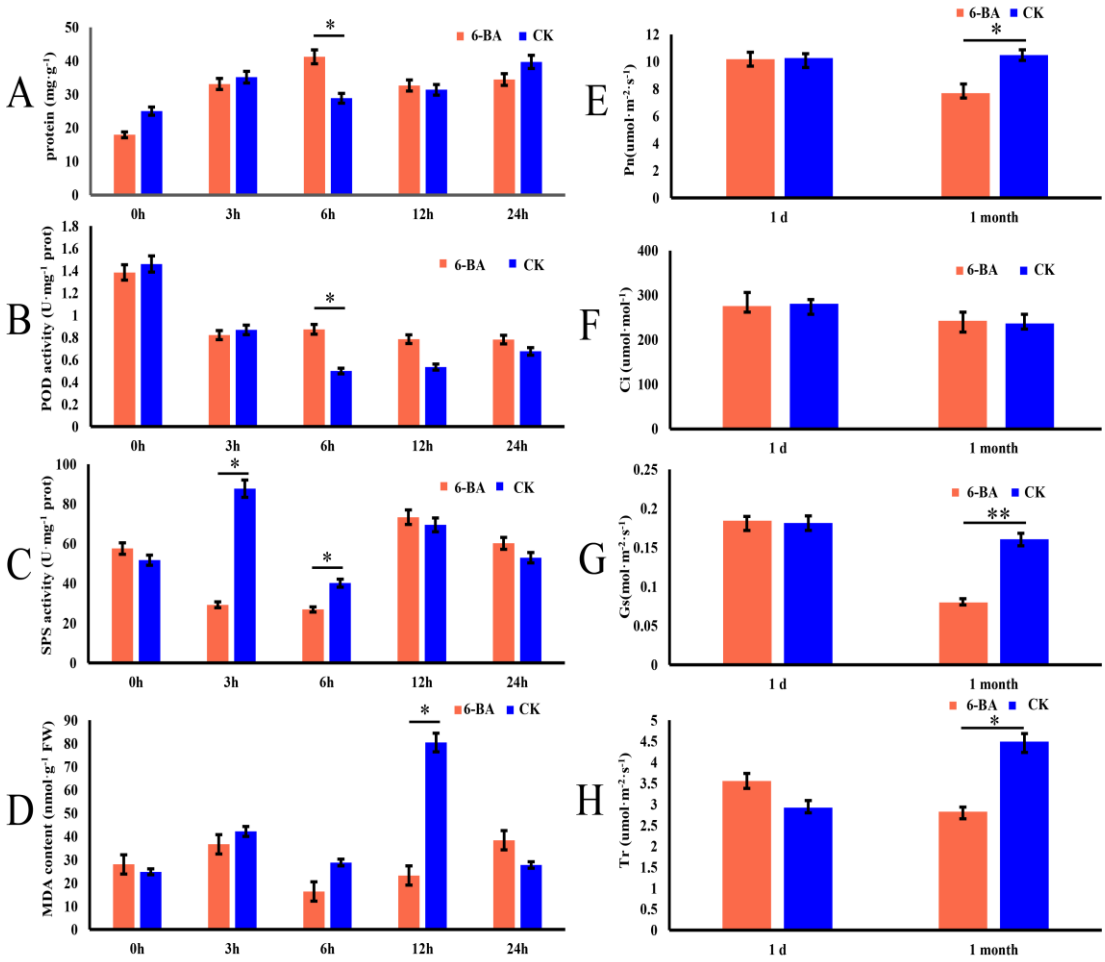


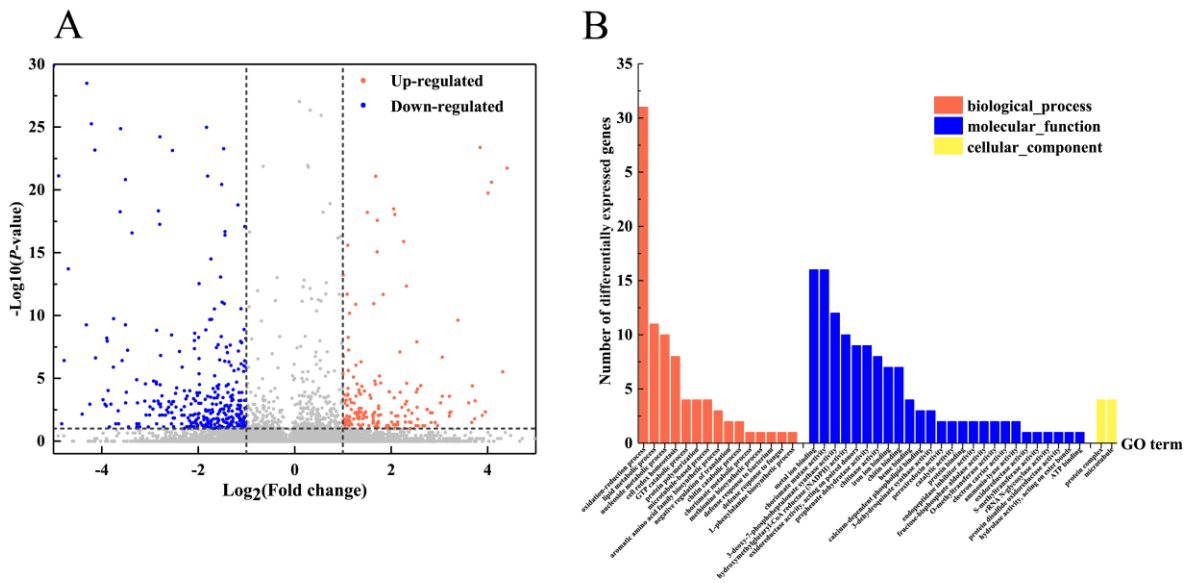
Figure 1 Effects of 6-benzylaminopurine treatment on physiological and photosynthetic characteristics of *P. tomentosa*. A–D Changes in total protein content, peroxidase (POD) activity, sucrose phosphate synthase (SPS) activity, and malondialdehyde (MDA) content in leaves at 0, 3, 6, 12, and 24 h of treatment. **E-H** Changes in net photosynthetic rate (Pn), intercellular CO_2 conductance (Ci),

stomatal conductance (Gs), and transpiration rate (Tr) measured 1 day and 1 month after treatment. Error bars represent standard deviation (SD) of three biological replicates (n=3). Asterisks indicate significant differences between the 6-BA-treated and control groups (*P < 0.05, **P < 0.01).

2.2. *Identification of 6-BA-responsive genes and lncRNAs in P. tomentosa*

Based on the RNA-seq data (Table S1), we identified differentially expressed genes (DEGs) between treatments based on normalized FPKM (Fragments Per Kilobase of exon model per Million mapped reads) values. A total of 501 genes were differentially expressed between the 6-BA-treated and control groups ($|\log_2(\text{fold change})|\geq 1$; and $P<0.05$) (Figure 2A), with 171 and 330 genes up- and downregulated, respectively (Table S2, Figure 3A). Gene ontology (GO) analysis of these 501 DEGs identified 129 enriched GO terms (Table S3, Figure 2B). Among these, 19 GO terms were associated with regulatory functions of biological processes, such as cell redox homeostasis, cellular amino acid biosynthetic process, cell wall macromolecule catabolic process, cellular glucan metabolic process, and cell wall modification. The enrichment of GO terms 'cell wall macromolecule catabolic process', 'cell wall modification', 'pectinesterase activity', and 'cell wall' suggests that the 6-BA-responsive genes affect plant growth and cell wall remodeling. In addition, the enrichment of the term '5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity' suggests that the differential expression of 6-BA-responsive genes is associated with the regulation of DNA methylation.

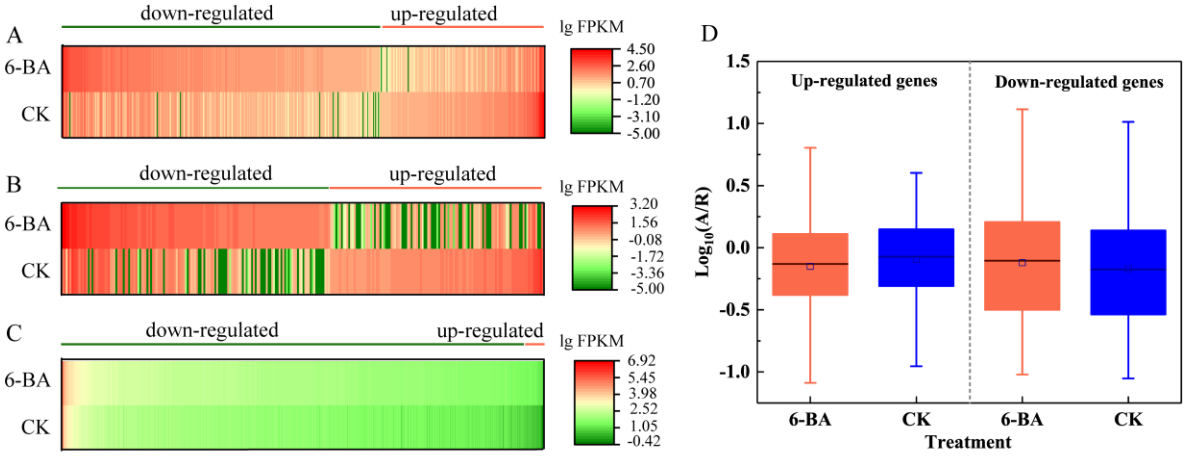
Genome-wide systematic analysis of lncRNAs revealed 6,351 and 4,875 lncRNAs (FPKM >1) from libraries from the control and 6-BA-treated groups, respectively. These 6,283 lncRNAs, which were stably expressed (FPKM>1 in at least one group, FPKM>0 in the other group), included 34 antisense lncRNAs, 69 intergenic lncRNAs, 19 intron lncRNAs, and 6,161 sense lncRNAs (Table S4). An investigation of the basic genomic characteristic of these 6,161 lncRNAs indicated that their transcript lengths ranged from 201 to 6,706 nucleotides, with a mean of 1,021 nucleotides, which are shorter than the transcripts of protein-coding genes of *P. tomentosa* (median length of 1800 nucleotides). In terms of expression levels, the lncRNAs showed fewer average counts (FPKM=8.29) than the protein-coding transcripts (FPKM=18.45). Of the 262 differentially expressed lncRNAs ($|\log_2(\text{fold change})|\geq 1$; and $P<0.05$) selected from the control and 6-BA treated groups, 116 were upregulated and 146 were downregulated (Table S5, Figure 3B). And



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 February 2020 doi:10.20944/preprints202002.0038.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 2117; doi:10.3390/ijms21062117

Figure 2 Analysis of 6-BA-responsive genes in *P. tomentosa*. A Differential expression levels (fold change) of 6-BA-responsive genes. B Functional categorization of 6-BA-responsive genes reveal enrichment in the biological process



gene ontology (GO) category.

Figure 3 Transcriptional variation in response to 6-BA treatment in *P. tomentosa*. A Heat map representing the expression levels of protein-coding genes in the control (CK) and 6-BA treatment group, respectively. B Heat map representing the expression levels of lncRNAs in the CK and 6-BA treatment group, respectively. C Heat map representing the expression levels of 24-nt siRNAs in the CK and 6-BA treatment group, respectively. D Boxplot representing allele-specific expression levels of 6-BA-responsive genes in the CK and 6-BA treatment group, respectively. Error bars represent standard deviation (SD) of three biological replicates (n=3).

2.3. Identification and characterization of 6-BA-responsive 24-nt siRNAs in P. tomentosa

We constructed two small RNA libraries from the control and 6-BA treated groups. After restrictive filtering of contaminating reads, an aggregate of 125,017,531 and 76,823,539 clean reads were obtained, respectively (Table S6). We identified a substantial number of 24-nt siRNAs tags by performing BLAST analysis of the small RNA sequencing reads to the Pln24NT database (see Table S7 for details). Regions containing a substantial number of 24-nt siRNAs tags were defined as 24-nt siRNA clusters. For each treatment group, the 24-nt siRNA tags within these clusters were standardized to reads per million. Genomic regions associated with 24-nt siRNA clusters will be referred to as 24-nt siRNA loci hereafter. Based on annotation with the Pln24NT database, we identified 7,494,036 and 6,545,777 unique tags of 24-nt siRNAs in the control and 6-BA-treated groups, which were derived from 76,690 and 76,072 clusters, respectively (24-nt siRNA clusters merged into 150 bp windows). The two treatment groups significantly differed in terms of the siRNA size classes of unique and total mapped reads. After 6-BA treatment, both the total number of tags and number of unique tags of 24-nt siRNAs decreased, with the number of unique tags decreasing by 45.29%. These findings indicate that the diversity and abundance of 24-nt siRNAs in poplar sharply decreased under 6-BA treatment.

Through expression analysis, we identified 32,836 and 33,101 reliably expressed 24-nt siRNAs in the control and 6-BA-treated groups, respectively. A total of 69,172 clusters occurred in both groups, with only 15,793 (~23%) showing significant differences in siRNA levels ($|\log_2(\text{fold change})| \geq 1$; $P<0.05$), with 38 (0.24%) and 15,755 (99.76%) up- and downregulated, respectively (Table S8, Figure 3C). Among the siRNA clusters with significantly different levels, 1,206 were located in 1,129 protein-coding genes, including 3′ UTR, 5′ UTR, and coding sequence (CDS) regions, whereas most occurred in intergenic regions (Table S8). The underrepresentation of TEs (90.61%) predominantly occurred in genic regions, whereas the overrepresentation of TEs (50.08%) occurred in intergenic regions (Figure S2). Moreover, 99.76% of the significant 24-nt siRNA clusters were downregulated after 6-BA treatment.

2.4. Identification of 6-BA-responsive ASE loci in P. tomentosa

As a diploid plant, poplar contains two sets of genomes, so there will be a small number of allele variation loci (heterozygous loci), most of the rest were homozygous, which are same in both sets of genomes. To explore the differential expression of alleles in poplar in response to 6-BA treatment, we used RNA-sequencing reads mapped to the reference genome to identify loci displaying differential expression between two alleles in the whole genome and genotyped samples from the 6-BA-treated and control groups. ASE analysis was conducted only for heterozygous loci that were expressed at detectable levels in the RNA-sequencing reads. Among these loci, 105,913 (15.61%) were distributed over 19,200 genes, with 1 to 22 heterozygous loci per gene (average 3.8); among these, 78.19% were located in annotated exons (Figure 4A). We detected 102,819 loci in 19,200 genes showing differences in ASE levels after 6-BA treatment (Table S9). To explore the variation in ASE of the 6-BA-responsive genes, we divided these genes into two groups: up- and downregulated genes. As shown in Figure 3D, under 6-BA treatment, the A/R value (Number of alternative loci / Number of reference loci) of the upregulated genes decreased, whereas that of downregulated genes increased, indicating that 6-BA treatment had different effects on genes with different expression patterns. These results indicate that 6-BA treatment causes a universal imbalance in the expression of alleles, as these values significantly deviated from a 1:1 ratio. Furthermore, most differential expression was detected in the exons of genes, thereby playing important roles in regulating gene expression and the translation of proteins controlling growth in *P. tomentosa*.

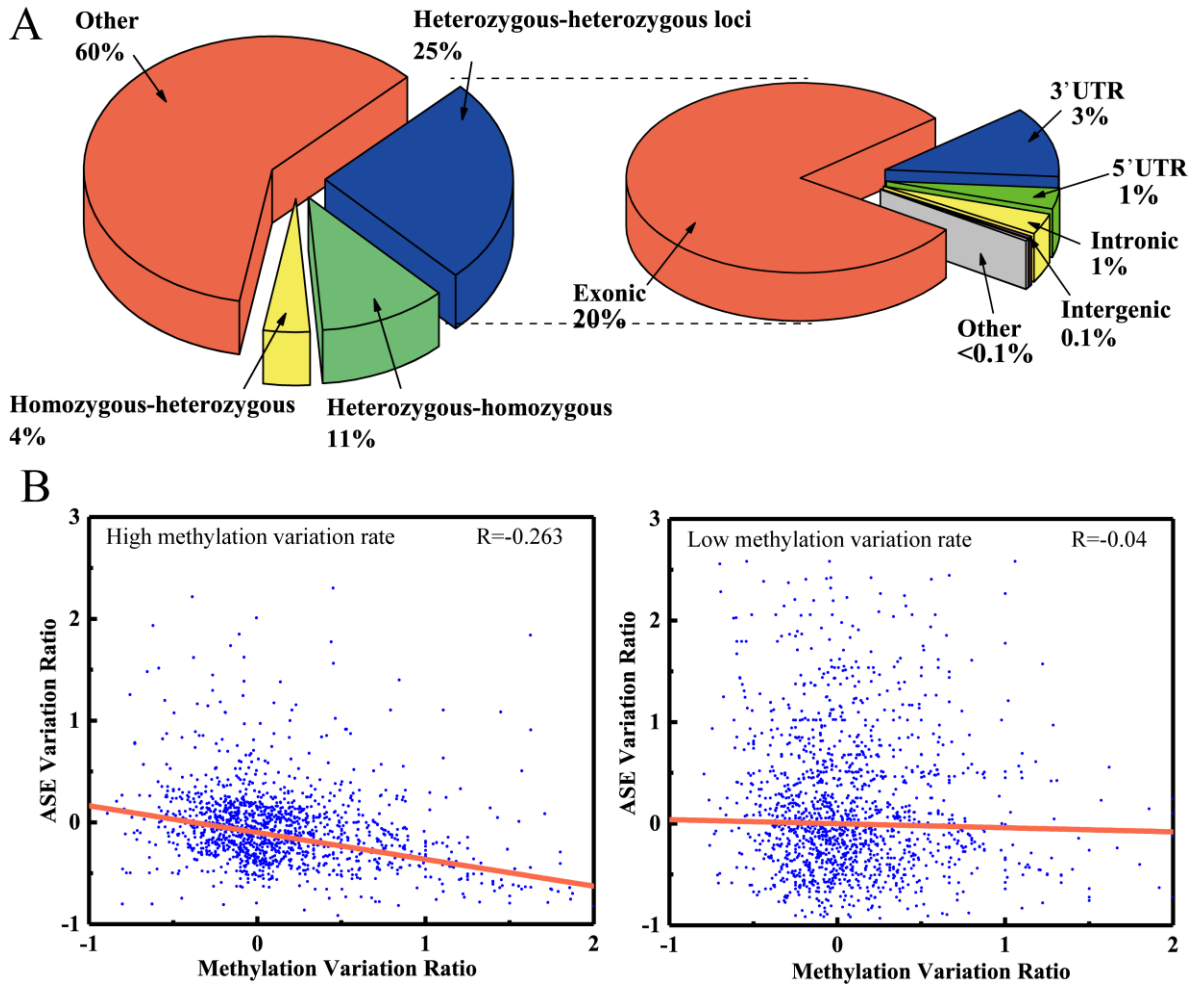


Figure 4 Analysis of *P. tomentosa* allele-specific loci in response to 6-BA treatment. A Pie charts representing the classification of loci showing allelic variation under 6-BA treatment, Heterozygous- heterozygous loci represent the loci was heterozygous in CK and 6-BA groups; Heterozygous-homozygous loci represent the loci was heterozygous in CK group and homozygous in 6-BA group; Homozygous-heterozygous loci represent the loci was homozygous in CK group and heterozygous in 6-BA group; Other represent the loci where no allele variation

occurred, which was homozygous in CK and 6-BA groups (left). And the distribution of allele-specific loci in different regions of the genome under 6-BA treatment (right). B Correlation analysis between ASE level and variation in DNA methylation level in the high and low methylation variation groups.

2.5. Variation of DNA methylation in P. tomentosa under 6-BA treatment

Based on the observed changes in physiological and photosynthetic traits, we used leaves from control and 6-BA treated plants at 6 h of treatment for whole-genome bisulfite sequencing (WGBS) (Table S10). For the control group, ~82 million clean reads were generated, and after filtering low-quality and duplicate reads, 28,762,783 reads were mapped to the *P. trichocarpa* genome (version 3.0) (~34.81%). 3.22% of the cytosine sites were methylated in the whole genome. These sites were classified as being in the CG, GHG (with H representing A, C, or T), or CHH contexts, representing 32.91, 30.63, and 36.45% of the total methylated cytosine sites, respectively. Among 6-BA treated plants, the mapping percentage was approximately 34.12%. After removing low-quality and duplicate reads, ~74 million uniquely mapped high-quality reads were obtained, with 25,367,876 reads mapped to the *P. trichocarpa* genome (version 3.0). 3.22% of the cytosine sites were methylated in the whole genome, including of 31.14% (CG), 29.05% (CHG), and 39.81% (CHH) (Figure 5A). Therefore, the total proportion of methylated cytosines sites did not change under 6-BA treatment, whereas the proportion of methylated cytosines in different contexts changed significantly. These results indicate that the response of *P. tomentosa* to the external environment at the epigenetic level is not only reflected in the number of methylated cytosines sites, but it is also related to the different contexts of methylation, which originate from different biological processes. The proportion of CHH methylation decreased after 6-BA treatment, while the proportion of methylation in the other contexts (CG and CHG) increased. CHH methylation is thought to be related to RNA-mediated DNA methylation (RdDM), which is mediated by 24-nt siRNAs, suggesting that 6-BA-responsive siRNAs might be involved in this progress. Moreover, the variation in methylation levels in different genomic features was consistent in the control and 6-BA-treated groups: both showed peak methylation levels in the promoter region close to the 5′ UTR of the gene, and the methylation level gradually decreased with increasing proximity to the exon region (Figure S3). The distribution of methylation levels in different contexts in the control (CK) and 6-BA-treated groups were shown in Figure 5C.

We then genotyped the methylated CpG sites at the allele level. Since most of the analyzed CpG sites (60.3%) showed little variability, with a low rate of methylation variation [$|(S_{6\text{-}BA}\text{-}S_{CK})| / S_{CK} \leq 0.5$)] under 6-BA treatment, we selected 9,148 CpG loci with allelic-specific expression (ASE) (Table S11) and divided them into two groups according to the S variation rate. These differentially methylated CpG sites were highly informative for identifying quantitative correlations between the methylation of CpG sites and allele-specific gene expression. We compared the ASE levels between the two groups containing CpG loci and performed correlation analysis between the rate of variation of ASE level vs. methylation level of the CpG site. The group with a high rate of methylation variation was more highly correlated with ASE level variation (r=-0.263, P=0.001) than the group with a low rate of methylation variation (r= -0.04, P=0.027) (Figure 4B). These results suggest that variation in methylation level influences the balance of ASE at single CpG loci.

2.6. Variation of differential methylation regions (DMR) under 6-BA treatment

To systematically explore the epigenetic response of poplar to 6-BA treatment, we identified and annotated DMRs based on bisulfite sequencing data. A total of 566 DMRs with differences in magnitude between the control and 6-BA treated groups were defined as DMRs (P<0.05) (23–1895 bp long (Table S12)). We mapped the DMRs to the *P. trichocarpa* genome (version 3.0) based on location information and found that they overlapped with various genomic elements, including protein-coding genes, lncRNAs, and 24-nt siRNA cluster sequences (24-nt siRNA database from Pln24NT). Specifically, we identified 210 protein-coding gene sequences, 22 lncRNA coding sequences, and 172 24-nt siRNA cluster sequences within the boundaries of these DMRs (Figure 5B). Most DMRs contained only a single element sequence, with 188 DMRs containing single

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 February 2020 doi:10.20944/preprints202002.0038.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 2117; doi:10.3390/ijms21062117

protein-coding gene sequences, 21 containing single lncRNA sequences, and 44 containing single 24-nt siRNA sequences. The remaining DMRs contained two genome elements, including 22 DMRs simultaneously occurring in protein-coding genes and siRNA sequences and one DMR simultaneously occurring in an lncRNA and siRNA (Table S13). Transposable elements were present in 87.8% of the 24-nt siRNA clusters within DMR boundaries (Table S8).

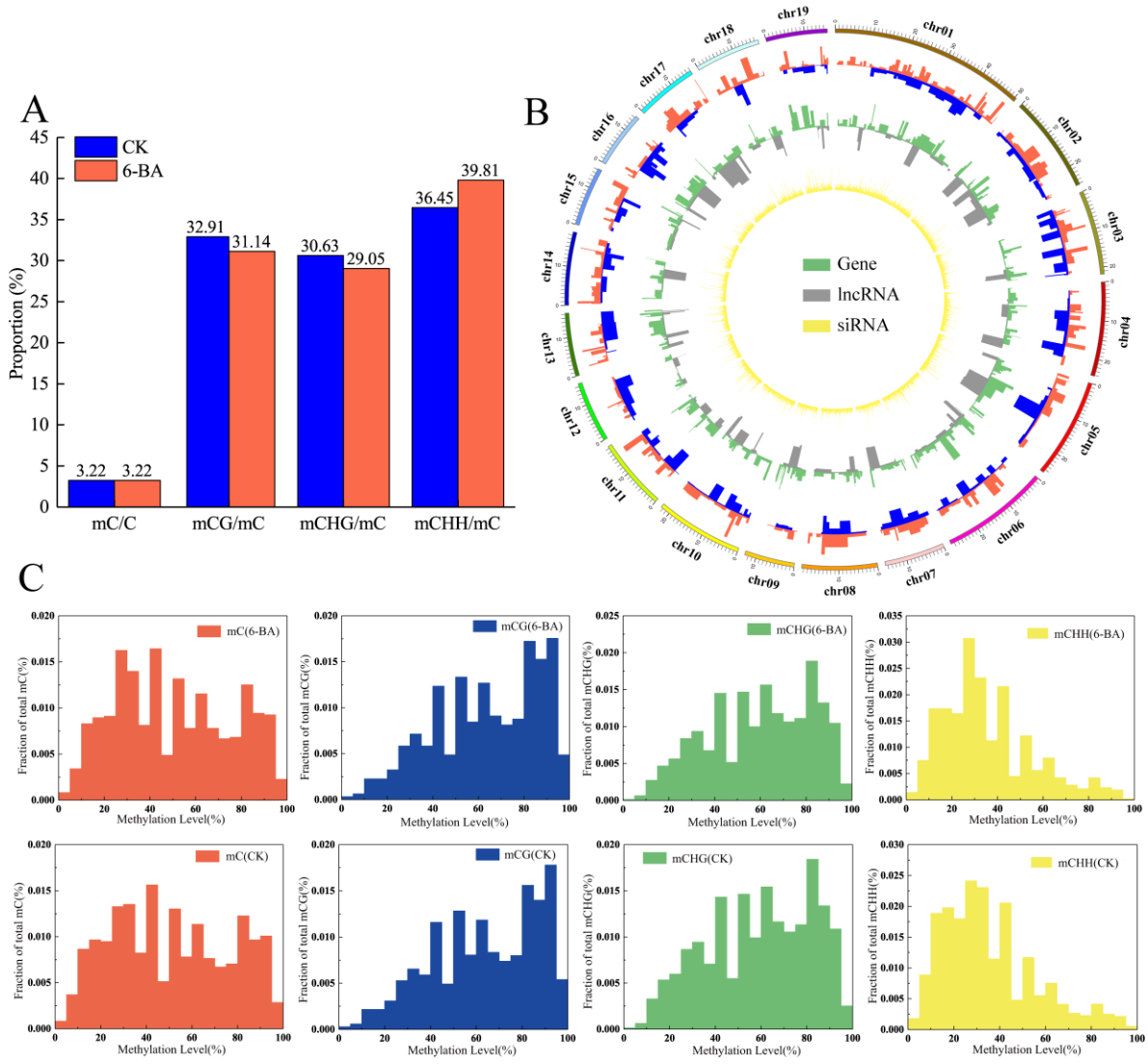


Figure 5 Patterns of variation in DNA methylation in response to 6-BA treatment in *P. tomentosa*. **A** Classification of methylated cytosines in the control (CK) and 6-BA-treated groups. **B** The outermost red ring represents the differential hypermethylation domains under 6-BA treatment; the blue ring represents the differential hypomethylation domains under 6-BA treatment; the green ring represents differentially expressed genes under 6-BA treatment; the gray ring represents differentially expressed lncRNAs under 6-BA treatment; the innermost yellow ring represents differentially expressed 24-nt siRNAs under 6-BA treatment. **C** Distribution of methylation levels in different contexts in the control (CK) and 6-BA-treated groups.

2.7. ASE analysis of transcription elements located in the 6-BA-responsive DMRs

To further explore the effects of DNA methylation on transcriptional regulation and ASE in poplar, we performed expression analysis and ASE analysis of transcription elements located in DMRs, which could potentially be regulated by DNA methylation, including 210 protein-coding genes, 22 lncRNA genes, and 172 siRNA cluster genes. Of the eight differentially expressed

protein-coding genes in response to 6-BA treatment, four were upregulated and four were downregulated (P<0.01). Among the lncRNAs, eight were expressed stably under 6-BA vs. control treatment, with Fold change values between 0.23 and 3.22. In addition, 67 24-nt siRNAs clusters were differentially expressed under 6-BA treatment, all of which were significantly downregulated ($|\log_2(\text{fold change})| \geq 1$, P<0.05, Table S2, S5 and S8).

To further explore the ASE patterns of these 6-BA-responsive genes and lncRNAs, we selected two 6-BA-responsive protein-coding genes and two lncRNA genes (one with maximum and the other with minimum Fold change values) to calculate the ASE levels of SNPs located in the genes. As shown in Figure 6, the ASE levels of both protein-coding genes (*Potri.011G140400* and *Potri.006G239700*) decreased in response to 6-BA treatment, which was accompanied by the downregulated expression of *Potri.011G140400* and the upregulated expression of *Potri.006G239700* (Figure 6A, B). The ASE levels of lncRNA *TCONS_00082911* and lncRNA *TCONS_00240091* simultaneously increased, with the former showing downregulated expression and the latter showing upregulated expression (Figure 6C, D). These results indicate that the differences in ASE of protein-coding genes decrease under 6-BA treatment. By contrast, there were greater differences in ASE in lncRNAs genes vs. protein-coding genes under 6-BA treatment compared to the control.

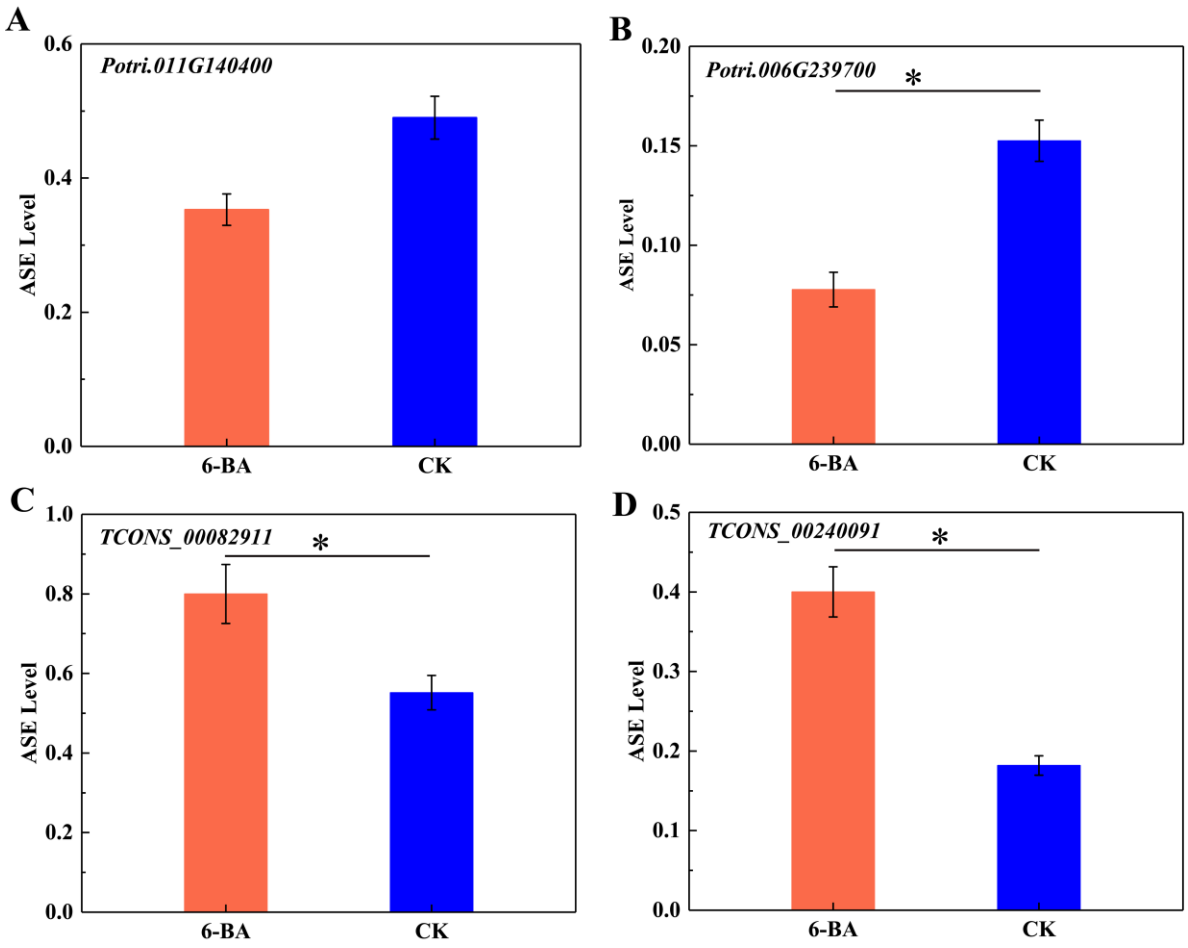


Figure 6 Analysis of allele-specific expression in a single gene in response to 6-BA treatment. A Variation of ASE levels in *Potri. 001G140400*. **B** Variation of ASE levels in *Potri.006G239700.* **C** Variation of ASE levels in lncRNA (*TCONS_0082911*). **D** Variation of ASE levels in lncRNA *TCONS_00240091*. Error bars represent standard deviation (SD) of three biological replicates (n=3). Asterisks indicate significant differences between 6-BA-treated (red) and control (blue) groups (*P < 0.05, **P < 0.01).

2.8. Variation in 6-BA-responsive DMRs is associated with phenotypic variation

We conducted a series of analyses to investigate the functional roles of the 6-BA-responsive elements. GO enrichment analysis showed that the 6-BA-responsive DNA methylated genes were enriched for the categories 'regulation of cellular metabolic process', 'cellular biosynthetic process' and 'cellular macromolecule metabolic process'. This enrichment suggests that 6-BA responsive genes are likely involved in phenotypic variation. To test this hypothesis, we conducted association analysis to investigate the potential functional roles of 6-BA-responsive DMRs in phenotypic variation. We selected 507 6-BA-responsive DMRs for SNP analysis by re-sequencing a natural population of *P. tomentosa*. After discarding SNPs with minor allele frequencies <5% and SNPs with >25% missing data, we identified 3,736 common SNPs. We used the SNPs within the DMR boundaries to perform association analysis relative to 10 growth and wood property traits, including height (H), diameter at breast height (DBH), stem volume (V), holocellulose (HC), hemicellulose (HEMC), α-cellulose (AC), lignin contents (LC), fiber length (FL), width (FW), and microfibril angle (MFA) (Table S14, S15). Using MLM in TASSEL5.0, we detected 182 significant associations between 152 SNPs from 507 DMRs and seven growth and wood property traits (DBH, V, HEC, HC, AC, LC, FW) (P<0.001, Q<0.1), with 12.39–27.12% of the phenotype variance (R^2) explained by each SNP (Table 1, Table S16).

Table 1. SNPs of DMRs associated with *P. tomentosa* growth and wood property traits.

Trait	Number of SNPs
Stem volume (V, m3)	46
Diameter at breast (D, cm)	107
a-cellulose content (Ac, %)	2
Hemicellulose content (HEMC, %)	4
Holocellulose content (HC, %)	11
Lignin content (LC, %)	1
Fiber width (FW, μm)	11
Total	182

Since dominant effects results from the interaction between alleles within a gene locus, the value of the dominant effect fully reflects the interaction between alleles. We subjected 102 ASE loci in the 6-BA-responsive DMRs to further analysis. Association analysis showed that 52.94% of the ASE loci prominently exhibited dominant effects on poplar traits (d>a). We identified 24 loci associated with multiple traits (DBH, V, HEC, HC), where dominant effects had a significant advantage over additive effects (d/a>2, Table S17). Interestingly, among these loci, four were jointly annotated to the lncRNA *TCONS_00053467* and were correlated with different traits (DBH, V); the ASE level changed significantly, accompanied by dominant effects (P<0.05, Figure 7C, D). Computational prediction identified six potential cis-regulated target genes and six potential trans-regulated target genes for *TCONS_00053467*. Among these target genes, the cis-target gene for *Potri.002G258000* and *TCONS_00053467* partially overlapped in the genome: *Potri.002G258000* was significantly downregulated and *TCONS_00053467* was significantly upregulated under 6-BA treatment (Figure 7B). These genes are located on sense and antisense strands, respectively. *Potri.002G258000* is homologous to *MIK2/LRR-KISS*, a *leucine-rich repeat receptor kinase* (*LRR-RK*) gene in *Arabidopsis*[21,22]. As shown in Figure 7A, multiple interacting elements exist in this candidate region, in which hypermethylated *DMR_Chr02_24663250* partially overlaps with *siRNA_cluster_32657* in the upstream region. The expression of *siRNA_cluster_32657* was significantly downregulated under 6-BA treatment (Figure 7B).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 February 2020 doi:10.20944/preprints202002.0038.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 2117; doi:10.3390/ijms21062117

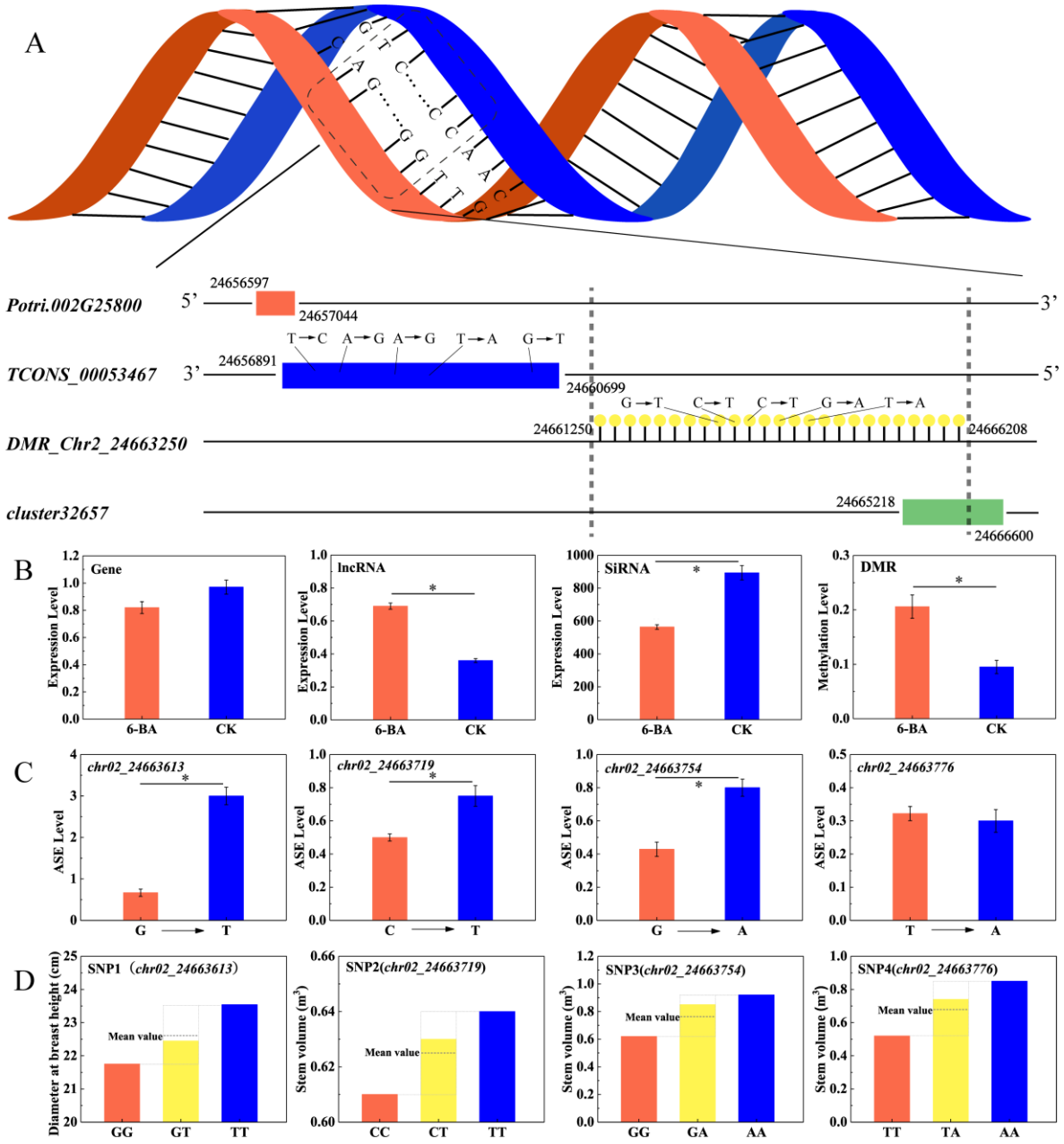


Figure 7 Analysis of patterns of genetic variation in candidate DMR regions and interactions between genomic elements. A The gene and lncRNA are located on complementary DNA strands, and a DMR partially overlapping with a siRNA cluster occurs in the upstream region of lncRNA. **B** Variation of expression levels of *Potri.002G258000*, *TCONS_00053467*, *siRNA_cluster_32657* and variation of DNA methylation level of *DMR_Chr02_24663250*. **C** Variation of ASE levels in SNPs located in lncRNA (TCONS_0053467). **D** Differences between the phenotypic values of the heterozygous genotype and the mean value of the homozygous allele indicate dominant effects. Error bars represent standard deviation (SD) of three biological replicates (n=3). Asterisks indicate significant differences between 6-BA-treated and control groups (*$P < 0.05$, **$P < 0.01$).

3. Discussion

3.1. The responses of physiological characteristics and photosynthetic indices to 6-BA treatment in P. tomentosa

Pn, Gs, and Tr significantly decreased in *P. tomentosa* under long-term 6-BA treatment, with Ci increasing slightly. This finding suggests that WUE increases under long-term 6-BA treatment.

However, this pattern is opposite that reported for cotton, where Pn and Tr significantly increased in response to 6-BA, while WUE remained stable[23]. Ding et al.[24] indicated that under normal light conditions, the foliar application of 6-BA had no effect on any photosynthetic parameter in cucumber plants. Thus, it appears that different species exhibit different photosynthetic patterns in response to 6-BA treatment. In our study, Pn decreased, while Ci increased slightly, which may be caused by non-stomatal restriction factors related with the limitation of photochemical activity, which hindered the utilization of CO2, resulting in the accumulation of Ci and the decrease of Pn. In addition, these changes may be caused by the decrease of Gs, resulting from stress effects of 6-BA treatment. On the other hand, 6-BA short-term treatment caused the slight increase of Tr, while Pn, Gs and Tr almost unchanged, which demonstrated that photosynthesis is insensitive to short-term 6-BA treatment and lagging behind physiological characteristics.

We previously compared the changes in physiological characteristics in poplar under GA treatment. The changes in POD and MDA under GA treatment were consistent with those observed under 6-BA treatment; however, the changes in total protein content and sucrose phosphate synthase activity were opposite those under 6-BA treatment. Interestingly, the most significant response to GA treatment was observed after 6 h[15]. Therefore, we reasoned that the physiological responses of *P. tomentosa* to different growth regulators are completely different. However, the responses were similar after 6 h, implying that responses to phytohormones occur rapidly in poplar, although the specific regulatory mechanisms remain unclear. In addition, Zhang et al.[25] found that 6-BA application increased peroxidase (POD) activities, and decreased the MDA contents in cottons, which are consistent with the changes in poplar. And it was found that 6-BA alleviated chilling injury in cucumber fruit through improving antioxidant enzyme activities and total antioxidant capacity and maintaining higher levels of ATP content and energy charge[26]. These changes under 6-BA treatment may be caused by that exogenous 6-BA affected the metabolism of related enzymes in short-term, so the physiological changes increased. While exogenous 6-BA, as stress factors, may trigger the endogenous homeostasis in plants to regulate the physiological indexes returning to normal level. Among these physiological characteristics, POD, a protective enzyme, removes hydrogen peroxide from cells to reduce oxidative damage, thus delaying leaf senescence[27]. POD activity in leaves increased significantly under 6-BA treatment, confirming the effect of cytokinin on delaying leaf senescence. Lipid peroxidation is an inherent feature of leaf senescence[28]. MDA activity in poplar leaves decreased significantly under 6-BA treatment, indicating that lipid membrane peroxidation was enhanced, leading to delayed leaf senescence. Therefore, 6-BA may delay leaf senescence by increasing the activity of photosynthetic and protective enzymes, thereby reducing secondary metabolic activity.

Most previous studies have simply explored the physiological effects of exogenous cytokinin on the growth of plants, such as stem elongation and rooting[26,29]. By contrast, in the current study, we explored the effects of cytokinin on poplar at the transcriptional, epigenetic, and allelic levels from a molecular perspective through sequencing technology. Exogenous cytokinin produced genome-wide changes, including changes in photosynthetic indicators, enzyme activity, and epigenetic modifications, which were not observed in previous studies, forming the basis for changes in poplar growth.

3.2. 6-BA responsive ASE analysis in poplar

Here, we performed the first systematic survey of ASE in poplar under 6-BA treatment. Since the identification of ASE loci is performed on the whole-genome scale, we faced several difficulties, such as the wide range of locus distribution, large amounts of data, and complex loci variation. It is particularly important to choose good identification criteria for ASE loci. Previous studies have used the NS-12 BeadChips assay to measure the average fluorescent signals of single alleles [A1/(A1+A2) or A2/(A1+A2)] in DNA and RNA, respectively. To obtain a quantitative measure of ASE, the authors subtracted the allele fractions [A1/(A1+A2)] measured in DNA from that in RNA and referred to this difference as the ASE level[30]. The advantage of this approach is that it eliminates

the effect of genetic variation on ASE; however, this approach is not suitable for ASE analysis of sequencing data.

In the present study, we used bioinformatics to compare RNA-seq data to identify ASE loci and selected stable heterozygous loci for further analysis. To minimize bias in SNP calling, we used a series of measures to increase the accuracy of genome mapping. First, we used software for RNA-specific comparisons in genome mapping, which uses an improved BWT algorithm that allows introns to be identified for cross-intron mapping to the genome. We then used the Picard tool to identify duplicates generated by PCR amplification. We also used the SplitNCigarReads tool in GATK to split the reads into exon segments by removing Ns and hard-clipping any sequences overhanging into intronic regions. Importantly, this analysis relied on the *P. trichocarpa* genome (version 3.0), which to some extent led to mapping bias of RNA-seq data in *P. tomentosa*. Therefore, we selected loci with high numbers of reads to reduce sequencing and alignment errors caused by the use of the *P. trichocarpa* genome (version 3.0). Actually, the reference genome served as an anchor template for indirect comparisons between the 6-BA treatment and control groups. The results should be accurate as long as the locations of loci between the 6-BA treatment group and control group are consistent.

We then performed ASE analysis based on the large-scale allelic variation loci identified by the SNP calling pipeline of RNA-seq. We selected loci from the 3′ UTR with strong specificity in a single gene for genotyping to score the ASE level of that gene. The advantage of this approach is that the non-conservative regions of the 3′ UTR are selected as candidate regions, which are sensitive to external environmental changes and growth regulators. However, due to the complexity of gene transcriptional regulatory elements, it is not clear whether this method can truly represent ASE patterns of genes; this notion requires further exploration. As we demonstrated based on 19,200 genes with identifiable functional variations, allele-specific gene expression is widespread under 6-BA treatment. Such strong, consistent imbalanced expression likely occurs because one of the alleles is favorable while the other is unfavorable under 6-BA treatment. Therefore, ASE is a crucial factor that helps plants to adapt to different environmental signals and epigenetic regulation during growth and development.

3.2. Effects of DNA methylation on the transcriptional regulation of alleles in P. tomentosa

The functional consequences of promoter methylation on transcriptional regulation are well known[31], DNA methylation in the promoter region inhibits gene expression. However, the influence of DNA methylation of specific alleles on gene expression remains largely unresolved, especially in plants. It was found that parental allele-specific expression (gene imprinting) is closely associated with DNA demethylation and siRNA-dependent RdDM pathway, which regulated seed development in maize[32]. And Kinoshita et al.[33] found that maintenance of endosperm-specific and parent of origin–specific FWA expression depends on maintenance DNA methyltransferase *MET1* in *Arabidopsis*, which indicated that the maintenance of FWA imprinting depends on the maintenance DNA methylation machinery. Therefore, DNA methylation plays an important role in the allele-specific expression of plants.

In the current study, we used the methylation support rate to genotype methylation loci and analyzed the changes in methylation at the allele level. The quantitative data for ASE and for the methylation levels of CpG loci allowed us to uncover a direct quantitative correlation between ASE and CpG locus methylation in response to 6-BA treatment. The correlation coefficient between ASE and the high methylation variation group was higher than that between ASE and the low methylation variation group, providing further evidence that DNA methylation affects the transcriptional regulation of alleles. Our identification of a large set of loci that exhibit ASE in poplar and a subset of loci with differential changes in methylation opens up new perspectives for more detailed studies on the molecular events that occur in response to 6-BA treatment.

3.3. The relationship between dominant-effect expression and ASE analyzed by association analysis

We conducted association analysis of the DMRs to analyze the genetic effects of the regions potentially regulated by DNA methylation. Among these, the dominant effect might be caused by allele-specific or imbalanced expression (the SNPs for two allelic sequences deviated significantly from a 1:1 ratio). The detection of ASE of genes under 6-BA treatment strengthens this hypothesis. Perhaps favorable alleles are expressed at higher levels in response to the growth regulator 6-BA, resulting in dominance. This finding suggests that two alleles can adapt to growth regulators in different ways.

We identified a key region of a DNA sequence (*DMR_Chr02_24663250*) with complex transcriptional regulation involving multiple transcriptional elements, where differential hypermethylation occurred under 6-BA treatment. *Potri.002G258000*, a homolog of *MIK2*, was identified by genetic analysis. This gene was targeted by *TCONS_00053467*, which contains four dominant SNPs associated with tree diameter at breast height (DBH) and stem volume (V). MIK2 is an important regulator of responses to cell wall damage triggered by the inhibition of cellulose biosynthesis[34], supporting the notion that 6-BA treatment promotes cell division through lncRNA-mediating transcriptional regulation.

In addition, we detected a siRNA cluster near this DMR. We further explored the expression levels of the above-mentioned transcriptional elements. The expression of MIK2 and the nearby siRNA cluster was negatively correlated with *TCONS_00053467* expression. SiRNA is closely associated with de novo methylation[35]. Our results indicate that *siRNA_cluster_32657*, which is located in *DMR_Chr02_24663250*, maintains the hemi-methylated state of this DNA region. Under normal growth conditions, hemi-methylated *DMR_Chr02_24663250* might regulate the ASE of adjacent *TCONS_00053467* expression. Transcription of the lncRNA *SRG1* driven by the promoter of the adjacent *SER3* gene repressed *SER3* expression by interfering with the binding of *Pol II* to its *cis*-element[28]. In the current study, the expression levels of *TCONS_00053467* and its *cis*-target gene were negatively correlated. *TCONS_00053467* partially overlaps (153 bp) with the 3′ terminus of *MIK2* and is located on the antisense strand. The concurrent expression of a pair of genes on complementary strands might induce the biosynthesis of double-stranded RNA (dsRNA), which would trigger RNA interference, resulting in adjacent gene silencing. However, no endogenous siRNAs were detected in this region. Thus, the upregulated *TCONS_00053467* might also repress *MIK2* expression via their shared flanking regions. However, this regulatory mechanism requires further study.

Collectively, the above results suggest that the hemi-methylation of *DMR_Chr02_24663250* might play important roles in the dominant effect of *TCONS_00053467* on D and V through mediating ASE of the *TCONS_00053467-MIK2* module. Under 6-BA treatment, *siRNA_cluster_32657* expression was significantly downregulated, along with hypermethylation of *DMR_Chr02_24663250*, pointing to possible negative feedback regulation between methylation level and the expression of adjacent siRNA clusters. Our study provides a reference for investigating allelic variation and epigenetic variation of various elements in the genome under 6-BA treatment at the molecular level, and it provides a new perspective on the effects of 6-BA on woody plants for future functional studies.

4. Materials and Methods

4.1. Plant Materials and 6-BA treatment

The experimental material consisted of 30 (2 treatments × 5 time points × 3 biological replicates) one-year-old *P. tomentosa* ramets (clone number ‘1316’), which were grown in a greenhouse at Beijing Forestry University, Beijing, China (40°0'N, 116°20'E). The ramets were grown in containers in a mixture of soil, organic matter, vermiculite, and perlite (1:1:1:1) under a 15h light/8h dark photoperiod and randomized in two separate groups: treatment (6-BA treated) and control. We conducted a preliminary experiment to determine the appropriate concentration of 6-BA by spraying different concentrations of 6-BA on poplar leaves. The concentration gradient of 6-BA was 0.1μmol/L, 1μmol/L, 10μmol/L, 100μmol/L, 1000μmol/L. The experiment used the test of least significant difference (LSD) multiple comparisons to compared different 6-BA concentration's

influence on physiological characteristic. The changes of total protein, SPS activity and MDA content appeared inflection point under 100μmol/L 6-BA treatment (Figure S4A-D). Thus 100μmol/L was selected as the concentration of 6-BA treatment. Leaves of the treatment groups were sprayed with 100μmol/L 6-BA (Sigma-Aldrich, St. Louis, MO, USA) until soaking (with drops of liquid dripping down), while the control group received the same treatment, but with 6-BA replaced by water. For both groups, mature leaves were harvested from the same position of each ramet at 0, 3, 6, 12, and 24 h after treatment, immediately frozen in liquid nitrogen, and stored at -80°C until further use. Three different clones were set up at different time points in both the treatment and control groups as biological replicates to eliminate the effect of deviation in leaf position.

4.2. Measurement of physiological, growth, and wood properties and photosynthetic indices

Various parameters were measured in plants from both treatment groups (6-BA treated and control) at five treatment times: i) four physiological traits. Total protein content was measured by the Bradford method using BSA as the standard. Sucrose phosphate synthase (SPS) activity was detected using a Sucrose Phosphoric Acid Synthase Assay Kit (Nanjing Jiancheng Bioengineering Institute, Jiangsu Province, China). Peroxidase (POD) activity was measured using a Plant POD Assay Kit (Nanjing Jiancheng Bioengineering Institute). And malondialdehyde (MDA) content was analyzed according to the method of (Heath and Packer., 1968). ii) three growth traits, including height (H), diameter at breast height (DBH), and stem volume (V), which following standard procedures described in Zhang et al. (2006). iii) seven wood property traits, including holocellulose (HC), hemicellulose (HEMC), α-cellulose (AC), lignin contents (LC), fiber length (FL), width (FW), and microfibril angle (MFA). LC was derived by wet chemistry analyses as described in Porth et al. (2013). The detailed methods for sampling and measuring the three growth traits (DBH, H, and V) and seven wood property traits are described in Du et al.[36]. And iv) four photosynthetic indices, including net photosynthetic rate (Pn), stomatal conductance (Gs), transpiration rate (Tr), and intercellular CO2 concentration (Ci); these parameters were measured from fully expanded leaves (three functional leaves, i.e., the top fourth to sixth leaves) using the LI-6400 Portable Photosynthesis System (LI-COR Inc., Lincoln, NE, USA) following the manufacturer's instructions at 9:00–11:00 AM on sunny days. Each leaf measurement was performed using three replications. WUE is calculated by formula: WUE=Pn/Tr.

4.3. DNA extraction and bisulfite sequencing

At 6 h after treatment, leaves of the 6-BA-treated and control groups were collected for bisulfite sequencing using three biological replicates for each treatment time. Total genomic DNA was extracted from young leaves using a DNase Plant Mini Kit (Qiagen China, Shanghai) following the manufacturer's protocol. The DNA was sonicated to generate fragments with a mean size of approximately 250 bp, followed by DNA repair of blunt ends (3′ ends) by the addition of dA and adaptor ligation. Bisulfite modification and conversion of genomic DNA were then performed using an EpiTect Bisulfite Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The resulting DNA from 6-BA-treated and control plants for all treatment time points was subjected to paired-end sequencing with a read length of 101 nt for each end using the ultrahigh-throughput Illumina Hiseq2000 platform as per the manufacturer's instructions.

4.4. RNA extraction and RNA-sequencing

At 6 h of treatment, leaves of 6-BA-treated and control plants were subjected to RNA sequencing using three biological replicates per treatment time. Total RNA was extracted from the samples using a Qiagen RNase Kit (Qiagen China, Shanghai, China) according to the manufacturer's instructions, followed by on-column DNase digestion to purify the RNA samples using an RNase-Free DNase Set (Qiagen). The RNAs were assessed with a NanoDrop ND-1000 and Agilent Bioanalyzer 2100 before being used to construct strand-specific RNA-seq libraries as described in the TruSeq RNA sample preparation guide. The strand-specific libraries were sequenced using an

Peer-reviewed version available at Int. J. Mol. Sci. 2020, 21, 2117; doi:10.3390/ijms21062117

Illumina HiSeq 2500 instrument in which 100-nucleotide paired-end reads were produced. Library construction and sequencing were performed by Shanghai Biotechnology Corporation (Shanghai, China).

4.5. Allele-specific expression analysis

The RNA-sequencing reads were mapped to the *P. trichocarpa* genome (version3.0) (http://www.phyozome.net/popalr.php) with TopHat (version: 2.0.9) using the spliced mapping algorithm (multi hits ≤ 1). The mapping results were converted to BAM format, and non-unique and unmapped reads were filtered with SAMtools. The AddOrReplaceReadGroups and MarkDuplicates tools in the Picard package (http://broadinstitute.github.io/picard/, version: 1.87) were used for acquiring read group information, sorting, and marking duplicates. The HaplotypeCaller tool in Genome Analysis Toolkit (GATK, version: 3.8) was used for variation calling in the 6-BA and control groups. The minimum phred-scaled confidence threshold for calling variants was set at 20. CombineGVCFs tool in GATK was used to integrate variations in all samples. All other parameters were set at default values. To filter the resulting callset, the VariantFiltration tool in GATK was used to filter clusters of at least 3 SNPs within a window of 35 bases and low-quality SNPs. Annotation of variation between the samples from the 6-BA-treated and control groups was based on gene model set v3.0 from Phytozome v11.0 (Figure S5). Loci within 500 bp of the 3′ untranslated region (UTR) sequences of 6-BA-responsive genes with maximum support reads were selected as candidate loci to represent the ASE levels of genes. The ASE level was calculated based on the allele fraction [ALT/(ALT + REF)], Alt refers to the number of reads that contain the alterative allele and REF refers to the number of reads that contain the reference allele.

4.6. Predicting lncRNAs and identifying 6-BA-responsive lncRNAs

FASTX-Toolkit version 0.0.13 was used for quality control of RNA-seq reads, including the removal of low-quality reads and adapter sequences shorter than 20 nucleotides. The clean reads were aligned to *Populus* rRNA sequences using the Short Oligonucleotide Analysis Package (SOAP2; http://soap.genomics.org.cn/soapaligner.html) to eliminate rRNAs. The clean data were aligned to the *P. trichocarpa* genome (version 3.0) with three base mismatches allowed. Cufflinks v2.1.1[37] was used to assemble transcripts based on the *P. trichocarp*a genome (version 3.0). Expression levels were calculated and normalized using fragments per kilobase of transcript per million fragments (FPKM). The prediction of lncRNAs from RNA-seq data was performed following the method of Sun et al. [38].

4.7. Predicting target genes of 6-BA-responsive lncRNAs

Potential target genes of the 6-BA-responsive lncRNAs were classified into two groups, *cis*-target genes and *trans*-target genes, according to their regulatory effects. Two different algorithms were used to predict the two types of target genes. The potential *cis* target genes, which are physically close to lncRNAs (within 10 kb), were predicted using genome annotation and a genome browser using the criteria described in Jia et al.[39]. Potential *trans*-targets were predicted by searching the *Populus* mRNA database based on mRNA sequence complementarity and RNA duplex energy prediction, assessing the impact of lncRNA binding on complete mRNA molecules. BLAST was used to select target sequences that were complementary to the lncRNA, setting *E*-value <1e-5 and identity ≥95%. RNAplex software was used to calculate the complementary energy between two sequences for further screening and to select potential *trans*-acting target genes (RNAplex -e-60)[40].

4.8. Identification of 6-BA-responsive 24-nt siRNAs

After filtering low-quality reads and adapters with FASTX (fastx_toolkit-0.0.13.2), clean siRNA sequence data were consolidated into an siRNA data set. The data set was further collapsed into FASTA format using a Perl script. Pln24NT (Pln24NT_analysis_local_version_v1.0)

(http://bioinformatics.caf.ac.cn/Pln24NT/) was used to carry out siRNA analysis: The 24-nt siRNA read sequences were retrieved and aligned to miRNAs, and RFAM noncoding RNAs were removed with Perl scripts[41]. Non-redundant 24-nt siRNA sequences were mapped to the *P. trichocarpa* genome (version 3.0) for each dataset using Bowtie[42]: a maximum of one mismatch (-v 1) was allowed, and the best alignments of reads with no more than 50 hits (-a -m 50 --best --strata) were reported. SiRNA cluster calling was processed by ShortStack v3.3 with a minimum coverage of ten 24-nt siRNA reads. Genomic regions associated with siRNA clusters are referred to as siRNA loci hereafter. If present, 24-nt siRNA clusters were merged in a 150-bp window[43] to generate the final set of 24-nt siRNA loci. Transposon (TE) overlap analysis of clusters was conducted by comparing the start and end positions with TEs annotated by RepeatMasker (http://www.repeatmasker.org/) using the intersectBed and subtractBed tools in Bedtools[44]. Mapping 24-nt siRNA reads were selected for expression level analysis by calculating reads per million (RPM) for each locus. The normalized number of reads mapped to the genome of all 24-nt siRNAs from each 24-nt siRNA cluster (RPM, reads per million) represents the expression level of the cluster.

4.9. Genome-wide identification of 6-BA-responsive methylated cytosine loci and DMRs

Quality control reads of the reads was performed using FASTX-Toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) with default parameters. Specifically, high-quality sequencing reads (Q >20) from the 6-BA-treated and control groups were selected and trimmed to 101 nt and mapped to the *P. trichocarpa* genome (version 3.0) (http://www.phyozome.net/popalr.php). To detect methylation sites, the Bismark alignment tool was used to detect cytosines that had been converted after bisulfite treatment and to conduct methylation site calling with default parameters. Fisher's Exact Test was used to test cytosine sites of the 6-BA-treated and control groups with a minimum coverage of ≥ 5.

The binomial distribution of the number of methylated cytosine and non-methylated cytosine frequencies at each site was tested to determine whether the site was a genuine methylated site. The methylation level (ML) of each identified methylated cytosine was calculated using the following formula: ML = mC / (mC + umC), where mC and umC represent the number of methylated cytosine and unmethylated sites, respectively. Due to the influence of bisulfite conversion rate, the methylation level was corrected using the following formula: ML_corrected = (ML-r)/ (1-r), where r represents the bisulfite conversion rate. swDMR software was used to identify differentially methylated regions (DMRs). Based on the methylation information for each site (read coverage ≥ 5), this software uses a sliding window method to scan the genome and identify DMRs. The sliding window size was set to 1,000 bp with 100 bp as a step size. Windows with probabilities < 0.05 were merged into larger regions to estimate the mean and variance of the entire methylation regions. ANOVA was used to filter DMRs with $P < 0.05$ (called candidate DMRs). To analyze the distribution of DMRs in various elements of genes, the DMRs were mapped to the *P. trichocarpa* genome (version 3.0) and annotated using Phytozome poplar genome v3.0 (http://www.phytozome.net/poplar.php). Chi-squared homogeneity tests were used to evaluate the effects on mapping distribution.

4.10. Genotyping of CpG loci

A quantitative measure of the methylation level for each analyzed CpG site was assigned, with methylation support rate (S) ranging from 0 to 1.0, corresponding to no methylation on either allele to complete methylation of both alleles (S=number of mC reads / (number of mC reads + number of C reads in single locus).

4.11. Genome re-sequencing and SNP calling

The SNP data used in this study were obtained by re-sequencing an association population composed of 435 unrelated *P. tomentosa* individuals. Total genomic DNA was extracted from individual leaves using a DNeasy Plant Mini Kit (Qiagen, Shanghai, China) according to the manufacturer's protocol, followed by re-sequencing of the raw data at a depth of >15× using the

Illumina GA II platform after the construction of library. Paired-end short reads (150 bp) were generated and filtered to obtain clean data by removing low-quality reads (≤ 50% of nucleotides with a quality score < Q20). The filtered reads were aligned and mapped to the *Populus* reference genome using BWA with default settings[45], with an 81 to 92% mapping rate and ~11× effective mapping depth for most individuals. Uniquely mapped paired-end reads was performed data quality control using fastqc software. BWA software was used for genome mapping of high-quality reads. After discarding the duplicate reads, SNP calling was performed using the HaplotypeCaller tool in GATK3.8[3]. The genotype data for the DMRs were obtained using location information. SNP markers with minor allele frequencies <5% were excluded. SNPs with >25% missing data, as well as SNPs with limited numbers of individuals represented in the alternative genotype class or classes, were also discarded from the association analysis.

4.12. Single SNP-based association analysis

Single SNP–trait association analysis was performed using the mixed linear model (MLM) in the software package TASSEL5.0 (https://www.maizegenetics.net/tassel) considering the effects of population structure (*Q*) and pairwise kinship coefficient (*K*). The pair-wise kinship coefficient (*K*) was evaluated with SPAGeDi 1.3[46], and the *Q* matrix was calculated with STRUCTURE 2.3.4 based on significant subpopulations (k = 3)[47]. The parameters were used to correct the relationships of associated individuals to minimize false positives obtained by association analysis. The genotypic effects of associated SNPs were effectively decomposed into additive and dominant effects under this model. The positive false discovery rate (FDR) was calculated to correct for errors related to multiple testing using QVALUE software[48]; a q-value of 0.10 was selected as the significance threshold. SNP markers with a value of dominance to additive effects ratio (d/a)> 2 were defined as prominent dominant-effect loci.

4.13. Statistical analysis

One-way ANOVA was performed using SPSS software (SPSS19, IBM Corporation, New York, USA), and significant differences between the 6-BA-treated and control groups were determined using Fisher's Least Significant Difference (LSD) test. Differences were considered statistically significant when P<0.05. The correlation between the ratios of variation in DNA methylation and ASE was determined at a significance level of P <0.01.

4.14. Data availability

The RNA sequencing raw data and bisulfite sequencing raw data have been submitted to Genome Swquence Archive in BIG Data Center (BIG, CAS, China) under accession number of CRA002030. The raw data of genome re-sequencing had been deposited in the Genome Sequence Archive in BIG Data Center (BIG, CAS, China) under accession number of CRA000903.

Supplementary Files:

Figure S1. Changes of WUE under long-term 6-BA treatment in *P. tomentosa*.

Figure S2. Statistics of transposons located in different genomic elements.

Figure S3. DNA methylation levels in different genomic features under control and 6-BA treatment.

Figure S4. Effects of different concentrations of 6-benzylaminopurine treatment on physiological and photosynthetic characteristics of *P. tomentosa*.

Figure S5. Pipeline for identifying allele-specific expression loci.

Table S1: Summary of *P. tomentosa* RNA-seq data.

Table S2: 6-Benzylaminopurine-responsive genes.

Table S3: Significant GO categories of 6-BA-responsive genes.

Table S4: Characteristics of all lncRNAs identified in this study.

Table S5: 6-Benzylaminopurine-responsive lncRNAs.

Table S6: Summary of *P. tomentosa* small RNA-seq data.

Table S7: Summary of *P. tomentosa* 24-nt siRNA data.

Table S8: 6-Benzylaminopurine-responsive 24-nt siRNA clusters.

Table S9: Summary of allelic variation of 6-benzylaminopurine-responsive loci.

Table S10: Summary of *P. tomentosa* bisulfite sequencing data.

Table S11: CpG loci with methylation variation and ASE variation.

Table S12: 6-Benzylaminopurine-responsive differentially methylated regions.

Table S13: Annotation of genomic elements located in DMRs.

Table S14: Phenotypic data used for association analysis.

Table S15: SNPs within DMRs used for association analysis in *P. tomentosa*.

Table S16: SNPs of DMRs associated with growth and wood property traits in a *P. tomentosa* natural population.

Table S17: SNPs of DMRs associated with growth and wood property traits with prominent dominant effects.

Author Contributions: D.Z. conceived and designed the experiments. A.X. carried out the gene expression analysis and drafted the manuscript. D.Z., Y.S. and El-Kassaby YA supervised and assisted with the writing. C.B. and P.C. performed the DNA and RNA extractions. All authors read and approved the final manuscript.

Acknowledgments: This work was supported by the Fundamental Research Funds for the Central Universities (No. 2017ZY04), the Project of the National Natural Science Foundation of China (Nos. 31670333 and 31872671), and the 111 Project (No. B20050). We are grateful for the sequence information produced by the U.S. Department of Energy Joint Genome Institute (http://www.jgi.doe.gov).

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Miller, C.O.; Skoog, F.; Okumura, F.S.; Von Saltza, M.H.; Strong, F.M. Isolation, Structure and Synthesis of Kinetin, a Substance Promoting Cell Division. *Journal of the American Chemical Society* **1956**, *78*, 1375-1380, doi:10.1021/ja01588a032.

2. Aloni, R.; Langhans, M.; Aloni, E.; Dreieicher, E.; Ullrich, C.I. Root synthesized cytokinin in Arabidopsis is distributed in the shoot by the transpiration stream. *J Exp Bot* **2005**, *56*, 1535-1544, doi:10.1093/jxb/eri148
3. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M., et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **2011**, *43*, 491-498, doi:10.1038/ng.806.
4. Durbak, A.; Yao, H.; McSteen, P. Hormone signaling in plant development. *Current Opinion in Plant Biology* **2012**, *15*, 92-96, doi:10.1016/j.pbi.2011.12.004.
5. Matsumoto-Kitano, M.; Kusumoto, T.; Tarkowski, P.; Kinoshita-Tsujimura, K.; Václavíková, K.; Miyawaki, K.; Kakimoto, T. Cytokinins are central regulators of cambial activity. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20027-20031, doi:10.1073/pnas.0805619105.
6. Zwack, P.J.; Robinson, B.R.; Risley, M.G.; Rashotte, A.M. Cytokinin Response Factor 6 negatively regulates leaf senescence and is induced in response to cytokinin and numerous abiotic stresses. *Plant Cell Physiol* **2013**, *54*, 971-981, doi:10.1093/pcp/pct049.
7. Argueso, C.T.; Ferreira, F.J.; Epple, P.; To, J.P.C.; Hutchison, C.E.; Schaller, G.E.; Dangl, J.L.; Kieber, J.J. Two-Component Elements Mediate Interactions between Cytokinin and Salicylic Acid in Plant Immunity. *Plos Genet* **2012**, *8*, doi:ARTN e100244810.1371/journal.pgen.1002448.
8. Seguela, M.; Briat, J.F.; Vert, G.; Curie, C. Cytokinins negatively regulate the root iron uptake machinery in Arabidopsis through a growth-dependent pathway. *Plant J* **2008**, *55*, 289-300, doi:10.1111/j.1365-313X.2008.03502.x.
9. Taniguchi, M.; Sasaki, N.; Tsuge, T.; Aoyama, T.; Oka, A. ARR1 directly activates cytokinin response genes that encode proteins with diverse regulatory functions. *Plant Cell Physiol* **2007**, *48*, 263-277, doi:10.1093/pcp/pcl063.
10. Massolo, J.F.; Lemoine, M.L.; Chaves, A.R.; Concellon, A.; Vicente, A.R. Benzyl-aminopurine (BAP) treatments delay cell wall degradation and softening, improving quality maintenance of refrigerated summer squash. *Postharvest Biology And Technology* **2014**, *93*, 122-129, doi:10.1016/j.postharvbio.2014.02.010.
11. Xu, F.; Chen, X.H.; Yang, Z.F.; Jin, P.; Wang, K.T.; Shang, H.T.; Wang, X.L.; Zheng, Y.H. Maintaining quality and bioactive compounds of broccoli by combined treatment with 1-methylcyclopropene and 6-benzylaminopurine. *J Sci Food Agr* **2013**, *93*, 1156-1161, doi:10.1002/jsfa.5867.
12. Yaish, M.W.; Colasanti, J.; Rothstein, S.J. The role of epigenetic processes in controlling flowering time in plants exposed to stress. *J Exp Bot* **2011**, *62*, 3727-3735, doi:10.1093/jxb/err177.
13. Kapranov, P.; Cheng, J.; Dike, S.; Nix, D.A.; Duttagupta, R.; Willingham, A.T.; Stadler, P.F.; Hertel, J.; Hackermuller, J.; Hofacker, I.L., et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **2007**, *316*, 1484-1488, doi:10.1126/science.1138341.
14. Zhang, J.; Mujahid, H.; Hou, Y.; Nallamilli, B.R.; Peng, Z. Plant Long ncRNAs: A New Frontier for Gene Regulatory Control. *American Journal of Plant Sciences* **2013**, *04*, 1038-1045, doi:10.4236/ajps.2013.45128.

15. Tian, J.X.; Song, Y.P.; QingzhangDu; Yang, X.H.; Ci, D.; Chen, J.H.; Xie, J.B.; Li, B.L.; Zhang, D.Q. Population genomic analysis of gibberellin-responsive long non-coding RNAs in Populus. *J Exp Bot* **2016**, *67*, 2467-2482, doi:10.1093/jxb/erw057.
16. Elvira-Matelot, E.; Bardou, F.; Ariel, F.; Jauvion, V.; Bouteiller, N.; Le Masson, I.; Cao, J.; Crespi, M.D.; Vaucheret, H. The Nuclear Ribonucleoprotein SmD1 Interplays with Splicing, RNA Quality Control, and Posttranscriptional Gene Silencing in Arabidopsis. *Plant Cell* **2016**, *28*, 426-438, doi:10.1105/tpc.15.01045.
17. Michel, A.P.; Sim, S.; Powell, T.H.; Taylor, M.S.; Nosil, P.; Feder, J.L. Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 9724-9729, doi:10.1073/pnas.1000939107.
18. Wittkopp, P.J.; Haerum, B.K.; Clark, A.G. Evolutionary changes in cis and trans gene regulation. *Nature* **2004**, *430*, 85-88, doi:10.1038/nature02698.
19. Ghotbi, R.; Gomez, A.; Milani, L.; Tybring, G.; Syvanen, A.C.; Bertilsson, L.; Ingelman-Sundberg, M.; Aklillu, E. Allele-specific expression and gene methylation in the control of CYP1A2 mRNA level in human livers. *Pharmacogenomics J* **2009**, *9*, 208-217, doi:10.1038/tpj.2009.4.
20. Tao, H.; Cox, D.R.; Frazer, K.A. Allele-specific KRT1 expression is a complex trait. *Plos Genet* **2006**, *2*, 848-858, doi:ARTN e9310.1371/journal.pgen.0020093.
21. Julkowska, M.M.; Klei, K.; Fokkens, L.; Haring, M.A.; Schranz, M.E.; Testerink, C. Natural variation in rosette size under salt stress conditions corresponds to developmental differences between Arabidopsis accessions and allelic variation in the LRR-KISS gene. *J Exp Bot* **2016**, *67*, 2127-2138, doi:10.1093/jxb/erw015.
22. Wang, T.; Liang, L.; Xue, Y.; Jia, P.F.; Chen, W.; Zhang, M.X.; Wang, Y.C.; Li, H.J.; Yang, W.C. A receptor heteromer mediates the male perception of female attractants in plants (vol 531, pg 241, 2016). *Nature* **2016**, *536*, 360-360, doi:10.1038/nature17985.
23. Pandey, D.M.; Goswami, C.L.; Kumar, B. Physiological effects of plant hormones in cotton under drought. *Biol Plantarum* **2003**, *47*, 535-540.
24. Ding, X.T.; Jiang, Y.P.; Wang, H.; Jin, H.J.; Zhang, H.M.; Chen, C.H., et al.; Yu, J.Z. Effects of cytokinin on photosynthetic gas exchange, chlorophyll fluorescence parameters, antioxidative system and carbohydrate accumulation in cucumber (Cucumis sativus L.) under low light. *Acta Physiol Plant* **2013**, *35*, 1427-1438, doi:10.1007/s11738-012-1182-9.
25. Zhang, H.; Li, C.; Xiao, K. Regulation effects of exogenous 6-BA on photosynthesis and leaf senescence in cotton. *Cotton Science* **2007**, *19*, 467-471, doi:10.3969/j.issn.1002-7807.2007.06.010.
26. Chen, B.; Yang, H. 6-Benzylaminopurine alleviates chilling injury of postharvest cucumber fruit through modulating antioxidant system and energy status. *J Sci Food Agric* **2013**, *93*, 1915–1921, doi:10.1002/jsfa.5990.
27. Salin, M.L. Toxic oxygen spcies and protective systems of the chloroplast. *Physiologia Plantarum* **1988**, *72*, 681-689, doi:10.1111/j.1399-3054.1988.tb09182.x.
28. Thompson, D.M.; Parker, R. Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. *Molecular and cellular biology* **2007**, *27*, 92-101, doi:10.1128/MCB.01023-06.

29. Kadota, M.; Niimi, Y. Effects of cytokinin types and their concentrations on shoot proliferation and hyperhydricity inin vitropear cultivar shoots. *Plant Cell Tissue and Organ Culture 72*, 261-265, doi:10.1023/a:1022378511659.
30. Milani, L.; Lundmark, A.; Nordlund, J.; Kiialainen, A.; Flaegstad, T.; Jonmundsson, G.; Kanerva, J.; Schmiegelow, K.; Gunderson, K.L.; Lonnerholm, G., et al. Allele-specific gene expression patterns in primary leukemic cells reveal regulation of gene expression by CpG site methylation. *Genome Res* **2009**, *19*, 1-11, doi:10.1101/gr.083931.108.
31. Weber, M.; Hellmann, I.; Stadler, M.B.; Ramos, L.; Paabo, S.; Rebhan, M.; Schubeler, D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **2007**, *39*, 457-466, doi:10.1038/ng1990.
32. Gehring, M.; Huh, J.H.; Hsieh, T.F.; Penterman, J.; Choi, Y.; Harada, J.J.; Goldberg, R.B.; Fischer, R.L. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **2006**, *124*, 0-506, doi:10.1016/j.cell.2005.12.034.
33. Kinoshita, T.; Miura, A.; Choi, Y.; Kinoshita, Y.; Cao, X.; Jacobsen, S.E.; Fischer, R.L.; Kakutani, T. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **2004**, *303*, 521-523, doi:10.1126/science.1089835.
34. Van der Does, D.; Boutrot, F.; Engelsdorf, T.; Rhodes, J.; McKenna, J.F.; Vernhettes, S.; Koevoets, I.; Tintor, N.; Veerabagu, M.; Miedes, E., et al. The Arabidopsis leucine-rich repeat receptor kinase MIK2/LRR-KISS connects cell wall integrity sensing, root growth and response to abiotic and biotic stresses. *Plos Genet* **2017**, *13*, doi:ARTN e100683210.1371/journal.pgen.1006832.
35. Elvira-Matelot, E.; Hachet, M.; Shamandi, N.; Comella, P.; Saez-Vasquez, J.; Zytnicki, M.; Vaucheret, H. Arabidopsis RNASE THREE LIKE2 Modulates the Expression of Protein-Coding Genes via 24-Nucleotide Small Interfering RNA-Directed DNA Methylation. *Plant Cell* **2016**, *28*, 406-425, doi:10.1105/tpc.15.00540.
36. Du, Q.Z.; Xu, B.H.; Gong, C.R.; Yang, X.H.; Pan, W.; Tian, J.X.; Li, B.L.; Zhang, D.Q. Variation in growth, leaf, and wood property traits of Chinese white poplar (Populus tomentosa), a major industrial tree species in Northern China. *Can J Forest Res* **2014**, *44*, 326-339, doi:10.1139/cjfr-2013-0416.
37. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks (vol 7, pg 562, 2012). *Nat Protoc* **2014**, *9*, 2513-2513, doi:10.1038/nprot1014-2513a.
38. Sun, L.; Zhang, Z.H.; Bailey, T.L.; Perkins, A.C.; Tallack, M.R.; Xu, Z.; Liu, H. Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse Klf1 knockout study. *Bmc Bioinformatics* **2012**, *13*, doi:Artn 33110.1186/1471-2105-13-331.
39. Jia, H.; Osak, M.; Bogu, G.K.; Stanton, L.W.; Johnson, R.; Lipovich, L. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *Rna* **2010**, *16*, 1478-1487, doi:10.1261/rna.1951310.
40. Tafer, H.; Hofacker, I.L. RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* **2008**, *24*, 2657-2663, doi:10.1093/bioinformatics/btn193.

41. Liu, Q.; Ding, C.J.; Chu, Y.G.; Zhang, W.X.; Guo, G.G.; Chen, J.F.; Su, X.H. Pln24NT: a web resource for plant 24-nt siRNA producing loci. *Bioinformatics* **2017**, *33*, 2065-2067, doi:10.1093/bioinformatics/btx096.

42. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **2009**, *10*, R25, doi:10.1186/gb-2009-10-3-r25.

43. El Baidouri, M.; Do Kim, K.; Abernathy, B.; Arikit, S.; Maumus, F.; Panaud, O.; Meyers, B.C.; Jackson, S.A. A new approach for annotation of transposable elements using small RNA mapping. *Nucleic Acids Res* **2015**, *43*, doi:ARTN e8410.1093/nar/gkv257.

44. Quinlan, A.R.; Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841-842, doi:10.1093/bioinformatics/btq033.

45. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754-1760, doi:10.1093/bioinformatics/btp324.

46. Hardy, O.; Vekemans, X. SPAGeDI: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2002**, *2*, 618-620, doi:10.1046/j.1471-8286.2002.00305.x.

47. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* **2005**, *14*, 2611-2620, doi:10.1111/j.1365-294X.2005.02553.x.

48. Storey, J.D.; Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *16*, 9440, doi:10.1073/pnas.1530509100.