Final Project Report

Epigenomics of Development in *Populus*

DOE Plant Feedstock Genomics, Award # ER64665

Steven H. Strauss, Michael Freitag and Todd Mockler, Oregon State University

Submitted December 2012

SYNOPSIS

We conductedresearch to determine the role of epigenetic modifications during tree development using poplar (*Populus trichocarpa*), a model woody feedstock species. Using methylated DNA immunoprecipitation (MeDIP) or chromatin immunoprecipitation (ChIP), followed by high-throughput sequencing, we are analyzed DNA and histone methylation patterns in the *P. trichocarpa* genome in relation to four biological processes: bud dormancy and release, mature organ maintenance, *in vitro* organogenesis, and methylation suppression.

Our project is now completed. We have 1) produced 22 transgenic events for a gene involved in DNA methylation suppression and studied its phenotypic consequences; 2) completed sequencing of methylated DNA from eleven target tissues in wildtype *P. trichocarpa*; 3) updated our customized poplar genome browser using the open-source software tools (2.13) and (V2.2) of the *P. trichocarpa* genome; 4) produced summary data for genome methylation in *P. trichocarpa*, including distribution of methylation across chromosomes and in and around genes; 5) employed bioinformatic and statistical methods to analyze differences in methylation patterns among tissue types; and 6) used bisulfite sequencing of selected target genes to confirm bioinformatics and sequencing results, and gain a higher-resolution view of methylation at selected genes 7) compared methylation patterns to expression using available microarray data.

Our main findings of biological significance are the identification of extensive regions of the genome that display developmental variation in DNA methylation; highly distinctive gene-associated methylation profiles in reproductive tissues, particularly male catkins; a strong whole genome/all tissue inverse association of methylation at gene bodies and promoters with gene expression; a lack of evidence that tissue specificity of gene expression is associated with gene methylation; and evidence that genome methylation is a significant impediment to tissue dedifferentiation and redifferentiation in vitro.

OBJECTIVES AND RESULTS

1. **Produce and characterize DNA methylation-deficient poplars generated by RNA-induced silencing of the poplar homologs of the *Arabidopsis* Decreased DNA Methylation 1 (*DDM1*) gene, *PtDDM1-1* and *PtDDM1-2*.** We have targeted the Decreased DNA Methylation 1 (*DDM1*) gene using a transgenic RNAi approach. *DDM1* is an ATPase in the ISWI family of chromatin remodeling factors. In *Arabidopsis thaliana*, DDM1 is required for the maintenance of CpG and non-CpG DNA methylation and histone H3 methylation patterns. *Arabidopsis ddm1* mutants accumulate developmental and other defects, most often caused by activation of normally silenced repetitive sequences. There are two

paralogs of the *Arabidopsis DDM1* gene in the *P. trichocarpa* genome: *PtDDM1-1 and PtDDM1-2.* We previously reported our success in producing an RNAi construct and using this construct in *Agrobacterium*-mediated transformation of *Populus tremula x alba* female clone (717-1B4), a model genotype used for transgenic studies. We confirmed 22 transgenic events. In addition, we previously reported our efforts toward chemical PtDDM1 suppression using the methyltransferase inhibitor 5-azacytidine. We have taken further steps to analyze both the transgenic events and the 5-azacytidine-treated tissues.

- To assess *PtDDM1* expression in the transgenic events, we performed reverse transcription quantitative RT-PCR (RT-qPCR) with all 22 transgenic events and the 717 control. Each event had 3 biological replicates (independent RNA isolation and cDNA synthesis), and 2 technical replicates (independent qPCR reaction). The qPCR experiments followed a randomized block design on each qPCR plate. We used *The real time PCR miner* (Zhao *et al.,*2005) to calculate normalized R0 from qPCR data. The assayed plants (including control) showed a wide range of *PtDDM1* expression. Transformed lines 3, 15, and 23 appeared to have the strongest suppression of *PtDDM1*, corresponding to approximately 1/3 of the *PtDDM1* expression of the 717 control (Fig. 1).

- A greenhouse-based phenotypic assessment of *Ptddm1* transformants was completed. Each event had 5 replicates. Although volume index was variable among the transgenic events, no obvious correlation was observed between volume index and *PtDDM1* expression (Fig. 2).

- An in vitro callus and shoot regeneration experiment included 21 transgenic events and the 717 control. For each event, 25 leaf or stem explants were collected and put on plates with or without 5-azacytidine. Each treatment had 4 replicate plates. Callus and shoot formation percentage was observed. An inverse relationship between the rate of callus growth and *PtDDM1* expression was observed, but only for stem explants without 5-azacytidine treatment (Fig 2B, 3).

2. **Prepare genomic DNA from tissues of the sequenced *Populus trichocarpa* clone and use commercially-available antibodies against 5-methylcytidine to isolate methylated DNA segments.**
3. **Sequence the methylated and enriched DNA on an Illumina GAII Genome Analyzer at the OSU CGRB and conduct bioinformatic and statistical studies to test the following hypotheses:**

   a. DNA and/or histone methylation differs between differentiation states.
   b. DNA and/or histone modifications are associated with RNA transcription levels.
   c. Chromatin states and gene expression are associated with small RNAs.
   d. Suppression of DNA methylation disturbs vegetative development and dormancy.

      - ***Methylated DNA immunoprecipitation and sequencing (MeDIP-seq) completed for 12 tissue types*** We previously reported our success in employing MeDIP-seq with

poplar. Deep sequencing has been completed for methylated genomic DNA from vegetative, reproductive and *in vitro* tissues using the Illumina GAIIx Genome Analyzer at Oregon State University's Center for Genome Research and Biocomputing. A non-immunoprecipitated ("input") control library was also sequenced. Numbers of reads from each tissue type that aligned to the reference genome with up to two mismatches are given below:

Tissue	Biological Reps	Platform	Lanes Sequenced	Total Reads (Illumina Yield)	Mapped reads	% mapped
Leaf	2	GA	4	70,264,008	13,524682	19.2%
Root	2	GA	5	81,017,743	9,284,891	11.5%
Xylem	1	GAII	5	33,649,277	12,278,525	36.5%
Phloem	2	GAII	4	38,358,479	22,737,382	59.3%
Female flower	2	GAII	5	63,861,351	17,212,810	27.0%
Male flower	2	GAII	5	78,702,151	24,870,095	31.6%
Spring bud	2	GAII	5	97,642,336	8,880,982	9.1%
Autumn bud	1	GAII	3	27,397,024	10,836,547	39.6%
Winter bud	1	GAII	3	26,301,421	9,734,237	37.0%
Internode explant	3	GAII	6	204,768,821	84,814,829	41.4%
Callus	3	GAII	6	214,470,025	48,813,758	22.8%
Regenerated internode	3	GAII	3	147,824,747	35,116,532	23.8%
		Total	**54**			

- ***Bioinformatic processing.*** Our previously-described bioinformatic processing pipeline was employed to generate summary genome methylation data for the poplar genome.

- ***Genome MeDIP-seq read coverage*.** Our MeDIP-seq coverage was variable among tissue types, covering ~40-80% of the 485 Mb poplar genome when both uniquely-mapping and redistributed k-mer repeat reads were included (Fig. 4). This coverage has provided a sufficient depth to allow us to identify methylated regions for each of the sampled tissue types.

- ***Gene model methylation patterns*.** When coverage data was averaged over all protein-coding genes, a pattern of higher coverage in the 2 Kb regions 5' and 3' of gene models, as well as higher coverage in the middle of the gene models than at the ends, is observed (Fig. 5). This pattern is consistent with what others have observed for poplar, as well as Arabidopsis and rice (Feng et al., 2010).

- ***Poplar Gbrowse updated for P. trichocarpa V2.2.*** We have developed an in-house version of Gbrowse, the widely-used viewer from the Genome Model Organism Database (GMOD)

project, to display genome scale data (http://poplar.cgrb.oregonstate.edu). We have now incorporated the complete set of *P. trichocarpa* gene models from the current JGI reference genome version, 2.2 (http://www.phytozome.net/poplar.php). Data from each tissue type is represented as a track that may be independently turned on or off by users, facilitating data viewing. Tissue data may be viewed on a per-nucleotide scale or on a 1 kb window RPKM scale. Additional tracks showing G/C content and transposable element categories are available, and users may also upload their own custom tracks. All Illumina data generated by this project has been shared with Phytozome in order to be uploaded on the main poplar genome Gbrowse site, where it will be compatible with genome assembly version 3.

- ***Bisulfite sequencing of selected regions.*** Treatment of double-stranded DNA with sodium bisulfite converts non-methylated cytosine bases to uracils, while methylated cytosines are unaffected. This principle has been used to obtain single-base resolution of cytosine methylation in regions of interest as well as for whole genomes, including the *Arabidopsis* genome (Feng et al., 2010). We performed bisulfite treatment of poplar bud genomic DNA, followed by PCR and cloning of selected regions showing varying levels of methylation intensity from MeDIP-seq alignments. This method allows us to confirm our MeDIP results, test the accuracy of methods for identifying methylated regions, and also provides the ability to resolve single cytosines as methylated or unmethylated. Eight regions were cloned and sequenced from bisulfite-treated DNA from the three sampled bud types, as well as from an untreated control. Results indicated a strong correlation between percentage of methylated cytosines and both maximum per-nucleotide MeDIP-seq coverage and calculated RPKM.

- ***Tissue-specific methylation differences.*** We have examined differences in methylation patterning among tissues at the whole-chromosome scale, at the scale of 1kb tiled genome windows and, ultimately, at individual genes. We have done this work comparing different poplar tissues from mature trees, from trees at different dormancy stages (paradormancy, endodormancy, ecodormancy) (Figure 7), and *in vitro* poplar tissues representing dedifferentiated and redifferentiated cellular states (Figure 8). We generated chromosome graphics highlighting regions of gross difference in mapped read density among tissue types. For methylated gene identification, we used two different approaches: The first approach set a false discovery rate (FDR) of 1% using comparisons of MeDIP-seq data to input; the second used a cumulative Poisson probability distribution (CPPD) to call methylated 1 kb windows across the genome. Using both approaches, we have generated slates of hundreds to thousands of gene models that are potentially methylated at particular levels and/or in particular tissue types.

  We expected transposable elements to be heavily methylated in all sampled *P. trichocarpa* tissues, since these features are generally methylated in plant genomes. Using an available catalog of poplar transposable elements (RepPop, Zhou 2009), we looked for enrichment of particular categories of transposable elements in our MeDIP-seq data, and for any

tissue-level differences in transposable element methylation, by comparing tallies of particular elements in our data to tallies of these elements in the genome as a whole (Fig. 9). While most transposon categories were overrepresented (methylated) as expected, the "simple repeat" category and two LINE element subcategories (LINE CR1, LINE L2) were underrepresented in the immunoprecipitated fractions of all tissues relative to their overall representation in the genome. In addition, three other element categories were underrepresented in male flowers, but overrepresented in all other tissues.

- ***Comparison of methylation with gene expression.*** In order to learn about potential biological function of methylation in *P. trichocarpa*, we used archival expression microarray data to look at the range of expression of genes that were called methylated by our methods. We compared expression of groups of genes that were methylated at promoters only, at gene bodies only, and at both promoters and gene bodies (Fig. 10). Genes with methylation at either feature had lower expression than unmethylated genes. Surprisingly, genes with body methylation had lower median expression than those with promoter methylation. Male catkins were different from other tissue types in that a higher number of genes were body-methylated, but these genes showed higher expression than body-methylated genes in other tissues.

- In addition to the planned grant objectives, PI Strauss collaborated with the Jacobsen laboratory at UCLA to add poplar to a genome-scale, sequence level analysis of DNA methylation in a wide variety of organisms (Feng et al. 2010). The work showed large differences between plants and animals, broad similarities among the three plant species included (rice, poplar, Arabidopsis), but also significant differences in the repetitive portions of the genome among the plant species. This is a likely consequence of poplar's extensive genome duplications, larger genome size, and thus its weaker genome v. 1.1 assembly compared to the other plant species analyzed.

Summary. We are now capitalizing on the foundational resources we established during the first two years of the project, and are currently producing manuscripts for publication. We have sequenced methylated DNA from all sampled tissue types, and used an additional technique, bisulfite sequencing, to confirm and further refine observations of methylation patterns. Using two different statistical analyses, we have generated lists of gene models that vary in cytosine methylation in different tissue types. We have determined levels of similarity and difference in gene methylation among tissues. We have produced chromosome-level views of gross methylation differences among tissue types and examined underlying loci. We have compared methylation levels of slates of genes to their expression levels in the studied tissues. We have found extensive differential methylation among poplar tissues, and lower expression of genes with promoter and/or gene body methylation. Finally, we have completed basic phenotypic and *PtDDM1* expression analyses of a population of *PtDDM1* transgenics. We have completed analysis of the *in vitro* tissues, and comparison of the methylome with the transcriptome from the three bud dormancy stages.

Publications and presentations

Kelly J. Vining, Kyle R. Pomraning, Larry J. Wilhelm, Henry D. Priest, Matteo Pellegrini Todd C. Mockler, Michael Freitag and Steven H. Strauss. **Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression**. (*BMC Genomics* 2012, **13**:27 (17 January 2012)).

Kelly Vining, Kyle R. Pomraning, Larry Wilhelm, Henry D. Priest, Cathleen MaYanming Di, Todd Mockler, Michael Freitag and Steven Strauss. **Methylome reorganization during in vitro dedifferentiation and regeneration of Populus trichocarpa** (Manuscript in submitted).

Kelly Vining, Kyle R. Pomraning, Larry J. Wilhelm, David Horvath, Yanming Di, Todd C. Mockler, Michael Freitag, Steven H. Strauss. **Methylome and transcriptome changes during seasonal dormancy in *Populus trichocarpa***. (Manuscript in preparation).

Kelly Vining. **Methylome Changes During In Vitro Regeneration of *Populus trichocarpa***. Forest Tree Workshop, Plant and Animal Genome XXI. San Diego, CA, USA. January 2013. (talk).

Kelly Vining. **Methylome-Transcriptome Interplay During Acquisition and Release From Seasonal Dormancy In *Populus trichocarpa***. Dormancy Workshop, Plant and Animal Genome XXI. San Diego, CA, USA. January 2013. (talk).

Kelly Vining, Kyle R. Pomraning, Larry Wilhelm, Henry D. Priest, Cathleen Ma, Ruoqing Zhu, Elizabeth Etherington, Matteo Pellegrini Todd Mockler, Michael Freitag and Steven Strauss. **Developmental variation in DNA methylation in poplar (*Populus trichocarpa*)**. IUFRO Tree Biotechnology Conference 2011. Arraial d'Ajuda, Bahia, Brazil. June 2011 (poster).

Steve Strauss. **Methylation in the poplar genome**. 26th New Phytologist Symposium: Bioenergy Trees, INRA-Nancy, France. May 2011 (talk).

Steve Strauss. **Genome and Developmental Variation in DNA Methylation in Poplar**. DOE Subcontractors Meeting, Crystal City, VA. April 2011 (talk).

Steve Strauss. **Genome-scale gene expression and DNA methylation during onset and release from vegetative dormancy in field-grown poplar**. Dormancy Workshop, Plant and Animal Genome XIX. San Diego, CA, USA. January 2011 (talk).

Steven Strauss, Kelly Vining, Kyle R. Pomraning, Larry Wilhelm, Palitha D. Dharmawardhana, Henry D. Priest, Olga Shevchenko, Cathleen Ma, Ruoqing Zhu, Elizabeth Etherington, Todd Mockler, and Michael Freitag. **Exploring the developmental epigenome of *Populus trichocarpa***. International Poplar Symposium. Orvieto, Italy. September 2010 (poster).

Steven Strauss, Kelly Vining, Kyle R. Pomraning, Larry Wilhelm, Palitha D. Dharmawardhana, Henry D. Priest, Olga Shevchenko, Cathleen Ma, Ruoqing Zhu, Elizabeth Etherington, Todd

Mockler, and Michael Freitag. **Exploring the developmental epigenome of *Populus trichocarpa***. 14th International Biotechnology Symposium and Exhibition - Biotechnology for the Sustainability of Human Society. Palacongressi, Rimini, Italy. September 2010 (poster).

Kelly Vining, Kyle R. Pomraning, Larry Wilhelm, Henry D. Priest, Peter Dolan, Cathleen Ma, Ruoqing Zhu, Elizabeth Etherington, Todd Mockler, Michael Freitag and Steven Strauss. **Genomic analysis of cytosine methylation in Populus trichocarpa tissues from differing developmental stages**. American Society of Plant Biologists, Montreal, Quebec, Canada. August 2010 (poster).

Steve Strauss. **DNA methylome of Populus trichocarpa during tree development**. FoResTTraC Workshop on Epigenetic Response to Climate Change, Aranjuez-Madrid, Spain, October 2010 (talk).

Project URLs

Project home page: http://poplar-epigenomics.cgrb.oregonstate.edu/
Genome browser: http://poplar.cgrb.oregonstate.edu/

References:

Suhua Feng, Shawn J. Cokus, Xiaoyu Zhang, Pao-Yang Chen, Magnolia Bostick, Mary G. Goll, Jonathan Hetzel, Jayati Jain, Steven H. Strauss, Marnie E. Halpern, Chinweike Ukomadu, Kirsten C. Sadler, Sriharsa Pradhan, Matteo Pellegrini, and Steven E. Jacobsen. (2010) Conservation and divergence of methylation patterning in plants and animals. Proc. Nat. Acad. Sci. U. S. A., 107:8689–8694

He, G. a. (2010). Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *The Plant Cell Online* , 17.

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

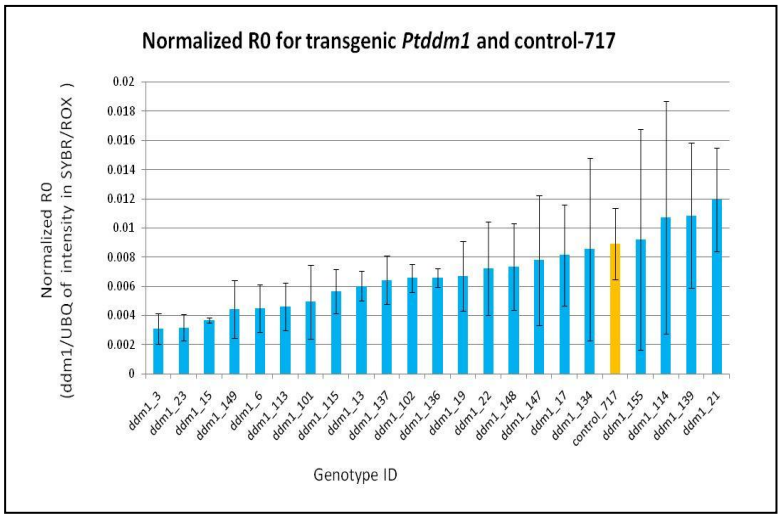

Figure 1. *PtDDM1* expression range in transformants determined by reverse transcription-qPCR. Each blue bar represents one transgenic event. The non-transformed control is represented by the yellow bar.

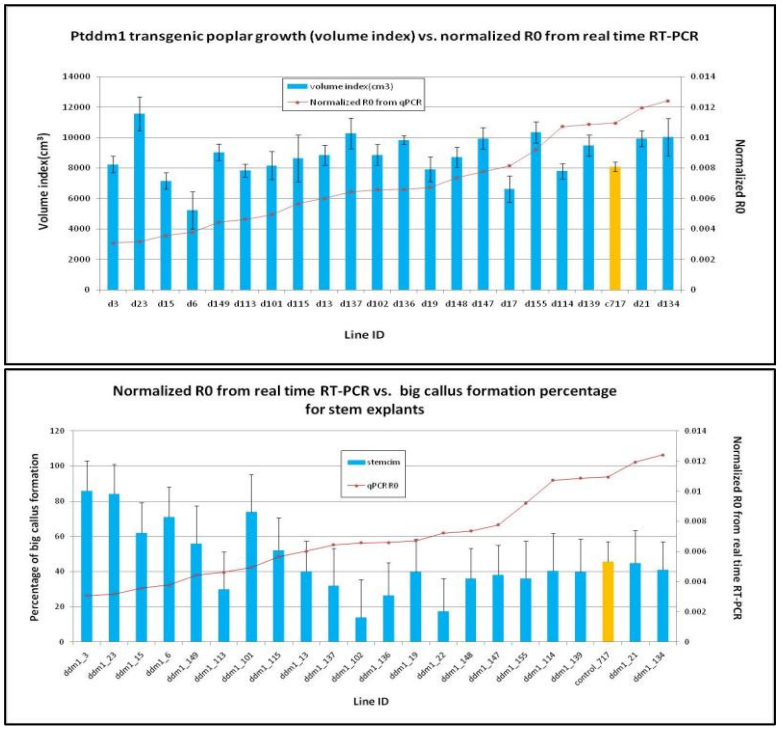

Figure 2. Comparison of *PtDDM1* expression with stem volume, *in vitro* callus size. (Top) Volume index of each transgenic line in the greenhouse shows no obvious correlation with *PtDDM1* expression. (Bottom) Callus size in tissue culture shows an inverse relation with *PtDDM1* expression.

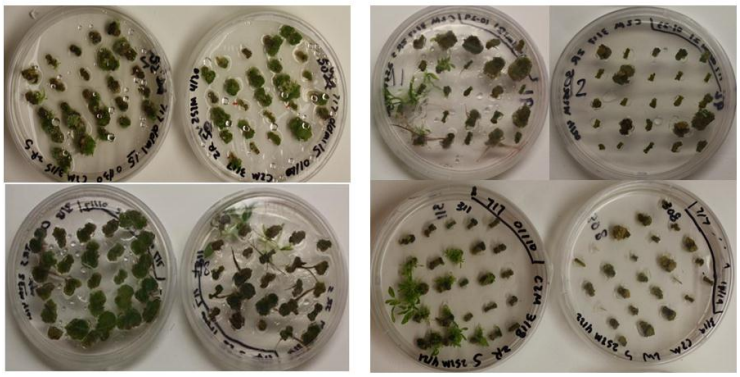

Figure 3. More large calli form in tissue culture when *PtDDM1* expression is lower.

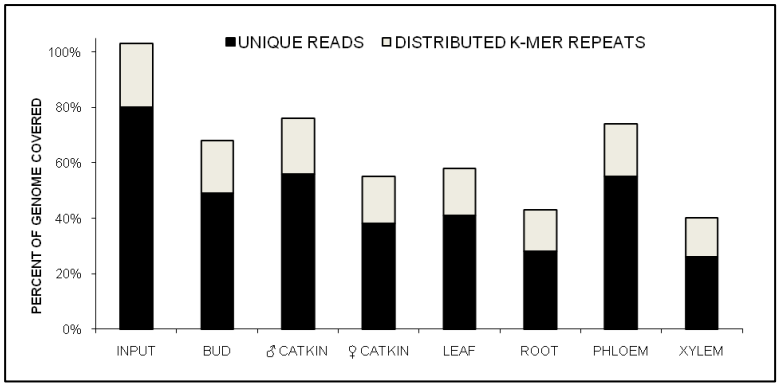

Figure 4. Overall MeDIP-seq genome coverage shown for each tissue type, as well as for non-immunoprecipitated control ("input"). The relatively higher coverage of input suggests that the MeDIP successfully enriched for methylated genomic regions.

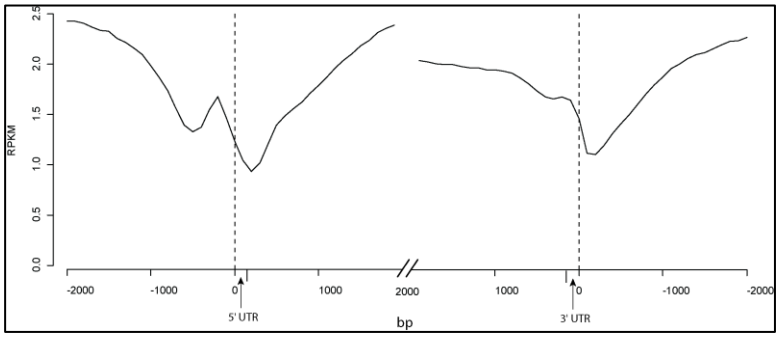

Figure 5. Methylation is lower at 5', 3' ends of genes. For MeDIP reads pooled across all tissues, the average reads per kilobase of target sequence per million input reads (RPKM) was plotted in tandem 100-bp windows surrounding gene models. The upstream and downstream regions were divided into windows of 50nt (5% of region). The gene model windows were 5% of gene model length. The average RPKM value for all such windows across the genome were plotted as shown. Note the reduced methylation by gene borders, and the increased methylation upstream and downstream of genes.

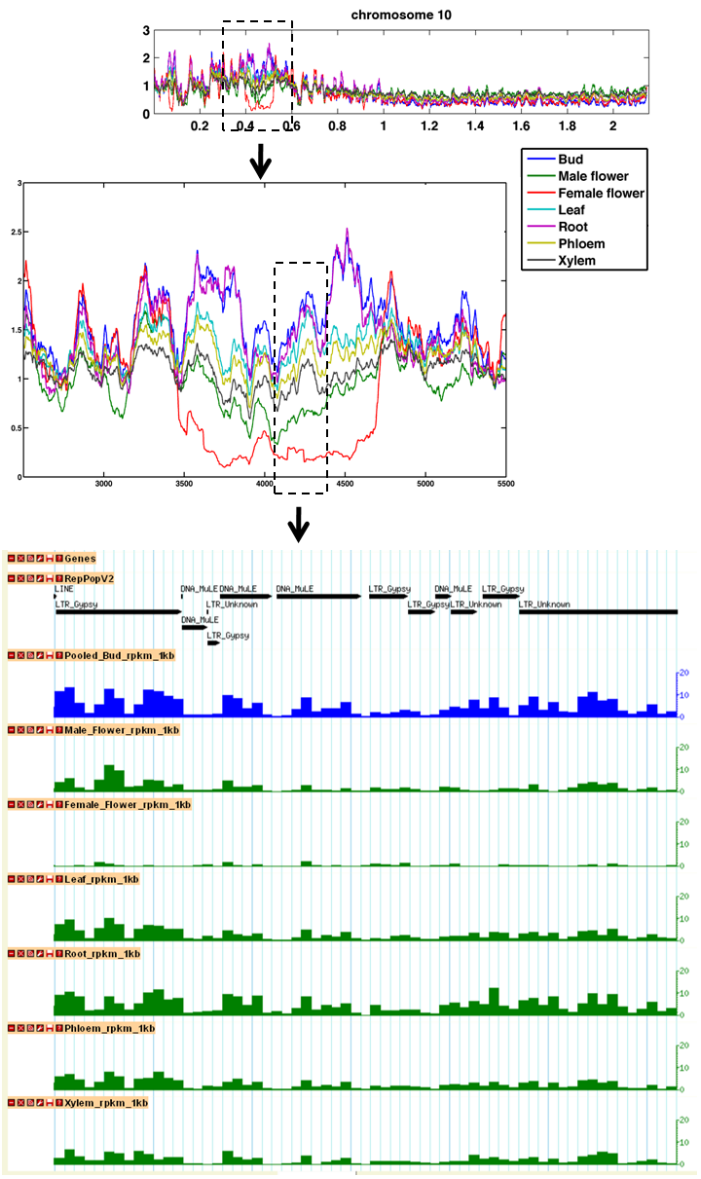

Figure 6. Chromosome-scale methylation patterns differ among tissue types. MeDIP-seq reads aligned to *P. trichocarpa* chromosome 10. Read counts are represented on the y axis, with one line per tissue type. Zooming in on a region with differential methylation among tissues shows that this region contains few genes and is rich in retroelements.

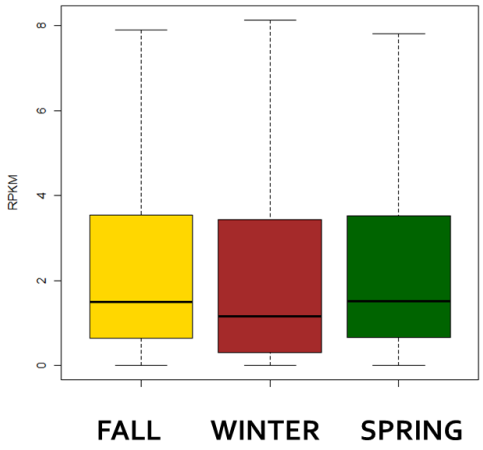

Figure 7. Median genome methylation decreases during annual dormancy. DNA methylation is shown for fall, winter and spring. Lines in boxes represent medians. Boxes contain 25th-75th percentile. Whiskers represent 1.5x the interquartile range.

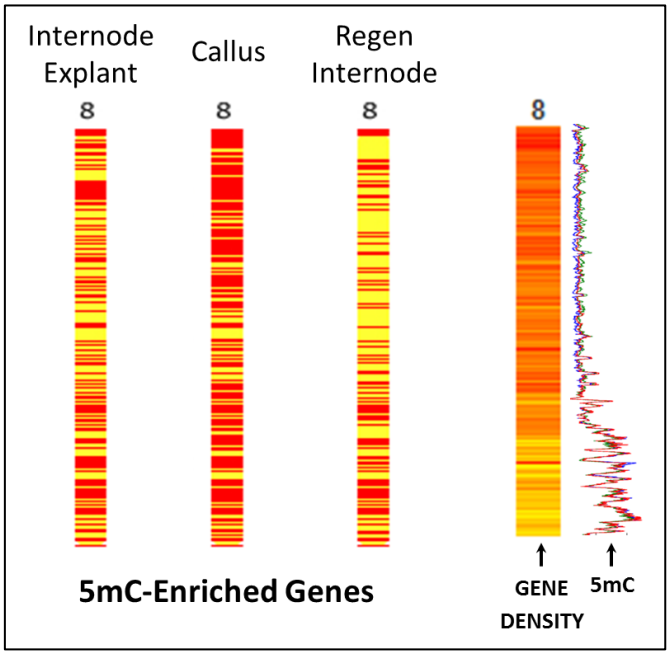

Figure 8. Gene body methylation increases in dedifferentiated callus, and this increase is visible on a chromosome scale in regions of high gene density. Left: 5mC-enriched genes are mapped (red) on P. trichocarpa chromosome 8 for internode explants, dedifferentiated callus tissue, and regenerated internode tissue. Right: Gene density from low (yellow) to high (red) is shown next to overall chromosome methylation signal.

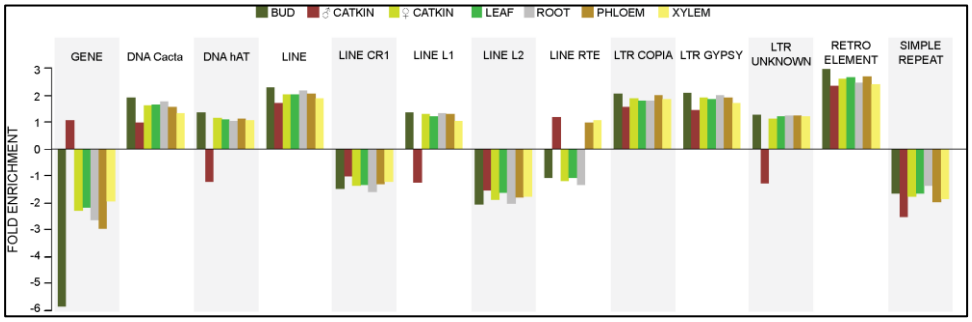

Figure 9. Male flowers (catkins) differ from other tissues in enrichment of some repeat element classes in methylated regions. The frequency of elements from each category among the methylated regions of each sample is compared to the frequency of that class of element in the entire genome. Element genome coordinates and annotations were obtained from the RepPop database.

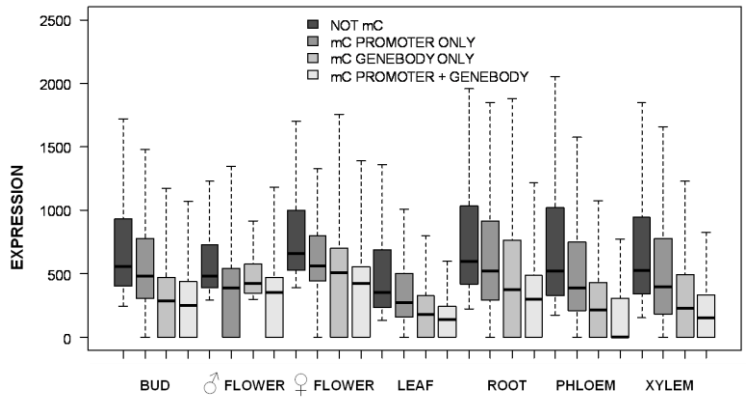

Figure 10. Expression of genes with methylated promoters is higher than expression of genes with methylated gene bodies in all tissues except male flowers. Box plot showing average expression of unmethylated genes, genes with methylated promoters, and genes with methylated gene bodies was compared for each tissue type using gene expression data from an available Nimblegen microarray. Each box encloses the middle 50% of the distribution (25th percentile – 75th percentile, or interquartile range (IQR)). Lines in boxes mark medians. Lines extending from boxes mark minimum and maximum values that fall within 1.5 times the IQR.