Journal of Experimental Botany, Vol. 66, No. 7 pp. 1891–1905, 2015
doi:10.1093/jxb/eru531 Advance Access publication 22 January 2015

RESEARCH PAPER

Methylation of microRNA genes regulates gene expression in bisexual flower development in andromonoecious poplar

Yuepeng Song[1,2], Min Tian[1,2], Dong Ci[1,2], 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, PR China
[2] Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, College of Biological Sciences and Technology, Beijing Forestry University, No. 35, Qinghua East Road, Beijing 100083, PR China

* To whom correspondence should be addressed. E-mail: DeqiangZhang@bjfu.edu.cn

Received 13 September 2014; Revised 15 November 2014; Accepted 10 December 2014

Abstract

Previous studies showed sex-specific DNA methylation and expression of candidate genes in bisexual flowers of andromonoecious poplar, but the regulatory relationship between methylation and microRNAs (miRNAs) remains unclear. To investigate whether the methylation of miRNA genes regulates gene expression in bisexual flower development, the methylome, microRNA, and transcriptome were examined in female and male flowers of andromonoecious poplar. 27 636 methylated coding genes and 113 methylated miRNA genes were identified. In the coding genes, 64.5% of the methylated reads mapped to the gene body region; by contrast, 60.7% of methylated reads in miRNA genes mainly mapped in the 5′ and 3′ flanking regions. CHH methylation showed the highest methylation levels and CHG showed the lowest methylation levels. Correlation analysis showed a significant, negative, strand-specific correlation of methylation and miRNA gene expression (r=0.79, P <0.05). The methylated miRNA genes included eight long miRNAs (lmiRNAs) of 24 nucleotides and 11 miRNAs related to flower development. miRNA172b might play an important role in the regulation of bisexual flower development-related gene expression in andromonoecious poplar, via modification of methylation. Gynomonoecious, female, and male poplars were used to validate the methylation patterns of the miRNA172b gene, implying that hyper-methylation in andromonoecious and gynomonoecious poplar might function as an important regulator in bisexual flower development. Our data provide a useful resource for the study of flower development in poplar and improve our understanding of the effect of epigenetic regulation on genes other than protein-coding genes.

Key words: DNA methylation, flower development, gene expression, microRNA, *Populus*.

Introduction

Cytosine methylation functions as an important epigenetic regulator of transposon silencing, heterochromatin organization, genomic imprinting, and gene expression (Zhang *et al.*, 2006; Suzuki *et al.*, 2008). Plant genomes have high, and highly variable, levels of cytosine methylation (Vaughn *et al.*, 2007), including interspersed methylated and non-methylated regions (Suzuki *et al.*, 2008; Zemach *et al.*, 2010). Roughly 30% of DNA methylation occurs in genic regions, but the majority of DNA methylation occurs in intergenic regions, in heterochromatin at pericentromeric and subtelomeric repeats, and at rDNA clusters (Zhang *et al.*, 2006). The relationship between DNA methylation and gene expression has been extensively studied. It appears that different methylated regions in genes have different effects on the regulation of gene expression. Analysis in *Arabidopsis thaliana* revealed that gene methylation in promoter regions generally associates with a greater degree of tissue-specific expression, whereas methylation in transcribed regions associates with higher levels of expression (Zhang *et al.*, 2006). By contrast, Vining *et al.* (2012) found that, in poplar, gene body methylation associated more with

a repression of transcription than did promoter methylation. This divergence might be affected by gene length and, in addition, by local epigenetic modification (Zilberman *et al.*, 2007; Vining *et al.* 2012).These studies show that gene length might be closely associated with the DNA methylation level, pattern, and the regulation of gene expression. Previous studies have focused on the cytosine methylation of genic regions, but the regulatory relationship between methylation and elements of intergenic regions remains unclear.

As important elements encoded by loci in the intergenic regions of the genome, miRNAs are short, non-coding RNA molecules that range from ~20–24 nt, and negatively regulate gene expression at the transcriptional and/or post-transcriptional levels by degrading or inhibiting the translation of target mRNAs (Chen, 2010). According to the data in the miRbase database, the poplar genome contains more than 400 miRNA genes. In humans, miRNAs participate in the regulation of about two-thirds of genes (Friedman *et al.*, 2009; Peter, 2009). Furthermore, transcription factors and epigenetic modification also regulate the expression of miRNA genes (Bracken *et al.*, 2008; Vrba *et al.*, 2010). For example, aberrant DNA methylation of miRNA promoters resulted in repressed expression in many cancer types (Li *et al.*, 2011). Down-regulation of the tumour suppressors miR355 and miR125b via aberrant DNA methylation is associated with increased malignancy or metastatic potential in breast cancer, illustrating the importance of the DNA methylation-mediated regulation of miRNAs in cancer (Png *et al.*, 2011).

DNA methylation and miRNA gene expression may have a complex interaction in plants. *Arabidopsis* miRNAs include canonical, ~21 nt miRNAs, and DCL3-dependent 24 nt variants referred to as long miRNAs, lmiRNAs (Wu *et al.*, 2010). The 21 nt RNAs generally repress expression of their target gene through mRNA cleavage, while lmiRNAs can direct cytosine DNA methylation at their own loci in *cis* and at their target genes in *trans*, resulting in transcriptional gene silencing (Wu *et al.*, 2010). Compared with protein-coding genes, miRNA genes are smaller and located in intergenic regions that may show differences in epigenetic modification and regulation of gene expression. However, DNA methylation of miRNA genes in plants remains poorly explored, particularly on the whole-genome scale.

Trees, with their long life spans and generation times, have to acclimate to different environments and are, therefore, models of interest for epigenetic studies (Hamanishi and Campbell, 2011). Poplar has been widely recognized as a model tree because of its available genome sequence, extensive transcriptome data and various molecular tools (Tuskan *et al.*, 2006; Jansson and Douglas, 2007). Poplar epigenomics research used HPLC to detect changes in global DNA methylation under stress conditions (Gourcilleau *et al.*, 2010; Raj *et al.*, 2011). Recently, with the development of next-generation sequencing techniques, research using high-resolution epigenomic methods has detected tissue-level variation in DNA methylation (Vining *et al.*, 2012; Lafon-Placette *et al.*, 2013). These studies used different methods to examine DNA methylation, including whole-genome shotgun bisulphite sequencing (WGsBs), Methylated DNA immunoprecipitation (MeDIP) sequencing, and DNase I-MeDIP sequencing. Among these methods, the comprehensive DNA methylation analysis technique of whole-genome bisulphite sequencing provides single-base resolution, good coverage of regions with low CpG density, and information on *cis* co-methylation (Laird, 2010). Thus whole-genome bisulphite sequencing is an ideal method for global methylation analysis.

Andromonoecious poplars and gynomonoecious poplars have been found in natural poplar populations. Unlike female and male poplars, which produce separate female and male flowers, andromonoecious poplars produce both male and hermaphrodite catkins, which include female and male flowers next to each other in bisexual flowers (Fig. 1A; see Supplementary Fig. S1A at *JXB* online). By contrast, gynomonoecious poplars produce female and hermaphrodite catkins, which include female and male flowers in a cup-shaped structure (see Supplementary Fig. S1B at *JXB* online). Our previous studies showed sex-specific DNA methylation and candidate gene expression in female and male flowers of andromonoecious poplar (Song *et al.*, 2012). In order to investigate whether DNA methylation affects the expression of miRNA and protein-coding genes at the whole-genome level, a comprehensive analysis of the poplar methylome, microRNAs, and transcriptome was conducted here, using male and female flowers from andromonoecious poplar. Methylated miRNA genes were identified and it was shown that their methylation patterns differ between miRNA and protein-coding genes. Correlation analysis showed a negative correlation between DNA methylation and miRNA gene expression. Moreover, it was shown that DNA methylation regulates a set of important flowering-related miRNA genes and their targets.

Materials and methods

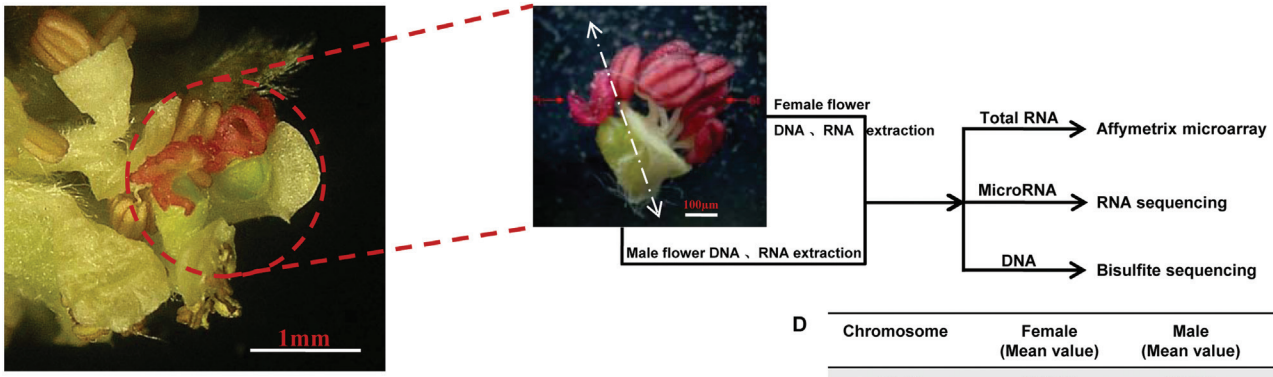
Plant materials

Observation of flower morphology identified 17 unrelated andromonoecious clones and 23 gynomonoecious clones in a natural population of 460 29-year-old unrelated individuals, representing almost the whole geographic distribution of *Populus tomentosa* in China. Three andromonoecious clones ('2–14', '3605', and '5103'), as biological replicates, were used in the methylome, microRNA, and transcriptome analyses. For sampling, female and male flowers were dissected from the hermaphrodite catkins at the last phase of flower development, before pollination (25 February).To enhance the reliability and association of methylome, microRNA, and transcriptome data, DNA, total RNA, and miRNA were simultaneously extracted from flowers using different extraction kits. Comparison of the female and male flower libraries of andromonoecious clones was used to identity differentially -methylated regions, expressed miRNAs, and expressed genes. The methylation and expression of candidate gene were validated in female and male flowers of gynomonoecious poplar, and in male poplar flowers and female poplar flowers, to validate the relationship of DNA methylation levels and bisexual flower development in poplar.

Bisulphite sequencing

Genomic DNAs were isolated using the Plant Genomic DNA Purification Kit (Zexing Inc., China). DNA was fragmented

A



B

	Female	Male
Biological replicates	3	3
Raw reads (Total)	228,643,605	250,454,484
Raw reads (Mean value)	76,214,535	83,484,828
Mapped reads (Total)	22,178,427	24,294,084
Mapped reads (Mean value)	7,392,809	8,098,028
Mapping efficiency (Mean value)	9.7%	11.6%

C

	Female			Male		
	Methylated	Unmethylated	Ratio	Methylated	Unmethylated	Ratio
C	31,552,817	197,090,787	13.8%	36,816,809	213,637,674	14.7%
CpG	15,781,873	17,065,669	47.9%	15,864,137	18,773,717	45.8%
CHG	10,568,822	38,818,196	21.4%	11,524,290	521,828,904	21.6%
CHH	5,276,179	141,284,371	3.6%	9,413,081	152,881,423	5.8%
Total	228,643,605			250,454,484		

D

Chromosome	Female (Mean value)	Male (Mean value)	DMR (Total)
Chr01	15.08%	13.90%	376±4
Chr02	17.37%	16.03%	169±7
Chr03	16.85%	15.49%	1±0
Chr04	14.76%	13.61%	1±0
Chr05	15.95%	14.66%	1±0
Chr06	16.70%	15.25%	173±5
Chr07	15.20%	14.08%	0±0
Chr08	18.83%	17.18%	115±9
Chr09	19.77%	18.13%	66±9
Chr10	18.33%	16.73%	0±0
Chr11	13.30%	12.30%	0±0
Chr12	13.86%	12.73%	3±1
Chr13	15.49%	14.26%	144±2
Chr14	16.69%	15.44%	144±8
Chr15	15.74%	14.55%	0±0
Chr16	14.80%	13.59%	114±5
Chr17	13.56%	12.48%	179±3
Chr18	14.17%	13.08%	0±0
Chr19	11.28%	10.46%	148±9

Fig. 1. The approaches used to obtain methylome, miRNA, and transcriptome data. (A) Description of the plant material and methylome, miRNA, and transcriptome approaches. The pictures show the procedure to isolate female and male flowers of andromonoecious poplar. The left red arrow shows the female flower and the right red arrow shows the male flower. (B) Summary of bisulphite sequencing samples and results. (C) Summary of the methylation ratio in different CG contexts between female and male flowers. (D) Summary of methylation on different chromosomes between female and male flowers. DMR represents differentially methylated regions.

by sonication with the Diagenome sonicator to a mean size of approximately 250 bp, followed by DNA repair of blunt ends, 3′-end addition of dA, and adaptor ligation according to the manufacturer's instructions (Illumina, San Diego, US). Adapter-ligated genomic DNA ranging from120–170 bp was isolated and subjected to sodium bisulphite conversion using the EpiTect Bisulfite kit (Qiagen, Valencia, CA). Conversion efficiency was approximately 99% using HPLC detection (see Supplementary Table S1 at *JXB* online). Bisulphite-treated DNAs were PCR amplified with 16 cycles. The resultant DNAs were used for paired-end sequencing with a read length of 101 nt for each end, using the ultra-high-throughput Illumina Hiseq2000 as per the manufacturer's instructions.

To verify the bisulphite sequencing data, primers specific for different methylated regions were designed using the Methyl Primer Express (v1.0) software (Herman *et al.*, 1996). According to the detailed procedure of Trap-Gentil *et al.* (2011), methyl-sensitive PCR (MS-PCR) was performed using bisulphite-treated DNA as templates, and was carried out for 25 cycles. PCR products were cloned into pMD18-T (Takara BioInc., Tokyo, Japan), and three positive clones for each individual were selected for sequencing. Primer sequences and annealing temperatures are listed in Supplementary Table S2 at *JXB* online.

Bioinformatic and statistical analyses

Raw data quality control reports were generated by FastQC Tools (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). High quality sequencing reads (Q >20) from female and male flowers were trimmed to 101 nt and mapped to the *P. trichocarpa* v2.2 reference genome (http://www.phyozome.net/popalr.php). The Bismark alignment tool (http://www.bioinformatics.babraham.ac.uk/projects/bismark/) was used for sequence alignment and methylation sites calling with default parameters. HashMatch was used to identify reads that align to multiple locations. To eliminate the bias produced by the alignment of duplicates generated by PCR, a de-duplication step was used to remove reads mapping to the same position of the reference genome. The coverage of cytosine sites and methylation levels were calculated in different conditions, such as different samples and on different chromosomes. Methylation level is defined as follows:

$$\text{Absolute methylation level} = \frac{\text{Total methylation level of mCs}}{\text{Total sequence length of the calculated region}}$$

$$\text{Relative methylation level} = \frac{\text{Total methylation level of mCs}}{\text{Total number cytosine sites in the calculated region}}$$

To detect differentially methylated sites, Fisher's Exact Test was used to test cytosine sites of female and male flowers with a minimum coverage of >=2. Minimum coverage represents the minimum threshold value of sequencing depth required to identify the differentially methylated sites between samples. To identify the most sites that are differentially methylated, two was chosen as the minimum threshold value of sequencing depth. After that, sliding windows and ANOVA were adopted to call differentially methylated regions (DMRs). The same coverage regions in all samples was used for identifying DMRs. Considering that the length of the DMR cannot be predicted before calling the DMRs, the size of the sliding windows was set to 1 000 bp, representing part of the continuous DMR sampling results. After that, those windows with probabilities of less than 0.05 were merged into larger regions to estimate the mean and variance of entire methylation regions. In these regions, ANOVA was used again to filter DMRs with P <0.05 (called candidate DMRs).

To analyse the distribution of methylation sites on different regions of the poplar genome, repeats sequences were detected by RepeatMasker (http://www.repeatmasker.org/) using the *P. trichocarpa* genome as the reference. Gene annotation from the Phytozome poplar genome v2.2 (http://www.phytozome.net/poplar.php) was used. After methylated reads were mapped to the reference genome, the difference in percentage of reads aligned to repeats and genes were compared. Genes with methylation at promoters, and/or within annotated transcribed regions were compared to archival microarray expression data in order to determine the correlation between methylation and expression. Pearson's correlation analysis (r_p) and evaluation of the statistical significance of methylation level and gene expression was performed as described by Lafon-Placette *et al.* (2013). Chi-squared homogeneity tests were used to evaluate the effects on mapping distribution.

Gene and miRNA expression data analysis

Three independent replicates at the final phase of flower development were used for each sample. Fresh tissue samples of flowers for RNA extraction were collected from the three andromonoecious poplars. Total RNAs were amplified, labelled, and purified using the GeneChip 3′IVT Express Kit (Affymetrix, Santa Clara, CA, US) following the manufacturer's instructions, to obtain biotin-labelled cDNA. One-channel chip hybridizations were performed by Shanghai Bio using the Affymetrix Genechip Poplar Genome Array. The chip hybridization results were scanned with a GeneChip Scanner 3000 and were normalized with the MAS 5.0 algorithm, and Gene Spring Software 11.0. The quality assurance/quality control measures had an average background of <100. Genes with more than a 2-fold change in the female flower compared with the male flower were selected. The fold-change analysis data were filtered by t-test (P <0.05), and gene expression pattern hierarchical clustering analysis was conducted using the complete linkage clustering (Cut off=0.6, Exponent=2) Cluster program. To avoid false positives, the q-value for the minimum false discovery rate at which the test may be called significant was calculated. Genes differentially expressed between the female and male flowers were selected at a q-value of <0.05.

For miRNA analysis in our study, total RNA was isolated from female and male floral tissue by a modified CTAB method (Chang *et al.*, 1993) with isopropanol instead of lithium chloride for RNA precipitation. Extracted RNA was used for small RNA (sRNA) library construction. The sRNAs were sequenced using an Illumina HiSeq 2000 at the Shanghai Bio Institute. Sequencing reads were aligned against the *P. trichocarpa* genome v2.2 using SOAP v2.21 (Li *et al.*, 2008). Sequences with a perfect match were used to BLAST against GenBank (http://www.ncbi.nih.gov/GenBank/) and Rfam v11.0 (http://rfam.sanger.ac.uk/) databases for annotation. rRNAs, scRNA, tRNAs, snRNAs, and snoRNAs were removed in sequencing reads. The remaining un-annotated sRNAs were searched against miRBase 19.0 with a maximum of two mismatches allowed to identify conserved miRNAs (Griffiths-Jones *et al.*, 2008). To predict novel miRNAs, miRCat (http://srna-tools.cmp.uea.ac.uk/plant/cgi-bin/srna-tools.cgi?rm=inputform&tool=mircat) was used to explore the secondary structures, DCL1 cleavage sites, and minimum free energies of the un-annotated sRNA tags that could be mapped to the *Populus* genome. miRNA expression was compared between female and male libraries (three biological replicates for each sex) to determine which miRNAs were differentially expressed. The expression levels of miRNAs within each library were normalized to get the expression in transcripts per million mapped reads. miRNA target genes were predicted using psRNATarget with settings: Maximum expectation <2.0, length for complementarity scoring <20, and target accessibility <20.

5′-RACE

RNA Ligase-Mediated 5′-RACE (RLM-RACE) was performed with the First Choice RLM-RACE Kit (Ambion, Austin, TX), as described by Song *et al.* (2013*b*). PCR was performed with 5′ adaptor primers and 3′ gene-specific primers using cDNA as the template (see Supplementary Table S3 at *JXB* online). The RACE products were gel-purified, cloned, and sequenced.

Gene functional enrichment analysis

The annotation information for differentially expressed genes was obtained from GenBank and KEGG (http://www.genome.jp/kegg). For enrichment analysis, the Biological Process (BP) branches of Gene Ontology (GO) were used. Applying the true-path rule, a gene annotated with a particular GO term was also annotated with all its parents. GO terms of BP were tested for statistical significance of enrichment using the cumulative hypergeometric test (Song *et al.*, 2013*a*). To adjust for multiple comparisons, a Benjamini–Hochberg false discovery rate (FDR; q value) (Benjamini and Hochberg, 1995) was calculated from the P values, and a q value threshold of 0.01 was used for significance. The graph was visualized using Cytoscape (Shannon *et al.*, 2003).

Gene expression analysis by qRT-PCR

The quantitative PCR amplifications were performed according to Song *et al.* (2012). The real-time PCR primer pairs are shown in Supplementary Table S4 at *JXB* online. The efficiency of the primer sets was calculated by performing real-time PCR on several dilutions of first-strand cDNAs. The different primer sets amplified with similar efficiencies. The specificity of each primer set was checked by sequencing PCR products (Zhang *et al.*, 2010). The results obtained for the different phases analysed were standardized to the levels of the *PtACTIN* and *UBIQUITIN* transcripts, which have stable expression during floral development (Song *et al.*, 2013*a*).

Results

DNA methylation sequence data

After removing low-quality and duplicate reads, ~76–83 million uniquely mapped high-quality reads were obtained for each line (Fig. 1B). After calling methylation sites, 22 785 332 and 24 381 556 methylated cytosines (mCs), on average, were identified in male and female flower libraries, accounting for 14.7% and 13.8%, respectively, of all covered cytosines throughout the reference genome (Fig. 1C). The percentages

of mCs in CG, CHG (with H being A, C, or T) and CHH contexts were 45.8%/47.9% (male/female), 21.6%/21.4% (male/female), and 5.8%/3.6% (male/female), respectively (Fig.1C).

Mapping of bisulphite sequencing reads to the genome

After filtering, 7 392 809 (female) and 8 098 028 (male) reads were mapped on the *P. trichocarpa* $_{v}$2.2 reference genome including protein-coding genes, microRNA genes, repeat elements, and intergenic regions. Mapped reads account for only 9.7% and 11.6% of male and female sequencing reads, respectively. Among the mapped reads, it was found that 74.1% of total sequence reads mapped to intergenic sequences, similar methylation patterns to the results observed in *A. thaliana* and *Oryza sativa* (Zhang *et al.*, 2006; Li *et al.*, 2012). It was also found that 10.1% of methylated reads mapped to repeats and 15.1% mapped to genes, but these differences were not statistically significant (Fig. 2A). At the scaffold level, among all 19 chromosomal scaffolds, methylated reads showed a significantly biased distribution: fewer reads than expected, based on scaffold sizes, mapped to scaffolds 4 (χ^2=4.36, P <0.05), 9 (χ^2=6.74, P <0.01), and 14 (χ^2=5.47, P <0.05) and more reads mapped to scaffolds 12 (χ^2=5.76, P <0.05), 15 (χ^2=6.41, P <0.05), and 17 (χ^2=4.82, P <0.05) (see Supplementary Fig. S2 at *JXB* online).

Mapping of bisulphite sequencing reads to genes

15.1% of the bisulphite sequencing reads were mapped to 66.8% of the v2.2 model genes (27 636 protein-coding genes and 113 miRNA genes), in the body regions and/or in the 5′ and 3′ flanking regions (±2kb). 95.7% of these reads were mapped to coding genes and 4.3% to miRNA genes (Fig. 2B). Most of methylated regions were enriched in the protein-coding gene body regions (64.5%), with exons and introns accounting for 56.3% and 23.5% of the reads, respectively (Fig. 2C, D). 7.4% and 9.6% of protein-coding genes were methylated in the 5′ and 3′ flanking regions (±2kb), respectively. Only 2.8% of the protein-coding gene body and

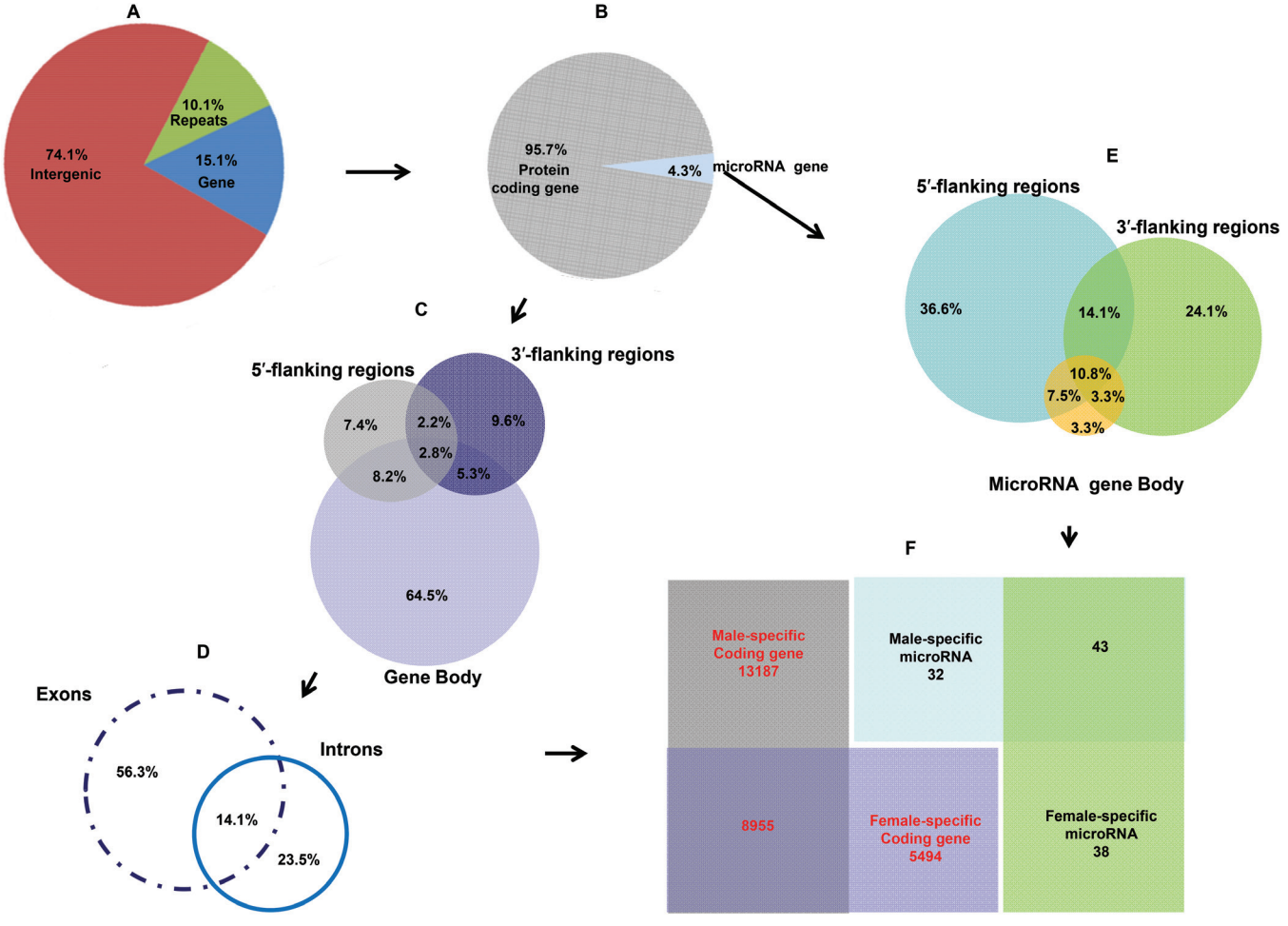


Fig. 2. Classification of methylation sites on different regions of protein-coding genes and microRNA genes. (A) Pie chart representing the proportion of regions covering v2.2 gene models, repeats or intergenic loci in the whole-genome bisulphite sequencing fraction. (B) Pie chart representing the proportion of protein-coding genes and microRNA genes in the whole-genome bisulphite sequencing fraction. (C) Venn diagram showing the percentage of protein-coding genes with differentially methylated regions. Protein-coding genes include the protein-coding gene body, 2kb 5′ flanking regions, and 2kb 5′ flanking regions. (D) Venn diagram showing the percentage of protein-coding gene bodies with differentially methylated regions. Protein-coding gene bodies include the protein-coding gene exon and intron. (E) Venn diagram showing the percentage of miRNA genes with differentially methylated regions. miRNA genes include the miRNA gene body, 2kb 5′ flanking regions, and 2kb 5′ flanking regions. (F) Venn diagram showing the number of protein-coding genes and microRNA genes with sex-specific DNA methylation.

flanking regions were methylated simultaneously (Fig. 2C). By contrast, most miRNA genes were methylated in the flanking regions including 36.6% and 24.1% in 5′ and 3′ flanking regions, respectively. Only 3.3% of miRNA genes were only methylated in gene bodies. 10.8% of miRNA gene bodies and flanking regions were methylated simultaneously (Fig. 2E).

Compared with different length protein-coding genes, miRNA genes showed significantly lower cytosine methylation levels (see Supplementary Fig. S3 at *JXB* online). In protein-coding genes, the average cytosine methylation level including all three cytosine methylation contexts (CG, CHG, and CHH) showed relative higher coverage in the 5′ and 3′ flanking regions, decreased dramatically at the borders of the genes, and increased to a peak in the central part of the gene body. CG (~16.54–46.89%) showed the highest methylation and CHG (~15.69–20.24%) showed the lowest. By contrast, in miRNA genes, the average methylation levels of all three cytosine methylation contexts showed relatively higher coverage in the 5′ and 3′ flanking regions, significantly decreased at the transcription start site, and remained lowest in the transcribed regions. CHH (5.91%~6.04%) showed the highest levels and CG (4.79%~5.64%) showed the lowest levels (Fig. 3; see Supplementary Fig. S4 at *JXB* online).

Methylation levels in different contexts between female and male flowers

The three cytosine methylation contexts showed significantly different methylation levels between female and male poplar flowers. CHH showed the highest methylation levels and CHG showed the lowest methylation levels. CHH methylation levels were higher in male flowers than in female flowers (Fig. 4A); CG did not differ between female and male flowers. By contrast, CHG showed site-specific methylation on C(T/A)G in male flowers and C(A/T)G in female flowers while CHH showed site-specific methylation on C(A/T)(A/T) in male flowers and C(A/C)(A/T) in female flowers. CHH methylation was principally enriched in intergenic regions and was significantly higher in male flowers than in female flowers. By contrast, CG methylation was mainly located in gene bodies and was not significantly different between female and male flowers (Fig. 4B).

Methylated regions were categorized by the genome feature of different tissues including promoter regions and the gene body of the protein-coding gene and miRNA genes. Among the protein-coding genes, the gene body of the andromonoecious poplar was the most frequently methylated (~31.5–48.5%). The percentage of methylated gene bodies was significantly higher in female flowers than in male flowers (Fig. 5). In the andromonoecious poplar, the promoter of miRNA genes was the most frequently methylated (~90–96%) and the fraction of methylated promoters was significantly higher than the fraction of methylated gene bodies in both female and male flowers (Fig. 5).

Regulatory roles of DNA methylation and miRNAs in gene expression

It was found that genes with intermediate expression levels tended to have higher methylation levels in the two tissues

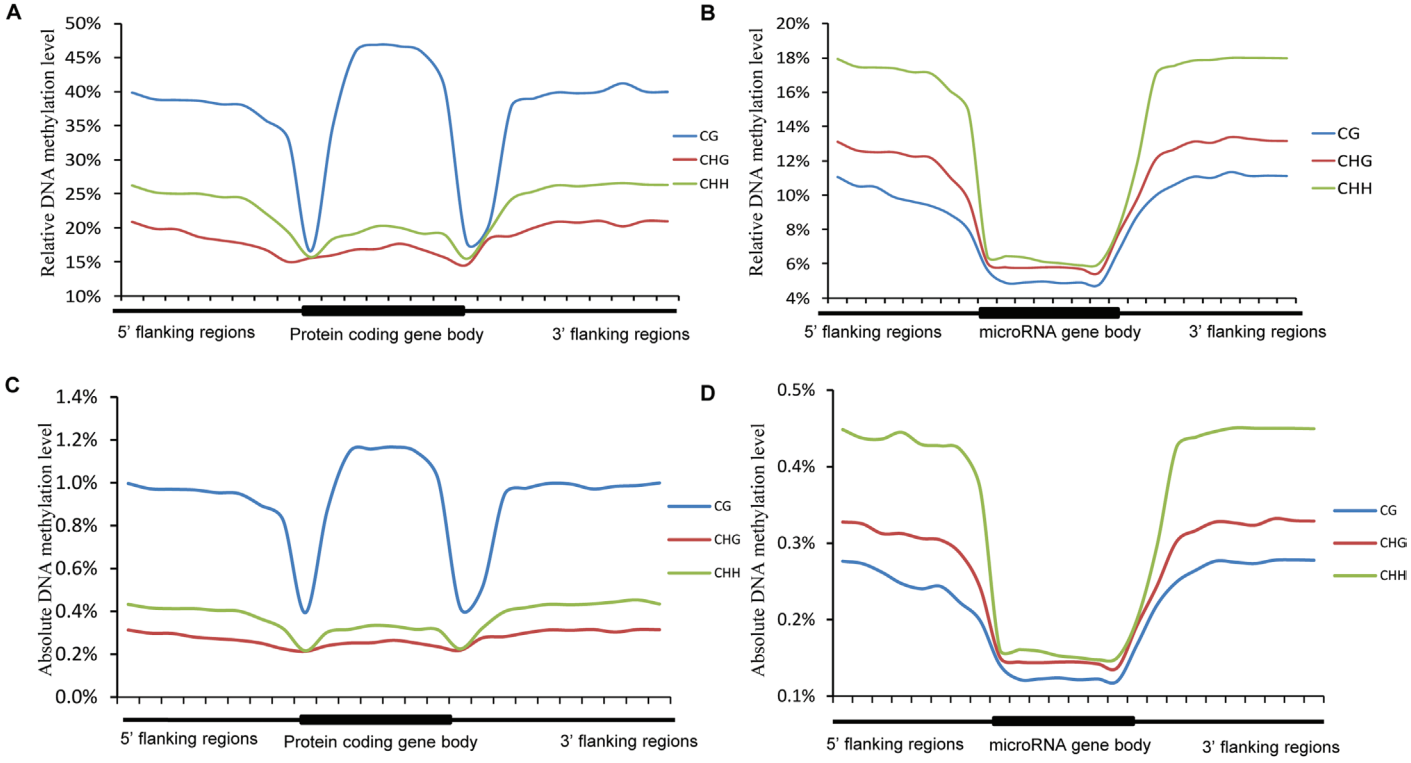


Fig. 3. DNA methylation patterns in different genomic regions. (A) Relative methylation level in protein-coding gene body and 2 kb flanking sequences on both sides. (B) Relative methylation level in microRNA gene bodies and 2 kb flanking sequences on both sides. (C): Absolute methylation level in the protein-coding gene body and the 2 kb flanking sequences on both sides. (D) Absolute methylation level in microRNA gene bodies and 2 kb flanking sequences on both sides.

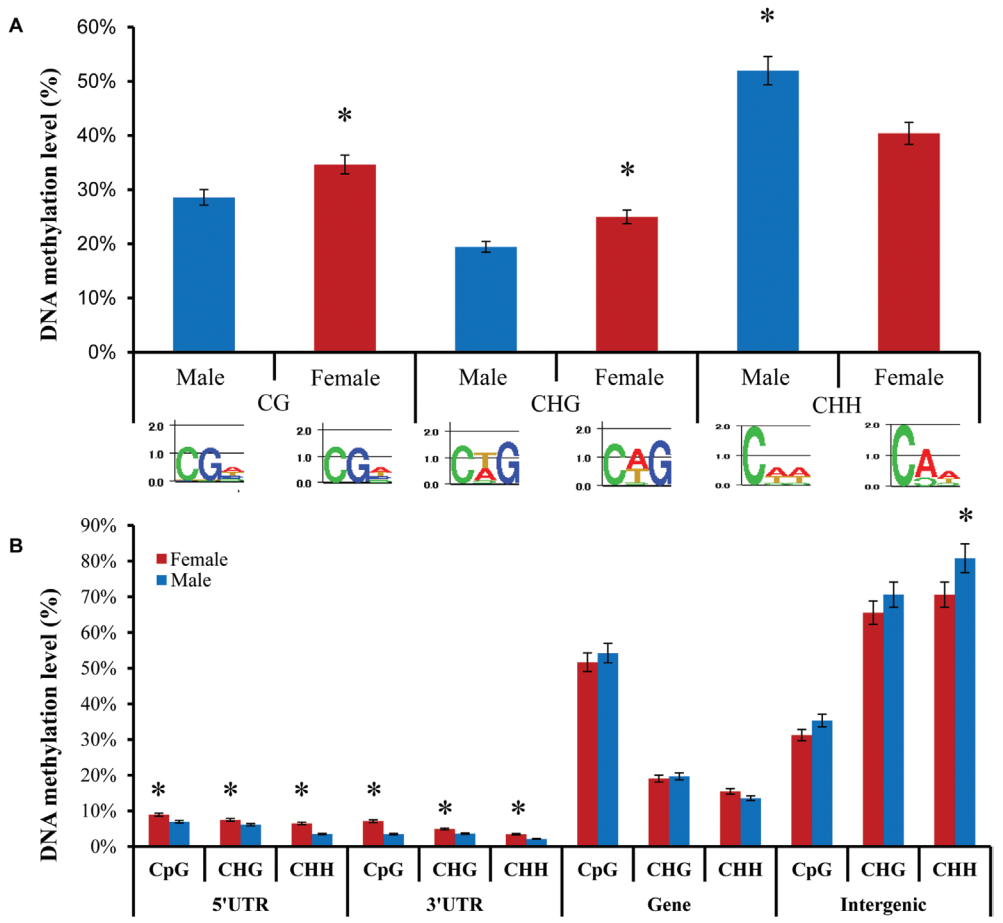


Fig. 4. The differential distribution of methylated CG and non-CG contexts between female and male flowers. (A) Preferentially methylated CG and non-CG contexts between female and male flowers. Sequence logos were obtained using GENIO/LOGO software (http://www.biogenio.com/logo/). Sequence logos were made with the same number of each CHG or CHH combination to eliminate site frequency-dependent bias. (B) CG and non-CG contexts differentially distributed along gene models between female and male flowers. Asterisk on error bars indicates significant differences at P <0.01.

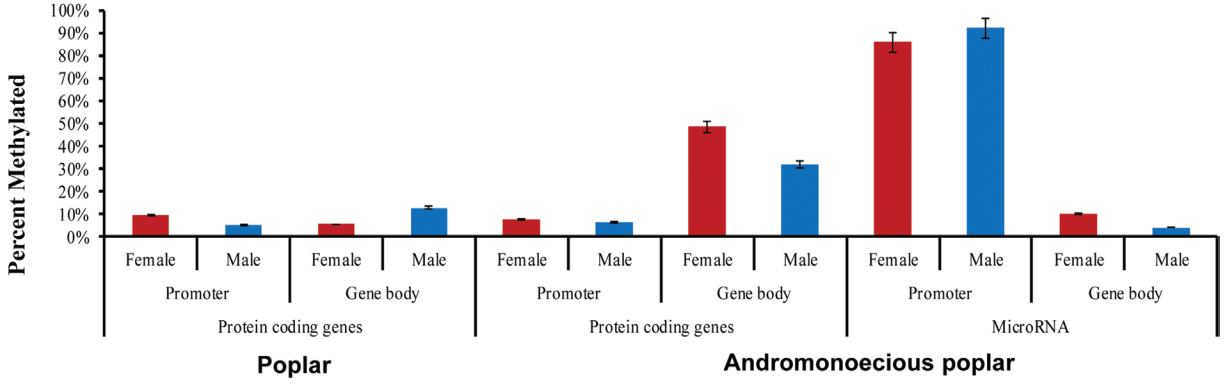


Fig. 5. The percentage of methylated promoter and gene body between female and male flowers of normal poplar and andromonoecious poplar. Bars show the percentage of each feature type determined to be significantly methylated. The data for female and male poplar flowers were obtained from Vining *et al.* (2012).

than those genes with higher or lower expression levels. At the genome level, a significant negative correlation was confirmed between miRNA gene expression and methylation (r=–0.79, P <0.05). Next, the expression levels of protein-coding genes and miRNA genes with different methylation statuses were compared. Among the protein-coding genes, methylated genes were significantly repressed to different levels. Protein-coding genes with methylated sites enriched in the 3' flanking regions and on both the 5′ and 3′ flanking regions were repressed more than genes with methylated sites enriched in the 5′ flanking regions and the gene body (Fig. 6A). Among the miRNA genes, expression of the miRNA genes with methylated sites enriched in the 5′ regions, 3′ flanking regions, or both 5′ and 3′ flanking regions, were significantly repressed (P <0.05) (Fig. 6B). In addition, since DNA methylation has strand specificity, the effects of methylation on the plus and minus

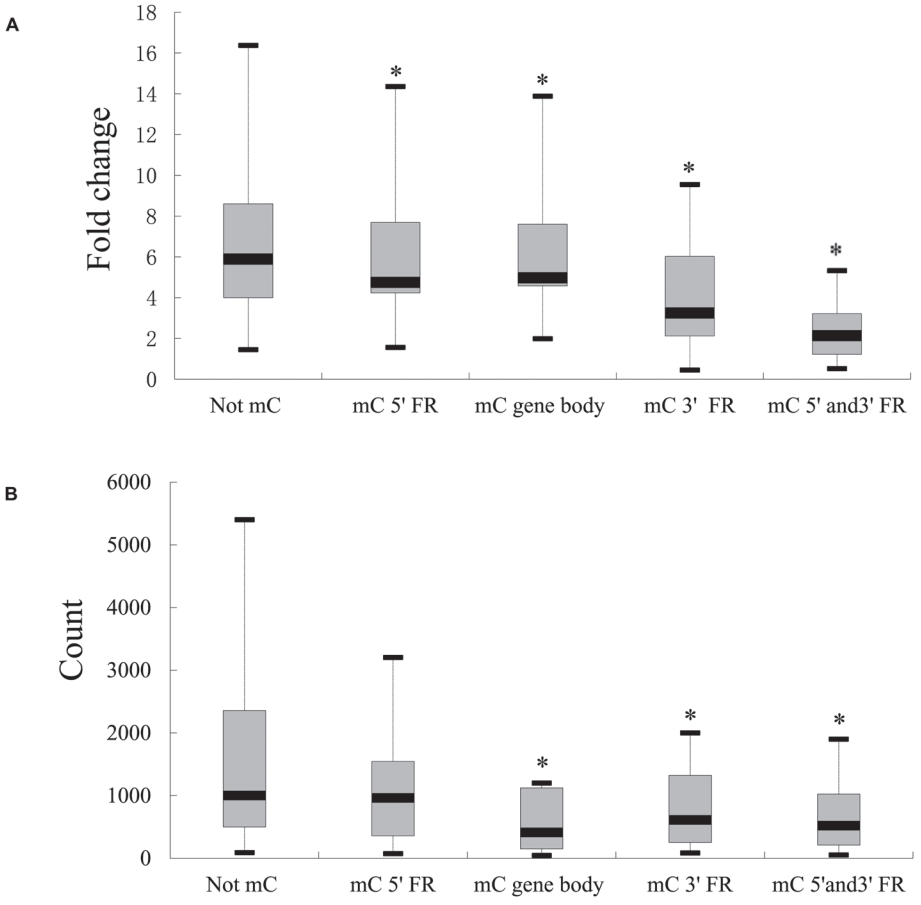


Fig. 6. Association of gene expression with methylation. (A) Box plots representing the association of protein-coding gene expression with DNA methylation in different regions. (B) Box plots representing the association of microRNA gene expression with DNA methylation in different regions. 'Not mC', unmethylated genes; 'mC 5′ FR', methylated 5′ flanking regions; 'mC gene body', methylated on gene body; 'mC 3′ FR', methylated on 3′ flanking regions; 'mC 5′and 3′ FR', methylated both on 5′ and 3′ flanking regions. Asterisks indicate significantly different means (P <0.05) based on a comparison of methylated coding genes and microRNA genes to unmethylated coding genes and microRNA genes.

strands were compared. It was found that methylation on the minus strand had a stronger correlation with repression of gene expression than methylation on the plus strand, both in protein-coding and miRNA genes (r_m=–0.89, r_p=–0.71, P <0.01, see Supplementary Fig. S5 at *JXB* online).

Among the 113 methylated miRNA genes (see Supplementary Excel S1 at *JXB* online), eight belong to the 24 nt long miRNAs (lmiRNAs). Of these, five lmiRNAs came from the miR478 gene family (miRNA478a, b, j, k, and u), and the others from miR474b, 481e, and 1449 (Table 1). Forty-three targets (including different transcript variants) of these microRNAs were predicted using psRNATarget tools (see Supplementary Excel S2 at *JXB* online). To confirm the regulatory patterns of these miRNAs and targets, 5′-RACE was used on a flower mRNA library to detect the cleavage products of the predicted targets but, for 17 targets, no cleavage products were found. Next, bisulphite sequencing was used to examine the status of DNA methylation at these predicted targets. Among these targets, it was found that 10 targets were methylated within 65 nt regions around the target sites (Table 1). These targets include miR1449 targets Potri.004G081200 and Potri.005G236500, the miR474b target Potri.011G001100, the miR478a and miR478b co-targets Potri.005G229000 and Potri.009G011200, the miR478j and miR478k co-targets Potri.001G284000 and Potri.016G032700, the miR478u target Potri.006G070000, and three miR481e targets Potri.013G108200, Potri.013G118500, and Potri.003G120800. Although no cleavage products were found, the expression of these target genes was still negatively associated with lmiRNA gene expression (see Supplementary Table S5 at JXB online). Moreover, Potri.003G120800, an miR481e target, which has four transcript variants, was methylated in two regions flanking splice sites. Potri.006G070000, an miR478u target, which has three transcript variants, was only methylated in one splice site flanking region (see Supplementary Fig. S6 at JXB online).

Gene ontology analysis of methylated genes and miRNA targets

Gene ontology (GO) analysis was used to characterize functionally the differentially methylated genes that had cytosine DNA methylation (5mC) in different gene regions. The GO terms of protein-coding genes with 5mC enrichment in the 5′ and 3′ flanking regions were enriched in DNA packaging, chromosome organization, the nitrogen compound metabolic process, the biosynthetic process, and the regulation of the cellular process for the biological process as well as

Table 1. *DNA methylation patterns and target genes of eight lmiRNAs*

Percentages represent relative cytosine methylation levels in these regions.

No.	Genome location	Female flower			Male flower			Target genes[a]	Inhibition
		5′ Flanking regions	Gene body	3′ Flanking regions	5′ Flanking regions	Gene body	3′ Flanking regions		
miR478a	Chr12:7133508-7133603	21.3%						Potri.009G011200[b]	Translation
miR478b	Chr12:7130082-7130177			16.4%			34.5%	Potri.005G229000[b]	Translation
miR478j	Chr06:2596810-2596868	38.1%						Potri.001G280400[c]	Cleavage
miR478k	Chr05:13672333-13672391	56.7%			74.4%		12.9%	Potri.016G032700[c]	Translation
miR478u	Chr19:14406415-14406509			31.1%			23.3%	Potri.006G070000[d]	Translation
miR474b	Chr01:6456792-6456898	1.9%					13.7%	Potri.T049300[d]	Cleavage
								Potri.004G044800[d]	Cleavage
								Potri.011G053700[d]	Cleavage
								Potri.004G018000[d]	Cleavage
								Potri.011G001100[d]	Translation
miR481e	Chr11:10598914-10599109	8.7%		46.2%		15.6%	9.7%	Potri.013G108200	Translation
								Potri.008G221700	Cleavage
								Potri.013G118500	Translation
								Potri.003G120800	Translation
miR1449	Chr02:24046728-24046842	5.9%						Potri.004G081200	Translation
								Potri.005G236500	Translation
								Potri.017G120500	Cleavage

[a] Target genes were obtained from psRNATarget tools (http://plantgrn.noble.org/psRNATarget).
[b] These represent co-targets of miR478a and miR478b.
[c] These represent co-targets of miR478j and miR478k.
[d] These represent co-targets of miR478u and miR474b.

oxidoreductase activity, nucleic acid binding, and ion binding for molecular function (see Supplementary Fig. S7 at *JXB* online). The GO terms of the protein-coding genes with 5mC enrichment in the gene body were enriched in fundamental processes known to be affected by methylation, including the primary metabolic process, the macromolecular metabolic process, and the cellular metabolic process for the biological process as well as transferase activity, hydrolase activity, nucleoside binding, and nucleotide binding for molecular function (see Supplementary Fig. S8 at *JXB* online). GO classification of the targets of the methylated miRNA genes showed that signal transducer activity, nucleoside binding, nucleotide binding, and transferase activity were enriched in GO terms of molecular function. In the biological process, programmed cell death, the cellular macromolecular metabolic process, the protein metabolic process, and the carbohydrate metabolic process terms were also enriched.

DNA methylation and expression of floral development-related miRNAs and their targets

Eleven known miRNAs related to flower development were detected in our bisulphite sequencing data, including miR156, miR159, miR164, miR169, miR172, and miR319. miR156l and k showed different methylation patterns in both female and male flowers (see Supplementary Table S6 and Supplementary Fig. S9 at *JXB* online). Both the 5′ and 3′ flanking regions of miR156l showed male-specific methylation and the miR156k gene body showed female-specific methylation. However, the expression of miR156l and k showed higher expression in female flowers than in male flowers. miR159a and miR319f were only methylated in the 5′ flanking regions in male flowers, and these genes were repressed in male flowers. Two members of the miR164 gene family were identified, miR164a and miR164e, which showed significantly higher methylation levels in male flowers than in female flowers and repressed expression in male flowers. Both the 5′ and 3′ flanking regions of miR164a were only methylated in male flowers and CHH methylation levels were higher than CpG and CHG levels. Three members of the miR169 gene family, miR169u, q and t, were also detected in the bisulphite sequencing data. miR169u and t were methylated in the 5′ flanking regions in male flowers and miR169q was methylated in the gene body regions in female flowers; all of these miRNA genes were repressed in male flowers. miR172b showed different methylation and expression levels in female and male flowers (Fig. 7; see Supplementary Table S10 at *JXB* online). However, miR172i showed female-specific methylation in the 3′ flanking regions and was repressed in female flowers.

The microarray data for the expression of targets of these miRNA genes were also examined (see Supplementary Table S6 at *JXB* online). The results showed that *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*), a target of miR156, was repressed in female flowers. *NUCLEAR TRANSCRIPTION FACTOR Y,* and *ALPHA* (*NFYA*), targets of miR169, were repressed in female flowers. miR172 was repressed in female flowers, and its target *APETALA 2* (*AP2*) was induced. Three MYB genes (*MYB33*, *MYB65*,

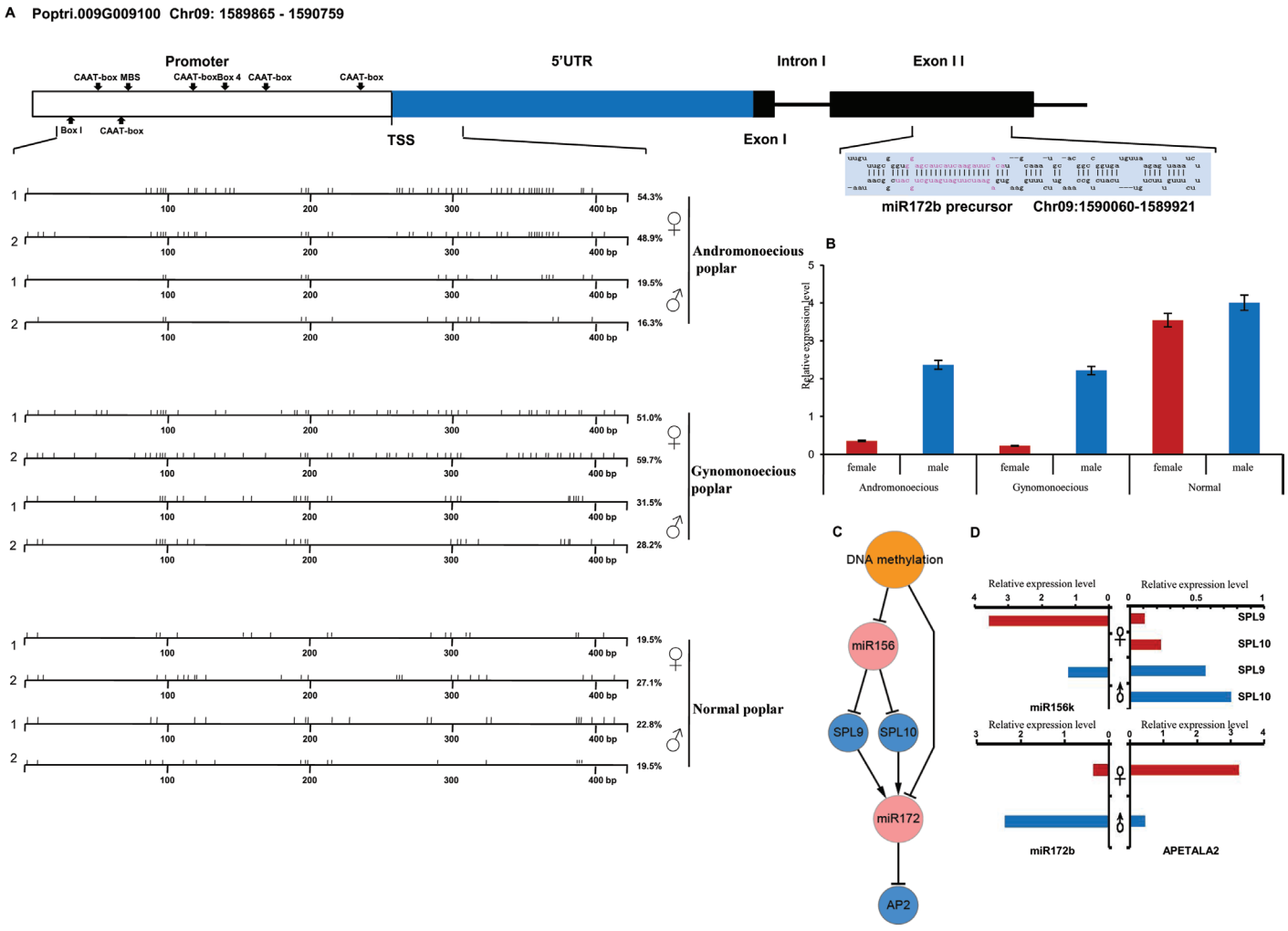


Fig. 7. DNA methylation changes in miR172b genes in andromonoecious, gynomonoecious, and normal poplar flowers. (A) Schematic representation of the differentially methylated levels upstream of the miR172b gene. Two biological replicates were used. Horizontal lines represent the gene sequence; the vertical bar represents the methylated CpG sites. The relative DNA methylation level was calculated. Arrows represent predicted *cis*-element in the promoter region. (B) miR172b gene expression pattern in female and male flowers of andromonoecious, gynomonoecious, and normal poplar. (C) Schematic diagram representing the network of DNA methylation, miR156, miR172, and their target. Line arrows represent positive regulation, and lines ending with a bar represent negative regulation. (D) miR156k, miR172b and their target gene expression patterns in female and male flowers of andromonoecious poplar.

and *MYB101*) and four TCP genes (*TCP2*, *3*, *10*, and *24*), targets of miR159 and miR319, respectively, were highly expressed in male flowers. Two targets of miR164, *CUP-SHAPED COTYLEDON* (*CUC1*) and *CUC2*, were induced in male flowers. Subsequently, real-time quantitative PCR was used to confirm the expression of miRNA target genes. The expression of all target genes was negatively correlated with the levels of a given miRNA, in accordance with the gene-silencing function of miRNAs (see Supplementary Fig. S11 at *JXB* online).

Validation of bisulphite sequencing and mapping

Among differentially methylated miRNA and target gene sequences, 41 regions were selected (9 036 bp in total) representing a variety of methylation patterns to confirm using the method of Trap-Gentil *et al.* (2011) (see Supplementary Table S2 and Supplementary Fig. S12 at *JXB* online). This validation suggested that bisulphite sequencing is an effective, stable, and reproducible technology for the detection of methylation sites in the poplar genome. To verify the results of cDNA microarray and microRNA sequencing, candidate protein-coding and miRNA genes representing a variety of functional categories and expression patterns were also selected for qRT-PCR analysis to validate the transcriptome data (see Supplementary Table S4 and Supplementary Fig. S11 at *JXB* online). The results showed that these genes have the same expression tendency as observed in the microarray data, although some absolute expression levels were different between the two analytical platforms, suggesting that the expression data are reproducible and reliable.

In order to verify the methylation pattern of miR172b in bisexual flowers, miR172b methylation was measured in gynomonoecious poplar bisexual flowers and compared with that in andromonoecious, female, and male poplar (Fig. 7A). The results showed that the relative methylation level of

miR172b was 53.47% in andromonoecious and gynomonoecious female flowers which was significantly higher than in male flowers (23.8%). In poplar, the relative methylation level of miR172b was 23.3% and 21.15% in female and male flowers, respectively. Although three SNPs also have been detected in miR172b flanking regions, a significant association of these SNP sites with candidate gene expression was not detected (see Supplementary Fig. S13 at *JXB* online).

Discussion

Distribution and preference of methylation

The coverage efficiency of reads mapping was 9.7% and 11.6% in male and female flower sequencing reads, respectively. This is higher than coverage in *P. trichocarpa* shoot apical meristematic cells and closer to 10% coverage in poplar apices, but significantly lower than DNA methylation in the male and female flowers of *P. trichocarpa* (27% in female flowers and 31.6% in male flowers) (Vining *et al.*, 2012; Lafon-Placette *et al.*, 2013). Thus, the lower coverage of methylated reads observed in andromonoecious poplar might be tissue-specific. Second, our study used a *P. tomentosa* genotype, not *P. trichocarpa* Nisqually-1, which was used as a reference for reads assembly and mapping. The genetic variability between *P. tomentosa* and *P. trichocarpa* might be a reason for the low mapping ratios. Finally, our observation that 14.45% of annotated protein-coding genes are methylated is consistent with the methylation levels of different tissues in *P. trichocarpa*, suggesting that most of the variation might come from non-coding regions. It is, therefore, speculated that whole-genome bisulphite sequencing might detect more intergenic regions which show a high degree of variation, compared with the sequences detected by MeDIP-seq or microarray techniques.

At the scaffold level, although no positive correlation was established between the number of mapped reads and the scaffold size of the *P. trichocarpa* $_{v}$2.2 genome, methylated reads showed a significantly biased distributions on scaffolds 12, 15, and 17, by contrast with methylated reads of poplar shoot apical meristematic cells, which showed significantly biased distributions on 8, 9, 14, and 17 scaffolds (Lafon-Placette *et al.*, 2013). Nonetheless, our results support the hypothesis that tissue-specific DNA methylation exists in poplar. Our results showed that 113 miRNA genes (~28.2% of all known miRNA genes of plants) are methylated in poplar, suggesting that miRNA genes are also extensively methylated in plants. In protein-coding genes, it was found that methylation is lowest near the transcription start and stop sites and increases in the gene body but, in miRNA genes, the gene body showed significantly lower methylation levels. In general, gene length is an important factor for the distribution of DNA methylation (Zilberman *et al.*, 2007; Lafon-Placette *et al.*, 2013). Thus, smaller gene size might be the main reason for low methylation levels in the gene bodies of miRNA genes.

Lafon-Placette *et al.* (2013) reported that contexts that contain at least two cytosines were enriched for non-CG contexts in shoot apical meristematic cells. This is inconsistent with our results that non-CG contexts tend to have cytosines and adenine. Also, different frequently methylated sites were enriched in male and female flowers, suggesting that frequently methylated sites show some tissue specificity. Moreover, different sequence types, including protein-coding genes, rDNA, and transposable elements, showed different levels and context preferences of methylation (Lafon-Placette *et al.*, 2013). In this study, different levels and context types of methylation were observed between protein-coding genes and miRNA genes, supporting the hypothesis that the preference for different methylation contexts depends on the type of sequences.

DNA methylation and gene expression

Several studies have demonstrated that upstream sequence, gene body, and downstream sequence methylation is negatively correlated with transcription (Zhang *et al.*, 2006; Vining *et al.*, 2012). Our results support this notion, as expression of downstream-methylated and upstream and downstream-methylated genes was significantly repressed compared with unmethylated genes. The interaction of RNA-binding proteins with 5′- or 3′-untranslated regions (UTRs) of mRNAs plays an important role in translational regulation (Mazumder *et al.*, 2003). The interactions between 5′- or 3′- UTRs mediated by proteins result in the formation of an RNA loop that can increase translational efficiency and tends to be regulated by the 3′ UTR (Mazumder *et al.*, 2003; Szostak and Gebauer, 2012). Multi-species analysis indicated that 3′ UTRs are substantially longer than 5′ UTRs, indicating that the 3′ UTRs have important functions in regulating gene expression (Kuersten *et al.*, 2003). Similarly, in our study, downstream-methylated genes showed significantly lower expression than upstream and gene body methylated genes, suggesting that downstream methylation might provide a stronger contribution to the regulation of gene expression. By contrast, in our study, expression of protein-coding genes with only promoter and gene body methylation did not significantly differ from unmethylated genes. This could be explained because the expression data were generated by microarray, which might inhibit the integration of methylome and transcriptome data, suggesting that further studies will require methylome/transcriptome studies on a single platform.

DNA methylation has strand specificity and strand-biased methylation occurs under environmental stress and in different regions of the genome (Ros and Kunze, 2001; Luo and Preuss, 2003). Under stressful environmental conditions, seedlings have enhanced strand-specific methylation, established during development and heritable over several days of seedling growth. For example, the *Arabidopsis* centromeric heterochromatin shows strand-biased methylation, suggesting that methylation might play important roles in centromere activity (Luo and Preuss, 2003). For transposable elements, the activator-dissociation transposon preferentially transposes from a non-methylated DNA strand, showing strongly chromatid selectivity (Ros and Kunze, 2001). These strand biases in methylation could modulate the transcript abundance of genes, transposons, and intergenic transcripts (Lister *et al.*, 2008). Thus, in this study, the

plus and minus strands for methylation and gene expression analysis were separated at the whole-genome scale. The results showed that methylation on the minus strand has a stronger correlation with repression of gene expression than methylation on the plus strand, for both coding and microRNA genes, suggesting that minus strand methylation might have an important role in regulation of gene expression. Although a significant correlation between minus strand methylation and gene repression was detected, it is believed that these are minimal estimates of the strength of these associations. Strand-specific transcript sequencing, which was not used in our current study, could increase the resolution and reduce the bias of our transcript mapping (Lister *et al.*, 2008). Thus, strand-specific transcript sequencing might be required for further strand-biased methylation analysis.

Previous studies have observed that the tissue-specific differentiation of gene expression relates to DNA methylation in plants, indicating that DNA methylation may play an important role in directing or maintaining differential gene expression, though its extent appears modest (Wang *et al.*, 2009; Vining *et al.*, 2012). Our results showed consistent tissue-level patterns, implying that differential methylation of miRNA genes might play a crucial role in directing or maintaining bisexual flower development-related gene expression of andromonoecious poplar.

In plants, DNA methylation is closely associated with miRNA gene expression. Miguel and Marum (2011) indicated that heterochromatin undergoes DNA hypomethylation resulting in the transcriptional activation of specific transposable elements accompanied by the production of 21 nucleotide smRNAs. miRNAs modulate their target gene expression by mRNA cleavage (Llave *et al.*, 2002) or by repressing translation (Chen, 2004; Lanet *et al.*, 2009). lmiRNAs can direct cytosine DNA methylation at their own loci in *cis* and at their target genes in *trans*, resulting in transcriptional gene silencing (Wu *et al.*, 2010). It was found that ten targets of eight lmiRNAs had lower methylation levels and increased expression, together with increased methylation of the miRNA genes, suggesting that these target genes might be regulated by the release of lmiRNA-directed DNA methylation. DNA methylation also plays an important role in the regulation of splicing events and thus in the final constitution of the protein sequence (Sati *et al.*, 2012). Our data showed that DNA methylation occurred on the different transcripts of target genes, implying that the patterns and levels of DNA methylation might regulate the ratios of different transcripts.

Flower development-related miRNA genes were extensively regulated by DNA methylation

Previous studies indicated that miR172 and miR156 are critical for the juvenile to adult phase change and the vegetative to reproductive transition in flowering plants (Lauter *et al.*, 2005; Gandikota *et al.*, 2007). miR156 expression showed a complementary temporal pattern to seven *SPL* targets during vegetative development (Wu and Poethig, 2006; Wu *et al.*, 2009). Two miR156 targets, *SPL9* and *SPL10*, positively regulate miR172 expression by binding to sequences in the regulatory region (Wu *et al.*, 2009). In our study, a decrease in miR172b expression was correlated with an increase in miR156l and k expression in female flowers, indicating that the indirect regulation of miR172 by miR156 exists in poplar as well. Also, our results showed no significant association of three SNP sites in candidate regions with candidate gene expression, suggesting that these mutation sites do not affect candidate gene expression. Also, miR156l, k, and miR172b are differentially methylated in male and female flowers, suggesting that DNA methylation might function mainly as a regulatory factor in flower development affecting gene expression in andromonoecious poplar. miR172 represses *AP2* in the centre of the flower, which is crucial for the proper development of the reproductive organs and for the timely termination of floral stem cells (Chen, 2004; Luo *et al.*, 2013). Our results showed that miR172b is repressed via methylation in the carpel (whorl 4) of andromonoecious poplar flowers, suggesting that methylation might be a key factor for hermaphrodite flower development.

Members of the miR164 gene family function as crucial factors in lateral organ separation, organ boundary formation, and lateral organ proliferation, via their regulation of a subset of NAC transcription factors (Nag and Jack, 2010). *CUC1* and *CUC2* belong to the NAC gene family and function in meristem maintenance and lateral organ separation (Aida *et al.*, 1997). In *cuc1* mutants, the sepals fuse slightly, and the cotyledons fuse to form a single cup-shaped structure, suggesting that *CUC* genes affect the formation of lateral organ boundaries by suppressing cell growth between lateral organs (Laufs *et al.*, 2004; Jones-Rhoades *et al.*, 2006). Our results showed that miR164a and e are specifically methylated in male flowers, leading to an increase in *CUC1* and *CUC2* expression, correlated with the decrease in miR164a expression. This might help to explain why boundaries between and within whorls of organs in male flowers are completely established, compared with female flowers.

The miR159 family targets *MYB33*, *MYB65*, and *MYB101*/*DUO POLLEN1* (*DUO1*) (Millar and Gubler, 2005). *Arabidopsis myb33 myb65* double mutants show tapetum hypertrophy and pollen abortion, implying that those genes have crucial functions in anther development. A decrease in *MYB33* gene expression correlated with an increase in miR159 gene expression would cause anther defects, male sterility, and delayed flowering (Achard *et al.*, 2004; Schwab *et al.*, 2005). Our results showed that miR159 is repressed via DNA methylation, leading to a higher expression level of *MYB33* in male flowers, implying that DNA methylation might protect the normal development of anthers and pollen in male flowers of andromonoecious poplar. miR319 and miR159 evolved from a common ancestor and have distinct expression patterns (Li *et al.*, 2011), but they target different genes and have different functions in plant development (Palatnik *et al.*, 2007). miR319 targets several TCP transcription factor genes that control leaf and flower growth (Ng *et al.*, 2009; Luo *et al.*, 2013). In this study, it was observed that miR319 and miR159 have distinct DNA methylation and expression patterns. Combined with the above analysis, it is speculated that DNA methylation might function as an important factor for the maintenance of miR319 and miR159 expression patterns and regulate their overlapping functions.

In antirrhinum and petunia, miR169 can activate C-class genes by modulating its target NF-YA transcription factors to control the development of reproductive organs (Cartolano *et al.*, 2007). NF-YA binding affects enhancement and maintenance of *AG* transcription (Hong *et al.*, 2003), and *AG* interacts genetically with other homeotic genes to specify carpel and stamen identity (Ng *et al.*, 2009). Our results showed that the miR169/NF-YA module also exists in poplar and is regulated by DNA methylation, but the relationship of this module with hermaphrodite flower development of andromonoecious poplar is still unclear.

Conclusion

Our observation that protein-coding genes are methylated is consistent with the methylation levels in *P. trichocarpa*, but less than 10% of reads could be mapped on the reference genome, supporting our speculation that most of the variation in methylation between *P. tomentosa* and *P. trichocarpa* might come from non-coding regions. So, the *de novo* assembly of the *P. tomentosa* genome might help to improve our results in the future. Regarding the distribution of methylation sites, miRNA and protein-coding genes have significantly different methylation patterns. It is speculated that gene sizes might be the major reason for the low methylation levels in miRNA gene body regions. Statistical analysis implied that downstream and minus strand methylation might provide a stronger contribution to the regulation of gene expression. DNA methylation not only affects microRNA gene expression via the modification of microRNA body and flanking regions, but also results in transcriptional gene silencing via the modification of target genes of lmiRNAs, implying that DNA methylation, microRNAs, and gene expression show various interactions.

Our study confirmed that the indirect regulation of miR172 by miR156 exists in poplar as well and that miR156l, k, and miR172b are differentially methylated in male and female flowers, implying that DNA methylation might play an important role in hermaphrodite flower development-related gene expression of andromonoecious poplar. However, more evidence should be gathered from future transgenic experiments to examine the mechanisms by which DNA methylation affects flower development.

Accession numbers

The sequencing and gene expression data reported here are available in NCBI with the SRA database accession numbers SRS557950 and SRS561387 and the GEO accession number GSE38432.

Supplementary data

Supplementary data can be found at *JXB* online.

Supplementary Table S1. The conversion ratio of bisulphite on poplar genomic DNA.

Supplementary Table S2. List of primers used for bisulphite sequencing of candidate genes.

Supplementary Table S3. Primers for 5′ RACE mapping of miRNA cleavage sites.

Supplementary Table S4. Information on real time-PCR primer sequences.

Supplementary Table S5. lmiRNA and their target genes expression patterns.

Supplementary Table S6. miRNA and target gene expression.

Excel S1. Summary of methylation patterns and levels on microRNA gene regions.

Excel S2. Different transcript variants were targeted by lmiRNA.

Supplementary Fig. S1. The pictures show the phenotype of male, female, and gynomonoecious poplar flowers.

Supplementary Fig. S2. Chromosome-level view of methylation between female and male flowers.

Supplementary Fig. S3. DNA methylation levels of genes of different sizes.

Supplementary Fig. S4. Histogram of DNA methylation patterns in different genomic regions.

Supplementary Fig. S5. Correlation analysis between strand-specific DNA methylation levels and gene expression.

Supplementary Fig. S6. Flanking regions of splice sites were methylated.

Supplementary Fig. S7. 5′ and 3′ flanking regions of methylated genes for statistically enriched GO terms in the 'Biological process' ontology.

Supplementary Fig. S8. Gene body methylated genes for statistically enriched GO terms in the 'Biological process' ontology.

Supplementary Fig. S9. Heat map indicating differentially expressed microRNA genes in female and male flowers of andromonoecious poplar.

Supplementary Fig. S10. Gene content in a region with methylation differences between female and male flowers.

Supplementary Fig. S11. Candidate miRNAs and their target gene expression patterns.

Supplementary Fig. S12. Correlation analysis of DNA methylation level detected by candidate sequence bisulphite sequencing and whole genome bisulphite sequencing.

Supplementary Fig. S13.The positions of single nucleotide polymorphisms.

Acknowledgements

This work was supported by grants from the following sources: the Forestry Public Benefit Research Program (No. 201304102), Fundamental Research Funds for Central Universities (No. TD2012-01), the Project of the National Natural Science Foundation of China (No. 31170622, 30872042), and Funds for Training Excellent Talent of Beijing City (No. 2013D009046000004). The authors would like to thank Professor Chung-Jui Tsai from The University of Georgia for critical reading of the manuscript. We are grateful for the sequence information produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov). The authors have no conflict of interest to declare.

References

Achard P, Herr A, Baulcombe DC, Harberd NP. 2004. Modulation of floral development by a gibberellin-regulated microRNA. *Development* **131,** 3357–3365.

Aida M, Ishida T, Fukaki H, Fujusawa H, Tasaka M. 1997. Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The Plant Cell* **9,** 841–857.

Benjamini Y, Hochberg Y.1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57,** 289–300.

Bracken CP, Gregory PA, Kolesnikoff N, *et al*. 2008. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial–mesenchymal transition. *Cancer Research* **68,** 7846–7854.

Cartolano M, Castillo R, Efremova N, Kuckenberg M, Zethof J, Gerats T, Schwarz-Sommer Z, Vandenbussche M. 2007. A conserved microRNA module exerts homeotic control over *Petunia hybrida* and *Antirrhinum majus* floral organ identity. *Nature Genetics* **39,** 901–905.

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

Chen X. 2004. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303,** 2022–2025.

Chen X. 2010. Small RNAs-secrets and surprises of the genome. *The Plant Journal* **61,** 941–958.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* **19,** 92–105.

Gandikota M, Birkenbihl RP, Hohmann S, Cardon GH, Saedler H, Huijser P. 2007. The miR156/157 recognition element in the 3′ UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition. *The Plant Journal* **49,** 683–693.

Gourcilleau D, Bogeat-Triboulot MB, Le Thiec D, Lafon-Placette C, Delaunay A, El-Soud WA, Brignolas F, Maury S. 2010. DNA methylation and histone acetylation: genotypic variations in hybrid poplars, impact of water deficit and relationships with productivity. *Annals of Forest Science* **67,** 208.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36,** 154–158.

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

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences, USA* **93,** 9821–9826.

Hong RL, Hamaguchi L, Busch MA, Weigel D. 2003. Regulatory elements of the floral homeotic gene *AGAMOUS* identified by phylogenetic foot printing and shadowing. *The Plant Cell* **15,** 1296–1309.

Jansson S, Douglas CJ. 2007. *Populus*: a model system for plant biology. *Annual Review of Plant Biology* **58,** 435–458.

Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology* **57,** 19–53.

Kuersten S, Goodwin EB. 2003. The power of the 3′ UTR: translational control and development. *Nature Reviews Genetics* **4,** 626–637.

Lafon-Placette C, Faivre-Rampant P, Delaunay A, Street N, Brignolas F, Maury S. 2013. Methylome of DNase I sensitive chromatin in *Populus trichocarpa* shoot apical meristematic cells: a simplified approach revealing characteristics of gene-body DNA methylation in open chromatin state. *New Phytologist* **197,** 416–430.

Laird PW. 2010. Principles and challenges of genome wide DNA methylation analysis. *Nature Reviews Genetics* **11,** 191–203.

Lauter N, Kampani A, Carlson S, Goebel M, Moose SP. 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proceedings of the National Academy of Sciences, USA* **102,** 9412–9417.

Laufs P, Peaucelle A, Morin H, Traas J. 2004. MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development* **131,** 4311–4322.

Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, Crete P, Voinnet O, Robaglia C. 2009. Biochemical evidence for translational repression by Arabidopsis microRNAs. *The Plant Cell* **21,** 1762–1768.

Li D, Zhao Y, Liu C, *et al.* 2011.Analysis of MiR-195 and MiR497 expression, regulation and role in breast cancer. *Clinical Cancer Research* **17,** 1722–1730.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligo nucleotide alignment program. *Bioinformatics* **24,** 713–714.

Li X, Zhu JD, Hu FY, Ge S, Ye MZ, Xiang H, Zhang GJ, Zheng XM, Zhang HY, Zhang SL. 2012. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* **13,** 300.

Li Y, Li C, Ding G, Jin Y. 2011. Evolution of MIR159/319 microRNAgenes and their post-transcriptional regulatory link to siRNA pathways. *BMC Evolutionary Biology* **11,** 122.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133,** 523–536.

Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297,** 2053–2056.

Luo S, Preuss D. 2003. Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **100,** 11133–11138.

Luo Y, Guo ZH, Li L. 2013. Evolutionary conservation of microRNA regulatory programs in plant flower development. *Developmental Biology* **380,** 133–144.

Mazumder B, Seshadri V, Fox PL. 2003. Translational control by the 3′-UTR: the ends specify the means. *Trends in Biochemical Sciences* **28,** 91–98.

Miguel C, Marum L. 2011. An epigenetic view of plant cells cultured *in vitro*: somaclonal variation and beyond. *Journal of Experimental Botany* **62,** 3713–3725.

Millar AA, Gubler F. 2005. The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant Cell* **17,** 705–721.

Nag A, Jack T. 2010. Sculpting the flower; the role of microRNAs in flower development. *Current Topics in Developmental Biology* **91,** 349–378.

Ng KH, Yu H, Ito T. 2009. AGAMOUS controls GIANT KILLER, a multifunctional chromatin modifier in reproductive organ patterning and differentiation. *PLoS Biology* **7,** e1000251.

Palatnik JF, Wollmann H, Schommer C *et al.* 2007. Sequence and expression differences underlie functional specialization of Arabidopsis microRNAsmiR159 and miR319. *Development Cell* **13,** 115–125.

Peter ME. 2009. Let-7 and miR-200 microRNAs: guardians against pluripotency and cancer progression. *Cell Cycle* **8,** 843–852.

Png KJ, Yoshida M, Zhang XH, *et al*. 2011. MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer. *Genes and Development* **25,** 226–231.

Raj S, Brautigam K, Hamanishi ET, Wilkins O, Thomas BR, Schroeder W, Mansfield SD, Plant AL, Campbell MM. 2011. Clone history shapes *Populus* drought responses. *Proceedings of the National Academy of Sciences, USA* **108,** 12521–12526.

Ros F, Kunze R. 2001. Regulation of Activator/Dissociation transposition by replication and DNA methylation. *Genetics* **157,** 1723–1733.

Sati S, Tanwar VS, Kumar KA, *et al*. 2012. High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. *PLoS ONE* **7,** e31621.

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. 2005. Specific effects of microRNAs on the plant transcriptome. *Development Cell* **8,** 517–527.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13,** 2498–2504.

Song YP, Ma KF, Ci D, Chen QQ, Tian JX, Zhang ZY, Zhang DQ. 2013*a*. Sexual dimorphism floral development in dioecious plants revealed by transcriptome, phytohormone, and DNA methylation analysis in *Populus tomentosa*. *Plant Molecular Biology* **83,** 559–576.

Song YP, Ma KF, Ci D, Zhang ZY, Zhang DQ. 2013*b*. Sexual dimorphism floral microRNA profiling and target gene expression in andromonoecious poplar (*Populus tomentosa*). *PLoS ONE* **8,** e62681.

Song YP, Ma KF, Bo WH, Zhang ZY, Zhang DQ. 2012. Sex-specific DNA methylation and gene expression in andromonoecious poplar. *Plant Cell Reports* **31,** 1393–1405.

Suzuki MM, Bird A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* **9,** 465–476.

Szostak E, Gebauer F. 2012.Translational control by 3′-UTR-binding proteins. *Briefings in Functional Genomics* **12,** 58–65.

Trap-Gentil MV, Hébrard C, Lafon-Placette C, Delaunay A, Hagège D, Joseph C, Brignolas F, Lefebvre M, Barnes S, Maury S. 2011. Time-course and amplitude of DNA methylation in the shoot apical meristem are critical points for bolting induction in sugar beet and bolting tolerance between genotypes. *Journal of Experimental Botany* **62,** 2585–2597.

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

Vaughn MW, Tanurdzić M, Lippman Z, *et al*. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biology* **5,** e174.

Vining KJ, Pomraning KR, Wilhelm LJ, Priest HD, Pellegrini M, Mockler TC, Freitag M, Strauss S. 2012. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics* **13,** 27.

Vrba L, Jensen TJ, Garbe JC, *et al*. 2010. Role for DNA methylation in the regulation of miR-200c and miR-141 expression in normal and cancer cells. *PLoS One* **5,** e8697.

Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW. 2009. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *The Plant Cell* **21,** 1053–1069.

Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. 2009. The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell* **138,** 750–759.

Wu G, Poethig RS. 2006. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* **133,** 3539–3547.

Wu L, Zhou HY, Zhang QQ, Zhang JG, Ni FR, Liu C, Qi YJ. 2010. DNA methylation mediated by a microRNA pathway. *Molecular Cell* **38,** 465–475.

Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328,** 916–919.

Zhang DQ, Du QZ, Xu BH, Zhang ZY, Li BL. 2010. The actin multigene family in *Populus*: organization, expression and phylogenetic analysis. *Molecular Genetics and Genomics* **284,** 105–119.

Zhang X, Yazaki J, Sundaresan A, *et al*. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126,** 1189–1201.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* **39,** 61–69.