Integrated Analysis of DNA Methylome and Transcriptome Reveals Epigenetic Regulation of CAM Photosynthesis in Pineapple

Haifeng Wang (wanghaifeng0613@foxmail.com)
Fujian Agriculture and Forestry University https://orcid.org/0000-0001-7213-927X
Yan Shi
Fujian Agriculture and Forestry University
Xingtan Zhang
Fujian Agriculture and Forestry University
Xiaojun Chang
Fujian Agriculture and Forestry University
Maokai Yan
Fujian Agriculture and Forestry University
Heming Zhao
Fujian Agriculture and Forestry University
Yuan Qin
Fujian Agriculture and Forestry University

Research article

Keywords: Epigenetics, DNA methylation, transcriptome, Ananas comosus, photosynthesis

Posted Date: October 8th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-81495/v1

Version of Record: A version of this preprint was published on January 6th, 2021. See the published version at https://doi.org/10.1186/s12870-020-02814-5.

Abstract

Background:

Crassulacean acid metabolism (CAM) photosynthesis is an important carbon fixation pathway especially in arid environments because it leads to higher water-use efficiency compared to C3 and C4 plants. However, the role of DNA methylation in regulation CAM photosynthesis is not fully understood.

Results:

Here, we performed temporal DNA methylome and transcriptome analysis of non-photosynthetic (white base) and photosynthetic (green tip) tissues of pineapple leaf. The DNA methylation patterns and levels in these two tissues were generally similar for the CG and CHG cytosine sequence contexts. However, CHH methylation was reduced in white base leaf tissue compared with green tip tissue across diel time course in both gene and transposon regions. We identified thousands of local differentially methylated regions (DMRs) between green tip and white base at different diel periods. We also showed that thousands of genes that overlapped with DMRs were differentially expressed between white base and green tip leaf tissue across diel time course, including several important CAM pathway-related genes, such as beta-CA, PEPC, PPCK, and MDH.

Conclusions:

Together, these detailed DNA methylome and transcriptome maps provide insight into DNA methylation changes and enhance our understanding of the relationships between DNA methylation and CAM photosynthesis.

Background

Drought is one of the most important abiotic stresses affecting the growth and development of plants and crops worldwide[1, 2], resulting in massive production losses. Compared to C3 and C4 plants, Crassulacean acid metabolism (CAM) plants have greater water-use efficiency (WUE) and are better adapted to arid and semi-arid regions.

Pineapple (*Ananas comosus*) is a major tropical crop, representing more than 20% of the world production of tropical fruits[3]. Pineapple fruit is used as a fresh and processed product, and global pineapple production is about 25.8 million metric tons fresh fruit[4]. Pineapple is also a model plant for studying CAM photosynthesis as an adaptation for increased water-use efficiency. In CAM plants, the stomata in the leaves are closed during the day, reducing evapotranspiration, but open at night to absorb carbon dioxide (CO_2). The CO_2 absorbed is stored in vacuoles in the form of four-carbon acid malate at night. During the daytime, malate can be transported into chloroplasts and converted into CO_2 for photosynthesis. The CO_2 is accumulated near the enzyme Rubisco, which reduces its oxygenase activity[5]. The benefits of this system are that CAM plants reduce water loss by closing stomata during

the day and also have increased photosynthetic efficiency. The mechanism and evolution of CAM have been extensively investigated[6–10]. For instance, many genes putatively involved in the carbon fixation module of CAM in pineapple have been identified[6], and many of them are differentially regulated in different leaf tissues[11].

DNA methylation is a heritable epigenetic modification found in most eukaryotic organisms including plants, animals, and fungi[12, 13]. In contrast to animals, where DNA methylation occurs mostly at the CG cytosine sequence context, DNA methylation in plants can occur at CG and CHG (where H is A, T, or C) as well as the CHH context[14–16]. To date, DNA methylome studies in plants have investigated the roles of methylation in seed development, flowering time, hybrid vigor, and gene evolution[17–20]. Many plants are also known to undergo genome-wide DNA methylation changes under different stress and environmental stimuli[2, 21–23]. For example, Liang and colleagues found that genome-wide DNA methylation levels in *Populus* were increased under drought stress compared to the control condition[24]. A recent study in rice highlighted the differences in DNA methylation and the impact on gene expression in three different cultivars (IR64, stress-sensitive; Nagina 22, drought-tolerant; and Pokali, salinity-tolerant). Specifically, extensive differences in DNA methylation among these three cultivars were observed, and numerous differentially methylated regions were found to be associated with differentially expressed genes[25].

In pineapple, DNA methylation was found to play an important role in ethylene-induced flowering[26]. The recently completed genome of pineapple has and will continue to facilitate many areas of research in this species, and the present work focuses on the regulatory impact of DNA methylation on CAM pathway-related genes using temporal and spatial methylome and transcriptome analysis. A comparison of DNA methylation patterns and levels between green and white leaf diel time course allowed us to broadly investigate the epigenetic variation at CAM pathway-related genes. Combined with RNA-seq data, many CAM pathway related genes showed differences in temporal (diel time course) and spatial (green and white leaf) expression, and associated with differentially methylated regions, especially in the CHH context. Taken together, these results suggest a critical role of DNA methylation in the regulation of CAM pathway related genes, and also provide insights into the potential of epigenetic modifications for CAM engineering for the scientific and agronomic community.

Results

DNA methylation pathway genes are conserved in the pineapple genome

Much of the knowledge about DNA methylation in plants comes from studies in the model plant *Arabidopsis thaliana*. With the recent availability of the pineapple genome, homolog searches of DNA methylation pathway genes in pineapple are now possible. To identify pineapple genes homologous to *A. thaliana* DNA methylation pathway genes, we searched the annotated protein-coding genes of pineapple by using the BLAST and HMM algorithms. Based on protein similarity and domain conservation, we

found that most of the DNA methylation pathway genes in *A. thaliana* are conserved in pineapple, including MET1, CMT2, CMT3, and DRM2, (Table 1). Genes involved in the RNA-directed DNA methylation (RdDM) pathway were also found in pineapple, including AGO4, DCL3, and NRPE5. RNA-seq data were analyzed as an indicator of the functionality of these genes in pineapple, and most of the DNA methylation pathway genes analyzed were expressed at a relatively high level (mean FPKM = 10.6; median FPKM = 7.1; 10am green tip sample) in the RNA-seq data. The results therefore indicate that the DNA methylation pathway is conserved and functional in pineapple.

Table 1
Putative DNA methylation pathway genes in Pineapple

		Pineapple (*Ananas comosus)*			
		Name (Arabidopsis)	Length (a.a)*	Copy1	Expression level (FPKM)
MET1		VIM1,2,3,4,5,6	645	Aco016212.1	5.9109
		MET1,2a,2b,3	1534	Aco005236.1	2.4835
CMT3		SUVH4	624	Aco004654.1	9.24099
		CMT2	1295	Aco015994.1	4.27537
		CMT3	839	Aco013381.1	0.0487694
	Pol IV recruit	CLSY1/CLSY2	1256	Aco011099.1	1.94309
		SHH1/SHH2	258	Aco017233.1	41.2662
	Pol IV	NRPD1	1453	Aco015559.1	6.11536
	Pol IV + V	NRPD2/NRPE2	1172	Aco009438.1	20.9551
	Pol IV + V	NRPD4/NRPE4	205	Aco018391.1	114.279
	Pol V	NRPE1	1976	Aco023234.1	13.3714
	Pol V	NRPE5	222	Aco000628.1	13.0982
	Pol V	NRPE9B	114	Aco027798.1	0
	Pol V recruit	DRD1	888	Aco017413.1	8.92269
		DMS3	420	Aco010843.1	42.9601
		RDM1	163	Aco004257.1	7.207
		SUVH2/9	650	Aco026941.1	14.5377
RdDM		RDR2	1133	Aco015280.1	2.83345
		DCL1	1910	Aco016650.1	21.5963
		DCL2	1388	Aco016157.1	7.79403
		DCL3	1580	Aco009891.1	10.8606
		DCL4	1702	Aco008189.1	10.0258

For each gene family, the length of protein indicates the longest protein in this gene family, and only one ortholog was identified in pineapple genome.

		Pineapple (*Ananas comosus*)		
	HEN1	942	Aco008418.1	2.91452
	AGO4	924	Aco017860.1	42.6909
	KTF1	1493	Aco005272.1	24.8359
	IDN2	647	Aco002916.1	10.2605
	SUVR2	740	Aco006945.1	7.52871
	DMS4	346	Aco012515.1	9.07365
	UBP26	1067	Aco013188.1	14.6001
	DRM2	626	Aco007653.1	9.03314
	LDL1	844	Aco002795.1	14.1758
	LDL2	746	Aco009449.1	6.63836
	JMJ14	954	Aco010573.1	25.3867
Others	HDA6	471	Aco016535.1	18.7015
	RDR6	1196	Aco016551.1	3.15965
	MOM1	2001	Aco008982.1	24.8269
	MORC6	663	Aco020316.1	0.0358978
	DDM1	764	Aco027111.1	56.4119
	DME	1987	Aco018501.1	12.2632
	ROS1	1393	Aco018808.1	0.219213
	DML2	1332	Aco002923.1	5.45417

For each gene family, the length of protein indicates the longest protein in this gene family, and only one ortholog was identified in pineapple genome.

Single-base resolution map of DNA methylation in the pineapple genome

To explore diel DNA methylation patterns between the photosynthetic (green tip) and non-photosynthetic (white base) leaf tissues of field-grown pineapple, we collected samples at 4 am (ante meridiem), 10 am, 4 pm (post meridiem) and 10 pm across a 24-h period from green tip (photosynthetic) and white base (non-photosynthetic) of leaf tissue and performed whole-genome bisulfite sequencing (WGBS) (Fig. 1). After trimming adapter sequences and filtering low-quality reads, a total of ~ 930 million paired-end reads were generated from the green and white tissue samples across four time course (each with two biological replicates), resulting in ~ 730-fold coverage of the reference genome (Supplemental Table 1). Approximately 70% of all cytosines in the reference genome were covered by at least four reads of

different replicates (10am as an example, Supplemental Fig. 1). We also observed an average bisulfite conversion rate of unmethylated C to T of more than 99.5% from the chloroplast control, and the BS-seq data of the two biological replicates of each sample were highly correlated with each other (correlation coefficients > 0.95) (Supplemental Table 2). In summary, our data were reproducible and sufficient for further analysis.

We next investigated the genomic CG, CHG, and CHH methylation levels and found that the genome-wide average DNA methylation levels were very similar between green tip and white base at different times, except for CHH methylation (Fig. 1a). This finding is consistent with data from studies in rice and *Arabidopsis*, in which average DNA methylation of different developmental stages of the same plant is very similar, except for highly specialized tissues, such as the endosperm and the pollen vegetative nucleus[17, 27–29]. The distribution of methylcytosine (mC) along the chromosomes was calculated using 500-kb sliding windows with step 100 kb (Fig. 1b-d). Consistent with the findings in other plants, the global DNA methylation profiles for the CG, CHG, and CHH sequence contexts revealed heavily methylated pericentromeric regions (Fig. 1b-d). Unlike the CG and CHG sequence contexts, the CHH methylation of green tip at different times is significant higher than that in white base, especially in pericentromeric regions, where TEs were also enriched (Fig. 1d). We found that the mC distribution in pineapple was different than the previously reported distribution in rice and sorghum[30, 31], and there were also differences between the green and white pineapple leaf tissues (10am as an example, Supplemental Fig. 2). Specially, while pineapple green leaf tissue showed bimodal patterns for all sequence contexts, bimodal patterns were observed only for CG and CHG in pineapple white leaf tissue and in rice and sorghum. This difference suggests that non-CG methylation, specifically, methylation in the CHH context, was maintained more effectively in green tip leaf tissue than in white leaf tissue.

Photosynthetic and non-photosynthetic pineapple leaf tissues have different DNA methylation patterns in gene and TE regions

DNA methylation is similar between green tip and white base at different time course in CG and CHG sequence contexts, but very different in CHH context (Fig. 1). To explore the methylation patterns of different genomic structures, we examined the DNA methylation profiles of both gene and TE regions. Consistent to the analysis of genome-wide average DNA methylation, we found that CG DNA methylation is similar of green tip and white base at any time. This phenomenon is consistent in the gene and TE regions (Fig. 2a and 2d). Significantly, non-CG DNA methylation is very different of green tip and white base across different time, especially CHH DNA methylation (Fig. 2b-c and 2e-f). The CHH DNA methylation of green tip is significantly higher than that of white base, and its shows temporal rhythm changes (methylation is increased at 4 pm and 10 pm, but decreased at 4 am and 10 am), suggesting that it may be related to the photosynthesis of pineapple leaves, such as the CAM pathway.

To further investigate diel DNA methylation levels of gene and TE regions, we focused on gene and TE body regions rather than flanking regions. From the temporal DNA methylation data, we compared the DNA methylation levels of gene and TE regions in green and white leaf tissue at different time, and

observed diurnal methylation changes of both gene and TE regions. All contexts DNA methylation decreased in the early morning (10am) and increased at the afternoon (4 pm) of both green and white leaf tissues across different time (Fig. 2g-j), except CHG and CHH DNA methylation of TE regions in white leaf tissue (Fig. 2k-l). Consistent to the metaplot analysis of gene and TE regions, we found non-CG DNA methylation of green tip is higher than that of white base, especially CHH context. Collectively, our results showed that diel DNA methylation patterns and levels of both green tip and white base, and CHH methylation should play a critical role into this day-night cycling.

Association Between Dna Methylation And Gene Expression

It is well established that DNA methylation is associated with gene expression[16, 32]. Moderately expressed genes have more methylation than lowly or highly expressed genes, and promoter DNA methylation is usually negatively correlated with gene expression, except in some cases in which it promotes gene expression[18, 33–35]. To investigate DNA methylation regulation of gene expression in pineapple leaf tissues, we initially examined the CG, CHG, and CHH methylation levels of all expressed genes in green leaf tissues (take 10am as an example). Expressed genes were proportionally divided into five groups according to the level of gene expression: genes with the lowest expression composed the first group, and the most highly expressed genes composed the fifth group. Consistent with previous studies in other plant species, moderately expressed genes were the most highly methylated in gene body regions for the CG sequence context, i.e., the fourth group of genes was the most highly methylated in our study (Fig. 3a)[16, 18]. Non-CG methylation was highest in the first group of genes in both the gene body and flanking regions, and the second groups of genes also had higher non-CG methylation than the remaining groups of genes. These data suggest that non-CG methylation has a role in repressing gene expression (Fig. 3b-c).

We also observed that CG, CHG, and CHH methylation levels in promoter and downstream regions were highest in the most lowly expressed genes. To further examine the relationship between DNA methylation and gene expression, we analyzed the DNA methylation levels of three groups of genes (unexpressed, lowly expressed, and highly expressed genes) at different genic regions (promoter, gene body, and downstream regions). Compared to unexpressed genes, lowly and highly expressed genes had significantly lower CG, CHG, and CHH methylation at all three genic regions (Mann-Whitney test, P-value < 0.001) (Fig. 3d-f). Additionally, the methylation levels of lowly expressed genes were significantly higher than those of highly expressed genes at the different genic regions, except for CG methylation in the gene body and CHH methylation in the promoter region. A recent study in cassava also positively correlated gene body CG methylation with gene expression[18]. Thus, based on the present analysis and studies in other plant species, gene body CG methylation may be positively associated with gene expression[36]. We next focused on lowly and highly methylated genes, and examined the expression level differences between them. Similar to the results described above, lowly methylated genes were more highly expressed than highly methylated genes, and this was consistent for all three sequence contexts in the three genic regions except for promoter CHH methylation (Mann-Whitney test, P-value < 0.001) (Fig. 3g-i).

Together, these data suggest that the relationship between DNA methylation and gene expression depended on the genic regions, as well as the cytosine sequence context.

Diel Dna Methylation Regulation Of Pineapple Leaf

Genome-wide DNA methylation levels and patterns in green tip leaf tissue were similar to those in white base leaf tissue. To investigate diel DNA methylation patterns of photosynthetic leaf tissue, we firstly calculated average weighted DNA methylation levels for each sequence context using 100-bp windows[37] and identified regions with significantly different methylation levels between different diel time course of green leaf tissues (hereafter referred to as differentially methylated regions, DMRs). For continuous time comparison, we found that the number of differentially methylated regions of CHH type is the largest, followed by CHG type and CG type. This also implies that the role of CHH methylation in the regulation of circadian rhythm of green leaf tissue is more important than CG and CHG methylation (Fig. 4a). In addition, we found that CHH DNA methylation changes were greatest of 10am to 4 pm during the day, and showed hypermethylation at 4 pm compared with 10am. To further examine the association between DNA methylation changes across different diel time course in three sequence contexts (CG, CHG and CHH), we compared DMRs of different sequence contexts, and found that DMRs are not overlapped of three sequence contexts. For example, the hypermethylated regions of CHH at 4 pm may not necessarily be hypermethylated on CG and CHG sequence contexts (Fig. 4b); these CHH hypermethylated regions of 4 pm become hypomethylated regions or no methylation changes regions at other times. This phenomenon is also consistent in CG and CHG methylation (Fig. 4c). This finding is consistent with the DMRs comparison between green tip and white base or white base at different diel periods (Supplemental Fig. 3). Our analysis indicates that DNA methylation changes are dynamic of green tip across different periods, and probably play an important role in this diel cycle, especially CHH context.

We next examined the genomic characteristics of regions that changes of DNA methylation across diel time course. We tested the extent of overlap between DMRs within gene promoters (2 kb upstream of transcriptional start site), exons, introns, downstream regions of genes and intergenic regions. Although differential DNA methylation occurs in many locations, we found that most DMRs are enriched in intergenic regions. In addition, the DMR-enriched regions differ greatly between different comparisons of various time courses (Fig. 4d). For example, Aside from DMRs in intergenic regions, DMRs were most enriched in promoter regions of comparison between 4am to 10 am, while exons were most enriched between 4 pm and 10 pm. This result is consistent with the above comparative analysis of DMRs, which shows the dynamic changes of DNA methylation in the green tip during the diel time course.

Characterization Of Dmr-associated Differential Expression Genes

To investigate the impact of DNA methylation variance on gene expression differences of pineapple leaf tissues, we generated RNA-seq libraries for the same tissues used in the DNA methylation analysis and identified differentially expressed genes (DEGs)(Supplemental Table 3). We also compared our transcriptome data with the transcriptome data from previous studies, and found that our data is highly reproducible and consistent with each other (Supplemental Fig. 4)[6]. We first focused on differentially expressed genes across different diel time courses between green tip and white base. Approximately 12,191 DEGs were identified across different time stages, and the most differentially expressed genes were identified at 4am (8,623) and the least were at 10am (6,916) (Fig. 5a). For these DEGs, we found that a large number of DEGs are associated with differential methylation regions (DMRs). For example, there are 10,708 (87.8%, 10,708 out of 12,191) DEGs overlapped with DMRs, most of which are CHH-type DMRs (Fig. 5b). This result is consistent with green tip and white base (Supplemental Fig. 5 and Supplemental Fig. 6). The results demonstrate that a large number of genes are differentially expressed across diel time course and that diel DNA methylation changes are critical for this transcriptional dynamics of pineapple leaf.

We defined DMR-associated genes as those overlapping with a DMR in the 2,000 bp upstream or 2,000 bp downstream region or within the gene body. 5,778, 2,710, and 10,708 DMR-associated DEGs at different times were identified in green tip, white base and comparison between green tip and white base, respectively. To understand the critical roles of DMR-associated DEGs for pineapple leaf tissue of diel time periods. We performed Gene Ontology (GO) term enrichment analysis of DMR-associated DEGs from different periods comparison between green tip and white base. For DMR-associated DEGs in comparison between green tip and white base, we divided these genes into up-regulated genes in green tip or up-regulated genes in white base, and performed GO categories analysis separately. The results showed that the two groups of genes had very different functional categories. The DMR associated up-regulated DEGs in green tip were significantly enriched in the pathway of photosynthesis, transmembrane, and ion transport (Fig. 5c). On the contrary, DMR associated up-regulated DEGs in white base were enriched in protein phosphorylation, carbohydrate metabolic and microtubule-based process (Fig. 5d).

For DMR-associated DEGs in green tip across different periods, genes involved in transmembrane transport, photosynthesis, light harvesting, ion homeostasis processes were highly enriched (Supplemental Fig. 5). For DMR-associated DEGs in white base, GO functional analysis showed an enrichment of genes associated with photosynthesis, oxidation-reduction process, and carbohydrate metabolic process (Supplemental Fig. 6). Our data suggest that DNA methylation changes regulate gene expression difference between green and white leaf tissue and probably involved in regulating the source-sink relationship of the CAM photosynthesis process.

Cam Pathway Related Genes Regulated By Dna Methylation

Because the GO analysis of DMR-associated DEGs indicated an enrichment of genes with major regulatory roles in photosynthesis, light signaling, carbohydrate metabolic process and transmembrane

transport pathways (Fig. 5, Supplemental Fig. 5 and Supplemental Fig. 6), we next focused on genes that are known to play critical roles in the CAM pathway. Previous studies have identified that 38 genes are involved in the carbon fixation module of pineapple CAM process through homologous search and expression profile analysis[6]. We found that the majority of them are differentially expressed in green and white leaf tissues at different diel periods, and divided into two clusters. Genes of cluster 2 showed rhythmic expression in green tip but low expression in white base, and genes of cluster 1 showed that diel expression patterns in white base but arrhythmic expression in green tip (Fig. 6a). Most of these CAM genes are DMR-associated DEGs between green and white leaf tissues at different diel periods, including carbonic anhydrase (CA), phosphoenolpyruvate carboxylase (PEPC), malate dehydrogenase (MDH), phosphoenolpyruvate carboxylase kinase (PPCK) and pyruvate orthophosphate diakineses (PPDK) (Fig. 6b, Supplemental Table 4), majority of them also have been verified by our qRT-PCR validation (Supplemental Fig. 7 and Supplemental Table 5). CA genes are responsible for carbon dioxide fixation in CAM photosynthesis. Three beta-CA and two gamma-CA genes are highly expressed in photosynthetic green tip, but lowly expressed in non-photosynthetic white base. Previous studies have shown that only beta-CA genes could be major protein for carbon fixation rather than other CA genes. In present study, we identified a greater number of CHH type DMRs located across gene body and flanking regions of all three beta-CA genes. For example, we identified several DMRs located in the Aco006181.1 (beta-CA) gene body region, and these regions showed reduced DNA methylation in green tip at different diel periods compared to white base (Fig. 6c). We suspected that these DMRs should be related to the gene expression difference of beta-CAs between green tip and white base.

In addition, we also found that many genes involved in CAM-related pathway are regulated by DNA methylation, such as transporters, glycolysis and gluconeogenesis. Several genes were studied to be candidates for pineapple to hydrolyze vacuolar sucrose to hexose, and showed peak expression at early or late morning, such as Aco023030.1, Aco023036.1 and Aco017533.1. All these three members of vacuolar acid invertase gene family showed diel expression patterns and associated many DMRs between green tip and white base (Supplemental Table 7). Aco005379.1, which is a candidate for a vacuolar hexose exporter, and was previously shown highly expressed in green tissue, but did not show significant day-night cycling[11]. But we found this gene was highly expressed in green tip with peak expression at night (10 pm). A total of 56 DMRs located across gene body and flanking regions of Aco005379.1, and most of them are CHH type DMRs (Supplemental Fig. 8, Supplemental Table 6).

In pineapple genome, there are nine enzymes involved in glycolysis and gluconeogenesis. All of these nine enzymes shown higher expression in green tip than that in white base[11], and most of them are expressed rhythmically in green tip and arrhythmic expression in white base. But the regulatory mechanisms are still unclear. We found that all these nine enzymes were overlapped with many DMRs. For example, there are 30 DMRs across Aco024971.1 (triose-phosphate isomerase) gene body and flanking regions (Supplemental Table 7). Taken together, these results suggest that a large number of genes involved into CAM-related pathway are regulated by DNA methylation.

Discussion

Many of the presently available research on CAM pathways has focused on evolutionary and transcriptome analyses of CAM pathway-related genes[7, 11]. In this study, we generated temporal and spatial genome-wide single-base resolution DNA methylome maps and transcriptome profiles of pineapple leaf tissues, which greatly enhance and complement previous knowledge of CAM pathway studies. Comparing the methylation levels in photosynthetic green tip and non-photosynthetic white base leaf tissues revealed no significant global differences in CG and CHG methylation, but CHH methylation was significantly reduced in white base leaf tissue compared with green tip leaf tissue across different diel time course. However, there are obvious dynamic local DNA methylation changes during diel time course of green tip or white base of pineapple leaf. We hypothesized that dynamic DNA methylation changes in green tip or white base leaf tissue should be related to the CAM-related pathway. Through sampling diel DNA methylation patterns in both green and white pineapple at different diel time course, we were able to identified differential methylation changes related to the CAM photosynthetic pathway. By combining DNA methylation data and transcriptome data, we could identified a large number of DMR-associated DEGs, which are often enriched in several important biological pathways of CAM cycle, such as photosynthesis, lighting harvest, carbohydrate metabolism, transporter and protein phosphorylation. There are three CA gene families annotated in pineapple genome (alpha-CA, beta-CA, and gamma-CA), but only beta-CAs were expressed highly in green tip, and showed diel expression patterns[6]. We found all three beta-CAs in pineapple genome showed different expression between green tip and white base tissues, and were associated with many DMRs, especially CHH contexts.

In addition, many genes involved in hexose transporter, glycolysis and gluconeogenesis were also associated with different methylation divergence, such as Aco005379.1, Aco023036.1, Aco005368.1 and Aco024987.1. Gene duplication and expansion was initially proposed to be the driven force for the evolution of the CAM pathway[38]. However, it is still controversial, many studies have shown that CAM pathway should be evolved by differential expression of CAM-related genes or neofunctionalization rather than gene dosage[7, 10, 11].

Conclusions

DNA methylation has been long recognized as a mechanism for gene expression regulation, repetitive element silencing, and plays a critical role for plant development and stress response. Here we showed that spatial and temporal DNA methylation and transcriptome changes during pineapple leaf CAM cycle. Our results strongly suggest that the transcription regulation of many key CAM-regulated genes involves DNA methylation and provide epigenetic insights for engineering of CAM in other crop improvement.

Methods

Plant material, library construction and sequencing

In this study, plant material of *Ananas comosus* var. *comosus* cultivar MD-2 was acquired from Guangxi Academy of Agricultural Sciences (GAAS), China (22.84°N, 108.48°E). Identification of the plant materials was made by the GAAS and original plant was acquired from Del Monte company (www.delmonte.com). The voucher specimen was deposited at the herbarium of the College of Plant Protection, Fujian Agriculture and Forestry University, China. DNA was extracted from 'D' leaf of pineapple (*Ananas comosus* var. *comosus* cultivar MD-2), and BS-seq libraries were prepared using the TruSeq Nano DNA LT kit (Illumina), as described previously[32]. For each tissue type, two libraries corresponding to two biological replicates were sequenced on a HiSeq X Ten system (Illumina) to obtain paired-end 150-bp reads per the manufacturer's instructions.

Total RNA was extracted from the same tissue used for the BS-seq libraries, and RNA-seq libraries were prepared using the TruSeq Preparation Kit with polyA mRNA selection, per the manufacturer's instructions (Illumina). Three libraries (only two libraries for 10am sample) were pooled and sequenced to obtain paired-end 150-bp reads on the Illumina HiSeq X Ten system.

Bs-seq Data Analysis

For each biological replicate of BS-seq data, bisulfite-converted reads were aligned to the pineapple reference genome using BSMAP v2.90[39]. Four mismatches were allowed per 100-bp read length, and only uniquely mapped reads were kept. The conversion rate was calculated by one minus the global average methylation level of the chloroplast genome. For methylcytosine, only coverage more than three reads was used to calculate the weighted methylation level. Correlation between replicates was calculated using the average DNA methylation level of non-overlapping 100-bp windows, and the *cor()* function in R. Metaplots for both gene and TE regions were generated using the average DNA methylation level of three separate regions: the upstream region (100-bp bin size), gene body (20 proportional bins), and downstream region (100-bp bin size). Differentially methylated regions (DMRs) were determined by DMRcaller, which is a versatile R/Bioconductor[40].

Rna-seq Data Analysis

RNA-seq reads were aligned to the pineapple reference genome by HISAT2 v2.1.0[41] with default parameters, and only uniquely mapped reads were kept. Expression value was quantified by StringTie v1.3.3b[41] as FPKM (fragments per kilobase per million mapped reads). Differentially expressed genes (DEGs) were identified by DESeq2 v1.22.2 with default parameters[42]. RNA-seq data of Ray M. et al (2015) were downloaded from NCBI BioProject PRJNA305042, and only data at 4am, 10am, 4 pm and 10 pm time points were considered for comparisons. The same RNA-seq analysis pipeline in present work was performed on these public dataset, genes with FPKM > 0.5 were considered as expressed, and used for heatmap and Pearson correlation coefficient analysis.

Qrt-pcr Validation

The fully expanded D leaf was collected from the pineapple that has at least 35 mature leaves. Green tissue at the leaf tip and white tissue at the leaf base were collected and quickly froze in liquid nitrogen at 10:00 am as Fig. 1a. The sample were stored in -80 °C refrigerator until total RNA extraction. The total RNA was extracted following manufacturer's protocol using RNA extraction kit (Omega Bio-Tek, Shanghai, China). Total RNA was diluted with nuclease-free water. 1 μg of purified total RNA was reverse transcribed to cDNA in a 25 μl reaction volume using M-MLV reverse transcriptase (Promega) according to the supplier's instructions. The qRT-PCR was performed with Quantitative kit (TRANS, Beijing, China) as described by Zhang with the program: 95 °C for 30 s; 40 cycles of 95 °C for 5 s and 60 °C for 30 s; 95 °C for 15 s. Three technical replicates and at least three independent biological replicates were performed in each condition. The specific primers used in the study are given in Supplemental Table 5.

Go Enrichment Analysis

GO enrichment of DMR-associated genes was performed by using the OmicShare tool, free online platform for data analysis (www.omicshare.com/tools) with hypergeometric test. Only GO terms with P-value less than 0.01 were used for further analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors approved the manuscript.

Availability of data and materials

The data reported in the present paper were deposited into the Gene Expression Omnibus (GEO) database (accession number GSE120401).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (No. U1605212 to Y.Q., and No. 31701874 to X.Z.), the Natural Science Foundation of Fujian Province (No. 2018J06006) to H.W.

Program for Excellent Youth Talents and Program for New Century Excellent Talents in Fujian Province University, and Pre-eminent Youth Fund and Distinguished Young Scholars in Fujian Province to H.W.

Author Contributions

H.W. conceived this project and designed the research; H.W. and Y.Q. designed the experiments; Y.S. and H.W. conducted methylation analysis; Y.S., X.Z., X.C. and M.Y. conducted transcriptome and DNA methylome analysis; Y.S. and H.Z. performed CAM analysis; M.Y. and X.C. collected the tissues. H.W. wrote the paper. All authors have read and approved the manuscript.

Acknowledgements

We thank laboratory members and colleagues for comments and discussions.

References

1. Wang WS, Pan YJ, Zhao XQ, Dwivedi D, Zhu LH, Ali J, Fu BY, Li ZK. Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). J Exp Bot. 2011;62(6):1951–60.
2. Yong-Villalobos L, Gonzalez-Morales SI, Wrobel K, Gutierrez-Alanis D, Cervantes-Perez SA, Hayano-Kanashiro C, Oropeza-Aburto A, Cruz-Ramirez A, Martinez O, Herrera-Estrella L. Methylome analysis reveals an important role for epigenetic changes in the regulation of the Arabidopsis response to phosphate starvation. Proc Natl Acad Sci USA. 2015;112(52):E7293–302.
3. Coveca. **Comision veracruzana de comercializacion agropecuaria**. *Gobierno del Estado de Veracruz, México* 2002.
4. Statista. **https://www.statista.com/statistics/298505/global-pineapple-production/**. **2016**.
5. Cushman JC. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. Plant physiology. 2001;127(4):1439–48.
6. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, et al. The pineapple genome and the evolution of CAM photosynthesis. Nat Genet. 2015;47(12):1435–42.
7. Zhang L, Chen F, Zhang GQ, Zhang YQ, Niu S, Xiong JS, Lin Z, Cheng ZM, Liu ZJ. Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. The Plant journal: for cell molecular biology. 2016;86(2):175–85.
8. Chomthong M, Griffiths H. Model approaches to advance crassulacean acid metabolism system integration. The Plant journal: for cell molecular biology. 2020;101(4):951–63.
9. Bai Y, Dai X, Li Y, Wang L, Li W, Liu Y, Cheng Y, Qin Y. Identification and characterization of pineapple leaf lncRNAs in crassulacean acid metabolism (CAM) photosynthesis pathway. Scientific reports. 2019;9(1):6658.

10. Chen LY, Xin Y, Wai CM, Liu J, Ming R. The role of cis-elements in the evolution of crassulacean acid metabolism photosynthesis. Horticulture research. 2020;7:5.
11. Wai CM, VanBuren R, Zhang J, Huang L, Miao W, Edger PP, Yim WC, Priest HD, Meyers BC, Mockler T, et al. Temporal and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple. The Plant journal: for cell molecular biology. 2017;92(1):19–30.
12. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics. 2010;11(3):204–20.
13. Chan SW, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature reviews Genetics. 2005;6(5):351–60.
14. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–9.
15. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci USA. 2010;107(19):8689–94.
16. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. 2006;126(6):1189–201.
17. Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D. Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci USA. 2010;107(43):18729–34.
18. Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc Natl Acad Sci USA. 2015;112(44):13729–34.
19. Lin JY, Le BH, Chen M, Henry KF, Hur J, Hsieh TF, Chen PY, Pelletier JM, Pellegrini M, Fischer RL, et al. Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proc Natl Acad Sci USA. 2017;114(45):E9730–9.
20. Lauss K, Wardenaar R, Oka R, van Hulten MHA, Guryev V, Keurentjes JJB, Stam M, Johannes F. Parental DNA Methylation States Are Associated with Heterosis in Epigenetic Hybrids. Plant physiology. 2018;176(2):1627–45.
21. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. **Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements**. *eLife* 2015, 4.
22. Xu R, Wang Y, Zheng H, Lu W, Wu C, Huang J, Yan K, Yang G, Zheng C. Salt-induced transcription factor MYB74 is regulated by the RNA-directed DNA methylation pathway in Arabidopsis. J Exp Bot. 2015;66(19):5997–6008.
23. Lu YC, Feng SJ, Zhang JJ, Luo F, Zhang S, Yang H. Genome-wide identification of DNA methylation provides insights into the association of gene expression in rice exposed to pesticide atrazine.

Scientific reports. 2016;6:18985.

24. Liang D, Zhang Z, Wu H, Huang C, Shuai P, Ye CY, Tang S, Wang Y, Yang L, Wang J, et al. Single-base-resolution methylomes of Populus trichocarpa reveal the association between DNA methylation and drought stress. BMC Genet. 2014;15(Suppl 1):9.
25. Garg R, Narayana Chevala V, Shankar R, Jain M. Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. Scientific reports. 2015;5:14922.
26. Wang J, Li Z, Lei M, Fu Y, Zhao J, Ao M, Xu L. Integrated DNA methylome and transcriptome analysis reveals the ethylene-induced flowering pathway genes in pineapple. Scientific reports. 2017;7(1):17167.
27. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. PLoS Genet. 2014;10(11):e1004785.
28. Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, Feng X. Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat Genet. 2018;50(1):130–7.
29. Kawakatsu T, Nery JR, Castanon R, Ecker JR. Dynamic DNA methylation reconfiguration during seed development and germination. Genome biology. 2017;18(1):171.
30. Li X, Zhu J, Hu F, Ge S, Ye M, Xiang H, Zhang G, Zheng X, Zhang H, Zhang S, et al. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. BMC Genomics. 2012;13:300.
31. Turco GM, Kajala K, Kunde-Ramamoorthy G, Ngan CY, Olson A, Deshphande S, Tolkunov D, Waring B, Stelpflug S, Klein P, et al. DNA methylation and gene expression regulation associated with vascularization in Sorghum bicolor. New Phytol. 2017;214(3):1213–29.
32. Wang L, Xie J, Hu J, Lan B, You C, Li F, Wang Z, Wang H. Comparative epigenomics reveals evolution of duplicated genes in potato and tomato. The Plant journal: for cell molecular biology. 2018;93(3):460–71.
33. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. Nature reviews Molecular cell biology. 2018;19(8):489–506.
34. Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu JK. Regulatory link between DNA methylation and active demethylation in Arabidopsis. Proc Natl Acad Sci USA. 2015;112(11):3553–7.
35. Lang Z, Wang Y, Tang K, Tang D, Datsenka T, Cheng J, Zhang Y, Handa AK, Zhu JK. Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. Proc Natl Acad Sci USA. 2017;114(22):E4511–9.
36. Takuno S, Seymour DK, Gaut BS. The Evolutionary Dynamics of Orthologs That Shift in Gene Body Methylation between Arabidopsis Species. Molecular biology evolution. 2017;34(6):1479–91.
37. Schultz MD, Schmitz RJ, Ecker JR. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. Trends Genet. 2012;28(12):583–5.

38. Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, Chen LJ, He Y, Xu Q, Bian C, et al. The genome sequence of the orchid Phalaenopsis equestris. Nat Genet. 2015;47(1):65–72.
39. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinform. 2009;10:232.
40. Catoni M, Tsang JM, Greco AP, Zabet NR. DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. Nucleic acids research. 2018;46(19):e114.
41. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature protocols. 2016;11(9):1650–67.
42. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome biology* 2014, **15**(12):550.

Figures

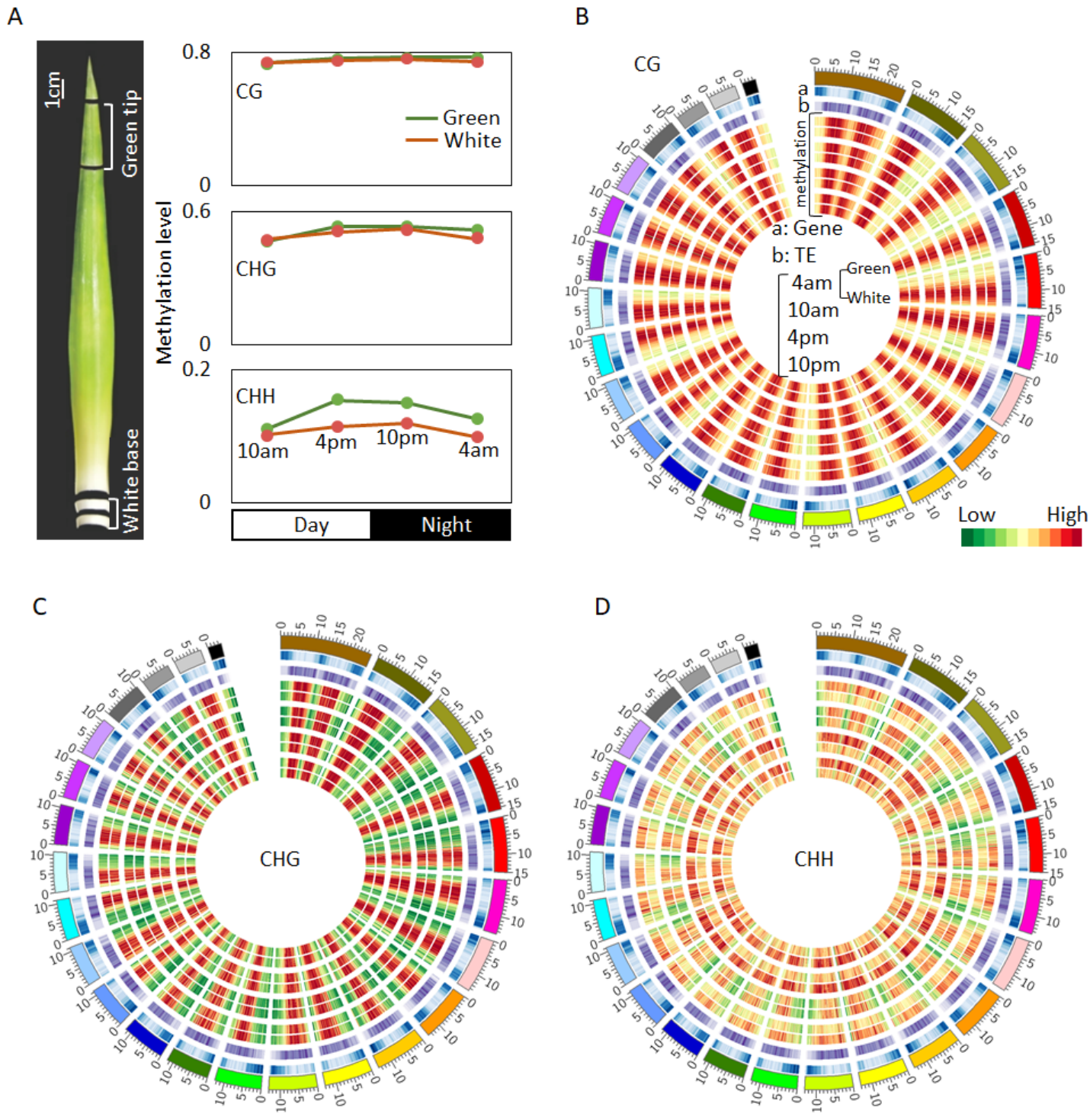

Figure 1

DNA methylation landscape of pineapple leaf. (a) Pineapple leaf tissue used to survey DNA methylation of CAM photosynthesis-related genes, and genome-wide average DNA methylation comparison between green and white leaf tissue across diel time course. Circos plot of gene density, TE density, and methylation levels of CG (b), CHG (c), and CHH (d) at different time periods between green and white leaf tissue across 25 chromosomes.

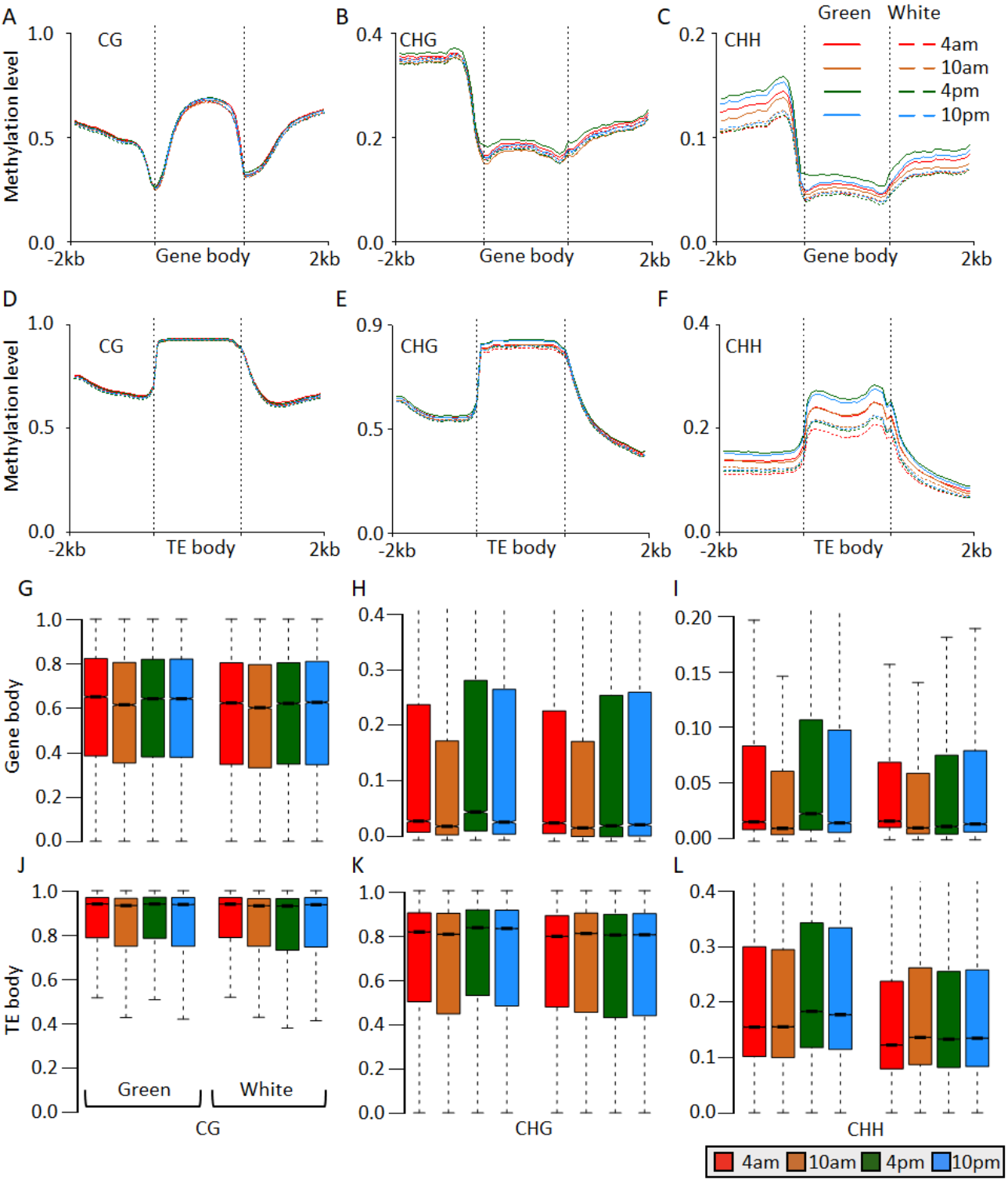

Figure 2

DNA methylation patterns of genes and transposable element regions of pineapple leaf. (a-c) Boxplot of CG, CHG, and CHH methylation of gene between green tip and white base across diel time course. (d-f) Boxplot of CG, CHG, and CHH methylation of transposable element regions between green tip and white base across diel time course. (g-i) DNA methylation patterns across gene regions of green tip and white base across different diel time course. (j-l) DNA methylation patterns across TEs of green tip and white

base across different diel time course. '-2kb' denotes the 2kb region upstream of transcription start site or TE body; '2kb' denotes the 2kb downstream of the transcription end site or TE body.

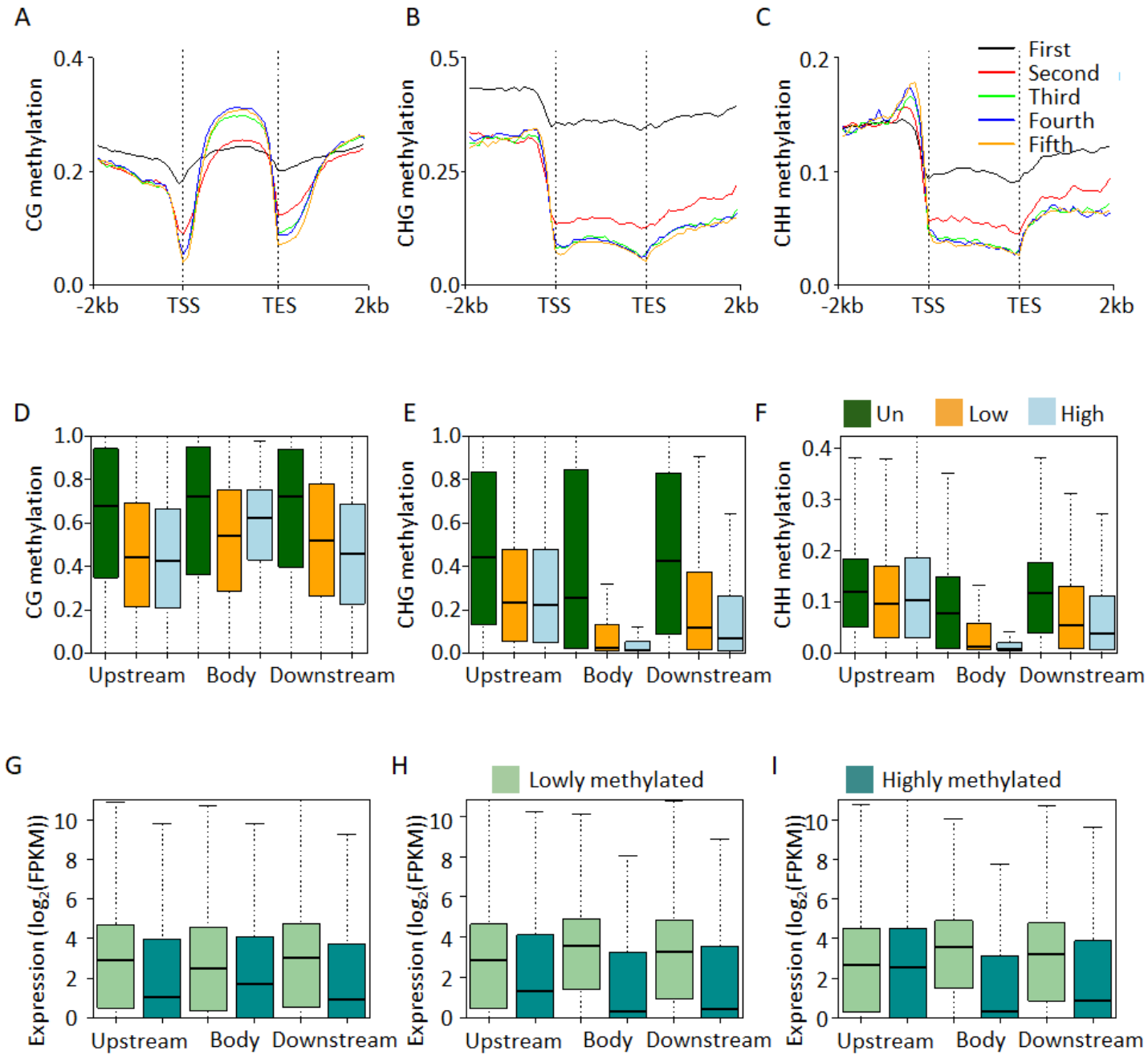

Figure 3

Association between gene expression and methylation at different genic regions. (a-c) Association between expression and DNA methylation for the CG (a), CHG (b), and CHH (c) sequence contexts. Genes were divided proportionally into five groups based on their expression level, such that the fifth group of genes contained the most highly expressed genes. The 2-kb were calculated for each bin. (d-f) Comparison of the methylation levels of unexpressed, lowly, and highly expressed genes for the region, gene body and 2-kb downstream region) are shown. (g-i) Comparison of the expression levels of lowly

and highly methylated genes for the CG (g), CHG (h), and CHH (i) sequence contexts in different genic regions. Unexpressed genes were defined as genes with FPKM 0, and highly and lowly expressed genes were defined as the one-third of genes with the highest and lowest expression levels, respectively.

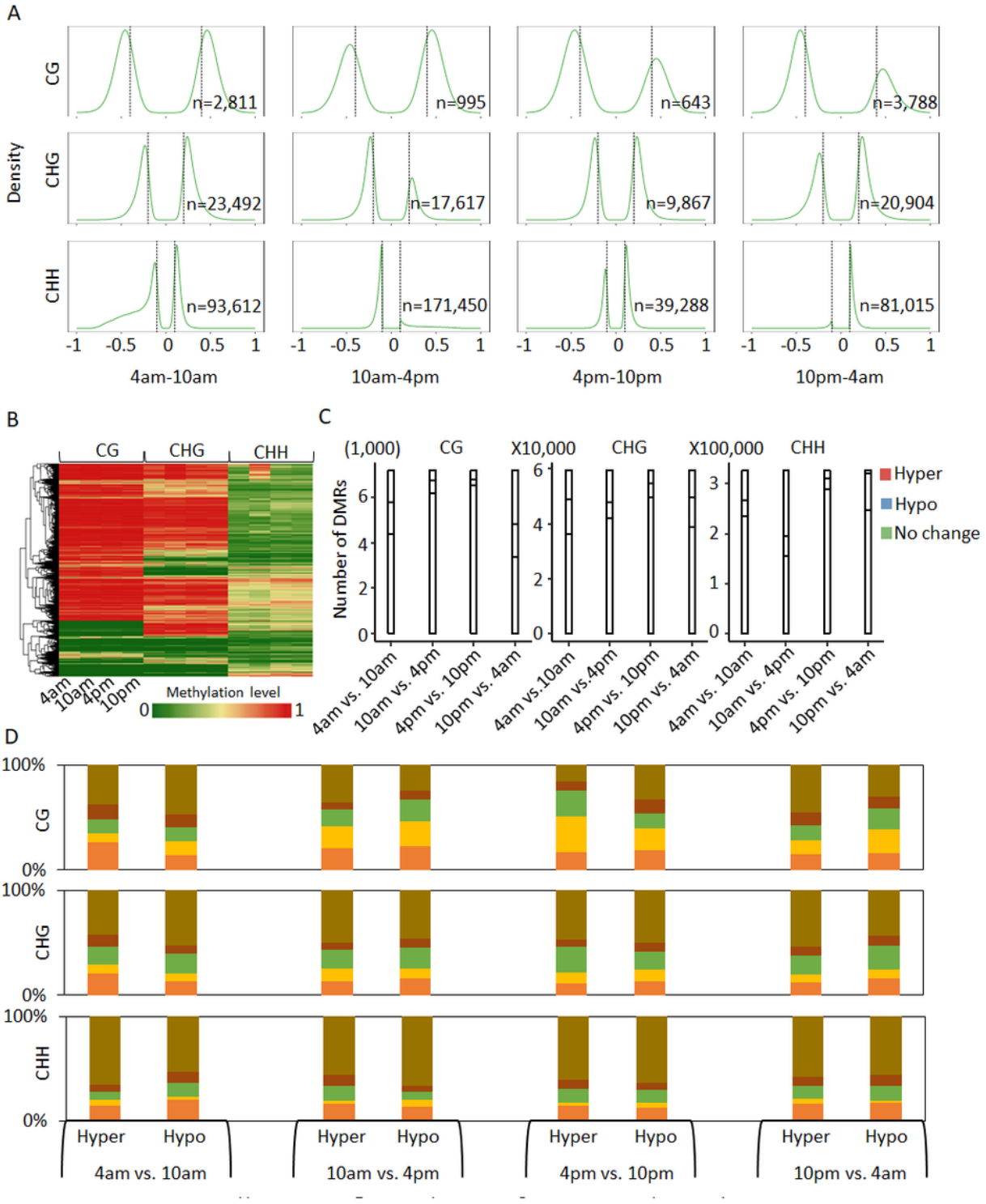

Figure 4

DNA methylation variance of green tip at different diel periods. (a) Density of differential methylation regions of CG, CHG, and CHH contexts across diel time course of green tip leaf tissue. (b) Heat map of

DNA methylation levels of all DMRs (CG+CHG+CHH) of green tip leaf tissue. (c) Sankey plot of DMR dynamics across different periods of green tip of CG, CHG, and CHH contexts. (d) DMR genomic distribution of green tip.

A

Green vs. White	DEG	DMR association	DEG Ratio
4am	8,623	6,744	0.78
10am	6,916	4,791	0.69
4pm	8,151	6,983	0.86
10pm	8,583	7,299	0.85
Total	12,191	10,708	0.88

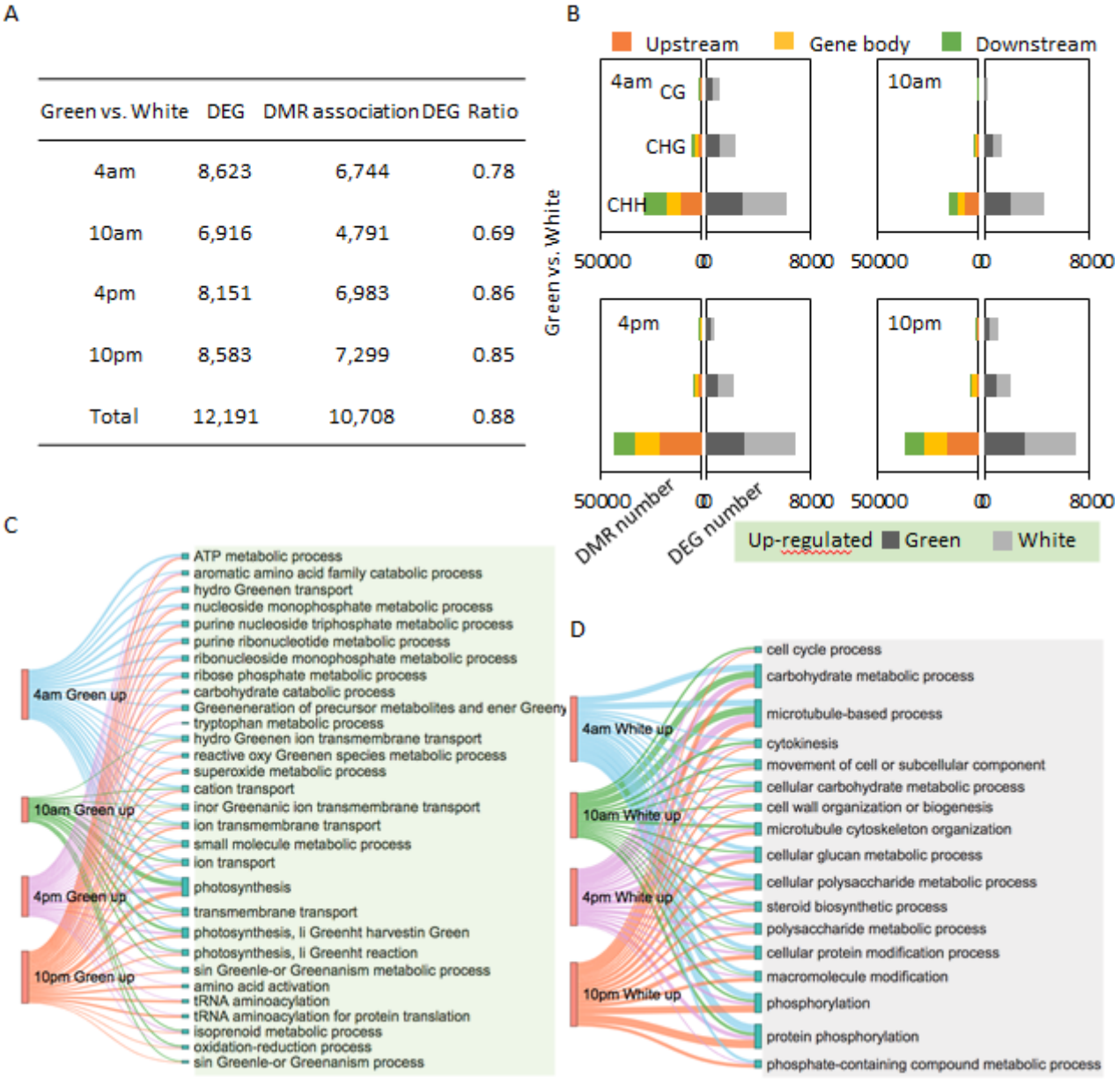

Figure 5

Differential methylation regions associated DEGs analysis. (a) Summary of DEGs and DMR-associated DEGs between green tip and white base at different diel periods. (b) The number of DMR-associated DEGs. DMRs were divided into up-/down-stream and gene body regions of CG, CHG and CHH contexts. (c) Enriched GO terms of DMR-associated up-regulated DEGs in green tip of different time course. (d) Enriched GO terms of DMR-associated up-regulated DEGs in white base of different time course.

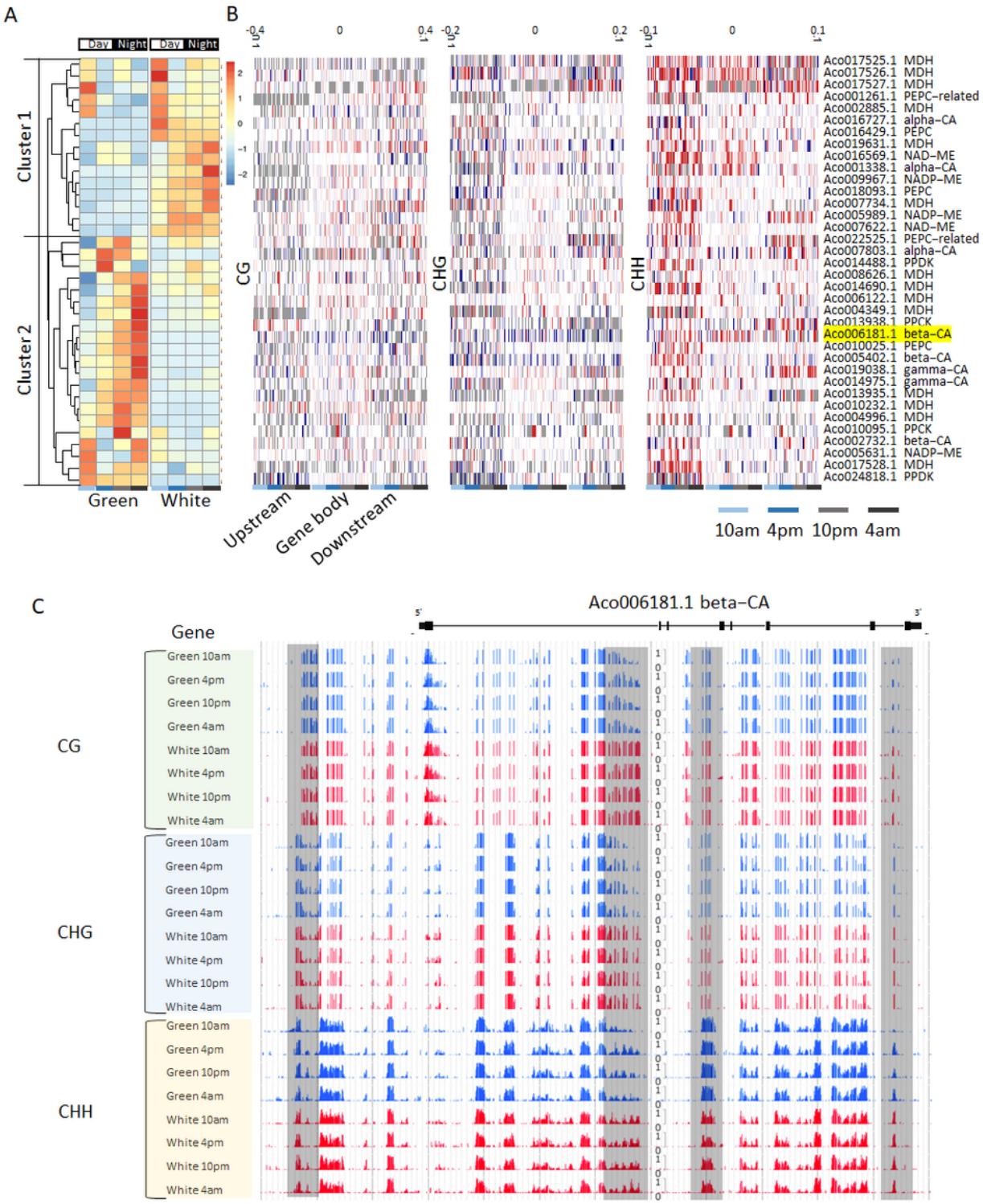

Figure 6

Diel expression patterns and DNA methylation of CAM related genes. (a) Expression patterns of CAM genes for the diel time course in green tip and white base. (b) DNA methylation divergence across CAM related genes between green tip and white base at different time course. Upstream, gene body and downstream regions were divided into 20bins, and methylation levels of each bin were calculated.

Methylation divergence between green tip and white base was calculated by Green-White. (c) Genome browser snapshot of DNA methylation changes of Aco006181.1 (beta-CA).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalTable7.xlsx
- SupplementalTable6.xlsx
- SupplementalTable4.xlsx
- SupplementalTable135SupplementalFig.18.docx