

RESEARCH ARTICLE

Open Access



Genome-wide identification of histone methylation (H3K9_{me2}) and acetylation (H4K12_{ac}) marks in two ecotypes of switchgrass (*Panicum virgatum* L.)

Vasudevan Ayyappan^{1†}, Venkateswara R. Sripathi^{2†}, Venu (Kal) Kalavacharla^{1,3*†}, Malay C. Saha⁴, Jyothi Thimmapuram⁵, Ketaki P. Bhide⁵ and Elizabeth Fiedler¹

Abstract

Background: Histone modifications play a significant role in the regulation of transcription and various biological processes, such as development and regeneration. Though a few genomic (including DNA methylation patterns) and transcriptomic studies are currently available in switchgrass, the genome-wide distribution of histone modifications has not yet been studied to help elucidate gene regulation and its application to switchgrass improvement.

Results: This study provides a comprehensive epigenomic analyses of two contrasting switchgrass ecotypes, lowland (AP13) and upland (VS16), by employing chromatin immunoprecipitation sequencing (ChIP-Seq) with two histone marks (suppressive- H3K9_{me2} and active- H4K12_{ac}). In this study, most of the histone binding was in non-genic regions, and the highest enrichment was seen between 0 and 2 kb regions from the transcriptional start site (TSS). Considering the economic importance and potential of switchgrass as a bioenergy crop, we focused on genes, transcription factors (TFs), and pathways that were associated with C4-photosynthesis, biomass, biofuel production, biotic stresses, and abiotic stresses. Using quantitative real-time PCR (qPCR) the relative expression of five genes selected from the phenylpropanoid-monolignol pathway showed preferential binding of acetylation marks in AP13 rather than in VS16.

Conclusions: The genome-wide histone modifications reported here can be utilized in understanding the regulation of genes important in the phenylpropanoid–monolignol biosynthesis pathway, which in turn, may help understand the recalcitrance associated with conversion of biomass to biofuel, a major roadblock in utilizing lignocellulosic feedstocks.

Keywords: Switchgrass, Epigenome, ChIP-Seq, Histone modifications, Differential binding, Phenylpropanoid pathway, And monolignols

Background

In eukaryotes, gene expression is a complex, coordinated process orchestrated by transcription factors (TFs) and chromatin proteins. As a result, chromosomal DNA is either compacted or relaxed to repress or activate the

transcription process, respectively. At the whole-genome level, genetic and epigenetic factors together regulate gene expression. The accessibility of DNA sequences and chromatin sites associated with nucleosome packaging is known to determine transcription efficiency. Epigenetic factors such as DNA methylation, histone modifications, and chromatin remodelers play a significant role in the regulation of gene expression and various biological processes, such as development [1], regeneration [2], and oncogenesis [3]. In plants, with recent advances, genome-wide epigenomic analysis methods are increasingly supporting the role of chromatin modifications in gene expression [4].

* Correspondence: vkavacharla@desu.edu

[†]Vasudevan Ayyappan, Venkateswara R. Sripathi and Venu (Kal) Kalavacharla contributed equally to this work.

¹Molecular Genetics and Epigenomics Laboratory, College of Agriculture and Related Sciences, Delaware State University, Dover, DE, USA

³Center for Integrated Biological and Environmental Research, Delaware State University, Dover, DE, USA

Full list of author information is available at the end of the article



Transcription factors are important in gene regulation; they bind to cis-regulatory sequences in the promoter region of their target genes to fine-tune gene expression [5]. In plants, transcription factors and histone modifications demonstrably play a role in development, vernalization, and response to biotic and abiotic stressors. [6–10]. The application of high-throughput technologies such as chromatin immunoprecipitation followed by genome-wide sequencing (ChIP-Seq) provides an opportunity to identify binding patterns for a protein(s) of interest in the entire genome. Genome-wide association and correlation of modifications, such as histone methylation and acetylation with biotic and abiotic stress responses in plants have been reported [8, 11–13].

Switchgrass (*Panicum virgatum* L.) is a native perennial and warm-season grass, widely used for biofuel, livestock feed, erosion control, and wildlife habitat. Carbon sequestration by switchgrass, a C4 plant with high cellulose conversion rates and drought tolerance, facilitates its use as a biofuel directly or by ethanolic conversion, making it a valuable, natural renewable energy resource [14]. To date, studies in switchgrass have pertained to molecular breeding [15, 16], genomic [17, 18] and transcriptomic [19, 20] analyses, and response to biotic and abiotic [21, 22] stresses. Using Bisulfite sequencing (BS-Seq) and Methylated DNA immunoprecipitation sequencing (MeDIP-Seq), similar DNA methylation patterns between AP13 and VS6 have been revealed, and the methylation levels were higher in Transposable elements (TEs) when compared to genic regions [23]. The recent bisulfite sequencing study suggested small interfering RNAs (siRNAs) positively regulate DNA methylation and thereby interferes with gene and long non-coding RNAs (lncRNAs) expression in switchgrass [24]. However, the genome-wide distribution of histone modifications in switchgrass has not been reported. The histone marks (a repressive mark, H3K9_{me2}, and an activation mark, H4K12_{ac}) utilized in this study have been studied and reported previously in several plant species [7, 8, 25, 26]. Methylation of histone H3K9 is preferentially localized to heterochromatin in *Arabidopsis* [25], and its profiling has been reported in other plants, for instance, rice [7], maize [26], and recently in the common bean [8]. On the other hand, H4K12_{ac} is primarily localized in active coding regions of the genome and therefore creates binding sites for regulatory factors to promote transcription [8].

Switchgrass has two main ecotypes, lowland and upland, with distinct morphological and genetic characteristics and specific geographical niches [27]. Two important switchgrass genotypes (AP13 and VS16) were selected to represent each of these two contrasting ecotypes. Among these, AP13, a lowland ecotype, was used for the developing the switchgrass reference

genome [28]. A female parent AP13 and male parent VS16 (an upland ecotype) were utilized for the construction of a genetic linkage map [18] and the identification of quantitative trait loci (QTL) associated with biomass yield and traits related to recalcitrance [22, 29]. Due to their adaptability and ability to withstand selection pressure from diverse environmental stressors, the inherent genetic potential of these genotypes at the level of gene expression has been recently explored [30]. However, little is known about the epigenetic regulation of gene expression associated with these genotypes. The H3K9_{me2} and H4K12_{ac} marks selected in this study regulate the closed and open chromatin states, respectively, during transcription. In this study, we report the first genome-wide profiling of histone methylation H3K9_{me2} and histone acetylation H4K12_{ac} marks by using ChIP-Seq in upland (VS16) and lowland (AP13) switchgrass ecotypes.

Methods

Plant materials and sample collection

The two contrasting, heterozygous, tetraploid genotypes ($2n = 4x = 36$) utilized in this study (AP13 and VS16) were derived from two ecotypes of switchgrass: the lowland cultivar, Alamo and the upland cultivar, Summer, respectively. The leaf materials were obtained from our collaborator and co-author on this manuscript Malay Saha at the Noble Research Institute (Ardmore, OK). Both AP13 and VS16 actively grow in summer, and hence the conditions (29/22 °C day/night temperatures and a 16 h photoperiod) considered in the greenhouses of the Noble Research Institute were uniform. Fully expanded leaves were randomly collected and pooled from five plants of one-month-old clonal ramets, cryopreserved in liquid nitrogen, and stored at -80°C . This experiment included two genotypes (AP13 and VS16), two marks (H3K9_{me2} and H4K12_{ac}), two conditions (input control and actual sample), and three replicates resulting in (2 X 2 X 2 X 3) a total of 24 samples. Of significance is that this current study on global ChIP-seq analyses utilized the same leaf samples that were collected, stored (-80°C), and used for RNA-seq analyses in our previous studies [30]. Since our goal is to understand the genome-wide histone modifications in regulating global gene expression rather than a specific focus on pathways related to lignin formation, we have used leaves instead of stems.

Isolation and immunoprecipitation of chromatin

All the lyophilized leaf samples were ground to a fine powder in liquid nitrogen, collected in frozen 50 mL falcon tubes (Corning Inc., Corning, NY). Next, 25 ml (mL) of cold nuclear isolation buffer containing 1% formaldehyde, and 20 μl (μl) of proteinase inhibitor (Thermo Scientific, IL, USA, Product #87786) were added. The

samples were incubated for 5 min at room temperature. Glycine (2 M) was used to terminate the crosslinking reaction. The lysate was filtered before nuclei pelletization at 3000 g for 10 min at 4 °C. The pellet was resuspended in cold nuclear isolation buffer (no formaldehyde included). An equal volume of 15% Percoll solution was added to the lysate, and the nuclei were pelleted again at 3000 g for 5 min at 4 °C. The pellet was then mixed in nuclear lysis buffer. This lysate was then sonicated using a Soniprep 150 (MSE UK Ltd., London, UK) for 6 pulses with 15 s between each pulse. The sheared DNA fragments collected after sonication ranged between 100 and 350 bp when visualized by agarose gel electrophoresis. ChIP dilution buffer was then added to the nuclear lysate to attain 10X dilution, and chromatin samples were then divided equally into two tubes. One of the samples was incubated with rabbit polyclonal antibody in a 20:1 mixture that specifically adhered to the N-terminus of histone subunit H4 acetylated at lysine 12 (H4K12_{ac}) [31]. Also, the chromatin sample was incubated with a mouse antibody in a 20:1 mixture, which adhered precisely to the epitope region of histone H3 di-methylated at lysine 9 (H3K9_{me2}) [32]. Twenty microliters of Pierce Protein A/G magnetic beads (ThermoFisher Scientific, Grand Island, NY) were added to the nuclear lysate. The samples were then incubated overnight at 4 °C. The chromatin samples that had no antibodies served as positive controls. Later, the centrifuged and pelleted magnetic beads were washed thrice with 100 µl ChIP dilution buffer. The magnetic beads bound to chromatin and antibody complex were then eluted with freshly prepared elution buffer. To terminate the crosslinking, 20 µl of 5 M NaCl was added and subsequently incubated at 65 °C overnight. After this, 5 µl of 2 M Proteinase-K (Thermo Scientific) was added and incubated for 2 h (h) at 45 °C. DNA was then collected by phenol-chloroform extraction followed by ethanol precipitation. The ChIP-DNA was resuspended in 20 µl of TE buffer. After verification by ChIP-PCR, each of these purified DNA samples were then used to generate a ChIP-Seq library using Illumina HiSeqTM 2500 (Illumina Inc., San Diego, CA). Libraries were sequenced at the Delaware Biotechnology Institute (DBI, Newark, DE).

Library construction and sequencing

The quality, purity, and size of the immunoprecipitated DNA samples were determined by an AATI Fragment Analyzer (Advanced Analytical, Ames, IA). The ChIP-Seq libraries (101 bp paired-end) were prepared by using Illumina TruSeq ChIP Sample Preparation kit (IP-202-1012; Illumina Inc., San Diego, CA). The NGS data generated in this study was submitted to the SRA section of NCBI with the project number PRJNA373877 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA373877>).

ChIP-Seq data analysis workflow

Sequence quality was assessed using FastQC (v0.11.2) [33] and quality trimming was carried out using the FASTX toolkit (v0.0.13) (http://hannonlab.cshl.edu/fastx_toolkit) to filter bases with less than the Phred33 score of 30 and to remove sequences with less than 50 bases. The quality trimmed reads were mapped to the AP13 v1.1 reference genome, *Panicum virgatum*, (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvirgatum), using bowtie2 (v2.2.6) [34] with default parameters, except the number of mismatches was set to 1. Peaks marked by H3K9_{me2} and H4K12_{ac} were identified using Spatial Clustering for Identification of ChIP-Enriched Regions (SICER; v1.1) [35] for both genotypes, using the following parameters: redundancy threshold of 10; window size of 200; gap size of 600; effective genome size as a fraction of reference genome of 0.7; and false discovery rate (FDR) threshold controlling significance at 0.01, for each replicate/sample with respect to their input control. The set of peaks present in at least two replicates was used for further analysis. A gene was considered methylated (H3K9_{me2}-marked) or acetylated (H4K12_{ac}-marked) only if it was overlapping (based on known annotated genes) with peak coordinates by at least one base. To study the pattern of histone binding in different regions of the genome, we also calculated distribution of peaks identified in different genomic regions (5' UTR, exon, intron, 3' UTR, upstream, downstream, and non-genic regions using PAVIS (v1.8) [36] based on the annotation of known genes. For each peak, the distance from the peak to the TSS was determined and plotted using annotations from Phytozome for the AP13 v1.1 reference genome. The percentage of genes for each modification was calculated as the total number of genes with peaks divided by the number of genes for each modification and multiplied by 100. As routinely used in expression profiling of genes associated with histone peaks in other biological systems, once we obtained the nearest annotated genes, we performed Gene Ontology (GO) analysis to assign gene functions such as biological processes, molecular functions, and cellular components, before conducting Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Among thousands of genes that were differentially marked or enriched using ChIP-Seq analysis, five genes (1–5) were selected for functional validation based on: i) their role in phenylpropanoid-monolignol pathway; ii) fold change (Log₂FC) > 2; and iii) FDR < 0.01.

ChIP-quantitative real time (RT)-PCR (qPCR)

To quantitatively measure the amplification of ChIP-DNA, quantitative real-time (RT)-PCR (qPCR) validation was performed with an ABI 7500 real-time PCR

(Applied Biosystems, Foster City, CA). The five genes selected here belong to the phenylpropanoid pathway that represents the active (H4K12_{ac}) and inactive (H3K9_{me2}) modifications from ChIP-Seq analysis. The primers were designed using the TaqMan qPCR primer design tool (GenScript USA Inc., Piscataway, NJ), and the primer pairs used to amplify the ChIP-enriched regions are presented in Additional file 1: Table S1. qPCR was carried out in a 25 μ L reaction that contained 10 ng of immunoprecipitated DNA, 10 μ M each of forward and reverse primers, and 12.5 μ L of SYBR Green PCR Master Mix (Germantown, MD). The qPCR reaction settings were: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 65 °C for 1 min.

We used three replicates to test the repeatability of qPCR validations. The universal cons7 primer was used as a constitutive control to determine normalized expression for all samples. Primer efficiency was determined by using 2- $\Delta\Delta$ CT method [37], where $\Delta\Delta$ CT (CT of genes - CT of cons7) = tissue to be observed - leaf tissue (CT of genes - CT of cons7). Minitab 17 [38] was used to analyze the normalized CT values ($\Delta\Delta$ CT), and the expression results were presented as mean \pm SE. To test the multiple comparisons between the sample means, one-way ANOVA was conducted on normalized CT values collected from qPCR analysis.

Results

Utilizing AP13 v1.1 as a reference [28], 138 million and 114 million clean 101 bp paired-end reads were obtained for H3K9_{me2} and H4K12_{ac} modifications for AP13, respectively. We also obtained 120 million and 108 million clean 101 bp paired-end reads marked respectively with H3K9_{me2} and H4K12_{ac} for VS16. Average mapping rates of 82.32% (H3K9_{me2}) and 81.20% (H4K12_{ac}) in AP13 and 73.42% (H3K9_{me2}) and 68.43% (H4K12_{ac}) in VS16 (Table 1), respectively, were observed.

Distribution of histone modifications

Overall distributions of H3K9_{me2} and H4K12_{ac} peaks were obtained using reads that mapped to all chromosomes and contigs greater than 5 kb, and contigs between 2 kb and 5 kb that have predicted genes while calling peaks (Additional file 2: Figure S1). A representative graph that shows the distribution of histone marks on phenylalanine ammonia lyase (PAL) gene in chromosome 7b was visualized in AP13 and VS16 for H3K9_{me2} and H4K12_{ac} modifications, using input (AP13 v1.1) as a positive control through the Integrative Genomics Viewer (IGV) [39] (Fig. 1). For H3K9_{me2} and H4K12_{ac} in both AP13 and VS16, numerous peaks were identified in non-genic regions (over 40%), and less than 40% of peaks were in genic regions (Fig. 2). The range of non-genic region for AP13 for H3K9_{me2} (399 bp - 27,799 bp)

and H4K12_{ac} (399 bp - 32,399 bp) modifications and VS16 for H3K9_{me2} (199-34,199) and H4K12_{ac} (199 bp - 29,399 bp) modifications has been observed in this study.

Genes related to peaks

By analyzing genes within 10 kb from the peaks, 24,163 and 21,918 genes were identified in AP13 and VS16, respectively. Of these, 5852 (AP13) and 3918 (VS16) gene-associated peaks were marked with H3K9_{me2}, while 18,311 (AP13) and 18,000 (VS16) gene-associated peaks were marked with H4K12_{ac} (Table 2). In AP13 and VS16, the percentage of methylated peaks located on genic regions was 24.22 and 17.88%, respectively. Also, the percentage of acetylated peaks located on genic regions in AP13 and VS16 was 75.78 and 82.12%, respectively. Furthermore, 71.86 and 63.31% of these genes carried both of the modifications (H3K9_{me2} and H4K12_{ac}) in AP13 and VS16, respectively (Table 2).

Distances from peaks to TSSs

Based on nearest gene distribution using BEDTools (v2.21.0) [40], over 18,000 methylated and acetylated peaks spanned between 0 to +2000 bases from TSSs in AP13. In addition, over 2000 acetylated peaks spanned between -1 to -2000 bases from TSSs in AP13. Over 12,000 methylated and 18,000 acetylated peaks spanned between 0 to +2000 bases from TSSs in VS16. (Fig. 3). Most enrichment was seen in the region 2 kb upstream or downstream of a TSS. The enrichment levels from both upstream and downstream of 2 Kb from TSS was seen to gradually decrease in both modifications for both genotypes (Additional file 3: Figure S2A-D). To understand the relationship between histone modifications (H3K9_{me2} and H4K12_{ac}) and TF binding, the peak distances to nearest genes were binned into 5 kb intervals (Additional file 4: Figure S3A-D). Of 30 bins with a nearest gene distance of 15 kb, 26 contain at least 100 peaks for H3K9_{me2} modification in AP13 and VS16 (Additional file 4: Figure S3A and B). All 30 bins with a nearest gene distance of 15 kb for the H4K12_{ac} modification contain at least 100 peaks for both genotypes (Additional file 4: Figure S3C and D).

Epigenomic factors

The five major categories of epigenetic factors identified, along with the number of genes marked with H3K9_{me2} in both genotypes (AP13/VS16), were methyl groups (186/159), methyl CpG-Binding domains (73/69), acetyltransferase groups (36/39), oxidase/desaturase groups (14/16), and chromatin remodeling factors (21/41). Similarly, genes marked with H4K12_{ac} and their corresponding counts for methyl groups, methyl CpG-Binding domains, acetyltransferase groups, oxidase/desaturase groups, and chromatin remodeling factors were 196, 69,

Table 1 Statistics of ChIP sequencing of switchgrass genotypes AP13 and VS16 and their alignments to the reference genome AP13 v1.1 (*P. virgatum*)

| Sample | Total Reads | Quality Control Reads | % Reads Passing QC | Total Reads Mapped | % Reads Mapped |
|--------------|-------------|-----------------------|--------------------|--------------------|----------------|
| AP13H3 | 138,072,842 | 132,672,004 | 96.08 | 113,596,549 | 82.32 |
| AP13H3 Input | 37,218,084 | 36,739,685 | 98.71 | 34,351,304 | 92.29 |
| VS16H3 | 119,727,844 | 115,043,859 | 96.05 | 87,947,726 | 73.42 |
| VS16H3 Input | 44,430,086 | 43,948,490 | 98.91 | 35,837,847 | 80.66 |
| AP13H4 | 113,969,918 | 108,106,092 | 94.9 | 92,471,201 | 81.20 |
| AP13H4 Input | 42,305,540 | 41,843,913 | 98.9 | 39,138,138 | 92.51 |
| VS16H4 | 107,606,766 | 103,864,501 | 96.5 | 73,455,893 | 68.43 |
| VS16H4 Input | 48,779,324 | 48,779,324 | 99.02 | 39,773,125 | 81.53 |

39, 14 and 21 in AP13 and 135, 95, 68, 16 and 41 in VS16, respectively (Additional file 5: Figure S4). In this study, epigenomic factors, methyl CpG binding domains, and acetyltransferase groups were most dominant.

Pathways and genes associated with peaks

Although several pathways and pathway-related genes were identified in the study, we only selected 16 pathways that had greater than or equal to 10 genes associated with both H3K9_{me2} and H4K12_{ac} in each genotype (Table 3). Of these, secondary metabolite biosynthesis pathways were predominant, with over 100 genes methylated and acetylated in both AP13 and VS16 genotypes. Therefore, greater emphasis was placed on genes, pathways, and TFs

associated with biomass/biofuel production as well as biotic and abiotic stresses. Functions for the nearest genes to these peaks were assigned using KEGG pathway analysis, and most of these genes were functionally associated with cellular and metabolic processes, responses to stimuli, other biological regulation, etc. (Additional file 1: Tables S2, S3, S4 and S5).

C4 photosynthesis-related enzymes

An important intermediate in C4 carbon-fixation pathway, phosphoenol pyruvate (PEP) and two key pathway enzymes; i.e., phosphoenolpyruvate carboxylase (PEPC), that catalyzes PEP and carbonic anhydrase (CA), and two transporter genes viz., tonoplast dicarboxylate transporter

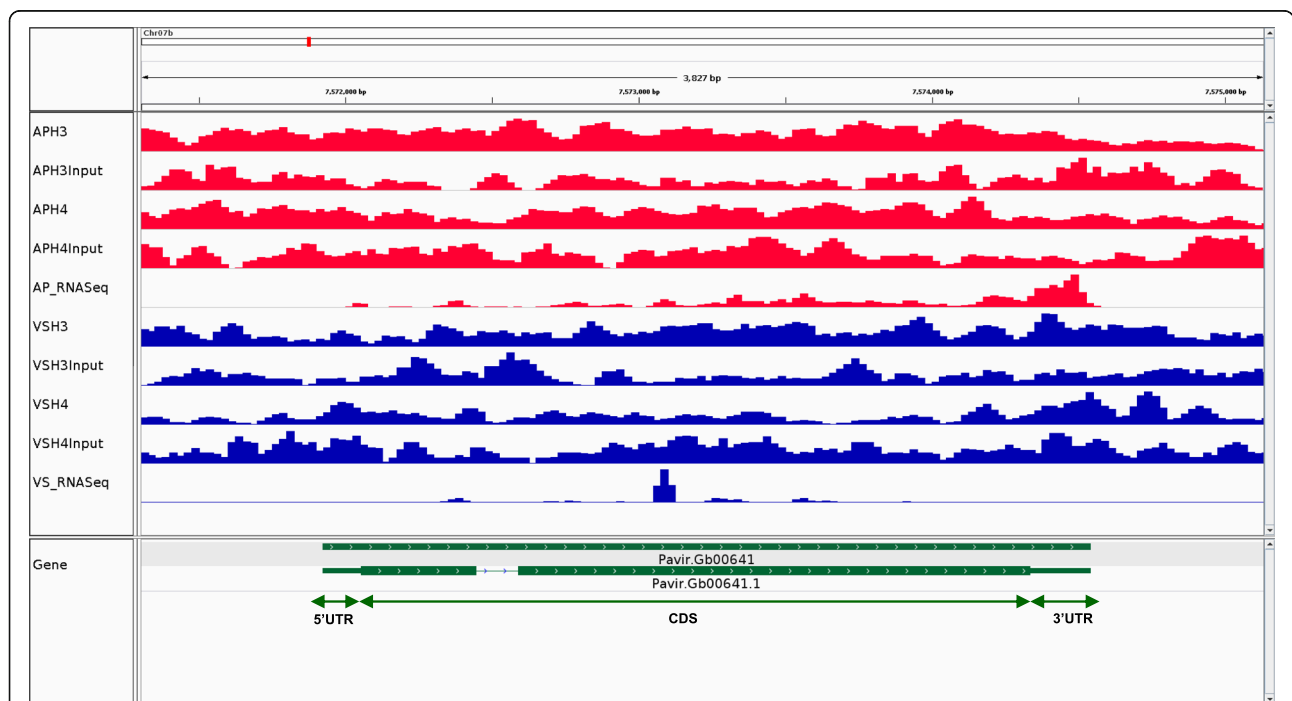
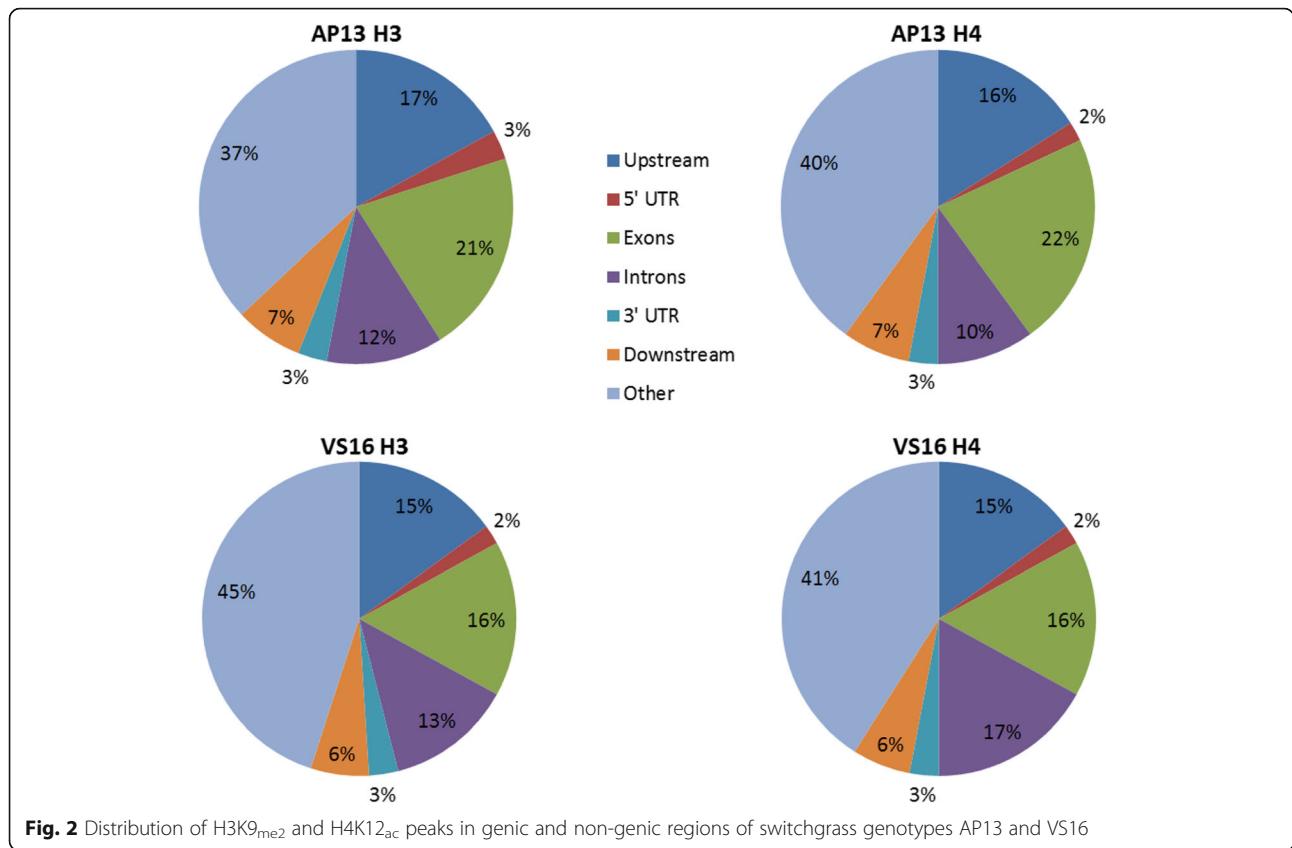


Fig. 1 Comparative visualization of a representative region of Phenylalanine Ammonia Lyase (PAL) gene on chromosome 7b in switchgrass genotypes AP13 (red) and VS16 (blue) with INPUT (positive control) as background for (A) H3K9_{me2} and (B) H4K12_{ac} modification using Integrative Genomics Viewer (IGV). R1-R3 represents the three biological replicates



(DIT) and bile acid sodium symporter/pyruvate transporter (BASS) were identified as marked in both AP13 and VS16 for both modifications, whereas thiophosphate isomerase was only identified in VS16 for both modifications (Table 4). The other genes, nicotinamide adenine dinucleotide phosphate (NADP)-dependent oxidoreductase, phosphoenolpyruvate carboxykinase (PEP-CK), and thiophosphate isomerase showed varied patterns in histone methylation and acetylation. For instance, in AP13, NADP-dependent oxidoreductase and PEP-CK were only methylated, while thiophosphate isomerase was neither methylated nor acetylated.

Photorespiration pathway-related genes

Marked by H3K_{9me2} and H4K12_{ac} in both the genotypes were eight key photorespiratory enzymes: 2-oxoglutarate (2-OG), glutamate (GLU), 3-phosphoglycerate (3-PGA),

ribulose 1,5-bisphosphate carboxylase/oxygenase (Ru-BisCo), glutamine (GLN), dicarboxylate transporter (DIT), catalase (CAT), and glutamate synthase (GS) (Table 4). Of these eight genes, 2-OG, GLU, and GLN showed binding affinity to H3K_{9me2} and H4K12_{ac} in both genotypes.

Phenylpropanoid pathway-related genes

Utilizing shikimate pathway intermediates, phenylpropanoid metabolism synthesizes a diverse array of secondary metabolites. Many genes and metabolites associated with phenylpropanoid pathways play a key role in feedstock quality and they were identified. Among these, the abundantly marked genes/metabolites in AP13 and VS16 for both H3K_{9me2} and H4K12_{ac} include hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase (HCT), cinnamyl alcohol dehydrogenase (CAD), phenylalanine

Table 2 Genes of AP13 and VS16 switchgrass genotypes associated with H3K_{9me2} and H4K12_{ac} regions¹

| Genotypes | Number of Genes associated with Peaks ¹ | Genes with H3K _{9me2} (%) | Genes with H4K12 _{ac} (%) | Genes with Both H3K _{9me2} and H4K12 _{ac} (%) |
|-----------|--|------------------------------------|------------------------------------|---|
| AP13 | 24,163 | 5852 (24.22) | 18,311 (75.78) | 17,364 (71.86) |
| VS16 | 21,918 | 3918 (17.88) | 18,000 (82.12) | 13,878 (63.31) |

¹H3K9 di-methylation and H4K12 acetylation regions include 10 kb upstream and downstream from the peak start position

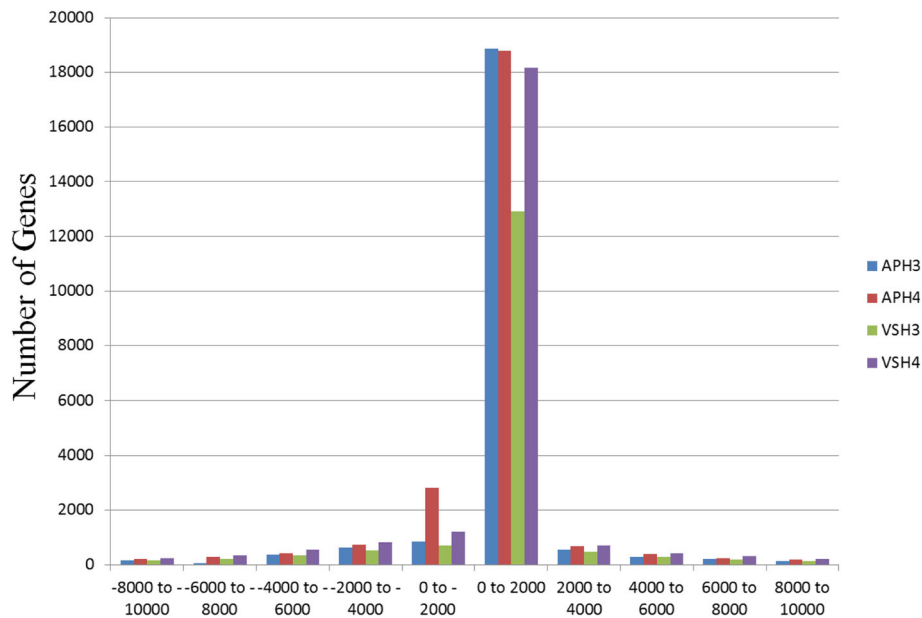


Fig. 3 Distance of peaks to transcriptional start sites (TSSs) for H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16

ammonia-lyase (PAL), 4-coumarate: CoA ligase (4CL), Cinnamate 4-hydroxylase (C4H), and caffeoyl-CoA 3-O-methyltransferase (CCoAMOT) (Table 4).

Peaks associated with stress-related genes

Even though we did not impart any stress to derive the reference epigenomes in this study, it was interesting

that many biotic and abiotic stress tolerance genes were prominent and inherently marked between AP13 and VS16 genotypes. In total, we identified histone marks on genes related to heat tolerance (176), drought tolerance (9), salinity tolerance (8), cold tolerance (3), flood tolerance (352), heavy metal tolerance (57), and disease resistance (277) in both genotypes (Additional file 1:

Table 3 KEGG pathway analysis for the genes with H3K9_{me2} and H4K12_{ac} associated-binding sites in switchgrass genotypes AP13 and VS16

| Pathway Description | Number of genes | | | |
|--|---------------------|------|---------------------|------|
| | H3K9 _{me2} | | H4K12 _{ac} | |
| | AP13 | VS16 | AP13 | VS16 |
| Metabolic pathways | 308 | 279 | 284 | 382 |
| Biosynthesis of secondary metabolites | 177 | 150 | 165 | 209 |
| Biosynthesis of antibiotics | 83 | 74 | 66 | 103 |
| Microbial metabolism in diverse environments | 67 | 65 | 50 | 72 |
| Carbon metabolism | 45 | 44 | 29 | 52 |
| Biosynthesis of amino acids | 42 | 37 | 38 | 54 |
| Plant hormone signal transduction | 29 | 24 | 28 | 28 |
| Starch and sucrose metabolism | 21 | 15 | 16 | 17 |
| RNA transport | 36 | 32 | 34 | 44 |
| Cell cycle | 16 | 22 | 20 | 26 |
| Plant-pathogen interaction | 15 | 13 | 14 | 15 |
| Carbon fixation | 15 | 14 | 10 | 10 |
| Phenylpropanoid biosynthesis | 14 | 11 | 12 | 11 |
| Fatty acid metabolism | 12 | 11 | 12 | 16 |
| Terpenoid backbone biosynthesis | 10 | 12 | 12 | 18 |
| Photosynthesis | 9 | 4 | 12 | 2 |

Table 4 List of C4-photosynthesis, photorespiration, and phenylpropanoid-monolignol pathway-related genes associated with H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16

| Genes Identified | Number of genes | | | |
|---|---------------------|------|---------------------|------|
| | H3K9 _{me2} | | H4K12 _{ac} | |
| | AP13 | VS16 | AP13 | VS16 |
| C4-Related Photosynthetic Genes | | | | |
| Tonoplast Dicarboxylate Transporter (DIT) | 17 | 1 | 4 | 2 |
| Phosphoenolpyruvate (PEP) | 19 | 15 | 20 | 0 |
| Phosphoenolpyruvate Carboxylase (PEPC) | 9 | 9 | 10 | 10 |
| Bile Acid Sodium Symporter/Pyruvate Transporter (BASS) | 4 | 2 | 3 | 1 |
| Carbonic Anhydrase (CA) | 6 | 5 | 9 | 2 |
| NADP-Dependent Oxidoreductase | 11 | 3 | 0 | 0 |
| Phosphoenolpyruvate Carboxykinase (PEP-CK) | 1 | 0 | 0 | 1 |
| Thiophosphate Isomerase | 0 | 2 | 0 | 4 |
| Photorespiration-Related Genes | | | | |
| 2-Oxoglutarate (2-OG) | 100 | 57 | 91 | 96 |
| Glutamate | 59 | 46 | 49 | 61 |
| 3-Phosphoglycerate (3-PGA) | 6 | 2 | 6 | 4 |
| Ribulose 1,5-bisphosphate carboxylase/Oxygenase (RuBisCo) | 4 | 3 | 7 | 5 |
| Glutamine (GLN) | 33 | 18 | 33 | 31 |
| Dicarboxylate Transporter (DIT) | 1 | 2 | 4 | 2 |
| Catalase (CAT) | 2 | 0 | 2 | 4 |
| Glutamate Synthase (GS) | 3 | 4 | 4 | 4 |
| Phenylpropanoid-Monolignol Pathway-Related Genes | | | | |
| Phenylalanine Ammonia Lyase (PAL) | 11 | 15 | 27 | 4 |
| 4-Coumarate:CoA Ligase (4CL) | 4 | 3 | 8 | 2 |
| Hydroxycinnamoyl CoA:Shikimate Hydroxycinnamoyl Transferase (HCT) | 32 | 17 | 32 | 25 |
| Cinnamyl alcohol dehydrogenase (CAD) | 33 | 26 | 42 | 31 |
| Caffeoyl-CoA 3-O-Methyl Transferase (CCoAOMT) | 6 | 1 | 5 | 1 |
| Cinnamate 4-hydroxylase (C4H) | 2 | 1 | 1 | 1 |

Tables S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32 and S33). For example, four drought-related genes significantly enriched in this study were drought-responsive family protein, DROUGHT SENSITIVE 1, chloroplast drought-induced stress protein of 32 kDa, and drought-induced 19 protein. A representative list of significantly marked (H3K9_{me2} and H4K12_{ac}) stress-related genes are highlighted (Table 5).

Peaks associated with transcription factors

In total, 653 and 650 significantly enriched peaks were annotated as TFs and were methylated and acetylated, respectively, in AP13, while 431 and 579 significantly methylated and acetylated peaks, respectively, were annotated as TFs in VS16. Among these, 17 TFs were

differentially marked (H3K9_{me2} and H4K12_{ac}) between the two genotypes (Table 6; Additional file 1: Tables S34, S35, S36 and S37). The most prominent TFs in descending order were Myeloblastosis (Myb), Myb-domain containing proteins, Gibberellic Acid Insensitive (GAI)-Repressor of Gal (RGA)-Scarecrow (SCR) (GRAS) family, Basic Leucine Zipper Domain (bZIP) family, GATA TF family protein (ability to bind to the DNA sequence GATA), W-box Containing TF (WRKY) family, TEO-SINTE BRANCHED1 (TB1) from maize (*Zea mays*), CYCLOIDEA (CYC) from snapdragon (*Antirrhinum majus*), as well as PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1 (PCF1) and PCF2 (TCP) family, heat shock TFs, Plant AT-rich Sequence and Zinc-binding Protein (PLATZ) TF family protein, and the Basic Helix-Loop-Helix (bHLH) family of TFs.

Table 5 List of abiotic and biotic stress-related genes for H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16

| | Number of genes | | | |
|---|---------------------|------|---------------------|------|
| | H3K9 _{me2} | | H4K12 _{ac} | |
| | AP13 | VS16 | AP13 | VS16 |
| Heat Tolerance Genes | | | | |
| DNAJ heat shock family protein | 115 | 111 | 126 | 199 |
| Heat shock proteins (HSPs) | 78 | 71 | 91 | 119 |
| DNAK-related heat shock proteins | 22 | 19 | 24 | 21 |
| HEAT repeat-containing protein | 10 | 6 | 11 | 7 |
| Heat shock transcription factors | 10 | 22 | 15 | 16 |
| Drought Tolerance Genes | | | | |
| Drought-responsive family protein | 5 | 5 | 6 | 7 |
| DROUGHT SENSITIVE 1 | 3 | 3 | 2 | 1 |
| Chloroplast drought-induced stress protein of 32 kDa | 1 | 1 | 2 | 1 |
| Drought-induced 19 protein | 5 | 5 | 5 | 7 |
| Salinity Tolerance Genes | | | | |
| Salt tolerance Zinc finger | 3 | 2 | 4 | 2 |
| Low temperature and salt responsive protein family | 3 | 7 | 7 | 2 |
| Salt tolerance homolog 2 | 3 | 2 | 6 | 2 |
| Cold Tolerance Genes | | | | |
| Cold regulated gene 27 | 2 | 1 | 0 | 1 |
| Cold, circadian rhythm, and rna binding 2 | 2 | 1 | 2 | 1 |
| Cold shock domain protein 1 | 1 | 1 | 1 | 1 |
| Soybean gene regulated by cold 2 | 0 | 2 | 0 | 1 |
| Cold shock domain protein 3 | 0 | 1 | 0 | 2 |
| Cold regulated 314 inner membrane 1 | 0 | 1 | 0 | 2 |
| Cold regulated 413 plasma membrane 1 | 0 | 1 | 0 | 0 |
| THO complex, subunit 5 | 0 | 0 | 0 | 1 |
| Flooding Tolerance Genes | | | | |
| Subtilase family protein | 57 | 28 | 69 | 88 |
| Leucine-rich repeat transmembrane protein kinase | 28 | 25 | 34 | 44 |
| Transducin family protein/WD-40 repeat family protein | 19 | 22 | 21 | 46 |
| Expansin | 24 | 17 | 21 | 41 |
| Adenine nucleotide alpha hydrolase-like superfamily protein | 25 | 13 | 12 | 123 |
| Heavy metal Tolerance Genes | | | | |
| Heavy metal transport/detoxification superfamily protein | 58 | 38 | 52 | 40 |
| Copper transport protein family protein | 6 | 6 | 6 | 6 |
| Heavy metal atpase | 3 | 5 | 3 | 3 |

Table 5 List of abiotic and biotic stress-related genes for H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16 (Continued)

| | Number of genes | | | |
|--|---------------------|------|---------------------|------|
| | H3K9 _{me2} | | H4K12 _{ac} | |
| | AP13 | VS16 | AP13 | VS16 |
| Chloroplast-targeted copper chaperone protein | 3 | 3 | 3 | 6 |
| Farneslated protein 3 | 1 | 2 | 5 | 4 |
| Farneslated protein 6 | 1 | 0 | 1 | 0 |
| Disease-Resistance Genes | | | | |
| NBS-LRR disease resistance protein | 68 | 50 | 59 | 96 |
| LRR-NB-ARC domains-containing disease | 44 | 38 | 38 | 66 |
| NB-ARC domain containing protein | 202 | 215 | 251 | 66 |
| Disease resistance protein (CC_NBS-LRR class) | 39 | 37 | 35 | 40 |
| Disease resistance responsive (dirigent-like protein) family protein | 21 | 18 | 32 | 27 |

Quantitative PCR validation

Genome-wide ChIP-Seq analysis revealed 17,364 and 13,878 genes associated with peaks between AP13 and VS16, respectively, for both modifications (H3K9_{me2} and H4K12_{ac}). Among these, five genes (PAL, 4CL, HCT, CAD, and CCoAOMT) from the phenylpropanoid-monolignol pathway, differentially marked with histone modifications, were selected to confirm the expression level by qPCR. The significant ($P < 0.05$) variations, in the relative expressions, of all five genes indicated that they were more marked in AP13 than in VS16. Further, for both H3K9_{me2} and H4K12_{ac} in these two genotypes the levels of expression were higher for CAD and HCT, while lower for CCoAOMT (Fig. 4).

Discussion

Histone modifications, such as H3K9_{me2}, H4K12_{ac}, H3K27_{me2/me3}, H3K27_{me1}, H3K4_{me3}, and H3K9_{me3}, bind either to heterochromatic or euchromatic regions and orchestrate epigenetic regulation of gene expression in a tissue-specific or genome-wide manner. These modifications have been suggested to play an important role in growth, development, biotic stresses, and abiotic stresses in plants [8, 11–13]. Similar to our previous study in common bean [8] and other studies in rice [7], Arabidopsis [41], cotton [42], and a diatom (*Phaeodactylum tricorutum*) [43], in this study, a genome-wide screening to understand the binding of two histone marks associated with gene inactivation (H3K9_{me2}) and activation (H4K12_{ac}) was performed using ChIP-Seq analyses in two diverse switchgrass genotypes (AP13 and VS16)

Table 6 Number of transcription factors associated with H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16

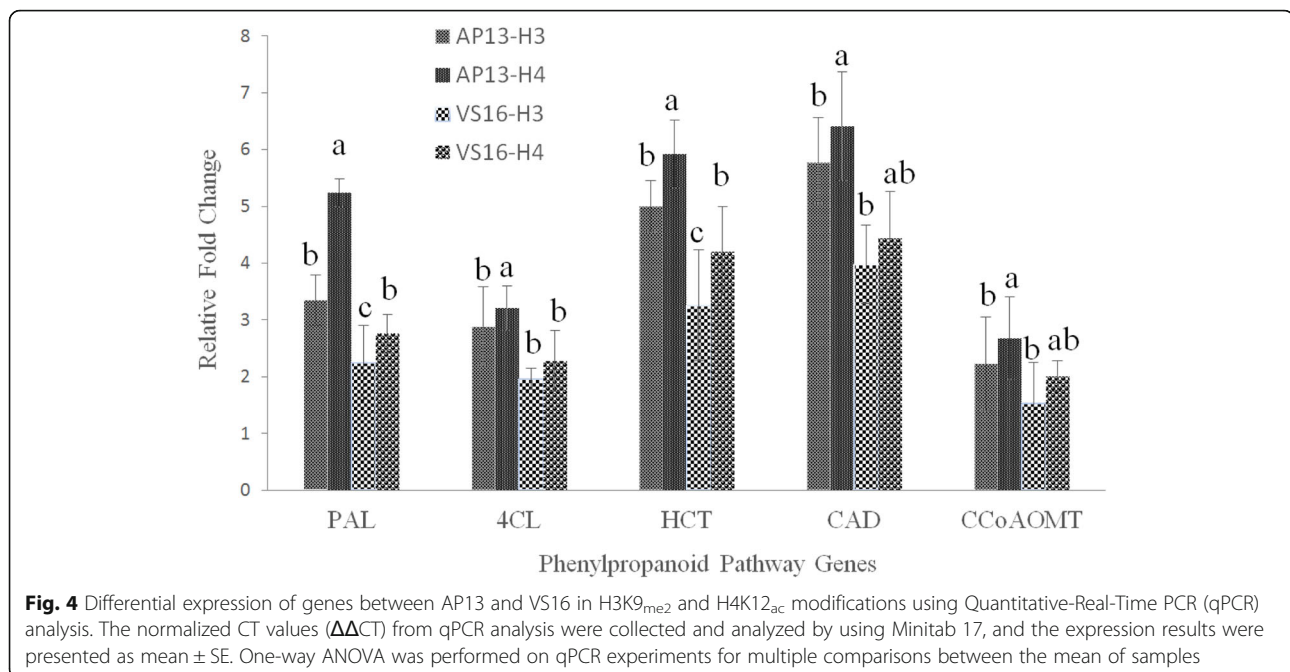
| Family of TFs | Function | Number of TFs | | | |
|---|---|---------------------|------|---------------------|------|
| | | H3K9 _{me2} | | H4K12 _{ac} | |
| | | AP13 | VS16 | AP13 | VS16 |
| Myb | Development, differentiation, metabolism, response to abiotic and biotic stresses | 103 | 58 | 95 | 82 |
| WRKY | Growth and development, secondary metabolism, leaf senescence, response to abiotic and biotic stresses | 37 | 22 | 27 | 38 |
| GRAS | Growth and development, gibberellin synthesis, cell division and endodermic specification | 65 | 38 | 52 | 40 |
| bHLH | Plant cell and tissue development, activation of genes in response to light, modulate gene expression, regulation of BR signaling pathway | 16 | 9 | 20 | 14 |
| bZIP | Photomorphogenesis, leaf and seed formation, energy homeostasis, response to abiotic and biotic stresses | 62 | 41 | 68 | 74 |
| TCP | Development, DNA binding, Morphogenesis, seed germination, leaf senescence, defense response | 27 | 14 | 19 | 19 |
| GATA | Nitrogen control, siderophore biosynthesis, photomorphogenesis, circadian regulation, mating-type switching | 41 | 21 | 61 | 41 |
| E2F TF3 | Cell cycle pathway, transcription activators and repressor, DNA replication | 2 | 0 | 0 | 0 |
| E2F FTF | Same as E2F TF3 functions | 0 | 4 | 6 | 4 |
| Redox Responsive TFs | Photosynthesis, Activation of gene expression, cross tolerance to abiotic and biotic stresses | 2 | 1 | 0 | 0 |
| Heat Shock TFs | Regulation of cellular homeostasis, promote survival and gene expression under environmental factors | 21 | 10 | 11 | 15 |
| Global TF Group A2 | Regulation of stress-responsive genes and <i>Hsps</i> | 6 | 4 | 5 | 10 |
| FRS TF | Same as Group A2 TF functions | 5 | 7 | 10 | 5 |
| PLATZ TF Family Protein | Response to abiotic and biotic stimuli, seed germination, transcriptional repression, cell cycle regulation and cell division | 17 | 21 | 11 | 13 |
| Ethylene Responsive Element Binding Factor 13 | Growth and development, hormonal regulation, development of ovule and seed coat | 2 | 5 | 2 | 4 |
| TF IIIC Subunit 5 | Transcription initiation, RNA Pol III binding | 5 | 4 | 12 | 9 |
| Plant-Specific TF YABBY Family Protein | Regulation of defense response, leaf senescence, trichome development, biosynthesis of specialized metabolites | 7 | 7 | 6 | 6 |

belonging to two ecotypes. The percentage of mapped reads to the reference genome in rice ranged between 91.70–98.90% [7], while in switchgrass we obtained mapping rates of 82.32 and 73.42% for H3K9_{me2} as well as 81.20 and 68.43% for H4K12_{ac} in AP13 and VS16, respectively, which are attributed to the genotype-specific adaptations, variations in GC-content, repeat regions, genome duplications and ploidies [6].

Histone modification patterns in switchgrass

With only slight variations between the two genotypes, genome-wide histone modification patterns were similar (Additional file 2: Figure S3A-D) when compared to

gene expression data from our recent effort [30]. More genes were associated with peaks marked with H4K12_{ac} than H3K9_{me2} in both genotypes, which is attributed to the activation of genotype-specific genes (Additional file 2: Figure S1). Even though the overall distribution pattern of histone marks is similar, slight variations existed between the two genotypes, which is reflected in the distribution of peaks in the respective genomes. Also, more genes were marked with H4K12_{ac} than H3K9_{me2} in both the genotypes. Moreover, bivalent histone modifications were commonly marked on specific genes for both genotypes, suggesting a significant degree of overlap between these two epigenetic marks. Our findings



are in agreement with genome-wide histone modification studies in rice [7, 44] and barley [45]. For example, in rice both activating (H3K4_{me3}) and repressive (H3K27_{me3}) histone modifications formed strong peaks within 1 kb downstream of the TSSs [7, 46]. ChIP-Seq analysis in *Populus* revealed that $\sim 90\%$ of ARK1 and ARK2 binding regions were within 1 kb of the TSSs [47]. In rice the enrichment of histone acetylation and methylation marks showed at 500 bp downstream from the TSSs [48]. The H3K4_{me3} distribution showed a clear peak around 300 bp downstream of the TSS in *Arabidopsis* [49]. Additionally, in *Arabidopsis* H3K9_{me2} and H3K4_{me} enrichment spanned around the entire gene body [50, 51].

Previous reports in rice [46] and humans [52] demonstrated that active histone marks are more enriched in bidirectional promoters compared to unidirectional promoters. A detailed analysis of the genomic locations associated with the peaks for H3K9_{me2} and H4K12_{ac} indicated higher affinity levels of H4K12_{ac} to switchgrass genes rather than H3K9_{me2}, suggesting the role of active histone marks in transcriptional activation of switchgrass genes. However, a greater number ($> 10\%$) of peaks associated with H4K12_{ac} and H3K9_{me2} marks spanned 5 kb upstream and downstream from TSSs in AP13 when compared to VS16.

Recently, the preferential binding of TFs ($\sim 86\%$) between -1000 to $+200$ bp from a TSS has been identified in *A. thaliana* [53]. In *Populus*, the majority of binding regions for all TFs analyzed have shown a general preference for binding near genes, and they were specifically enriched within 5 kb of the TSSs [47]. Histone-mediated transcription initiation/inactivation occurs by binding to

promoter and enhancer regions and also by modulating transcription elongation [48]. Analysis of histone modification profiles with bins of increasing gene length in *Arabidopsis* revealed that most peaks have been localized around 480 bp downstream of the TSSs. The shape and width of the H3K36_{ac} peaks were independent of gene length [12]. Similarly, in this study most H3K9_{me2} and H4K12_{ac} peaks were localized between 0 and 499 bp downstream, supporting the idea that peak position, shape, and length are independent of gene length. Transcription factors and DNA binding proteins were highly enriched in the cluster of genes associated with H3K27_{me3} in tissue-specific repression as reported in *Arabidopsis* (6). Genome-wide comparative analyses using two histone marks, H3K9_{ac} and H3K4_{me3}, in *A. thaliana* and *A. arenosa* revealed coordination between histone modifications and gene expression within and between these species, including allopolyploids [6].

Pathways and genes related to biomass/biofuel production

Further, we narrowed our search to identify histone markings on genes related to C4-photosynthesis and photorespiration pathways, and phenylpropanoid-monolignol pathway, as they play a key role in biomass and biofuel production [54]. Using three histone acetylation marks (H3K9_{ac}, H4K5_{ac}, and H3K18_{ac}) and two histone methylation marks (H3K4_{me3} and H3K4_{me2}) and targeting five C4 genes (CA, PPKK, ME, PEPCK, and RBCS2), comparative ChIP-Seq analyses reported in maize, sorghum (*Sorghum bicolor*), and foxtail millet (*Setaria italica*) revealed a role for histone modifications in light-specific and cell type-

specific responses [55]. Supporting this idea, our analyses here showed preferential binding of histone marks (H3K9_{me2} and H4K12_{ac}) between AP13 and VS16 in five prominent C4-pathway genes (DIT, PEP, PEPC, BASS and CA) five photorespiration-related genes (2-OG, Glu, 3-PGA, RuBisCo, and GLN); and, six monolignol pathway-related genes (C4H, PAL, 4CL, HCT, CAD, and CCoAOMT). Genes that mediate phenylpropanoid pathways play a key role in feedstock quality. A recent histone binding study in *Eucalyptus grandis* reported that about 8% of phenylpropanoid pathway genes were marked by H3K4_{me3} [56]. In this study, the histone modification patterns for 430 monolignol pathway-related genes between two ecotypes were mostly correlated (Fig. 5), supporting the idea of a common histone modification code between genotypes and possibly across species, as has been suggested.

Interestingly, this study identified several abiotic and biotic stress-related genes differentially marked by H3K9_{me2} and H4K12_{ac} in VS16 and AP13, which may contribute to differing genetic potential of these two contrasting ecotypes (upland and lowland). Here, we identified H3K9_{me2} and H4K12_{ac} modifications on different stress-related genes abundantly marked in both genotypes such as disease-resistance, heat tolerance, and flood tolerance genes. Our recent transcriptome analyses in switchgrass genotypes AP13 and VS16 identified over 200 significantly enriched stress-related genes [8]. Similarly in this study, prominently associated with H3K9_{me2} and H4K12_{ac} were

genes that encode the leucine rich repeats (LRR)-family of proteins, NBS-LRR, NB-ARC, LRR-NB-ARC, and CC-NBS-LRR (disease resistance); the heat shock proteins (HSPs), DNAJ and DNAK (heat tolerance); and the subtilase family protein, LRR-transmembrane protein kinase and adenine nucleotide alpha hydrolase-like superfamily protein (flooding tolerance) (Table 5). Even though the overall binding patterns were similar, slight variations in markings were observed between the two ecotypes. Therefore, changes noticed in histone bindings represent inherent differences between the upland (VS16) and lowland (AP13) switchgrass ecotypes.

Quantitative PCR validation

It is important to validate differentially marked genes in switchgrass, a prominent biofuel crop, especially in regard to the phenylpropanoid–monolignol biosynthesis pathway. Previous reports suggested that over ten different genes play an important role in the phenylpropanoid–monolignol pathway in switchgrass [57–59], including the five genes selected in this study (PAL, 4CL, HCT, CAD, and CCoAOMT). Though transcriptional networks associated with the phenylpropanoid pathway have been reported [60], studies regarding the regulation of phenylpropanoid genes by histones in plants are limited. In maize, acetylation of H3K9/K14 in the promoter region resulted in activation of the anthocyanin biosynthetic gene, A1, by interacting with EMSY-related factor, thus linking TFs with chromatin modifications [61]. Our

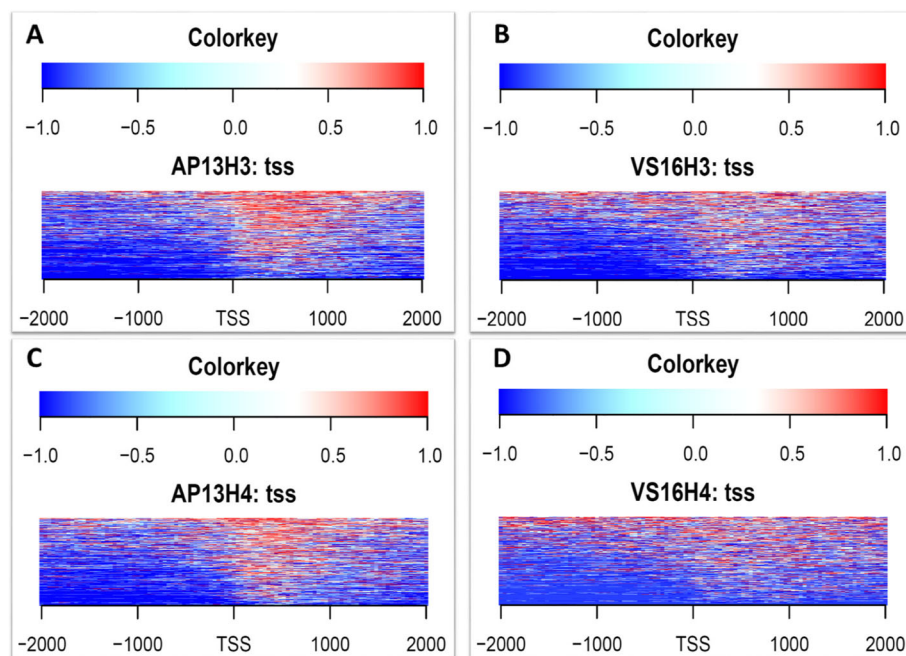


Fig. 5 Differentially marked histone modifications (H3K9_{me2} and H4K12_{ac}) on 430 selected phenylpropanoid-monolignol pathway-related genes between AP13 and VS16 using heatmap analysis

results suggested that all five genes selected for ChIP-qPCR from the phenylpropanoid pathway were significantly marked by both histone modifications in both genotypes. Recently, comparative RNA-Seq analyses in switchgrass suggested that phenylpropanoid pathway genes were abundantly expressed in crowns and rhizomes of lowland cultivar, Kanlow, when compared to upland cultivar, Summer [62]. Similarly, comparative ChIP-Seq and ChIP-qPCR analyses as reported here in switchgrass suggest that differentially methylated and acetylated histone modification regions associated with the phenylpropanoid pathway genes were relatively abundant in lowland cultivar, AP13, when compared to upland cultivar, VS16. This suggests the role of active and inactive histone modification marks in regulating genes associated with biomass yield and biofuel production. Further, it is reasonable to assume that phenylpropanoid pathway genes may undergo a combinatorial transcriptional regulation involving TFs and histone modifications.

Conclusions

Our genome-wide screening for histone methylation (H3K9_{me2}) and acetylation (H4K12_{ac}) marks in two contrasting ecotypes of switchgrass identified similarities and disparities between AP13 and VS16 at the epigenome level. The distribution of significantly enriched peaks spanned 0–2000 bp around TSSs, and most of these peaks were observed in non-genic regions. Our comprehensive analysis of annotated peaks facilitated the identification of key regulatory marks, genes, and TFs involved in biosynthetic pathways, including C4 photosynthesis, photorespiration, phenylpropanoid–monolignol, as well as biotic and abiotic stresses.

In summary, we integrated ChIP-seq and RNA-seq datasets to explore the interactions between histone modifications and TFs/transcripts related with gene expression and their patterns across two contrasting genotypes of switchgrass, AP13 and VS16. Furthermore, the integrated methodology and results highlighted here will guide future studies in exploring similar interactions in gene expression and regulation.

Additional files

Additional file 1: Tables S1-S37. Supplementary tables with legends in the first sheet called “Supplementary Table Legends”. (XLSX 479 kb)

Additional file 2: Figure S1. Number of H3K9_{me2} and H4K12_{ac} binding regions in switchgrass genotypes AP13 and VS16. (PPTX 320 kb)

Additional file 3: Figure S2. Heatmap of preferentially marked H3K9_{me2} and H4K12_{ac} genes along with transcriptional activity 2 kb upstream and downstream from TSSs in switchgrass genotypes AP13 and VS16. (A) H3K9_{me2}-AP13 (B) H3K9_{me2}-VS16 (C) H4K12_{ac}-AP13 and (D) H4K12_{ac}-VS16. Genes were sorted according to their expression level from mRNA analysis. The histone modification intensities are displayed along with –2 kb to +2 kb regions around TSSs. (PPTX 1474 kb)

Additional file 4: Figure S3. Peak width distributions of H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16. (A) H3K9_{me2}-AP13 (B) H3K9_{me2}-VS16 (C) H4K12_{ac}-AP13 and (D) H4K12_{ac}-VS16. The numbers 499 and 999 in peak width bin column refer to the number of peaks having a peak width between 0 to 499 and 500 to 999, respectively. (PPTX 369 kb)

Additional file 5: Figure S4. List of epigenomic factors with H3K9_{me2} and H4K12_{ac} modifications in switchgrass genotypes AP13 and VS16. (PPTX 325 kb)

Abbreviations

2-OG: 2-oxoglutarate; 3-PGA: 3-phosphoglycerate; 4CL: 4-coumarate: CoA ligase; BASS: Bile acid sodium symporter/pyruvate transporter (BASS); bHLH: The Basic Helix-Loop-Helix; BS-Seq: Bisulfite sequencing; bZIP: Basic Leucine Zipper Domain Containing TF; C4H: Cinnamate 4-hydroxylase; CA: Carbonic anhydrase (CA); CAD: Cinnamyl alcohol dehydrogenase; CAT: Catalase; CCoAMOT: Caffeoyl-CoA 3-O-methyltransferase; ChIP-Seq: Chromatin immunoprecipitation sequencing; CT: Cycle threshold; DIT: Dicarboxylate transporter; FDR: False discovery rate; GLN: Glutamine; GLU: Glutamate; GO: Gene Ontology; GRAS: Gibberellic Acid Insensitive (GAI)-Repressor of Ga1 (RGA)-Scarecrow (SCR); GS: Glutamate synthase; H3K9_{me2}: Histone H3 di-methylated at lysine 9; H4K12_{ac}: Histone H4 acetylated at lysine 12; HCT: Hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase; IGV: Integrative Genomics Viewer; KEGG: Kyoto Encyclopedia of Genes and Genomes; lncRNAs: Long non-coding RNAs; Log₂FC: Fold change; MeDIP-Seq: Methylated DNA immunoprecipitation sequencing; Myb: Myeloblastosis; NADP: Nicotinamide adenine dinucleotide phosphate; PAL: Phenylalanine ammonia-lyase; PEP: Phosphoenol pyruvate; PEP-CK: Phosphoenolpyruvate carboxylase; PEP-CK: Phosphoenolpyruvate carboxylase; PLATZ: Plant AT-rich Sequence and Zinc-binding Protein; qPCR: Quantitative Real-Time (RT)-PCR; QTL: Quantitative trait loci; RuBisCo: Ribulose 1,5-bisphosphate carboxylase/oxygenase; SICER: Spatial clustering for identification of ChIP-enriched regions; siRNAs: Small interfering RNAs; TCP: TEOSINTE BRANCHED1 (TB1), CYCLOIDEA (CYC), PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR1 (PCF1) and PCF2; TEs: Transposable elements; TF: Transcription factor; TSS: Transcriptional start site; WRKY: W-box Containing TF

Acknowledgments

We thank the Department of Energy Joint Genome Institute for pre-publication access and permission to publish this analysis using the reference genome of *Panicum virgatum*.

Authors' contributions

VK designed and supported the study and edited the manuscript. VA performed the experiments and wrote the manuscript. VRS analyzed the data and wrote the manuscript. MCS provided plant samples, designed the study and edited the manuscript. JT and KP analyzed the data, assisted in writing and edited the manuscript. EF assisted in undertaking experiments and data analysis. All authors read and approved the final manuscript.

Funding

This work was supported by Delaware EPSCoR with funds from the National Science Foundation Grant EPS-0814251 and the State of Delaware to VK. These funds have been utilized in the design of the study, data collection, analysis, interpretation, and in writing the manuscript.

Availability of data and materials

All data sets generated as part of this study are available at the SRA section of the NCBI with the bio-project number PRJNA280864 for ChIP-Seq experiment.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Molecular Genetics and Epigenomics Laboratory, College of Agriculture and Related Sciences, Delaware State University, Dover, DE, USA. ²Molecular Biology and Bioinformatics Laboratory, College of Agricultural, Life and Natural Sciences, Alabama A&M University, Normal, AL, USA. ³Center for Integrated Biological and Environmental Research, Delaware State University, Dover, DE, USA. ⁴Noble Research Institute, Ardmore, OK, USA. ⁵Bioinformatics Core, Purdue University, West Lafayette, IN, USA.

Received: 31 May 2018 Accepted: 16 August 2019

Published online: 22 August 2019

References

- Jin B, Li Y, Robertson KD. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer*. 2011;2:607–17.
- Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol*. 2015;16:519–32.
- Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010;31:27–36.
- Li W, Liu H, Cheng ZJ, Su YH, Han HN, Zhang Y, Zhang XS. DNA methylation and histone modifications regulate *De Novo* shoot regeneration in *Arabidopsis* by modulating WUSCHEL expression and auxin signaling. *PLoS Genet*. 2011;7:e1002243.
- Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci*. 2014;217–218:109–19.
- Ha M, Ng DWK, Li WH, Chen ZJ. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res*. 2011;21:590–8.
- Du Z, Li H, Wei Q, Zhao X, Wang C, Zhu Q, Yi X, Xu W, Liu XS, Jin W, Su Z. Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. *Japonica*. *Mol Plant*. 2013;6:1463–72.
- Ayyappan V, Kalavacharla V, Thimmapuram J, Bhide KP, Sripathi VR, Smolinski TG, Manoharan M, Thurston Y, Todd A, Kingham B. Genome Wide Profiling of Histone Modifications (H3K9_{me2} and H4K12_{ac}) and Gene Expression in Rust (*Uromyces appendiculatus*) infected Common Bean (*Phaseolus vulgaris* L.). *PLoS ONE*. 2015;10:e0132176.
- Pfluger J, Wagner D. Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol*. 2007;10:645–52.
- Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev*. 2004;18:1263–71.
- Charron JB, He H, Elling AA, Deng XW. Dynamic landscapes of four histone modifications during deetiolation in *Arabidopsis*. *Plant Cell*. 2009;21:3732–48.
- Mahrez W, Arellano MS, Moreno-Romero J, Nakamura M, Shu H, Nanni P, Köhler C, Gruissem W, Hennig L. H3K36ac is an evolutionary conserved plant histone modification that Marks active genes. *Plant Physiol*. 2016;170:1566–77.
- Rodríguez-Sanz H, Moreno-Romero J, Solís MT, Köhler C, Risueño MC, Testillano PS. Changes in histone methylation and acetylation during microspore reprogramming to embryogenesis occur concomitantly with Bn HKMT and Bn HAT expression and are associated with cell totipotency, proliferation, and differentiation in *Brassica napus*. *Cytogenet Genome Res*. 2014;143:209–18.
- Van der Weijde T, Alvim Kamei CL, Torres AF, et al. The potential of C4 grasses for cellulosic biofuel production. *Front Plant Sci*. 2013;4:107.
- Bouton JH. Molecular breeding of switchgrass for use as a biofuel crop. *Curr Opin Genet Dev*. 2007;17(6):553–8.
- Rao MN, Soneji JR, Kwit C, Stewart CN Jr. Advances in biotechnology and genomics of switchgrass. *Biotechnol Biofuels*. 2013;6(77):1–15.
- Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, Tobias CM. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics*. 2010;185:745–60.
- Serba D, Wu L, Daverdin G, Bahri BA, Wang X, Kilian A, Bouton JH, Brummer EC, Saha MC, Devos KM. Linkage maps of lowland and upland tetraploid switchgrass ecotypes. *Bioenerg Res*. 2013;6:953–65.
- Palmer NA, Donze-Reiner T, Horvath D, Heng-Moss T, Waters B, Tobias C, Sarath G. Switchgrass (*Panicum virgatum* L) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics. *Funct Integr Genomics*. 2015;15:1–16.
- Wang Y, Zeng X, Peal L, Tang Y, Wu Y, Mahalingam R. Transcriptome analysis of nodes and buds from high and low Tillering switchgrass inbred lines. *PLoS One*. 2013;8:e83772.
- Li YF, Wang Y, Tang Y, Kakani VG, Mahalingam R. Transcriptome analysis of heat stress response in switchgrass (*Panicum virgatum* L.). *BMC Plant Biol*. 2013;13:153.
- Serba DD, Uppalapati SR, Mukherjee S, Krom N, Tang Y, Mysore KS, Saha MC. Transcriptome profiling of rust resistance in switchgrass using RNA-Seq analysis. *Plant Genome*. 2015;8:2.
- Dworkin M, Xie S, Saha M, Thimmapuram J, Kalavacharla VK. Analyses of methylomes of upland and lowland switchgrass (*Panicum virgatum*) ecotypes using MeDIP-seq and BS-seq. *BMC Genomics*. 2017;18:851.
- Yan H, Bombarely A, Xu B, Frazier TP, Wang C, Chen P, Chen J, Hasing T, Cui C, Zhang X, Zhao B, Huang L. SiRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnol Biofuels*. 2018;11:208.
- Zhou J, Wang X, He K, Charron JB, Elling A, Deng X. Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol Biol*. 2010;72:585–95.
- West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW. Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One*. 2014;9:e105267.
- Hultquist SJ, Vogel KP, Lee DJ, Arumuganathan K, Kaeppler S. Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. *Crop Sci*. 1996;36:1049–52.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40(D1):D1178–86.
- Serba DD, Uppalapati SR, Krom N, Mukherjee S, Tang Y, Mysore KS, Saha MC. Transcriptome analysis in switchgrass discloses ecotype difference in photosynthetic efficiency. *BMC Genomics*. 2016;17(1):1040.
- Ayyappan V, Saha MC, Thimmapuram J, Sripathi VR, Bhide KP, Fiedler E, Hayford RK, Kalavacharla VK. Comparative transcriptome profiling of upland (VS16) and lowland (AP13) ecotypes of switchgrass. *Plant Cell Rep*. 2016;36:129–50.
- Medina R, Ghule PN, Cruzat F, Barutcu AR, Montecino M, Stein JL, van Wijnen AJ, Stein GS. Epigenetic control of cell cycle-dependent histone gene expression is a principal component of the abbreviated pluripotent cell cycle. *Mol Cell Biol*. 2012;32:3860–71.
- Nitzsche A, Steinhauber C, Mucke K, Paulus C, Nevels M, Steinh C, Katrin M, Steinhäusser C. Histone H3 lysine 4 methylation marks post replicative human cytomegalovirus chromatin. *J Virol*. 2012;86:9817–27.
- Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet*. 2013;4:288.
- Langmead B, Salzberg S. Fast-gapped read alignment with bowtie 2. *Nat Methods*. 2012;9:357359.
- Xu S, Grullon S, Ge K, Peng W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods Mol Biol*. 2014;1150:97–111.
- Huang W, Loganantharaj R, Schroeder B, Fargo D, Li L. PAVIS: a tool for peak annotation and visualization. *Bioinformatics*. 2013;29(23):3097–9.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(Delta Delta C(T))} method. *Methods*. 2001;25:402–8.
- Minitab 17 Statistical Software. Computer software. State College: Minitab, Inc.; 2010. <http://www.minitab.com>
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Lander S, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
- Wang C, Liu C, Roqueiro D, Grimm D, Schwab R, Becker C, Lanz C, Weigel D. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res*. 2014;25:246–56.
- Zheng D, Ye W, Song Q, Han F, Zhang T, Chen ZJ. Histone modifications define expression Bias of Homoeologous genomes in Allotetraploid cotton. *Plant Physiol*. 2016;172:1760–71.
- Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, Dingli F, Rivarola M, Ott S, Liu X, Sun Y, Rabinowicz PD, McCarthy J, Allen AE, Loew D, Bowler C, Tirichine L. An integrative analysis of posttranslational histone modifications in the marine diatom *Phaeodactylum tricorutum*. *Genome Biol*. 2015;16:102.

44. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* 2012;22:151–62.
45. Baker K, Dhillon T, Colas I, Cook N, Milne I, Milne L, Bayer M, Flavell AJ. Chromatin state analysis of the barley epigenome reveals a higher order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J.* 2015;84:111–24.
46. Fang Y, Wang L, Wang X, You Q, Pan X, Xiao J, Wang XE, Wu Y, Su Z, Zhang W. Histone modifications facilitate the coexpression of bidirectional promoters in rice. *BMC Genomics.* 2016;17:768.
47. Liu L, Ramsay T, Zinkgraf M, Sundell D, Street NR, Filkov V, Groover A. A resource for characterizing genome-wide binding and putative target genes of transcription factors expressed during secondary growth and wood formation in *Populus*. *Plant J.* 2015;82:887–98.
48. Li X, Wang X, He K, Ma Y, Su N, He H, Stolc V, Tongprasit W, Jin W, Jiang J, Terzaghi W, Li S, Deng XW. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell.* 2008;20:259–76.
49. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* 2009;10:R62.
50. Zhou J, Wang X, He K, Charron JF, Elling AA, Deng XW. Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol Biol.* 2010;72:585–95.
51. van Dijk K, Ding Y, Malkaram S, Riethoven JM, Liu R, Yang J, Laczko P, Chen H, Xia Y, Ladunga I, Avramova Z, Fromm M. Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. *BMC Plant Biol.* 2010;10:238.
52. Bornelov S, Komorowski J, Wadelius C. Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. *BMC Genomics.* 2015;16:300.
53. Yu CP, Lin JJ, Li WH. Positional distribution of transcription factor binding sites in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2016;6:25164.
54. Van der Weijde T, Alvim Kamei CL, Torres AF, Vermerris W, Dolstra O, Visser RGF, Trindade LM. The potential of C4 grasses for cellulosic biofuel production. *Front Plant Sci.* 2013;4(107):1–18.
55. Heimann L, Horst I, Perduns R, Dreesen B, Offermann S, Peterhansel C. A common histone modification code on C4 genes in maize and its conservation in *Sorghum* and *Setaria italica*. *Plant Physiol.* 2013;162:456–69.
56. Hussey SG, Mizrahi E, Groover A, Berger DK, Myburg AA. Genome-wide mapping of histone H3 lysine 4 trimethylation in *Eucalyptus grandis* developing xylem. *BMC Plant Biol.* 2015;15:117.
57. Simmons BA, Loqué D, Ralph J. Advances in modifying lignin for enhanced biofuel production. *Curr Opin Plant Biol.* 2010;13:313–20.
58. Fu C, Xiao X, Xi Y, Ge Y, Chen F, Bouton J, Dixon RA, Wang ZY. Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass. *Bioenerg Res.* 2011a;4:153–64.
59. Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC, He J, Shen H, Srivastava AC, Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK. Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). *Plant J.* 2013;74:160–73.
60. Zhao Q, Dixon RA. Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 2011;16:227–33.
61. Hernandez JM, Feller A, Morohashi K, Frame K, Grotewold E. The basic helix loop helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. *Proc Natl Acad Sci U S A.* 2007;104:17222–7.
62. Palmer NA, Saathoff AJ, Tobias CM, Twigg P, Xia Y, Vogel KP, et al. Contrasting metabolism in Perenniating structures of upland and lowland switchgrass plants late in the growing season. *PLoS One.* 2014;9:e105138.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

