JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

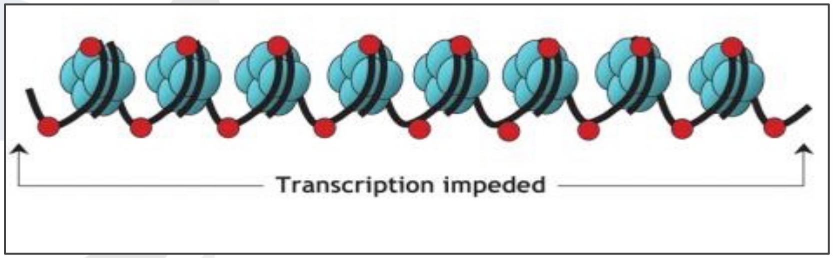
DNA Methylation Sequencing (Bisulfite-seq)

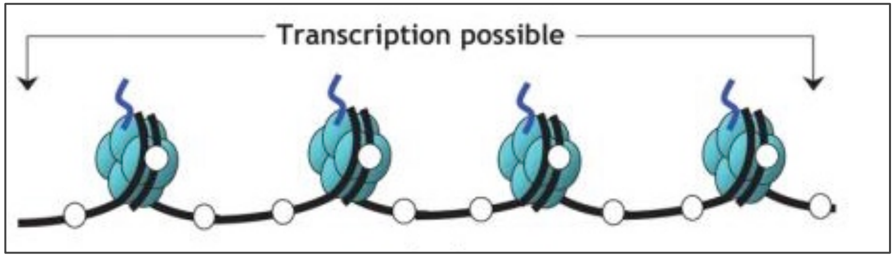
Kunde Ramamoorthy Govindarajan
(Govind)

Overview

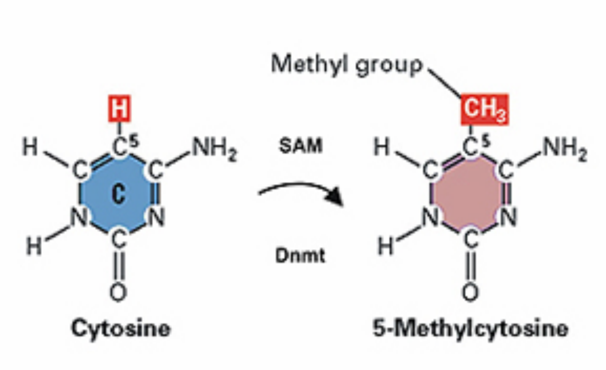
- Importance of DNA methylation
- Bisulfite-seq Technology
- Mapping challenges
- QC & analysis workflow
- Case study: *Sorghum bicolor*
- Reads & Mapping statistics
- Results & Summary

Why DNA methylation?





Major roles in fundamental cellular activities.
- -control of gene expression
- -maintenance of genome integrity
- -control of genomic imprinting



70-80% of all CpG di-nucleotides are methylated.
- -most of this occurs in repetitive elements or
- -regions of low CpG density

CpG rich regions (CpG Islands):
- -often found in gene promoters
- - ‘generally’ unmethylated

<u>Context:</u>
CpG, CHG, CHH
H – stands for A, T or C

Bisulfite Modification of DNA

Unmodified sequence:

GTC^{m}GAACCGTTCATGTTGC^{m}GAGCTG

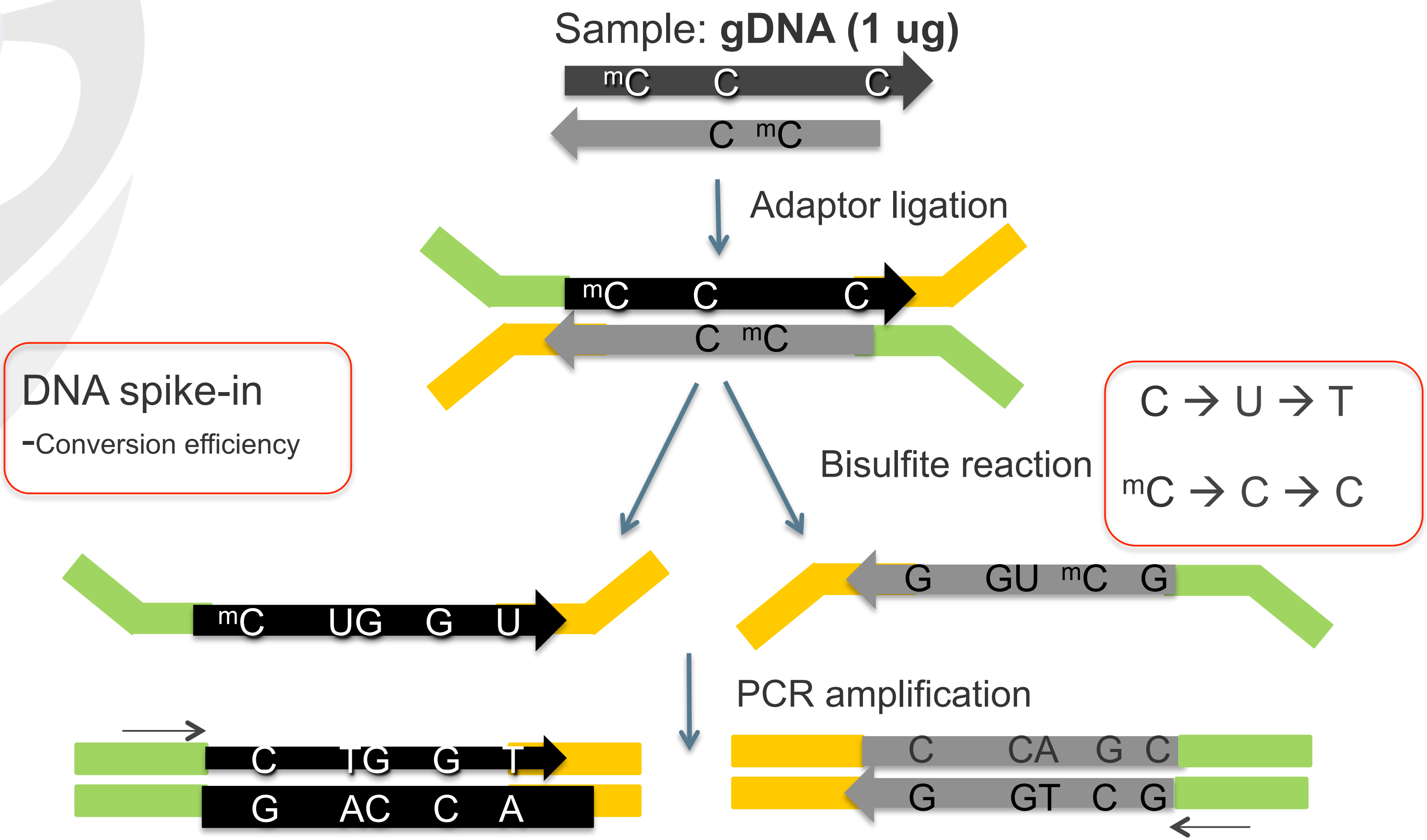
↓ Bisulfite Modification

GTC^{m}GAAUUGTTUATGTTGC^{m}GAGUTG

↓ PCR

GT CG AATTGTTTATGTTG CG AGTTG

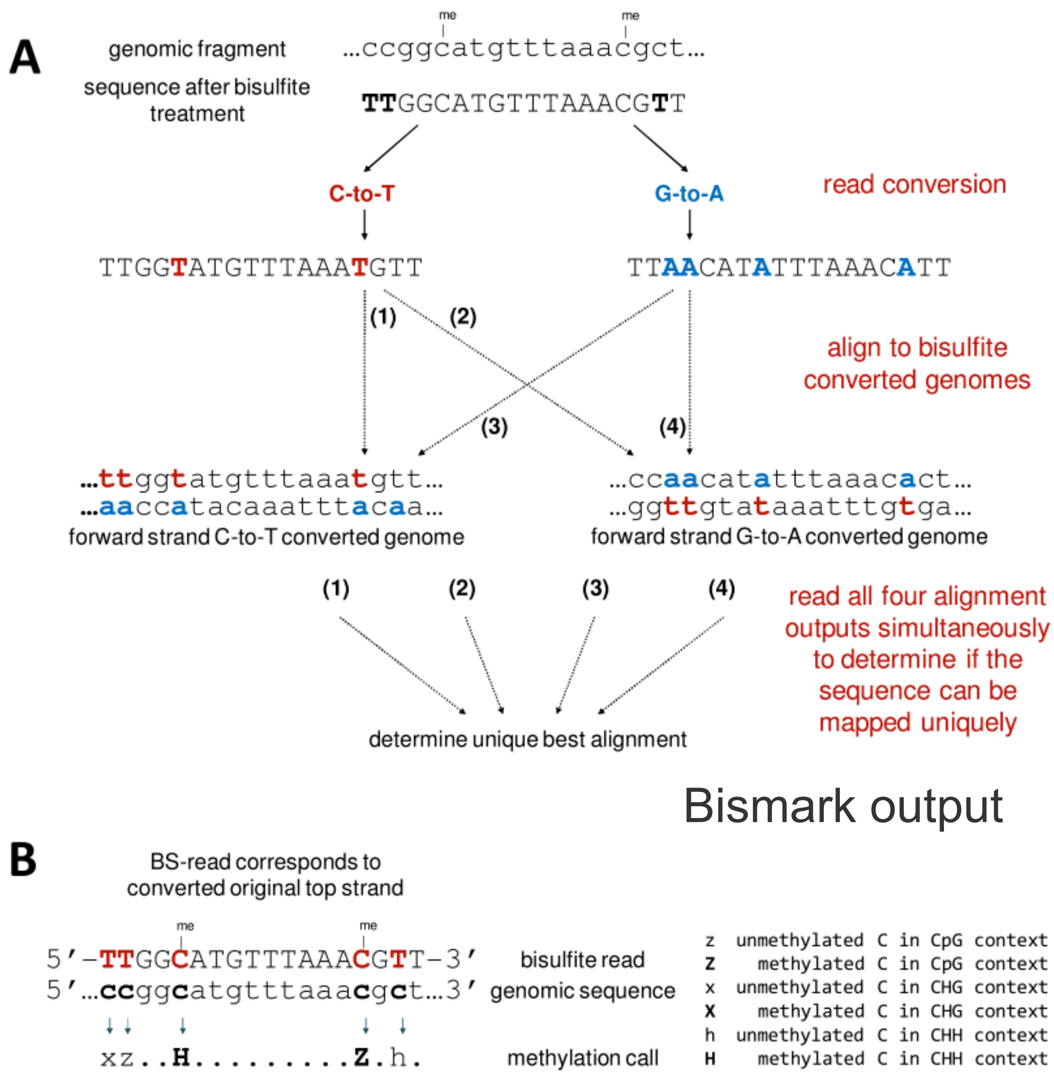
Bisulfite-seq



Methylation Data Analysis Software

Software	Features
BISMARK	Supports both single end and pair-end reads. Uses bowtie aligner.
PASH 3.0	Methylation & SNP's. Uses low memory & Slow speed alignment
BSMAP	Maps both single/pair-end reads. Uses SOAP aligner.
Methylcoder	Maps both single/pair-end reads. Handles also color space reads (SOLiD).
BS-Seq	Uses Gaussian Mixture model (GMM) to identify the probability of A vs G vs C vs T. GMM available only to Arabidopsis genome
BRAT	Maps both single/pair-end reads. Trims low quality bases. Improves unique mapping for pair-end reads.
Kismeth	Web-based tool. Designed for plant methylation data.

BISMARK algorithm



Felix Krueger & Simon R. Andrews Bioinformatics 2011

BISMARK algorithm

- Simple to use and widely used in the DNA methylation community.
- Bismark uses Bowtie2 mapper for alignment.
- Post-processing scripts to parse aligned reads to identify methylated and unmethylated C's.
- Handles both single and pair-end libraries.
- Output both C's context and % methylation.

JGI Workflow: QC & Analysis

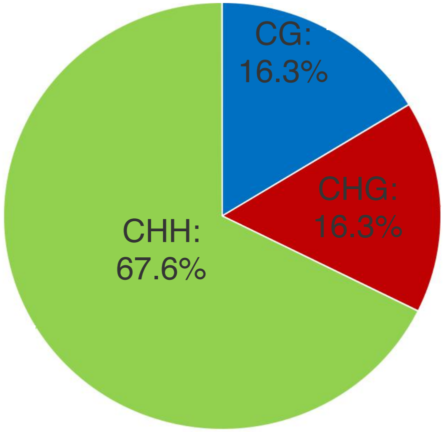
JOINT GENOME INSTITUTE

Methyl-seq (DNA seq) → Contaminants screening → Adaptor trimming/ Quality filtering → Bisulfite conversion (spike-ins) → Genome mapping → Remove duplicate reads (PCR amplification) →

- Coverage
- Methylation calls & % methylation
  - CpG, CHG, CHH

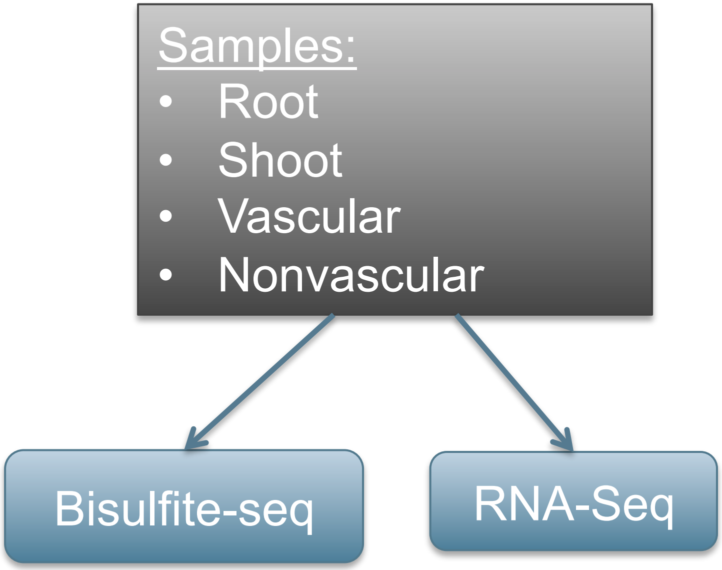
Sorghum bicolor

Genome Information:
Genome size: ~ 700 Mbp
# of Cytosines: 306.4 millions



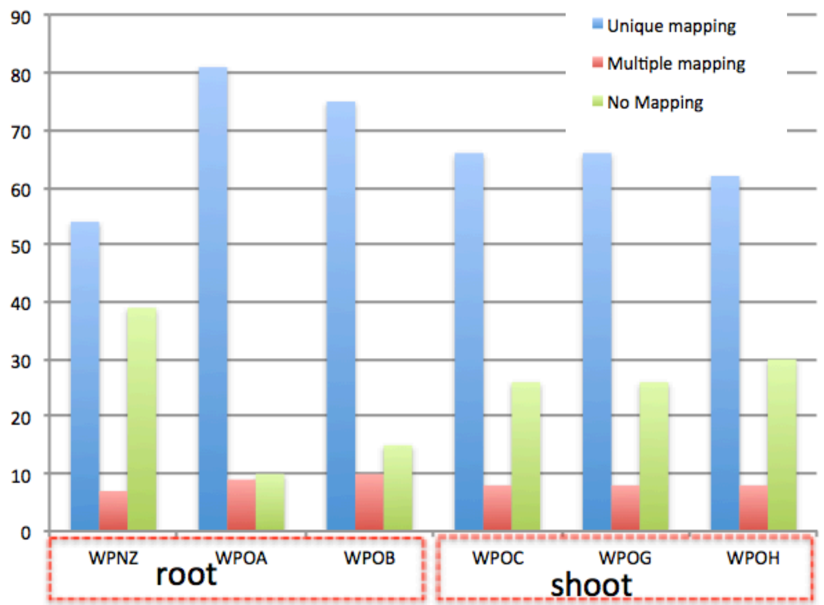
Plant secondary cell walls (SCWs) contains important polymers for the production of biofuels.

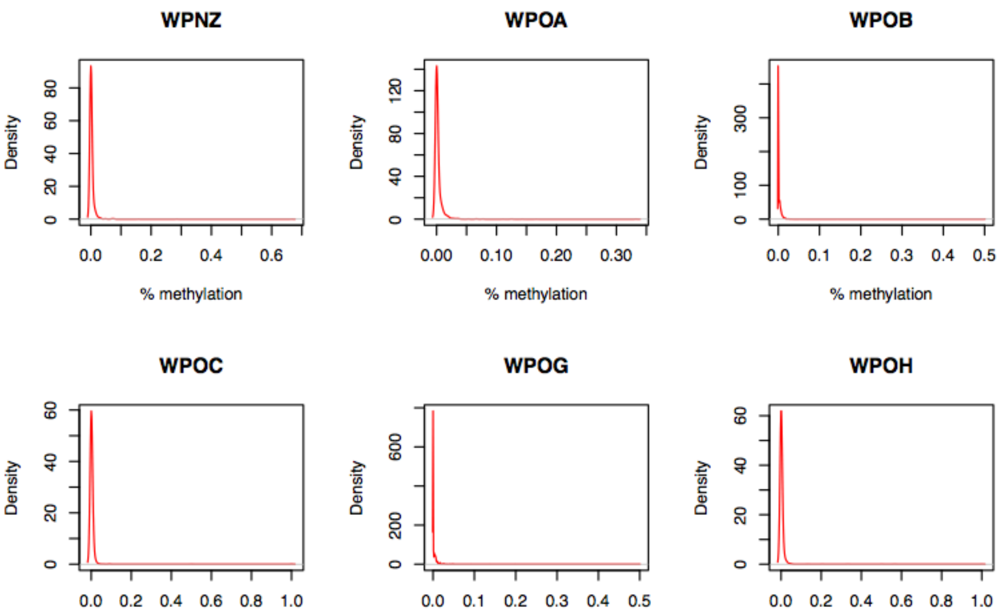
Identification of Tissue specific TF's through gene expression and methylations that regulate SCW's.



Siobhan M Brady (In prep)

Mapping & Bisulfite-conversion efficiency

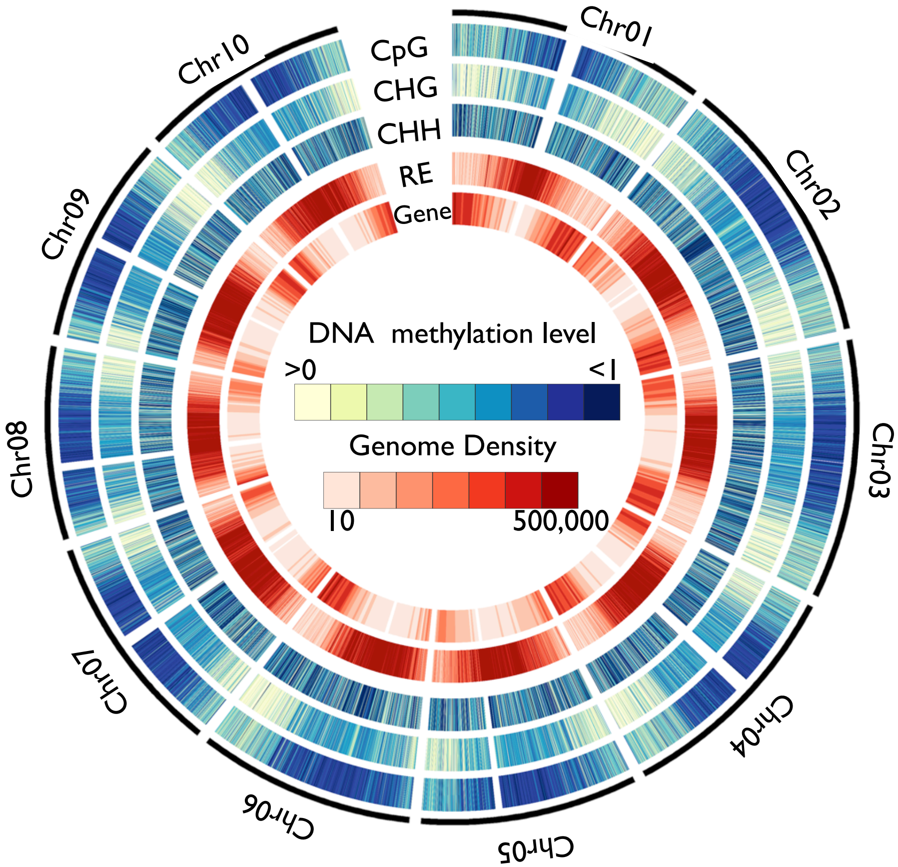




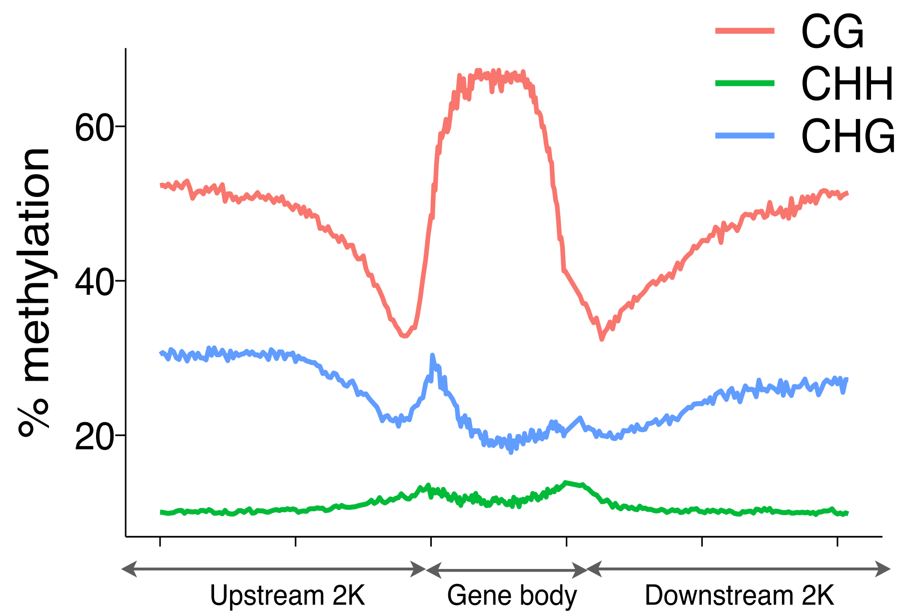
~ 75 % of the reads are mapped to genomes (unique and multiple hits)

Bisulfite conversion efficiency is very high (> 99 %)

Example: Global DNA methylation



CG and CHG methylation marks correlate with repetitive elements (RE) and reversely correlate with gene density.

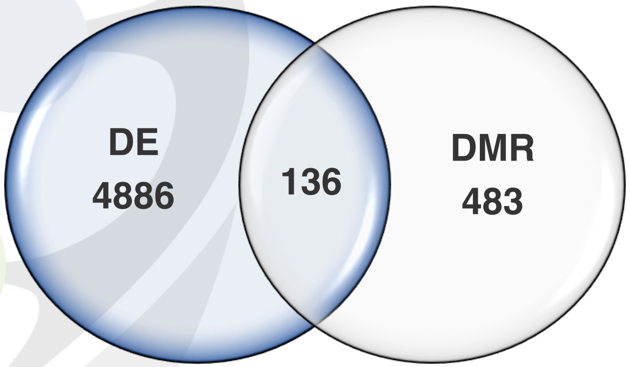


CGs are hypermethylated in gene bodies and hypomethylated near TSS as expected.

Siobhan M Brady (In prep)

Differential methylation

overlap between DMR and DE



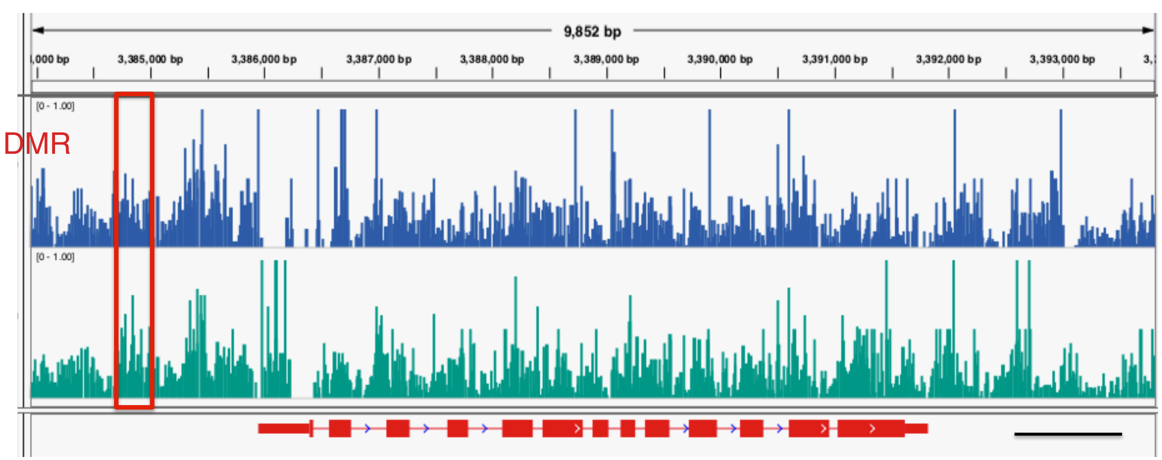
DMR = Differentially methylated regions
DE = Differentially expressed genes

Selected differentially methylated regions in vascular and nonvascular tissues

Gene Name	DMR	Annotation	DE
Sb06g026480	VAS	expansin precursor, putative, expressed	NON
Sb01g027880	VAS	CSLD1 - cellulose synthase-like family D, expressed	NON
Sb03g005570	VAS	gibberellin receptor GID1L2, putative, expressed	NON
Sb06g019290	VAS	GASR4 - Gibberellin-regulated GASA/GAST/Snakin family protein precursor, expressed	VAS
Sb10g024830	VAS	STRUBBELIG-RECEPTOR FAMILY 7 precursor, putative, expressed	VAS
Sb04g012910	VAS	nodulin MtN3 family protein, putative, expressed	VAS
Sb10g021750	VAS	BES1/BZR1 homolog protein, putative, expressed	VAS
Sb03g037510	VAS	OsFBX27 - F-box domain containing protein, expressed	VAS
Sb10g009640	NON	GASR7 - Gibberellin-regulated GASA/GAST/Snakin family protein precursor, expressed	VAS
Sb10g028580	NON	glycosyl hydrolases family 16, putative, expressed	VAS
Sb07g020920	NON	linker histone H1 and H5 family protein, expressed	VAS
Sb01g004210	NON	CESA2 - cellulose synthase, expressed	VAS

Nonvascular DNA Methylation

Vascular DNA Methylation



Sb01g004210

1Kb

Siobhan M Brady (In prep)

Summary

At JGI, we have the capability to carry out DNA methylation projects/studies.

Sorghum bicolor Bisulfite-seq data shows high genome mapping and Bisulfite conversion efficiency.

Identification of Tissue specific TF's through gene expression and methylation that regulate SCW's, we can develop stronger and thicker walls resulting in higher biomass.

Acknowledgements

Siobhan M. Brady, PhD
Gina Turco, Graduate Student
All Collaborators

JGI:

Atif Shahab

Chia-Lin Wei, PhD

Matthew Blow, PhD

Chee-Hong Wong

Chew Yee Ngan, PhD

Len Pennacchio, PhD & Management